	٠
1.3 Kelease n	
	tes
	ers
	ch administrator
1.5.1	Adding a new language
1.5.1	Configuration modification
	Crawling Files with ManifoldCF
	Solr configuration customization
	i Open local files
	openSearch
	Security Management
	B Extracting file content in ManifoldCF and apply OCR
	Manage Solr configuration with Zookeeper
	ch experts
	Alerts management - Mail Configuration
	? Analytics
	5.2.2.1 Dynamic analytics
1	5.2.2.2 ELK analytics
1	5.2.2.3 Search navigation analytics
1.5.2	Autocomplete configuration
1.5.2	Connectors configuration
	i Crawling
	5.2.5.1 Add a JDBC connector (MySQL, Oracle)
	5.2.5.2 Add the JCIFS Connector to Datafari
	5.2.5.3 Reset a MCF job
	Deduplication management
	' Facets management
	Favorites management
	) Fieldweights
1.5.2	0 Highlighting configuration
1.5.2	1 Likes management
	2 Promolinks management
	3 Query Elevator
	5.2.13.1 Admin UI feature
	5.2.13.2 Search View Feature
	4 Stopwords management
	5 Synonyms
	Alerts
	Autocomplete
	Beduplication
	Facets
	Favorites
1.5.3	i Highlighting
1.5.3	' Likes
1.5.3	Promolinks
1.0.0	) Searching
	0 Stopwords
1.5.3	1 Alerts (from Datafari 3.1)
1.5.3 1.5.3	
1.5.3 1.5.3 1.5.3	)ers
1.5.3 1.5.3 1.5.3 1.6 For develo	Ders
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajax	rancelabs graphical framework
1.5.3 1.5.3 1.6 For develo 1.6.1 Ajax 1.6.1	rancelabs graphical framework
1.5.3 1.5.3 1.6 For develo 1.6.1 Ajao 1.6.1 1.6.1	rancelabs graphical framework Displaying thumbnails for autocomplete and results list
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajax 1.6.1 1.6.1 1.6.1	rancelabs graphical framework Displaying thumbnails for autocomplete and results list
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajav 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1	rancelabs graphical framework Displaying thumbnails for autocomplete and results list Pramework architecture and mechanism Quickly changing the main colors
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajav 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1	rancelabs graphical framework Displaying thumbnails for autocomplete and results list Framework architecture and mechanism Quickly changing the main colors Tutorial
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajao 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add	rancelabs graphical framework Displaying thumbnails for autocomplete and results list Framework architecture and mechanism Quickly changing the main colors Tutorial Widgets and Modules custom fields in Datafari and display them
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajao 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add	rancelabs graphical framework Displaying thumbnails for autocomplete and results list Framework architecture and mechanism Quickly changing the main colors Tutorial
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajao 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler	rancelabs graphical framework Displaying thumbnails for autocomplete and results list Framework architecture and mechanism Quickly changing the main colors Tutorial Widgets and Modules custom fields in Datafari and display them
1.5.3 1.5.3 1.6 For develo 1.6.1 Ajax 1.6 For develo 1.6.1 Ajax 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler 1.6.4 Arcl	rancelabs graphical framework Displaying thumbnails for autocomplete and results list Framework architecture and mechanism Quickly changing the main colors Tutorial Widgets and Modules custom fields in Datafari and display them
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajay 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler 1.6.4 Arcl 1.6.5 Cus	rancelabs graphical framework Displaying thumbnails for autocomplete and results list Framework architecture and mechanism Quickly changing the main colors Tutorial Widgets and Modules custom fields in Datafari and display them s - Architecture and mechanism tecture
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Aja) 1.6.1 Aja) 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler 1.6.4 Arcl 1.6.5 Cus 1.6.6 Dec	rancelabs graphical framework Displaying thumbnails for autocomplete and results list Pramework architecture and mechanism Quickly changing the main colors Tutorial Widgets and Modules Sustom fields in Datafari and display them S - Architecture and mechanism tecture
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Aja) 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler 1.6.4 Arcl 1.6.5 Cus 1.6.6 Dec 1.6.7 ELM	rancelabs graphical framework Displaying thumbnails for autocomplete and results list Framework architecture and mechanism Quickly changing the main colors Tutorial Widgets and Modules Sustom fields in Datafari and display them s - Architecture and mechanism tecture
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajax 1.6.1 Ajax 1.6.1 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler 1.6.4 Arcl 1.6.5 Cus 1.6.6 Dec 1.6.7 ELK 1.6.8 Fac	rancelabs graphical framework Displaying thumbnails for autocomplete and results list
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajax 1.6.1 Ajax 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler 1.6.4 Arcl 1.6.5 Cus 1.6.6 Dec 1.6.7 ELK 1.6.8 Fac 1.6.8	rancelabs graphical framework Displaying thumbnails for autocomplete and results list
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajax 1.6.1 Ajax 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler 1.6.4 Arcl 1.6.5 Cus 1.6.6 Dec 1.6.7 ELk 1.6.8 Fac 1.6.8 Fac 1.6.9 Fiel	rancelabs graphical framework Displaying thumbnails for autocomplete and results list
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajax 1.6.1 Ajax 1.6.1 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler 1.6.4 Arcl 1.6.5 Cus 1.6.6 Dec 1.6.7 ELk 1.6.8 Fac 1.6.8 Fiel 1.6.9 Fiel 1.6.10 Lik	rancelabs graphical framework Displaying thumbnails for autocomplete and results list
1.5.3 1.5.3 1.5.3 1.6 For develo 1.6.1 Ajax 1.6.1 Ajax 1.6.1 Ajax 1.6.1 1.6.1 1.6.1 1.6.2 Add 1.6.3 Aler 1.6.4 Arcl 1.6.5 Cus 1.6.6 Dec 1.6.7 ELk 1.6.8 Fac 1.6.8 Fac 1.6.9 Fiel 1.6.10 Lik 1.6.11 Se	rancelabs graphical framework Displaying thumbnails for autocomplete and results list

1.6.12 Semantics	140
1.6.12.1 Link an ontology	140
1.6.12.2 Promolinks - architecture and mechanism	142
1.6.12.3 Synonyms and stopwords 1	143
1.6.13 Setup development environnement with Eclipse (Linux)	144
1.6.14 Setup development environment with Eclipse (Linux) for Datafari 2.0	150
1.6.15 Setup development environment with Eclipse (Linux) for Datafari 2.x - Maven	
1.6.16 Unit Testing	
1.6.17 Utility logs	175
1.6.17.1 Core Monitoring logs 1	176
1.6.17.2 Statistics logs	177
1.6.18 Widgets and CSSs good practices 1	178
1.6.19 Add new icon for search results 1	178
1.6.20 Setup development environment with Eclipse (Linux) for Datafari 3.x - Maven	179
1.6.21 Save search - technical doc (enterprise version) 1	196
1.6.22 Logging	197
1.6.22.1 ELK logs exploitation1	
1.7 Articles de conseil	199

# Datafari

#### Datafari v2.2.1 is available !

We released in April 2016 version 2.2.1 of Datafari. Loads of changes, and we are happy to share it with you. Available for Linux, enjoy !! We released a Docker version as well. A windows version will soon be also available, but we didn't test all the functionalities on it.

#### Welcome to the Datafari documentation!

This is the home page for the Datafari documentation space within Confluence. This serves as a central repository for anything that relates to Datafari, be it for users or for developers.

Datafari is the ideal product for those who want to search through their data, while using advanced open source technologies. Datafari combines Apache ManifoldCF and Apache Solr and ELK, to allow its users to analyse and search through many different and diverse data sources: their file shares, their cloud shares (dropbox, google drive...), their databases, but also their emails and many more. Available as community or enterprise version, Datafari has its specificities:

- Its open source licence is not aggressive as it's under Apache v2 licence: you are free to do whatever you want with it, you just need to mention that you're using it.
- It combines three popular Apache projects: Solr for the indexing and searching, ManifoldCF for the crawling, and Cassandra for the user management. The usage of these products guarantees stability over time.
- It provides analytics capabilities, through the integration of ELK.

### Search this documentation



## Recently Updated Pages

- Unit Testing il y a 4 minutes • mis à jour par Cédric Ulmer • afficher les modifications
- Setup development environment with Eclipse (Linux) for Datafari 2.x - Enterprise

il y a 5 minutes • mis à jour par Cédric Ulmer • afficher les modifications

Setup development environment with Eclipse (Linux) for Datafari 2.0 il y a 8 minutes • mis à jour

par Cédric Ulmer • afficher les modifications

Setup development environnement with Eclipse (Linux) il y a 16 minutes • mis à jour

par Cédric Ulmer • afficher les modifications

Search View Feature hier, à 6:49 PM • mis à jour par Olivier Tavard • afficher les modifications

## Foreword

Datafari is the ideal product for those who want to analyse and search through their data, while using advanced open source technologies. Datafari combines Apache ManifoldCF and Apache Solr, to allow its users to search through many different and diverse data sources: their file shares, their cloud shares (dropbox, google drive...), their databases, but also their emails and many more. Available as community or enterprise version, Datafari has its specificities:

- Its open source licence is non aggressive as its under Apache v2 licence: you are free to do whatever you want with it, you just need to
  mention that you're using it.
- It combines several popular Apache projects: Solr, ManifoldCF, and Cassandra, which guarantees stability over time.
- It provides analytics capabilities, through the integration of Banana and ELK.

## Introduction

Be it for users or for companies, the amount of data is increasing exponentially. On top of it, the challenge of the cloud multiplies the number and heterogeneity of systems hosting data. Search engines are here to tackle this challenge. They can connect to many systems, propose a single view on the entirety of available data. Contrarily to web search engines, enterprise search engines are 100% controlled by you, and have access to data which belong to you, and which are not necessarily public. These search engines guarantee the security of access to these data.

Among the many existing solutions, a majority is proprietary : you acquire a licence, you pay for support, and you buy the integration. However, a large chunk on the basic functionalities are now available as open source, and don't require massive investments. The big players of the web have made this choice: Linkedin, eBay, Twitter, Salesforce, Kelkoo... They all use open source tools.

However, the most well knowned tools, Apache Lucene and Apache Solr, are only the heart of a search solution. They do not provide any framework to manage the access to the data sources, they do not handle security, and they do not manage backup or monitoring activities. Other complementary open source projects are available, but the integration is not always easy. This is where Datafari is standing: it integrates these technologies, using as much as possible projects using an Apache licence (or equivalent), in order to remain non aggressive for companies.

We wanted to offer to the community an easy to use tool, affordable – even free – for many use cases, but also able to scale up in order to manage hundreds of millions of indexed documents, thanks to SolrCloud. You will find in this document an overview of Datafari, in order to better understand it, use it, even extend it. Obviously, we encourage the users community to help us in the evolution of this open source tool. In the architecture section, you will discover the architecture of Datafari and its main components. In the crawling section, we will detail the crawling section and its usage. In the indexing section, we present the content indexing part of the search engine. In the search section, we present the default user interface used by Datafari. In the security section, we cover the security challenges. In the analytics section, we will learn how to monitor Datafari. In the use cases section, we present use cases which you can use as a way to kickstart you own projects.

#### User documentation

In this user documentation section, you will learn all you need about Datafari. Whether you are the search manager, a standard or the Datafari system administrator, we cover in this section the functionnalities of Datafari.

#### **Developer documentation**

In this developer documentation section, you will learn how to set up a proper development environment, as well as the test environment into place. For now, only the development environment is documented.

## Release notes

Target release	2.0
Document status	RELEASE
Document owner	Olivier Tavard
Designer	
Developers	Olivier Tavard, Aurélien Mazoyer
QA	Olivier Tavard

#### Major changes from v1.0

- · Ajaxfrancelabs has been completly revamped:
  - New responsive design
  - New widgets
- · Apache Cassandra is integrated to handle user management
- Migration to Solr 5
- The admin UI has been completly revamped:
  - Harmonisation of the different components, leveraging the great Devoops V2 admin theme

- New stable functionnalities:
  - User management
  - Connection to AD through LDAP
  - Promolinks
  - Synonyms
  - Size limitation
  - Autocomplete configuration
- New experimental functionalities:
  - Likes and favorites
  - Alerts
  - Deduplication
  - OCR
  - · Facets configuration
  - Field based boosts

#### Background and strategic fit

The UI was looking a bit old, it was important to make it look more "2016". We were also lacking a proper user management system, as well as graphical admin functionalities for a better search experience. With this release, we will set aside the branch 1.x and focus mainly on branch 2.x, so we strongly recommend you to migrate if you want to benefit from all the improvements we are planning for the future.

#### Main changes

#	Title	Description	Notes
1	New UI framework	For the admin part, we leverage the Devoops framework. For the search UI, big update for Ajaxfrancelabs	Many thanks to the devoops dev team to share this framework in a licence compatible with Apache v2 !
2	Integration of Apache Cassandra	Used for user management, favorites, likes and alerts.	
3	More documentation	We have added many new chapters to the documentation, be it for users or for developers	

## Types of users

# Starting with v2.0, Datafari proposes four distinct roles, which have access to different sets of functionalities.

Those four roles are: **anonymous search user**, **connected search user**, **search expert**, **search administrator**. The last two roles have been created to distinguish the activity of optimising the search engine from a relevance/usage perspective and from a technical perspective. Some details about what these four roles represent.

#### 1. Anonymous search user :

This is the most basic type of user. This user can only use the search functionality, and gets only two possible displays: the search bar (including autocomplete), and the results list (including facets). He has no access to functionalities such as alerts, statistics, saving favorites, or liking results. Also, since he is anonymous, he can only search through public data. Anything that has been labelled as restricted in terms of visibility (e.g. through ACLs for files) will not be searchable. If searching through these secured documents is required, the search user will need to switch to the "connected search user" role.

#### 2. Connected search user :

This is a role similar to the anonymous search user role, except that it proposes more search functionalities. Obviously, the user can search using the search bar (including autocomplete) and the results list (including facets). Contrarily to the anonymous user, he has access to functionalities such as alerts and saving favorites. To reach these functionalities, the searcher needs to go in his administration panel, where he will see the required tabs. Also, he can search through secured documents. This means that when searching, Datafari does a security check on the documents and the user, and displays in the search result all the documents that the connected search user is allowed to see. Check the pages detailing the alerts functionality and likes and favorites functionality to learn how to use it.

#### 3. Search expert :

The search expert is responsible for maintaining the relevance of Datafari, monitoring its usage, and managing its community of users. He ensures that queries with no clicks are not left alone. He checks the vitality of the system in terms of queries. He manages the promolinks. Beyond Datafari, he also ensures that people are aware of the tool, and he trains them if necessary. The expert has access to a range of functionalities from his administration interface: statistics through a graphical dashboard, promolinks management, synonyms and stopwords management, field based boosts, documents elevation, duplicate identifier, OCR activation, user management. We don't detail these functionalities here, they have dedicated sections in the documentation.

#### 4. Search administrator :

The admin is responsible for maintaining the datafari system. He is the one in charge of restarting the server, monitoring the performances, checking the logs, ensuring that the hardware specs are still compatible with the number of documents indexed and the number of queries per second. He has a range of administration functionalities available from his admin panel: statistics through a graphical dashboard, logs search dashboard, system alerts (both for connectors and for the search engine), user alerts overall performance, user alerts scheduler activation, size limitations for the indexing phase. These functionalities are detailed in their respective sections of the documentation.

### For users

Welcome to the users documentation about Datafari. Datafari is a packaged enterprise search solution, in open source under Apache v2 Licence, combining Apache Solr, Apache ManifoldCF, ELK and Ajaxfrancelabs. To learn more about Datafari as a product, you can go to the main page of Datafari. Or if you want to know more about the creators - and main contributors - of Datafari, visit the corporate site of France Labs.

For clarity, we have decided to split this documentation based on the types of users. This structure may not be optimal, but at least it gives a bit of structure.

#### Search administrator

#### This section lists the management functionalities of Datafari for the search administrator.

The admin is responsible for maintaining the datafari system. He is the one in charge of restarting the server, monitoring the performances, checking the logs, ensuring that the hardware specs are still compatible with the number of documents indexed and the number of queries per second. He has a range of administration functionalities available from his admin panel: statistics through a graphical dashboard, logs search dashboard, system alerts (both for connectors and for the search engine), user alerts overall performance, user alerts scheduler activation, size limitations for the indexing phase. These functionalities are detailed in their respective sections of the documentation.

#### Adding a new language

For now Datafari is preconfigured for English, French, Italian, Arabic and Brazilian Portuguese. Still, its aim is to have a global reach, so the steps to enable additional languages is rather straightforward and can be found here. In case you did it, please contact us so that we can integrate it in the next releases of Datafari !

#### Step-by-step guide

For the internationalization of the user interface, we use the i18n java library:

- 1. The folder to store the i18n config files is here : datafari/WebContent/js/AjaxFranceLabs/locale. Just open it and add your new language. For example if you add German translation, put here a file named de.json.

```
public static final List<String> availableLanguages =
Arrays.asList("en", "fr", "it", "ar");
```

- 3. Add the new language in WebContent/js/AjaxFranceLabs/i18njs.js :

```
availableLanguages : [ 'en', 'fr', 'it', 'ar' ],
```

4. Finally launch the ant script datafari-dev.xml to take into Datafari the modifications.

For the internationalization of the language detection, indexing and search by the Datafari Solr engine, follow these steps:

- 1. Modify the dedicated Solr updateprocessor which is declared in the **DATAFARI\_HOME/solr/solr\_home/FileShare/conf/solrconfig.xml**, that you can find at *updateRequestProcessorChainDatafari*, which detects the languages based on the fields content and title. By default, we use English and French. In order to add a new language, modify the new language to the element "*langid.whitelist*".
- Modify the Solr schema to handle the new language, which you will find at DATAFARI\_HOME/solr/solr/solr\_home/FileShare/conf/schema. xml. You will notice that we already have the following two fields which are language specific, namely content and title. Therefore, we have "content\_en", "title\_en", "content\_fr", "title\_fr". You need to create your specific "content\_xy" and "title\_xy" fields for your new language.
- 4. Now you can restart your Datafari for the changes to be taken into account.

You can either send us your new language either directly or using github, either way is fine by us as the modifications are not huge. We will send you a **cool Datafari T-Shirt** if you share that with us, so that you can show the community you are a real Datafarian

#### **Related articles**

• Adding a new language

#### **Configuration modification**

You can edit the file DATAFARI\_HOME/tomcat/conf/datafari.properties to edit the configuration of Datafari.

• ALLOWLOCALFILEREADING

It is related to the content that you index with ManifoldCF : local filesystem crawling. It allows you to change the way that you can open local files in the Datafari Search interface.

If the setting is at **true**, it means that you do not need any extension installed on your browser to open local files (see this page for information about it). Indeed Datafari opens the file for you and sends it to you directly when you click on the document link.

If the setting is at **false**, you need to have specific settings enabled on your browser to open the document because it becomes a normal link to the resource.

#### **Crawling Files with ManifoldCF**

1 The date of LibreOffice documents is not crawled correctly. Dates extraction is not available on plain text documents.

The local file crawler is a DEMO connector that shouldn't be used for production environments. It has a bug in the extraction of the modification date of the documents (making Date facet not working).

A good alternative to the local file crawler is the "Windows share" (a.k.a. JCIFS connector), that may be used to crawl files in a file shared directory (it works under Linux as well 🙂 ).

Documentation on how to setup a samba share on Linux here

Download jcifs-1.3.xx.jar from http://jcifs.samba.org/src/ to DATAFARI\_SOURCE\_DIR/mcf/mcf\_home/connector-lib-proprietary Then edit the file DATAFARI\_SOURCE\_DIR/mcf/mcf\_home/connectors.xml and uncomment the line : <!--repositoryconnector name="Windows shares" class="org.apache.manifoldcf.crawler.connectors.sharedrive.SharedDriveConnector"/->

To take into account your modification, do as follows:

- If you have already installed Datafari:
  - stop Datafari if running,
  - change the installation status of Datafari to "STATE=installed" in the file DATAFARI\_INSTALL\_DIR/bin/common/init\_state.properties, start Datafari.
- If you are running Datafari in development environment: Restart Datafari.

Use the files presented below to add whatever you want to the existing Solr configuration of Datafari. They will never be erased by an update in Datafari 2.x if you perform a modification, plus they are automatically included in the default conf. You can save them in order to backup your custom conf and easily port them on other Datafari installations.

Keep in mind that if you chose not to use these files and to directly modify the standard solrconfig.xml file and/or schema.xml file, you will have troubles saving and restoring your modifications on a new Datafari installation or an update process.

In order to have a smooth update behavior for our users, the Solr configuration files of Datafari are configured by default to include custom files that represent specific parts of Solr. Here is the custom files list and their purpose:

Location = {Datafari\_home}/solr/solr\_home/FileShare/conf/customs\_schema/
 custom\_fields.incl - Used to add new fields to the FileShare Solr Core. Structure of the file:

```
<fields>
<field name="my_field" indexed="true" stored="true"
type="string" multiValued="true"/>
<field name="my_field2" indexed="true" stored="true"
type="string" multiValued="true"/>
....
</fields>
```

• custom\_copyFields.incl - Used to add custom copy fields to the FileShare Solr Core. Structure of the file:

```
<copyField source="my_field" dest="my_copy" />
<copyField source="my_field2" dest="my_copy2" />
...
```

·-----

• custom\_dynamicFields.incl - Used to add custom dynamic fields to the FileShare Solr Core. Structure of the file:

```
<dynamicField name="my_dynamic_field*" type="string"
multiValued="true"/>
<dynamicField name="my_second_dynamic_field*" type="string"
multiValued="true"/>
...
```

• custom\_fieldTypes.incl - Used to add custom field types to the FileShare Solr Core. Structure of the file:

```
<fieldType name="myType" class="solr.TextField"
positionIncrementGap="100">
        <analyzer>
        <tokenizer class="solr.KeywordTokenizerFactory" />
        <filter class="solr.LowerCaseFilterFactory" />
        </analyzer>
        <fieldType name="myType2" class="solr.TextField"
        positionIncrementGap="100">
        <analyzer>
        <tokenizer class="solr.KeywordTokenizerFactory" />
        <filter class="solr.LowerCaseFilterFactory" />
        <filter class="solr.LowerCaseFilterFactory" />
        </fieldType>
....
```

- Location = {Datafari\_home}/solr/solr\_home/FileShare/conf/customs\_solrconfig/
  - custom\_libs.incl Used to add additional libraries that Solr must load. Structure of the file:

```
<lib dir="./lib/mylibs"/>
<lib dir="./lib/mylibs2"/>
...
```

• custom\_request\_handlers.incl - Used to add new request handlers for the FileShare Solr Core. Structure of the file:

```
<requestHandler class="my.requestHandler"
name="/myRequestHandler">
<lst name="defaults">
<str name="param1">true</str>
<str name="param2">json</str>
</lst>
</requestHandler>
<requestHandler class="my.requestHandler2"
name="/myRequestHandler2">
<lst name="defaults">
<str name="defaults">
<str name="param1">true</str>
<str name="param2">json</str>
</lst>
</requestHandler>
```

• custom\_search\_components.incl - Used to add new search components for the FileShare Solr Core. Structure of the file:

• custom\_search\_handler.incl - Used to replace the default Datafari search handler with a custom one. This file must contain only one search handler which respect the following structure:

```
<requestHandler class="solr.SearchHandler" name="/select">
....
</requestHandler>
```

When this file is filled with a custom search handler, you MUST comment the original one in the *{Datafari\_home}/solr/so lr\_home/FileShare/conf/solrconfig.xml* file. Otherwise you will have troubles with Solr and Datafari !

• custom\_update\_processors.incl - Used to add new update processors for the FileShare Solr Core. Structure of the file:

```
<processor class="my.OwnUpdateProcessorFactory">
    <bool name="enabled">true</bool>
    <str name="param">something</str>
    </processor>
<processor class="my.OwnUpdateProcessorFactory2">
    <bool name="enabled">true</bool>
    <str name="param">something</str>
    </processor>
...
```

#### **Open local files**

- 1. Open local files with Chrome
- 2. Open local files with Firefox
- 3. Open local files with Internet Explorer

#### 1. Open local files with Chrome

By default (and for security !), Chrome forbids the opening of urls pointing towards local files, from a web browser. There is a workaround using the Localinks extension:

https://chrome.google.com/webstore/detail/locallinks/jllpkdkcdjndhggodimiphkghogcpida

Then activate the extension.

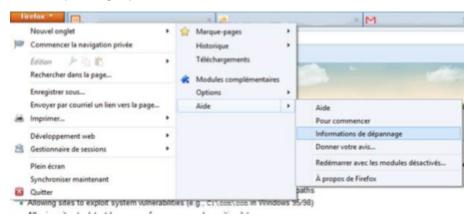
2. Open local files with Firefox

Again, by default and for security, Firefox forbids the opening of URLs pointing at local files, from a web browser. Here is a work around:

• Open the AppData folder

To locate it easily in Firefox, click on :

Firefox->Help->Debug help



Then click on Open the corresponding folder

#### Paramètres de base de l'application

Nom	Firefox
Version	18.0.1
Agent utilisateur	Mozilla/5.0 (Windows NT 6.2; WOW64; nr:18.0) Gecko/20100101 Firefox/18.0
Dossier de profil	Ouvrir le dossier correspondant
Plugins activés	about:plugins
Configuration de compilation	aboutbuildconfig
Rapports de plantage	aboutcrashes
Utilisation mémoire	aboutmemory

The path sould like the following : C:\Users\Utilisateurs\AppData\Roaming\Mozilla\Firefox\Profiles\xxx.default

2 possibilities :

• Either there is already a user.js file, to which you need to add the following lines :

user\_pref("capability.policy.policynames", "localfilelinks");

user\_pref("capability.policy.localfilelinks.sites", "http://192.168.1.88:8080");

user\_pref("capability.policy.localfilelinks.checkloaduri.enabled", "allAccess");

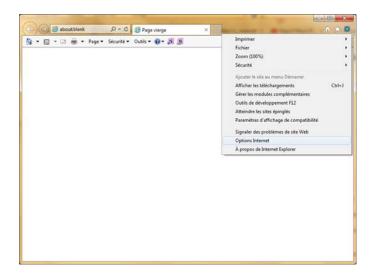
• Or there is no file named user.js, in which case you need to create it and put the lines detailed above.

You then need to copy user.js into the folder, to stop Firefox (check that the process has been killed) and restart it. The local files should open normally.

3. Open local files with Internet Explorer

Open Internet Explorer

#### Go to Internet Options



#### Go to the security tab

#### Click on Trusted Sites



Click on Sites and add the ip address of your datafari, starting with http:// (do not add the port number, just the IP)

Vous pouvez ajouter ou suppriv zone. Tous les sites Web prése les paramètres de sécurité de l	nts dans cette zone utiliseront
jouter ce site Web à la gone :	
nttp://192.168.0.74	Ajouter
ites Web :	
fle://127.0.0.1	* Supprimer
http://192.168.0.120 http://192.168.0.74	1
http://192.168.0.92	
Asses. (1100 110 0 100	•
Exiger un serveur sécurisé (https:) p	gur tous les sites de cette zone

#### OpenSearch

In this page, we will show you how to generate OpenSearch standard responses from your Datafari's Solr indexes.

Therefore you can use it on your Windows explorer windows directly if you want.

You will need two files in order to do that :

- A OSDX file
- A XSLT file

The files are present in DATAFARI\_SRC/dev-tools/opensearch

On your Datafari server

#### XSLT

We need now to transform the response sent by Solr with the XSLT writer to be compliant with the OpenSearch format.

To do that, we add the opensearch.xsl file in DATAFARI-SRC/solr/solr\_home/FileShare/conf/

The fields displayed will be the title, the last modification date and the snippet highlighted for each document.

To take the file into account, you need either to reload the FileShare Solr core or restart Datafari.

**On your Windows PC desktop** 

#### **OSDX** file

Copy the file datafari.osdx into your PC. This file contains the Solr connection information.

Edit the file to change the address of your Datafari server :

```
<Url type="application/rss+xml"
template="http://localhost:8983/solr/FileShare/opensearch?q={searchTerms}&
amp;wt=xslt&tr=opensearch.xsl"/>
```

NB : in this example, we allow remote access of Solr and we do not use the search proxy which is not a recommanded way to do it in production.

To enable remote access of Solr, edit the file DATAFARI\_SRC/solr/server/etc/jetty-http.xml and comment the line :

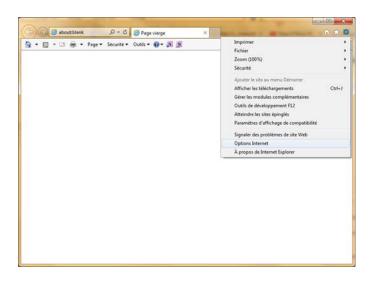
<Set name="Host">127.0.0.1</Set>

Then, to install it on your Windows system, just double click on it. You will see a new shortcut on your Windows explorer.

In case you get an "element blocked by Internet security parameters" error in windows explorer when searching, do the following:

Open Internet Explorer

Go to Internet Options



#### Go to the security tab

#### **Click on Trusted Sites**



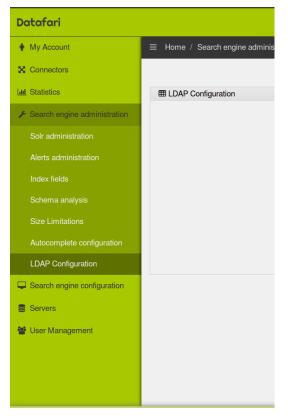
Click on Sites and add the ip address of your datafari, starting with http:// (do not add the port number, just the IP)



Restart your windows explorer, do another search and voila !

#### **Security Management**

As mentioned in the section about Realms in the Datafari developers documentation, the *JNDIRealm* allows Datafari to use an AD for authentification, via an LDAP call. This functionality requires a specific configuration for the LDAP used. To set that, the administrator can use the admin page "*LDAP Configuration*" in the section "*Search engine administration*" (see the image below).



This functionality requires that you have the Search admin role. Once logged in and into the admin environement, go to the left menu. In "Search engine administration", click on the LDAP configuration submenu.

The first time you open this page, the switcher is off.



This means that the ADconfiguration is not used. By clicking on the switcher you will activate the process. Datafari will then be integrated with your AD. This click is not sufficient enough to make AD work, you also need to configure it correctly. Configure the following:

- "connection URL" following this syntax: Idap://URLTOSERVER:PORT.
- "userBase". This is the path to root. Please follow this syntax: cn=GROUPOFUSERS,dn=ROOT2,dn=ROOT2...
- "ConnectionUsername": admin username for the AD
- "ConnectionPassword": admin password for the AD. This will allow Datafari to check the authentification of users by accessing the AD. It
  is also used to see the permission of the user in the AD so that for an example we can allow some files to be viewed in FileShare and
  some others not by checking the ACL permissions.

When the configuration is done and you click save, the server will test if the configuration is correct by attempting to connect to the AD. If the attempt is successful, then the modifications are saved and the corresponding files are changed. If the attempt fails, an error message is displayed below the Save button :

Admin Datafari - Firefox Developer Edition	. 🔢 Downl 🚺 httain 🛛 Online 🔌 HTTP 🧐 Java 🔮	Solr A   Proble Admin × Apache M   Keydo   C Q Search  C ■ Q Search	<ul> <li>Expre</li> <li>▶ ☆ 自 ♥ ↓</li> </ul>	
Datafari				Welcome, admin
My Account	Home / Search engine administration / LDAP Configurati	ion		
X Connectors				
Lul Statistics	IDAP Configuration			^
Search engine administration			on 🕑	
Solr administration	Connection URL	Idap://52.16.74.128:389		
Alerts administration	Connection Username	reda@corp.francelabs.com		
Index fields	Connection Password UserBase	fakepassword cn=Users,dc=corp,dc=francelabs,dc=com		
Schema analysis				
Size Limitations			SAVE	
		■ Fail to connect to LDAP with the setting given		
Autocomplete configuration		A run to connect to EDA with the county given		
LDAP Configuration				
Search engine configuration				
Servers				
嶜 User Management				
	Da	tafari © Copyright France Labs		

Once the AD is correctly configured, a user will be able to connect with its AD identifiers without suffix. For instance, assuming there is an "admin" user and the domain is corp.francelabs.com, he can connect to Datafari by giving only admin as username instead of admin@corp.francelabs.com.

#### Extracting file content in ManifoldCF and apply OCR

By default, Datafari uses the Apache Solr Extracting Handler (aka Solr Cell) that leverages Apache Tika to extract the content that will be indexed from the crawled files. To limit the resource consumption (especially the network if ManifoldCF is installed on an external server), it is possible to use Tika directly in ManifoldCF. In this case, the content is extracted directly in ManifoldCF, and only the content that should be indexed is sent to Apache Solr. A Tika Transformation Connection is configured by default in Datafari. To use it, you simply have to add it to your crawling job in ManifoldCF :

Datafari									Page	le recherche Bienver admin
🛊 Mon compte		ministration	de MCF							
X Connecteurs										
Administration de MCF	<u>بشر</u>		Document In	gestion						16 juin 2016 09:20
Sauvegarde et restauration config	💧 📄 Apache									
dd Statistiques	Manifo	dCF™								
Administration du moteur de										
echerche	Outputs		1	1			1.			Edit job 'Crav
Configuration du moteur de	List Transformation Connections	Name	Connection	Scheduling	Hop Filters		Repository Path	s		Edit job Crav
cherche	List Output Connections				Stage	Type	Precedent	Description		Connection name
Serveurs	Authorities				1.	Reposito				LocalFiles
	List Authority Groups	Pipeline:	Delete Insert tra	nsformation before	2.	Output	4.			DatafariSoir
Gestion d'utilisateurs	List User Mapping Connections									
	List Authority Connections		Transfort			nation			None selected •	
	Repositories	Priority: 5	•			Start n	nethod:	Don't outo	matically start this job	TikaOCR
	List Repository Connections							Doint auto	iniaucally start this job	
	List Notification Connections	Save	Cancel							
	Jobs									
	List all Jobs									
	Status and Job Management									

Then, click on "Insert Transformation Before". You can now go to the tab "Boilerplate" and select "Extract Everything" :

Datafari									Page de recherche	Bienvenue, admin
🛊 Mon compte	⊟ Home / Connecteurs / Ad	ministration de MCF								
Administration de MCF	····	Document Ingestion Apache							16 juin 2016 09:35:22	
Sauvegarde et restauration confi										
Lull Statistiques	Manifo	dCF™								
recherche	Outputs	Name Connection	Scheduling	Hop Filters		Field manning		Boilemlate		Edit job 'Crawl'
	List Transformation Connections		y			. ing mapping	Encophono	Benerbiate		
recherche	List Output Connections									
Serveurs	Outputs List Transformation Connections List Output Connections	Denotphate extraction	Extract everythin	ig		•				
	List Authority Groups	Save Cancel								
Gestion d'utilisateurs	List User Mapping Connections									
	List Authority Connections									

Now your crawling job is ready to use Apache Tika directly in ManifoldCF to extract the content of the crawled files. You can also use Tesseract OCR to analyse images in order to extract data from image and pdf files. Tesseract OCR is bundled in Datafari. In order to make the ManifoldCF TikaOCR transformation connector able to use Tesseract to do the OCR analysis, you have to change the property "**OCR**" in */opt/datafari/tomcat/ conf/datafari.properties* from "**false**" to "**true**". Then restart Datafari.

#### Manage Solr configuration with Zookeeper

The search administrator has the task to manage the Solr configuration via upload it or download it from/to Zookeeper.

Indeed with Datafari 3+ and the SolrCloud mode by default we cannot directly edit the configuration on the file system and reload the core. The Solr configuration is now stored in ZK.

1. Modify the configuration, send it back to Zookeeper and reload the Solr collection

Edit the Solr configuration
 By the UI admin

You can easily modify some parameters of the Solr configuration by the UI admin like the synonyms.

• by the file system

You can change the configuration in the filesystem directory to change precise parameters that you cannot change directly in the UI for example. The path is : /opt/datafari/solr/solr/solr/cloud/FileShare/conf

• Send the new configuration to ZK

When you are done with your changes, you have to send the new configuration to ZK.

You can do it directly by the UI : click on the tab Zookeeper located in Search engine configuration.

Doto Afficher les applications		Search Page	Welcome, admin
n My Account			
X Connectors			
LIII Statistics	Download configuration from Zookeeper		^
<ul> <li>Search Engine</li> <li>Administration</li> </ul>	Download the FileShare core configuration from the Zookeeper server Download		
Configuration			
Query Elevator	Upload configuration to Zookeeper		^
Promolinks	Upload the FileShare core configuration to the Zookeeper server		
Synonyms	Upload the FileShare core conliguration to the Zookeeper server Upload		
Fields Weight			
Facet Configuration	Reload Zookeeper configuration		^
Deduplication	Reload the Zookeeper configuration		
Likes and favorites	Reload the Zookeeper Configuration Reload		
Zookeeper			
Servers			
📽 User Management			

Now click on the button Upload configuration to ZK.

Doing that, your modified configuration is now stored in ZK.

The final step is to load it into Solr.

Reload Collection

We now tell to the Solr collection FileShare to reload itself to have the latest version of Solr configuration from Zookeeper.

For that, click on the button Reload.

Be careful to not do it during heavy load of Solr, especially during indexing time.

2. Download Solr configuration from ZK

If you want to replace the Solr configuration present on the filesystem by the Solr configuration present in ZK, you can do it also by the Zookeeper Admin UI.

Click on the button Download. The Solr configuration present in /opt/datafari/solr/solrcloud/FileShare/conf will be replaced by the Solr configuration from the ZK server.

Note : if you add new files on the local file system they will not be deleted from the Solr configuration from the ZK server. in order to have a clean copy of the Solr configuration you can erase all the content of the folder /opt/datafari/solr/solrcloud/FileShare/conf and after that download the configuration from ZK server.

#### **Search experts**

# This section presents the configuration, management and analytics tools that the search expert can leverage in Datafari.

The search expert is responsible for maintaining the relevance of Datafari, monitoring its usage, and managing its community of users. He ensures that queries with no clicks are not left alone. He checks the vitality of the system in terms of queries. He manages the promolinks. Beyond Datafari, he also ensures that people are aware of the tool, and he trains them if necessary. The expert has access to a range of functionalities from his administration interface: statistics through a graphical dashboard, promolinks management, synonyms and stopwords management, field based boosts, documents elevation, duplicate identifier, OCR activation, user management. We don't detail these functionalities here, they have dedicated sections in the documentation.

#### Alerts management - Mail Configuration

Search admin: managing the alerts mechanism

Using the search admin role, you first need to configure the alert mechanism so that it uses the smtp server to send emails.

The Alerts Administration UI allows to graphically edit the config file alerts.properties which is stored in InstallationDirectoryDatafari/tomcat/conf/

It needs the following information:

- SMTP = the name of the SMTP host
- Address = the sender address (exemple@exemple.com)
- **User** = the username (usually it will be the address)
- **Password** = the password of the mail address

#### For Instance :

- smtpHost = smtp.gmail.com;
- from = datafari.test@gmail.com;
- username = datafari.test@gmail.com;
- password = \*\*\*\*\*\*;

Once this is configured and saved, Datafari will automatically restart the alerts to take into account your modifications.

The search admin has more access rights than a connected search user. He gets the same type of information, except that he sees it for everyone. The only thing he cannot do is to create alers on behalf of other users.

#### Analytics

Datafari proposes by default two analytics page:

- a "static" analytics UI, focusing on search behavior, taking into account a notion of "navigation" of users when they do their search.
- a "dynamyc" analytics UI, focusing on the search log, allowing the user to create new dashboards, more flexible but with no notion of "navigation" of users.

#### **Dynamic analytics**

This page allows for a flexible analysis of search related information. It consumes a datafari solr core dedicated to Solr logs. It leverages the Bana na analytics framework, provided by the Lucidworks team, and which is a port of the Kibana analytics framework. It provides a graphical tool to perform analytics on search logs. For a deeper explanation of Banana, refer to its reference documentation.

We detail here the panels we have pre-configured for Datafari v0.9

- The first panel on the upper left, which shows a search bar, is preconfigured with the \*:\* value. This means that the remainder of the widgets refer to the empty query. As for Google, an empty query means that all documents are returned. Since we use a search index that stores all the queries done on Datafari, this empty query returns all the queries that users have done on this Datafari.
- The second panel, below the first panel on the left, is a technical widget. Please refer to the Banana documentation for details about it.
  The next panel, on the right of the second panel, allows to configure the time window used by some of the widgets. It goes from 5 min to 30d. For instance, it impacts the Event Counts widgets, which presents a timeline.
- Event counts: this chart presents a timeline of the number of search queries that have been triggered on the server. One count corresponds to one search query. If you left the empty query on the first panel, you get the total number of queries done on the server. If you have entered another value, you will only get on the timeline the queries that contain this term.
- The next panel, on the right of "Event counts", presents with a colored histogram all the terms that were co-occuring with the term of the first panel in the search queries. If you left the empty query in the first panel, this histogram gives you a ranking of the most popular terms. If you put a term in the first panel, you will get a ranking of the most popular terms colocated with this term. This widget is dynamic. If you click on one bar, which means one term, you will get an new histogram that shows the most popular terms that are colocated with both the term of the first panel, and the term of the colored bar you've clicked on.
- The last panel at the bottom, is a chart presenting the content of the search queries index. Each line thus represents a user search query. You can play with the left column, which lets you pick which fields of the index you want to see on the results list.

	DA	TAFART				
					Searc	h
•	Search Engine Cra	wler Server				
ũ	• ••		٩		12	
•		0 0 ×	TILTERING			• • × TIMEPICKER
	querystring <u>must</u>	time <u>must</u> Cr Sr 🗴 🔮	5m	15m 1h 6h 12h	24h 2d 7d	30d
		field date from:150/WHOUR-1HOUR from/Date/Obj "2015-01-15121-47-12.4292" to:150/WHOUR%281HOUR teDate/Obj: "2015-01-15122-47-12.4292"	Relative	Absolute   Since   Auto-refresh		
•	View           Q.Zoom Out           (12)         count per 3           35			munique (s) © comm (s) © presse (s) = Massing Reid (b) © Oher val		0 0 x TERMS
	QUERIES					🖹 🚯 🌣 🗙 TABLE
•	Fields <b>O</b>		0 to 10	of 12 available for paging		>
		♦ (date ♥)	∢q_search⊁	<nohits> <numfo< th=""><th>und &gt; <numclicks></numclicks></th><th><qtime< th=""></qtime<></th></numfo<></nohits>	und > <numclicks></numclicks>	<qtime< th=""></qtime<>
	©_version_		datafari			
	C dick	2015-01-15T22:39:36.561Z	france labs			
	✓ date □ history		communique			
	Did	2015-01-15T22:37:48.171Z	communique			
	⊗ noHits ⊗ numClicks	2015-01-15T22:25:14.657Z	"communique presse"	0 2	0	

#### **ELK analytics**

The 2.2 release of Datafari introduces the ELK layout (Elasticsearch, Logstash, Kibana) to provide a full customizable analytic UI through Kibana.

By default, three dashboards are provided with Datafari:

- Content Analysis
- Content Analysis Over Time
  Usage Statistics

To open or switch between the available dashboards, click on the "Dashboard" tab, then click on the folder button next to the filter bar and select your dashboard in the list that appears

Discover	Visualize	Dashboard	Settings					ł	② Last:	24 hours
					٩	B	B	ľ	Ð	٠
								mar	nage das	hboards
									3 dash	boards
	Discover	Discover Visualize	Discover Visualize Dashboard	Discover Visualize Dashboard Settings	Discover Visualize Dashboard Settings					Q B B C C O manage das

#### **Configure ELK**

This feature is only available from the v2.2 of Datafari !

By default, aside from Kibana, Elasticsearch and Logstash are automatically configured by Datafari on the first start, to fit with your installation directory and be ready to run.

If for some reasons (like an architecture decision), you move one or more of these components to a different place than the installation directory of Datafari, you must modify the ELK environment parameters located in [DATAFARI\_HOME]/elk/scripts/set-elk-env.sh and maybe also the [DAT AFARI\_HOME]/elk/scripts/start-elk.sh and [DATAFARI\_HOME]/elk/scripts/start-elk.sh and [DATAFARI\_

The first time you run Datafari, go to the admir	UI and to the 'ELK Configur	ration' page (Statistics/ELK Co	nfiguration). You will find this UI:

Home / User Management / ELK Configuration		
	ол 📀	
Kibana URI	http://localhost:5601	
	SAVE	
	SAVE	

Let us go through the 2 options:

- "ON/OFF" button : Should ELK start when Datafari is running ? (Note that if you swich it ON, it will immediately start ELK)
- Kibana URI : The URI to reach Kibana. By default it is http://localhost:5601

Once you have started ELK (be aware that the average time for ELK to be 100% operational is approx. 15 seconds on a "standard" PC), click on any statistic tab ('Usage Statistics', 'Corpus Statistics' or 'Corpus Over Time Statistics'). You should see the following page :

Datafari		Search Page 🕖 admin
My Account	$\equiv$ Home / Statistics / Usage Sta	tistics
X Connectors		
Left Statistics	<b>kibana</b>	Discover Visualize Dashboard Settings
On Demand Statistics	Indices Advanced Objects	
Usage Statistics	Index Patterns	
Corpus Statistics	warning No default index pattern. You must select or create one to continue.	Configure an index pattern
Corpus Over Time Statistics		In order to use Kibana you must configure at least one index pattern. Index patterns are used to identify the Elasticsearch index to run search and analytics against. They are also used to configure fields.
Query Statistics		conquire news.
ELK Configuration		
Search Engine Administration		Use event times to create index names
Search Engine Configuration		Index name or pattern
Servers		Patterns allow you to define dynamic index names using * as a wildcard. Example: logstash-*
📽 User Management		logstash-*
Ser Wanagement		Unable to fetch mapping. Do you have indices matching the pattern?

You need to configure the Elasticsearch indexes which will be used for the dashboards. To do so, type 'monitoring' as Index name and click 'Create':

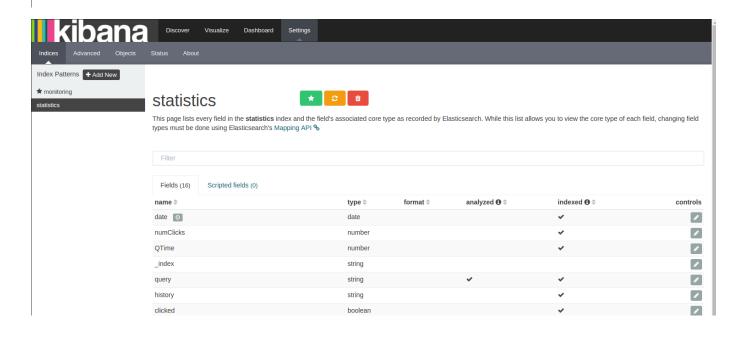
kibana	Discover Visualize Dashboard Settings
Indices Advanced Objects	- Status About
Index Patterns	
warning No default index pattern. You must select or create one to continue.	Configure an index pattern
	In order to use Kibana you must configure at least one index pattern. Index patterns are used to identify the Elasticsearch index to run search and analytics against. They are also used to configure fields.
	<ul> <li>✓ Index contains time-based events</li> <li>□ Use event times to create index names</li> </ul>
	Index name or pattern
	Patterns allow you to define dynamic index names using * as a wildcard. Example: logstash-*
	monitoring
	Time-field name 🚯 refresh fields
	timestamp
	Create

The index should have been found by Kibana and you should see the following page:

	bjects Status About							
Add New								
onitoring	📕 ★ mor	litoring	*	<b>c</b> 1				
	This page lists ev	5		associated core t	type as recorded by El	asticsearch. While this list	allows you to view the core type of	each field, changing
	Filter							
	Fields (9)	Scripted fields (0)						
	name 🌩			type 🌲	format ‡	analyzed	indexed 🔁 ≑	cont
	_index			string				
	facetType			string			~	
	numDocs			number			✓	
	_source			_source				
	facet			string			~	
	timestamp 💿			date			~	
	_id			string				
	_type			string				
	_score			number				

Now click on the "Add New" button, and repeat the same steps with the index name 'statistics':

kibana	Discover Visualize Dashboard Settings
Indices Advanced Objects	Status About
Index Patterns	
* monitoring	Configure an index pattern
	In order to use Kibana you must configure at least one index pattern. Index patterns are used to identify the Elasticsearch index to run search and analytics against. They are also used to configure fields.
	Index contains time-based events Use event times to create index names
	Index name or pattern Patterns allow you to define dynamic index names using * as a wildcard. Example: logstash-*
	statistics
	Time-field name 0 refresh fields
	date
	Create



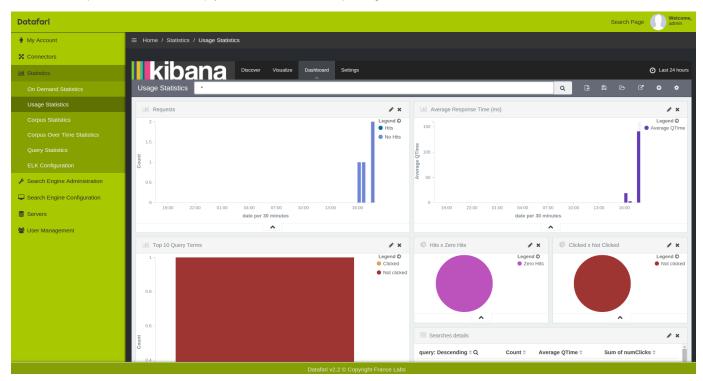
Now click on the "Objects" tab and then on the "Import" button:

kiba	Discover Visualize Dashboard Settings
	Objects Status About
Edit	Saved Objects Lexport Everything
	e you can delete saved objects, such as saved searches. You can also edit the raw data of saved objects. Typically objects are only modified via their associated application, which is what you should use instead of this screen. Each tab is limited to 100 results. You can use the filter to find objects not in the default list.
Filter	
Dashb	oards (0) Searches (0) Visualizations (0)
Select	tAll 🖞 Delete 🕹 Export
No "das	hboards" found.

And import the [DATAFARI\_HOME]/elk/save/datafari-kibana.json file. If you don't have access to the folders local to the server, you can retrieve it from our github repository https://github.com/francelabs/datafari/raw/zk/datafari-elk/save/datafari-kibana.json :

kit	Dana Discover Visualize Dashboard Settings	
Indices Adva		
	Edit Saved Objects	
	From here you can delete saved objects, such as saved searches. You can also edit the raw data of saved objects. Typically objects are only modified via their associated applic probably what you should use instead of this screen. Each tab is limited to 100 results. You can use the filter to find objects not in the default list.	ation, which is
	Filter	
	Dashboards (3) Searches (3) Visualizations (13)	
	Select All ( Delete Lexport	
	Content Analysis	A (
	Content Analysis Over Time	
	Usage Statistics	/ O

Check that you have exactly the same things as the screenshot below: 3 Dashboards, 3 Searches and 13 Visualizations. If it is the case congratulations ! Your ELK is now fully configured and if you click, in the Datafari admin UI, on any statistic tab ('Usage Statistics', 'Corpus Statistics' or 'Corpus Over Time Statistics'), you will now see the corresponding dashboard:



#### **Content Analysis**

This dashboard is based on the Core Monitoring logs generated by Datafari, which are inserted by Logstash in Elasticsearch under the

"monitoring/logs" index.



#### It shows the documents distribution in the core.

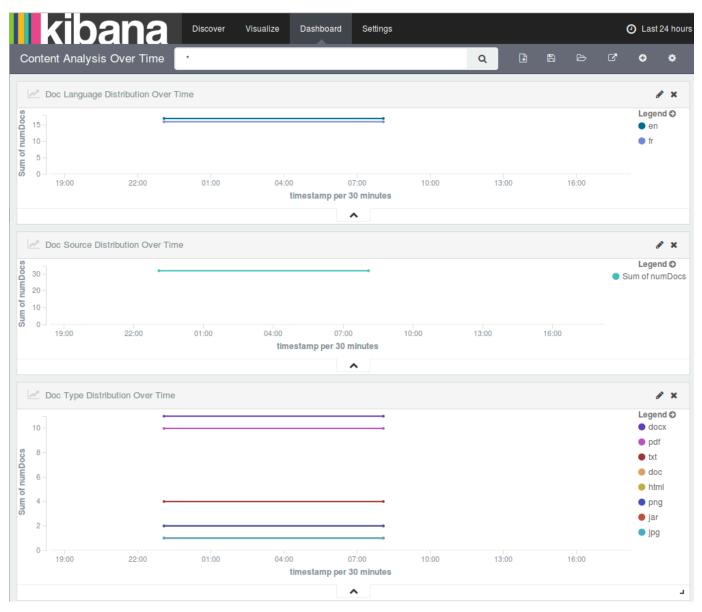
There is one pie chart per distribution type, and by default the types are doc language, doc type and doc source. As you may have understood, each pie is based on a facet : language, source and extension. You can create more charts, based on other facets by learning how Kibana works, and how our ELK layout works as well.

The important thing on this dashboard is the selected time frame. You can see it at the top right of the Kibana, it is set to "Today" by default. There is a really good reason why it is set to this value. Kibana works with time events and we decided by default to have one time event by day. It means that we have only one Elasticsearch document by facet value per day. Thus, those "daily documents" are updated each hour by Datafari when a monitoring iteration happens, and we keep the data displayed in Kibana uptodate. This also means that **if you change the time frame of this dashboard to another value than "Today", the data used by the charts will be wrong**. Indeed, the time frame of this dashboard must be set to one time event (so the current day if the time event is daily, the current hour if the time event is hourly, and the current minute if the time event is minutely)

For the moment, the Kibana time event is not configurable from the admin UI of Datafari but only by the code (see ELK layout for a better understanding).

#### **Content Analysis Over Time**

This dashboard is based on the Core Monitoring logs generated by Datafari, which are inserted by Logstash in Elasticsearch under the "monitoring/logs" index.



It shows, like the Content Analysis dashboard, the documents distribution in the core, but over the time.

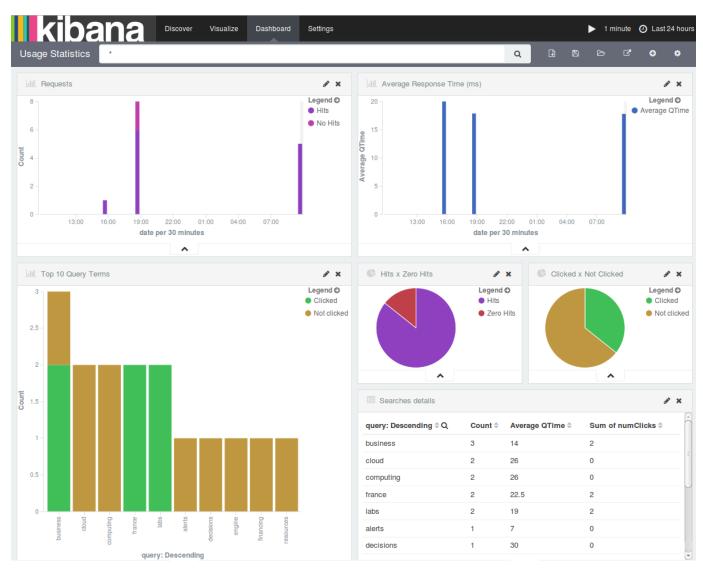
There is one line chart by distribution type which are by default doc language, doc type and doc source. As you may have understood, each chart is based on a facet : language, source and extension.

You can create more charts, based on other facets by learning how Kibana works, and how our ELK layout works as well.

Kibana works with time events and we decided by default to have one time event by day. It means that we have only one Elasticsearch document by facet value by day. Thus, those "daily documents" are updated each hour by Datafari when a monitoring iteration happens, and we keep the data to display in Kibana up to date. This also means that **you will only see one dot per day for each value**. If you want more values per day, you need to change the Kibana time event which is yet only configurable in the code of Datafari.

#### **Usage Statistics**

This dashboard is based on the Statistics logs generated by Datafari, which are inserted by Logstash in Elasticsearch under the "statistic/logs" index.



It shows, various informations about the behaviour of users on Datafari.

In the default dashboard, six visualizations are available:

- Requests: represents the number of requests over the time, with, in addition, the distribution between the requests that had at least one hit and those that did not have any hit
- Average Response Time (ms): shows the requests average response time in milliseconds over the time
- Top 10 Query Terms: it is the top 10 of the most searched terms in Datafari. For each term the chart shows the total number of time it was queried composed by the number of queries where users have clicked on a result and the number of queries where they did not.
- Hits x Zero Hits: shows the distribution between requests that had at least one hit and those that did not have any
- Clicked x Not Clicked: shows the distribution between queries where users have clicked at least on one result and those where they did not.
- · Searches details: it is a data table showing all terms queried with some infos

The data used to build this dashboard are based on data corresponding to the selected time frame, which is 24 hours by default.

You can create more charts, based on other data, by learning how Kibana works, and how our ELK layout works as well.

#### Search navigation analytics

This static page focuses on search analysis, with a notion of "navigation" from users.

Going through the functionnalites, you can first decide which type of analysis you want to see, but selecting it with the buttons on the first line titled "Type":

- Top Queries: it will display in the chart below only the most frequent queries. This allows you to identify trends in terms of what users are searching for.
- No Hits Queries: it will display in the chart below only the most frequent queries where the search engine found no corresponding

documents in its index. This may be a sign that you may have missed a relevant data source repository (and then you should add it in ManifoldCF), or maybe some documents have not been taken into account at the indexing phase, in which case you should contact your system administrator to investigate the issue.

- No Clicks Queries: it will display in the chart below only the most frequent queries with results, but where users did not click on any result from the list. This may be a sign that users are not satisfied with the documents displayed. If you have an idea of which document should be displayed for a given query, you can use the admin functionnality that allows to promote documents based on identified queries.
- **Too long Queries**: it will display in the chart below the queries sorted by their execution time, from the slowest to the fastest. This is can be used to evaluate the performance of the engine. But it can also be used as hint for the system administrator if some specific queries unexpectedly consume a very long time to be executed.



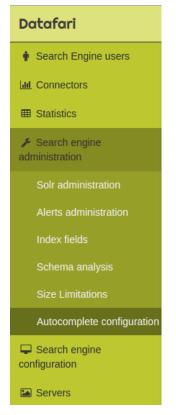
Type : O Top Queries No Hits Queries No Clicks Queries Too long Queries Filter by keyword : Filter by date : O All Last 12 months Last 6 months Last 30 days Process

Statistics on 8 distinct queries									
query	count	frequency	AVGHits	numNoHits	withClick	withClickRatio	AVGClickPosition	AVGQTime	MaxQTime
communique	4	33.33	12	0	1	25	1	260	367
comm*	2	16.67	0	2	0	0	-	16	21
"communique presse"	1	8.33	2	0	0	0	-	60	60
comm	1	8.33	0	1	0	0	-	12	12
comminuque	1	8.33	0	1	0	0	-	120	120
communique presse	1	8.33	10	0	0	0	-	48	48
datafari	1	8.33	8	0	1	100	2	17	17
france labs	1	8.33	14	0	0	0	-	47	47

#### Autocomplete configuration

Auto-complete configuration allows you to set the proportion of documents in the index where a term should appear to be suggested in the auto-complete. The higher this proportion, the less terms are going to be suggested.

In order to configure auto-complete, you have to be the search administrator, and access to the UI.



Now you can set the value to something between 0 and 1, then click the "confirm" button.

^
irm

#### Core reload required

Keep in mind that Datafari takes account of the changes only after a reload of the core or a complete restart.

Datafari relies on Apache ManifoldCF for the connection to data sources and retrieval of the data. It has a standalone documentation, which you can refer to in order to properly setup your connectors: Apache ManifoldCF official documentation.

#### Crawling

The crawling phase focuses on contacting the remote data repositories, retrieving the data, pre-processing it, and handing it over to the indexing engine. Using Apache ManifoldCF, it handles full or partial data retrieval, can connect to many different types of data repositories, and handles the security phase of indexing and searching. The full documentation for the crawl administrator is naturally located on the apache manifoldCF website, more precisely on the ManifoldCF 1.5.1 documentation.

## 1 The date of LibreOffice documents is not crawled correctly. Date metadata is not extracted from plain text documents.

We detail here how to kick off quickly in Datafari with a crawling of a local file folder.

Local and shared files crawling

#### A The local file crawler is a DEMO connector that shouldn't be used for production environments.

A good alternative to the local file crawler is the "Windows share" (a.k.a. JCIFS connector), that may be used to crawl files in a file shared directory (it works under Linux as well  $\bigcirc$ ). You can check its official documentation on the Apache ManifoldCF website. Documentation on how to setup a samba share on Linux here.

Download jcifs-1.3.xx.jar from http://jcifs.samba.org/src/ to DATAFARI\_SOURCE\_DIR/mcf/mcf\_home/connector-lib-proprietary Then edit the file DATAFARI\_SOURCE\_DIR/mcf/mcf\_home/connectors.xml and uncomment the line : <!--repositoryconnector name="Windows shares" class="org.apache.manifoldcf.crawler.connectors.sharedrive.SharedDriveConnector"/->

To take into account your modification, do as follows:

If you have already installed Datafari:

- stop Datafari if running,
- change the installation status of Datafari to "STATE=installed" in the file
- DATAFARI\_INSTALL\_DIR/bin/common/init\_state.properties,
- start Datafari.

From the admin menu, click on Crawler and then on ManifoldCF Admin.



ManifoldCF then asks you for your credentials, which you have setup either in the config xml file, or in the install phase if you used it.

	DATAFAR	
Search Engine	Crawler	Server
Apach Mani		nt Ingestion
User ID: Password:		
Login		

You are now in the MCF admin page. Since this is the MCF of Datafari, an output connection is already configured to point to the Solr of Datafari. You can check that by clicking on "List Output Connections" on the left menu. You will then see on the right, the DatafariSolr output connection.

	DATAFAR				Search
Search Engine	Crawler	Server			
Apache Manife	oldCF <sup>™ Docume</sup>	ent Ingestion			15 janv: 20
List Output Connections	List of Output Con	nections			
Authorities		Name	Description	Connection Type	Мах
List Authority Groups	View Edit Delete	DatafariSolr		Solr	10
List Authority Connections	Add a new output cor	nnection			
Repositories					

Since we want to crawl local files, we first need to declare a new repository connection (to define the type of connection), and then we will instantiate a job that user this repository connection, pointing a the desired local folder. To declare a new repository connection, click on List Repository Connections on the left menu, then click the Add new connection button on the right panel.

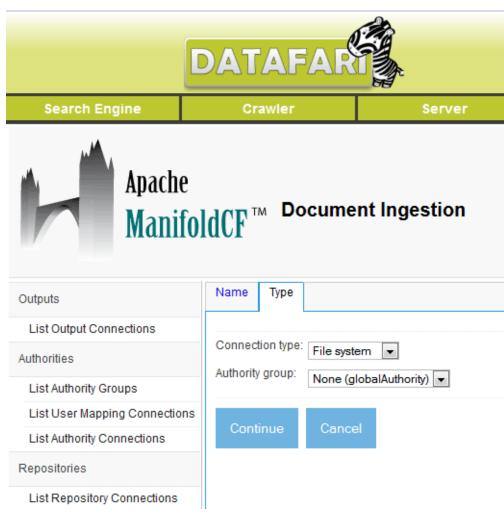
DATAFARI					
Search Engine	Crawler	Server			
Apache Manif		ent Ingestion			
Outputs	List of Dopository (	annastiona			
List Output Connections	List of Repository C	onnections			
Authorities					
List Authority Groups	Name	Description	Соп		
List User Mapping Connectior	Add new connection				
List Authority Connections					
Repositories					
List Repository Connections					

Type in the name and the description, then click on the Type tab.

DATAFARI						
Search Engine	Crawle	∋r	Server			
Apache Manif	oldCF™	Docume	nt Ingestion			
Outputs	Name Ty	be				
List Output Connections						
Authorities	Name:	Files conne	ctor			
List Authority Groups	Description:	Samples file	es connector for local do	cs crawling		
List User Mapping Connection	IS					
List Authority Connections	Cancel					
Repositories						
List Repository Connections						

In the type tab, you pick in the dropdown list File System (note that this is a DEMO connector with known bugs, don't use it for production

<u>environments</u>), and leave the Authority group to None, since it is quick start exemple. Again, refer to the full documentation of Apache ManifoldCF to dig deeper into these options.



Click on Continue, and a Throttling tab appears, as well as a Save and a Cancel buttons. Click on Save.

DATAFAR       Search Engine     Crawler     Server					
Search Engine	- Cr	awier		Server	
Apache Manifol	dCF™	M Do	ocumen	t Ingestion	
Outputs	Name	Туре	Throttling		
List Output Connections	Connect	ion type:	File system		
Authorities	Authority		-	alAuthority) 💌	
List Authority Groups					
List User Mapping Connections	Save	e c	Cancel		
List Authority Connections					
Repositories					
List Repository Connections					

Clicking on save displays the View Repository Connection Status. You should have something that looks like the illustration below.

DATAFARI					
Search Engine	Crawler		Server		
Apache Manifo	)ldCF <sup>™ Do</sup>	cume	nt Ingestion		
Outputs View Repository Connection Status					
List Output Connections	List Output Connections				
Authorities					
List Authority Groups	Name:	Files co	onnector		
List User Mapping Connections	1				
List Authority Connections	Connection type:	File sys	stem		
Repositories	Authority group:	None (	global authority)		
List Repository Connections	Throttling:		ular expression		
Jobs		No thro	mes		
List all Jobs	Connection status	: Conne	ction working		

Status Renorts

Status and Job Management

In the menu, click again on List Repository Connections to check that your new connection is now listed. You should have something like this:

	DATAFA	RI			Search	
Search Engine	Crawler	Ser	ver			
Apache Manifol	List of Reposito	ment Ingestion	on		15 נייען 15	2016 22:3
List Output Connections Authorities						
List Authority Groups		Name	Description	Connection Type	Authority Group	Max
List User Mapping Connections	View Edit Dekite	Files connector	Samples files connector for local docs crawling	File system	None(globalAuthority)	10
List Authority Connections						
Repositories	Add new connects	on				
List Repository Connections						

Now that we have declared a repository connection of type local file system, we can now "instantiate" it through a job. For that, click on the menu on "List all Jobs". The right panel will display an empty job list. Click on the button "Add a new job"

DATAFARI					
Search Engine	Crawler	Server			
,		ment Ingestion			
Outputs	Job List				
List Output Connections					
Authorities	Name	Output Connection	Rep		
List Authority Groups	Hame	ouput connection	Kep		
List User Mapping Connection	Add a new job				
List Authority Connections					
Repositories					
List Repository Connections					
Jobs					
List all Jobs					

You then need to declare a name. Do it and click on the Connection tab

	DAT	AFA			
Search Engine	Cr	awler	Server		
Apache Manifo	ldCF'	M Docum	ent Ingestion		
Outputs	Name	Connection			
List Output Connections					
Authorities	Name:	Name: Crawl local temp			
List Authority Groups					
List User Mapping Connections	Can	icel			
List Authority Connections					
Repositories					
List Repository Connections					
Jobs					
List all Jobs					

In this tab, click on the Output connection dropdown and select the Datafari specific Output connection "DatafariSolr". On the right, click on the Repository connection dropdown and pick the one we just created, in our case it is "Files connector". Then click on Continue.

DATAFA	R			Search
Search Engine	Crawler	Server		
Apache Manifo	ldCF <sup>™ Docume</sup>	nt Ingestion		16 janv. 2015 22:38:
Outputs	Name Connection			Edit job 'Crawl local temp
List Output Connections				
Authorities	Output connection: Datafa	riSolr 💌	Repository connection:	Files connector
List Authority Groups	Priority: 5		Start method:	Don't automatically start this job
List User Mapping Connections				
List Authority Connections	Continue Cance	il .		
Repositories				
List Repository Connections				
Jobs				
List all Jobs				

A list of tabs then appears. Check the ManifoldCF official documentation to get an understanding of all of them.

DAT	٩FA	RI					
Search Engine	Cr	awler		Server			
Apache Manifo							
Outputs	Name	Connection	Scheduling	Hop Filters	Forced Metadata	Solr Field Mapping	Repository Paths
List Output Connections	Output o	onnection: Dat	afariSolr			Repository conr	nection: Files connect
Authorities	Priority:	5				Start method:	Don't auton
List Authority Groups		_					
List User Mapping Connections	Save	Cancel					
List Authority Connections							
Repositories							
List Repository Connections							
Jobs							

This is where you define the actual location in your local system, that you want the job to crawl and send to Datafari Solr for indexing. For our example, click on the Repository Paths tab. In the illustration below, we enter the local path "d:\Temp" and click on Add (not Save yet!).

DAT	AF/	ART	:					
Search Engine	Cr	awler		Server				
Apache Manifo		M Docun						
Outputs	Name	Connection	Scheduling	Hop Filters	Forced Metadata	Solr Field Mapping	Repository Paths	
List Output Connections						Convert path	to LIBI?	
Authorities				Root path		(e.g. http:/xyz/index.html => http://x		
List Authority Groups	Reposit	ory Paths:		d:\Temp		No documents specified		
List User Mapping Connections			Add	artrempt				
List Authority Connections								
Repositories	Save	Cancel						
List Repository Connections								
contrepresently contrepresents								

A list of potential regex appears on the right, but we won't care with it here. Click now on the Save button.

	DATA	FARI						Search		
Search Engine	Crawler		Server					Search		
Apache Manifo	ldCF <sup>™ Doc</sup>	ument Inge	stion							
Outputs	Name Connects	on Scheduling	Hop Filters	Forced Metadata	Solt Field Mapping	Repository Paths			Editjo	6 'Crawl Io
List Output Connections										
Authorities		Red	path		Convert path to URI? (e.g. http://gitindex.html +>	http://spatindex.html)	Rules			
List Authority Groups		Delete			N			Indude/exclude	Fileldiredory	Match
List User Mapping Connections							Insert Here	include ·	File •	•
List Authority Connections										
Repositories							Delete	Polula	-	
List Repository Connections							_	include •	File	
Jobs	Repository Paths:						Insert Here	10.000	1.00	
List all Jobs							P. I. I.	noute	disettery.	
Status and Job Management							Delete			
Status Reports							Add	include ·	File •	•
Document Status										
Queue Status		Add								
History Reports										
Simple History	Seve Cer	-								
Maximum Article										

You get the confirmation that the job is succesfully created when you see the summary of the created job, in the "View a Job" window. This should like the following illustration:

	DATAFA	S						
Ľ							Search	
ngine	Crawler		Server					
Apache Manifo	oldCF <sup>™ Doc</sup>	cument Inge	estion					15 janv. 2015 22:44:
nections	View a Job							
	Name:		Crawl local temp					
roups	Output connection:		DatafariSolr			Repos	tory connection:	Files connector
ng Connections	Priority:		5			Start m	ethod:	Don't automatically sta
Connections	Schedule type: Minimum recrawl in Expiration interval:	iterval:	Scan every documen Not applicable Not applicable	t once			um recrawl interval d interval:	Not applicable Not applicable
	No scheduled run	n times						
Management	No forced metad		Inlimited					
	Hop count mode:	it for link type child.	Delete unreachable	focuments				
IS	Field mappings:	Metadata field nam	•		So	Ir field name		
	Keep all metadata:	No field mapping	specified					
			Convert path to I	IB12				
y		Root path		iex.html => http://xyz/index.html)		Rules		
width	Repository Paths:	d:\Temp	No			Include/exclude	File/directory	
im						Include	file directory	
	Edit Delete	е Сору						

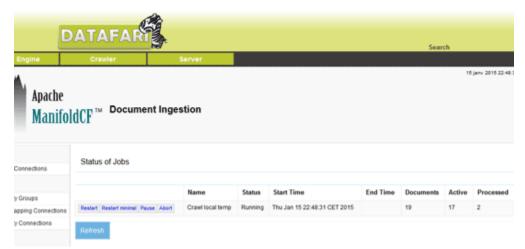
To ensure things really went ok, go on the left menu and select "List all Jobs". You should see on the right panel a job list with only one entry, the job you just created, as in the illustration below.

D	ATAFAR	\$			Search
Search Engine	Crawler	Server			
Apache Manifol	dCF <sup>™</sup> Document	Ingestion			
Outputs List Output Connections	Job List				
Authorities					
List Authority Groups		Name	Output Connection	Repository Connection	Schedule Type
List User Mapping Connections List Authority Connections	View Edit Delete Copy	Crawl local temp	DatafariSoir	Files connector	Specified time
Repositories	Add a new job				

It is now time start the actual crawling of the folder. For that, click in the left on menu on "Status and Job Management". You should get a Jobs status list on the right pane, with only one entry, which is the job you recently created. Click on Start to trigger the crawling, and wait for the magic to happen.

	DATAFAR	1					Search	
Search Engine	Crawler	Server					Jearch	
Apacho Manif		nent Ingestion						15 janu 2015
Outputs								
List Output Connections	Status of Jobs							
Authorities								
List Authority Groups		Name	Status	Start Time	End Time	Documents	Active	Processed
List User Mapping Connection	5 Start Start ninimal	Crawl local temp	Not yet run	Not started	Never run	0	0	0
List Authority Connections								
Repositories	Refresh							
List Repository Connections								
Jobs								
List all Jobs								
Status and Job Management								

If you want, you can click several time on the Refresh button to get a feeling that things are being done. This can be observed by looking the last 3 fields of the list: "Documents", "Active" and "Processed", whose numbers should be evolving as you click on Refresh. Wait for a few documents to be counted in the "Processed" field, and you can then start searching for it in the Search view, which is accessible at any time through the Search link on the upper right of the Datafari header.



Add a JDBC connector (MySQL, Oracle)

In order to crawl databases as MySQL or Oracle databases, please respect these steps (this example is for Debian and for a MySQL database but it is almost the same for Windows version) :

- Download the Java connector for your database
- Add the JAR file into the folder /opt/datafari/mcf\_home/connector-lib-proprietary

```
root@datafaridebian8:/opt/datafari/mcf/mcf_home/connector-lib-proprietary# ls -lah
total 972K
drwxr-xr-x 2 statd lpadmin 4,0K juin 30 19:29 .
drwxr-xr-x 7 statd lpadmin 4,0K juin 30 19:21 ..
-rwxr-xr-x 1 statd lpadmin 1,1K juin 11 17:58 alfresco-README.txt
-rwxr-xr-x 1 statd lpadmin 1,3K juin 11 17:58 jcifs-README.txt
-rwxr-xr-x 1 statd lpadmin 1,5K juin 11 17:58 livelink-README.txt
-rwxr-xr-x 1 statd lpadmin 1,5K juin 11 17:58 livelink-README.txt
-rwxr-xr-x 1 statd lpadmin 1,5K juin 11 17:58 livelink-README.txt
-rw-r--r-- 1 root root 946K juin 30 18:09 mysql-connector-java-5.1.35-bin.jar
-rwxr-xr-x 1 statd lpadmin 1,3K juin 11 17:58 README.txt
root@datafaridebian8:/opt/datafari/mcf/mcf_home/connector-lib-proprietary#
```

Here : mysql-connector-java-5.1.35-bin.jar

• Edit the file : options.env.unix (in /opt/datafari/mcf/mcf\_home ) (if you are on Windows, the file is options.env.win)

And add the path to the new lib in the -cp parameter :

./connector-lib-proprietary/mysql-connector-java-5.1.35-bin.jar:

#### You will have this :

-Xns256a -Xns256a

-Dorg.spache.manifoldcf.configfile=./properties.xml -cp

-ep .../lib/mcf-core.jari./lib/mcf-agents.jari./lib/mcf-pull-agent.jari./lib/hsqldb-2.3.2.jari./lib/comptos-lib/jari./lib/comptos-codec-1.9.jari./lib/comptos-codec-1.9.jari./lib/comptos-codec-1.9.jari./lib/comptos-codec-1.2.3.jari./lib/comptos-codec-1.9.jari./lib/codec-1.3.jari./lib/comptos-codec-1.9.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-1.3.jari./lib/codec-2.3.jari

- · Restart Datafari to load the library
- · You can now index your Database in ManifoldCF
  - Configure the repository connection :

Apache ManifoldCF™	Document Ingestion
Outputs	Name Type Throttling Database Type Server Credentials
List Transformation Connections	
List Output Connections	Database type: Sybase (>= V10)
Authorities	MySQL (>= V5)
List Authority Groups	Access Method by name 🔺
List User Mapping Connections	by label v
List Authority Connections	
Repositories	Save Cancel
List Repository Connections	
Jobs	
List all Jobs	
Status and Job Management	

View Reposi	itory Connection Status				
Name	bod			Description.	
Connection type:				Max connections:	10
	None (global authority)				
Throtting:	Bin regular expression No throttles	Description	Max ang Mohasimin		
Host Raw Parameters: Pass JOB JOB	óase name-lest slocational alardof-meter zalodo-meter D'houdeximpad: D'houdeximpad:				
Connection status	: Threw exception: "Driver class not found: com.mysql.jdbc.Driver"				
Refresh	dit Dekte Clear all related history				

You will see the message : 'Driver class not found' .But it is not the case : the driver is correctly loaded.

• Next configure the job as usual :

Apache ManifoldCF™	Document In	gestion					
Outputs	Name	Connection	Schedu		Queries	s. Security	
List Transformation Connections	Name	Connecson	acheou	and a	utenes	security	
List Output Connections				SELECT id AS	\$/100010990	EP/H test	
Authorities				ACCESS AN AD	P(APC/LOCH)		
List Authority Groups	Seeding que (return ids that	ry: at need to be checked)					
List User Mapping Connections							
List Authority Connections							
Repositories		Version check query: (return ids and versions for a set of documents;					
List Repository Connections	leave blank if	fno versioning capability)					
obs							
List all Jobs	Access token	query:					
Status and Job Management		d access tokens for a set of o fno security capability)	socuments;				
Status Reports							
Document Status				SELECT 1d AS	S(IDCOLUMN),	id AS \$(URLCOLUMN) HERE id in \$(IDLIS	, text AS
Queue Status	Data query:	ris, and data for a set of docu	man (a)				
istory Reports	(return ids, ur	rs, and oata of a secondocu	and the second s				
Simple History							6
Maximum Activity	Save	Cancel					
Maximum Bandwidth	Save	Cancer					
Result Histogram							

# And you will see in the history that the job is correctly executed.

- 2.21 N.2.1 m	Artista .	Identifier schlermaderen	Result Code	p.cea	Time	Result Description
0-10-201193231421	MU1	2	OK			
06-30-20-8 19:30 14:427	BALP.		OK .			
9-9-201 9-20-20 9-9-201 9-20-27 9-9-201 9-20-42	eriena gany	SELECT HAD UP HERADIAF ON BUILDING AD UP HERADIAL.	0K		3	
3-3-2014221 3-3-2014221	depunent ingest (Saturborn)		04	10	990	
09-30-2010 19:30 27:627	NUT	3	0X	10	1020	
3-3-201922194	eneral query	<pre>dt_ttp://www.station.com/action_</pre>	ox	0		
28-30-2018 19:30 28:481	enternal (suery	DELECT WAD UP_GPROM test arguments = ()	OK	0	401	
0-3-24 922468	10 start	12060 - 10000 (will)		4		

#### You can check it in Solr admin :

Search Engine	Crawler	Server	
Arache Solr Dashboard Doging Core Admin	Crawler - common 9 1 1 1 1 1 1 1 1 1 1 1 1 1	Server	"responsativader": {     "tations": 6,     "tations": 6,     "grammer": 6,     "aparemer": 6,     "mundfrows": 6,     "mundfrows": 6,     "aparemer": 6,     "mundfrows": 6,     "aparemer": 6,     "mundfrows": 6,     "mundfrows": 6,     "mundfrows": 6,     "mundfrows": 6,     "mundfrows: 6,     "mu
Java Properties	0 fi	10	"start": 0, "docs": [ {
FileShare •	đ		"Ad" ("1") "ad": "1") "attession" (") "aanged": "att
T Analysis	Raw Query Parameters key1=val16key2=val2		"contect.en": ( ) " (n in each of a sello in " ],
Documents	wt json ⊠ indent	•	"_version_"1 1505426552013081500, "allow_take_mshare": [ "nosccurity"
Ping	debugQuery     dismax		], "deny_token_document": [ "nosecurity"
ng Replication	edismax hl facet spatial		]. "Geny,token_share": [ "oosecurity" ]. "allow_token_document": [
	Execute Query		"_neseculty_" ] }, e

# Add the JCIFS Connector to Datafari

Due to licences issues, we do not have the right to package the JCIFS library with Datafari. So you have to download it on your own and to add it to Datafari.

If you want to add the JCIFS connector to the list of existing connectors, here it is the procedure in order to do it. We distinct 2 cases :

- you just downloaded Datafari and you NEVER launched it
- you have an existing Datafari running

The beginning is the same for the two cases. If you are in the second case, you just need to follow few more steps that we will detail (please stop Datafari : bash stop-datafari.sh before beginning).

1. Download the jcifs-1.3.xx.jar from http://jcifs.samba.org/src/

C f https://jcifs.samba.org/src/

# Index of /src

	Name	Last modified	<u>Size</u>	<b>Description</b>
	Parent Directory		-	
	LICENSE.txt	29-Oct-2014 23:28	26K	
	build.xml	29-Oct-2014 23:28	10K	
	docs/	29-Oct-2014 23:26	-	
	examples/	29-Oct-2014 23:28	-	
?	jcifs-1.1.11.jar	05-May-2005 01:26	253K	
D	jcifs-1.1.11.tgz	05-May-2005 01:27	609K	
D	jcifs-1.1.11.zip	05-May-2005 01:28	864K	
?	jcifs-1.2.25.jar	19-Oct-2008 23:43	373K	
D	jcifs-1.2.25.tgz	19-Oct-2008 23:43	1.0M	
D	jcifs-1.2.25.zip	19-Oct-2008 23:43	1.7M	
?	jcifs-1.3.18.jar	29-Oct-2014 23:22	387K	
D	jcifs-1.3.18.tgz	29-Oct-2014 23:22	860K	
D	jcifs-1.3.18.zip	29-Oct-2014 23:22	1.2M	
D	jcifs-krb5-1.3.17.zip	10-Nov-2011 23:20	1.3M	
D	midlc-0.6.1.tar.gz	29-Nov-2006 12:04	333K	
D	ntstatus-0.3.1.tar.gz	04-Sep-2004 18:06	146K	
Ō	patches/	11-Jan-2015 22:34	-	
	src/	29-Oct-2014 23:28	-	

Apache/2.2.3 (CentOS) Server at jcifs.samba.org Port 443

2. Copy the JAR to DATAFARI\_SOURCE\_DIR\mcf\mcf\_home\connector-lib-proprietary

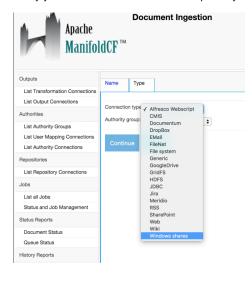
So if you are in Debian for example, copy to /opt/datafari/mcf/mcf\_home/connector-lib-proprietary# cp /root/jcifs-1.3.18.jar . root@datafaridebian:/opt/datafari/mcf/mcf\_home/connector-lib-proprietary# cp /root/jcifs-1.3.18.jar . root@datafaridebian:/opt/datafari/mcf/mcf\_home/connector-lib-proprietary# ls -lah total 412K drwxr-xr-x 2 sshd messagebus 4,0K févr. 8 23:45 . drwxr-xr-x 7 sshd messagebus 4,0K févr. 8 23:34 .. -rwxr-xr-x 1 sshd messagebus 1,1K janv. 21 18:24 alfresco-README.txt -rw-r--r-- 1 root root 388K févr. 8 23:45 jcifs-1.3.18.jar -rwxr-xr-x 1 sshd messagebus 1,3K janv. 21 18:24 jcifs-README.txt -rwxr-xr-x 1 sshd messagebus 1,5K janv. 21 18:24 livelink-README.txt -rwxr-xr-x 1 sshd messagebus 1,3K janv. 21 18:24 livelink-README.txt

3. Rename it into jcifs.jar mv jcifs-1.3.18.jar jcifs.jar

4. Edit the file /opt/datafari/mcf/mcf\_home/connectors.xml and uncomment the line : <!--repositoryconnector name="Windows shares" class="org.apache.manifoldcf.crawler.connectors.sharedrive.SharedDriveConnector"/

<pre></pre>	Connector"/>
If you just downloaded Datafari and never launched it, you are done !	
5. Additional step if you have an existing Datafari :	
a. You need to initialize the connectors in MCF. To do it the simplest is to add this script in /opt/datafari/bin :	
initialize_connectors.sh	
b. Then authorize the execution :	
chmod u+x initialize_connectors.sh	
c. Finally launch the script :	
bash initialize_connectors.sh	
You will see that the connector is now in the list of MCF connectors :	
Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.wiki.WikiConnector' Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.mebcrawler.WebcrawlerConnect Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.sharepoint.SharePointReposit Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.meridio.MeridioConnector' Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.jira.JiraRepositoryConnector Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.jibc.JDBCConnector' Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.jdbc.JDBCConnector' Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.gidfs.GridFSRepositoryConnector Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.gogledrive.GogleDriveConnector Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.gogledrive.GogleDriveRepos Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.generic.GenericConnector' Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.filesystem.FileConnector' Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.dropbox.DropboxRepositoryCon Successfully registered repository connector 'org.apache.manifoldcf.crawler.connectors.dropbox.DropboxRepositoryCon Suc	ory' ' tor' ' ctor' itoryConnector' nector'

Finally you can launch Datafari. In the repository menu of MCF, the new connector is now present :



# Reset a MCF job

In this case, we assume that you indexed some data with a MCF job and you want to delete the indexed data before restarting the MCF job. So in order to do it, the steps are :

• Make sure that your job is in "Done" mode :

Go to Jobs -> Status and Job management

Outputs	Status of Jobs							
List Transformation Connections								
List Output Connections		Name	Status	Start Time	End Time	Documents	Active	Processed
Authorities	Start Start minimal	Enron files	Done	Wed Feb 03 18:32:24 CET 2016	Wed Feb 03 18:33:15 CET 2016	381	0	381
List Authority Groups List User Mapping Connections List Authority Connections Repositories List Repository Connections Jobs	Refresh							
List all Jobs Status and Job Management • Copy your existing Go to Jobs -> List all Jobs	job :							

Then click on Copy next to the job name

Apache Manifol	dCF <sup>™</sup>	suon						
Outputs	Job List							
List Transformation Connections								
List Output Connections								
Authorities		Name	Output Connection	Repository Connection	Schedule Type			
List Authority Groups	View Edit Delete Copy	Enron files	DatafariSolr	Files	Specified time			
List User Mapping Connections								
List Authority Connections	Add a new job							

In the next screen, change the job name, here we add "2" at the end of the name then click on the Save button.

Apache Manifol	dCF		nent inge	estion		
Outputs					1.	
List Transformation Connections	Name	Connection	Scheduling	Hop Filters	Repository Paths	
List Output Connections						
Authorities	Name:	Enron files 2				
List Authority Groups						
List User Mapping Connections	Save	e Cancel				
List Authority Connections						
Repositories						
List Repository Connections						
Jobs						
List all Jobs						
Status and Job Management						

• Go back to the List jobs menu

So click on Jobs -> List all jobs.

Outputs	Job List				
List Transformation Connections					
List Output Connections					
Authorities		Name	Output Connection	Repository Connection	Schedule Type
List Authority Groups	View Edit Delete Copy	Enron files	DatafariSolr	Files	Specified time
List User Mapping Connections	View Edit Delete Copy	Enron files 2	DatafariSolr	Files	Specified time
List Authority Connections					
Repositories	Add a new job				
List Depository Connections					

Now we have the 2 jobs present on the List all jobs menu. We now can delete the original job.

• Delete the original job

Click on the Delete button front of the "old" job, so here Enron files. Confirm your choice on the popup.

• Go to the Status and job management

Click on Jobs -> Status and Job management. You will see that the Enron files job status is cleaning up. It means that MCF gives the order to Solr to delete all the documents crawled per this job.

Apache Manifol		ment Inge	estion					4 févr. 2016 12:35
Outputs	Status of Jobs							
List Transformation Connections								
List Output Connections		Name	Status	Start Time	End Time	Documents	Active	Processed
Authorities		Enron files	Cleaning up	Thu Feb 04 12:35:19 CET 2016	Wed Feb 03 18:33:15 CET 2016	0	0	0
List Authority Groups	Start Start minimal	Enron files 2	Not yet run	Not started	Never run	0	0	0
List User Mapping Connections								
List Authority Connections	Refresh							
Repositories								
List Repository Connections								

Wait a little and click periodically on the resfresh button until the cleaning process ends. At the end, only the new job is present.

Apache Manifol		ent Ingestion						4 févr. 2016 12:3
Outputs	Status of Jobs							
List Transformation Connections								
List Output Connections		Name	Status	Start Time	End Time	Documents	Active	Processed
Authorities	Start Start minimal	Enron files 2	Not yet run	Not started	Never run	0	0	0
List Authority Groups								
List User Mapping Connections	Refresh							
List Authority Connections								
Repositories								
List Repository Connections								
Jobs								
List all Jobs								
Status and Job Management								

· Start the new job

Finally you can launch your new job. Just click on the start button next to it and the data will be crawled and indexed into Solr.

# **Deduplication management**

BETA Version: Instable as of 2.2.1

#### Administration:

To enable or disable this functionality, we have not done the user interface for now (as of Datafari 1.2). You need to open the datafari.properties file, which is available in the folder \[Datafari\_install\_Folder]\bin.

In this file, set the parameter to DEDUPLICATION\_IS\_ENABLED=true or false depending on your needs. By default, it is set to true as it is resource consuming. Note that you need to restart Datafari in order for the change to be taken into account.

If you need more technical information, please refer to Deduplication : Technical note

#### **Facets management**

BETA version, Instable as of version 2.2.1

From the admin interface, you can create, delete, or modify the order of appearance of the facets displayed in the searchview.

I Add a basic facet		^
Field :	Pagination :  Name of the facet (en) :	Multiple selection :
HADING OF THE LEVEN (TIL) .	name of the lacer (eff).	Submit
Add a query facet		^
Field :	Pagination :	Multiple selection :
Name of the facet (fr) :	Name of the facet (en) :	
Query number*1 :	Query Label (en) :	Query Label (fr) :
Query number*2 :	Query Label (en) :	Query Label (fr) :
Add a query slot : 🛨	Submit	
List of the facets		^
Order	Field	
1	last_modified	8
2	extension	Ê
3	source	8
4	language	8
	Submit	

For that, you need to have the search expert or search administrator role to access the admin interface.

⊞ Add a basic facet			
	agination : 🔲 lame of the facet (en) :	Multiple selection : 🗆	Submit
⊞ Add a query facet			
	agination : 🗆 lame of the facet (en) :	Multiple selection : 🗖	
	uery Label (en) :	Query Label Query Label	
Add a query slot : +	Submit		
⊞ List of the facets			
Order	Field		
1	last_modified		<b>@</b>
2	extension		۵
3	source		Û
4	language		<b>•</b>
	Submit		

There are two distinct boxes to add a facet:

- "Add a basic facet" box: The first one is to add simple facets. The only mandatory form field is the Solr field. You can optionally:
   set the pagination to true,
  - allow the user to select multiple elements of this facet,
  - customize the name of this facet.
- "Add a query facet" box: The second one is for the query facet. It adds custom queries, that you will have to fill, at least one, otherwise it makes no sense, customize the name of the request.
  - If you don't customize the name, they will be called by the name of the field which is not explicit.
  - If you click on the "add a query slot" button, it will add another field for a query and for its query label. In this context, query label is the label that will be printed next to the checkbox in the search UI. Note that if the query slot is empty, Datafari will not take into account the field dedicated to this query label. Also, if no query is declared, Datafari will ignore the request. We do not do validity checks on the queries, so it is your responsibility to ensure they are correct. If they are invalid, it is highly probable that the global search will not work anymore. But you can still delete them by clicking on the trash at the end of the line of the bottom table.

Note that you can neither have two facets on a field, nor can you have a basic facet and a query facet on a single field. Thus the fields that already faceted are removed from the selection

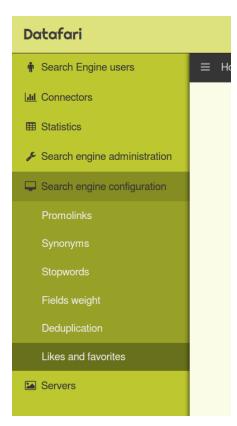
The "List of the facets" box allows to do the following on existing facets:

- · Deletion, by clicking the trash icon on the right
- Change the printing order of the facet. To change it you can drag and drop lines of the table and once the order fits your needs, click the submit button.

# **Favorites management**

Especially useful for librarians or researchers, the Favorites functionality allows users to save documents that appear in the result.

This functionality is disabled by default. To enable it, to login to the admin section. Browse to "Search Engine Configuration" and then click on "Likes and favorites" as shown in the image at the left.



Now click on the switcher to activate or deactivate the functionality. If the switcher displays the "OFF" statement, it means that the functionality is deactivated.



If you click on it, it will be activated. The switcher should display the "ON" statement.



# **Fieldweights**

Every modification you do with this UI will be saved in a custom searchHandler which is located in the {Datafari\_home}/solr\_home/FileShare/conf/customs\_solrconfig/custom\_search\_handler.incl, and the original searchHandler from the standard solrconfig.xml file will be (if not already) commented. Thanks to this behavior you can "rollback" your modifications and re-apply the default settings by uncommenting the original searchHandler in {Datafari\_home}/solr/solr\_home/FileShare/conf/solrconfig.xml and removing the custom one in {Datafari\_home}/solr/solr\_home/FileShare/conf/customs\_solrconfig/custom\_search\_handler.incl

This admin UI allows you to modify the weight of any field, in order to optimize the persistence of Datafari regarding your documents.

There are two kinds of weights: those for unique word search and those for multiple words search. So, depending on the kind of weight you want to modify, select your desired field in the corresponding list. The current weight of the field will be displayed in the text area and you will be free to set a new value.

Datafari					Page de recherche Blenvenue, admin
Mon compte	Home / Configuration du moteur de recherche / Pondé	ération des champs			
X Connecteurs					
Lat Statistiques	I Requêtes à mot unique				^
Administration du moteur de recherche	test	-	500 ‡		
Gonfiguration du moteur de recherche	1051	i	500 +	Confirmer	J
Promoliens					
Synonymes	⊞ Requêtes à mots multiples				^
Pondération des champs					
Configuration des facettes				Confirmer	
Deduplication					
1 Barris and the designation					

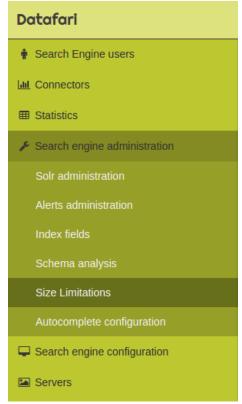
Once you are satisfied with the new value, click on the "Confirm" button.

Notice that in order to take your modifications into account, you will need to reload the Solr Core conf through the Solr admin UI or through a restart.

# **Highlighting configuration**

Size limitations allow you to restrict the number of characters analyzed by the highlighter. The higher the number, the better the highlight and the slower the queries.

To set the size limitation, you have to be the search administrator. Go in the admin UI.



Now set the value to a positive integer, then click the "confirm" button

III Limitation of the fields size for highlighting	^
Maximum number of analyzed chars by the highlight : 40000 Confirm	

Keep in mind that Datafari will take the changes into account only after a reload of the core or a complete restart.

#### Likes management

Inspired from social networks, the Likes functionality allows users to like a search result. This may be used as part of the relevancy algorithm. In terms of administration, the search expert can decide to activate or deactivate the functionality.

This functionality is disabled by default. To enable it, you have to log into the admin section. Browse to "Search Engine Configuration" and then click on "Likes and favorites" as shown in the image at the left.



Now click on the switcher to activate or deactivate the functionality. If the switcher displays the "OFF" statement, it means that the functionality is deactivated.



If you click on it, it will be activated. The switcher should display the "ON" statement.



## **Promolinks management**

Starting with Datafari v1.2, promolinks can be conceptually assimilated to google adwords: independently from the indexed content, it allows the search expert to display text in a dedicated place on the screen (usually on top or on the side). This display is triggered by the presence of specific terms in the search query.

For instance, the illustration below shows a promolink entitled "Nice title" which has the content "report" and is displayed because the keyword "report" was used in the search query::

report		
		⊙ All words
Results 0 - 0 de 0 pour report (	0.004 Seconds)	
Last Modification	~	Answer Nice title
Less than 1 month (0) Less than 1 year (0) Less than 5 years (0)		report No result found
Туре	~	
Source	~	
Language	~	

In order to create Promolinks, you need to have the search expert role.

You have access to a dedicated web interface in the admin UI. First, go to the administration menu, click on Search Engine Configuration and select Capsule in the dropdown list.

Datafari
Lul Connectors
Statistics
✤ Search engine administration
Search engine configuration
Capsules
Synonyms
Stopwords
Search Engine users
Servers

You then have access to the promolink admin UI, which shows the list of all existing promolinks. You can do the following operations on the existing promolinks:

- Delete by clicking the red button,
- Edit by clicking the keyword,

Using the search bar on top, you can filter out the results by typing in something and pressing enter.

To add a promolink, use the form at the bottom of the page, and fill in the fields to add a promolink.

Note that the dates are optional, that you may fill only one out of those fields, and that if you choose to fill both, the starting date must be prior to then ending date. If dates have been filled during the add, Datafari will check if, when a search is made on the matching keyword, the current date is posterior to the starting date and prior to the ending date

=	Home / Se	earch engine configuration	/ Capsules					8•f ≯ in ằ
	Search the ca	psules by keyword, if you w	ant all of them lea	ave 1				
F	esults							
	E List of t	he capsules						~ 2 ×
	#	Keyword :		Title :	Content :	Starting date :	Ending date :	
	0	report		Nice title	report	07/01/2015	07/15/2015	0
A	dd or edi	t a capsule						
	Add or edit							∧ Z
F	arameter	ſS						
		Keyword :	Keyword					
		Title :	Title					
		Content :	Content					
0	ates							
		Starting date :	Date					
		Ending date :	Date					
				Confirm				

# **Query Elevator**

The version 2.3 of Datafari comes with the new "Query Elevator" feature. It allows users with the "SearchExpert" role to elevate documents for specific queries, in order to make them appear in the top of the results list.

Two possibilities are provided to elevate documents:

- Directly from the search view
- Thanks to the admin UI

#### **Admin UI feature**

In order to access the admin UI described bellow, you will need to be authenticated in Datafari as a user with the "SearchExpert" or the

From the 2.3 version of Datafari, a new feature is available for the "SearchExpert" and "SearchAdministrator" users in the admin UI, the "Query Elevator" :

Datafari				Search Page Welcome admin
My Account		vator		
S Connectors				
Lat Statistics	Modify an existing query elevation			
F Search Engine Administration         Select the query				
Search Engine Configuration				
Query Elevator	Modify the documents order			
Promolinks	Documents list	<b>B</b> 111		~ Z
Synonyms	ld	Position		
Fields Weight	CONFIRM			
Facet Configuration				
Deduplication				
Likes and favorites	Create a new query elevation and/or add new docu	ments		
Servers	Query		Documents IDs	
😤 User Management				0
	CONFIRM			

This UI allows you to either modify/remove existing elevate boosts, or create new ones.

Let's start with the modify/remove part :

elect the query	
docx •	
lodify the documents order	
Documents list	^
d	Position
ile:////localhost/Datafarishare/offre_France_Labs_ExpertRecherche_Q12015.docx	1
ile:////localhost/Datafarishare/Template_stage_developpement.docx	2
ile:////localhost/Datafarishare/Communique_de_presse_FranceLabs_Datafari.docx	3

As you can see, it is composed by two sections :

- The first one is a "Select" where you will be able to select a query which has elevated documents. The list only contains queries that are present in the elevate.xml file located in [Datafari\_Home]/solr/solr\_home/FileShare/conf.
- The second is a table where will appear the documents which have an elevate boost for the query you have selected.
   The documents are ordered by priority of appearance in the search results, which is also indicated by their "Position" number.
   Here you can simply drag and drop the documents to change their position in the search results. You can also remove some of them if you want by clicking on their associated trash icon.

Once you are satisfied with the order of the documents for the selected query, you can click on the "confirm" button. Datafari will update the elevate.xml file to match your configuration and will reload the Solr core so your modifications will take effect at the end of the process. Once the Solr core is reloaded, a confirmation message will appear next to the "confirm" button.

⊞ Create a new query elevation and/or add new documents					
Query	Documents IDs				
docx	[file:////localhost/Datafarishare/Stage_template_letudiant.docx				
	[file:////localhost/Datafarishare/enterprise_document.docx				
		0			
CONFIRM					

This part allow you to create or add elevate boosts for a query. Notice that if you add new elevate boosts to a query which already has ones, the new ones will be added at the top of the list in the same order as you have entered them.

Simply type a query in the corresponding text field and then enter the id of the documents you want to elevate for this query. You can add several documents by clicking on the green cross which will add a new document ID line.

Once your are done with your configuration, click on the "confirm" button. Datafari will update the elevate.xml file to match your configuration and will reload the Solr core so your modifications will take effect at the end of the process. Once the Solr core is reloaded, a confirmation message will appear next to the "confirm" button.

## **Search View Feature**

NB : The feature is temporarily unavailable for Datafari 3.0. It will be integrated again in Datafari 3.2.

To see the feature described bellow in the search view of Datafari, you must be connected to Datafari with a user having either the "SearchExpert" role, or the "SearchAdministrator" role. Refer to the Types of users for more details.

Once you are connected as a "SearchExpert" or a "SearchAdministrator", if you perform a search in Datafari, you will notice two news icons at the right of each result:

	docx				Q
Changer pour R	echerche Avancee			💽 Tous les mots 🔵 Au moins un	mot e Expression exacte
Résultats 1 -	6 sur 6 pour docx (	0.007 Sec	ondes)		
Date		~	12	Internship_Offer_FranceLabs_2016.docx	5
				file:////localhost/Datafarishare/Internship_Offer_FranceLabs_2016.docx	
Moins de ur				offre_France_Labs_ExpertRecherche_Q12015.docx	<b>^</b>
Moins de ci				file:////localhost/Datafaoffre_France_Labs_ExpertRecherche_Q12015.docx	iO.
			tore 5		•
Туре		~	12	enterprise_document.docx	-
docx (6)				Enterprise document , file:////localhost/Datafarishare/enterprise_document.docx	E)
		_			
Source		~	12	Template_stage_developpement.docx	<b>^</b>
		· ·		Stage développement informatique à France Labs	
file (6)				file:////localhost/Datafarishare/Template_stage_developpement.docx	K)
			E000	Stage template letudiant.docx	A
Langue		~	12	/docProps/thumbnail.jpeg Stage de aeuewoppemenc mfnrmatique (1 poste) au" n un un: un: n» b uu»a. m» m	
fr (4)				Mp-«u-ræmnmmrnun v	
en (2)				file:////localhost/Datafarishare/Stage_template_letudiant.docx	κ.)
			toce .	Communique de presse FranceLabs Datafari.docx	
			12	France Labs Palais Jacqueline 4 bis bouleva	
				file:////localhost/DatafaCommunique_de_presse_FranceLabs_Datafari.docx	кЭ

These icons are used to elevate or remove an elevate boost of the associated document for the current query:

۰

• : Is used to elevate the associated document. It will add a boost to the document for the current query, that will make it appear as the first result of the list.

When you click on this button, Datafari modifies the elevate.xml file located in [Datafari\_Home]/solr/solr\_home/FileShare/conf to add or move the document as the first of the list and then restart the Solr core. This is the reason why you will have a delay before the search view is reloaded and the document appears as the first of the results.

If you elevate several documents with this method, notice that it will always be the last elevated doc which will be on top of the results list. It works like a stack where the last added element is added on top of the stack so it has the priority on the previous inserted one which also has the priority on the previous etc etc.

This feature is intended to quickly and easily elevate only one or very few documents. As on each click the Solr core is reloaded and that it may be annoying to find the doc you want to elevate or to wait after each click to elevate several docs, you will prefer to manage the elevate configuration through the Admin UI feature.

. K⊃

 : Is used to remove the elevate boost of the associated document for the current query. If the document has an entry for the current query in the elevate.xml file located in [Datafari\_Home]/solr/solr\_home/FileShare/conf, it will be removed from the list of elevated documents and the Solr core reloaded.

As for the elevate feature, this search view incorporated feature is intended to quickly and easily remove elevate boost on only one or very few documents. For better control and ergonomic interface, you will prefer the Admin UI feature.

### **Stopwords management**

In order to create Stopwords, you need to be connected with the search expert role.

Go to the administration interface. Then, go to the administration left menu, click on Search Engine Configuration and select Stopwords in the dropdown list.

M Connectors
Statistics
⊁ Search engine administration
Generation Search engine configuration
Capsules
Synonyms
Stopwords
Search Engine users
Servers

Stopwords are language specific. You need to select the language for which you want to manage the stopwords. Once this is done, you get a text window allowing you to edit the stopwords list. Note that only one search expert at a given time can edit this file. Any other simultaneous tentative will end up with an error message on the screen.

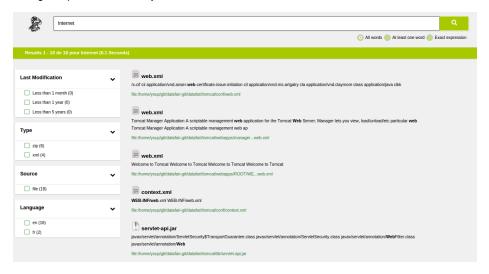
-		
Select the la	nguage	
	en 🔻	
Stopwords		
	9 P X 4 2 2 3 4 4 9	
	#Standard english stop words taken from Lucene's StopAnalyzer a	
	an	
	and are	
	a5 10	
	the of e	
	it.	

When editing the text file, please respect the format used by the Solr 5.1 for its stopwords file :

#Standard english stop words taken from Lucene's StopAnalyzer
a
an
and
are
as
at
be
but
by
for
if
in
into
is
it

# Synonyms

Synonyms are a way to create a link between words. If a word A is declared as a synonym of word B, then any search for word B will end up looking also for word A. This allows for instance, to link the words *internet* and *web*, which have an identical meaning in a given context, but have a totally dissimilar spelling. Web is a good exemple of the usefulness of giving "directions" to your synonyms. You may want searches for internet to look for documents with web in it as well, but you may not want searches for web to look for documents with internet in it, in case you are looking for spider webs. Here you can see that the term "Internet" is linked with the word "web"



Search Expert: managing synonyms

In order to create Synonyms, you need to be connected with the search expert role.

Once in the administration interface, go to the administration menu, click on Search Engine Configuration and select Synonyms in the dropdown list.

Datafari
Lill Connectors
E Statistics
Search engine administration
Search engine configuration
Capsules
Synonyms
Stopwords
Search Engine users
Servers

Synonyms are language specific. You need to select the language for which you want to manage the synonyms. Once this is done, you get a nice interface allowing you to edit the synonyms list. Note that only one search expert at a given time can edit this file. Any other simultaneous tentative will end up with an error message on the screen.

Datafari		1	Search Page Welcome, admin
My Account			
Connectors			
Land Statistics			
Search Engine Administration	Choose the language		
Search Engine Configuration	en		
Query Elevator	Synonyms		
Promolinks	⊞ List of synonyms     Words	Synonyms	~ 2
Synonyms			
Fields Weight	Enter a word and press Enter X Internet	Enter a word and press Enter Web	8
Facet Configuration	Add new synonyms		
Deduplication			
Likes and favorites	Validate modifications		
Zookeeper			
Servers			
🖀 User Management			

Here you can delete/add words and add/delete their corresponding synonyms or even delete a synonym entry by clicking on the 'trash' button at the end of a line.

To add new synonym entry, simply click on the 'Add new synonyms' button, a new line will be created and you will be able to edit this new entry:

Datafari			Search Page Welcome,
My Account	Home / Search Engine Configuration / Synonyms		
X Connectors			
Lud Statistics			
Search Engine Administration	Choose the language		
🖵 Search Engine Configuration	en 💌		
Query Elevator	Synonyms		
Promolinks	⊞ List of synonyms Words	Synonyms	^ <i>/</i>
Synonyms			
Fields Weight	Enter a word and press Enter	Enter a word and press Enter Web	â
Facet Configuration	Enter a word and press Enter	Enter a word and press Enter	â
Deduplication			
Likes and favorites	Add new synonyms		
Zookeeper			
Servers	Validate modifications		
嶜 User Management			
Datafari			Search Page 🕢 Welcome, admin
Datafari My Account	≡ Home / Search Engine Configuration / Synonyms		Search Page Welcome,
			Search Page Welcome, admin
My Account			Search Page Welcome, admin
My Account     Connectors     Let Statistics			Search Page Welcome, admin
My Account       Connectors     Lat Statistics       ✓ Search Engine Administration	⊞		Search Page Welcome, admin
My Account     Connectors     Lat Statistics     Search Engine Administration     Search Engine Configuration	■ Choose the language		Search Page Welcome, admin
<ul> <li>My Account</li> <li>Connectors</li> <li>Mi Statistics</li> <li>Search Engine Administration</li> <li>Guery Elevator</li> </ul>	Choose the language en Synonyms It List of synonyms		Search Page Welcome, admin
<ul> <li>My Account</li> <li>Connectors</li> <li>Statistics</li> <li>Search Engine Administration</li> <li>Search Engine Configuration</li> <li>Query Elevator</li> <li>Promolinks</li> </ul>	Choose the language	Synonyms	
<ul> <li>My Account</li> <li>Connectors</li> <li>Mi Statistics</li> <li>Search Engine Administration</li> <li>Guery Elevator</li> </ul>	Choose the language en Synonyms It List of synonyms	Synonyms Enter a word and press Enter	
<ul> <li>My Account</li> <li>Connectors</li> <li>Lat Statistics</li> <li>Search Engine Administration</li> <li>Search Engine Configuration</li> <li>Ouery Elevator</li> <li>Promolinks</li> <li>Synonyms</li> </ul>	Choose the language en Synonyms Ust of synony	Enter a word and press Enter X Web	
<ul> <li>My Account</li> <li>Connectors</li> <li>Statistics</li> <li>Search Engine Administration</li> <li>Search Engine Configuration</li> <li>Query Elevator</li> <li>Promoliniks</li> <li>Synonyms</li> <li>Fields Weight</li> </ul>	Choose the language en Synonyms Mords Words		
<ul> <li>My Account</li> <li>Connectors</li> <li>Statistics</li> <li>Search Engine Administration</li> <li>Search Engine Configuration</li> <li>Query Bevator</li> <li>Promolinks</li> <li>Synonyms</li> <li>Fields Weight</li> <li>Facet Configuration</li> </ul>	Choose the language en Synonyms Ust of synony	Enter a word and press Enter X Web	
<ul> <li>My Account</li> <li>Connectors</li> <li>Statistics</li> <li>Search Engine Administration</li> <li>Search Engine Configuration</li> <li>Query Elevator</li> <li>Promotinks</li> <li>Synonyms</li> <li>Fields Weight</li> <li>Facet Configuration</li> <li>Deduplication</li> </ul>	Choose the language en Synonyms Ulst of synonyms Under a word and press Enter Customize	Enter a word and press Enter X Web	
<ul> <li>My Account</li> <li>Connectors</li> <li>Statistics</li> <li>Search Engine Administration</li> <li>Search Engine Configuration</li> <li>Query Elevator</li> <li>Promolinks</li> <li>Synonyms</li> <li>Fields Weight</li> <li>Facet Configuration</li> <li>Deduplication</li> <li>Likes and favorites</li> </ul>	Choose the language en Synonyms Ulst of synonyms Under a word and press Enter Customize	Enter a word and press Enter X Web	
<ul> <li>My Account</li> <li>Connectors</li> <li>Statistics</li> <li>Search Engine Administration</li> <li>Search Engine Configuration</li> <li>Query Elevator</li> <li>Promolinks</li> <li>Synonyms</li> <li>Fields Weight</li> <li>Facet Configuration</li> <li>Deduplication</li> <li>Likes and favorites</li> <li>Zookeeper</li> </ul>	Choose the language en Synonyms Uods Enter a word and press Enter Enter a word and press Enter Customize Add new synonyms	Enter a word and press Enter X Web	

Once you are ok with your modifications, click on the 'Validate modifications'. If everything is ok you will be noticed by a message to reload the Solr core in order to apply the new configuration:

Datafari		Search Page Welcome, admin
My Account		
X Connectors		
Lat. Statistics	· ⊞	
Search Engine Administration	Choose the language	
Search Engine Configuration	en	
Query Elevator	Synonyms	
Promolinks	⊞ List of synonyms Words	Synonyms
Synonyms		
Fields Weight	Enter a word and press Enter X Internet	Enter a word and press Enter Web
Facet Configuration	Enter a word and press Enter × Customize	Enter a word and press Enter X Modify
Deduplication		
Likes and favorites	Add new synonyms	
Zookeeper		
Servers	Validate modifications	
🖀 User Management	Modifications done. Reload the Solr Core to apply them	

Otherwise you will get an error:

Datafari		Search Page 🕕 Welcome
My Account	Home / Search Engine Configuration / Synonyms	
X Connectors		
Lat. Statistics	I	
Search Engine Administration	Choose the language	
Search Engine Configuration	en 💆	
Query Elevator	Synonyms	
	⊞ List of synonyms Words	Synonyms
Synonyms		Synonyms
Fields Weight	Enter a word and press Enter x Internet	Enter a word and press Enter Web
	Enter a word and press Enter	Enter a word and press Enter X Modify
	Add new synonyms	
Servers	Validate modifications	
嶜 User Management	Error: some words do not have synonyms or some synonyms do not have corresponding words !	

# Search users

In this section, we present all the functionalities that can be used by both anonymous and connected search users.

#### 1. Anonymous search user :

This is the most basic type of user. This user can only use the search functionality, and gets only two possible displays: the search bar (including autocomplete), and the results list (including facets). He has no access to functionalities such as alerts, statistics, saving favorites, or liking results. Also, since he is anonymous, he can only search through public data. Anything that has been labelled as restricted in terms of visibility (e.g. through ACLs for files) will not be searchable. If searching through these secured documents is required, the search user will need to switch to the "connected search user" role.

# 2. Connected search user :

This is a role similar to the anonymous search user role, except that it proposes more search functionalities. Obviously, the user can search using the search bar (including autocomplete) and the results list (including facets). Contrarily to the anonymous user, he has access to functionalities such as alerts and saving favorites. To reach these functionalities, the searcher needs to go in his administration panel, where he will see the required tabs. Also, he can search through secured documents. This means that when searching, Datafari does a security check on the documents and the user, and displays in the search result all the documents that the connected search user is allowed to see. Check the pages detailing the alerts functionality and likes and favorites functionality to learn how to use it.

# Alerts

#### Alerts on plain text (eg .txt) files

1. For the moment, the alerts cannot be used on plain text files, as Tika cannot extract any metadata like the modification date from them.

Rather than actively searching for documents using queries, alerts allow you to receive information about documents via email. Once you define a query and a frequency, you regularly receive information about new and modified documents that refer to your query. For instance, the screenshot below illustrates an email alert on the keyword *Nice*.



#### datafari.test@gmail.com

À moi 🖃

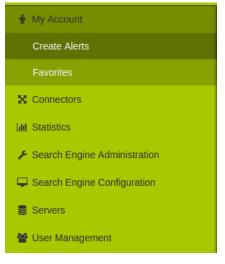
141 new or modified document(s) have been found for the key Nice; Festival Crossover : Nice autrement ! Nice : Les Moulins poursuivent leur métamorphose L'OGC Nice crucifié par l'AS Monaco L'OGC Nice héroïque fait chuter l'OM L'OGC Nice impuissant face à Reims Nice : Restaurant Franchin, une adresse à redécouvrir L'OGC Nice subit la loi de Lyon L'OGC Nice offre la victoire à Bastia Nice : le bassin olympique s'appellera Camille Muffat Nice : l'art s'expose sur la Prom

Each document is composed of three pieces of information : the document title, where to find it, and when it has been created or modified.

Connected Search user: managing your alerts

In order to create an alert, you need to be connected with a connected search user role.

We expose a web interface. First, go to the administration menu, click on My Account and select Create Alerts in the dropdown list.



You then have access to the UI that allows you to create Alerts by filling the fields, and choosing a frequency.

The form asks for the following information:

- Keyword is the word that will be used by the Solr request,
- Subject is the subject of the email,
- · Mail is the recipient's email address,
- Core is the Solr core in which the request will be made,
- Frequency is the frequency for triggering your search query and sending the corresponding emails.

If you want to know which alerts you have already created, use the search bar on top of the page, and type in what you want to search for (it searches in the queries keywords). If you want to see all your searches, leave the search bar empty and click on the search icon.

The identified results are then displayed in the results table. You can edit an existing alert by clicking on the keyword, or delete one by clicking on the red button.

Note that the alert results are related to the currently connected user. Only the search admin role and the search expert role can see all the existing alerts for all the users.

List of the alerts						· · · · · · · · · · · · · · · · · · ·
Id	Keyword :	Subject :	Core :	Frequency :	Mail :	User :
7203d23e-62d0-451e-a1d1-4bc11683f	5cd datafari	datafari modified docs	FileShare	Hourly	giovanni.usai@francelabs.com	admin
Add or edit an alert						
Add or edit an alert						^
Keyword :	datafari					
Subject :	datafari modified docs					
Mail :	giovanni.usai@francelabs.com					
Core :	FileShare					
Frequency :	Hourly	•				

# **Autocomplete**

Autocomplete displays a dropdown window below the main search bar. It suggests a list pf search queries based on what the user is currently typing, more precisely by suggesting words that start with what the user is currently typing. To pick a suggestion, the user clicks on it with the left mouse button.

# Deduplication

Datafari can allow a user to see wich documents are duplicated in the result of the search.

The deduplication functionnality uses the MD5 Algorithm for hashing the documents so that solr could recognize which documents are duplicated.

When activated, users have a special "duplication" facet that appears on the bottom left of the results page. Each item in this facets represents a set of duplicated documents, with a name and the number of duplications in parenthesis.

When clicking on a facet item, the results display will show all the duplicated documents related to the clicked facet item. This functionality can be useful to find out how many duplicated documents are present in the corpus.

## **Facets**

Facets allow you to only see the results matching a certain criteria.

It can be the type of a document or its modification date, or its content language. Facets are available in the search and you can click on the criteria you want to activate or desactivate this functionnality.

Date	~	video6_4.mp4
Duit	•	HandBrake 0.9.8 2012071700 HandBrake 0.9.8 2012071700
Less than 1 month (0)		file:/home/youp/Documents/9781783983926/Package/videos/video6_4.mp4
Less than 5 years (40)	_	video5_3.mp4
Turne		HandBrake 0.9.8 2012071700 HandBrake 0.9.8 2012071700
Туре	~	file:/home/youp/Documents/9781783983926/Package/videos/video5_3.mp4
🗹 mp4 (40)		
pdf (17)		video1_4.mp4
odt (4)		HandBrake 0.9.8 2012071700 HandBrake 0.9.8 2012071700
		file:/home/youp/Documents/9781783983926/Package/videos/video1_4.mp4
Source	~	
		video6_1.mp4
file (40)		HandBrake 0.9.8 2012071700 HandBrake 0.9.8 2012071700
Language	~	file:/home/youp/Documents/9781783983926/Package/videos/video6_1.mp4
🗌 en (40)		video3_1.mp4
		HandBrake 0.9.8 2012071700 HandBrake 0.9.8 2012071700
		file:/home/youp/Documents/9781783983926/Package/videos/video3_1.mp4

# Favorites

The favorites functionnality allows connected user to save search results that they appreciate or want to look at later on.

When activated, the search interface will be as follows:

Search - Firefox Developer Edition						🤶 F 🔀 💌 🕪) 13:04 🕇
🕻 🔺 Listeni 💥 Worki 🗦	🗶 Worki 🛛 Adr	in Da 🗴 Syntax MongoCli Admin Da 🌰 Sea 🗙	🖸 Outpu http:cted MongoCli 🛛	🗈 🕨 Fast 🔡 ping g 🔛 diag	gra 🞦 Copie 🖇 UML 2	2 🎲 Bing T 🕨 EDM > 🕂
🗧 🕈 🕘 localhost:8080/Datafari,	/Search?searchTy	e=allWords&query=	<del>-</del> ୯ ୧ ୨୦	arch	🖌 🏠 自 🔍 🦊	♠ ♥ ♥ ♥ ■ ■ ■
		<ul> <li>Tous les</li> </ul>	mots 🔵 Au moins un mot 🔵 Expression e	Q		Administration
Resultats 1 - 10 de 109 pou	ır *:* (0.502 Sec	ondes)				Tri par : 🛛 Pertinence 🗸
Dernière Modification	~	Reda Rahal SABIR A106, Avenue des Facultés Maison des	scientif			
Moins de un mois (0) Moins de un an (6)		file:/home/rsabir/Documents/testSearch/lettre de motivation.c unlike - 🕁 1	odt			
Moins de 5 ans (6)		T Enseirb - WebXML - Evaluation 1 - 2014-15.	pdf 📃			
Туре	~	file:/home/rsabir/DocumentEnseirb - WebXML - Evaluation unlike - 🖬 1	1 - 2014-15.pdf			
<ul> <li>pdf (42)</li> <li>xml (12)</li> <li>doc (9)</li> </ul>		lettre de resiliation.odt     Résiliation d u contrat N: 167483850 Bordeaux , I	۵			
odt (6)		file/home/rsabir/Documents/testSearch/lettre de resiliation.or unlike · 🖬 1	dt			
zip (3)	~	file:/home/rsabir/Documents/testSearch/crypt.doc				
file (108)	-	unlike -	Д			
Langage	~	file:/home/rsabir/Documents/testSearch/file/home/rsabir/Cv Like · 🖬 0	/_francais.pdf			
🗌 en (93)		. Attactation store at bouid bouls a "	_			
☐ fr (15)		attestation stage al barid bank.pdf file:/home/rsabir/Documents/attestation%20stage%20al%2 Like - 👍 0	Q0barid%20bank.pdf			

The only difference from a standard search result interface, is that it now includes the "favorite" icon next to each result title. This icon is a little red flag.

# $\square$

When the flag is empty, it means the corresponding document is not part of the user's favorites.



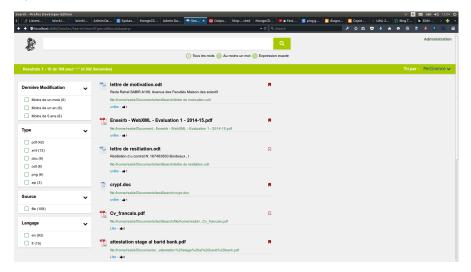
When the flag is fully red, it means it is part of the user's favorites. Clicking on the flag will add or delete the document into the favorites.

In the screenshot above, crypt.doc is saved as a favorite (fully red flag), Cv\_français.pdf is not (empty red flag).

# Highlighting

When doing a search, Datafari highlights parts of the results list. The highlight is set on terms of the user queries that can be found in the results list, be it in the title or in the snippet. This allows for a quick understanding of the document context for the query terms stand. **Likes** 

#### Inspired from social networks, Likes allows users to like selected search result entries. When activated, the search interface will be as follows:



The only difference from a standard search result interface, is that it now includes the "like" text below each result entry, as well as the number of likes, identified with a "thumbs-up" icon.

Like

The display of the "like" button means that the connected user has not liked the document yet.

### unlike

The display of the "unlike" button means that the connected user has already liked the document. By cliking on the (un)like button, you will switch the state of the document to liked or unliked.



Next to the (un)like button, the thumbs-up symbol indicates the number of cumulated likes for the document.

In the screenshot above, crypt.doc is liked and only one person likes this document.

#### **Promolinks**

Promolinks can be conceptually assimilated to google adwords: independently from the indexed content, it allows the search expert to display text in a dedicated place on the screen (usually on top or on the side). This display is triggered by the presence of specific terms in the search query.

For instance, the illustration below shows a promolink entitled "Nice title" which has the content "report" and is displayed because the keyword "report" was used in the search query:

report	Q
	O All words At least one word Exact expression
Results 0 - 0 de 0 pour report (0.004 Sec	onds)
Last Modification	
Less than 1 month (0) Less than 1 year (0) Less than 5 years (0)	report No result found
Туре	
Source	
Language	

Search expert: managing promolinks

In order to create Promolinks, you need to have the search expert role.

You have access to a dedicated web interface in the admin UI. First, go to the administration menu, click on Search Engine Configuration and select Capsule in the dropdown list.

Datafari
Lul Connectors
I Statistics
✤ Search engine administration
Search engine configuration
Capsules
Synonyms
Stopwords
Search Engine users
Servers

You then have access to the promolink admin UI, which shows the list of all existing promolinks. You can do the following operations on the existing promolinks:

- Delete by clicking the red button,
- Edit by clicking the keyword

Using the search bar on top, you can filter out the results by typing in something and pressing enter.

To add a promolink, use the form at the bottom of the page, and fill in the fields to add a promolink.

Note that the dates are optional, that you may fill only one out of those fields, and that if you choose to fill both, the starting date must be prior to then ending date. If dates have been filled during the add, Datafari will check if, when a search is made on the matching keyword, the current date is posterior to the starting date and prior to the ending date.

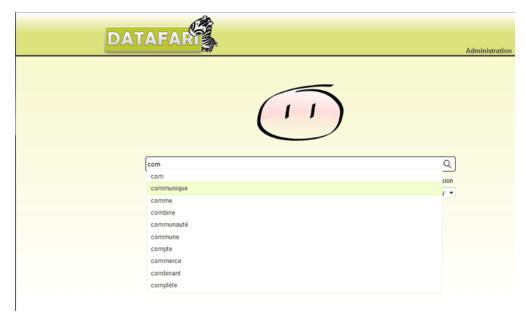
≡н	ome / Se	earch engine configuration	/ Capsules					8⁺f ≯in ằ
Sea	ch the ca	psules by keyword, if you w	ant all of them lea	ave 1				
Res	ults							
8	E List of th	he capsules						~ Z ×
#		Keyword :		Title :	Content :	Starting date :	Ending date :	
0		report		Nice title	report	07/01/2015	07/15/2015	0
Add	or edit	t a capsule						
		it a capsule						∧ Z
Para	ameter	ſS						
		Keyword :	Keyword					
		Title :	Title					
		Content :	Content					
Date	es							
		Starting date :	Date					
		Ending date :	Date					
				Confirm				

In the creation phase, if you type in a keyword that is already used as a keyword for another capsule, the existing capsule will be displayed and you can choose to replace it or not. This also means that you cannot have two capsules using the same keyword.

# Searching

The search phase is the most "intuitive" one, in a sense that it leverages user interface paradigms that are already widespread thanks to web search engines such as Google or Bing, and also thanks for ecommerce search engines such as eBay or Amazon. We explain briefly here the different capabilities that Datafari exposes, thanks to the graphical framework Ajaxfrancelabs, and of course thanks to the Apache Solr engine working behind the scene.

The first page you see when you use the search functionnality, is a simple search bar. Start typing text in it, and the autocomplete automatically starts proposing you terms that are present in the search index, ranked by "relevance" (the relevance here being mainly correlated to the terms frequency in the corpus). You can either continue typing, or at any point in time you can click on one of the proposed term. The autocomplete by default is a term by term autocomplete. If you select a term, and type a whitespace and start typing a second term, the autocomplete will work on the second term.



Once you've clicked the search icon, the display takes you to the search result page. This page is composed as follows: on the upper part, you get the standard search bar. On the left, you get the default facets to filter and navigate through the results. And on the right, you get the list of results ranked by relevance, related to the query terms. These panels use defaults views, and you can refer to the Ajaxfrancelabs framework documentation if you want to modify the display. The deeper the modifications, the more probable it is that you will need to modify also the configuration of the Solr of Datafari, in which case you should refer to the reference Apache Solr 4.8 documentation.

- Facets panel: by default, the following facets have been configured:
  - Last modifications: allows to filter based on time windows compared to the current date. Less than 1 month old documents, Less than 1 year old documents, Less than 5 years old documents.

- Type: allows to filter based on document types (pdf, doc, docx, xls, html ...)
- Source: allows to filter based on the source repository configured in Apache ManifoldCF
- Language: based on Solr capability to autodetect languages, allows to filter based on the documents language.
- Results list: by default, each result in this list is composed as follows:
  - Graphic icon symbolising the document type
  - Document file name in bold font / Title from the HTML header title in case of web page
  - Text snippet, 3 lines maximum, surrounding the query terms found in the document. These query terms are highlighted in bold fonts.
  - Document path in the source repository (or URL for web pages)

ATAFAR	Administ
communique	Q
	🗿 All words 🛑 At least one word 🛑 Exact expression
	Sort by : Relevancy -
sult 1 - 10 of 10 for comm	nunique (0.367 seconds)
st Modification	
Less than 1 month (0)	Communique_de_presse_FranceLabs_Datafari.docx
Less than 1 year (1)	formation sur les technologies de recherche Lucene / Solr, déjà utilisées par ebay, linkedin et twitter. Ses clients sont variés, et vont de la <b>commune</b> voulant équiper ses employés d'une solution de
Less than 5 years (10)	file:/D:/Temp/Sortie Datafari/Communique_de_presse_FranceLabs_Datafari.docx
Less man s years (10)	
vpe	Communique_de_presse_FranceLabs_B2B_Teched.pdf
docx (6)	commune voulant équiper ses employés d'une solution de recherche, aux sociétés de eCommerce en passant par les entreprises de veille et d'archivage. Sa dernière solution de recherche, Datafari,
pdf (4)	aide les
partity	file:/D:/Temp/SAP_TechedCommunique_de_presse_FranceLabs_B2B_Teched.pdf
ource	Communique de presse FranceLabs B2B Teched.docx
file (10)	formation sur les technologies de recherche Lucene / Solr, déjà utilisées par ebay, linkedin et twitter. Ses clients sont variés, et vont de la commune voulant équiper ses employés d'une solution de
anguage	file:/D:/Temp/Sortie DatCommunique_de_presse_FranceLabs_B2B_Teched.docx
fr (10)	
	Communique_de_presse_FranceLabs_SelectionTeched.pdf
	linkedin et twitter. Ses clients sont variés, et vont de la commune voulant équiper ses employés d'une solution de recherche, aux sociétés de eCommerce en passant par les entreprises de veille et
	file:/D:/Temp/SAP_TeCommunique_de_presse_FranceLabs_SelectionTeched.pdf
	Communique de presse FranceLabs SelectionTeched.docx
	Communique_ue_presse_rranceLabs_Selection reched.docx

twitter. Ses clients sont variés, et vont de la commune voulant équiper ses employés d'une solution de recherche, aux sociétés de eCommerce en passant par les entreprises de veille et d'archivage.

Facets have a standard way of working: if you click on a facet value, it will filter out all the results that don't satisfy the facet value condition. In the illustration below, selecting the pdf value in the Type facet only displays pdf documents in the result panel.

DATAFAR		Administration
communique	Q	
	<ul> <li>All words</li> <li>At least one word</li> <li>Exact expression</li> <li>Sort by : Relevancy</li> </ul>	
Result 1 - 4 of 4 for communi	que (0.049 seconds)	
Last Modification	- <del>11</del>	
Eess than 1 month (0)	Communique_de_presse_FranceLabs_B2B_Teched.pdf commune voulant équiper ses employés d'une solution de recherche, aux sociétés de eC	ommerce
Eess than 1 year (0)	en passant par les entreprises de veille et d'archivage. Sa dernière solution de recherche,	
Eess than 5 years (4)	aide les file:/D:/Temp/SAP_TechedCommunique_de_presse_FranceLabs_B26_Teched.pdf	
🛛 Туре	🔀 Communique_de_presse_FranceLabs_SelectionTeched.pdf	
docx (6)	linkedin et twitter. Ses clients sont variés, et vont de la commune voulant équiper ses empl	
🗸 pdf (4)	d'une solution de recherche, aux sociétés de eCommerce en passant par les entreprises et	de veille
	file:/D:/Temp/SAP_TeCommunique_de_presse_FranceLabs_SelectionTeched.pdf	
Source		
🧱 file (4)	20130703_Communique_presse- FranceLabs_Formations_Hadoop_Solr_BigData.pdf	
() anguaga	FranceLabs_Formations_Fladoop_Solf_BigData.pdf	
S Language	20130703_Communique_presse-FranceLabs_Formations_Hadoop_Solr_BigData.pdf	
fr (4)	-	
	🖾 Communique_de_presse-	
	FranceLabs_Gagnant_Categorie_SCS_V0.9-20121214.pdf	
	COMMUNIQUE DE PRESSE COMMUNIQUE DE PRESSE France Labs est le vainqueur du régional de	concours

...Communique\_de\_presse-FranceLabs\_Gagnant\_Categorie\_SCS\_V0.9-20121214.pdf

Datafari also proposes the spellchecker functionnality of Apache Solr. In the illustration below, we enter the query term "comminuque" instead of "communique". The spellchecker automatically proposes a correct word. By default, Datafari does not do a search on this suggestion, but it can do it with slight modifications.

DATAFAR			
	05		Administratio
comminuque		Q	
		All words At least one word Exact expression	
		Sort by : Relevancy -	
Result 0 - 0 of 0 for comminue	que (0.12 seconds)		
Last Modification			
Less than 1 month (0)	Try with : communique ?		
Less than 1 year (0)	No result found		
Less than 5 years (0)			
🛛 Type			
> >>			
Source			
> >>			
S Language			
> >>			

In the search bar, you can type in several query terms, and you can put operators to fine tune your search, such as AND and OR. Check the Apache Solr reference documentation for the full list of operators.

DATAFAR	Admin
communique pr	resse Q
	All words All east one word Exact expression
	Sort by : Relevancy -
Last Modification Less than 1 month (0)	20130703_Communique_presse- FranceLabs Formations Hadoop Solr BigData.pdf
	20130703 Communique presse-
	FranceLabs_Formations_Hadoop_Solr_BigData.pdf Hadoop et Constellio. Plus d?informations sur France Labs : www.francelabs.com Contact presse :
Less than 1 month (0)	FranceLabs_Formations_Hadoop_Solr_BigData.pdf
Less than 1 month (0) Less than 1 year (1) Less than 5 years (10)	FranceLabs_Formations_Hadoop_Solr_BigData.pdf Hadoop et Constellio. Plus d?informations sur France Labs : www.francelabs.com Contact presse : Cédric Ulmer Tél. : +33 4 92 91 14 27 Email :
Less than 1 month (0) Less than 1 year (1) Less than 5 years (10)	FranceLabs_Formations_Hadoop_Solr_BigData.pdf Hadoop et Constellio. Plus d?informations sur France Labs : www.francelabs.com Contact presse : Cédric Ulmer Tél. : +33 4 92 91 14 27 Email :
Less than 1 month (0) Less than 1 year (1) Less than 5 years (10)	FranceLabs_Formations_Hadoop_Solr_BigData.pdf Hadoop et Constellio. Plus d'informations sur France Labs : www.francelabs.com Contact presse : Cédric Ulmer Tél. : +33 4 92 91 14 27 Email :20130703_Communique_presse-FranceLabs_Formations_Hadoop_Solr_BigData.pdf
Less than 1 month (0) Less than 1 year (1) Less than 5 years (10)	FranceLabs_Formations_Hadoop_Solr_BigData.pdf Hadoop et Constellio. Plus d'informations sur France Labs : www.francelabs.com Contact presse : Cédric Ulmer Tél. : +33 4 92 91 14 27 Email :20130703_Communique_presse-FranceLabs_Formations_Hadoop_Solr_BigData.pdf 20130703_Communique_presse-
Less than 1 month (0) Less than 1 year (1) Less than 5 years (10) <b>Type</b> docx (6)	FranceLabs_Formations_Hadoop_Solr_BigData.pdf Hadoop et Constellio. Plus d'informations sur France Labs : www.francelabs.com Contact presse : Cédric Ulmer Tél. : +33 4 92 91 14 27 Email :20130703_Communique_presse-FranceLabs_Formations_Hadoop_Solr_BigData.pdf 20130703_Communique_presse- FranceLabs_Formations_Hadoop_Solr_BigData.docx Labs propose aux entreprises son expertise sur les couches open source Lucene, Solr, Hadoop et

Right below the search bar, three buttons allow to modify the search behaviour:

- All words: forces the search engine to return only the documents that contain ALL the terms present in the query term
- At least one word: that's the standard behavior. The search engine will favor documents that contain all the terms, but it not all terms are present, it will display the documents that contain a subset of the query terms
- Exact expression: forces the search engine to return only the documents that contain the exact expression in the query term. In the illustration below, only documents containing the exact expression "communique presse" will be return. So a document containing "presse communique" won't be returned.

DATAFARI			
communique pres	sse Q		
	All words 🛑 At least one word 🧿 Exact expression		
	Sort by : Relevancy -		
Result 1 - 2 of 2 for "communique presse" (0.06 seconds)			
Less than 1 month (0)	20130703_Communique_presse- FranceLabs_Formations_Hadoop_Solr_BigData.pdf		
Eess than 1 year (0)	France Labs Pal		
Less than 5 years (2)	20130703_Communique_presse-FranceLabs_Formations_Hadoop_Solr_BigData.pdf		
🛛 Туре	20130703_Communique_presse-		
docx (1)	FranceLabs_Formations_Hadoop_Solr_BigData.docx 20130703_Communique_presse-FranceLabs_Formations_Hadoop_Solr_BigData.docx		
🧱 pdf (1)			

# Stopwords

Stopwords allows you to ignore identified words at the indexing phase.

If a word listed in the stopword file is found, then Datafari will ignore it when doing the indexing of the corresponding document, except if they are only words listed as stopwords in the query.

# Alerts (from Datafari 3.1)

Since the version 3.1.x of Datafari, the Alerts feature has evolved. You can now create your alert directly from the search view of Datafari. Here are the steps to follow in order to create an alert:

- Authenticate yourself in Datafari
- Perform the search on which you want to create an alert
- Click on the "Create alert" button
- Enter you e-mail address in the corresponding field and select a send frequency
- Validate

#### [Picture]

You can now consult the alerts you have created in the "Parameters" view in the "Alerts" tab.

#### [Picture]

Here you can click on the "modify" button to change the send frequency of the corresponding alert or delete it.

# For developers

This section of the documentation is targeting a technical audience, more precisely developers who want to understand or extend the capacities of Datafari.

 $\stackrel{O}{=}$  For an overview of the available Datafari's build and deployment flows, have a look here Distributions HOW-TO

# Ajaxfrancelabs graphical framework

AjaxFranceLabs is a contribution from our company so that the community can manipulate Solr, Datafari and Constellio through an Ajax Framework.

Spiritually, it is a child of the AjaxSolr framework, and you will see many similarities, in particular at the architecture level. The AjaxSolr has been a great source of inspiration for us.

Still, the code itself has been done from scratch (except for 2 classes which are almost copy pastes), as there were parts that were rather different, because of the differences between Constellio and Datafari/Solr.

Although it is not part of the framework per se, we gave a particular care to the widgets that make use of our framework, because we know that you'll probably start with these widgets in order to get a hands on on our framework (you may even want to keep them as they are).

### Displaying thumbnails for autocomplete and results list

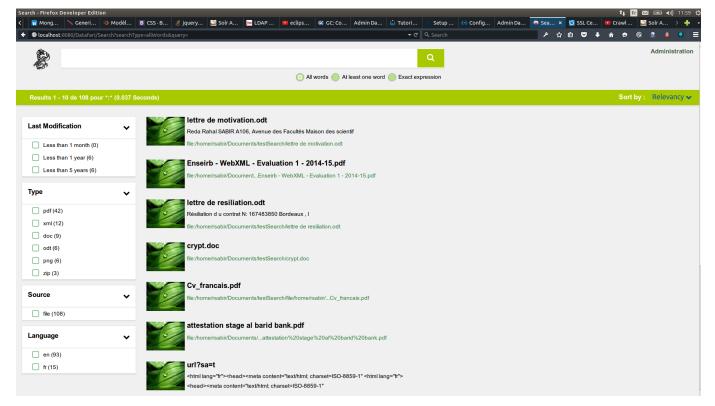
This documentation explains how to enable the « pictures » mode for the results display, for instance if you are indexing videos or pictures.

Note that this functionality is not complete yet, as of Datafari 2.0. For now the pointer to the image used is hardcoded. We still need to create a dedicated field in Solr that will hold the urls to the image, and to modify the widget so that it uses this field.

- 1. In searchView.jsp, you need to include « results-illustrate.css » which is in the css folder of datafari. Note that the css is already present, you just need to uncomment it.
- 2. In search.js, you need to replace resultWidget by with ResultIllustratedWidget, using the same attributes that resultWidget had.

This will make space for the image and show it at the left of the results in desktop mode and at the right of the results in the mobile mode. You can see below how it will appear :



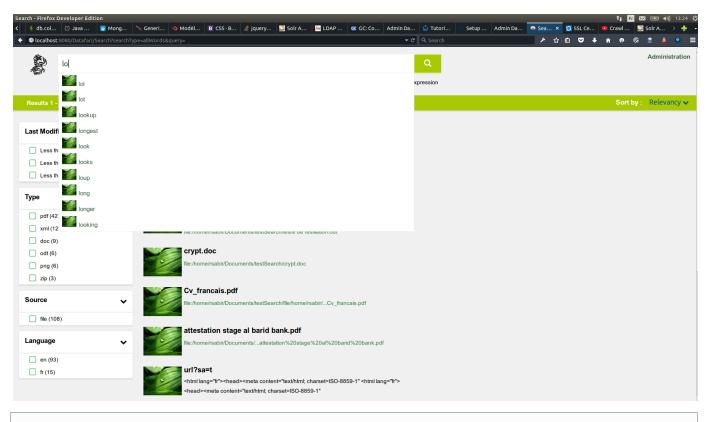


Now to have the pictures in the autocomplete functionnality, do the following :

1. In search.js, replace « AjaxFranceLabs.SearchBarWidget » with « AjaxFranceLabs.SearchBarIllustratedWidget », using the same parameters. And then you have it !

Note that this functionnality uses the following two javascripts, which are included by default « *js/AjaxFranceLabs/widgets/SearchBarlllustrated.wi dget.js* » and « *js/AjaxFranceLabs/modules/Autocompletelllustrated.module.js* » .

Note also that this functionality is a better fit for suggestions rather than for search autocomplete. For instance, in a digital asset management scenario, the autocomplete with the image would propose me when clicking on it directly the image rather than triggering a search on the solr. For now, we have not coded such a suggestion widget.



This widget is currently visually optimised for a fixed image size. If the source images do not have this size, we do a basic resize operation, which may lead to not-so-good-looking displays of the thumbnails.

# Framework architecture and mechanism

**Code Structure** 

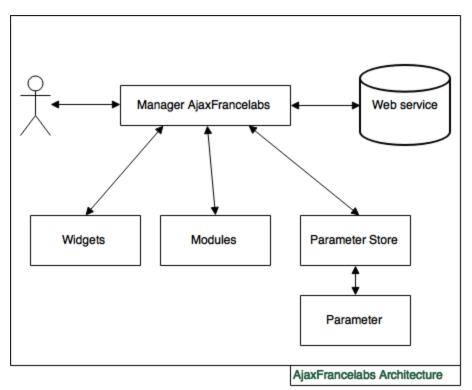
### The framework contains the following folders:

- core: the main classes of the framework
- manager: implemented managers
- **module**: the different modules
- widgets: the different interface widgets

A widget is a part of the user interface (UI) which is attached to a DOM element. A module is an extension of the widgets.

Eg: my result widget displays all the documents found by the search query and my favorite module will add "stars" to those documents title in order to save them to my favorite documents.

Architecture diagram



Interaction between the different components

The manager acts as a master controller. The widgets and modules are dealing with the manager in order to get the information they need and to modify the query request.

The parameter store stores the different parameter necessary to process the search query.

When a request query has to be processed, the manager calls a web service which receives the search query. This web service answers back to the manager, which stores the result in a variable (response) that the different components can access.

#### **Class diagram**

#### Core

AjaxFranceLabs offers different utility methods:

• string tinyUrl(url, length)

A method that reduces the length of the url given in parameter. Adds '...' at the end if the url have been tined. Leaves everything after the last / (page name) in the url. length is set to 75 by default.

string tinyString(str, length)

A method that reduces the length of the given string in parameter. Adds '...' at the end if the string have been tined. length is set to 25 by default.

• **string** truncate(string, length)

Reduces the size of a string if the length of it is superior at length. This method will not slice a word in two if the max length is reach but keep it and add '...' at the end.

<void addMultiElementClasses(source)

Adds the following classes to the different elements of the source:

- first | middle | last
- odd | even
- e-\* where \* is the index of the element in the source
- Thus you have all the necessary selectors for JavaScript and CSS.
- void clearMultiElementsClasses(source)

Remove the classes first, middle, last, even, odd, e-\* from the elements of source.

• boolean isString(obj)

Returns true if obj is a string.

• **boolean** isRegExp(obj)

Returns true if obj is a RegExp **boolean** equals(foo, bar)

- Compares the two objects given. Evaluates RegExp.
- boolean empty(mixed\_var)

Returns true if the mixed\_var is empty ('0', 0, null, undefined, false, "")

• string trim(str, charlist)

Removes the invisible characters on the beginning and end of the string str. charlist is an optional parameter.

string extractLast(term)

Returns the last term of term (used for the autocomplete widget).

• array split(val)

Splits val on the invisible character \u200c (used for the autocomplete widget).

• object extend(target, ...)

Extends the object target with the other arguments given.

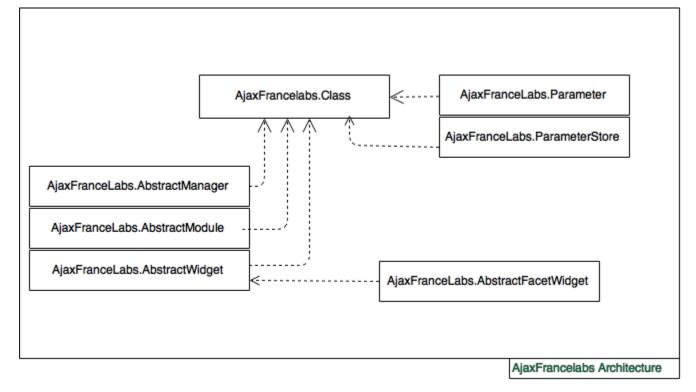
• string capitalize(string)

Sets the first letter of a word to upper case. The word is given as a parameter.

The main classes of the framework are:

- AbstractManager
- AbstractWidget
- AbstractFacetWidget
- AbstractModule
- Parameter
- ParameterStore

Any class class has to extend one of these or the AjaxFranceLabs.Class. For this, use the .extend method provided by the AjaxFranceLabs.Class and any of the previous one.



Architecture of the main classes of the framework

AbstractManager

The manager is the main class of the framework. It is the master controller, the crossroad of all the information going in and out. This class has to be extended in order to implement executeRequest() which have to make the call to the web service.

Insert architecture of the abstract manager here

AjaxFranceLabs.AbstractManager
<ul> <li>constellio</li> <li>serverUrl</li> <li>servelet</li> <li>response</li> <li>widgets</li> <li>modules</li> <li>store</li> <li>initialized</li> <li>collections</li> <li>connectionInfo</li> </ul>
+ init() + addModule(module) + addWidget(widget) + handleResponse() + makeRequest() + executeRequest()

Manager Class Diagream

AbstractManager class diagram

# Variables

- constellio: boolean set to true by default. Set to false when using Solr. When set to true, the manager will load your constellio collections
- serverUrl: address where the web service calls have to be made
- **servlet:** servlet that have to be used for the web service calls
- response: contains the last response from a search query on the search engine
- widgets: contains all the widgets of the manager
- modules: contains all the modules of the manager
- store: a store that contains all the parameters of the search query
- collections: the list of collections of the Constellio
- **collection:** the collection currently used

```
connectionInfo: {
  autocomplete: {
    serverUrl: 127.0.0.1,
    servlet: 'select',
    queryString: 'callFunction=getInformations&type=user
  }
}
```

- init: import collections if using constellio, initializes the widgets and modules
- · addModule: adds the module provided in parameter to the module collection ("array")
- addWidget: adds the widget provided in parameter to the widget collection ("array")
- handleResponse: executes the afterRequest method from all the widgets and modules after the search query has been processed
- makeRequest: executes the beforeRequest method from all the widgets and modules before processing the search query
- executeRequest: must be implemented in a Manager object, this is the method which calls the web service.
- getWidgetByID: gets the handle to the widget with ID specified as input.
- makeDefaultRequest: performs a select all (\*:\*) request and updates the address bar with the search query.

# Parameter and ParameterStore

Insert architecture parameter/parameterstore

AjaxFranceLabs.Parameter	AjaxFranceLabs.ParameterStore
- name - value - locals	- params
+ val(value) + local(name,value) + remove(name) + string() + parseString(str) + valueString(value) + parseValue(value) + escape(value)	+ isMultiple(name) + get(name) + values(name) + add(name,param) + remove(name,index) + addByValue(name,value,locals) + removeByValue(name,value) + string() + parseString(str) + find(name,value)

# Parameter and Parameter Store Class Diagram

Parameter and ParameterStore class diagram

# Parameter

#### Variables

- name: name of the parameter
- value: value of the parameter
- locals: local variables for nested request

#### Methods

- val: if a parameter is given, change the value of the parameter, return the value instead
- local: if value is given, changes the value of the local parameter, returns its value instead
- remove: removes a local parameter
- string: toString method of the parameter
- parseString: parses a string to create the parameter
- valueString: transforms value into a string if it is an array and encodes special characters
- parseValueString: decodes special characters and transforms the string in array if necessary
- escapeValue: quotes the string

# **ParameterStore**

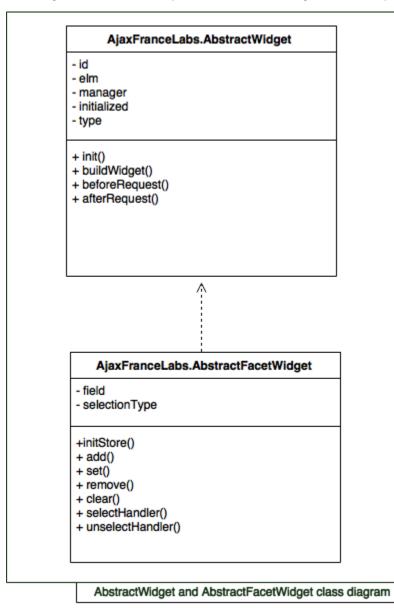
# Variables

• params: a collection of the parameters

- isMultiple: returns true if the parameter has multiple value
- get: returns the parameter, creates it if it does not exists
- values: return the value(s) of the parameter
- add: add a new parameter, creates an empty Parameter if only name is given
- remove: deletes a parameter
- addByValue: creates a new Parameter
- removeByValue: removes a parameter with the given name and value
- string: the toString method of the parameterStore that returns all the parameters on a string form
- parseString: parses a string to creates the parameters
- find: check if a parameter exists with the given name and value
- isParamDefined: helper function to check whether the given parameter name in input is defined or not inside the parameter store.

# AbstractWidget

All the widgets have to extend the AjaxFranceLabs.AbstractWidget class or the AjaxFranceLabs.AbstractFacetWidget class.



AbstractWidget and AbstractFacetWidget class diagram

# AbstractWidget

# Variables

- id: a unique id for the widget
- elm: element where the widget will be build
- **manager:** a reference to the widget's manager
- **type:** the type of the widget

- buildWidget: a method that build the main DOM structure of your widget.
- beforeRequest: code that have to be executed before the manager send the search query to the web service. This is mainly used when you have to add parameters to the search query, like the content of the search bar.
- afterRequest: code that will be executed after the search query is done. This is mainly used when you have to update the content of
  your widget, for example, empty previous result list and include the new result instead.

• reset: implement this method to reset the widget (notably after user input).

# AbstractFacetWidget

# Variables

- **field:** the referring field
- selectionType: can take 3 values depending on how you can select facet values:
  - AND: intersection of facet values
  - OR: document contains at least one facet value
  - ONE: can select one facet value at a time
- returnUnselectedFacetValues: if set to true, values for the unselected facet will be returned anyway (default: false)

# Methods

- selectHandler: callback method when a facet element is selected.
- unselectHandler: callback method when a facet element is unselected.

# AbstractModule

Modules are extension of widgets, they will modify them once they are built.

# AjaxFrancelabs.AbstractModule

- id

- manager

+ init()

- + beforeRequest()
- + afterRequest()

# AbstractModule class diagram

AbstractModule class diagram

# Variables

- id: a unique id for the module
- manager: a reference to the module's manager

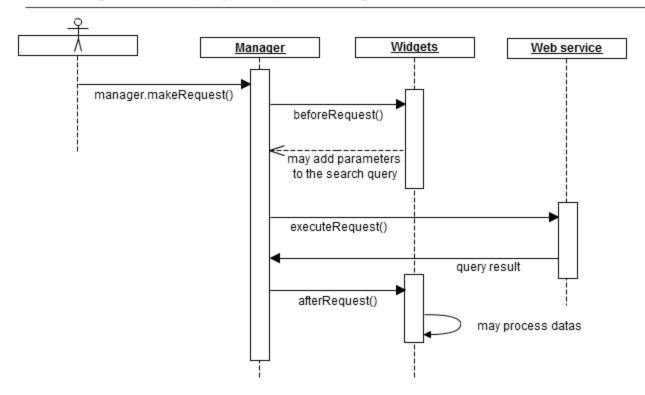
#### Methods

- beforeRequest: code that have to be executed before the manager send the search query to the web service.
- afterRequest: code that will be executed after the search query is done.

# Sequence diagram

Execution of a search query

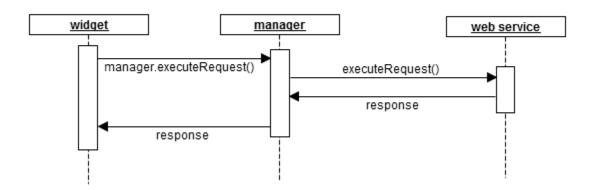
# Executing a search query - Sequence Diagram



- 1. The element (widget, external user or program) triggers a request to the search web service from the manager using the method **makeR** equest()
- 2. The manager first calls the **beforeRequest()** method of all the widgets and modules it controls. Thus the different widget can, for example, add parameters to the search query
- 3. The manager then executes the search query on the web service using the method executeRequest() and wait for its response
- 4. When the manager receives the response, it stores it into a local variable (reponse) accessible from all the widgets and modules it controls
- 5. Finally, the manager executes the afterRequest() method of all the widgets and modules it controls, thus, they can update their data

# Widget calling a web service

# Widget asking for informations to the web service - Sequence Diagram



In order to illustrate this example, I will take the case of the AutocompleteWidget

- 1. A user types some text in an input field which has an autocomplete widget attached to it
- 2. The autocomplete widget retrieves the data entered by the user into the field and triggers the **executeRequest()** method from its manager to retrive the suggestions

- 3. The manager answers to the autocomplete widget when it receives the results
- 4. The autocomplete widget displays the results

# Quickly changing the main colors

To customize the main colors of the UI, you should modify results-color.css and color.css :

- To change the green that it is used as background, you should change the attribute background of .bc-color in *color.css*. You should also change the footer.bc-color. We have separated the color of the footer from the others because of the effect of the shadow that impacts the apparent color.
- To change the green used in the pagination and the checkbox of the facets, you should change the line 126 in *color.css* where the css selectors are :

```
#facets .checkboxIcon,
#facets_mobile input[type="checkbox"]:checked + label
span.checkboxIcon,
.pagerModule .pages .page,
.pagerModule i
```

- To change the green used in the Searchbox icons (all words, at least one word), edit the file *check\_radio\_sheet.png* located in images folder with a graphics editor.
- To change the color of each document's link, edit the file results-color.css. Change the color attribue at the line :

```
.resultWidget .doc .res .address { color: #378140; }
```

# **Tutorial**

#### This demonstration is an introduction to the framework AjaxFranceLabs and will demonstrate how to use it.

In order to do that, you will need to download the AjaxFranceLabs framework and include it to your workspace.

Here are the first steps:

- 1. Download the latest update of the version 1 of jQuery on jQuery's website
- 2. Create a main.js file where we will add all our code to create the interface
- 3. Include this two JavaScript files to the html page
- 4. Include the main.css file, the result.css from the framework and the font awesome library as bellow.

In a first place, we will create the AjaxFranceLabs manager, configure it and include the searchBar widget which will allow us to process searches using keywords.

We are going to include the required files as well as their dependencies,

```
<!-- index.html - Adding a SearchBar widget -->
 <!DOCTYPE html>
 <html lang="en">
                   <head>
                                      <link rel="stylesheet" type="text/css" href="css/main.css">
                                      <link rel="stylesheet" type="text/css" href="css/results.css">
                                      <link rel="stylesheet"
href="//maxcdn.bootstrapcdn.com/font-awesome/4.3.0/css/font-awesome.min.cs
s">
                  </head>
                  <body>
                                      <script type="text/javascript" src="js/jquery-1.8.1.min.js"></script>
                                      <script type="text/javascript"</pre>
src="js/jquery-ui-1.8.23.min.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script><
                                      <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/uuid.core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                      <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/i18njs.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                                      <!-- The framework main file -->
                                      <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/Core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                      <!-- Required file for the manager -->
                                      <script type="text/javascript"
src="js/AjaxFranceLabs/core/Parameter.js"></script></script></script></script></script></script></script>
                                      <!-- Required file for the manager -->
                                      <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/ParameterStore.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                      <!-- The abstract manager -->
                                      <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractManager.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></sc
                                      <!-- An implementation of the manager -->
                                      <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/manager/Manager.js"></script>
                                      <!-- The AbstractWidget file -->
                                      <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractWidget.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                      <!-- The SearchBar widget, implementing AbstractWidget -->
                                      <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/widgets/SearchBar.widget.js"></script>
                                      <script type="text/javascript" src="js/main.js"></script>
                                      <script type="text/javascript" src="js/searchTutorial.js"></script>
                                                                                             <div id="solr">
                                                                                                                 <div id="searchBar"></div><!-- the widget will be inserted here</pre>
  -->
                                                                                             </div>
                                                                           </body>
                                     </html>
```

Now let's create the main.js file .

```
/* main.js */
$(function($){
 /* Now, we had the widget to the manager */
 //Version 1 - if you do not need to have a variable refering to the widget
 Manager.addWidget(
    new AjaxFranceLabs.SearchBarWidget({
      elm: $('#searchBar'),
         id: 'searchBar'
     })
 );
    //Version 2
    var searchBarWidget = new AjaxFranceLabs.SearchBarWidget({
        elm: $('#searchBar'),
        id: 'searchBar'
    });
    Manager.addWidget(searchBarWidget);
    /* And we initialize the manager */
    Manager.init();
});
```

```
The display should look like something like this:
```

Firefox Developer Edition http://localari/test.jsp × 💠		② F ■ ● (* 14:37 役)
🔶 💿 localhost:8080/Datafari/test.jsp	← C Q Search	▶ ☆ 自 ♥ ↓ 合 巻 ◎ 📓 💄 ♥ 🚍
	Q	
G	All words O At least one word O Exact expression	
		Sort by : Relevancy 🗸

```
In the html file
```

```
<!-- index.html - Adding a Result widget -->
<!DOCTYPE html>
 <html lang="en">
                       <head>
                                                <link rel="stylesheet" type="text/css" href="css/main.css">
                                                <link rel="stylesheet" type="text/css" href="css/results.css">
                                                <link rel="stylesheet"
href="//maxcdn.bootstrapcdn.com/font-awesome/4.3.0/css/font-awesome.min.cs
s">
                       </head>
                       <body>
                                                <script type="text/javascript" src="js/jquery-1.8.1.min.js"></script>
                                                <script type="text/javascript"</pre>
src="js/jquery-ui-1.8.23.min.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script><
                                                <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/uuid.core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                                <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/i18njs.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                                                <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/Core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                                <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/Parameter.js"></script></script></script></script></script></script></script>
                                                <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/ParameterStore.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                                <script type="text/javascript"
src="js/AjaxFranceLabs/core/AbstractManager.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></sc
                                                <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/manager/Manager.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                                                <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractWidget.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                                <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/widgets/SearchBar.widget.js"></script>
                                                <script type="text/javascript" src="js/main.js"></script>
                                                <!-- The ResultWidget -->
                                                <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/widgets/Result.widget.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></s
                                                <script type="text/javascript" src="js/searchTutorial.js"></script>
                                                <div id="solr">
                                                                       <div id="searchBar"></div><!-- the widget will be inserted here -->
                                                                       <div class="col left"></div>
                                                                       <div class="col right">
                                                <div id="results"></div><!-- the widget will be inserted here -->
                                                                       </div>
                                                                       <div class="clear"></div>
                                              </div>
                       </body>
 </html>
```

```
/* searchTutorial.js */
$(function($){
    Manager.addWidget(new AjaxFranceLabs.SearchBarWidget({
        elm: $('#searchBar'),
        id: 'searchBar'
    }));
    /* The result widget */
    Manager.addWidget(new AjaxFranceLabs.ResultWidget({
        elm: $('#results'),
        id: 'documents',
        afterRequest : function() { /* implements the method displaying the
documents */
           var data = this.manager.response;
            $(this.elm).find('.doc_list').empty();
            if (data.response.numFound === 0) {
                $(this.elm).find('.doc_list').append('<span</pre>
class="noResult">No document found.</span>');
            } else {
                var self = this;
                $.each(data.response.docs, function(i, doc) {
                    var title = (doc.title === undefined) ? doc.url :
doc.title;
                    if(data.highlighting === undefined ||
(data.highlighting[doc.id].content_fr === undefined &&
data.highlighting[doc.id].content_en === undefined)){
                        if(doc.content_fr)
                            content = doc.content_fr;
                        else
                            content = doc.content_en;
                        content = AjaxFranceLabs.truncate(content, 200)
                    }
                    else {
                        if(data.highlighting[doc.id].content fr)
                            content = data.highlighting[doc.id].content_fr;
                        else
                            content = data.highlighting[doc.id].content_en;
                    $(self.elm).find('.doc_list').append('<div class="doc</pre>
e-' + i + '" doc_id="'+doc.id+'">');
                    $(self.elm).find('.doc:last').append('<div</pre>
class="res">');
                    $(self.elm).find('.doc:last .res')
                        .append('<a class="title" target="_blank" href="' +
doc.url + '">')
                        .append('')
                        .append('');
                    $(self.elm).find('.doc:last .title')
                        .append('<span>' + title + '</span>');
                    $(self.elm).find('.doc:last .description')
                        .append('<span>' + content + '</span>');
                    $(self.elm).find('.doc:last .address')
```

/\* Lets simulate an empty search query in order to fill the results when loading the page  $\,^{*/}$ 

/\* The makeRequest method actually initializes the manager if this has not been done before \*/

Manager.makeRequest();

});

# You should get the following result:

Firefox Developer Edition		🤤 Fr 🖂 📧 🕸 🕸
http://localari/test.jsp 🗙 🕂		
🗧 💿 localhost:8080/Datafari/test.jsp	← ♂ Q Search	ト ☆ 自 マ キ 合 😕 📓 💄 🔹 🚍
	🗿 All words 🥚 At least one word 🔴 Exact expression	
		Sort by : Relevancy 🗸
	lettre de motivation.odt	
	Reda Rahal SABIR A106, Avenue des Facultés Maison des scientif	
	file:/home/rsabir/Documents/testSearch/lettre%20de%20motivation.odt	
	Diagrame de classe.png	
	file:/home/rsabir/Documents/testSearch/Diagrame%20de%20classe.png	
	PPetri1.jpp	
	file:/home/rsabir/Documents/testSearch/PPetri1.jpp	
	attestation dactivité.pdf	
	Expéditeur : Mme Amilha	
	file:/home/rsabir/Documents/testSearch/attestation%20dactivité.pdf	
	le recit des decouvertes.docx	
	file:/home/rsabir/Documents/testSearch/le%20recit%20des%20decouvertes.docx	
	index.png	
	file:/home/rsabir/Documents/testSearch/index.png	
	Practical-Work-Tank_CORRIGEE.jtr	
	Test du reservoir, utilisation des événements répétés, calcul	
	file:/home/rsabir/Documents/testSearch/Practical-Work-Tank_CORRIGEE.jtr	
	lettre de resiliation.odt	
	Résiliation d u contrat N: 167483850 Bordeaux , I	

Now we have our search bar and the results displayed, why not adding the search informations like the number of documents returned, the query execution time. Lets add the SearchInformation widget.

# The html file

\_\_\_\_\_

```
<!-- index.html - Adding a SearchInformation widget -->
  <!DOCTYPE html>
 <html lang="en">
                  <head>
                                     <link rel="stylesheet" type="text/css" href="css/main.css">
                                     <link rel="stylesheet" type="text/css" href="css/results.css">
                                     <link rel="stylesheet"
href="//maxcdn.bootstrapcdn.com/font-awesome/4.3.0/css/font-awesome.min.cs
 s">
                  </head>
                  <body>
                                     <script type="text/javascript" src="js/jquery-1.8.1.min.js"></script>
                                     <script type="text/javascript"</pre>
  src="js/jquery-ui-1.8.23.min.js"></script>
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/uuid.core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/i18njs.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                                     <script type="text/javascript"
 src="js/AjaxFranceLabs/core/Core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/core/Parameter.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script><
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/core/ParameterStore.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/core/AbstractManager.js"></script>
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/manager/Manager.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/core/AbstractWidget.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/widgets/SearchBar.widget.js"></script>
                                     <script type="text/javascript" src="js/main.js"></script>
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/widgets/Result.widget.js"></script>
                                     <!-- The SearchInformationWidget -->
                                     <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/widgets/SearchInformation.widget.js"></script>
                                     <script type="text/javascript" src="js/searchTutorial.js"></script>
                                     <div id="solr">
                                                               <div id="searchBar"></div>
                                                              <div id="result information"></div><!-- the widget will be inserted</pre>
here -->
                                                              <div class="col left"></div>
                                                              <div class="col right">
                                                                   <div id="results"></div>
                                                              </div>
                                                               <div class="clear"></div>
                                    </div>
                  </body>
  </html>
```

And the javascript file :

```
/* searchTutorial.js */
$(function($){
    Manager.addWidget(new AjaxFranceLabs.SearchBarWidget({
        elm: $('#searchBar'),
        id: 'searchBar'
    }));
   Manager.addWidget(new AjaxFranceLabs.ResultWidget({
        elm: $('#results'),
        id: 'documents',
        afterRequest : function() { /* implements the method displaying the
documents */
           var data = this.manager.response;
            $(this.elm).find('.doc_list').empty();
            if (data.response.numFound === 0) {
                $(this.elm).find('.doc_list').append('<span</pre>
class="noResult">No document found.</span>');
            } else {
                var self = this;
                $.each(data.response.docs, function(i, doc) {
                    var title = (doc.title === undefined) ? doc.url :
doc.title;
                    if(data.highlighting === undefined ||
(data.highlighting[doc.id].content_fr === undefined &&
data.highlighting[doc.id].content_en === undefined)){
                        if(doc.content_fr)
                            content = doc.content fr;
                        else
                            content = doc.content_en;
                        content = AjaxFranceLabs.truncate(content, 200)
                    }
                    else {
                        if(data.highlighting[doc.id].content_fr)
                            content = data.highlighting[doc.id].content_fr;
                        else
                            content = data.highlighting[doc.id].content_en;
                    $(self.elm).find('.doc_list').append('<div class="doc</pre>
e-' + i + '" doc id="'+doc.id+'">');
                    $(self.elm).find('.doc:last').append('<div</pre>
class="res">');
                    $(self.elm).find('.doc:last .res')
                        .append('<a class="title" target="_blank" href="' +
doc.url + '">')
                        .append('')
                        .append('');
                    $(self.elm).find('.doc:last .title')
                        .append('<span>' + title + '</span>');
                    $(self.elm).find('.doc:last .description')
                        .append('<span>' + content + '</span>');
```

```
$(self.elm).find('.doc:last .address')
                        .append('<span>' + AjaxFranceLabs.tinyUrl(doc.url)
+ '</span>');
                });
AjaxFranceLabs.addMultiElementClasses($(this.elm).find('.doc'));
                if (this.pagination !== false) {
                    this.pagination.afterRequest(data);
                }
            }
        }
    }));
    /* The SearchInformation widget */
    Manager.addWidget(new AjaxFranceLabs.SearchInformationWidget({
        elm: $('#result_information'),
        id: 'searchInformation'
    }));
    Manager.makeRequest();
```

});
You should get the following display:

Firefox Developer Edition					🗟 🖪 🤤		≪ 14	1:31 3
http://localari/test.jsp 🗙 🕂								
🗧 💿 localhost:8080/Datafari/test.jsp		ਦ Q Search	▶ ☆ 自	 â	9 🖗		<b>a</b> 🔹	
	c	λ						
	All words At least one word Exact express	sion						
Results 1 - 10 de 117 pour *:* (0.058 Se	ronde)				Sort by :	Rele	evancy	
	lettre de motivation.odt							
	Reda Rahal SABIR A106, Avenue des Facultés Maison des scientif							
	file:/home/rsabir/Documents/testSearch/lettre%20de%20motivation.odt							
	Diagrame de classe.png							
	file:/home/rsabir/Documents/testSearch/Diagrame%20de%20classe.png							
	PPetri1.jpp							
	file:/home/rsabir/Documents/testSearch/PPetri1.jpp							
	attestation dactivité.pdf							
	Expéditeur : Mme Amilha							
	file:/home/rsabir/Documents/testSearch/attestation%20dactivité.pdf							
	le recit des decouvertes.docx							
	file:/home/rsabir/Documents/testSearch/le%20recit%20des%20decouvertes.docx							
	index.png							
	file:/home/rsabir/Documents/testSearch/index.png							
	ne.nonersabir/boodinentareatoealorandex.prg							
	Practical-Work-Tank_CORRIGEE.jtr							
	Test du reservoir, utilisation des événements répétés, calcul							
	file:/home/rsabir/Documents/testSearch/Practical-Work-Tank_CORRIGEE.jtr							
	lettre de resiliation.odt Résiliation du contrat N: 167483850 Rordeaux							

For this part of the tutorial, we will continue by adding the pagination feature.

The html file:

```
<!-- index.html - Adding the pagination -->
<!DOCTYPE html>
<html lang="en">
               <head>
                             <link rel="stylesheet" type="text/css" href="css/main.css">
                             <link rel="stylesheet" type="text/css" href="css/results.css">
                             <link rel="stylesheet"
href="//maxcdn.bootstrapcdn.com/font-awesome/4.3.0/css/font-awesome.min.cs
s">
               </head>
               <body>
                              <script type="text/javascript" src="js/jquery-1.8.1.min.js"></script>
                             <script type="text/javascript"
src="js/jquery-ui-1.8.23.min.js"></script>
                             <script type="text/javascript"
src="js/AjaxFranceLabs/uuid.core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                             <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/i18njs.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                             <script type="text/javascript"
```

```
src="js/AjaxFranceLabs/core/Core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                                           <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/core/Parameter.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script><
                                                           <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/core/ParameterStore.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                                           <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/core/AbstractManager.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></sc
                                                         <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/manager/Manager.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                                                           <!-- Add this class -->
                                                           <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/core/AbstractModule.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                                           <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/core/AbstractWidget.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                                           <!-- Add this module -->
                                                           <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/modules/Pager.module.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></sc
                                                           <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/widgets/SearchBar.widget.js"></script>
                                                           <script type="text/javascript" src="js/main.js"></script>
                                                           <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/widgets/Result.widget.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></s
                                                           <!-- The SearchInformationWidget -->
                                                           <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/widgets/SearchInformation.widget.js"></script>
                                                           <script type="text/javascript" src="js/searchTutorial.js"></script>
                                                           <div id="solr">
                                                                                       <div id="searchBar"></div>
                                                                                      <div id="result_information"></div> <!-- the widget will be inserted</pre>
here -->
                                                                                      <div class="col left"></div>
                                                                                      <div class="col right">
                                                         <div id="results"></div>
                                                                                      </div>
                                                                                       <div class="clear"></div>
                                                           </div>
                               </body>
```

#### And the javascript file

```
/* searchTutorial.js */
$(function($){
Manager.addWidget(new AjaxFranceLabs.SearchBarWidget({
    elm: $('#searchBar'),
    id: 'searchBar'
}));
/* The result widget */
Manager.addWidget(new AjaxFranceLabs.ResultWidget({
    elm: $('#results'),
    id: 'documents',
   pagination : true, /* Add this */
    afterRequest : function() {
        var data = this.manager.response;
        $(this.elm).find('.doc list').empty();
        if (data.response.numFound === 0) {
            $(this.elm).find('.doc_list').append('<span class="noResult">No
document found.</span>');
        } else {
           var self = this;
            $.each(data.response.docs, function(i, doc) {
                var title = (doc.title === undefined) ? doc.url :
doc.title;
                if(data.highlighting === undefined ||
(data.highlighting[doc.id].content_fr === undefined &&
data.highlighting[doc.id].content_en === undefined)){
                    if(doc.content fr)
                        content = doc.content_fr;
                    else
                        content = doc.content_en;
                    content = AjaxFranceLabs.truncate(content, 200)
                }
                else {
                    if(data.highlighting[doc.id].content_fr)
                        content = data.highlighting[doc.id].content_fr;
                    else
                        content = data.highlighting[doc.id].content_en;
                }
                $(self.elm).find('.doc list').append('<div class="doc e-' +</pre>
i + '" doc_id="'+doc.id+'">');
                $(self.elm).find('.doc:last').append('<div class="res">');
                $(self.elm).find('.doc:last .res')
                    .append('<a class="title" target="_blank" href="' +
doc.url + '">')
                    .append('')
                    .append('');
```

```
$(self.elm).find('.doc:last .title')
                    .append('<span>' + title + '</span>');
                $(self.elm).find('.doc:last .description')
                    .append('<span>' + content + '</span>');
                $(self.elm).find('.doc:last .address')
                    .append('<span>' + AjaxFranceLabs.tinyUrl(doc.url) +
'</span>');
            });
AjaxFranceLabs.addMultiElementClasses($(this.elm).find('.doc'));
            if (this.pagination !== false) {
                this.pagination.afterRequest(data);
            }
        }
    }
}));
Manager.addWidget(new AjaxFranceLabs.SearchInformationWidget({
    elm: $('#result_information'),
    id: 'searchInformation'
}));
```

```
Manager.makeRequest();
```

});

# You should get the following result:

Firefox Developer Edition	🤝 🇗 🐱 🖜 💷 🕸
• Olocalhost:8080/Datafari/test.jsp	- C 9. Search 🛛 🖉 🖡 🏟 🛞 🔝 🛔 🔹 🚍
	PPetri1.jpp
	file:/home/rsabir/Documents/testSearch/PPetri1.jpp
	attestation dactivité.pdf
	Expéditeur : Mme Amilha
	file:/home/rsabir/Documents/testSearch/attestation%20dactivité.pdf
	le recit des decouvertes.docx
	file:/home/rsabir/Documents/testSearch/le%20recit%20des%20decouvertes.docx
	index.png
	file:/home/rsabir/Documents/testSearch/index.png
	Practical-Work-Tank_CORRIGEE.jtr
	Test du reservoir, utilisation des événements répétés, calcul
	file:/home/rsabir/Documents/testSearch/Practical-Work-Tank_CORRIGEE.jtr
	lettre de resiliation.odt
	Résiliation d u contrat N: 167483850 Bordeaux , I
	file:/home/rsabir/Documents/testSearch/lettre%20de%20resiliation.odt
	BChapter_2.doc
	file:/home/rsabir/Documents/testSearch/BChapter_2.doc
	Enseirb - WebXML - Evaluation 1 - 2014-15.pdf ENSEIRB 0
	file:/home/rEnseirb%20-%20WebXML%20-%20Evaluation%201%20-%202014-15.pdf
	1 2 3 4 5 6 7 8 9 10 >>>

Easy right? Regarding to the documentation, the pagination variable can take 3 values, true, false, a PagerWidget. By default this value is set to true and will automatically create a PagerWidget with the default options. But, if you want to modify some methods or values of the widget, you can create another one and assign it to the pagination variable.

# The next thing we will do is including the autocomplete widget to the search bar.

The autocomplete widget is a direct extension of the autocomplete plugin from jQuery UI, so we will need to download it and include it to our javascript files.

```
src="js/jquery-ui-1.8.23.min.js"></script>
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/uuid.core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/il8njs.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/Core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                           <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/Parameter.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script><
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/ParameterStore.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractManager.js"></script>
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/manager/Manager.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractModule.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractWidget.js"></script>
                                           <!-- Add this module -->
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/modules/Autocomplete.module.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/modules/Pager.module.js"></script>
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/widgets/SearchBar.widget.js"></script>
                                            <script type="text/javascript" src="js/main.js"></script>
                                            <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/widgets/Result.widget.js"></script>
                                            <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/widgets/SearchInformation.widget.js"></script>
                                            <script type="text/javascript" src="js/searchTutorial.js"></script>
                                            <div id="solr">
                                                                <div id="searchBar"></div>
                                                                  <div id="result_information"></div><!-- the widget will be inserted</pre>
here -->
                                                                  <div class="col left"></div>
                                                                <div class="col right">
                                            <div id="results"></div>
                                                                </div>
                                                                  <div class="clear"></div>
                                            </div>
                     </body>
```

```
</html>
```

# And the javascript file

```
/* searchTutorial.js */
$(function($){
    Manager.addWidget(new AjaxFranceLabs.SearchBarWidget({
        elm: $('#searchBar'),
        id: 'searchBar',
        autocomplete : true //add this line for the autocompletion
    }));
    Manager.addWidget(new AjaxFranceLabs.ResultWidget({
        elm: $('#results'),
        id: 'documents',
        pagination : true,
        afterRequest : function() {
            var data = this.manager.response;
            $(this.elm).find('.doc_list').empty();
            if (data.response.numFound === 0) {
                $(this.elm).find('.doc_list').append('<span</pre>
class="noResult">No document found.</span>');
            } else {
                var self = this;
                $.each(data.response.docs, function(i, doc) {
                    var title = (doc.title === undefined) ? doc.url :
doc.title;
                    if(data.highlighting === undefined ||
(data.highlighting[doc.id].content_fr === undefined &&
data.highlighting[doc.id].content_en === undefined)){
                        if(doc.content_fr)
                            content = doc.content_fr;
                        else
                            content = doc.content_en;
                        content = AjaxFranceLabs.truncate(content, 200)
                    }
                    else {
                        if(data.highlighting[doc.id].content_fr)
                            content = data.highlighting[doc.id].content_fr;
                        else
                            content = data.highlighting[doc.id].content_en;
                    $(self.elm).find('.doc_list').append('<div class="doc</pre>
e-' + i + '" doc_id="'+doc.id+'">');
                    $(self.elm).find('.doc:last').append('<div</pre>
class="res">');
                    $(self.elm).find('.doc:last .res')
                         .append('<a class="title" target="_blank" href="' +
doc.url + '">')
```

```
.append('')
                       .append('');
                   $(self.elm).find('.doc:last .title')
                       .append('<span>' + title + '</span>');
                   $(self.elm).find('.doc:last .description')
                       .append('<span>' + content + '</span>');
                   $(self.elm).find('.doc:last .address')
                       .append('<span>' + AjaxFranceLabs.tinyUrl(doc.url)
+ '</span>');
               });
AjaxFranceLabs.addMultiElementClasses($(this.elm).find('.doc'));
               if (this.pagination !== false) {
                   this.pagination.afterRequest(data);
               }
           }
       }
   }));
   Manager.addWidget(new AjaxFranceLabs.SearchInformationWidget({
       elm: $('#result_information'),
       id: 'searchInformation'
    }));
   Manager.makeRequest();
```

});

# You should get the following:

	fox Developer Edition							(î:	Fr 🖂		4  ×	14:27 🏌
	://localari/test.jsp 🗙 🕂											
÷	localhost:8080/Datafari/test.jsp		→ <sup> C</sup> <sup> Q</sup> <sup> Search <sup> </sup> </sup>	۶	습   I	• •	Â	₿	۲		•	●   =
	le		Q									
	let		xpression									
_	least											
	length								t by :	Rele	evan	cy 🗸
	less											
	les											
	level											
	letter											
	lengths											
	levels	file:/home/rsabir/Documents/testSearch/Diagrame%20de%20classe.png										
		nie./nome/rsabir/bocuments/testSearch/blagrame %20de %20classe.png										
		PPetri1.jpp										
		file:/home/rsabir/Documents/testSearch/PPetri1.jpp										
		attestation dactivité.pdf										
		Expéditeur : Mme Amilha										
		file:/home/rsabir/Documents/testSearch/attestation%20dactivité.pdf										
		le recit des decouvertes.docx										
		file:/home/rsabir/Documents/testSearch/le%20recit%20des%20decouvertes.docx										
		index.png										
		file:/home/rsabir/Documents/testSearch/index.png										
		Practical-Work-Tank_CORRIGEE.jtr										
		Test du reservoir, utilisation des événements répétés, calcul										
		file:/home/rsabir/Documents/testSearch/Practical-Work-Tank_CORRIGEE.jtr										
		lettre de resiliation.odt										
		Résiliation d u contrat N: 167483850 Bordeaux , I										

Easy right? Regarding to the documentation, the pagination variable can take 3 values, true, false, a PagerWidget. By default this value is set to true and will automatically create a PagerWidget with the default options. But, if you want to modify some methods or values of the widget, you can create another one and assign it to the pagination variable.

#### Lets go a bit further now and add faceting.

#### The html file

```
src="js/AjaxFranceLabs/uuid.core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                                   <script type="text/javascript"
src="js/AjaxFranceLabs/il8njs.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script>
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/Core.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/Parameter.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script><
                                                  <script type="text/javascript"
src="js/AjaxFranceLabs/core/ParameterStore.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractManager.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></sc
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/manager/Manager.js"></script>
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractModule.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractWidget.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scr
                                                   <!-- Add this class which our widget extends -->
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/core/AbstractFacetWidget.js"></script>
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/modules/Autocomplete.module.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></scri
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/modules/Pager.module.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></sc
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/widgets/SearchBar.widget.js"></script>
                                                   <script type="text/javascript" src="js/main.js"></script>
                                                   <script type="text/javascript"</pre>
 src="js/AjaxFranceLabs/widgets/Result.widget.js"></script>
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/widgets/SearchInformation.widget.js"></script>
                                                   <!-- Add this widget -->
                                                   <script type="text/javascript"</pre>
src="js/AjaxFranceLabs/widgets/Table.widget.js"></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></script></sc
                                                   <script type="text/javascript" src="js/searchTutorial.js"></script>
                                                   <div id="solr">
                                                                             <div id="searchBar"></div>
                                                                             <div id="result information"></div>
                                                                             <div class="col left">
                                                   <!-- Add this block -->
                                                   <div id="facets">
                                                                             <div id="facet_langue"></div><!-- First widget -->
                                                                           <div id="facet source"></div><!-- Second widget -->
                                                   </div>
                                                                           </div>
                                                                             <div class="col right">
                                                   <div id="results"></div>
                                                                             </div>
                                                                             <div class="clear"></div>
                                                   </div>
                          </body>
```

```
</html>
```

# And the javascript file:

```
/* searchTutorial.js */
$(function($){
    Manager.addWidget(new AjaxFranceLabs.SearchBarWidget({
        elm: $('#searchBar'),
        id: 'searchBar',
        autocomplete : true //add this line for the autocompletion
    }));
    Manager.addWidget(new AjaxFranceLabs.ResultWidget({
        elm: $('#results'),
        id: 'documents',
        pagination : true,
        afterRequest : function() {
            var data = this.manager.response;
            $(this.elm).find('.doc_list').empty();
            if (data.response.numFound === 0) {
                $(this.elm).find('.doc_list').append('<span</pre>
class="noResult">No document found.</span>');
            } else {
                var self = this;
                $.each(data.response.docs, function(i, doc) {
                    var title = (doc.title === undefined) ? doc.url :
doc.title;
                    if(data.highlighting === undefined ||
(data.highlighting[doc.id].content_fr === undefined &&
data.highlighting[doc.id].content_en === undefined)){
                        if(doc.content_fr)
                            content = doc.content_fr;
                        else
                            content = doc.content_en;
                        content = AjaxFranceLabs.truncate(content, 200)
                    }
                    else {
                        if(data.highlighting[doc.id].content_fr)
                            content = data.highlighting[doc.id].content_fr;
                        else
                            content = data.highlighting[doc.id].content_en;
                    $(self.elm).find('.doc_list').append('<div class="doc</pre>
e-' + i + '" doc_id="'+doc.id+'">');
                    $(self.elm).find('.doc:last').append('<div</pre>
class="res">');
                    $(self.elm).find('.doc:last .res')
                         .append('<a class="title" target="_blank" href="' +
doc.url + '">')
```

```
.append('')
                        .append('');
                   $(self.elm).find('.doc:last .title')
                        .append('<span>' + title + '</span>');
                   $(self.elm).find('.doc:last .description')
                       .append('<span>' + content + '</span>');
                   $(self.elm).find('.doc:last .address')
                       .append('<span>' + AjaxFranceLabs.tinyUrl(doc.url)
+ '</span>');
               });
AjaxFranceLabs.addMultiElementClasses($(this.elm).find('.doc'));
               if (this.pagination !== false) {
                   this.pagination.afterRequest(data);
               }
           }
        }
    }));
   Manager.addWidget(new AjaxFranceLabs.SearchInformationWidget({
        elm: $('#result_information'),
        id: 'searchInformation'
    }));
   Manager.addWidget(new AjaxFranceLabs.TableWidget({
        elm : $('#facet_langue'),
        id : 'facet_langue',
       field : 'language',
       name : window.i18n.msgStore['language'],
       pagination : true,
       selectionType : 'OR',
       returnUnselectedFacetValues : true
   }));
Manager.addWidget(new AjaxFranceLabs.TableWidget({
        elm : $('#facet_source'),
        id : 'facet_source',
        field : 'source',
       name : window.i18n.msgStore['source'],
       pagination : true,
        selectionType : 'OR',
       returnUnselectedFacetValues : true
    }));
   Manager.makeRequest();
```

});

# You should get the following:

Firefox Developer Edition							4	🗟 Fr		N 4×	14:23 🚶
http://localari/test.jsp ×	+										
🗧 🕘 localhost:8080/Datafai	ri/test.jsp		۶	合自	◙	÷	â	<b>e</b> (	8 🖸		• =
		Q									
		All words All least one word Exact expression									
Results 1 - 10 de 117 p								Sort b	/: R	elevai	ncy 🗸
Language	~	lettre de motivation.odt									
		Reda Rahal SABIR A106, Avenue des Facultés Maison des scientif									
en (90)		file:/home/rsabir/Documents/testSearch/lettre%20de%20motivation.odt									
🗌 fr (27)		Diagrame de classe.png									
Source		file:/home/rsabir/Documents/testSearch/Diagrame%20de%20classe.png									
Source	~	me.momensabin bocumentarestoear un bragrame //zoorase.prg									
file (117)		PPetri1.jpp									
		file:/home/rsabir/Documents/testSearch/PPetri1.jpp									
		attestation dactivité.pdf									
		Expéditeur : Mme Amilha									
		file:/home/rsabir/Documents/testSearch/attestation%20dactivité.pdf									
		le recit des decouvertes.docx									
		file:/home/rsabir/Documents/testSearch/le%20recit%20des%20decouvertes.docx									
		index.png									
		file:/home/rsabir/Documents/testSearch/index.png									
		Practical-Work-Tank_CORRIGEE.jtr									
		Test du reservoir, utilisation des événements répétés, calcul									
		file:/home/rsabir/Documents/testSearch/Practical-Work-Tank_CORRIGEE.jtr									
		lettre de resiliation.odt									
		Résiliation d u contrat N: 167483850 Bordeaux , I									

# Widgets and Modules

# SearchBar widget

A widget displaying a search bar and search options.

It contains a link useful to switch to advanced search (if that widget is activated): : in this case a default query is performed and the inputs and URL reset.

#### Variables

- autocomplete: can take 3 values:
  - false(default): no autocomplete.
  - true: autocomplete automaticaly created.
  - AutocomepleteWidget: an autocomplete widget.
- autocompleteOptions: options for the autocomplete widget if autocomplete set to true. Check the code to see the default structure.
- removeContentButton: if set to true, creates a button which appears when there is text in the search bar in order to remove it and execute a request (default: false).
- **noRequest:** if true, inhibits the request to be sent to Solr (default: false).
- updateBrowserAddressBar: if true, the address bar URL is updated with the query parameter, e.g. ?searchType=AllWords&query=datafari (default: true).
- removeContentButton: if true, adds a button that clears the search bar (default: false).

- buildWidget: an implementation of the parent class method. It contains the onClick event handler for the advanced search link.
- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method; closes the autocomplete menu.
- reset: an implementation of the parent class method: it resets the inputs of the widget (useful when switching between basic and

advanced search).

- updateAddressBar: updates the address bar URL with the search term, e.g. ?searchType=AllWords&query=datafari
- clean: resets the pagination and the fq and start parameters of the manager's store.
- makeRequest: if the noRequest variable is false, cleans the results area and facets, updates the address bar and performs the search request.

Note that the Autocomplete widget is also an extension of the Autocomplete plugin from jQuery UI and so it has all its variables and methods too.

# **Result widget**

A widget displaying the search result.

# Variables

- pagination: can take 3 values
  - false(default): no pagination.
  - true: pagination automaticaly created.
  - PagerWidget: a pager widget.

#### Methods

- buildWidget: an implementation of the parent class method.
- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.

# SubClassResult widget

A widget that implements Datafari specific code and parameters the Result widget. We did not test this widget towards a direct Solr system. This widget displays the search result. It displays results differently depending on the activation of the LikesAndFavorites widget.

# Variables

- pagination: can take 3 values
  - false(default): no pagination.
  - true: pagination automaticaly created.
  - PagerWidget: a pager widget.
- elmSelector: used to store the results to be displayed. Corresponds to the jQuery selector.
- firstTimeWayPoint: used in the mobile mode, to know when we reach the bottom of the page, to trigger the spinner in order to load and display the further results.
- isMobile: checks whether we are in mobile mode or not.
- mutex\_locked : is a variable used as a mutex for the mobile mode

# Methods

- **buildWidget:** an implementation of the parent class method.
- **beforeRequest:** an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.

# **ResultIllustrated widget**

A widget that extend Result Widget and that displays image in result.

# Variables

- pagination: can take 3 values
  - false(default): no pagination.
  - true: pagination automaticaly created.
  - PagerWidget: a pager widget.

#### Methods

- **buildWidget:** an implementation of the parent class method.
- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.

# LikesAndFavorites Widget

#### Methods

- buildWidget: an implementation of the parent class method.
- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.
- · function that is runned after getting the likes and favorites of the user
- showError: function that analyze the error and display the corresponding error message

#### SearchInformation widget

A widget displaying the search informations like the number of document found, the query execution time, ...

# Methods

- buildWidget: an implementation of the parent class method.
- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.

# **Table widget**

A widget displaying a facet values in a table.

# Variables

- name: a name for the facet to be displayed.
- pagination: can take 3 values:
  - false(default): no pagination.
  - true: pagination automaticaly created.
  - PagerWidget: a pager widget.
- nbElmToDisplay: the number of facets to display (default: 10).
- sort: can take 3 values
  - occurences(default): sort the facets by the number of occurences.
  - AtoZ: sort the facets name by alphabetical order.
  - ZtoA: sort the facets name in a reversed akphabetical order.
- maxDisplay: maximum number of facets (default: 40).
- checkedOnTop: if set to true, the checked facets will be sorted on the top before the unchecked facets (default: true).

- buildWidget: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.
- update: a method updating the content of the widget.
- **assocTags:** a method maping the search query facets result.
- sortBy: the method sorting the facets, takes the sort mode as parameter.
- clickHandler: a callback called when a facet is selected or unslected.

# FacetDuplicates widget :

This widget displays duplicate documents that are in the index. It is very specific, not recommended for instance if you have an ecommerce website. It is more useful in an intranet context, where you may be interested to know which documents are present several times in your repositories.

The class is of type « final ». For each hash of a given document, the widget retrieves the corresponding document name (the corresponding field name for the document name needs to be identified. By default in Datafari, it is *title\_en or title\_fr*). For more information, refer to the specific deduplication documentation.

# Variables

- name: a name for the facet to be displayed.
  - pagination: can take 3 values:
    - false(default): no pagination.
    - true: pagination automaticaly created.
    - PagerWidget: a pager widget.
- nbEImToDisplay: the number of facets to display (default: 10).
- sort: can take 3 values
  - occurences(default): sort the facets by the number of occurences.
    - AtoZ: sort the facets name by alphabetical order.
    - ZtoA: sort the facets name in a reversed akphabetical order.
- maxDisplay: maximum number of facets (default: 40).
- checkedOnTop: if set to true, the checked facets will be sorted on the top before the unchecked facets (default: true).

# Methods

- buildWidget: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.
- update: sends an ajax request for each hash to get the name of the document. It will also build the html that will be used to display the facet. If there is no duplicate document, the facet will not be displayed.
- assocTags: a method maping the search query facets result.
- sortBy: the method sorting the facets, takes the sort mode as parameter.
- clickHandler: a callback called when a facet is selected or unslected.

#### HierarchicalFacet widget

A widget displaying a facet values in a hierarchical way. Available as of Datafari v2.2.

# Prerequisites

In order to make this widget work, it is mandatory that the selected facet field represents a path tokenization where each value is prefixed by its depth level.

For example, the path '/home/france/labs' will be tokenized as this : 0/home, 1/home/france, 2/home/france/labs

# Variables

- **name:** a name for the facet to be displayed.
- pagination: can take 3 values:
  - false(default): no pagination.
    - true: pagination automaticaly created.
    - PagerWidget: a pager widget.
- nbElmToDisplay: the number of facets to display (default: 10).
- sort: can take 3 values
  - occurences(default): sort the facets by the number of occurences.
  - AtoZ: sort the facets name by alphabetical order.
  - ZtoA: sort the facets name in a reversed akphabetical order.
- maxDisplay: maximum number of facets (default: 40).

- checkedOnTop: if set to true, the checked facets will be sorted on the top above the unchecked facets (default: true).
- **rootLevel:** the root depth level of the hierarchical tree (default: 0)
- maxDepth: the maximum depth to be displayed, starting from the rootLevel (default: 3)
- separator: the separator used between each depth level ( default '/' )

# Methods

- buildWidget: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.
- update: a method updating the content of the widget.
- **assocTags:** a method mapping the search query facets result.
- sortBy: the method sorting the facets, takes the sort mode as parameter.
- clickHandler: a callback called when a facet is selected or unslected.
- selectNoRequest: a method that does the same as the selectHandler parent class method but that doesn't perform the request at the end
- unselectNoRequest: a method that does the same as the unselectHandler parent class method but that doesn't perform the request at the end
- displayLevel: a method which constructs the hierarchy tree with the search query facets result

#### PrevisualizeResult widget

A widget that extend Result Widget and that displays a preview window at mousover on the document.

B	240-annuaire-general-de-montpellier.htm	
1	Infos supplémentaires	
	ID : http://www.montpellier.fr/240-annuaire-general-de-montpellier.htm	

# Prerequisites

In order to make this widget work, you have to customize the content that you want in the window and to activate the widget.

To activate the widget :

• in datafari/WebContent/js/search.js :

Uncomment the line :

```
Manager.addWidget(new AjaxFranceLabs.PrevisualizeResultWidget());
```

• in datafari/WebContent/searchView.jsp

Uncomment the lines :

```
<script
type="text/javascript" src="js/AjaxFranceLabs/widgets/PrevisualizeResult.w
idget.js" charset="utf-8"></script>
```

and

<div id="previsualize"></div>

To customize the information displayed :

Go to datafari/WebContent/js/AjaxFranceLabs/widgets/PrevisualizeResult.widget.js and edit the line to display whatever you want instead of the ID of the document :

```
$($('.doc_list .previsualizetemplate')[index]).append('<div
id="previsualizeid">ID : '+docs[index].id+' </div>' );
```

Methods

- buildWidget: an implementation of the parent class method.
- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.
- showError: function that analyze the error and display the corresponding error message

#### ExternalResult widget

A widget that allows to have a federated search in Datafari ie display results from another search engine (Solr, Sharepoint for example) and display below the normal results from Datafari.



Home

Some figures: Lorem psum dolor sit amet, consectetur adipiscing elit

http://localhost

See more results

# Rest of results :

#### www.francelabs.com

France Labs: Moteurs de recherche open source pour entreprise

http://www.francelabs.com

# Prerequisites

In order to make this widget work, you have to parse the response from your external datasource.

To activate the widget :

• in datafari/WebContent/js/search.js :

Uncomment the line :

Manager.addWidget(new AjaxFranceLabs.ExternalResultWidget());

• in datafari/WebContent/searchView.jsp

Uncomment the lines :

```
<script
type="text/javascript" src="js/AjaxFranceLabs/widgets/ExternalResult.widge
t.js" charset="utf-8"></script>
```

#### and

#### <div id="external"></div>

to select the data to parse :

in datafari/WebContent/js/AjaxFranceLabs/widgets/ExternalResult.widget.js

#### Edit the line :

var urldatasource = 'http://localhost:8080/Datafari/externaldata.json';

Change the url for the data you want to parse. And adapt the code below to parse the info you want.

We provide some example data into Datafari if you want to test it out of the box :

· Copy datafari/dev-tools/example-data/externaldata.json into datafari/WebContent

#### Methods

- buildWidget: an implementation of the parent class method.
- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.
- showError: function that analyze the error and display the corresponding error message

#### **Table Facet Queries widget**

A widget displaying facet queries in a table.

# Variables

- **name:** a name for the facet to be displayed.
- field: the field that will be used in the queries
- **queries:** the list of queries
- pagination: can take 3 values:
  - false(default): no pagination.
  - true: pagination automaticaly created.
  - PagerWidget: a pager widget.
- nbElmToDisplay: the number of facets to display (default: 10).
- maxDisplay: maximum number of facets (default: 40).

#### Spellcheck widget

A widget spellchecking a query string. (eg: can offer you to correct constellia in constellio).

This widget works with both basic and advanced search.

Note: this requires enabling a spellchecker component in the searchhandler of Solr (done in the xml config files), whether you use a pure Solr or a Constellio Solr.

# Variables

There are no specific variables: only the ones inherited from AjaxFranceLabs.AbstractWidget.

# Methods

- buildWidget: an implementation of the parent class method: builds the widget.
- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method: here there is the logic of the spellchecker itself. It checks whether the result
  of the search query has some suggestions and, if so, it calls doSpellCheckerQuery to perform a new search with the spellchecked input,
  contained in the collations array.
- doSpelicheckerQuery: performs a new search with the spelichecked input.

# AdvancedSearch widget

A widget enabling you to specify on which fields to perform a search.

The result of the search query is spellchecked by SpellChecker widget (if that widget is activated).

It contains a link useful to switch to basic search (if that widget is activated): in this case a default query is performed and the inputs and URL reset.

By default the AdvancedSearch is desactivated from the splash page.

The AdvancedSearch widget is built with 2 other classes:

- AdvancedSearchTable: which is a part of the advanced search form, containing fields.
- AdvancedSearchField: which are the different fields provided by the advanced search.

#### Variables

- tables: an array containing all the tables.
- resizable: if set to true, display a resize bar on the bottom of the advanced search form. Requires jQuery UI Resizable plugin.
- **options:** options given to the jQuery UI Resizable plugin.
- optionsExtend: if set to true, extends the default parameters defined in this widget (default: true).

# Methods

- **buildWidget:** an implementation of the parent class method: builds the widget and links the basic search link to the basic search bar.
- beforeRequest: an implementation of the parent class method: builds the search query with the inputs from AdvancedSearchFields and calls updateAddressBar.
- addTable: adds the AdvancedSearchTable object given in parameter to the table array.
- reset: an implementation of the parent class method: resets the inputs of the widget and calls the reset method on the underlying elements table and fields (useful when switching between basic and advanced search).
- makeRequest: cleans the results area (list and facets) and performs a default search (\*:\*)
- cleanResults: cleans the results area (list and facets)
- updateAddressBar: updates the URL in the address bar with the advanced search query created by beforeRequest.

# AdvancedSearchTable

Extends AjaxFranceLabs.Class

#### Variables

- parent: the advancedSearch DOM element.
- elm: the AdvancedSearchTable DOM element itself.
- title: a name for the table (default: empty string).
- **description:** a description for the table (default: empty string).
- fieldStore: an array containing the AdvancedSearchField fields related to the table.
- manager: a reference to the widget's manager.

- · init: an init method that creates the class.
- addField: a method adding an AdvancedSearchField field to the fieldStore array.
- reset: an implementation of the parent class method: resets the inputs of this class and calls the reset method on the underlying fields.

# AdvancedSearchField

Extends AjaxFranceLabs.Class

## Variables

- parent: the advancedSearchTable DOM element.
- elm: the AdvancedSearchField DOM element itself.
- label: a name for the field (default: empty string).
- range: if set to true, will create two inputs instead of one to enable an input range (default: false).
- description: a description for the field (default: empty string).
- multiInput: if set to true, will add a link which allows the user to add more inputs for this field (default: false).
- addinputLabel: text to append to the link adding inputs (default: "Add input").
- autocomplete: can take 3 values:
  - false(default): no autocomplete.
  - true: autocomplete automaticaly created.
  - AutocomepleteWidget: an autocomplete widget.
- autocompleteOptions: options for the autocomplete widget if autocomplete set to true. Check the code to see the default structure.
- manager: a reference to the widget's manager.
- filter: if set to true, the widget will add the content of the field to a fq parameter instead of appending it to the q parameter (default: false).
- field: the field in the search engine index this AdvancedSearchField refers to.

## Methods

- init: an init method, that creates the class.
- getValue: a method returning the value of this field, after call to buildSearchTerm method.
- reset: an implementation of the parent class method: resets the inputs of this class.
- **buildSearchTerm:** expands the field's search term with the available languages: e.g. content:giovanni becomes ((content\_en:giovanni) OR (content\_it:giovanni))

# Information about dynamic query composition:

Since on Solr side we store the title and content on separated fields based on the detected language, the query in input is then translated into the logic OR between the Solr fields that are functionally the same thus having language specialization: E.g.: content:datafari => (content\_fr:datafari) OR (content\_en:datafari) OR (content\_it:datafari) OR (...).

Then every "expanded" logical term is composed with the other "expanded" search terms by means of the radio buttons.

Please note that the query expansion based on languages is dynamic and it's based on the available languages array in i18n variable.

#### How to add a new field:

You need create a new field object and add it to the table object, as follows: in createAdvancedSearchTable function of both search.js

```
Add new field to advanced search
// Table is already created
var ast = new new AjaxFranceLabs.AdvancedSearchTable({ ... });
// Add your field NEW_FIELD
var asf = new AjaxFranceLabs.AdvancedSearchField({
    parent : '#advancedSearchTable',
    label : window.i18n.msgStore['advancedSearch-NEW_FIELD-label'],
    description : window.i18n.msgStore['advancedSearch-NEW_FIELD-descr'],
    field : 'NEW_SOLR_FIELD'
    });
ast.addField(asf);
```

## Menu widget

A widget building a dropdown menu.

#### Variables

- name: the parent element text to display.
- menultems: a collection of the elements to add to the menu ( {name: ..., link: ...}).

## Methods

• buildWidget: an implementation of the parent class method.

#### Slider widget (beta)

A widget that will help you to select values as range or min/max for a chosen field.

This widget uses the jQuery UI Slider plugin.

## Variables

- name: String name of the widget that will appear as the title in the search view
- field: String the field name in the Solr core that will be used by the widget
- fieldType: String the type of the chosen field. Only two values are accepted by the widget, 'date' and 'numeric'. (notice that the lowercases matter)
- elm: String the id of the div in the search view that will contain the widget
- range: Boolean set to true if you want a slider with a range, false otherwise
- min: Integer the minimum value of the slider (also used as the default minimum value if the range slider has been selected). Can be
  negative
- max: Interger the maximum value of the slider (also used as the default maximum value if the range slider has been selected). Can be negative
- unit: String this field is only used when the fieldType has been set to 'date'. In that case you have three options 'YEAR', 'MONTH' or 'DAY'. When a date field type is used with this widget, the resulting Solr query filter that will be applied by the widget will be "(field):[NOW(+/-)(slider\_value)(unit)]. So keep in mind that the slider will be based on the current date to create the query. For now we didn't have implemented a parameter to change this behavior
- defaultValue: Integer only used when the slider is not in the 'range mode'. Default value when the widget is loaded for the first time

- step: Integer defines the number which represents the minimum step performed when the user slide
- comparator: String only used when the widget is not in the 'range mode'. Two possible values, 'greater' or 'less' (lowercases matter). When the widget is not in the 'range mode' you have the choice to define if the chosen field value must be greater than or less than the selected one in the slider. For now we didn't have implemented the 'equal' comparator.

# Methods

- buildWidget: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.

#### Suggest widget

A widget creating a result widget that proposes other suggested documents to your search query. The name may be misleading, this is not the suggest module for an autocomplete, it is rather for a "more like this" behavior.

#### Variables

- **boots:** a list of boosts to apply. { field1: boost\_value1, field2: boost\_value2, ... }.
- resultWidget: the result widget which will contain the suggestions, internal use.

## Methods

- init: an implementation of the parent class method.
- buildWidget: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.

# OntologySuggestion widget

A widget creating a result widget that proposes the most popular parent and child label of all the results. The name may be misleading, this is not the suggest module for an autocomplete, it is rather for a "related subjects" behavior.

## Prerequisites

In order to make this widget work, it is mandatory that the *ontologyEnabled* parameter of the 'datafari.properties' file is set to true, as the documents (at least a few of them) in the results have been enriched by the OntologyUpdateProcessor

## Variables

- · parentLabelsField: the field in your schema that contains the parent documents labels of the ontology
- childLabelsField: the field in your schema that contains the child documents labels of the ontology
- useLanguage: (boolean) should use the Datafari selected language for both parentLabelsField and childLabelsField ? If *true*, be sure that you used the **OntologyUpdateProcessor** with the parameter **userLanguages** set to *true* during the crawl of the documents by MCF

# Methods

- init: an implementation of the parent class method.
- buildWidget: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.

Pager module

A module creating a pagination.

# Variables

- source: the source elements to apply the pagination on. Optional if the elements are retrieved by an Ajax request.
- prevLabel: the content of the previous nav button (default: <).
- nextLabel: the content of the next nav button (default: >).
- firstLabel: the content of the first nav button (default: <<).
- lastLabel: the content of the last nav button (default: >>).
- nbElmToDisplay: the number of elements to display (default: 10).
- **nbPage:** the number of pages, internal use.
- nbPageDisplayed: the number of pages to display in the navigation bar.
- nbElements: the number of elements in source, internal use.
- fastNavigationAlwaysVisible: if set to true, will always display the navigation button even if they are useless (eg: first and previous button when you are on the first page. Default: false).
- **display:** the css display value of the pagination elements.
- pageSelected: the page actually selected (default: 0).

#### Methods

- init: an implementation of the parent class method.
- updatePages: the method updating the navigation bar (eg: elements to display)
- clickHandler: a callback function called when a click on an element of the navigation is detected.
- updateList: a method always called when a click on an element of the navigation is detected. This method is empty by default and must be implemented if you need it.

#### Autocomplete module

This module creates a list of suggestions when you type in an input field. Note: this requires enabling a suggest component as a requesthandler of Solr (done in the xml config files), in case you use a pure Solr. If you are using Constellio, this config is normally already activated in its Solr. For Constellio 1.3, it also requires a modification of the web service servlet in order to reach the Constellio autocomplete service. Contact us if you need help.

This module is also an extension of the jQuery UI Autocomplete plugin and also has all its parameters and methods.

#### Variables

- options: options given to the jQuery UI Autocomplete plugin.
- optionsExtend: if set to true, extends the default parameters defined in this module (default: true).
- render: an empty variable. You can define it a function changing the way an element is rendering in the autocomplete.
- field: if the values retrived must be focused on a specific field of your search engine index. Note that this option will not work with a standard 1.3 Constellio. It requires a modification of its web service. This field variable becomes the variable "f" in the generated query to the server
- openOnFieldFocus: if set to true, executes a request to retrieve data when cursor entering the input field (default: false).
- valueSelectFormat: a function that format the data selected in the autocomplete results (eg: if you want to put it between double quotes).
- singleValue: if set to true, when a value is choson in the autocomplete result, the actual value in the input field will be replaced by the selected one.
- searchForltSelf: if set to true, also search the term itself (default: false). This means that the list will contain the word currently typed in addition to the words which have at least one more letter

### Methods

• init: an implementation of the parent class method.

### **Collection widget**

A widget displaying your constellio's collections. Obviously, this widget works only for Constellio. Note that this requires that your manager was initialised with the array of Constellio collections. See AbstractManager.collections for that. Note also that it requires a modification of the Constellio 1.3 web service servlet. Contact us to know what to modify.

#### Methods

• buildWidget: an implementation of the parent class method.

#### **Capsule widget**

A widget that gives you tips about your search. Does not work for Solr. Note that this only works for Constellio 1.3, with a modified web service in order to support the request.

### Methods

- init: an implementation of the parent class method.
- buildWidget: an implementation of the parent class method.
- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.

#### **Favorite documents and searches**

A module adding the possibility to create a favorite system on your searches and documents. Does not work for Solr. Note that this only works for Constellio 1.3, with a modified web service in order to support the request.

# Variables

• userId: the id of the user manipulating its favorites. This is actually optional if your server is managing the user sessions.

#### Methods

- beforeRequest: an implementation of the parent class method.
- afterRequest: an implementation of the parent class method.
- removeFavoriteDocument: takes as parameter the id of the favorite document and a callback method.
- removeFavoriteSearch: takes as parameter the id of the favorite search and a callback method.
- removeSearchInHistory: takes as parameter the id of the search to remove in the history and a callback method.
- getFavoriteDocuments: takes as parameter a callback function, the first element to start on and the number of elements to retrieve.
- getFavoriteSearches: takes as parameter a callback function, the first element to start on and the number of elements to retrieve.
- getSearchHistory: takes as parameter a callback function, the first element to start on and the number of elements to retrieve.

## LoginDatafariFormWidget

The widget containing the login to admin UI form.

The code of this widget has been migrated from embedded JS of login.jsp to a separate widget.

# Variables

No specific variables for this widget, only the ones inherited from AjaxFranceLabs.AbstractWidget.

## Methods

- buildWidget: an implementation of the parent class method. Creates the widget.
- JQuery anonymous function \$(function(\$) {...}: adds the widget to the manager.

## LoginDatafariLinksWidget

The widget containing the login to admin UI form.

The code of this widget has been migrated from embedded JS of index.jsp to a separate widget.

## Variables

No specific variables for this widget, only the ones inherited from AjaxFranceLabs.AbstractWidget.

# Methods

• buildWidget: an implementation of the parent class method. Creates the widget.

#### LanguageSelectorWidget

This widget enables the user to change the language of the UI, among the available ones.

When a language change is detected by the application, it gets reset and a new splash page is displayed.

To add a language, you need to add an option in the combo box (option DOM element of buildWidget).

# Variables

No specific variables for this widget, only the ones inherited from AjaxFranceLabs.AbstractWidget.

#### Methods

• buildWidget: an implementation of the parent class method. Creates the widget.

# Add custom fields in Datafari and display them

Let's say we want to add a new metadata field in Solr named meta:author in the source and named **authorname** in Datafari. Let's see each step to display the field into the Datafari UI and permit to search into the field with Solr.

#### 1) Edit solrconfig.xml

We want to map the original metadata of the source file : **meta:author** to a new Solr field named authorname. So we have to edit the Solr cell request handler :

The correct syntax is meta\_author (and not meta:author) because of the line <str name="lowernames">true</str> The documentation says : "lowernames=true|false - Map all field names to lowercase with underscores"

You can also see in the configuration that we store all the ignored metadata in the dynamic field ignored. I invite you to change the configuration of the field in the schema.xml and to change stored=false to store=true to see all the metadata found by Tika (and to see the correct syntax to map the fields into Solr) For example :

```
"ignored last save date": [
  "2009-03-12T13:48:39"
Ъ
"ignored_meta_creation_date": [
  "2008-02-07T11:33:11"
1.
"ignored_flags_0": [
 "192"
1,
"ignored_resolution_units": [
  "inch"
Ъ
"ignored_data_precision": [
  "8 bits"
Ъ,
"ignored_tiff_bitspersample": [
 "8"
Ъ,
```

2) Edit schema.xml

We want now to add the new field into the Solr schema. So add the following line :

<field name="authorname" type="text\_en" indexed="true" stored="true" multiValued="true"/>

Ok so far we can launch the indexation with ManifoldCF and the new field is well present in Solr.

3) Add the new field to the search Edit solrconfig.xml, in the select request handler add the field :

After the core reloads, we can now search and find the data of the new field.

4) Configure the Datafari UI into datafari/tomcat/webapps/Datafari/js/main.js (source code) or Datafari/tomcat/webapps/Datafari/js/main.js (installed version) Change the line :

Manager.store.addByValue("fl", 'title,url,id,extension');

And add the field you want to add, here autorname

Manager.store.addByValue("fl", 'title,url,id,extension, authorname');

The last step is to change the Javascript file search.js : datafari/WebContent/js/search.js (source code) or Datafari/tomcat/webapps/Datafari\js/search.js (installed version) Add the display of your field adding the code : doc.subject where you want to add it. For example if you want to add it after the URL of the document : (I made a mistake in my previous answer, it is correct now)

elm.find('.doc:last .address').append('<span>' + AjaxFranceLabs.tinyUrl(decodeURIComponent(url)) +
'</span>')
elm.find('.doc:last .address').append('<div id="author">' + doc.author );

And finally your Datafari UI should be like that with the field author at the end of each document :

6		
DATAFAR		
		Administration
	Q	
	All words     At least one word     Exact expression	
	Sort by : Relevancy 🔻	
	0.044	
Result 1 - 8 of 8 for *:* (	u.utt seconasj	•
Last Modification	— 🗟	
Less than 1 month (0)	🖹 Tulips.jpg	
Less than 1 year (0)	file:/C:/Users/Public/Pictures/Sample Pictures/Tulips.jpg David Nadalin	
Less than 5 years (0)	🖹 Koala.ipg	
	1.0	
🖾 Туре	file:/C:/Users/Public/Pictures/Sample Pictures/Koala.jpg Corbis	
jpg (7)		
	🖹 Lighthouse.jpg	
Source	file:/C:/Users/Public/Pictures/Sample Pictures/Lighthouse.jpg Tom Alphin	
file (8)		
	🖹 Desert.jpg	
🖾 Language	file:/C:/Users/Public/Pictures/Sample Pictures/Desert.jpg Corbis	
en (8)		
	🖻 Hydrangeas.jpg	
	file:/C:/Users/Public/Pictures/Sample Pictures/Hydrangeas.jpg Amich Patel	

# Alerts - Architecture and mechanism

The alerts are stored in a local Apache Cassandra. You can create, edit or delete them from the *alerts.html* page which communicates with the *Al erts* servlet. This servlet creates a connection towards the database with the info contained in the *alerts.properties* file.

1. For the moment, the alerts cannot be used on plain text files, as Tika cannot extract any metadata like the modification date from them.

An alert is composed of seven fields:

- an \_id created by the Cassandra,
- a keyword which will be used to make the request on the solr core,
- the core on which the query will be made, default : FileShare,
- a *frequency*: Hourly, Daily or Weekly. Frequency changes the time lapse between two requests but also the recency of the documents returned.
- a mail to which the results (if any) of the query will be sent,
- a subject which describes the subject of the email
- a user: the one who created the alert in the UI. It is only used for now to allow for the monitoring of the alerts by the admin and the search expert.

The alertsAdmin UI and servlet are used to modify the alerts.properties file. It is used to:

- configure the initial delay of the execution of the alerts of a specific frequency,
- print the date of the previous execution
- print the date of the next planned executions.

- configure the mail address from which Datafari will send its emails.
- configure the host and port of the Cassandra and the name of the Database and of the collection where the alerts are.

The AlertsManager class is a Singleton and is called at each start/stop of the server by the StartAlertsListener.

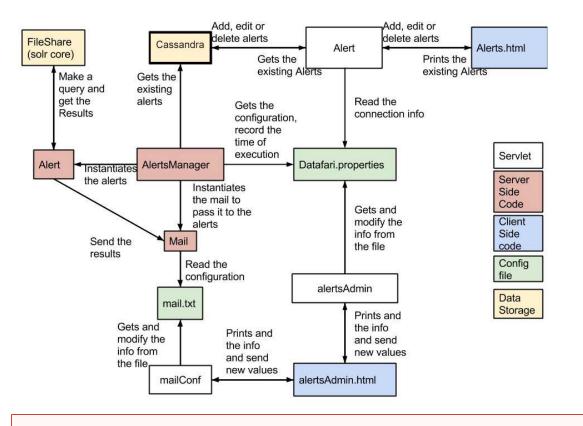
At each start it reads the *alerts.properties* file to fill its own fields, and check if the alerts were on or off at the end of the previous execution of Datafari. If it was on On, the alerts Manager schedules the task according to the dates in the file. However if those are invalid, it will calculate the delays according to the original execution schedule of the alerts.

For instance the hourly alerts are set to begin at midnight, if everything is going smoothly, the hourly alerts will run at 00:00 then at 01:00 and so on. Now let's say that the date was already past at the start of Datafari. Then the AlertsManager will plan the execution for the next expected hour.

This allows the alerts to be run regularly, even after one or more restart of the server.

In order to run an alert:

- the AlertsManager creates a connection with the Cassandra database at each execution
- the AlertsManager then retrieves all the alerts with the matching frequency.
- For each alert, the AlertsManager creates an Alert object, and delegates the execution to them.
- An object alert will use its attributes to make a solrQuery
- If the solrQuery contains results, it will use the email object passed at its creation to send a mail. To avoid re reading of the *alerts.properti* es file for each alerts, this object is initialized in the alerts manager and passed as a parameter to the Alert constructor. If the configuration file is modified through the admin UI, Datafari will automatically restart the alerts, if it is manually modified, you will need to turnOff/turnOn the alerts in the alertAdmin UI to take into account the modifications.



The schema is wrong concerning the Datafari.properties file and mail.txt file, both files are no more used by the AlertsManager and they have been replaced by one configuration file named alerts.properties located in Datafari\_Home/tomcat/conf

#### More info on our code?

Please refer to our code in github, as our source code contains additional technical details avaiable as comments

# Architecture

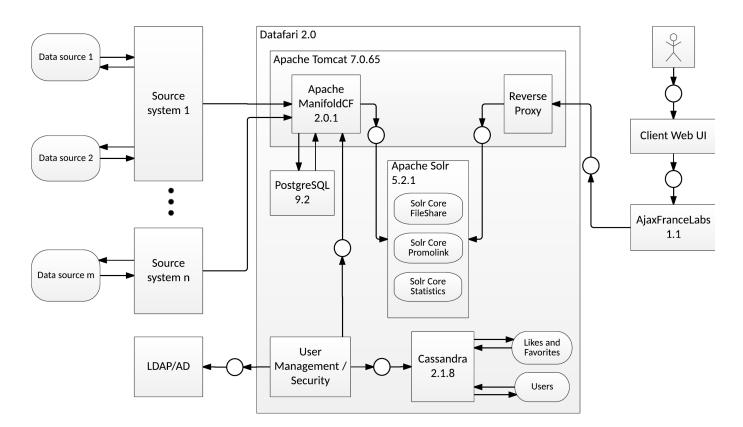
# Valid for V2.0

Datafari uses a typical search engine architecture. It is based on the triptych crawling, indexing and search. The crawling part is using Apache ManifoldCF. There is another opensource connectors framework under Apache licence v2, the Google Connector Framework. Yet the latter is only being supported and developed by Google, so it appeared more reasonable to us to leverage ManifoldCF, proposed by the Apache foundation, and which benefits from the support of several committers from different entities.

The indexing and search parts both use Apache Solr. Again, there is another popular indexing and search engine available under Apache licence V2. It is Elasticsearch, but similarly to the Google Connector Framework, it is led by the eponym entity, and as such does not guarantee such a longevity as Apache Solr.

ManifoldCF, although being independent from Apache Solr, has the advantage of being conceived from the start as Solr's connectors framework. It is thus conceptually "naturally" connected to Apache Solr.

The figure below illustrates the system architecture that we are using for Datafari. Its v2 is becoming rather large in terms of components, so this architecture is rather high level and intentionally avoids some connections and components for the sake of clarity.



## Datafari 3.0

On Datafari 3.0 we chose to use directly SolrCloud and to isolate the different parts of our product that can be installed on different servers.

If your Solr index becomes very huge or if your nomber of users becomes too big, you can easily shard your index and/or replicate it with SolrCloud.

# **Custom Solr configuration**

Datafari allows you to add your own Solr configuration to the existing one, without modifying the original configuration files. To do so, both solrconfig.xml and schema.xml files are including custom files that are organized by sections:

Note that if for some reason you need to modify directly the solrconfig.xml file or the schema.xml file, you can do it but we strongly advise you not to do so. These files may change in future versions of Datafari and you will have to manually manage the conflict during the update process.

solrconfig.xml includes 4 custom files located in {Datafari\_Home}/solr\_solr\_home/FileShare/conf/customs\_solrconfig.

- custom\_libs.xml: Used to add new jar libraries that Solr should load. For exemple, to use the ontology feature of Datafari, the Apache Jena lib directory will be added in this file
- custom\_request\_handlers.xml: Used to add new request handlers
- custom\_search\_components.xml: Used to add new searchComponents
- custom\_update\_processors.xml: Used to add new updateProcessors. The updateProcessors that you will specify in this file will be added to the 'datafari' updateRequestProcessorChain (refer to the solrconfig.xml)

schema.xml also includes 4 custom files located in {Datafari\_Home}/solr/solr\_home/FileShare/conf/customs\_schema:

- **custom\_copyFields.xml**: Used to add new copyFields to the FileShare core
- custom\_dynamicFields.xml: Used to add new dynamicFields to the FileShareCore
- custom\_fields.xml: Used to add new fields to the FileShare core
- custom\_fieldTypes.xml: Used to add new fieldTypes to the FileShare core

Those 8 custom files contain by default an xml tag <to\_remove/> This tag serves to pass the xml validity check of the file when Solr loads it, and avoid exceptions that will crash Solr. When you fill those files with your needs, you should remove or comment this tag and put it again when you remove everything from a file.

# **Deduplication : Technical note**

The deduplication functionality is one of the functionalities that is simplified since Solr 1.4 and the above versions, as it proposes to enable it through its config files. Datafari used this functionality to implement it and make it available in the front-side.

- For the backend side, we have applied the wiki that is available here : https://cwiki.apache.org/confluence/display/solr/De-Duplication . The hash we generate is stored in a solr field labelled *signature*. By default, the hash is computed on the solr *content* field.
- we set the overwriteDupes parameter to false :

<bool name="overwriteDupes">false</bool>

in the *solrConfig.xml*. You can set it to true if you want the index to contain only one instance per set of duplicated documents. This is not our goal here (we want to let the search admin the possibility to remove duplicates after the fact). we have set the fields parameter of the *solrConfig.xml* to:

e have set the helds parameter of the soll config.xhill to.

<str name="fields">content, content en, content fr</str>

so that the hash will take in consideration only the content of each document. You can change this parameter to put the fields you want, for instance the title so that the hash will take into consideration the title.

The processors of the dedupe are in :

<updateRequestProcessorChain name="datafari">

We used also - for max precision - the MD5 Algorithm of solr for hashing. Still, you can always change it by changing the parameter:

<str name="signatureClass"> solr.processor.MD5Signature</str>

in solrConfig.xml.

For the front-end, we want to expose duplicates using a facet:

For this, we have created a new class called FacetDuplicates and which is located in /datafari/WebContent/js/AjaxFranceLabs/widgets/. This class inherits from TableWidget and overloads the update method. This was achieved due to the fact that duplication is a facet and so returns only the hashes and that we wanted to return the names of a document from the duplicated documents. So what happens is that we send for every hash in the facet a get query. We have also set a mincount for the facet so we will show only duplicated file names in the facet which is not the case when the mincount is equal to 0 (which is the default configuration). You can find in this link a short doc for the parameter mincount : https://cwiki.apache.org/confluence/display/solr/Faceting. We also make the facet disappear if it doesn't contain a duplicated document : a simple \$('# facet\_signature').show/hide had done the trick.

# **ELK** layout

Feature only available from the v2.2 of Datafari

#### **ELK configuration notes**

By default, aside from Kibana, Elasticsearch and Logstash are automatically configured by Datafari on the first start, to fit with your installation directory and be ready to run.

If for some reasons (like an architecture decision), you move one or more of these components to a different place than the installation directory of Datafari, you must modify the ELK environment parameters located in [DATAFARI\_HOME]/elk/scripts/set-elk-env.sh and maybe also the [DATAFARI\_HOME]/elk/scripts/start-elk.sh and [DATAFARI\_HOME]/elk/scripts/stop-elk.sh scripts.

Furthermore, if you want to manage by yourself the instance of ELK you just need to set the ELKactivation property to "true" in [DATAF ARI\_HOME]/tomcat/conf/datafari.properties and enter the correct Kibana URI in the ELK configuration page of the admin UI

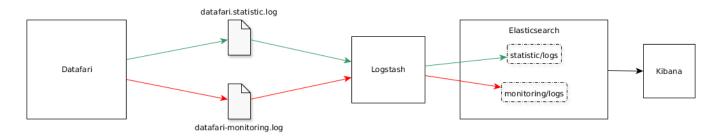
The 2.2 version of Datafari comes with an Elasticsearch Logstash Kibana layout in order to bring customizable monitoring and analytic views on Datafari.

In order to use the ELK layout, you need to configure it first !

Two new kind of logs are generated each in separate files, in order to be exploited in Kibana:

- Statistics logs
- Core Monitoring logs

The diagram below shows the flow between Datafari and Kibana:



As Datafari writes lines in the log files, Logstash, in near real time, catches each new line and insert it in the "statistic/logs" or the "monitoring/logs" index of Elasticsearch, according to the log line comes from the "datafari.statistic.log" file, or the "datafari-monitoring.log".

The "logstash.conf" file which describe the path of the input log files and how to process the lines and insert them into Elasticsearch can be found in the Logstash main repository. Refer to the Logstash documentation for a deep understanding on how it works. You will also find the templates used for each Elastisearch index in the "templates" directory under the Logstash main directory. The files are called "datafari-statistic-template.json" and "datafari-monitoring-template.json".

# 1. Statistics logs

Thoses logs give informations about the queries and actions performed in Datafari by users. For each query or action, a log line is written in the *datafari.statistic.log* file, Logstash detects the new line, applies some filters (defined in the *logstash.conf* file), and inserts it as a document in the statistic/logs index of Elasticsearch.

As described in the Statistics logs page, an id is generated for each query (the actions related to a query, like clicking on a result, keep the id of the query) by Datafari. This id is used as the document id in Elasticsearch. By doing this, one has the guarantee, in Elasticsearch and then Kibana, to not have duplicates data for the same query/action.

So, if several lines in the log file are related to the same query/action, they will be inserted in the same document in Elasticsearch, which consequently means that the document in Elasticsearch will correspond to the most recent log line. However, this does not mean that informations are lost, because a log line contains the history of the query and his related actions.

Thanks to this flow, you can see what kind of charts you can create on the Usage Statistics page.

### 2. Core Monitoring logs

Those logs give informations on the content of the main Solr core of Datafari at a fixed rate. The default rate is once per hour and currently can only be changed by modifying it in the code.

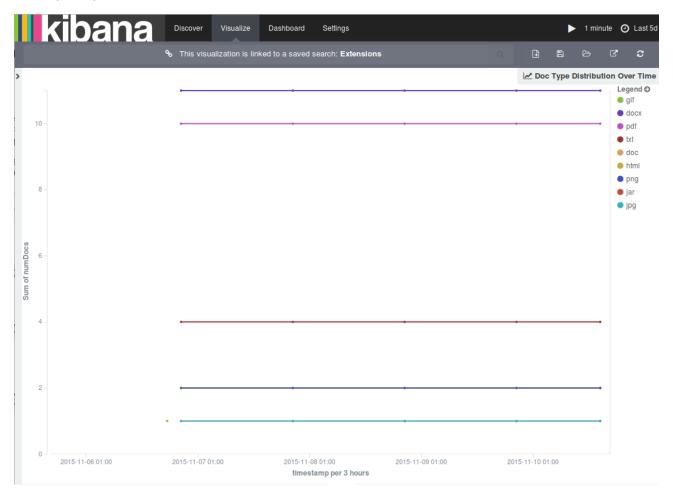
So, at each iteration, Datafari performs some queries on Solr, format the results and writes several logs lines in the *datafari-monitoring.lo g file*. In near real time, Logstash detects the new lines, applies some filters (defined in the *logstash.conf* file), and inserts them as documents in the *monitoring/logs* index of Elasticsearch.

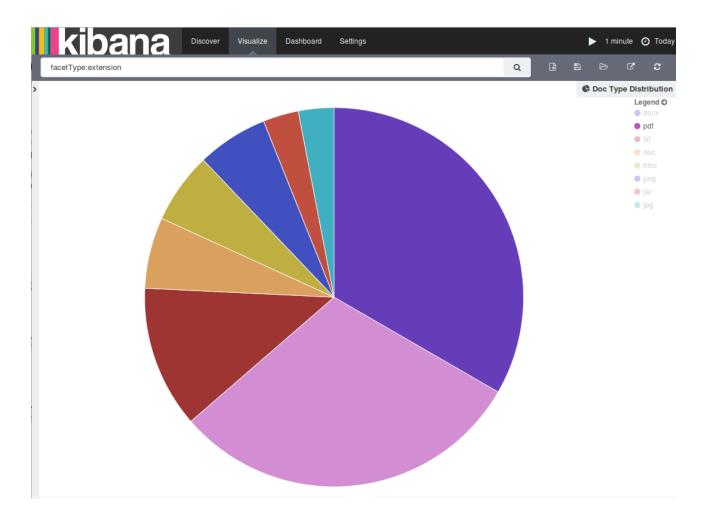
As described in the Core Monitoring logs page, a log line contains an id which is based on the facet value, the facet field and the time event unity. This id is used as the document id in Elasticsearch.

The time event, which is "daily" by default, correspond to the unity that will be used to visualize data in Kibana. For example, with the default "daily" unity, you will have one Elasticsearch document by facet value and field, by day. If you set the time event unity to "hourly" you will have one Elasticsearch document by facet value and field by hour.

This means that as long as a new time event has not started, for each new log line related to a couple of facet value and field, only one document will be created/updated. When a new time event starts, a new document is created/updated for the same couple of facet value and field.

Concretely, a daily time event unit allows such Kibana visualizations:





As you can see on the time based line chart on the left, which represents the document type distribution over the time, we only have one dot per doc type per day, so, for the last 5 days, 5 values for each facet value of the face field "extension".

On the global doc type distribution pie chart on the right, one have the global distribution of document types in the Solr core. This chart is not based on time unlike the other one, so, to be able to have a correct distribution view, one need to set the time frame of Kibana to the time event unity evoked before : the current day

Notice that, as our monitoring log iterations are hourly (by default), our data are updated each hour without having more than one dot per day in the time base line chart.

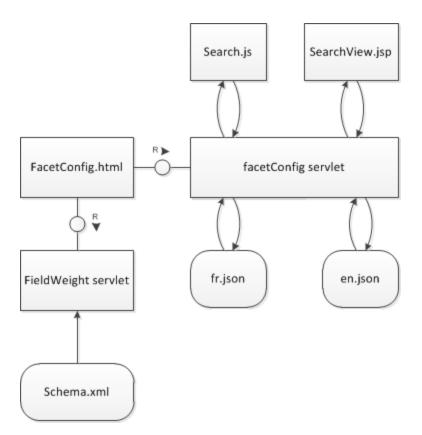
Now you have a better understanding on the time event unity that you can change in the code and his effect on Elasticsearch and the visualization of data in Kibana. Up to you to set you own time event unity and adapt your visualizations in consequence. You can find more example of time based charts and "normal" charts you can create thanks to this flow in the Content Analysis Over Time and Content Analysis pages.

# Faceting

FacetConfig allows you to create custom facets, to delete them, or to modify their printing order.

The UI interacts with the fieldWeight and the facetconfig servlet:

- fieldWeight reads the schema.xml,
- · facetconfig reads and writes the internationalisation file, the searchview.jsp and the search.js



# At creation:

- a request is made towards FieldWeight servlet to get the list of fields,
- another request is made towards the facetConfig servlet to get all the existing servlets, removing them from the selection.
- the submit button sends a post request to facet config, with all the facets. Those are differenciated between themselves by their parameters. Both types of facets are added pretty much the same way, except in the modification of the search.js.

Note that adding a facet modifies four files : the searchview.jsp, the search.js, then en and fr.json. You need to modify or add a json file after you have changed or added another language.

For query facets, the construction of the String added to the js is different, it includes every query that is filled, and if there is no query passed as parameters then it ignores the request.

# Manually adding facets to the Datafari UI

In this tutorial, we will see how to add a facet in the Datafari UI. The facet here will be a facet field, that means that the source of the faceting are the different values of the Solr field.

- Indexation
  - Add a metadata value in a MCF job
  - Field already added to the Datafari schema
- Configure the Datafari User interface

# Indexation

We distinguih two cases here :

- you want to add some metadata from you ManifoldCF job
- or you already modified your schema and you indexed a custom Solr field

Remark : in the tutorial we write the absolute path for Debian version wich starts by /opt/datafari but it is the same for Windows paths : just replace the beginning of the path by your local installation of Datafari

# Add a metadata value in a MCF job

In our example, let's say that we want to have a facet with the source of the ManifoldCF job. Here I want to index 2 websites with ManifoldCF :

francelabs.com and datafari.com. I want to propose to the users to have a facet with the origin of the document : francelabs.com or datafari.com.

Let's do this !

Add the custom field to the Solr schema

Edit custom\_fields.xml located in /opt/datafari/solr/solr\_home/FileShare/conf/customs\_schema/custom\_fields.xml and add your new field. It needs to be at least indexed (not necessarily stored) and not tokenized so if I want to add the field named job, the configuration will be :

```
<field name="job" type="string" indexed="true" stored="true" multiValued="false"/>
```

Don't forget to save your changes into the file.

Reload the FileShare core for Solr to load your new schema

Go to the Datafari Admin User Interface (UI) then click on Search Engine Administration then on Solr Administration

Datafari							
Utilisateurs		administration / So	ir administration				
Lill Connecteurs							
Statistiques							
Administration du moteur de recherche		Instance			E System 0.00 0.01 0.05		G
Administration de Solr	Soir	(# Start	about 8 hours ago		Physical Memory 36.0%		
Administration des alertes	📾 Dashboard	😤 Versions			2.82 GB		
Champs de l'index	Logging	solr-spec solr-impl	4.8.1 4.8.1 1594670 - rmuir - 2014-05-14 19:38:41				
Analyse du schéma	Java Properties	ag lucene-spec	4.8.1		Swap Space 0.0%		
Limitation des tailles	Thread Dump	lucene-impl	4.8.1 1594670 - muir - 2014-05-14 19:22:52	0.00 M	8		
Configuration de l'autocompletion	Core Selector				File Descriptor Count 8.7%		
Configuration du moteur de recherche							
Serveurs					355		
		jvm			JVM-Memory 24.3%		
		Runtime Processors	Oracle Corporation Java HotSpot(TM) 64-Bit Server VM (1.7.0_71 24.71-b01)		100 C		
		Args	Opva ko tropfar-logitalizationexteep Octationa konev exploitalizationextee Octationa konev exploitalizationextee Opvaa vantoread dis-voiptolaisfarintima cateroloread Xmm2044m Opvaa valto gearte manofeld: configitie=J.inclinet jonexipropertees.em Opvas.valto gearte manofeld: configitie=J.inclinet jonexipropertees.em Opvas.valto gearte manofeld: configitie=J.inclinet jonexipropertees.em Opvas.valto geartees		479.84 MB		1.93 GB 1.93 GB
				Documentation	n 🔹 Issue Tracker 🛛 🙇 IRC Channel	🖂 Community forum 🛛 👩 Solr Qi	uery Syntax

Then click on Core Admin on the main window, check if FileShare is highlighted then click on the Reload button

Datafari				
Utilisateurs		administration / Solr admin	nistration	
Lat. Connecteurs				
E Statistiques				
Administration du moteur de recherche	Solr	Add Core	💥 Unload 📑 Rename	😹 Swap 🖏 Reload 🔀 Optimize
Administration de Solr	3011 🥪	FileShare	Core	
Administration des alertes	Dashboard	Promolink	startTime:	about 7 hours ago
Champs de l'index	Logging	Statistics	instanceDir: dataDir:	(opt/datafari/tomcat/bin///soir/soir_home/FileShare/ (opt/datafari/soir/soir_home/FileShare/data/
	Core Admin			to be to measure (neor taxes - nor ref). Let come a tomana)
Analyse du schéma	🧝 Java Properties		👔 Index	
Limitation des tailles	Thread Dump		lastModified:	about 6 hours ago
Configuration de l'autocompletion	Core Selector 💌		version:	748
Configuration du moteur de recherche			numDocs: maxDoc:	495 495
			deletedDocs:	
Serveurs			optimized:	*
			current:	*
			directory:	org.apache.lucene.store.NRTCachingDirectory.NRTCachingDirectory(MMapDirectory@/opt)datafari/soir/soir_home/FileShare/data/index lockFactory=NativeFSLockFactory@/opt)datafari/soir/soir_home/FileShare/data/index; maxCacheMB=48.0 maxMergeSizeMB=4.0)
				📄 Documentation   🏨 Issue Tracker 🛛 🔌 IRC Channel 🛛 🖂 Community forum 🛛 👩 Solr Query Syntax

The button is now green, it means that the new configuration is correctly loaded into Solr.

• So now we can configure ManifoldCF. First thing is obviously to configure the repository connection and the job configuration.

I invite you to go to this section of the wiki if you are not familiar with this step : Crawling

Here we add a Web repository connection and then added a FranceLabswebsite job in which we configured the seeds tab with the url of the France Labs website.

Datafari											
• Utilisateurs	Home / Connectors										
(at Connecteurs									3 and 2015 DR 12 16		
I Statistiques	<b></b>	Document Ingestion Apache ManifoldCF™									
F Administration du moteur de recherche	Apache										
Configuration du moteur de recherche	ManifoldCF <sup>™</sup>										
E Serveurs											
	Outputs	View a Job									
	List Transformation Connections	view a Job									
	List Output Connections	Name:	FranceLabsWebSit								
	Authorities	Name									
	List Authority Groups	Pipeline:	Stage	Type	Precedent	Description	Connection name				
	List User Mapping Connections		2	over			Depresention				
	List Authority Connections	Priority:	5					Start method:	Don't automatically		
	Repositories	Schedule type:	Scan every docume						start		
	List Repository Connections	Minimum recrawl interval: Expiration interval:	Not applicable Not applicable	ni orce				Maximum recrawl inter- Reseed interval	val: Not applicable Not applicable		
	Jobs	No scheduled run times									
	List all Jobs										
	Status and Job Management	Maximum hop count for link type 'link': Maximum hop count for link type 'redire	Unlimited ct: Unlimited								
	Status Reports	Hop count mode:	Delete unreachable	documents							
	Document Status	1									
	Queue Status	Seeds		ttp://www.francelabs.com							
	History Reports			lered and have all sessions removed							
	Simple History	No canonicalization opecned - al	I ONLS WILL DE TEOR	lereu anu nave ali sessions remover	,						
	Maximum Activity	No mappings specified; will acce	al URLs								
	Maximum Bandwidth										
	Result Histogram	Include only hosts matching seeds?	2	8							
	Miscellaneous	Include in crawl: Include in index:									
	Help	Exclude in index:									
	Log Out	Exclude from crawc Exclude from index									
	<	And a second second							•		

We did the same thing for Datafari website (seed: datafari.com) :

iurs									
teurs									3 and 2011
lues	· · · · · · · · · · · · · · · · · · ·	Document Ingestion							0 800 201
tration du moteur de recherche	Apache 🖌 🖌								
	ManifoldCF	TM							
ration du moteur de recherche	Mailloluur								
5									
	Outputs	View a Job							
	List Transformation Connections	view a 500							
	List Output Connections								
	Authorities	Name:	Datafor/WebSite2						
	List Authority Groups	Pipeline:	Stage	Type	Precedent	Description	Connection name		
	List User Mapping Connections		2	outer			Depresention		
	List Authority Connections	Priority:	5					Start method:	Don't autom
	Repositories	Schedule type:	Scan every document						start
	List Repository Connections	Minimum recrawl interval Expiration interval	Not applicable Not applicable					Maximum recrawl in Reseed interval	terval: Not applicat Not applicat
	Jobs	No scheduled run times							The approx
	List all Jobs	No scheduled run times							
	Status and Job Management	Maximum hop count for link type 'link' Maximum hop count for link type 'red	Unlimited						
	Status Reports	Hop count mode:	Delete unreachable d	ocuments					
	Document Status	1							
	Queue Status	I. Seeds		ttp://www.datafari.com					
	History Reports								
	Simple History	No canonicalization specified -	all URLs will be reorder	red and have all session	is removed				
	Maximum Activity	No mappings specified; will acc	out all LIDLa						
	Maximum Bandwidth	No mappings specified, will act	ept all OKLS						
	Result Histogram	Include only hosts matching seeds?	y	es					
	Miscellaneous	Include in crawl:	1						
		Include in index:	1						
	Help	Exclude from crawl:							

• So for now we have our 2 ManifoldCF (MCF) jobs well configured. We need to indicate to ManifoldCF to add a particular metadata for each job to distinguish the documents. In order to to that, we need to configure a Transformations connector. Click on the right menu on List transformations connections in the Output sections.

Then add a name and click on the Type tab. Select Metadata adjuster on the dropdown list then click save.

Datatari			
• Utilisateurs			
Lat. Connecteurs			3 août 2015 08:28 17
III Statistiques		Document Ingestion	3 800 2010 05 25 11
Administration du moteur de recherche	Apache		
Configuration du moteur de recherche	ManifoldCF <sup>™</sup>		
Serveurs	, interiore of		
an serveurs	Outputs		
	List Transformation Connections	View Transformation Connection Status	
	List Output Connections		
	Authorities	Name: Transformation/HebSite	Description:
	List Authority Groups	Connection type: Metsdata adjuster	Max connections: 10
	List User Mapping Connections	Connection status: Connection working	
	List Authority Connections	Refresh Edit Delete	
	Repositories		
	List Repository Connections		
	Jobs		
	List all Jobs		
	Status and Job Management		
	Status Reports		
	Document Status		
	Queue Status		
	History Reports		
	Simple History		
	Maximum Activity		
	Maximum Bandwidth		
	Result Histogram		
	Macellaneous		
	Help		
	Log Out		
	<		•

Now go back to the job configuration. Click on List all jobs on the Jobs section. Then click on the edit link for the "FranceLabsWebsite" job.

Go to the Connections tab and click on the dropdown select list in front of the Transformation line then select the new transformer freshly created. Then click on the button named "Insert transformation before".

So you should have the following configuration :

Datafari															
• Utilisateurs															
Lat Connecteurs															3 audi 2015 08:32:43
E Statistiques		Docume	ent Inge	stion											3 2001 2015 06:32 43
F Administration du moteur de recherche	Apache														
Configuration du moteur de recherche	ManifoldCF <sup>™</sup>														
Serveurs															
	Outputs	Name C	Connection	Scheduling Ho	Filters Seed	Canonicalization	URL Mapping	s Inclusions	Exclusions	Security	Metadata	Move metadata			Edit job 'Datafari/VebSite'
	List Transformation Connections														
	List Output Connections						Stage	Type	Precedent	Deso	ription			Connection name	
	Authorities							Republicy						Wes	
	List Authority Groups				ation before			Tanatumation						Tanatumation/twodite	
	List User Mapping Connections	Pipeline:			_			0.004						Depressor	
	List Authority Connections				e			C. Salar							
	Repositories	1						Terreturnation						None selected	
	List Repository Connections											method			
	Jobs	Priority:	5	٠							Start	method.	Don't automa	atically start this job 🔻	
	List all Jobs	S	Cancel												
	Status and Job Management	Same	Canton												
	Status Reports														
	Document Status														
	Queue Status														
	History Reports														
	Simple History														
	Maximum Activity														
	Maximum Bandwidth														
	Result Histogram														
	Miscellaneous														
	Help														
	Log Out														
	<														E.

New tabs appeared called Metadata, Move metadata and Add metadata. Click on Add metadata. Into parameter name you have to write the same name that the field you configured above so in our case it is named job. For the parameter value, I wrote FranceLabs. Then click on save.

The global configuration of the FranceLabs job is this one now :

Datafari									
Utilisateurs									
(a) Connecteurs	/								
I Statistiques									
& Administration du moteur de recherche	Outputs	View a Job							
	List Transformation Connections	view a 500							
Configuration du moteur de recherche	List Output Connections								
Serveurs	Authorities	Name:	FranceLabsWel						
	List Authority Groups		Stage	Type Recordery	Precedent	Description	Connection name		
	List User Mapping Connections	Pipeline:	2	Transformation			Terreformation/vectore		
	List Authority Connections		3.	Overvi	1		OesterDor		
	Repositories	Priority:	5					Start method:	Don't automatically start
	List Repository Connections	Schedule type: Minimum recrawl interval	Scan every door Not applicable	ument once				Maximum recrawl inter-	ust Not applicable
	Jobs	Expiration interval	Not applicable					Reseed interval:	Not applicable
	List all Jobs	No scheduled run times							
	Status and Job Management	Maximum hop count for link type 'link':	University						
	Status Reports	Maximum hop count for link type 'redire	st: Unlimited						
	Document Status	Hop count mode:	Delete unreacha	able documents					
	Queue Status	1.							
	History Reports	Seeds:		http://www.francelabs.com					
	Simple History	No canonicalization specified - all URLs will be reordered and have all sessions removed							
	Maximum Activity								
	Maximum Bandwidth	No mappings specified; will acce	pt all URLs						
	Result Histogram	Include only hosts matching seeds?		Y85					
	Miscellaneous	Include in crawt		1					
	Help	Include in Index:		2					
	Log Out	Exclude from crawl:							
		Exclude from index:							
		No access tokens specified							
		Excluded headers:							
	<	· · · · · · · · · · · · · · · · · · ·							•

We do the same thing for the Datafari job, the parameter name is job too and the value is Datafari.

We can now launch the crawl job for the two jobs. Go to the Jobs section and click on "Status and Jobs management". Than click on Start for the two jobs.

teurs	Home / Connectors								
cteurs									3 août 20
iques	<u>بد</u> .	Document Ingestion							3 8001 25
stration du moteur de recherche	🔺 📄 Apache								
	ManifoldCF	тм							
uration du moteur de recherche	<b>Mainfolu</b>								
irs	Outputs	Status of Jobs							
	List Transformation Connections	Status of Jubs							
	List Output Connections		Name	Status	Start Time	End Time	Documents	Active	Processed
	Authorities	Restart Restart minimal Pause Abort	Datafar/WebSite	Running	Mon Aug 03 07:30:17 CEST 2015	200 1000	57	11	57
	List Authority Groups	Restart Restart minimal Pause Abort	FranceLabsWebSite	Running	Mon Aug 03 05:57:15 CEST 2015		542	49	542
	List User Mapping Connections								
	List Authority Connections	Refresh							
	Repositories								
	List Repository Connections								
	Jobs								
	List all Jobs								
	Status and Job Management								
	Status Reports								
	Document Status								
	Queue Status								
	History Reports								
	Simple History								
	Maximum Activity								
	Maximum Bandwidth								
	Result Histogram								
	Miscellaneous								
	Help								

We have now to wait a little for some documents to be indexed.

• We are going to check if all is OK in the Solr Administration interface. Click on Search engine administration then on Solr administration in the right menu.

In the main window, select now FIIeShare in the dropdown select list then click on Query in the tabs list below.

Then click on the blue button Execute Query : it will launch a default Solr query : we search anything on all the Solr corpus. We should obtain :

Datarari			
Utilisateurs		administration / Solr administration	
Lat Connecteurs			
I Statistiques			
F Administration du moteur de recherche	Solr	Request-Handler (qt) /select	13 http://178.32.145.133.8080/Datafar/ledmin/proxy/solr/FileShare/select?q=%3Ar&wt=json&indent=true
Administration de Solr		- common	{     "responseleader": {
Administration des alertes	Dashboard     Desping	9  x:x	"status": 0, "QTIME": 181,
Champs de l'index	E Core Admin		"params": ( "indent": "true",
Analyse du schéma	🧕 Java Properties	fq.	"@": ""::"; "_": "1438663255445",
Limitation des tailles	Thread Dump	sort	"hdt": "json" )
Configuration de l'autocompletion	FileShare +	start, rows	), "response": ( "numerand": 405.
🖵 Configuration du moteur de recherche	T Overview	0 10	"maneous: +0%, "start":0, "60:5":[
Serveurs	T Analysis	fi	( ("lat_modified": "2013-12-20117:41:052",
	C Documents	df	"194": "Trip://www.francelabs.com/bootstrap.min.cos", "194": "Trip://www.francelabs.com/bootstrap.min.cos",
	🔜 Files	Raw Query Parameters	"wel": "http://ww.francelabs.com/bootstrap/css/bootstrap.min.css", "source": "http://ww.francelabs.com/bootstrap.css/bootstrap.min.css",
	Ping	key1=val16key2=val2	"extension": "cs", "language": "en",
	L Query	ison 🔻	ampang - um , "content_en": [ " 'in in //"lin " Bootstrap v2.3.1in "in " Copyright 2012 Twitter, Incin " Licensed under the Apache License v2.0in " http://www.
	12 Replication	indent     debugOuery	<pre>); (iii); ); (title_m*; [</pre>
	🔛 Schema Browser	dismax	"bootstrap.nin.css"
		edismax	"title": {     "bootstree.min.css"
		🗉 N	b
		facet	"_wersion_": 150464402440250000, "allow_token_share": [
		spatial	allow_conde_share: 1 "nosecurity_"
		spellcheck	b
		Execute Query	"deny_token_document": [ " noscurity "
			nosecurity'
			"deny_token_share": {

So we notice that in the fields list, we have the field job with the value FranceLabs. It is exactly what we expected : we have a new field with a different value fo each job configured in MCF.

We can now configure the Datafari User Interface !

# Field already added to the Datafari schema

If you have already modified your indexation code and added your metadata, you have to respect these requirements for the Solr field :

```
<field name="job" type="string" indexed="true" stored="true" />
```

The field needs to be indexed and not tokenized. You can now configure the Datafari User Interface.

Configure the Datafari User interface

Our field is correctly present in our Solr schema and the values are present for each document. We just now have to configure the faceting in the Datafari UI.

We need to modify the search.js file in order to add the TableFacet widget, we also have to add the facet display in the searchView.jsp and finally add the label of the facet in the i18n files : en.json and fr.json.

First, edit the file search.js into /opt/datafari/tomcat/webapps/Datafari/js, add the widget in the code :

```
Manager.addWidget(new AjaxFranceLabs.TableWidget({
elm : $('#facet_job'),
id : 'facet_job',
field : 'job',
name : window.il8n.msgStore['job'],
pagination : true,
selectionType : 'OR',
returnUnselectedFacetValues : true
}));
```

We need to have an unique identifier in elm and id, we choose to call it facet\_job, in the field parameter we indicate job, for the name we put job.

The file seems like that :



• Edit now the file searchView.jsp located in /opt/datafari/tomcat/webapps/Datafari

We add the facet element to display on the page.We add it on the div section called col left :

<div id="facet\_job"></div>

The order is important, by default we have the facets in this order : by date, type and source. If I want to the job facet just after the source facet, the file seems like that :

<pre>cscript type="text/javascript" are="ja/desktop_design.ja"&gt;</pre>	
et and the second s	
if (ScriptConfiguration.getFroperty("DEDUFFICATION_IS_ENABLED")'=null && ScriptConfiguration.getFroperty("DEDUFFICATION_IS_ENABLED").equals("true") )(	
<pre>cscript type="text/javascript" src="js/doublon.js"&gt;c &lt;{}</pre>	
<acript arc="ja/search-ja" type="text/javascript"></acript>	
<script <="" td="" type="text/javascript"><td></td></tr><tr><td><pre>src="js/mav_mobile.js"></pre>/script></pre></td><td></td></tr><tr><td><jspinolude page="header.jsp" /></td><td></td></tr><tr><td><pre><div id="facets_mobile" style=""></pre></td><td></td></tr><tr><td><pre><div id="mar mobile" class="bc-exitor"> </pre></td><td></td></tr><tr><td>c:== radu sectre un trud dynamique par rapport 1/2 ia iangue =>> <a here=#iprovious nobiler/sic labas=#if achetron-left=></a></td><td></td></tr><tr><td>(/divs</td><td></td></tr><tr><td><div id="facet source mobile"></div></td><td></td></tr><tr><td><div id="facet_type_mobile"></div></td><td></td></tr><tr><td></day></td><td></td></tr><tr><td><div id="solr"></td><td></td></tr><tr><td><divid="searchBar"> <a href="/Datafari/Search">cimg src="ccs/images/logo_zebre.png"/></a>/div></td><td></td></tr><tr><td><pre><div id="results make" class="bc-closs"></td><td></td></tr><tr><td><pre></pre></td><td></td></tr><tr><td><pre></pre></td><td></td></tr><tr><td><pre></pre></td><td></td></tr><tr><th></th><th></th></tr><tr><th><div class="col left"></th><th></th></tr><tr><td><div id="facets"></td><td></td></tr><tr><td><pre><dirui="face_date"></dir></td><td></td></tr><tr><td><pre></pre></td><td></td></tr><tr><td>divid="facet language"></td><td></td></tr><tr><td><pre>cdiv id="facet job"></pre>/div></pre></td><td></td></tr><tr><td><pre><div id="facet signature"></div></pre></td><td></td></tr><tr><td></data/</td><td></td></tr><tr><td></div></td><td></td></tr><tr><td><div class="col right"></td><td></td></tr><tr><td><pre><div id="promoline"></div> </pre></td><td></td></tr><tr><td></td><td></td></tr><tr><td>c/div</td><td></td></tr><tr><td><div class="clear"></div></td><td></td></tr><tr><td></41x></td><td></td></tr><tr><td><jsp:include page="footer.jsp"/></td><td></td></tr><tr><td><div id="spinner_mobile"×i class="fa fa-spinner fa-spin"×/1×/div></td><td></td></tr><tr><td></br></td><td></td></tr><tr><td>s/html></td><td></td></tr></tbody></table></script>	

• Ok now the final step is to localize the label name of our new facet. Let's edit the two files en.json and fr.json in /opt/datafari/tomcat/webapps/Datafari/js/AjaxFranceLabs/locale :

We add a new line for the parameter 'job'. Here we add the same value for both the languages : Job. So the line to add is :

"job" : "Job",

Screenshot of the fr.json file :

GNU nano 2.2.6	FichierÅ : fr.json	
weeklyDelay" : "Quand lancer les alertes hebdomadaires",		
port" : "NumÁ@ro du port",		
database" : "Nom de la base de donnÃ@es",		
collection" : "Nom de la collection",		
missingParameter" : "Certains paramk"tres sont vides, des valeurs par dÀDfault	seront utilisÅDes Å leur place",	
preateAlerts" : "Ajouter ou ÄDditer une alerte",		
addAlerts" : "Ajouter des alertes",		
alertsSearch" : "Rechecher d'alertes via le mot-clef, si vous voulez toutes le	s aletes, laissez de champ vide",	
previous": "PrAScA0dent",		
duplicates" : "Doublons",		
undefined" : "indk@finie",		
mongoError" : "Le serveur n'a pas pu se connecter À la base de donnÀ®es Mongol	DB, assurez-vous que la configuration soit correcte.",	
	nt les alertes, avant de se rÅ@enclencer en fonction de leur frÅ@quence, si la date est celle d'aujourd'hui et l'heure actuelle Å	5 minutes environ le lanc
hint2" : "Ici vous pouvez spĥDcifier l'adresse, le port, le nom, de la base de	donnADes ainsi que la collection dans laquelle se trouvent vos alertes.",	
previousExecution" : "Ces alertes ont ADTAD ADXADcutADs la dernià re fois le :		
nextExecution" : "La prochaine ÄdxÄdcution est prÄdvue le : ",		
takingAccount" : " ne tient pas compte d'A@ventuelles modifications postA@rieu	res A la derniA're ABxecution*,	
parameterSaved" : "ParamĀ"tres sauvegardĀDs!",		
fieldWeight" : "PondADration des champs",		
pf" : "RequÂ*tes  mots multiples",		
gf" : "RequÃ*tes  mot unique",		
fieldList" : "Liste des champs",		
indexed" : "IndexAD",		
multivalued" : "Valeurs multiples", stored" : "StockÅ@".		
stored" : "StockAB", required" : "Requis".		
required" : "Requis", name" : "Non".		
type" : "Type", indexField" : "Champs de l'index",		
indexrieid" : "Unamps de l'index", schemaànalvsis" : "Analvse du schäGma.		
sizeLimitation" : "Limitation des tailles".		
autocompleteConf" : "Configuration de l'autocompletion" .		
labelhl" : "Nombre maximum de caractà"res analvsåßs par le surlignage".		
mailConf" : "Configuration du mail".		
address" : "Address : ",		
pass" : "Password : ",		
deduplication" : "Deduplication",		
limitHL" : "Limitation de la taille des champs pour le surlignage".		
autocompleteConf" : "Configuration de l'autocompletion".		
labelth" : "Fraction".		
iob": "Job".		
limitTH" : "Pourcentage minimal de documents, oã' le terme doit apparaã@tre"		
interin i rearceneage minimax ac accaments, on re terme doit apparametre		

• The configuration is now over. We can check if our new facet is correctly displayed in Datafari :

sold	٩	Administra
	🕐 Al words 🥌 Al least one word 🌞 Exist expression	
Results 1 - 10 de 265 pour soir (i	A03 Seconds	Sort by : Relevancy
ast Modification	souther interview of the state of the sta	
Less than 1 month (0)	name**laywooti' content**lusere solt, apache solt, moturd a moture observation or internationality of an annual solutionality of an annual solution of a sol	
Less than 1 year (78)	ied* langs*ien*> <	
Less than 6 years (84)	http://www.funneelabs.com/solt.html	
	soft.html	
Гуре	Littlefth vigitarisative vi France Laboritative vineta narva*ilevisorisi contente*lucene solt, apache solt, solt, open source search ergine* />	
prg (22)	Kineta name="description" contant KIDOCTYPE html: KI-jf k IE 7 (>html: desa="e #d" lang="en"> 41	
html (8)	http://www.faracelabe.com/en/lab/html	
	logo_sofr-small.png	
Source	http://www.france/abb.com/magesthame-picelogo_sol-warall.png	
web (205)		
	formation_sofr_unlesshed.html #\idea disperfunctions bid Unlesshed in Strate Laboration - create annex*"/seported/or "	
anguage	<ul> <li>a use - Approximations per functional constructions for the previous contract - bandwards - structures and - Annual names - bandwards - bandwa constant-list - bandwards - bandwa - bandwards - ban</li></ul>	
🗌 en (255)	http://www.farcolubia.com/formation_asit_unitexasitati	
	formation solr avances.html	
Job	Inimation_son_article_texture     */std> * stdy=states	
FranceLabs (255)	content="Le formation \$oir Avancée <idoctype dl<="" dlig.tll.com="" ent="" http:="" s="" td=""><td></td></idoctype>	
Datafari (10)	http://www.fancelabs.com/formation_solv_evancee.html	
	formation solr unleashed.html	
	"/16%" robe-36/t Unleashed twinings << Finne Labsr/Dite> rmeta name*"keywords" content*"taining self unleashed" /> rmeta name#"description"	
	contente"Sofe unlesshed Tashings with Lucideoks	
	http://www.francelabs.com/enformation_sob_univasited.html	
	formation solr initiale.html	
	Vertile#Ve	
	content#"La formation Solr Initiale «IDOCTYPE html» <=/for Life #0" lange="entity" = 0" lange="entity" <=	
	http://www.fanceiabs.com/formation_sole_initiale.html	

It seems to be the case. We can now filter the results by the ManifoldCF job : documents from Datafari website or France Labs website !

# Fieldweights - Architecture and mechanism

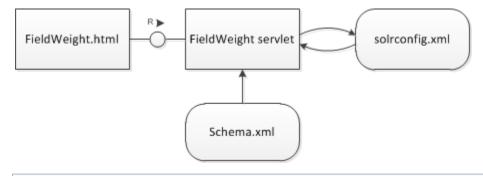
The fields weight are stored in the *solrconfig.xml* file of the FileShare core.

The admin UI proposed in Datafari communicates with the FieldWeight servlet, which in turn:

- reads the schema and solrconfig files,
- modifes the solrconfig.

When you go on the UI, it immediately sends a request towards the servlet to get the fields that are appended to the select.

Selecting a field will acquire a semaphore to prevent people from reading or writing a value that you are about to change.



Note that there are two types of query with their own weight for each fields, so there is one semaphore per type, since you will not have to modify the same line in the file.

Once you selected a file, you will get its current weight (0 if not weighted). Confirming your modifications will send the request and release the semaphore. Due to the particular structure of the lines involved, this functionnality has its dedicated servlet, but if you want to modify the content of a node, you can use the ModifyNodeContent servlet.

#### ModifyNodeContent

It is a servlet that opens solrconfig.xml and looks for a node with a specified attribute. For instance, assuming there is a node with parameter "name" equal to "threshold", and we want to set the node value to 0.01:

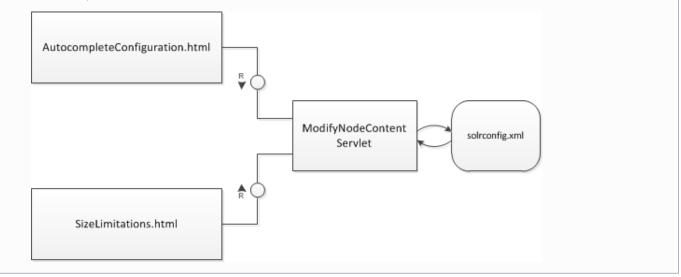
This is done by applying the following parameters:

• <float name="threshold">0.005</float>

This line will then become :

• <float name="threshold">0.01</float>

This servlet is currently used by the size Limitation UI, and by the autocomplete configuration UI. This allows to prevent two or more users from modifying the same value. This servlet creates a semaphore on the doGet. The semaphore is based on the value of the attribute: in our previous example, "threshold", someone can modify the size limitation and someone else can modify the threshold of the autocomplete.



# **Likes and Favorites**

## **Problematic :**

Starting with Datafari 2.0, users have the possibility to store pointers to documents shown in results list.

In the following document, "likes" corresponds to the functionality where a connected user can like or unlike a document, and "favorites" corresponds to the functionality where a connected user can save a pointer to a document he found in the results list.

Technically wise, we use Apache Cassandra for saving all the likes and favorites and Solr for saving the counters of likes of each document.

Using Cassandra, we cannot use standard relational db schemas, like the following:

id_fav orite	id_article	id_user
11	111	1235
12	112	1456

id_like	id_article	id_user
11	111	1235
12	112	1456

Modeling with Nosql in mind:

For our nosql modeling, we took our inspiration from this article http://docs.mongodb.org/manual/tutorial/aggregation-with-user-preference-data/.

Applied to our problematic, we get the following :



That means we have dedicated to likes and favorites (which is distinct from the collection dedicated to user authentication and authorization).

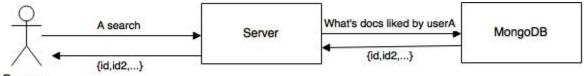
#### The request :

The aim is to get for the connected user the list of documents he liked out of the ones which are currently displayed. After the standard search query, we send only one http request which contains only the username. In return, we get all the documents in the index that are liked, as well as the list of favorites.

On the browser side, we go through this documents list and pick the ones that match the currently displayed documents to identify the favorites and the liked.

In order to display the total number of likes per documents, we use Solr and its External File Fields mechanism (https://cwiki.apache.org/confluen ce/display/solr/Working+with+External+Files+and+Processes).

Every time a connected user like or unlikes a document, we increment or decrement the counter of the document in the file located in SOLR\_HOME/FileShare/data/externalk\_nbLikes.





## The Implementation :

There are two parts: Frontend and backend. For the frontend, we added a widget that builds the inteface and makes all the Ajax Requests. For the backend, we created two classes: Like and Favorite. We created the corresponding servlets that use these two classes and send back the result.

1. Frontend :

The widget is called LikesAndFavoritesWidget. This widget inherits from ResultWidget and is only included in SearchView.jsp if the functionality of likes was activated by the admin.

This inheritance is important because we need to execute first the code of the ResultWidget and then the LikesAndFavoriteWidget.

Once the buildWidget is executed, we execute the beforeRequest which will add the fl nbLikes:field(nbLikes) in the Solr query so we can retrieve as part of the results the number of likes per document.

Then the afterRequest will fire. The afterRequest retrieves the documents liked and saved by the user from the server (ajax request) and stores it in window.globalVariableLikes and window.globalVariableFavorites. Then we parse the results retrieved from Solr and put the correct values in html by checking if the doc belong in the window.globalVariableLikes and window.globalVariableFavorites.

Next, we put a listener to the "Like" button and a listener for saving a favorite. Depending on which button is clicked, we send a request to add/delete the favorite or the like from the server and from the lists saved in the global variable window of javascript.

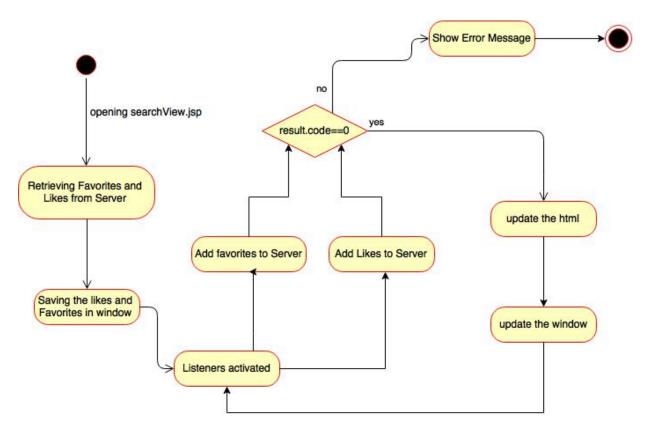


Illustration 1: State Diagram of AfterRequest of LikesAndFavoritesWidget

#### 2. Backend :

We use two major classes which are Like And Favorite. They provide the method that will get, delete and add a Favorite or a Like by querying the Cassandra database. Whenever a change is needed to likes and favorites, we use directly these classes.

Like.class and Favorite.class belong to the package com.francelabs.datafari.user which also contains UserConstants.class where we put all the constant Strings used in this package.

We also use the following servlets:

- AddFavorite: add a document to favorites,
- · AddLikes: add document to likes,
- · DeleteFavorite: delete a document from favorites,
- Unlike: delete a document from the likes,
- GetLikesAndFavorites: gets all the likes and the favorites of a user,.,.

AddLike, AddFavorite, DeleteFavorite and Unlike are very similar. Therefore we will detail only one of them, the others can be derived from it. This is the sequence diagram involving the AddLike servlet:

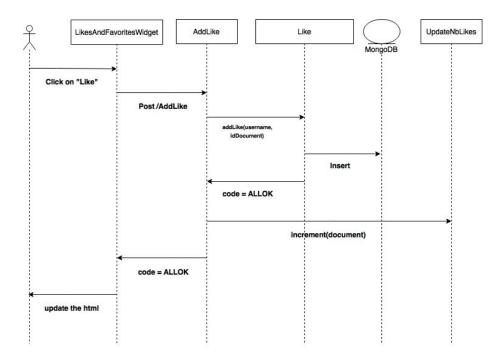


Illustration 2: Sequence Diagram of AddLike for a normal Behaviour

Here is the sequence diagram that involves the LikesAndFavoritesWidget to retrieve the likes and favorites of a user :

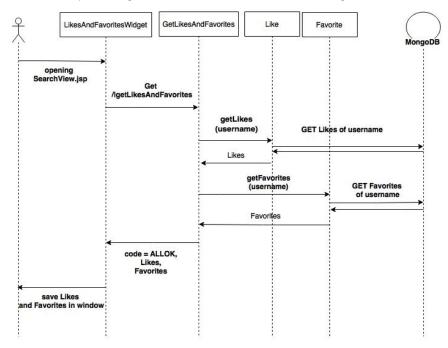


Illustration 3: Sequence Diagram of getting the likes and favorites from server

# Security

This section details the way Datafari manages security.

# CustomCombinedRealm - Architecture and mechanism

In Datafari, we propose a user management following the standard Role Based Access Control mechanism.

We use two steps:

- the authentication step, to ensure a user is who he claims to be (using credentials validation),
- the authorization step, to ensure the authenticated user is allowed to see a certain part of Datafari (be it admin pages or secured documents in the search results list).

To manage that propery, we rely on the Tomcat Realm mechanisms. Still, there was no existing realm to satisfy our needs, which are to do the authentication either on a remote LDAP/AD or on our Cassandra, and to do the authorization on our Cassandra (and Solr for the search part, but this is another story). This is why we have created CustomCombinedRealm and a user data model in our Cassandra.

CustomCombinedRealm is a class that belongs to the package com.francelabs.datafari.realm.

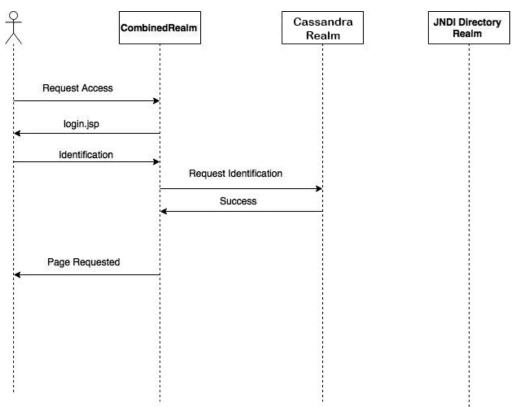
This class inherits from CombinedRealm and differs from its super class by collecting the roles.

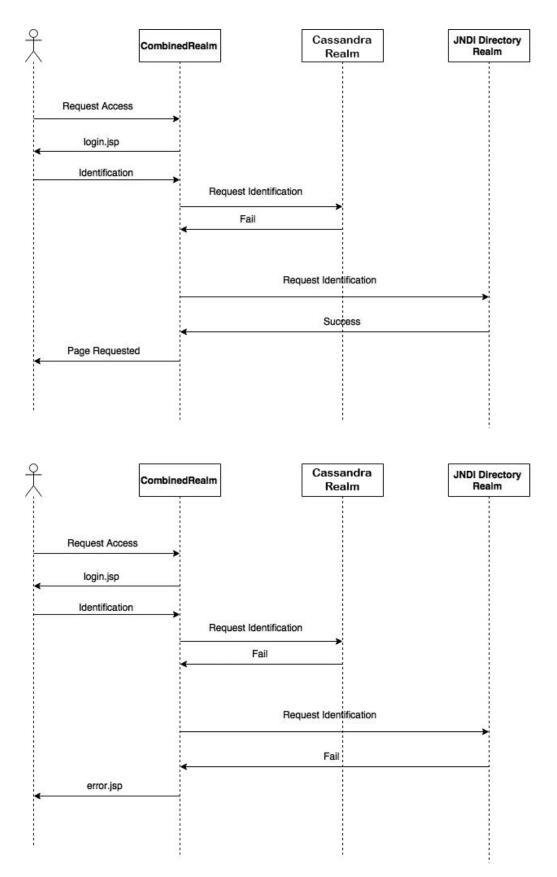
We developed this class to retrieve the roles of a user from our Cassandra database, and the user authentication either from AD/LDAP or from Cassandra. That means that even if we use LDAP for authentication, we will get the corresponding roles from Cassandra and not from LDAP.

The process is as follows:

- we start by requesting an authentication in Cassandra.
- In case of success, we get the corresponding roles from it but if it fails we use LDAP for authentication.
- If the authentication succeeds, we get then the roles from Cassandra and if it fails we return an error to the user.

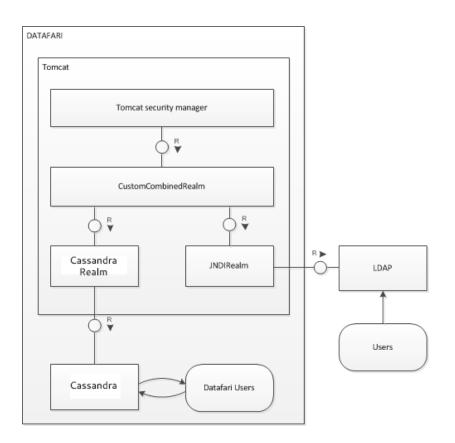
You can see below the illustration of these steps.





As shown above, CustomCombinedRealm communicates with the two Realm : CassandraRealm and the JNDIRealm (AD/Ldap Realm). These classes will communicate respectively to there databases and check if the authentication succed and return the response to the CustomCombinedRealm.

Here is the overall architecture to understand the way it works:



# **Understanding Security Realms**

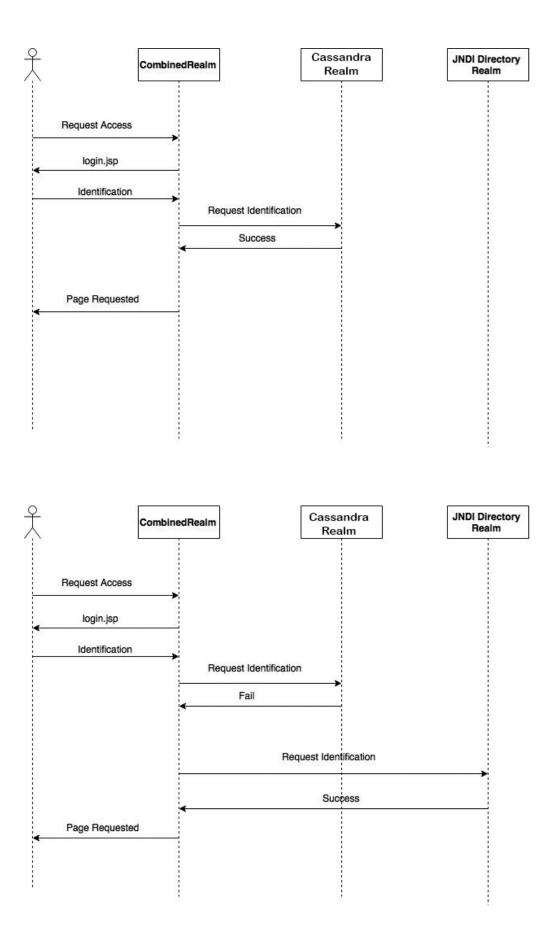
This document explains how user management is done in Datafari, following the RBAC (Role Based Access Control) model.

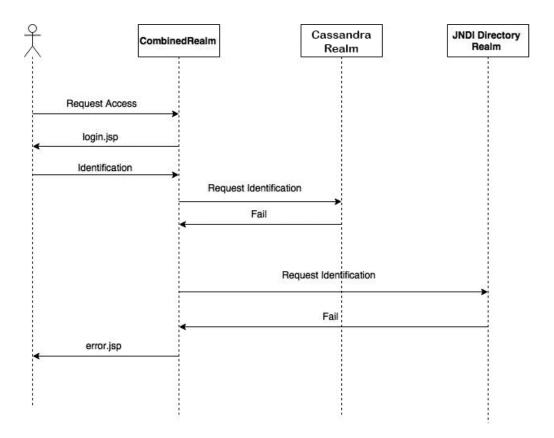
Datafari uses three Tomcat Realms for the authentication. This functionality is used to recognize the user and give him a personalized environment.

Also, every user has a specific role which will allow the user to have access to specific areas of Datafari and to prevent others from having this ability.

This functionality is of no use for persons who install Datafari on their personal computers.

The steps for authentication is illustrated in the three pictures below.





As shown above, when a user wants to authenticate, Datafari first calls CombinedRealm.

CombinedRealms triggers CassandraRealm, which will check the authentication using the database of Cassandra. If it fails, CombinedRealm trigg ers JNDI DirectoryRealm that will check the authentifcation using the AD/LDAP you have configured. These steps are done using the user credentials provided.

It is important to note that AD/LDAP is not required for Datafari to work : You can use only Cassandra without changing any setting.

Since Datafari uses 3 realms, they need to be configured. *CustomCombinedRealm* and *GenericCassandraRealm* should not require you to modify their configuration, as they are embedded in the Datafari package. Still, in case you want to customise them, it is feasible. The realm that requires configuration is the *JNDIRealm*, for the remote AD/LDAP connection. These configurations are done in *context.xml*, which is in *WebCont ent/META-INF* :

```
<?xml version="1.0" encoding="UTF-8"?>
<Context>
 <Realm
 className="com.francelabs.realm.CustomCombinedRealm">
  <Realm
     authDB="db-containing"
     authCollection="users"
     authUserField="username"
     authPasswordField="password"
     authRoleField="role"
     className="com.francelabs.realm.GenericMongoRealm"
     defaultDbHost="localhost"
     defaultDbPass=""
     defaultDbUser=""
     defaultRole="user"
     digest="SHA-256"/>
  <Realm
     className="org.apache.catalina.realm.JNDIRealm"
     connectionURL="ldap://ldap.forumsys.com:389"
     userPattern="uid={0},dc=example,dc=com"
     connectionName="cn=read-only-admin,dc=example,dc=com"
     connectionPassword="password"/>
 </Realm>
</Context>
```

Regarding the last Realm above, which relates to the AD/LDAP, its configuration depends on your AD/LDAP system. To configure it properly, please follow the Tomcat 7 JNDIRrealm howto page (url: https://tomcat.apache.org/tomcat-7.0-doc/realm-howto.html )

Regarding the Cassandra realm configuration: by default, the Cassandra database requires no username and password. Still, you should definitely add a password and a username for the connection your Cassandra. Once this is done, fill in these information in *defaultDbUser* and *def aultDbPass*.

# **Semantics**

In this section, you will find information on the "semantics" capabilities of Datafari. We put semantics in brackets, because some of you may consider that certain functionalities are not real semantics. But since we noticed some of our competitors don't hesitate to flag as semantic the simple use of a dictionary to handle synonyms, we've decided to create this section. So you will find here details about the management of synonyms and stopwords, promolinks, and ontologies.

# Link an ontology

Datafari gives you the opportunity to link an ontology to it, in order to enrich your documents with the ontology data and then use these additional infos in the search phase.

To do so, the first step is to add the OntologyUpdateProcessor as a custom UpdateProcessor in the file {Datafari\_Home}/solr/solr\_home/FileShar e/conf/customs\_solrconfig/custom\_update\_processors.xml

The OntologyUpdateProcessor will index, for each document inserted in Solr, the labels of the corresponding node into the specified ontology but also its parents and children labels and URIs.

It has several parameters that can be set:

- **enabled** : boolean value to enable/disable the OntologyUpdateProcessor
- annotationField : [REQUIRED] the field in the input document that contains the annotation URI used as the reference in the ontology.

Note that this field must be added to the document by yourself: Datafari does not provide any plugin or other mechanism to do it. You may want to modify the DatafariUpdateProcessor or make your own UpdateProcessor to do so.

• ontologyURI : [REQUIRED] the location of the ontology, such as http://francelabs.com/ontology/owl.rdf or file:///ontology/owl.rdf

Datafari only supports OWL ontology format for now, so the ontologyURI parameter has to reference a OWL formated ontology

- labelField : the field in your schema that should be used for the annotation's label(s). Default: ontology\_labels
- **childField** : the field to be used for child document references. These are direct (ie. single-step) relationships *down* the hierarchy. Default: *ontology\_children*
- parentField : the field to be used for parent document references. These are direct relationships up the hierarchy. Default: ontology\_pare nts
- childLabel : the field to be used for child documents labels. Default: ontology\_children\_labels
- parentLabel : the field to be used for parent documents labels. Default: ontology\_parents\_labels
- **useLanguages** : (boolean) should language distinction for label fields be used? If this is set to *true*, the processor will create additional label fields for each language found, using the language as a suffix, for instance *ontology\_parents\_labels\_fr*, *ontology\_parents\_labels\_e* n

The field format is [Original\_Label\_Field]\_[Language]

This is useful if you want to have, in addition to the original label fields which will contain all values, specific label fields containing values of the concerned language. But remember that the more the Ontology contains languages, the more it will create label fields. Default: *false* 

- includeIndirect : (boolean) should indirect parent/child relationships also be indexed ? If this is set to *true, all* ancestor and descendant labels and URIs will also be added to the document. Default: *false*.
- descendantsField : the field to be used for the full set of descendant references. These are direct AND indirect relationships down the hierarchy. Default: ontology\_descendants
- ancestorsField : the field to be used for the full set of ancestor references. These are direct AND indirect relationships up the hierarchy. Default: ontology\_ancestors
- descendantsLabel : the field to be used for descendant documents labels. Default: ontology\_descendants\_labels
- ancestorsLabel : the field to be used for ancestor documents labels. Default: ontology\_ancestors\_labels

If you decide not to use the default values for the fields, DO NOT FORGET to add them in the appropriate custom schema file(s) which are located in {Datafari\_Home}/solr/solr\_home/FileShare/conf/customs\_schema Otherwise none of the ontology features will work !

To implement the OntologyUpdateProcessor, follow these steps:

- open the 'custom\_libs.xml file in {Datafari\_installation\_folder}/solr/solr\_home/FileShare/conf/customs\_solrconfig/
- add the following line (replace the '<to\_remove\_tag>') :

\_\_\_\_\_

<lib dir="./lib/jena"/>

Datafari uses Apache Jena to load an ontology

- open the 'custom\_update\_processors.xml file in the same location
- add the following lines(still replace the '<to\_remove\_tag>'):

Of course you have to replace the values by yours and also add the optional parameters described up above with what you want

- Restart Datafari
- Crawl (again) your documents to add to them the new fields generated by the OntologyUpdateProcessor

Then, you can configure Datafari to use the new Ontology infos on the search requests :

- open the 'datafari.properties' file in {Datafari\_installation\_folder}/tomcat/conf/
- Modify those lines to set the ontologyEnabled property to true and the other lines to match with your ontology configuration:

ontologyEnabled=true ontologyLanguageSelection=true ontologyNodeLabels=ontology\_labels ontologyChildrenLabels=ontology\_children\_labels ontologyParentsLabels=ontology\_parents\_labels

ontologyEnabled : (boolean) Enable or disable the ontology use ontologyLanguageSelection : (boolean) should use the Datafari selected language for the ontology labels ? If *true*, be sure that you used the OntologyUpdateProcessor with the parameter userLanguages set to *true* during the crawl of the documents by MCF ontologyNodeLabels : the field in your schema that contains the annotation's label(s) ontologyChildrenLabels : the field in your schema that contains the child documents labels ontologyParentsLabels : the field in your schema that contains the parent documents labels

Your Datafari is now ready to use the ontology additional fields during search queries thanks to the OntologySuggestion.widget

# **Promolinks - architecture and mechanism**

Promolinks are indexed in a solr Core. The configuration is located in \$DATAFARI\_HOME/solr/solr\_home/Promolinks/conf

A Promolink contains five fields :

- a keyword (mandatory field) which is unique since it is also the id of the PromoLink,
- a title (mandatory field),
- a content (mandatory field),
- a starting date (before this date the PromoLink will not be displayed) (optional),
- an ending date (after this date the capsule will not be printed) (optional).

You can see all the existing Promolinks in the Admin UI. You can also add or delete or edit a Promolink on this page. This UI communicates with the admin servlet which communicates with the solr core.

When you make a search on Datafari, once SearchProxy's doGet, two queries are sent:

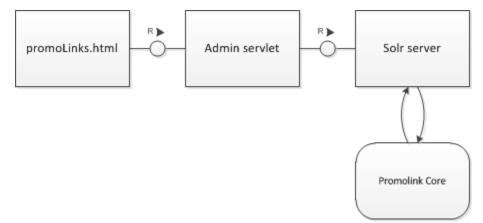
- One is sent to FileShare to retrieve the documents
- Another is sent to the PromoLink core.
  - If a promolink is found, there will be a check of the dates (if those are filled).

- If dates are not filled, the promolink will be considered as valid and it will be put in the promolinkSearchComponent key of the json object returned at the end.
- Then the promolink widget will print it at the beginning of the result list.

The Solr fieldtype chosen for the field keywork is named capsuleType :

```
<fieldType name="capsuleType" class="solr.TextField"
positionIncrementGap="100">
<analyzer>
<tokenizer class="solr.WhitespaceTokenizerFactory"/>
<filter class="solr.LowerCaseFilterFactory" />
<filter class="solr.ASCIIFoldingFilterFactory"/>
</analyzer>
</fieldType>
```

Basically it splits the token at each blank character, it puts the string in lowercases and removes the accentuated characters. So it supports multiple words.



# Synonyms and stopwords

Those two functionnalities are very similar since they use the corresponding Solr basic functionnalities that rely on a text file.

Unless it has been modified, those text files are located in the conf folder of the FileShare core.

The Admin UI allows the user to see and modify the files.

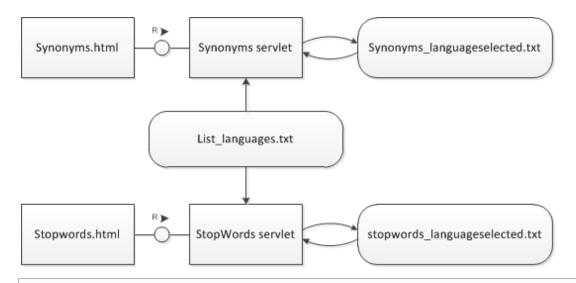
The syntax for the synonyms is as follows :

```
i-pod, i pod => ipod,
sea biscuit, sea biscit => seabiscuit
```

The syntax for the stopwords file is the following one :

a an to from

To prevent two people from modifying the same text file at the same moment, the servlets dedicated to these functionnalities use a mutex semaphore. There is also a list\_language.txt that you have to fill, which defines where to apply the semaphores.



You can see the effect of the stopword filter in the solr admin UI analysis.

# Setup development environnement with Eclipse (Linux)

**DEPECRATED** This page is deprecated as of Datafari 2.0

**Clone repository** 

git clone https://github.com/francelabs/datafari.git

Open the project in Eclipse

Open a terminal. Navigate to the folder datafari. We now call it DATAFARI\_SRC.

Run an : "ant eclipse" in DATAFARI\_SRC.

In eclipse, do an « import existing project into workspace ».

Select the DATAFARI\_SRC folder and click Finish.

8	Import		- + ×
Import Projects			
Select a directory to sea	arch for existing Eclipse projects.		
Select root directory	/home/mazoyer/Projets/datafari/dataf	•	Browse
<ul> <li>Select <u>a</u>rchive file:</li> </ul>		-	B <u>r</u> owse
Projects:			
🗹 Datafari (/home/m	nazoyer/Projets/datafari/datafari/dev-tools	s/ecl	Select All
			Deselect All
			R <u>e</u> fresh
Options			
Searc <u>h</u> for nested p	rojects		
<u>C</u> opy projects into v			
	lready exist in the workspace		
Working sets Add projec <u>t</u> to wor	ling cots		
Working sets:	king sets		Select
Working Secs.			<u>Je</u> rceen.
oc 🗟 Declaration			
at this time.			

# Setup the Tomcat server in Eclipse

Go to Window => Preferences => Server => Runtime Environments => Add...

and select Apache Tomcat V. 7.0

Select DATAFARI\_SRC/tomcat for the Tomcat Installation directory and Finish.

≅	<u>م م م بالا بال بالم بالم بالم</u> Preferences	- + ×	Quick	Access
	type Filter text 🐼 Server Runtime Environments	(> → → → →		
	Edit Server Runtime Environment			
		- + ×		
	Tomcat Server     Specify the installation directory			
	Name: Apache Tomcat v7.0			
	Tomcat installation directory:			
	/home/mazoyer/Projets/datafari/datafari/tomcat	Browse		
	apache-tomcat-7.0.47 Dowr	nload and Install		
		stalled JREs		
	• f [		🖳 v 📑 v 🖳 🗖	🗄 Out An out
	? < Back Next > Cancel	Finish		
	▶ Team			
	Terminal			
	Validation Veb			
	Web Services			
	XML     /home/mazoyer/Projets/datafari/datafari/tom	ncat		
	Cance	ОК		
6				

In a server windows (for example in Debug Perspectives), click on "add a server", select a "tomcat 7.0 server" and set the server name to "Datafari":

4 101 100 110 1 =0 😽   🎋 🗸 🔿 🗸					
	🖻 🎋 O 🖉 🔳 🗄		Breakpoints		
create a new server	۲	New Server	- + ×		
	Define a New Server				
	Choose the type of server t	o create			
	Select the server type:	Download add	ditional server adapters		
	type filter text				
	i Iomcat vs.5 server				
	🗄 Tomcat v6.0 Server				
	🗄 Tomcat v7.0 Server				
	Tomcat v8.0 Server				
	Publishes and runs J2EE and Tomcat server.	d Java EE Web projects and server con	figurations to a local		
				<b>→</b>	🗄 Outline 🖾
	Server's host name:	° localhost			An outline is not availab
	Server name:	Datafari			-
	Server runtime environmen	t: Apache Tomcat v7.0	¢ Add		
		Configure	runtime environments		
	?	< Back Next > Cance	el Finish		
	<u> </u>		,		

#### Double click on Datafari to change its configuration :

🗄 Datafari 🛛					-
Overview					
General Information	and other common settings.		Publishing		
Server name:	Datafari		Timeouts		
Host name:	localhost		▼ Ports		
Runtime Environme	Apache Tomcat v7.0	:	Modify the server ports.		
Configuration path:	/Servers/Datafari-config	Browse	Port Name	Port Number 8005	
Open launch configu	ration		HTTP/1.1	8080	
<ul> <li>Server Locations</li> </ul>			🛱 AJP/1.3	8009	
Specify the server pat make changes.	h (i.e. catalina.base) and deploy path. Server must be published with no modules p	esent to			
-	netadata (does not modify Tomcat installation)		MIME Mappings		
	Ilation (takes control of Tomcat installation)				
🔿 Use custom loca	tion (does not modify Tomcat installation)				
Server path: .met		Browse			
Set deploy path to t	he default value (currently set)				
Deploy path: wtpv	<i>r</i> ebapps	Browse			
<ul> <li>Server Options</li> </ul>					
Enter settings for the	server.				
Serve modules w	ithout publishing				
Publish module	ontexts to separate XML files				
🗹 Modules auto re	oad by default				
Enable security					
Enable Tomcat debug logging (not supported by this Tomcat version)					

```
Overview Modules
```

Click on « Use tomcat installation » and on « Open launch configuration ».

In argument, add this to VM arguments :

-Dsolr.solr.home="DATAFARI\_SRC/solr/solr\_home" -Dorg.apache.manifoldcf.configfile="DATAFARI\_SRC/mcf/mcf\_hom e/propertiesDev.xml"

And replace the parameter catalina.home by the path of your DATAFARI\_SRC/tomcat :

#### -Dcatalina.home="DATAFARI\_SRC/tomcat"

Finally, if you want to access to the manager page of Tomcat and the Banana page, you have to copy/paste the webapps into the webapps folder of your Eclipse workspace :

cp -R DATAFARI\_SRC/tomcat/webapps/ROOT your-eclipse-workspace/.metadata/.plugins/org.eclipse.wst.server.core/tmp0/webapps/

cp -R DATAFARI\_SRC/tomcat/webapps/manager your-eclipse-workspace/.metadata/.plugins/org.eclipse.wst.server.core/tmp0/webapps/

cp -R DATAFARI\_SRC/tomcat/webapps/banana.war your-eclipse-workspace/.metadata/.plugins/org.eclipse.wst.server.core/tmp0/webapps/

In java EE perspective, select File Import and select Web War File :

÷	Import	- + ×
Select		
Import an external WAR file into	a Web Project	
Select an import source: Lype filter text Java EE Mayen		2
<ul> <li>Image: Plug-in Development</li> <li>Remote Systems</li> </ul>		
<ul> <li>Prince Run/Debug</li> <li>Prince SVN</li> <li>Prince Tasks</li> </ul>		
<ul> <li>▷ Team</li> <li>▼ ▷ Web</li> </ul>		
ै WAR file ► ➢ Web services		
? < Bac	Next >	Cancel Finish

Browse to DATAFARI\_SRC/solr/war/solr.war and set name of web project to datafari-solr.

Do the same for DATAFARI\_SRC/mcf/war/mcf-api-service.war with the name datafari-mcf-api-service, for DATAFARI\_SRC/mcf/war/mcf-crawler-ui.war (web project name : datafari-mcf-crawler-ui) and for DATAFARI\_SRC/mcf/war/mcf-authority-service.war (web project name : datafari-mcf-authority-service).

Go back to the server window. Right click on Datafari, and do an Add and Remove... Add all the web app application (Datafari, datafari-solr, datafari-mcf-crawler-ui, datafari-mcf-api-service, datafari-mcf-authority-service).

	E 🌣 🖸 🖉 I	🔲 🗊 🔽 🗖 💷 💷	les 🛛 🂁 Breakpoints		0
		Add and Remove	- + ×	1	
	Add and Remove	that are configured on the server		📮 v 📑 v 🖳 🗖	E Outline ⊠
sr/lib/jvm/java-8-oracle/bin/java (17 févr. 2015 16:14: Marga prosebo potal inc. corso Arga atalina.properties 🖹 context.xml 🖹 web.xml >n settings.	Move resources to the Available:	e right to configure them on the sec Add > Add All >>	erver Configured: Datafari datafari-mcf-api-service datafari-mcf-authority-service datafari-mcf-crawler-ui datafari-solr		An outline is not available.
Ifari-config Browse e) and deploy path. Server must be published es. t modify Tomcat installation) trol of Tomcat installation)		, publish changes immediately			
Jify Tomcat installation) p.eclipse.wst.server.core/tmp0 Browse urrently set) Browse	0	< Back Next >	Cancel Finish		
to VML Filos				I	

Go to the « Servers » « Datafari Config » directory in the « Package explorer » of eclipse. Change the Datafari admin password in the tomcat-users.xml file. Copy DATAFARI\_SRC/tomcat/conf/datafari.properties in the Datafari Config directory. You can now run the Datafari tomcat server

## Run database and manifold agents

Open a terminal and set JAVA\_HOME environement variable and navigate to DATAFARI\_SRC/mcf/mcf\_home.

Run bash start-database-dev.sh.

In another terminal, run bash initialize-dev.sh and then start-agent-dev.sh.

Keep the two terminals opened.

## **Configure connectors**

Run Datafari tomcat. Go to the Datafari URL (http://localhost:8080/Datafari). Login in administrator. Go to Crawler tab-> Manifold CF Admin.

Log in and click on « List Output Connections » and create a new « Output Connection ». Set the name to « DatafariSolr ». Set Port to « 8080 », set web application name to « datafari-solr » and « Core/Collection name » to FileShare. You can now save:

Apache ManifoldCF™										23 Wr. 2015 18294
Outputs	Name Type Throttling	Solr type	Server	ZooKeeper	Paths	Schema	Arguments	Documents	Commits	Edit an Output Connection
List Output Connections	Protocol:									
Authorities		http 💌								
List Authority Groups	Server name:	localhost								
List User Mapping Connections	Port:	8080								
List Authority Connections										
Repositories	Web application name:	datafari-s	olr							
List Repository Connections	Core/Collection name:	FileShare								
Jobs										
List all Jobs	Connection timeout (seconds):	60								
Status and Job Management	Socket timeout (seconds):	900								
Status Reports	Realm:									
Document Status	neain.									
Queue Status	User ID:									
History Reports	Password:									
Simple History										
Maximum Activity	SSL trust certificate list:	No certifica	tes pres	ent						
Maximum Bandwidth		Add Ce	rtificate:	Parcourir	Aucu	n fichier s	électionné.			
Result Histogram										
Miscellaneous	Save Cancel									

Your development environnement is now ready.

# Setup development environment with Eclipse (Linux) for Datafari 2.0

## DEPRECATED

Deprecated : this page is for versions of Datafari older than 2.1, the one for 2.x is with Maven environment : Setup development environment with Eclipse (Linux) for Datafari 2.x - Maven

## **Pre-requisites**

- Have a JDK installed on your PC (suggested version Java 7)
- Have an IDE installed on your PC (in the guide we use Eclipse) and be able to lunch it as root
- · Have git installed on your PC

#### **Clone repository**

Open a terminal and navigate to the folder where you want to have Datafari source code to be checkouted (usually your workspace folder).

#### Perform a

git clone https://github.com/francelabs/datafari.git

to checkout the code. The root directory name is datafari.

## Open the project in Eclipse

Navigate to the folder datafari. We now call it DATAFARI\_SRC.

Run an : "ant eclipse" in DATAFARI\_SRC.

i If you get an error related to tools.jar not found, please check your JAVA\_HOME environment variable.



In eclipse, go to File -> Import... , expand General and do an « Import existing projects into workspace ».

Select the DATAFARI\_SRC folder as root directory and click Finish.

	٢	Import		- + ×					
	Import Projects           Select a directory to search for existing Eclipse projects.								
	Select roo <u>t</u> directory:	/home/mazoyer/Projets/datafari/dataf	•	Browse					
	○ Select <u>a</u> rchive file:		▼	Browse					
	<u>P</u> rojects:								
	🗹 Datafari (/home/ma	zoyer/Projets/datafari/datafari/dev-tools	s/ecl	<u>S</u> elect All					
				Deselect All					
				R <u>e</u> fresh					
	Options <ul> <li>Search for nested pro</li> </ul>	ieste							
	Copy projects into we								
		eady exist in the workspace							
	Working sets								
	Add project to working	ing sets							
	Working sets:		÷	S <u>e</u> lect					
🛚 Javadoc 🗟 Declaration									
display at this time.									
	?	< Back Next > Cancel		Finish					

## Setup the Tomcat server in Eclipse

Go to Window -> Preferences -> Server -> Runtime Environments -> Add...

and select Apache Tomcat V. 7.0

Select DATAFARI\_SRC/tomcat for the Tomcat Installation directory and Finish.

) .R   🗟 🕱   🎋 🔻 🔾 🔻 💁	م م م ب کا ب کا ب ک د ب ک ب Preferences –	+ ×
publish]	type filter text 🛛 Server Runtime Environments 🔶 👻 💠	▼
Jubrishj	Edit Server Runtime Environment - + >	×
	Tomcat Server Specify the installation directory	
	Name: Apache Tomcat v7.0	
	Tomcat installation directory:	
	/home/mazoyer/Projets/datafari/datafari/tomcat Browse	
	apache-tomcat-7.0.47 Download and Install	
	Workbench default JRE ‡ Installed JREs	
	P	📮 🕆 🖶 🗖 🚼 Outlin An outlin
	(?) < Back Next > Cancel Finish	
	Team Terminal Validation Web Web Services	
	XML /home/mazoyer/Projets/datafari/datafari/tomcat	
	Cancel   OK	

In a Servers window (for example in Debug Perspectives), click on "add a server", select a "tomcat 7.0 server" and set the server name to "Datafari":

4 LOL 10 LIK I HO 😽 😽 K Y 🗸 🗸 🖌 🖉 🐇				Quick	
		▽ □ □ (×)= Variables 🛙	Breakpoints		
create a new server		New Server	- + ×		
	Define a New Server				
	Choose the type of server to	create			
		Download	additional server adapters		
	Select the server type:				
	type filter text		$\boxtimes$		
	<ul> <li>Tomcac vs.s server</li> <li>Tomcat v6.0 Server</li> </ul>				
	Tomcat v7.0 Server				
	Tomcat v8.0 Server				
	Basic				
	Publishes and runs J2EE and J Tomcat server.	lava EE Web projects and server o	configurations to a local		
				🗉 v 📑 v 😐 🗖	🗄 Outline 🖾
	Server's host name:	localhost	]		An outline is not availab
	Server name:	Datafari			
	Server runtime environment:	Apache Tomcat v7.0	\$ <u>Add</u>		
		Configu	ire runtime environments		
	?	Back Next > Ca	Finish		

Double click on Datafari to change its configuration :

<sup>8</sup> Datafari ⊠				- ,
Overview				
Overview     General Information     Specify the host name a     Server name:     Host name:     Host name:     Configuration path:     Open launch configura     Server Locations     Specify the server path     Was romcat install     Use custom locatio     Server path: Instance     Server path: Instance     Server path: Instance     Server path: Instance     Server paths:     Enter settings for the se     Server modules wit     Publish modules core     Modules auto reloi     Enable security	// Servers/Datafari-config tion i.e. catalina.base) and deploy path. Server must be published with no modules pre tadata (does not modify Tomcat installation) tition (takes control of Tomcat installation) n (does not modify Tomcat installation) default value (currently set) appps rver. nout publishing texts to separate XML files	: Browse	<ul> <li>Publishing</li> <li>Timeouts</li> <li>Ports</li> <li>Modify the server ports.</li> <li>Port Name</li> <li>Format admin port</li> <li>HTTP/1.1</li> <li>AJP/1.3</li> <li>MIME Mappings</li> </ul>	Port Number           8005           8080           8009
Overview Modules				

In Server Locations section: select « Use Tomcat installation ».

In General Information section, click on « Open launch configuration ».

```
In Arguments tab, add this to VM arguments :
```

```
-Dorg.apache.manifoldcf.configfile="DATAFARI_SRC/mcf_mcf_home/propertiesDev.xml"
-Dsolr.solr.home="DATAFARI_SRC/solr_home/"
```

[Optional] And replace the parameter catalina.home by the path of your DATAFARI\_SRC/tomcat :

```
-Dcatalina.home="DATAFARI_SRC/tomcat"
```

Finally, if you want to access to the manager page of Tomcat and the Banana page, you have to copy/paste the webapps into the webapps folder of your Eclipse workspace :

```
cp -R DATAFARI_SRC/tomcat/webapps/ROOT
your-eclipse-workspace/.metadata/.plugins/org.eclipse.wst.server.core/tmp0/webapps/
```

```
cp -R DATAFARI_SRC/tomcat/webapps/manager
your-eclipse-workspace/.metadata/.plugins/org.eclipse.wst.server.core/tmp0/webapps/
```

cp -R DATAFARI\_SRC/tomcat/webapps/banana.war your-eclipse-workspace/.metadata/.plugins/org.eclipse.wst.server.core/tmp0/webapps/

In Java EE perspective, select File -> Import and select Web -> War File :

Import Import	- + ×
Select Import an external WAR file into a Web Project	RN I
Select an import source:	
type filter text	$\boxtimes$
<ul> <li>&gt; Aven</li> <li>&gt; Plug-in Development</li> </ul>	
<ul> <li>&gt; Remote Systems</li> </ul>	
P ≧→ Run/Debug	
• 🗁 SVN	
P → Tasks	
> 🗁 Team	
* 🗁 Web 🗟 WAR file	- 11
Web services	
? < Back Next > Cancel	Finish

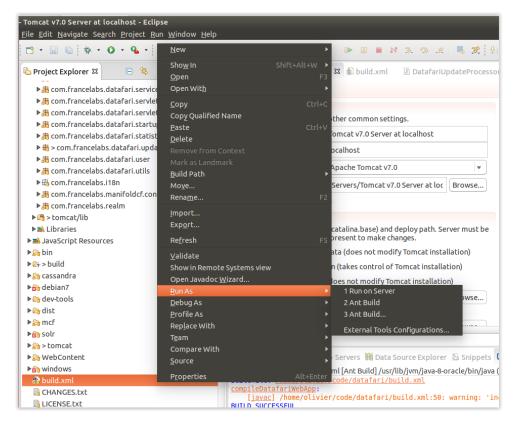
Browse to DATAFARI\_SRC/mcf/war/mcf-api-service.war and set name of web project with the name datafari-mcf-api-service, for DATAFARI\_SRC/mcf/war/mcf-crawler-ui.war (web project name : datafari-mcf-crawler-ui) and for DATAFARI\_SRC/mcf/war/mcf-authority-service.war (web project name : datafari-mcf-authority-service).

Go back to the Servers window. Right click on Datafari, and do an Add and Remove... Add all the web app applications (Datafari, datafari-mcf-crawler-ui, datafari-mcf-api-service, datafari-mcf-authority-service).

	😕 🛷 • 🛛 🕥 🖧 📮 🔌	9. 6 5 54 8 11	🔍 🗷 🕴 - 🕅	• <b>*</b> > <b>/</b> > • •> •	
ि Project Explorer 🛛 📄 🔄 🐨 🗖 🗖	🖥 Tomcat v7.0 Server at local	😣 🗉 Add and Remove			🛙 Mail.java 🗖
▼∰ > Datafari [datafari master]	Overview	Add and Remove			
<ul> <li>Deployment Descriptor: Datafari</li> <li>JAX-WS Web Services</li> </ul>	General Information Specify the host name and o	Modify the resources that are cor	nfigured on the serve	r 📃	
▼證 Java Resources ▼	Server name:	Move resources to the right to cor Available:			
▼	Host name:			Configured:	
AlertFrequencyFirstExecution.java	Runtime Environment:		Add >	datafari-mcf-api-service	Number
▶ 🛺 AlertsManager.java ▶ 🖓 Mail.java	Configuration path: /		< Remove	🔓 datafari-mcf-authority-service 🔓 datafari-mcf-crawler-ui	105
▶ 🛛 StartAlertsListener.java	Open launch configuration				180
▶ 冊 com.francelabs.datafari.constants ▶ 册 com.francelabs.datafari.service.db ▶ 册 com.francelabs.datafari.service.ldap	<ul> <li>Server Locations</li> <li>Specify the server path (i.e. published with no modules published with no modul</li></ul>	5	Add All >>		
▶過 com.francelabs.datafari.service.search ▶過 com.francelabs.datafari.servlets ▶過 com.francelabs.datafari.servlets.admin	<ul> <li>Use workspace metada</li> <li>Use Tomcat installation</li> </ul>		<< Remove All		
<ul> <li>▶ ∰ com.francelabs.datafari.startup</li> <li>▶ ∰ com.francelabs.datafari.statistics</li> </ul>	<ul> <li>Use custom location (c</li> <li>Server path: .metadata</li> </ul>	🕑 If server is started, publish cha	anges immediately		
▶ 曲 > com.francelabs.datafari.updateprocessor ▶ 過 com.francelabs.datafari.user ▶ 過 com.francelabs.datafari.utils	Set deploy path to the def				-
▶⊕ com.francelabs.i18n ▶⊕ com.francelabs.manifoldcf.configuration.api	Overview Modules	? < Back	Next >	Cancel Finish	-
▶∰ com.francelabs.realm ▶∰ > tomcat/lib	· · ·	lhost [Stopped, Synchronized]			
▶ ➡ Libraries ▶ ➡ JavaScript Resources					

Open either Package explorer or the Navigator view: Go to Servers Datafari-config directory. Change the Datafari admin password in the tomcat-users.xml file (look for @PASSWORD@). Copy DATAFARI\_SRC/tomcat/conf/datafari.properties in the Datafari-config directory.

Run Ant on DATAFARI\_SRC/build.xml in order to build the required libraries (Right click on build.xml of datafari web project, then select Run As -> Ant build) :



## **Run Solr and Cassandra**

Start Datafari: Open a terminal and navigate to DATAFARI\_SRC/dev-tools/script/debian.

You need to be root to run the command.

bash start-datafari-devmode.sh

After that, you will need to initialize Cassandra. To do it, go to DATAFARI\_SRC/dev-tools/script/debian.

You also need to be root to run the command.

bash configure-cassandra-dev.sh

The command ends without any feedback.

The Cassandra keyspace and the tables are created. A new user 'admin' is created too. The credentials to login to Datafari admin are : admin/admin

Check if the folder DATAFARI\_SRC/pid is created and if the Solr and Cassandra PID files are present. If not create it manually and affect enough rights for the folder. For example :

mkdir DATAFARI\_SRC/pid

sudo chmod -R 777 DATAFARI\_SRC/pid

#### Run database and ManifoldCF agents

[Optional] Set JAVA\_HOME environment variable pointing to your installation of Java.

Open a terminal and navigate to DATAFARI\_SRC/mcf/mcf\_home.

(you may need to change the permissions on the MCF folder if you are not root)

#### Run

bash start-database-dev.sh.

In another terminal, run (ONLY ONCE, the first time you set-up Datafari dev environment)

bash initialize-dev.sh

#### and then

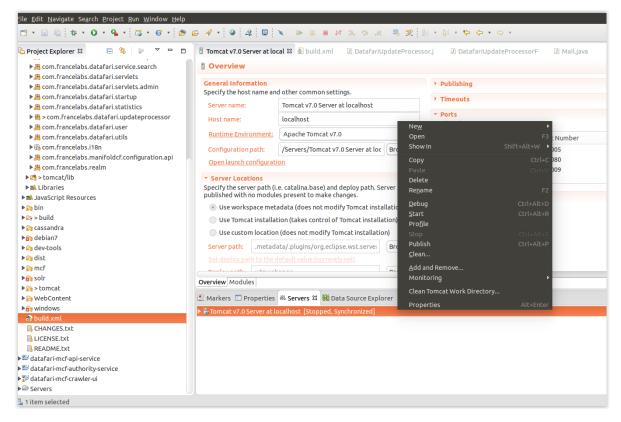
bash start-agents-dev.sh.

#### Keep the two terminals open.

#### **Run Tomcat**

Go to Eclipse and start the Tomcat server (from Servers view)

🕕 If you get FileNotFoundExceptions on the Datafari and/or Solr log files, please run Eclipse as root.



Go to the Datafari URL (http://localhost:8080/Datafari)

To login into Datafari, the credentials are : admin/admin

#### **Configure connectors**

Run Datafari tomcat. Go to the Datafari URL (http://localhost:8080/Datafari). Login in administrator (admin/admin).

In the left menu, click on Connectors; the Apache ManifoldCF login page shows up. Enter admin/admin to login (click on Login button, as the Enter key may not work).

Click on « List Output Connections » of Output section, and add a new « Output Connection ».

In Name tab, set the name to « DatafariSolr ».

In Type tab, set the connection type to Solr and then click on Continue button.

In the Server tab, set port to « 8983 », set web application name to « solr » and « Core/Collection name » to FileShare.

You can now save:

Datafari													
🛊 Mon compte													
X Connecteurs	[												
Lul Statistiques		Docu	imei	nt Inges	tion								
Administration du moteur de recherche	Apache												
Configuration du moteur de recherche	ManifoldCF	ΓM											
Serveurs													
🔮 Gestion d'utilisateurs	Outputs	Name	Туре	Throttling	Solr type	Server	ZooKeeper	Paths	Schema	Arguments	Documents	Commits	
	List Transformation Connections	-											
	List Output Connections	Protocol:			http 🔻								
	Authorities	Server na	mo:										
	List Authority Groups	Serverna	ne.		localhost								
	List User Mapping Connections	Port			8983								
	List Authority Connections				N								
	Repositories	Web appl	cation	name:	solr								
	List Repository Connections	Core/Coll	ection r	name:	FileShare								
	Jobs	Connection timeout (seconds):			is): 60								
	List all Jobs	-			1800								
	Status and Job Management	Socket tim	neout (s	seconds):									
	Status Reports	Realm:											
	Document Status												
	Queue Status	User ID:											
History Reports		Password											
	Simple History Maximum Activity								_				
			certifica	ate list:	No certific								
	Maximum Bandwidth				Add Ce	rtificate: 0	Choisissez un	fichier A	Aucun fichie	r choisi			
	Result Histogram												
	Miscellaneous	Save	Ca	ancel									
	Help												

Your development environment is now ready.

For more information about ManifoldCF configuration, please have a look here: Crawling

# Setup development environment with Eclipse (Linux) for Datafari 2.x - Maven

#### **Pre-requisites**

- Have a JDK installed on your PC (suggested version Java 7)
- Have an IDE installed on your PC (in the guide we use Eclipse)
- Have git installed on your PC
- Have Maven installed on your PC (at least version 3.0.5)
- Have Ant installed on your PC
- Have an Internet connection available (No offline installation permitted)

Undering what are the possible Datafari build flows ? Have a look here Distributions HOW-TO

## **Clone repository**

Open a terminal and navigate to the folder where you want to have Datafari source code to be checkouted (usually your workspace folder).

#### Perform a

git clone https://github.com/francelabs/datafari.git -b maven datafari

to checkout the code. The root directory name must be "datafari": we now call it DATAFARI\_SRC.

## Open the project in Eclipse

Start Eclipse by calling the executable from the workspace folder (the one directly containing DATAFARI\_SRC).

In eclipse, go to File -> Import... , expand Git and do a « Projects from Git ».

😣 🗊 Import	
Select	
Import one or more projects from a Git Repository.	
Select an import source: type filter text © General © EJB © Git © Projects from Git © Install © Java EE © Maven © Oomph	
<ul> <li>Plug-in Development</li> <li>Remote Systems</li> </ul>	
Dun/Debug	
? < Back Next > Cancel	Finish

Select "Existing local repository" and add the repo we have just cloned (by means of browse pop-up).

🕲 🗉 Import Projects from Git	
Select a Git Repository	GIT
😣 No repositories found, please clone or add a repository	
type filter text	⊠ <u>A</u> dd
Add Git Repositories	
Search and select Git repositories on your local file system	
Search for local Git repositories on the file system	GIT
Search of total of repositories of the hysteric	
Search criteria	
Directory: /home/gusai/mvn-git-workspace/datafari	rowse Search
Look for nested repositories	
Search results	
type filter text	
I/home/gusai/mvn-git-workspace/datafari/.git	
0	
? Cancel	Finish

Use "Import as general project" as Wizard. (Don't worry, the classpath and Eclipse settings will be set-up by Ant script).

😢 🗉 Import Projects from Git	
Select a wizard to use for importing projects	
Depending on the wizard, you may select a directory to determine the wizard's scope	
Wizard for project import Import <u>e</u> xisting Eclipse projects	
Import using the New <u>P</u> roject wizard	
O Import as general project	
<ul> <li>Working Directory - /home/gusai/workspace/datafari</li> <li>.git</li> <li>.settings</li> <li>.settings</li> <li>WebContent</li> <li>&gt; bin</li> <li>&gt; cassandra</li> <li>&gt; datafari-cassandra</li> <li>&gt; datafari-core</li> <li>&gt; datafari-dependencies</li> </ul>	
? < Back Next > Cancel Finish	)

Select the DATAFARI\_SRC folder as root directory and click Finish.

😣 💷 Import Projects from Git	
Import Projects	GIT
Import projects from a Git repository	
<u>P</u> rojects:	
type filter text to filter unselected projects	<u>S</u> elect All
🗹 🗁 datafari (/home/gusai/mvn-git-workspace/datafari)	Deselect All
Searc <u>h</u> for nested projects	
Working sets	
Add projec <u>t</u> to working sets	
Working sets:	S <u>e</u> lect
? < Back Next > Cancel	Finish

Leave Eclipse build the project automatically.

You should get this output, in project explorer view:

# 🛛 📅 datafari 2.0-SNAPSHOT [datafari maven]

- AX-WS Web Services
- 🕨 🛅 Deployment Descriptor: datafari
- 🔻 🐉 Java Resources
  - ▶ ∰ src/main/java
  - Libraries
- JavaScript Resources
  - 🗟 Deployed Resources
- bin
- 🕨 🎧 datafari-cassandra
- datafari-core
- datafari-dependencies
- 🕨 🗁 datafari-elk
- datafari-manifoldcf-scripts
- 🕨 🎧 datafari-mcf
- 🕨 🗁 datafari-realm
- datafari-searchcomp-capsule
- 🕨 🚮 datafari-solr
- 🕨 📴 datafari-tomcat
- datafari-updateprocessor
- debian7
- dev-tools
- macosx
- 🕨 🚮 SLC
- tomcat
- WebContent
- windows
  - 🔒 CHANGES.txt
  - 🚮 datafari-dev.xml
  - LICENSE.txt
  - LICENSES THIRD-PARTY.txt
  - 💦 maven-ant-tasks.xml
  - \Lambda pom.xml
  - 🔒 README.txt

As you can see, there are some "red crosses". This is not a problem as they are coming from external packages like ManifoldCF and Solr UI.

## **Eclipse configuration**

In order to properly configure the classpath and the facets of the project, navigate to DATAFARI\_SRC folder and run the following script: ant -f ./dev-tools/setup-devenv/eclipse/setup-eclipse.xml

## Check the Eclipse config

The Datafari Git repo contains some Eclipse config as well.

It's better to check that it has been correctly set-up =P:

Right click on datafari project -> Properties

## Java Build path should be the following:

8	Properties for data	afari	
ty	/pe filter text 🛛 🗷	Java Build Path	⟨¬ ¬ ¬ ¬ ▼
►	Resource Builders	Bource Berojects ■Libraries �Order and Export	
	Deployment Asser	Source folders on build path:	
	Git	▶ ﷺ datafari/src/main/java	Add Folder
	Java Build Path		Link Source
	Java Code Style Java Compiler		Edit
•	Java Editor Javadoc Location		Remove
•	JavaScript JSP Fragment		
•	Maven Project Facets Project Reference:		
	Refactoring Histor Run/Debug Setting	<ul> <li>Allow output folders for source folders</li> <li>Default output folder:</li> </ul>	
	Server Service Policies	datafari/dist/WEB-INF/classes	Browse
	Targeted Runtimes		Apply
(	?)	Ca	ncel OK

Deployment assembly:

8	Properties for dat	afari				
ty	pe filter text 🛛 🗷	Web Deployment Asse	embly			
►	Resource Builders	Define packaging structu	re for this Java	EE Web Application proje	ect.	
	Deployment Assemb	Source	De	ploy Path		A <u>d</u> d
	Git	🗀 /dist	۵,	/		Edit
	Java Build Path	🗀 /src/main/java		WEB-INF/classes		
►	Java Code Style					<u>R</u> emove
►	Java Compiler					
►	Java Editor					
	Javadoc Location					
►	JavaScript					
	JSP Fragment					
►	Maven					
	Project Facets					
	Project References					
	Run/Debug Settings					
	Server					
	Service Policies					
	Targeted Runtimes					
	Task Repository					
	Task Tags					
	Validation					
	Web Content Setting					
	Web Page Editor					
	Web Project Setting: WikiText					
•	XDoclet					
-	ADOCIEL					
		Advanced				
					Revert	Apply
(	?)				Cancel	ОК

Java compiler should be as follows (Java 7):

# 🙆 🗉 Properties for datafari

	X	Java Compiler	⇔ - ⇒	• •
►	Resource	Senable project specific settings	kspace Set	tings
	Builders			
	Deployment Assemb	JDK Compliance		
	Git	<b>U</b> se compliance from execution environment 'JavaSE-1.7' on the	Java Build	Path
	Java Build Path	Compiler compliance level:	1.7	-
	Java Code Style Java Compiler	Use default compliance settings		
	Java Editor	Generated .class files compatibility:	1.7	~
	Javadoc Location	deneraced reason neo compacionary.		
►	JavaScript	Source compatibility:	1.7	•
	JSP Fragment	Disallow identifiers called 'assert':	Error	
►	Maven			
	Project Facets	Disallow identifiers called 'enum':	Error	~
	Project References	Classfile Generation		
	Run/Debug Settings	Add variable attributes to generated class files (used by the de	ouaaer)	
	Server	Add line number attributes to generated class files (used by the		
	Service Policies			
	Targeted Runtimes	Add source file name to generated class file (used by the debug	ger)	
►	Task Repository	Preserve unused (never read) local variables		
	Task Tags	✓ Inline finally blocks (larger class files, but improved performance)	e)	
	Validation	Store information about method parameters (usable via reflect		
	Web Content Setting		liony	
	Web Page Editor			
	Web Project Setting: WikiText			
	XDoclet			
	ADOCIEL			
		Restore Defaults	Арр	y
(	?)	Cancel	ОК	

## Project facets:

Dynamic Web Module V 3.0

Java V 1.7

JavaScript V 1.0

8	Properties for dat	əfəri		
ty	pe filter text 🛛 🗷	Project Facets		
►	Resource Builders	<u>C</u> onfiguration: <custom></custom>		▼ Save As Delete
	Deployment Assemt	Project Facet	Version	De <u>t</u> ails <u>R</u> untimes
	Git	🔲 🖹 Application Client module	6.0	Application Client module 6.0
	Java Build Path	Axis2 Web Services		Enables the project to be deployed as a Java
	Java Code Style	CXF 2.x Web Services	1.0	EE Application Client module.
	Java Compiler	🗹 🗔 Dynamic Web Module	3.0	Requires the following facet:
	Java Editor	🗆 🗎 EAR	6.0	Requires the rollowing facet:
	Javadoc Location	🗆 🗔 EJB Module	3.1	<ul> <li>Java 1.6 or newer</li> </ul>
	JavaScript	🗌 📄 EJBDoclet (XDoclet)	1.2.3	Conflicts with the following facets:
	JSP Fragment	🔽 🗾 Java	1.7	•
	Maven	Value and Script	1.0	Application Client module
	Project Facets	JavaServer Faces	2.2	EAR
	Project References	□ 🗎 JAX-RS (REST Web Services)	1.1	🖕 🖾 EJB Module
	Run/Debug Settings	□ ↔ JAXB	2.2	JCA Module
	Server	🗆 🗎 JCA Module	1.0	Static Web Module
	Service Policies		2.1	Utility Module     Web Fragment Module
	Targeted Runtimes	B Static Web Module		
	Task Repository	Dtility Module		
	Task Tags Validation	B Web Fragment Module	5.0	•
	Web Content Setting	🗌 🗎 WebDoclet (XDoclet)	1.2.3	•
	Web Page Editor			
	Web Project Setting:			
	WikiText			
	XDoclet			
				Revert <u>Apply</u>
$\subseteq$				Кеченс Дррну
0	?)			Cancel
	•			

## Run Maven build

Run Maven on DATAFARI\_SRC/pom.xml in order to build the required libraries (Right click on pom.xml of datafari web project, then select Run As -> Maven build) :

Leave empty the goals section, the default one will be run.

	Edi	t (	C	on	Fi	αι	ıra	ti	o	n

Edit configura	tion and launch.
Name: datafa	ri
🖹 Main  🛋	JRE 🧇 Refresh 🦻 Source 🖾 Environment 🔲 <u>C</u> ommon
Base directory	:
/home/gusai/	/mvn-git-workspace/datafari
	Browse Workspace Browse File System Variables
Goals:	
Profiles:	
User settings:	File
	Offline Update Snapshots
	Debug Output Skip Tests Non-recursive
	Resolve Workspace artifacts
	1 Threads
Parameter Na	am Value Add
	Edit
	Revert Apply
?	Close Run

A console shows up with the output of Maven compilation.

At the end of the build process (it may take some time due to the download of external dependencies), you should see this output:

[INFO]
[INFO] Reactor Summary:
[INFO]
[INFO] Datafari Enterprise Search - Dependencies POM SUCCESS [ 0.441 s]
[INFO] Datafari Enterprise Search - Parent POM SUCCESS [ 32.289 s]
[INFO] Datafari Enterprise Search - Core module SUCCESS [ 3.689 s]
[INFO] Datafari Enterprise Search - UpdateProcessor module SUCCESS [ 0.467 s]
[INFO] Datafari Enterprise Search - Capsule Search Component module SUCCESS [ 0.342 s]
[INFO] Datafari Enterprise Search - Realm module SUCCESS [ 0.355 s]
[INFO] Datafari Enterprise Search - ManifoldCF Backup/Restore Scripts module SUCCESS [ 4.674 s]
[INFO]

Now check to have the some jars listed in Maven dependecies. They should point to your Maven's repo directory, eg user\_home/.m2/repository.

🔻 🚰 datafari 2.0-SNAPSHOT [datafari maven]
JAX-WS Web Services
🕨 🛅 Deployment Descriptor: datafari
🔻 🎦 Java Resources
▶ ∰ src/main/java
▼ ■ Libraries
🛋 EAR Libraries
JRE System Library [JavaSE-1.7]
🔻 🛋 Maven Dependencies
Iog4j-1.2.17.jar - /home/gusai/.m2/repository/lo
solr-solrj-5.2.1.jar - /home/gusai/.m2/repository,
commons-io-2.4.jar - /home/gusai/.m2/repositor

httpmime-4.4.1.jar - /home/gusai/.m2/repositor

If you see nothing in there or you have red crosses on some classes under Java Resources / src/main/java, right click on Datafari project, then select Maven entry and Update project.

Eclipse is not refreshing dependencies automatically.

The dependencies should show up under Maven Dependencies (if it's not the case, refresh the project).

Setup Tomcat server in Eclipse

Delease note that we don't rely on Eclipse WTP environment, so the publish, clean, etc ... targets of server view are discouraged. The only target you should run is the "start" of Tomcat.

Add Server view: Window -> Show view -> Servers

and click on the link to create a new server.

Select Apache Tomcat V. 7.0 and put:

host name: localhost

server name: Datafari

😣 🗊 New Server		
Define a New Server Choose the type of server to	create	
	Show additional server adapters Rel	fresh
Select the server type: type filter text		×
<ul> <li>Tomcat v5.0 Server</li> <li>Tomcat v5.5 Server</li> <li>Tomcat v6.0 Server</li> <li>Tomcat v7.0 Server</li> <li>Tomcat v8.0 Server</li> <li>Basic</li> <li>BM</li> <li>OW2</li> <li>ObjectWeb</li> </ul>	Java EE Web projects and server configurations to a local	
Server's host name:	localhost	
Server name:	Datafari	
?	< Back Next > Cancel Finis	h

Select DATAFARI\_SRC/tomcat for the Tomcat Installation directory and then next

😣 🗊 New Server	
Tomcat Server	
Specify the installation directory	
Na <u>m</u> e:	
Datafari	
Tomcat installation <u>directory</u> :	
/home/gusai/mvn-git-workspace/datafari/tomcat	B <u>r</u> owse
apache-tomcat-	7.0.47 Download and Install
JRE:	
Workbench default JRE	▼ Installed JREs
? < Back Next >	Cancel Finish

DO NOT add Datafari web application to the server and then Finish.

If Datafari-config project pops-out, you are safe to delete it (generated by Eclipse WTP).

Double click on Datafari server to change its configuration :

In Server Locations section, select "Use custom location" (to use the Tomcat embedded in Datafari).

Check that all the other configs are the same as in screenshot. You shouldn't see the Modules tab on the right of overview, or it should be empty (no deployed apps).

Notably, be careful that the paths are exactly as follows: datafari/tomcat/config and datafari/tomcat (no leading /)

Overview													
General Information							▼ Publishing						
Specify the host nar	Modify settings for												
Server name:	Datafari			O Never publish a	automaticall	У							
Host name:	localhost		O Automatically			nge							
Runtime Environm	nent: Apache To	omcat v7.0		O Automatically									
Configuration pat	:h: datafari/t	omcat/conf				Browse	Publishing inter		ds):				
<u>Open launch confi</u>	<u>iguration</u>						Select publishing a						
make changes.	oath (i.e. catalina.l	ase) and deploy p	ath. Server must be pu t installation)	ıblished with ı	no module:	s present to	Vpdate con	LEAC Patris					
🔘 Use Tomcat in:	stallation (takes o	ontrol of Tomcat i	nstallation)				Timeouts						
🗿 Use custom lo	cation (does not r	nodify Tomcat inst	allation)										
Server path: da	atafari/tomcat					Browse							
Set deploy path to	o the default value												
Deploy path: w	rebapps					Browse							
• Server Options Enter settings for th													
Serve modules	s without publishi	ng											
📃 Publish modul	le contexts to sep	arate XML files											
🗹 Modules auto	reload by default												
🗌 Enable securit	<sup>L</sup> y												
Enable Tomcal	t debug logging (n	ot supported by th	is Tomcat version)										
Overview													
🕄 Markers 🔲 Prope	erties 🥵 Servers	🛛 🕄 Problems	🕅 Data Source Exp	Progress	🔗 Search	🛱 Package E	xplor 🔒 History 📥 G	it Staging	Git Repositor	ies 🏞 Call	Hierarchy	Console	
											E 🎋	0 🖉 🔳	1
🖡 Datafari [Started	d, Synchronized]												ĺ,

## In General Information section, click on « Open launch configuration ».

## In Arguments tab, add this to VM arguments :

-Dorg.apache.manifoldcf.configfile="DATAFARI\_SRC/mcf/mcf\_home/propertiesDev-tomcat.xml" -Dsolr.solr.home="DATAFARI\_SRC/solr\_home/"

😣 🗊 🛛 Edit Confi	guration
Edit launch con	figuration properties
Name: Datafari	
🗄 Server 🕅 Ar	guments 🔸 Classpath 🦆 Source 🖾 Environment 🗉 <u>C</u> ommon
Program argum	ents:
datafari/tomca home/gusai/m workspace/dat	Variables =="/home/gusai/mvn-git-workspace/datafari/tomcat" -Dcatalina.home="/home/gusai/mvn-git-workspace/ at" -Dwtp.deploy="/home/gusai/mvn-git-workspace/datafari/tomcat/wtpwebapps" -Djava.endorsed.dirs="/ vn-git-workspace/datafari/tomcat/endorsed" -Dorg.apache.manifoldcf.configfile="/home/gusai/mvn-git- cafari/mcf/mcf_home/propertiesDev-tomcat.xml" ue="/home/gusai/mvn-git-workspace/datafari/solr/solr_home/"]
	Variables
Working direct	ory:
O Default:	/home/gusai/mvn-git-workspace
Other:	Workspace File System Variables
	Vanabies.
	Revert Apply
?	Cancel

and then OK.

Modules tab should be empty or not visible at all.

# Run ZK, Solr and Cassandra

Start Datafari: Open a terminal and navigate to DATAFARI\_SRC/dev-tools/script/debian.

bash start-datafari-devmode.sh

After that, you will need to initialize Cassandra. To do it, go to DATAFARI\_SRC/dev-tools/script/debian.

bash configure-cassandra-dev.sh

The command ends without any feedback.

The Cassandra keyspace and the tables are created. A new user 'admin' is created too. The credentials to login to Datafari admin are : admin/admin

Finally we have to initialize Solr : send the Solr configuration to Zookeeper and create the Solr collections. Launch the script configure-zk-solr-dev.sh in DATAFARI\_SRC/dev-tools/script/debian

## Run database and ManifoldCF agents

Open a terminal and navigate to DATAFARI\_SRC/mcf/mcf\_home.

Run

bash start-database-dev.sh.

In another terminal, run (ONLY ONCE, the first time you set-up Datafari dev environment)

bash initialize-dev.sh

and then

bash start-agents-dev.sh.

Keep the two terminals open.

**Run Tomcat** 

Go to Eclipse -> Servers view.

Start it (right click on the server name -> start). Datafari application is already deployed by Maven/Ant.

UPlease note that we don't rely on Eclipse WTP environment, so the publish, clean, etc ... targets of server view are discouraged. The only target you should run are the "start / stop" of Tomcat.

Note: if you get this exception

INFO: Exception performing authentication. Retrying... javax.naming.CommunicationException: 52.16.74.128:389 [Root exception is java.net.SocketTimeoutException: connect timed out]

it's the, let's say "normal" behaviour =P! (Fix on-going)

Go to the Datafari URL (http://localhost:8080/Datafari)

To login into Datafari, the credentials are : admin/admin

HAPPY SEARCHING 📛 !!!

Need to get Datafari ready for production? Have a look here Distributions HOW-TO

Datafari incremental build and redeployment

For redeployment of Datafari Java sources and web resources ONLY (e.g. during an on-going dev), you can use the handy Ant targets "all" or "redeploy" of datafari-dev.xml.

This script recompiles the Datafari core sources and deploys the application on Tomcat.

🕕 If you get strange errors while running Ant, double check to run it on a JDK and not on a JRE (Oracle one, for instance).

## **Configure connectors**

Run Datafari tomcat. Go to the Datafari URL (http://localhost:8080/Datafari). Login in administrator (admin/admin).

In the left menu, click on Connectors; the Apache ManifoldCF login page shows up. Enter admin/admin to login (click on Login button, as the Enter key may not work).

Click on « List Output Connections » of Output section, and add a new « Output Connection ». In Name tab, set the name to « DatafariSolr ».

In Type tab, set the connection type to Solr and then click on Continue button.

In the Server tab, set port to « 8983 », set web application name to « solr » and « Core/Collection name » to FileShare.

You can now save:

Datafari													
🛉 Mon compte													
X Connecteurs	r												
Lul Statistiques		Doci	umen	t Inges	tion								
Administration du moteur de recherche	Apache												
Configuration du moteur de recherche	ManifoldCF <sup>®</sup>	И											
Serveurs													
Gestion d'utilisateurs	Outputs	Name	Туре	Throttling	Solr type	Server	ZooKeeper	Paths	Schema	Arguments	Documents	Commits	
	List Transformation Connections												
	List Output Connections	Protocol:											
	Authorities				http 🔻								
	List Authority Groups	Server na	ame:		loc alhost								
	List User Mapping Connections	Port:			8983								
	List Authority Connections				000								
	Repositories	Web application name:		ame:	solr								
	List Repository Connections			FileShare									
	Jobs	Connecti	on timeou	ut (seconds)	60								
	List all Jobs				ds): 60								
	Status and Job Management	Socket tin	meout (se	conds):	1800								
	Status Reports	Realm:											
	Document Status	Troum.											
	Queue Status	User ID:											
	History Reports	Password:											
	Simple History												
	Maximum Activity	SSL trust	certificat	e list:	No certificates present								
	Maximum Bandwidth Add Certificate: Choisissez un fichier Aucun fichier choisi												
	Result Histogram												
	Miscellaneous	Save	Car	ncel									
	Help												

Your development environment is now ready.

U For more information about ManifoldCF configuration, please have a look here: Crawling

# **Unit Testing**

Datafari 3.x comes with some Unit Tests (at last !)

Three frameworks for Unit Testing are embedded in this version of Datafari :

- JUnit : the standard JAVA Unit Testing framework
- · Mockito : famous framework that allow to mock JAVA objects
- XMLUnit : very useful framework for XML comparison

We think that these three frameworks cover more than 90% of the needs when creating a Unit Tests, so now you have no good excuses to create Unit Tests when it is possible for your development.

Now, let's try to explain our usages and good practices of the Unit Tests in Datafari:

- The first thing to know is that all the existing and future Unit Tests have to be located in Datafari\_Home/src/test. As you will notice, there are two sections in this directory:

- · main/java : That contains all the JUnit JAVA classes
- resources : That contains the resource files needed by the Unit Tests. Each Test which is using one or several resource files, must have his own repertory in this section. It is then easier to access to the resource files corresponding to a specific Test

To be able to run the tests from the eclipse environment, you will need to add a Variable to the classpath of Datafari :

Right click on the project Build Path Configure Build Path

8	Properties for datafa	ri	
	×	Java Build Path 🗘 🔻 🖒 🔻 😽	,
•	Resource Builders Deployment Assembly	Bource Brojects ■Libraries Arrow Order and Export JARs and class folders on the build path:	
	Git	▶ ● DATAFARI_CLASSPATH - /data/datafari-maven/datafa Add JARs	)
	Java Build Path	► ► EAR Libraries	í.
►	Java Code Style	▶ ➡ JRE System Library [JavaSE-1.7] Add External JARs	J
►	Java Compiler	▶ ➡ Maven Dependencies Add Variable	
►	Java Editor	▶ ➡ Web App Libraries Add Library	1
	Javadoc Location		2
	JavaScript	Add Class Folder	
	JSP Fragment	Add External Class Folder	1
	Maven	Add External class Folder	J
	Project Facets	Edit	
	Project References		
	Refactoring History	Remove	
	Run/Debug Settings		
	Server	Migrate JAR File	
	Service Policies		
	Targeted Runtimes		
	Task Repository	Apply	
C	?)	Cancel OK	)

Now go to the libraries tab and click on "Add Variable", then "Configure Variables" and click on "New". Name it "DATAFARI\_CLASSPATH" and for the Path browse to the Datafari\_Home/dist/WEB-INF/classes.Then click "OK" till you return to the Libraries tab and click "Apply" then "OK"

- The second point is that the ant script "datafari-dev.xml" which is used to compile and deploy the web part of Datafari (inluding the servlets) to the tomcat server, is also configured to automatically run all the JUnit tests found in the Datafari\_Home/src/test/main/java directory. So when you are done with a new development or a modification of Datafari, check the output of this ant script to see if all the tests have passed with success.

- The third and last thing to understand is the kind of Unit Tests we are expecting in Datafari. The usage of Unit Testing is limited in Datafari, because the product has many interactions with tier applications (ELK, Solr, ManifoldCF, Cassandra) and that it makes no sense to try to mock those components as they represents the core of Datafari. So, based on this observation, we have focused our Unit Tests on another important part : the manipulation of the configuration files. It applies for now to the config files of Datafari, Solr and ELK. Our tests generate calls to the servlets in charge of the modifications/access to the config files and compare their behavior and their outputs to the expected ones. In that way we insure well formed configuration files and presence of every expected parameters and values.

So now that you know what is our goal with the Unit Tests and that Datafari provides you the tools to make your owns, we hope and encourage you to enrich Datafari with a lot of Unit Tests.

# **Utility logs**

The 2.2 version of Datafari introduces utility logs, which are distributed in two types:

- the Statistics logs
- the Core Monitoring logs

## **Core Monitoring logs**

Datafari periodically generates monitoring logs about the main Solr core. Those logs are intended to provide general informations on the core contents (like number of indexed doc by types, by source, etc.).

By default, the generation rate is once per hour, and can be currently modified through the java class "com.francelabs.datafari.monitoring.IndexMonitoring". Here is how monitoring logs look like:

monitoring logs
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:175 - ec6d7ff440530d73e020766403a3e058 2015-11-06T18:14:00.000+0100 40 no no 2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 -
dd3fb399e47dced897b708d9c6e78d2f 2015-11-06T18:14:00.000+0100 11 docx extension
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - 039435d1512eb763234a2690fb06c63c 2015-11-06T18:14:00.000+0100 6 pdf extens ion
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - 611555203c75e7880c8dc67afd41062d 2015-11-06T18:14:00.000+0100 4 txt extens ion
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - 2ce23ed298b044a3eda0a0e695921024 2015-11-06T18:14:00.000+0100 3 png extens ion
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - 0fc7bbcad0efcfc2037ae2a2f1fbfe04 2015-11-06T18:14:00.000+0100 2 doc extens ion
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - 31f8884b788c7872e5b2a286faa42747 2015-11-06T18:14:00.000+0100 2 html exten sion
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - e155fbb2f018035d65553e763e0d8e63 2015-11-06T18:14:00.000+0100 1 gif extens ion
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - c9dc854a334e1dc339d6cdfd10cb37cf 2015-11-06T18:14:00.000+0100 1 jar extens ion
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - c10e628086bdc9fe4126c5f9c3b222c6 2015-11-06T18:14:00.000+0100 1 jpg extens ion
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - 24ca7afe128eb7b33b36c4919e830509 2015-11-06T18:14:00.000+0100 16 en langua ge
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - 8879abd5830038c4a630d07526857faf 2015-11-06T18:14:00.000+0100 15 fr langua ge
2015-11-06 18:14:00 MONITORING IndexMonitoring\$FileShareMonitoringLog:184 - dc8834aa358f02cff45e71052dd648c0 2015-11-06T18:14:00.000+0100 30 file sour ce

They respect a specific format which is: [log\_id] | [timestamp] | [num\_doc] | [facet\_value] | [facet\_field]

Let's explain each field:

- [log\_id] : the log id. This id IS UNIQUE per time event, per facet value and field, and is generated from these 3 values. The reason behind this is the ELK layout.
- [timestamp] : the timestamp of the log. It is formatted and could be different from the log4j timestamp so it is mandatory
- [num\_doc] : the number of documents found according to the facet
- [facet\_value] : the facet value
- [facet\_field] : the facet field used, in relation with the facet value

At each monitoring log generation iteration, a log line is created for each facet value of the selected facet fields, plus a log line for the global number of documents indexed (in that case, the [facet\_value] and [facet\_field] are both set to "no"). The default selected facet fields are "extension", "language" and "source". One can change the facet fields selected in the class "com.francelabs.monitoring.IndexMonitoring.FileShareMonitoringLog".

## **Statistics logs**

Datafari generates a statistic log each time a query is performed by a user or he/she clicks on a facet/result/page. Here is how a statistic log looks like:

statistic log
2015-11-05 15:24:45 STAT StatsPusher:95 508e9b3e-fdb0-4831-8fab-2acad81f1cb5|2015-11-05T14:21:29.740+0100|engine|0
|4|1|7|2|1|[engine/////4//7//1////, engine//(extension:docx
)//2//6//11////,
///////file:/home/youp/Downloads/doc/Alertes.docx///2]|file:/home/
youp/Downloads/doc/Alertes.docx

It respects a specific format which is :

[query\_id] | [query\_timestamp] | [query] | [noHits] | [numFound] | [numClicks] | [QTime] | [positionClickTot] | [click] | [history] | [spell] | [suggest] | [url]

Let's explain each field:

• [query\_id] : the id of the query. This id is used to keep track of the user behaviour.

A query id is generated each time the user clicks on the search button, or presses the "Enter" key when the focus is on the search field of Datafari. However, when the user performs a "sub-query" like selecting a facet or a page or even clicking on a result, the "sub-query" keeps the id of the root query.

For example, if the user searches for the word "engine", a query id is generated. Then, if the user clicks on the facet "fr" on the query results, the facet "sub-query" will keep the id of the root "engine" query. So several log lines can concern one query id as one log is generated on each action performed

- [query\_timestamp] : the full timestamp of the query, measured by Datafari
- [query] : the literal query performed by the user
- [noHits] : indicates if the query does or does not have hits. Two values are possible, '0' if the query has hits, '1' if the query does not have any hit
- [numFound] : the number of documents found for the query
- [numClicks] : the number of clicks on the found documents
- [QTime] : the query time in milliseconds. This value is directly provided by Solr
- [positionClickTot]: represents the sum of the positions of the documents clicked.
   For example, if the user clicks on the first result and the third, the positionClickTot will be 1+3=4
   The positions are absolute and are not based on the page. If there are 10 results by page and the user clicks on the first link of the second page, the position number will be 11
- [click] : like [noHits], it is a boolean value where '0' indicates that the user did not click on any result, and '1' indicates that the user has clicked on at least on one result.
- [history] : keeps track of the user behaviour. It represents a list of user "actions" : [user\_action1,user\_action2,...]
  - A user action is formatted as follows: "query"///"facet\_used"///"num\_doc\_found"///"query\_time"///"num\_of\_page"///"url"///"url\_position" • "query" : the query performed
    - "facet\_used" : filled if the user has clicked on a facet, it is formatted like this : ("facet\_field":"facet\_value") For example, if you look at the second "user action" in the history of the example log : (extension:docx), you can deduce that the

user has clicked on the facet "docx" which is based on the field named "extension")

- "num\_doc\_found" : number of docs returned by the query
- "query\_time" : the query time measured by Solr
- "num\_of\_page" : the number of the page where was the user
- "url" : the URL of the clicked document
- "url\_position" : the position of the clicked document (first result = 1, second = 2, etc.). Remember that the position is absolute and based on all the available results.
- When the user clicks on a document, only two values of the history are set : the "url" and the "url\_position"
- [url] : the clicked document URL. This value is only set if the log itself represents the click

Now that the basics are settled, let's decode the example log. Here is what it says :

- The log has been generated on a click of the user on a document (the [url] field is not empty)
- The original query was "engine", which returned 4 results and took 7 milliseconds
- The user has clicked on one document of which the URL is "file:/home/youp/Downloads/doc/Alertes.docx"
- Relying on the history, we can say that :
  - the original query was "engine", returned 4 results and took 7 milliseconds
  - then the user clicked on the facet "docx" which is based on the field named "extension". This "sub-query" returned 2 results and took 6 milliseconds
  - the last action (which is the log subject) is a click on the 2nd result of which the URL is "file:/home/youp/Downloads/doc/Alertes.docx"

By default, those logs are not displayed in the console, but are written into specifics log files. The configuration of the log files (path, size, number etc.) can be set in the log4j properties located in tomcat/lib/log4j.properties.

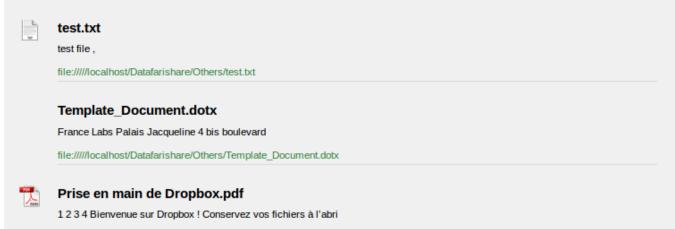
# Widgets and CSSs good practices

Datafari allows you to create your own widgets an CSSs and add them to enrich and customize the UI. The cost of such power is of course the possibility to break everything in the UI. So by reading carefully and following the few steps presented here, you will ensure an easy development, more compliant code for next Datafari updates, and fast rollback in case of problems:

- Modify an existing AjaxFranceLabs widget for enhancement or bugfix On one hand, if you find an improvement or a bug fix on an existing widget, our recommandation is for you to apply your modifications directly on it and create a ticket on our Jira. On another hand, If you want to change the main behavior or main features of an existing widget, we recommand that you make your own and replace the original widget by yours in the searchView.jsp and search.js
- Modify an existing CSS for enhancement or bugfix Same thing as the widgets: On one hand, if you find an improvement or a bug fix on an existing CSS, you can apply your modifications directly on it and create a ticket on our Jira. On another hand, if you need to change an existing CSS of Datafari, we recommand that you make your own copy and include it in the proper file. A custom css already exists for the searchView.jsp, it is available in {Datafari\_Home}/tomcat/webapps/Datafari/css/custom/customSearchView.css
- Create your own CSS per widget To make things clean and avoid mess, it is strongly recommended to have at least one separate CSS file per widget you made. Furthermore, there is a dedicated location for widget CSSs which is {Datafari\_Home}/tomcat/webapps/Datafari/css/widgets
- Make one file per widget If you develop multiple widgets, be sure to have each of them in a separate file. It is easier to maintain and to deactivate in case of problems.

## Add new icon for search results

It may happen that you have indexed documents with a specific extension which is not known by Datafari. If this is the case, the concerned documents will be displayed in the result view without any icon (here Template\_Document.dotx):



file:////localhost/Datafarishare/Others/Prise en main de Dropbox.pdf

To correct/avoid this, you will have to add a new icon for your concerned extension in 'DATAFARI\_HOME\_DIR/tomcat/webapps/Datafari/images/icons'. The file must be a png type and respect the following pattern : [extension\_nam e]-icon-24x24.png

For example, the icon file to add to the icons folder in order to correct the problem showed above would be : dotx-icon-24x24.png

Of course you can make a copy of an existing icon from the icons directory and rename it.

Once the icon is available, it will be automatically used by Datafari.

## Setup development environment with Eclipse (Linux) for Datafari 3.x - Maven

#### **Pre-requisites**

- Have a JDK installed on your PC (suggested version Java 7)
- Have an IDE installed on your PC (in the guide we use Eclipse)
- Have git installed on your PC
- Have Maven installed on your PC (at least version 3.0.5)
- Have Ant installed on your PC
- · Have an Internet connection available (No offline installation permitted)

Wondering what are the possible Datafari build flows ? Have a look here Distributions HOW-TO

## **Clone repository**

Open a terminal and navigate to the folder where you want to have Datafari source code to be checkouted (usually your workspace folder).

#### Perform a

git clone https://github.com/francelabs/datafari.git -b master datafari

to checkout the code. The root directory name must be "datafari": we now call it DATAFARI\_SRC.

## Open the project in Eclipse

Start Eclipse by calling the executable from the workspace folder (the one directly containing DATAFARI\_SRC).

In eclipse, go to File -> Import... , expand Git and do a « Projects from Git ».

😣 💷 Import	
Select	
Import one or more projects from a Git Repository.	
Select an import source:	
type filter text	×
<ul> <li>▷ General</li> <li>▷ EJB</li> <li>□ Cito</li> </ul>	
▼	
<ul> <li>► Install</li> <li>► Java EE</li> </ul>	
<ul> <li>▶ Aven</li> <li>▶ Oomph</li> </ul>	
<ul> <li>Plug-in Development</li> <li>Remote Systems</li> </ul>	
▶ 🎘 Dun/Debug	
? < Back Next > Cancel	Finish

Select "Existing local repository" and add the repo we have just cloned (by means of browse pop-up).

🕲 🗉 Import Projects from Git	
Select a Git Repository	GIT
😣 No repositories found, please clone or add a repository	
type filter text	⊠ <u>A</u> dd
Add Git Repositories	
Search and select Git repositories on your local file system	
Search for local Git repositories on the file system	GIT
Search of total of repositories of the hysteric	
Search criteria	
Directory: /home/gusai/mvn-git-workspace/datafari	rowse Search
Look for nested repositories	
Search results	
type filter text	
I/home/gusai/mvn-git-workspace/datafari/.git	
0	
? Cancel	Finish

Use "Import as general project" as Wizard. (Don't worry, the classpath and Eclipse settings will be set-up by Ant script).

😣 🗊 Import Projects from Git	
Select a wizard to use for importing projects	GIT
Depending on the wizard, you may select a directory to determine the wizard's scope	
Wizard for project import Import <u>e</u> xisting Eclipse projects Import using the New <u>P</u> roject wizard Import as <u>g</u> eneral project	
<ul> <li>Working Directory - /home/gusai/workspace/datafari</li> <li>git</li> <li>settings</li> <li>webContent</li> <li>webContent</li> <li>bin</li> <li>cassandra</li> <li>datafari-cassandra</li> <li>datafari-core</li> <li>datafari-dependencies</li> </ul>	
? < Back Next > Cancel	Finish

Select the DATAFARI\_SRC folder as root directory and click Finish.

😣 💷 Import Projects from Git	
Import Projects	GIT
Import projects from a Git repository	
<u>P</u> rojects:	
type filter text to filter unselected projects	<u>S</u> elect All
🗹 🗁 datafari (/home/gusai/mvn-git-workspace/datafari)	Deselect All
Searc <u>h</u> for nested projects	
Working sets	
Add projec <u>t</u> to working sets	
Working sets:	S <u>e</u> lect
? < Back Next > Cancel	Finish

Leave Eclipse build the project automatically.

You should get this output, in project explorer view:

# 🛛 📅 datafari 2.0-SNAPSHOT [datafari maven]

- AX-WS Web Services
- 🕨 🛅 Deployment Descriptor: datafari
- 🔻 🐉 Java Resources
  - ▶ ∰ src/main/java
  - Libraries
- JavaScript Resources
  - 🗟 Deployed Resources
- bin
- 🕨 🎧 datafari-cassandra
- datafari-core
- datafari-dependencies
- 🕨 🔄 datafari-elk
- datafari-manifoldcf-scripts
- 🕨 🎧 datafari-mcf
- 🕨 🗁 datafari-realm
- datafari-searchcomp-capsule
- 🕨 🚮 datafari-solr
- 🕨 📴 datafari-tomcat
- datafari-updateprocessor
- debian7
- dev-tools
- macosx
- 🕨 🚮 SLC
- tomcat
- WebContent
- windows
  - 🔒 CHANGES.txt
  - 🚮 datafari-dev.xml
  - LICENSE.txt
  - LICENSES THIRD-PARTY.txt
  - 💦 maven-ant-tasks.xml
  - \Lambda pom.xml
  - 🔒 README.txt

As you can see, there are some "red crosses". This is not a problem as they are coming from external packages like ManifoldCF and Solr UI.

# **Eclipse configuration**

In order to properly configure the classpath and the facets of the project, navigate to DATAFARI\_SRC folder and run the following script: ant -f ./dev-tools/setup-devenv/eclipse/setup-eclipse.xml

# Check the Eclipse config

The Datafari Git repo contains some Eclipse config as well.

It's better to check that it has been correctly set-up =P:

Right click on datafari project -> Properties

## Java Build path should be the following:

8	Properties for data	afari	
ty	/pe filter text 🛛 🗷	Java Build Path	⟨¬ ¬ ¬ ¬ ▼
►	Resource Builders Deployment Asser	Bource Berojects ■Libraries �Order and Export	
		Source folders on build path:	
	Git	▶ ﷺ datafari/src/main/java	Add Folder
	Java Build Path		Link Source
<ul><li>Java Code Style</li><li>Java Compiler</li></ul>			Edit
•	Java Editor Javadoc Location		Remove
•	JavaScript JSP Fragment		
•	Maven Project Facets Project Reference:		
	Run/Debug Setting Default o	<ul> <li>Allow output folders for source folders</li> <li>Default output folder:</li> </ul>	
	Server Service Policies	datafari/dist/WEB-INF/classes	Browse
	Targeted Runtimes		Apply
(	?)	Ca	ncel OK

Deployment assembly:

8	Properties for dat	afari				
ty	pe filter text 🛛 🗷	Web Deployment Asse	embly			
►	Resource Builders	Define packaging structu	re for this Java	EE Web Application proje	ect.	
	Deployment Assemb	Source	De	ploy Path		A <u>d</u> d
	Git	🗀 /dist	۵,	/		Edit
	Java Build Path	🗀 /src/main/java		WEB-INF/classes		
►	Java Code Style					<u>R</u> emove
►	Java Compiler					
►	Java Editor					
	Javadoc Location					
►	JavaScript					
	JSP Fragment					
►	Maven					
	Project Facets					
	Project References					
	Run/Debug Settings					
	Server					
	Service Policies					
	Targeted Runtimes					
	Task Repository					
	Task Tags					
	Validation					
	Web Content Setting					
	Web Page Editor					
	Web Project Setting: WikiText					
•	XDoclet					
-	ADOCIEL					
		Advanced				
					Revert	Apply
(	?)				Cancel	ОК

Java compiler should be as follows (Java 7):

# 🙆 🗉 Properties for datafari

	X	Java Compiler	⇔ - ⇒	• •
►	Resource	Senable project specific settings	kspace Set	tings
	Builders			
	Deployment Assemb	JDK Compliance		
	Git	✓ Use compliance from execution environment 'JavaSE-1.7' on the	Java Build	Path
	Java Build Path	Compiler compliance level:	1.7	-
	Java Code Style Java Compiler	Use default compliance settings		
	Java Editor	Generated .class files compatibility:	1.7	~
	Javadoc Location	deneraced reason neo compacionary.		
►	JavaScript	Source compatibility:	1.7	•
	JSP Fragment	Disallow identifiers called 'assert':	Error	
►	Maven			
	Project Facets	Disallow identifiers called 'enum':	Error	~
	Project References	Classfile Generation		
	Run/Debug Settings	Add variable attributes to generated class files (used by the de	ouaaer)	
	Server	Add line number attributes to generated class files (used by the		
	Service Policies			
	Targeted Runtimes	Add source file name to generated class file (used by the debugger)		
►	Task Repository	Preserve unused (never read) local variables		
	Task Tags	✓ Inline finally blocks (larger class files, but improved performance)	e)	
	Validation	Characieformation about mathed as sometars (usable via soflastion)		
	Web Content Setting		liony	
	Web Page Editor			
	Web Project Setting: WikiText			
	XDoclet			
	ADOCIEL			
		Restore Defaults	Арр	y
(	?)	Cancel	ОК	

# Project facets:

Dynamic Web Module V 3.0

Java V 1.7

JavaScript V 1.0

😣 🗉 Properties for datafari					
ty	type filter text 🗷 Project Facets 🗘 🗘 🗸				
►	Resource Builders	<u>C</u> onfiguration: <custom></custom>		▼ Save As Delete	
	Deployment Assemb	Project Facet	Version	De <u>t</u> ails <u>R</u> untimes	
	Git	📃 🖹 Application Client module	6.0	Application Client module 6.0	
	Java Build Path	Axis2 Web Services		Enables the project to be deployed as a Java	
	Java Code Style	🗌 🗎 CXF 2.x Web Services	1.0	EE Application Client module.	
	Java Compiler	🗹 🗔 Dynamic Web Module	3.0	Requires the following facet:	
	Java Editor	EAR	6.0		
	Javadoc Location	🗆 🗟 EJB Module	3.1	Java 1.6 or newer	
	JavaScript	EJBDoclet (XDoclet)	1.2.3	Conflicts with the following facets:	
	JSP Fragment	Java	1.7	•	
	Maven	VavaScript	1.0	<ul> <li>Application Client module</li> <li>Dynamic Web Module</li> </ul>	
	Project Facets	JavaServer Faces	2.2	EAR	
	Project References	□ 🖹 JAX-RS (REST Web Services)	1.1	Subscription of the second sec	
	Run/Debug Settings	□ ↔ JAXB	2.2	JCA Module	
	Server Service Policies	C D JCA Module	1.6	Static Web Module	
			2.1	E Utility Module     Web Fragment Module	
	Targeted Runtimes Task Repository	C Static Web Module			
		Dtility Module			
	Task Tags Validation	B Web Fragment Module	3.0	•	
	Web Content Setting	🗌 🗎 WebDoclet (XDoclet)	1.2.3		
	Web Page Editor				
	Web Project Setting:				
	WikiText				
	XDoclet				
				Revert <u>Apply</u>	
				Revert <u>A</u> ppiy	
	?)			Cancel OV	
	1)			Cancel OK	

# Add mcf-core lib to your maven repo

Datafari needs the mcf-core lib as a dependency. The problem is that Apache ManifoldCF does not have a maven repository so the library must be added manually to the maven repository. To do so, open a terminal and change directory to the main directory of Datafari, then execute the following command:

```
mvn install:install-file -Dfile=./datafari-mcf/lib/mcf-core-2.4.jar
-DgroupId=org.apache.manifoldcf -DartifactId=mcf-core -Dversion=2.4
-Dpackaging=jar -DgeneratePom=true
```

## **Run Maven build**

Run Maven on DATAFARI\_SRC/pom.xml in order to build the required libraries (Right click on pom.xml of datafari web project, then select Run As -> Maven build) :

Leave empty the goals section, the default one will be run.

8  Edit Configuration				
Edit configuration and launch.				
Name: datafari				
Base directory:				
/home/gusai/mvn-git-workspace/datafari				
Browse Workspace Browse File System Variables				
Goals:				
Profiles:				
User settings: File				
Offline Update Snapshots				
Debug Output Skip Tests Non-recursive				
Resolve Workspace artifacts				
1 v Threads				
Parameter Nam Value Add				
Edit				
Revert Apply				
Close Run				

A console shows up with the output of Maven compilation.

At the end of the build process (it may take some time due to the download of external dependencies), you should see this output:

Now check to have the some jars listed in Maven dependecies. They should point to your Maven's repo directory, eg user\_home/.m2/repository.

🔻 🚰 datafari 2.0-SNAPSHOT [datafari maven]
JAX-WS Web Services
🕨 🛅 Deployment Descriptor: datafari
🔻 🎦 Java Resources
▶ ∰ src/main/java
▼ ■ Libraries
🛋 EAR Libraries
JRE System Library [JavaSE-1.7]
🔻 🛋 Maven Dependencies
Iog4j-1.2.17.jar - /home/gusai/.m2/repository/lo
solr-solrj-5.2.1.jar - /home/gusai/.m2/repository,
commons-io-2.4.jar - /home/gusai/.m2/repositor

httpmime-4.4.1.jar - /home/gusai/.m2/repositor

If you see nothing in there or you have red crosses on some classes under Java Resources / src/main/java, right click on Datafari project, then select Maven entry and Update project.

Eclipse is not refreshing dependencies automatically.

The dependencies should show up under Maven Dependencies (if it's not the case, refresh the project).

Setup Tomcat server in Eclipse

Delease note that we don't rely on Eclipse WTP environment, so the publish, clean, etc ... targets of server view are discouraged. The only target you should run is the "start" of Tomcat.

Add Server view: Window -> Show view -> Servers

and click on the link to create a new server.

Select Apache Tomcat V. 7.0 and put:

host name: localhost

server name: Datafari

😣 💷 New Server	
<b>Define a New Server</b> Choose the type of server to	create
Select the server type:	Show additional server adapters Refresh
<ul> <li>Tomcat v5.0 Server</li> <li>Tomcat v5.5 Server</li> <li>Tomcat v6.0 Server</li> <li>Tomcat v7.0 Server</li> <li>Tomcat v8.0 Server</li> <li>Basic</li> <li>BBM</li> <li>OW2</li> <li>ObjectWeb</li> </ul>	Java EE Web projects and server configurations to a local
Server's host name: Server name:	localhost Datafari
?	< Back Next > Cancel Finish

Select DATAFARI\_SRC/tomcat for the Tomcat Installation directory and then next

😣 💷 New Server		
Tomcat Server		
Specify the installation directory		
Na <u>m</u> e:		
Datafari		
Tomcat installation <u>directory</u> :		
/home/gusai/mvn-git-workspace/datafari/to	omcat	B <u>r</u> owse
	apache-tomcat-7.0.47	Download and Install
JRE:		
Workbench default JRE	▼	Installed JREs
? < Back	Next > Can	cel Finish

DO NOT add Datafari web application to the server and then Finish.

If Datafari-config project pops-out, you are safe to delete it (generated by Eclipse WTP).

Double click on Datafari server to change its configuration :

In Server Locations section, select "Use custom location" (to use the Tomcat embedded in Datafari).

Check that all the other configs are the same as in screenshot. You shouldn't see the Modules tab on the right of overview, or it should be empty (no deployed apps).

Notably, be careful that the paths are exactly as follows: datafari/tomcat/config and datafari/tomcat (no leading /)

Overview													
General Information							▼ Publishing						
Specify the host nar	me and other com	mon settings.					Modify settings for						
Server name:	Datafari						O Never publish a	automaticall	У				
Host name:	localhost						O Automatically			nge			
Runtime Environm	nent: Apache To	omcat v7.0				-	O Automatically	publish afte	r a build event				
Configuration pat	:h: datafari/t	omcat/conf				Browse	Publishing inter		ds):				
<u>Open launch confi</u>	<u>iguration</u>						Select publishing a						
make changes.	oath (i.e. catalina.l	ase) and deploy p	ath. Server must be pu t installation)	ıblished with ı	no module:	s present to	Vpdate con	LEAC Patris					
🔘 Use Tomcat in:	stallation (takes o	ontrol of Tomcat i	nstallation)				Timeouts						
🗿 Use custom lo	cation (does not r	nodify Tomcat inst	allation)										
Server path: da	atafari/tomcat					Browse							
Set deploy path to	o the default value												
Deploy path: w	rebapps					Browse							
• Server Options Enter settings for th													
Serve modules	s without publishi	ng											
📃 Publish modul	le contexts to sep	arate XML files											
🗹 Modules auto	reload by default												
🗌 Enable securit	<sup>L</sup> y												
Enable Tomcal	t debug logging (n	ot supported by th	is Tomcat version)										
Overview													
🕄 Markers 🔲 Prope	erties 🥵 Servers	🛛 🕄 Problems	🕅 Data Source Exp	Progress	🔗 Search	🛱 Package E	xplor 🔒 History 📥 G	it Staging	Git Repositor	ies 🏞 Call	Hierarchy	Console	;
											5	0 🖉	
🖡 Datafari [Started	d, Synchronized]												

## In General Information section, click on « Open launch configuration ».

## In Arguments tab, add this to VM arguments :

-Dorg.apache.manifoldcf.configfile="DATAFARI\_SRC/mcf/mcf\_home/propertiesDev-tomcat.xml" -Dsolr.solr.home="DATAFARI\_SRC/solr\_home/"

😣 🗉 🛛 Edit Confi	guration
Edit launch con	iguration properties
Name: Datafari	
🗄 Server 🔲 Arg	guments 🚸 Classpath 🦻 Source 🖾 Environment 🔲 <u>C</u> ommon
Program argum	ents:
start	Variables
VM arguments:	
datafari/tomca home/gusai/my workspace/dat	="/home/gusai/mvn-git-workspace/datafari/tomcat" -Dcatalina.home="/home/gusai/mvn-git-workspace/ it" -Dwtp.deploy="/home/gusai/mvn-git-workspace/datafari/tomcat/wtpwebapps" -Djava.endorsed.dirs="/ /n-git-workspace/datafari/tomcat/endorsed" -Dorg.apache.manifoldcf.configfile="/home/gusai/mvn-git- afari/mcf/mcf_home/propertiesDev-tomcat.xml" e="/home/gusai/mvn-git-workspace/datafari/solr/solr_home/"
	Variables
Working directo	pry:
O Default:	/home/gusai/mvn-git-workspace
Other:	
	Workspace File System Variables
	Revert Apply
?	Cancel OK

and then OK.

Modules tab should be empty or not visible at all.

# Run ZK, Solr and Cassandra

Start Datafari: Open a terminal and navigate to DATAFARI\_SRC/dev-tools/script/dev.

• Edit the variable DEV\_OS in dev-datafari.properties. if you are on Debian (let debian 7 for all versions of Debian) or MacOSX :

DEV\_OS=debian7

The two possible variables are : debian7 or macosx.

The property STATE indicates if you launch for the first time the development environment. The possible values are : installed if you never launched it and initialized if you have already launched at least one time the environment.

• Then start the script start-datafari-devmode.sh :

```
bash start-datafari-devmode.sh
```

For the first launch, it will need some time to launch and initialize Zookeeper, Solr adn Cassandra.

A new user 'admin' will be created in Cassandra database.

The credentials to login to Datafari admin are : admin/admin.

• To stop datafari :

bash stop-datafari-devmode.sh

Note that a convenient script is also present : restart-solr-devmode.sh in order to reload only Solr and not other services like ZK.

#### Run database and ManifoldCF agents

Open a terminal and navigate to DATAFARI\_SRC/mcf/mcf\_home.

Run

bash start-database-dev.sh.

In another terminal, run (ONLY ONCE, the first time you set-up Datafari dev environment)

bash initialize-dev.sh

and then

bash start-agents-dev.sh.

Keep the two terminals open.

## **Run Tomcat**

Go to Eclipse -> Servers view.

Start it (right click on the server name -> start). Datafari application is already deployed by Maven/Ant.

• Please note that we don't rely on Eclipse WTP environment, so the publish, clean, etc ... targets of server view are discouraged. The only target you should run are the "start / stop" of Tomcat.

Note: if you get this exception

INFO: Exception performing authentication. Retrying... javax.naming.CommunicationException: 52.16.74.128:389 [Root exception is java.net.SocketTimeoutException: connect timed out]

it's the, let's say "normal" behaviour =P! (Fix on-going)

Go to the Datafari URL (http://localhost:8080/Datafari)

To login into Datafari, the credentials are : admin/admin

HAPPY SEARCHING 📛 !!!

Need to get Datafari ready for production? Have a look here Distributions HOW-TO

#### Datafari incremental build and redeployment

For redeployment of Datafari Java sources and web resources ONLY (e.g. during an on-going dev), you can use the handy Ant targets "all" or "redeploy" of datafari-dev.xml.

This script recompiles the Datafari core sources and deploys the application on Tomcat.

🕕 If you get strange errors while running Ant, double check to run it on a JDK and not on a JRE (Oracle one, for instance).

## **Configure connectors**

Run Datafari tomcat. Go to the Datafari URL (http://localhost:8080/Datafari). Login in administrator (admin/admin).

In the left menu, click on Connectors; the Apache ManifoldCF login page shows up. Enter admin/admin to login (click on Login button, as the Enter key may not work).

Click on « List Output Connections » of Output section, and add a new « Output Connection ».

In Name tab, set the name to « DatafariSolr ».

In Type tab, set the connection type to Solr and then click on Continue button.

In the Server tab, set port to « 8983 », set web application name to « solr » and « Core/Collection name » to FileShare.

You can now save:

Datafari												
🛉 Mon compte												
X Connecteurs												
Lul Statistiques	. 🐴	Documen	nt Inges	tion								
Administration du moteur de recherche	Apache											
Configuration du moteur de recherche	ManifoldCI	P <sup>IM</sup>										
Serveurs												
Gestion d'utilisateurs	Outputs	Name Type	Throttling	Solr type	Server	ZooKeeper	Paths	Schema	Arguments	Documents	Commits	
	List Transformation Connections											
	List Output Connections	Protocol: http   Protocol: Intervention Server name: Iocalhost Port: 9992		http 🔻								
	Authorities			localhost				7				
	List Authority Groups			_								
	List User Mapping Connections List Authority Connections			8983								
		Web application r	name:	solr					L	à		
	Repositories	Core/Collection n	amo:									
	List Repository Connections	Core/Correction in	ame.	FileShare								
	Jobs	Connection timeo	ut (seconds)	60								
	List all Jobs	Socket timeout (se										
	Status and Job Management	SUCKET IIIIEUUT (SE	econus).	1800								
	Status Reports	Realm:						1				
	Document Status	Lines ID:										
	Queue Status	User ID:										
	History Reports	Password:						]				
	Simple History							_				
	Maximum Activity	SSL trust certificat	te list:	No certific								
	Maximum Bandwidth			Add Ce	rtificate: C	Choisissez un	fichier A	ucun fichie	r choisi			
	Result Histogram	Cours	Pacel									
	Miscellaneous	Save Ca	ncel									
	Help											

Your development environment is now ready.

For more information about ManifoldCF configuration, please have a look here: Crawling

# Save search - technical doc (enterprise version)

The enterprise version of Datafari introduces a new feature which allows you to save a search. This feature is divided in four parts:

- The UI: a new div element has been added to the file WebContent/searchView.jsp. The id of this div element is 'save\_search' and the div is only added when an authenticated user is detected to prevent unsupported access to the feature.
- The widget: the javascript of the widget that adds the search button and implements its behavior is located in WebContent/js/AjaxFranceLabs/widgets/SaveSearch.widget.js and is instanciated in the file WebContent/js/search.js. The widget javascript file needs also to be imported in the file WebContent/searchView.jsp. The widget function is to save all the parameters of the query performed by Solr, thanks to the manager.store object.
- The servlet: Once the user clicks on the "Save search" button, he needs to enter a name and validate. When the 'Validate' button is pushed, the widgets sends the chosen search name and the search infos (the query performed by Solr) to the servlet "/saveSearch". The servlet retrieve the query parameters, the name entered by the user and the username, and insert everything in the Cassandra database. The servlet java code is located in src/main/java/com/francelabs/datafari/servlets/SaveSearch.java

- The database table: The servlet insert the saved search data into the datatable "search" of the Datafari database of Cassandra. This table is created thanks to the script datafari-cassandra/conf/dev-env/tables and is composed of three fields:
  - username: the user name
  - name: the search name entered by the user
  - request: the Solr query parameters corresponding to the saved search

# Logging

Datafari 3.0 comes with a revised way of logging. All the logs generated by Datafari and the other components (Manifoldcf, Solr, ELK) have been formated in order to be exploited by ELK.

Here is the new format :

```
[level] [timestamp] [thread name] - [process name] | [component name] | [package name] | [message]
```

It is implemented through a log4j properties file for each Datafari component. The corresponding log4j pattern is the following:

```
%5p %d{ISO8601} (%t) - [process name]|[component name]|%c{3}|%m%n
```

To understand the specific log4j patterns used, refer to the official doc: https://logging.apache.org/log4j/1.2/apidocs/org/apache/log4j/PatternLayout.html

For example, the following log4j conversion pattern has been defined in the log4j.properties file of Tomcat for the Datafari appender:

```
log4j.appender.datafari.layout.ConversionPattern=%5p %d{ISO8601} (%t) -
Tomcat|Datafari|%c{3}|%m%n
```

This configuration will generate log lines that look like this:

```
INFO 2016-07-25 16:38:08,374 (localhost-startStop-1) -
Tomcat|Datafari|datafari.startup.UserManagementLauncher|UserManagement info
INFO 2016-07-25 16:38:08,851 (localhost-startStop-1) -
Tomcat|Datafari|service.db.DBContextListerner|Cassandra client initialized
successfully
INFO 2016-07-25 16:38:08,936 (localhost-startStop-1) -
Tomcat|Datafari|datafari.alerts.AlertsManager|Alert config file
successfully read
INFO 2016-07-25 16:38:08,945 (localhost-startStop-1) -
Tomcat|Datafari|datafari.alerts.AlertsManager|Alert scheduler started
```

With this kind of format, Datafari can push every generated log line to Elasticsearch thanks to Logstash, and Kibana can be used to visualize and exploit them with efficiency.

## **ELK logs exploitation**

Starting with Datafari 3.1

When ELK is activated in Datafari, every log file generated by Datafari and its components is pushed to Elasticsearch thanks to Logstash. The configuration file of Logstash is the following: DATAFARI\_HOME/elk/logstash/logstash-datafari.conf and the template used to create the Elasticsearch logs indices is DATAFARI\_HOME/elk/logstash/templates/datafari-logs-template.json

If you open the logstash-datafari.conf file, you will notice that the configuration also includes the statistics and monitoring logs. It is because logstash can only manage one config file by process and one of the best practice is to limit the number of logstash processes. This is the reason why you will also notice that the inputs are "marked" with a "type" which can be 'log' or 'exploit' and is used to apply a different parsing and insertion strategy for the inputs.

The other thing to notice in the logstash conf file is that the logs are inserted in a daily index and the logs with a "DEBUG" level are ignored. These lines are used to specify that the logs must be pushed to the index corresponding to the log date (variable %{index\_date} which is created earlier in the conf file):

```
if [type] == "log" {
    if [level] != "DEBUG" {
        elasticsearch {
            hosts => ["localhost:9200"]
            index => "logs-%{index_date}"
            template => "./templates/datafari-logs-template.json"
            template_name => "datafari-logs"
            template_overwrite => true
            flush_size => 100
            idle_flush_time => 1
        }
    }
}
```

For example, the following log line:

```
INFO 2016-07-25 16:38:08,936 (localhost-startStop-1) -
Tomcat|Datafari|datafari.alerts.AlertsManager|Alert config file
successfully read
```

will be inserted in the Elasticsearch index named "logs-2016.07.25". This behavior has been implemented to facilitate the purge of Elasticsearch when necessary. All you need to do is to delete the indices that you consider too old and all the logs corresponding to the index date will be deleted. This is a quick way, with a low resource consumption, to purge Elasticsearch and it is recommended by the Elasticsearch best practices.

Once the logs are inserted in Elasticsearch, you will be able to consult them and create dashboards with Kibana which make it a powerful tool for monitoring. To create dashboards or visualizations for the logs, simply connect to Kibana either by the direct url http://[HOSTNAME]:5601 or through the admin UI of Datafari.

Go to the indices settings, click on the 'Add New' button, and add the new pattern 'logs-\*'

kibana Discove	r Visualize Dashboard Settings	
Indices Advanced Objects Status A		
Index Patterns Warning No default index pattern. You must select or create one to continue.	Configure an index pattern In order to use Kibana you must configure at least one index pattern. Index patterns are used to identify the Elasticsearch index to run search and analytics against. They are also used to configure fields.	
	S Index contains time-based events          □ Use event times to create index names         Index name or pattern         Patterns allow you to define dynamic index names using * as a wildcard. Example: logstash-*         logs-*         Time-field name ● retesh fields         date	1
	Create	-

Check that Kibana detects the field 'date' as in the above screenshot and click on 'Create'. You should obtain the following result:

Notice         Advanced         Opposed         Status         Advanced         Opposed         Status         Advanced         Opposed         Status         <	kibana 🔤	ver Visualize	Dashboard Settings					
Image: Normal Sector         Image: Sector								
Instruction       Specific fields (0)       Instruction (0) <thi< th=""><th></th><th></th><th></th><th></th><th>h. While this list allows you t</th><th>o view the core type of each field, cf</th><th>nanging field types must be done using Elasticsea</th><th>rch's Mapping API %</th></thi<>					h. While this list allows you t	o view the core type of each field, cf	nanging field types must be done using Elasticsea	rch's Mapping API %
Imme			entitlant fields (II)					
processstingpackagesting_indexstingkevelstingcontentstingpathstingcomponentsting				type ≎	format ≑	analyzed 🔀 🗘	Indexed 🔀 ≑	controls
packagesting_indexstingkevelstingcontentstingpathstingcomponentsting		date 💿		date			<b>√</b>	/
Indexstingkevelstingcontentstingpathstingcomponentsting		process		string			*	2
keel     sting        content     sting        path     sting        component     sting		package		string			*	×
contentstringpathstringcomponentstring		_index		string				1
path string Component string Component		level		string			*	/
component string		content		string			*	
		path		string			*	/
host string 🗸 🗾		component		string			*	1
		host		string			*	
		_source		_source				2
		_ld		string				
		_type		string				2
_score number		_score		number				

Now you can create visualizations and dashboards, but you can already some visualizations and dashboards made by France Labs by importing the file datafari-logs-kibana.json located in DATAFARI\_HOME/elk/save. To import them, go to Setting => Objects in Kibana, click on the import button then select the json file.

# Articles de conseil

Add how-to article

Title	Creator	Modified
Adding a new language	Cédric Ulmer	août 03, 2016