everteam

everteam.analytics

Guide

Référence : 52 ECA.V1

Ever Team
336 rue Saint-Honoré - 75001 PARIS - France
17 quai Joseph Gillet - 69316 LYON Cedex 04 - France
33 (0)1 72 71 33 00 • www.everteam.com

Terms and conditions

Documentation

everteam software

Table of contents

Introduction

The everteam.analytics solution allows rapid analysis:

- of structured and unstructured content
- from any type of source and in any data format,
- internal or external to the organization.

It carries out linguistic and analytical processes on data from very diverse origins: databases, file systems, internet sites, etc., using an external indexing server, Solr.

It offers the user specific features when he is viewing the data found by his searches in the everteam application.

Data is consulted via a simplified search, using the following tools:

- Facets,
- Statistical graphics (pie charts, bar charts, etc.),
- Dynamic filters on the search result, using the values in a facet,
- Predefined static filters (configurable).

The solution offers these tools for assisting searches:

- Automatic term suggestion (autocomplete)
- Determining intent ("did you mean?")
- Suggesting more terms ("more like this")

everteam.analytics offers the enterprise three advantages:

- Simplified decision support system (Content Intelligence)
    - Rapid responses
    - Real time access (no complicated data warehouse)
    - Analyzes of content and data
- Improved enterprise content management (ECM)
    - Detects signals (information monitored by the user), monitoring and alerts
- Improved Enterprise Search
    - Authoritative results
    - Results which lead to actions: management information reports, workflow processes, content publication.

This guide presents these features for the user using examples of the interface. Each application may make use of this interface as required.

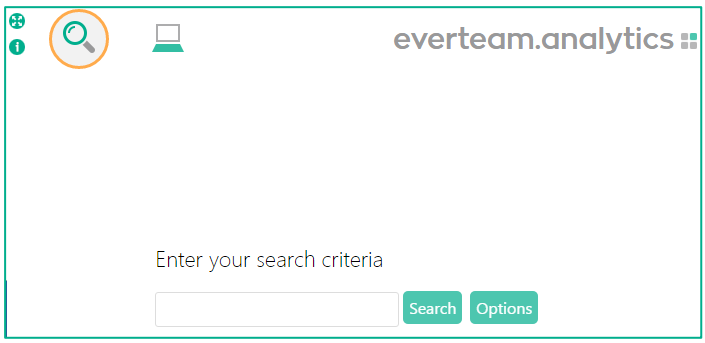
The second part is for administrators. It covers configuration of indexes and results display.

1. Search interface

1.1. Overview

The everteam.analytics user interface offers a single search box for accessing the various sources declared for the application.

A single search box is used, whichever objects are being queried (database, file system or web sites).



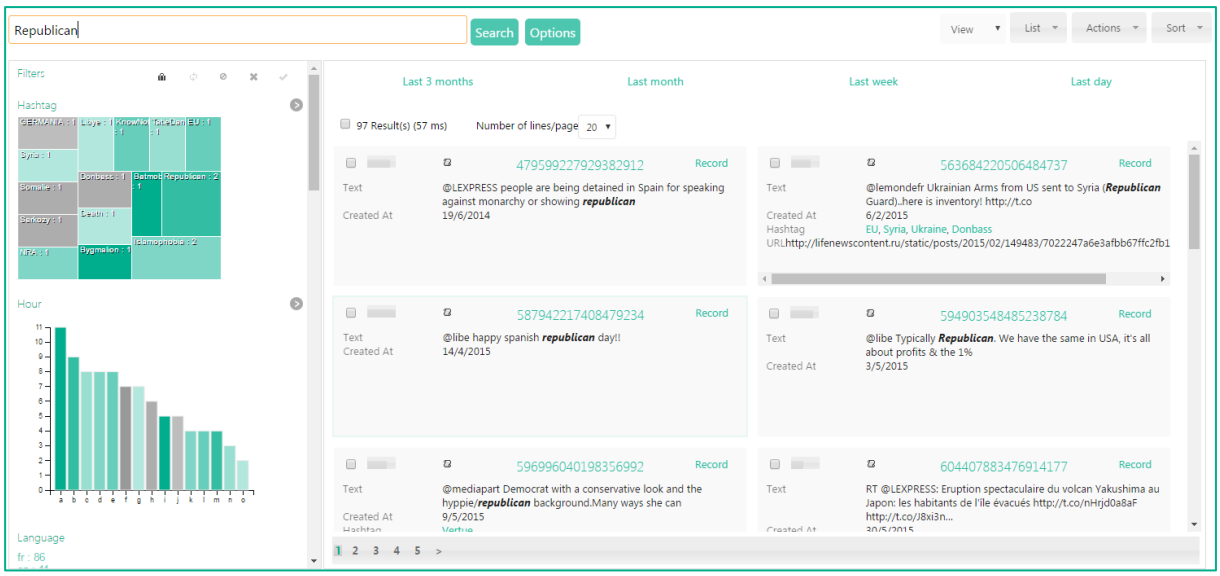
Search interface home page

The results list is displayed, along with any facets, which allow the user to refine her search easily and effectively. All display options are the result of index settings done in the administration module.

After the first search, the window is redrawn, but still with a search box. Below is an example of layout of an everteam.analytics window.

The screen shot below is the example of a search application which will be used in this general introduction to the everteam.analytics search interface.

Information
This guide is based on the demonstration application which the screenshot below comes form. All the elements used can be modified to meet the requirements of your application.



Layout of the everteam.analytics *window*

With the following items in the table above (from left to right)

- Dynamic filters
- Static filter (limits the number of words analyzed)
- Language facet as a bar chart in descending order
- Date facet as a pie chart » Ascending order
- Status bar
- Static filters on the results list
- Results list

1.1.1. Search box

The user can indicate the term he wishes to expand by putting a + after it.

Example: guerre+ could be transformed into guerre OR combat OR war OR etc.

Search box

Under the search box there is a configurable static filter on ranges of numeric values. Several search filters may be available, each one pointing to a different numeric field (number of words in press articles, price of objects displayed, etc.).

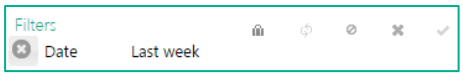
1.1.2. Facet area

Results are displayed in facets based on a particular criterion, common to the indexed data, for example language, publication date or source. Up to six modes may be available:

- Pie chart

- Bar chart
- List (list if identifiers associated with the number of occurrences)
- Palette (equal-sized areas of color)
- ColorMap (areas of color with proportionate sizes)
- Path (tree)
-

Choosing a facet element applies a filter to the results list, which is immediately refreshed. The filter appears above the facets.



Filter display

1.1.3. Menu bar

Results are displayed in the display mode selected in the menu bar.

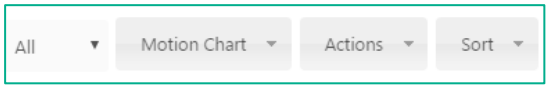
You can define the general display mode in the [View] menu:

- [Facets] displays all the graphical facets (pie chart, bar chart, list, etc.) in full screen mode, replacing the results list.
- [All] restores the standard layout: search box, facets and results list.
- [Results] menu displays the results list in full screen mode.

The [List] menu gives the choice of display mode for results. The default mode is [List] (for results list).

The [Actions] menu offers actions to carry out on the results list or a selection, configured by the administrator (for example: CVS export, Excel export, document export), send by email.

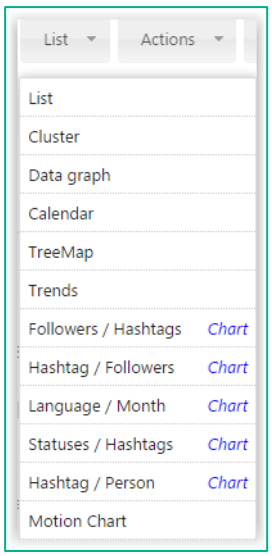
The [Sort] menu displays the list of fields on which the results list may be sorted, defined in the administration interface (default sort, sort fields, sort on multiple fields).



Menu bar for the results list

1.2. List menu

The [List] lets you choose the results display mode from various graphical options, groupings and cross tabs (suffixed with *Chart*). By default, the List option is selected (standard results list: see the next chapter).



List menu

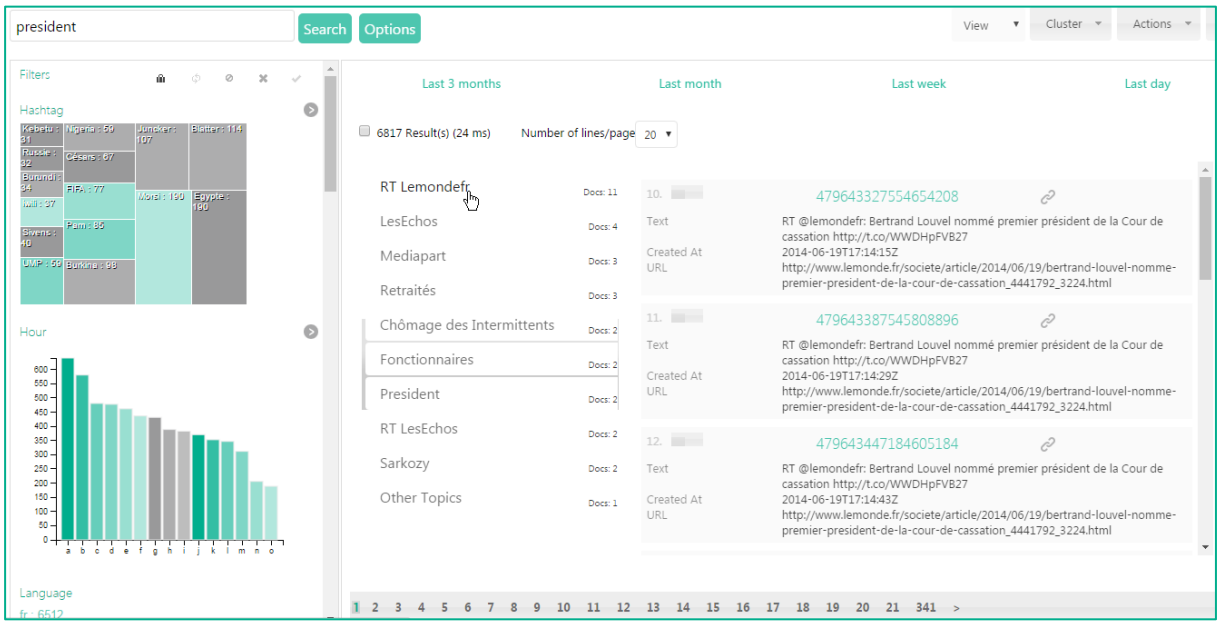
1.2.1. Cluster and Data graph

The Cluster and Data graph options show data grouped by a field (this option depends on how the Cluster and Data graph blocks have been set up in the [Application Params] tab).

This automatic document classification is based on statistics applied to one or more text field(s). The administrator can choose the field from which everteam.analytics will extract the group titles.

- Cluster: grouping in a list

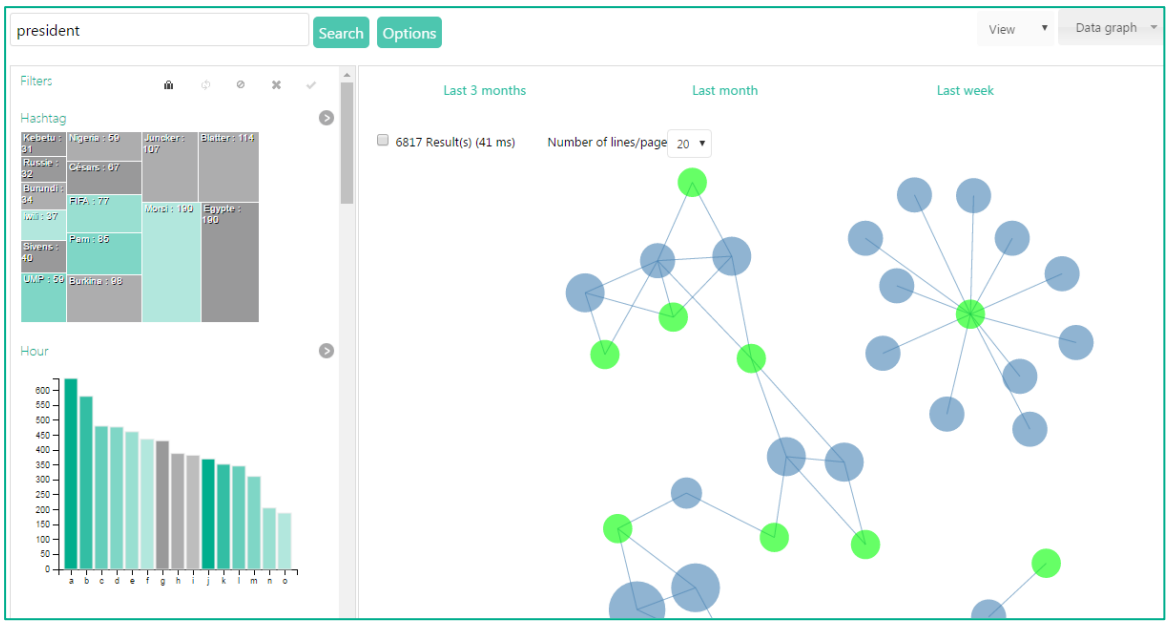
A click in the list displays the matches.



Cluster

- Data graph: grouping in a graph

A click on an item displays the list of matches for this item.

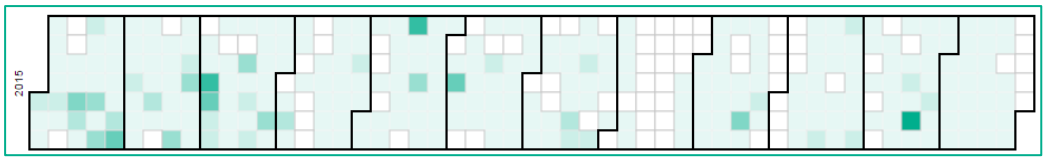


Data graph

1.2.2. Calendar

The Calendar option shows the results in a calendar.

The darker the squares, the more matches there are.



Calendar

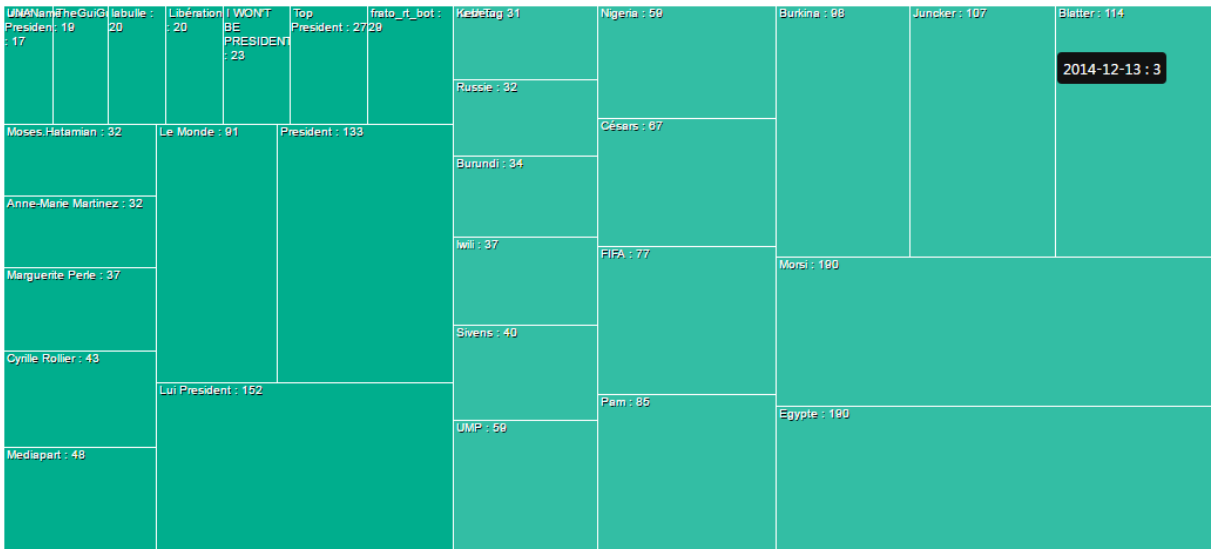
1.2.3. TreeMap

The TreeMap option displays results data in a way that suits a limited space, like a computer screen for example.

This option may also show the data hierarchy.

Each colored area corresponds to a criterion (geographical, personality, organization, etc.).

The areas are sized relative to each other.



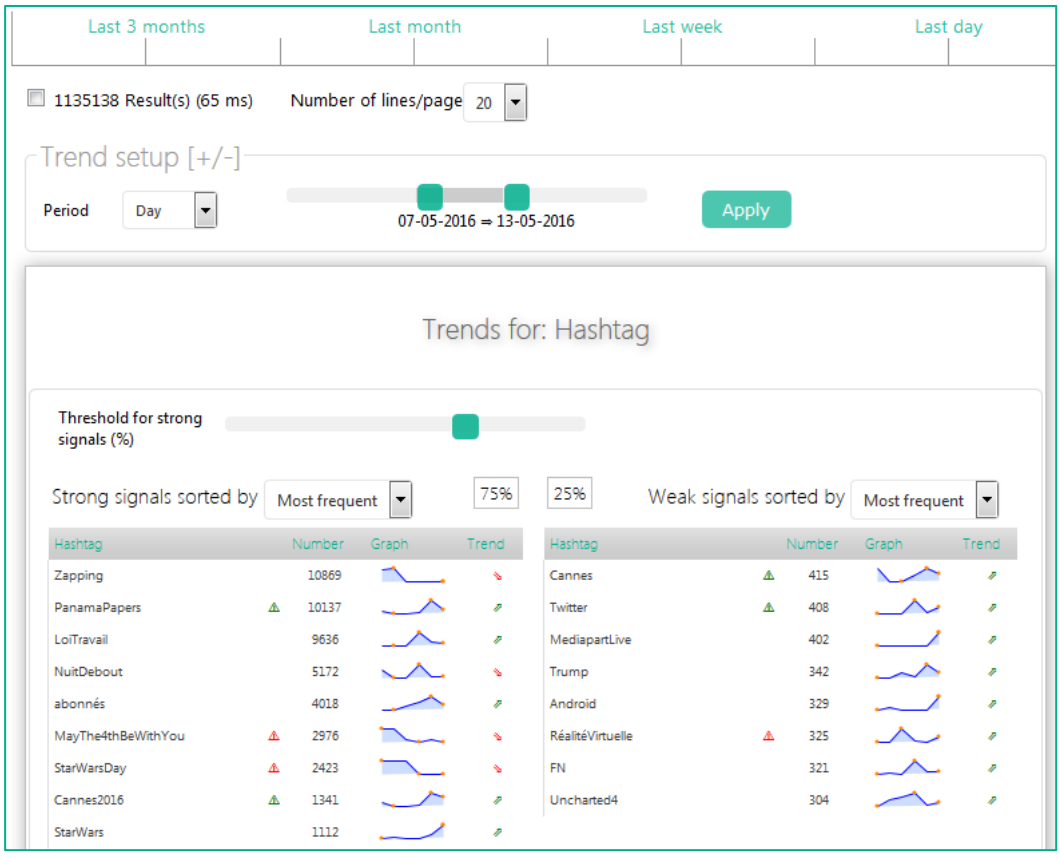
TreeMap

1.2.4. Trends

The Trends option allows you to view trends, potential changes in the targeted data, in a given period.

Select the period by adjusting the gauge then click on [Apply].

The threshold separating strong signals from weak signals is configured (here 75 %), but can be changed by sliding the cursor on the gauge.



Trends

Trends are presented in blocks, here: people, organizations and locations.

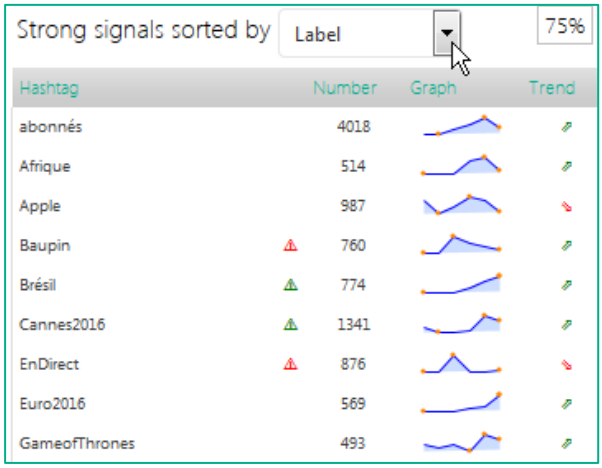
Lists (limited to 10 items) are in decreasing frequency order by default, but other sorts are available, in order to get a different view of the results.

Strong signals can be sorted by:

- Decreasing frequency (from more to less frequent)
- Label
- Increasing trend (positive)
- Decreasing trend (negative)

Weak signals can be sorted by:

- Decreasing frequency (from more to less frequent)
- Label
- Increasing trend (positive)

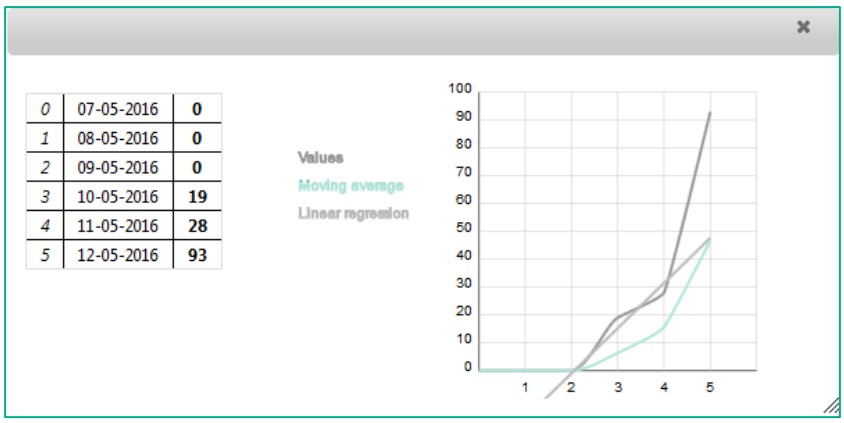


Sort by label

The red or green signal, to the right of the name, indicates a change in the trend, possibly a change of direction, to be correlated with the information in the Trend column.

A click in the Graph column plots several lines:

- Values (bold gray): number of appearances/j
- Moving average (green): rolling three day average to smooth the curve
- Linear regression (gray): general trend.



Trend graph

In the example above, we can see that a trend is strengthening, as shown by the information in the associated graph (the Moving average and the Linear regression converge).

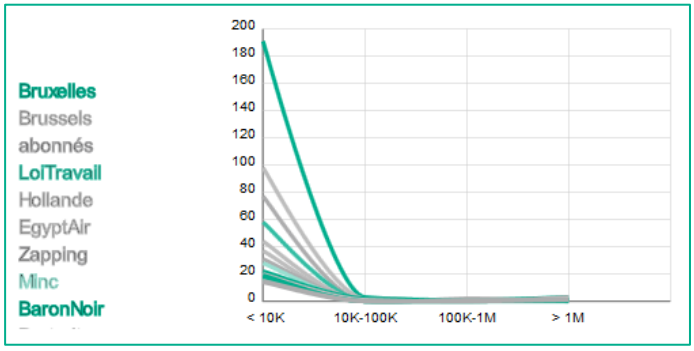
1.2.5. Pivots

Cross tabs give the number of matches for two criteria (configured in the definition of the index ([Application Params] tab – Pivots block)).

> 1M	< 10K	10K-100K	100K-1M	> 1M	Total	Chart
Bruxelles	29688	415	122	302	30527	
Zapping	9995	79	795		10869	
PanamaPapers	9686	170	33	248	10137	
LoiTravail	9153	88	69	326	9636	
Brussels	5448	83	9	70	5610	
NuitDebout	4934	36	30	172	5172	
abonnés	3796	63	6	153	4018	
panamapapers	2948	41	11	44	3044	
MayThe4thBeWithYou	2944	5	23	4	2976	
StarWarsDay	2409	5	5	4	2423	
EgyptAir	1672	22	15	54	1763	
Syrie5ans	1563	23		95	1681	
ONPC	1432	16	8	24	1480	
Cannes2016	1178	27	10	126	1341	
VivreAvec	1054	33	1	64	1152	
Others	*978094*	*9574*	*18481*	*37160*	***1043309***	
Total	**1065994**	**10680**	**19618**	**38846**	**1135138**	

Hashtag / Followers cross tab

A click on Chart on the right of the label, in the [List] menu displays the same data as a graph:



Hashtag / Followers cross tab as a graph

In the example below, the table shows references to people for each month. The date ranges are predefined, declared in the [Application Params] - Ranges of values block.

34719 result(s) (62 ms)

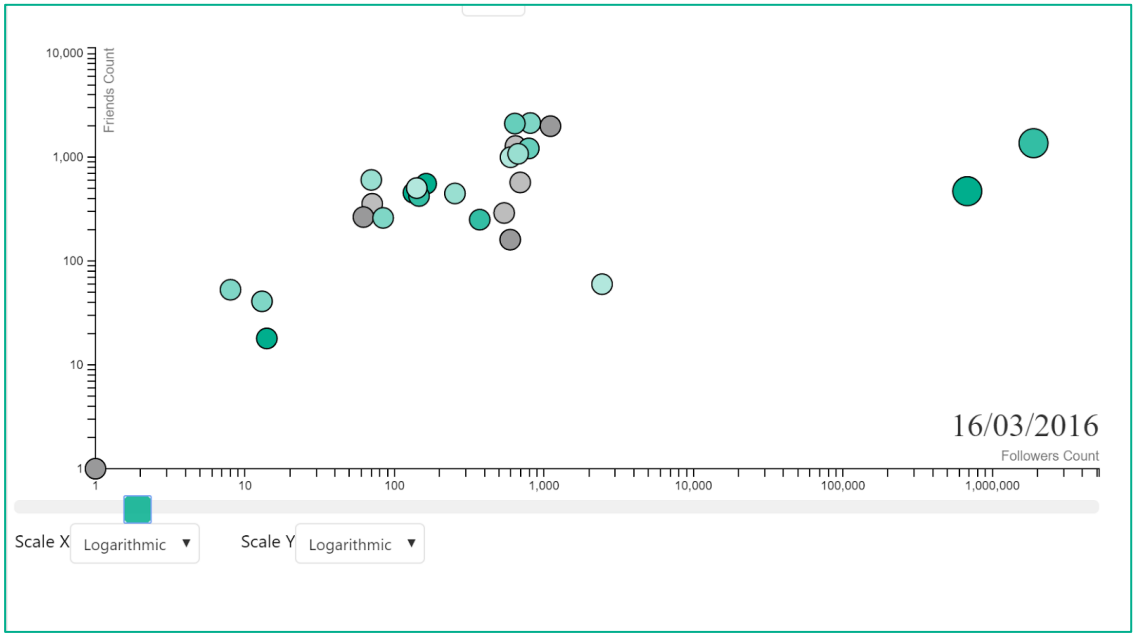
Person / Month	12/2011	05/2012	06/2012	07/2012	08/2012	09/2012	10/2012	11/2012	12/2012	01/2013	02/2013	03/2013	05/2013	06/2013	07/2013	09/2013	10/2013	11/2013	12/2013	Total	Diagramme
François Hollande		17	1297	1586	24	5				9	45	44	70	270	352	360	298	142	9	**4528**	
Jean-Marc Ayrault		66	459	781	14	3				5	10	20	18	127	137	158	158	126	4	**2086**	
Nicolas Sarkozy			577	533	7	3				3	9	72	25	125	246	156	153	62	3	**1974**	
François Fillon			167	375	8	1				1	23	6	11	45	68	160	86	19	4	**974**	
Valérie Trierweiler			597	273							1			13	6	5	12	6	2	**915**	
Jean-François Copé			141	189	11	3				1	10	9	29	83	76	101	90	39	1	**783**	
Manuel Valls			118	55	4	3				5	4	4	20	72	92	162	206	25	2	**772**	
Jean-Luc Mélenchon			383		4						4	18	8	32	25	44	46	13	3	**580**	
Claude Bartolone			395	62							6	2	7	27	24	22	29	5		**579**	
Ségolène Royal			397	84		2					1	2	2	10	28	25	12	8	1	**572**	
Cécile Duflot			241	65	1							3	1	24	33	106	59	8	1	**542**	
Nathalie Kosciusko-Morizet			208	57							4	6	29	32	35	43	25	16	3	**458**	
Arnaud Montebourg			18	189	10	2					2	12		51	48	42	24	16	1	**415**	
Xavier Bertrand			163	171		1					1	1		6	4	12	15	8		**382**	
Barack Obama			1	4	1								2	42	43	131	101	22		**347**	

Person/Month cross tab

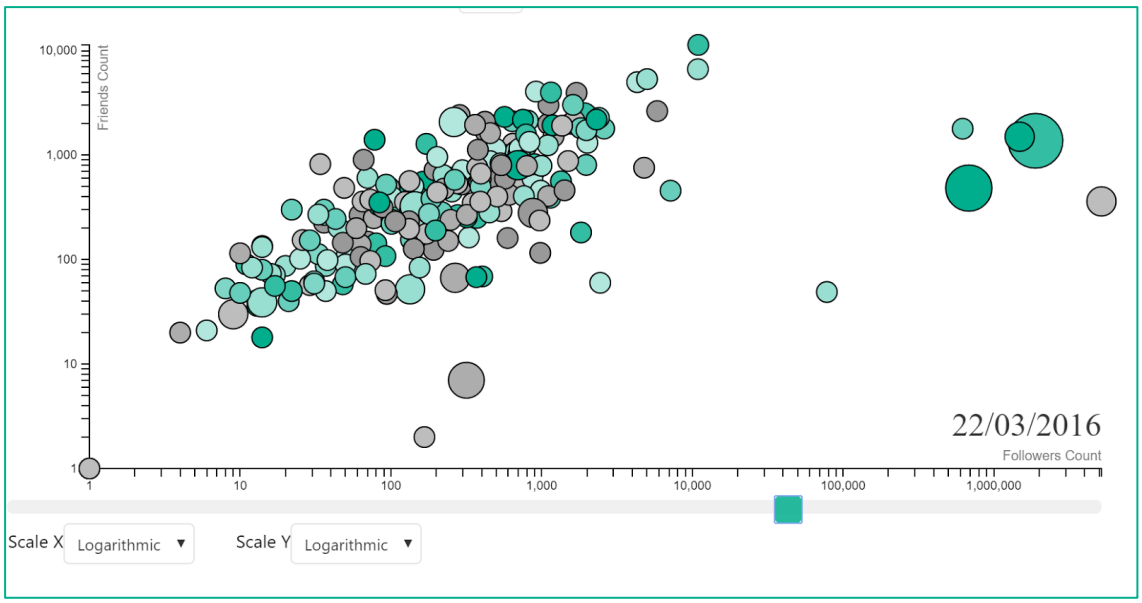
1.2.6. Motion chart

A motion chart shows how data has evolved over several periods.

For example:



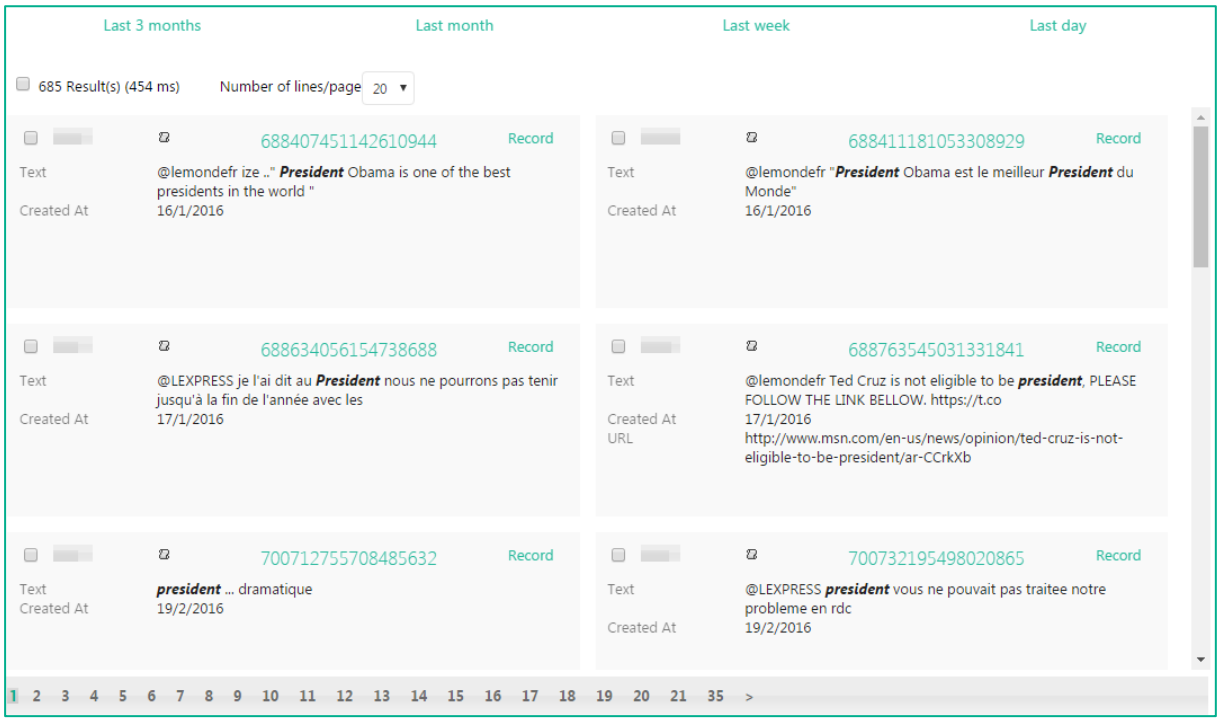
Motion chart at day D



Motion chart at D + 6

1.3. Results list

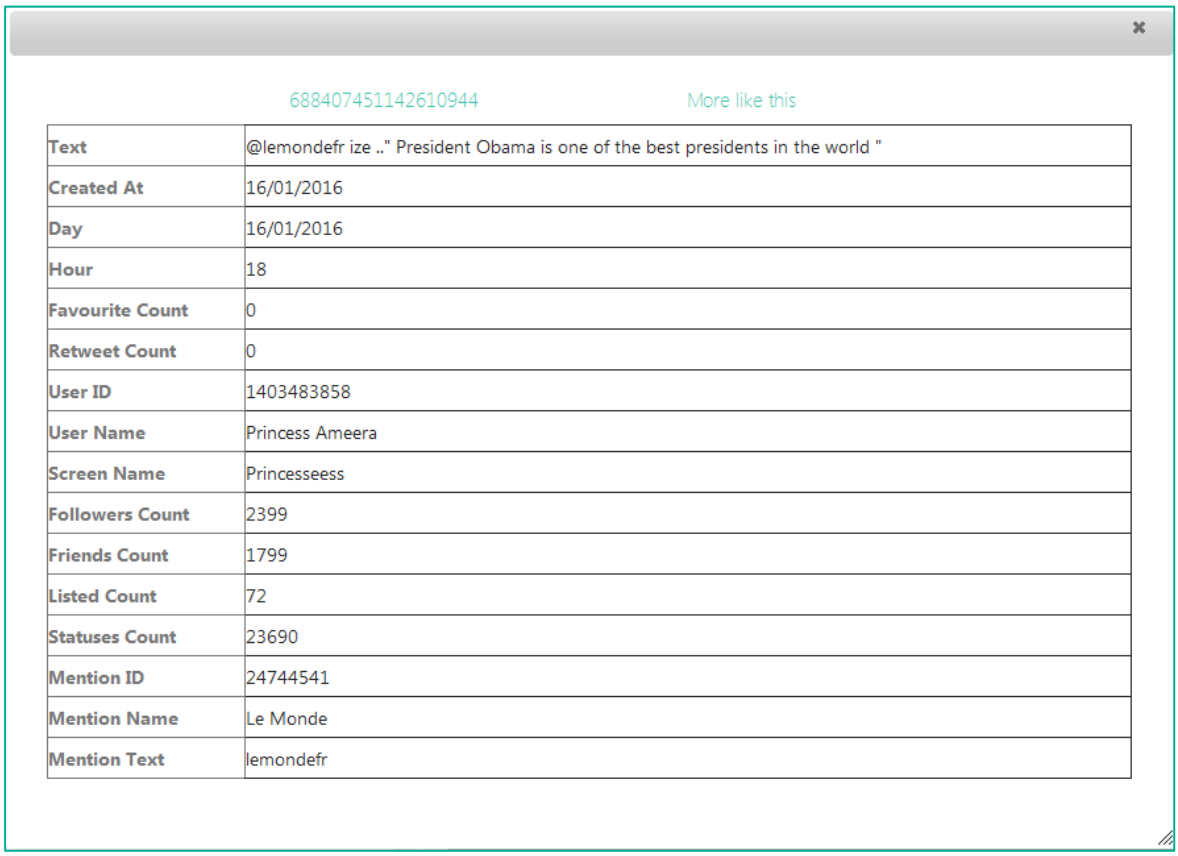
The results list gives access to the record, similar items or attached documents. Results are displayed in order of relevance.



Results list display

1.3.1. Record display

A click on the hyperlink (often the title) displays the everteam.analytics record:



688407451142610944 More like this

Text	@lemondefr ize .." President Obama is one of the best presidents in the world "
Created At	16/01/2016
Day	16/01/2016
Hour	18
Favourite Count	0
Retweet Count	0
User ID	1403483858
User Name	Princess Ameera
Screen Name	Princesseess
Followers Count	2399
Friends Count	1799
Listed Count	72
Statuses Count	23690
Mention ID	24744541
Mention Name	Le Monde
Mention Text	lemondefr

Record display

Information

- The parameter which permits everteam.analytics to display the everteam record is esAccess=true. This is an http request parameter which must be added to the call in the everteam.analytics search form:

apps/esci/ui/htm/index.est?name=/core-news&esAccess=true

- To display the record in another everteam view, its name must simply be included in the ViewName parameter.

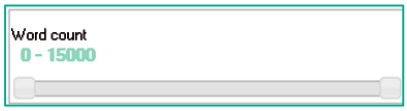
1.3.2. Displaying similar items

A click on [More like this] returns records from the database which are associated with the same words:

1.3.3. Static filters

The results list can be filtered on criteria predefined by the administrator (declared in the [Application Params] tab - block Static filters): date, number of words, price, etc.

Static filter on a date



Static filter on a range of values (range type)

1.4. Viewing facets

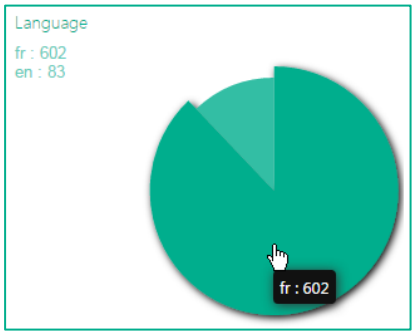
Facets make it possible to show search results from different angles.

A click on an element applies a dynamic filter (shown above the facets) and the results list is updated automatically or manually (depending on how the Filters block is being used).

Six display modes are available, depending on what has been configured: pie, bar, list, palette, colorMap, path. The display mode for results in the TreeMap (under the [List] menu is based on putting several facets together.

1.4.1. Pie chart mode

The pie mode displays result data by sector in a circle, according to respective proportions.



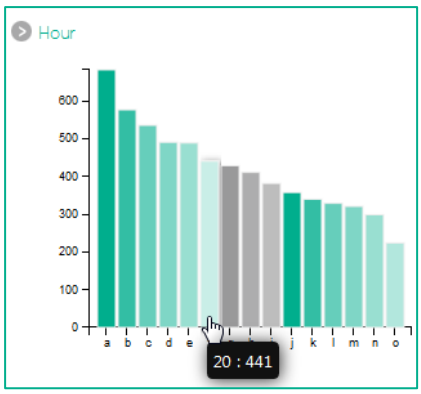
Pie mode

It displays up to nine sectors representing the items with the most occurrences. The other values are all included in the 10th sector, *Others*.

The legend contains the labels for the identifiers associated with the occurrences. When the mouse hovers on a sector, its label is highlighted in the legend.

1.4.2. Bar chart mode

The bar mode displays facet values in a histogram, based on occurrences. Labels are displayed under each bar. The number of bars is restricted in the setup ([Application Params] tab: General parameters /Max rows block).



Bar mode

1.4.3. List mode

The list mode displays facet values as a list: identifiers with their number of occurrences (sorted by occurrence). The number of rows is restricted in the setup ([Application Params] tab: General parameters /Max rows block).

List mode

1.4.4. Palette mode

The palette mode must be applied to a field containing a name or a color code (example: red, black, #c125d4).

It displays a palette of small colored squares like the one used for selecting a color in various applications.

Palette mode

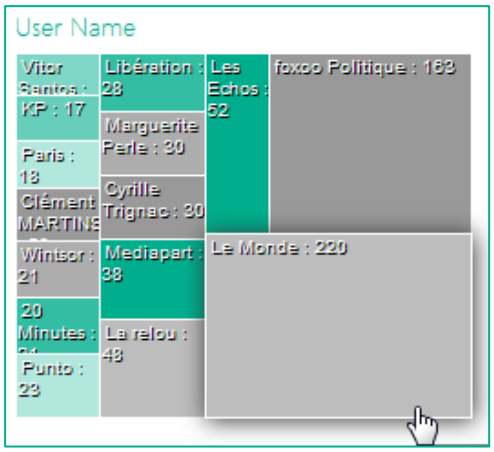
The areas are in different colors but are all the same size. The number of areas is restricted in the setup ([Application Params] tab: General parameters / Max rows block).

The number of occurrences appears in each area.

1.4.5. colorMap mode

The colorMap mode must be applied to a field containing a name or a color code (example: red, black, #c125d4). It displays a palette of colored areas whose size is in proportion to the number of occurrences. This number appears in each area.

The number of areas is restricted in the setup ([Application Params] tab: General parameters / Max rows block).

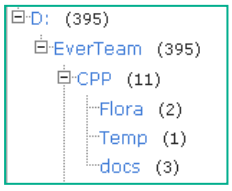


colorMap mode

1.4.6. path mode

The path mode is for facets based on fields containing one or more fragment(s) of text separated by a " /" (as in UNIX paths).

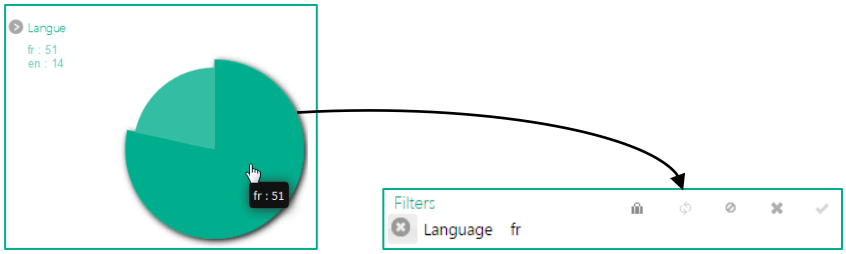
This mode displays a tree view (example: paths of indexed files). The number of occurrences is given in brackets.



path mode

1.5. Dynamic filters

Dynamic filters are created from facets, by clicking on an item.



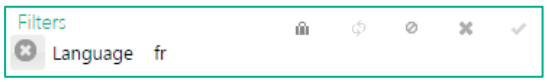
Creating a dynamic filter

They are activated either automatically (single selection) or manually (multiple selection).

1.5.1. Single selection

If there is a single selection per filter, the interface includes these tools:

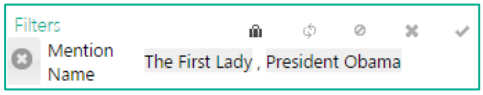
- go into multiple selection
- exclude the item from filters (*NOT*)
- delete the filter (or use the icon next to the filter)



Dynamic filters on a single selection

1.5.2. Multiple selection

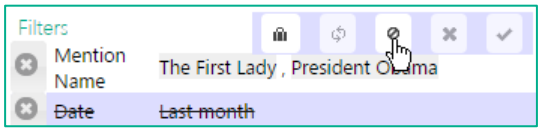
When you go into multiple selection, you can define several items for the same filter, and adapt the operator used.



Dynamic filters on a multiple selection

It has these tools:

- go into single selection
- define operators between items (*AND*/*OR*)
- exclude the item from filters (*NOT*)
- delete all filters
- apply filters



Excluding an item from the filters Configuration

2. Configuration

Specific everteam.analytics parameters can be configured in the everteam.core > Parameters > Services > everteam.analytics menu.

This screen contains all the parameters which the administrator can modify for his/her application.

Parameter	Description	Default value
esci.export.rootPath	Export path	*/tmp/esca/export*
esci.fs.extractNamedEntities	Extract named entities for file system indexing *yes* | *no*	*yes*
esci.indexRefreshDelay	Delay before refreshing the index (seconds)	*7200*
esci.indexScriptsIntoDB	Store the everteam.analytics index description files in the system database *true* | *false*	*false*
esci.solr.home	Solr directory	*solr (*)*
esci.solr.qretry	Number of attempts when there is an error when querying (default: 5)	*5*
esci.solr.qretry.sleep	Interval between each attempt (milliseconds)	*100*
esci.solr.secure	Secure connection to Solr *true* | *false*	*true*
esci.solrIndexing.sliceDelay	Delay before sending a batch of Solr documents to the index (seconds)	*5*
esci.solrIndexing.sliceSize	Number of documents in a batch sent to the Solr server	*1000*
esci.web.extractNamedEntities	Extract named entities for Web indexing	*yes*

	yes | *no*	
esci.zookeeper	Address and port of the ZooKeeper server	*localhost:2181*
esci_database	Deployment database name	*dPortal*
esciStartService	Start the ESCI service *true* / *false*	*false*

ESCI service parameters

(*) To be changed at the start of the project so that this directory is no longer in the webapp.

3. Security

Security in everteam.analytics is managed using:

- rights on indexes, defined in the SIS service: search (*SEARCH*), modification (*MODIFY*), creation (*CREATE*), deletion (*DELETE*),
- search filters configured in everteam.analytics
- secure access: /dataimport must be removed from the solrconfig.xml file (advanced configuration).

If a user has no access rights, an http message is sent with http return code: 401.

Reference

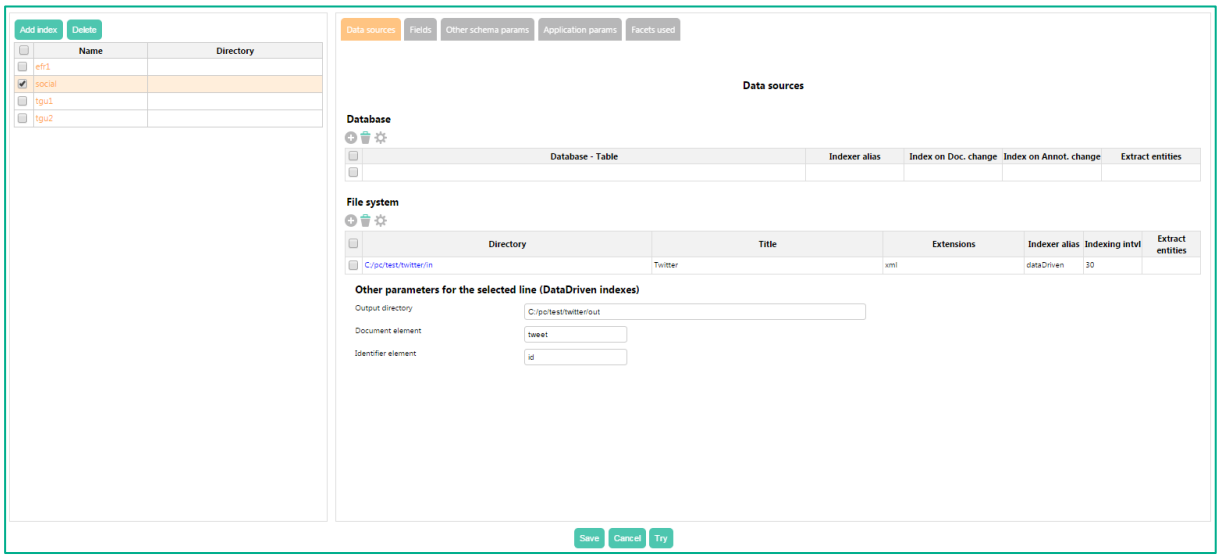
For more information on index rights, please refer to the SIS Indexing and search guide.

For more information on everteam security, please refere to the CORE Security guide.

4. Administration interface

everteam.analytics offers a single interface for administering all the indexes.

To administer the everteam.analytics solution, click on the everteam.analytics > Administration menu:



Index configuration window

The list of existing indexes is on the left. There are several possible actions:

- Add a new index
- Delete an index
- To configure an index, click on the link in the Name column.

The section on the right is for configuring different aspects of everteam.analytics indexes.

4.1. Managing indexes

It is easy to manage indexes in the administration interface: the two most basic operations, adding and deleting an index can be carried out with the [Add index] and [Delete] buttons.

To modify the characteristics of an existing index, just cllck on its name (on the left). All the setup is displayed in the center of the screen.

Generic buttons, enabled for all updates are at the bottom:

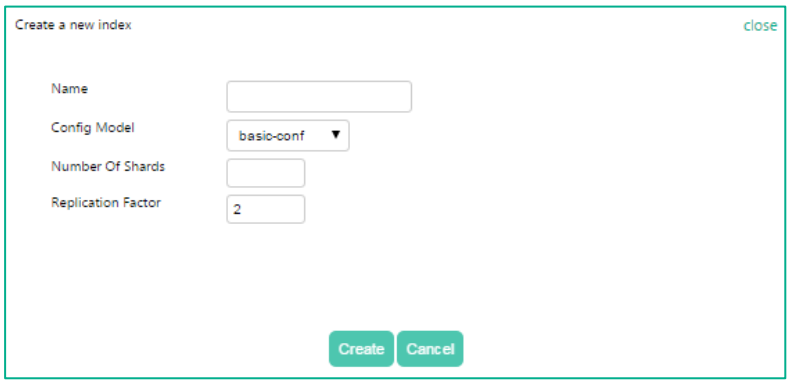
- The [Save] button is for saving all configuration data, from any of the tabs, distributed over several files and tables.
- The [Cancel] button comes out of input mode without applying any of the updates in any of the tabs. It refreshes the display with data from the server.

- The [Test] button gives access to the search interface to test the current index set up. You must use it after saving the parameters. Caution, some setups require the application server to be restarted.

Actions requested from the server are stored in a Log which is kept until the end of the session. After saving an index, a click on the [Message Log] button (at the bottom of the screen) displays the latest messages in a window.

4.1.1. Add a new index

When you click on the [Add index] button in the pane on the left, a dialog box opens:



Adding a new index

The administrator enters the name of the new index which will be created using XML files supplied.

Config Model: collection model to be selected from the list of defined configuration models (stored in apps/esci/conf/solr).

Number of Shards: number of shards to configure (depends on the number of nodes in the cluster). SolrCloud is designed to function in cluster mode (a cluster contains several nodes which are different servers). A collection is a set of indexes distributed over several nodes. Each part of the collection on a different node is a shard. The number of shards to configure depends on the number of nodes in the cluster and must be less than or equal to this number. (Example: on a cluster with 3 nodes, you can have between 1 and 3 shards). Logically, you would have 3 nodes in order to split the collection into 3 shards.

Replication Factor: To handle fail-over or a large number of users carrying out searches, a shard may be replicated one or more times. In a normal production configuration, we recommend a replication factor of at least 2. This means that each shard exists twice, with one active copy and one passive.

For every index created, a subdirectory is created for the directory defined in the esci.solr.home parameter (default: */solr*) It has the same name as the index and contains:

A <index>/conf subdirectory containing 2 JavaScript files:

- index_name.js: this file is generated and modified automatically by the everteam.analytics administration interface.
- index_name.esfct.js: created by the administration interface, but it will be modified manually to meet project requirements. It replaces the "script" block which has been withdrawn from the administration interface for security reasons.

A conf-copy subdirectory: This contains a copy of the configuration in ZooKeeper. It can be modified exceptionally by an administrator with good knowledge of Solr before being pushed to ZooKeeper

4.1.2. Delete an index

The administrator selects one or more line(s) and clicks on the [Delete] button.

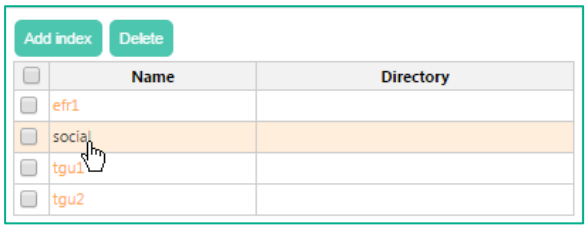
This action deletes the index, the index directory and the corresponding configuration files and updates the everteam tables.

A message asking for confirmation is shown in a pop up: *"Are you sure you want to delete these indexes permanently?"* This allows you to confirm the action or to go back, by clicking on "Cancel").

4.1.3. Viewing and updating an index

When you select an index from the list on the left, you get access to all its parameters.

They are described in the next paragraph.



Selecting an index

4.2. Characteristics of an index

The main part of the screen is dedicated to the setup of the index selected on the left.

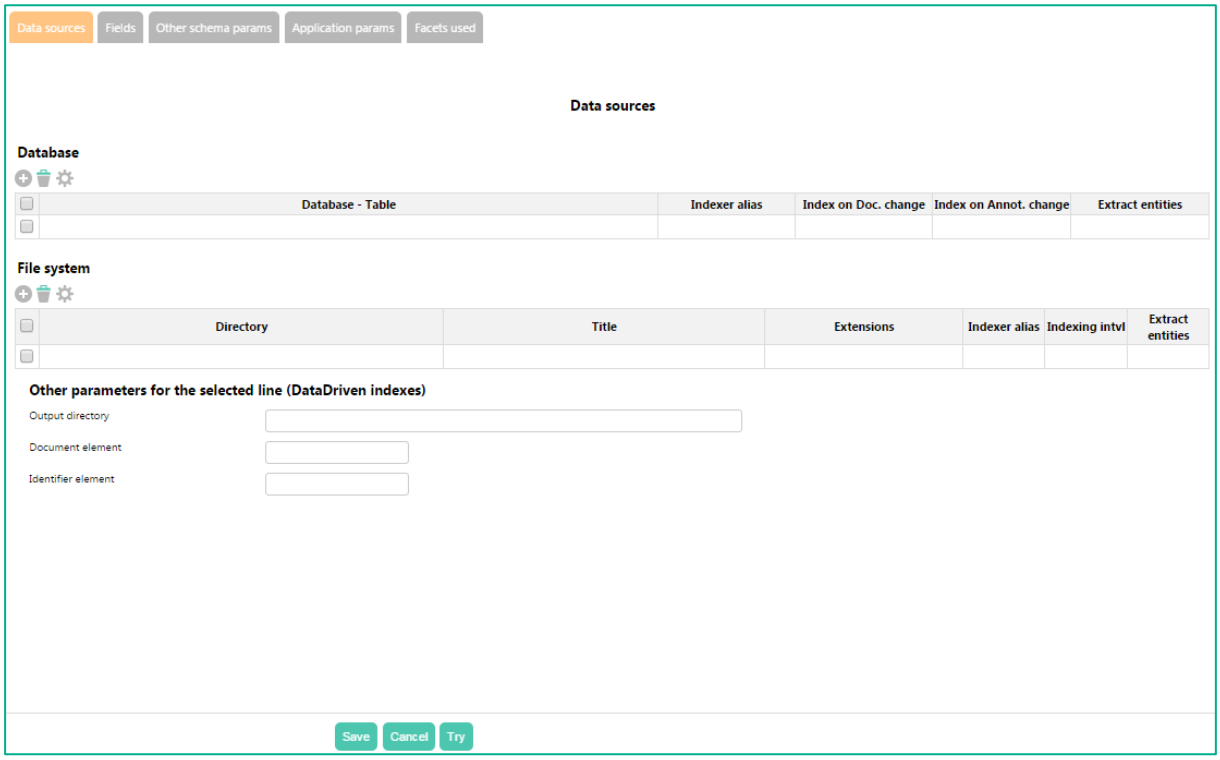
The tabs display values from the schema.xml and solrconfig.xml files, by default in the ./solr/multicore/<*index-name*>/*conf* folder (as well as in the everteam system database and other XML files).

Information
At the beginning of a project, it is recommended to change the location of the index directory, so that it is no longer in the webapp. **See the esci.solr.home parameter**

All the labels and titles defined in these configuration tabs - and displayed in the search interface - can be translated, provided that this localization is saved in the application's dictionary.

4.2.1. Data sources

The [Data sources] tab contains the following window:



Data sources] tab

The interface lets you configure the two types of data source which are permitted, in the [Data sources] tab:

- Database (table)
- File system (directory)

The administrator may index each data source specifically. Two different sources cannot be indexed at the same time. The different procedures for each data source must be run successively.

For each type of source, the available tools are:

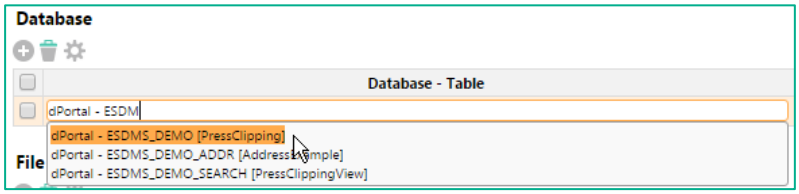
- [Add] Add a new source of the same type
- [Delete] Delete selected (checked) sources
- [Index all] Generate index for source(s) of the same type. In this interface, this is only for the initial indexing. Afterwards, the indexes will be updated by incremental indexing.

Once the different data sources have been declared, you should save using the [Save] button and start generating the index using the [Index all] button.

Database

Indexing databases

Database - Table: the source is selected from the list of tables declared in the selected database (by auto-completion). The drop-down list displays the name of the table or the view, preceded by the name of the database.



Selecting a table using auto-complete

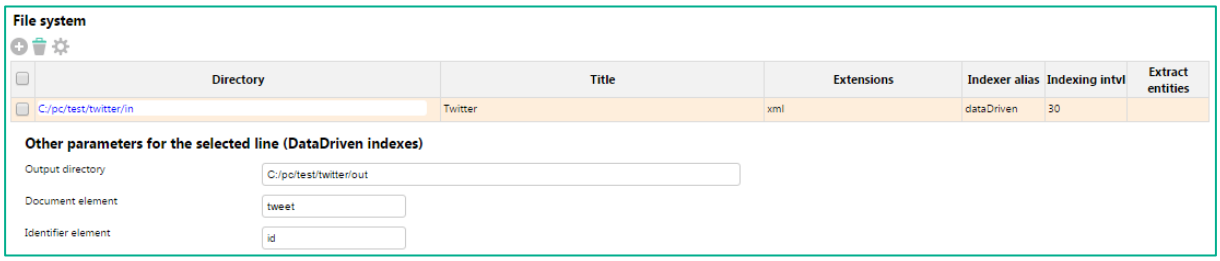
Indexer alias: indexer used: *default* (for indexing records), *dms* (for indexing records + attached files), *item* (RSS feeds), *spl* (spools), *hds* (HDS folders), *docs* (CSDOCS documents), *cws* (Collaboration Space).

Index on Doc change: check this field to reindex the table record if attached documents have been added, updated or deleted (*true*|*false*).

Index on annot. change: check this field to reindex annotations added to records (*true*|*false*).

Extract entities: check this field to index entity extractions (*true*|*false*).

File system



File system indexing

Directory: enter the directory path, in plain text.

Title: title associated with the index for display in everteam.analytics.

Extensions: extensions of the files to be indexed, separated by semi-colon (example: doc;pdf;xls;docx). This information is saved in the CSFTDIR table.

Indexer alias: two possible values:

- *default* (indexes all files)
- *datadriven*: indexing of a document stream which arrives in a working directory before being moved or deleted. The contents of an XML file indicate to the indexer which operation to carry out (Index/Delete). In this case, complete the fields in the Other parameters for the selected line (DataDriven Indexes) sub-block.

Indexing intvl: incremental indexing interval.

Extract entities: check this field to index entity extractions (*true*|*false*).

Other parameters for the selected line (DataDriven indexes:

- Output directory: this directory will receive files after they have been indexed. If this parameter is empty, the files are deleted after indexing.
- Document element: the physical XML file (DataDriven only handles XML files) may contain several logical XML documents. To be able to distinguish them, you must give here the name of the element at the root of each document.
- Identifier element: an element is designated as containing the "unique identifier".

4.2.2. Fields

The [Fields] tab displays in alphabetical order the fields to be indexed and the fields to be shown in the results list, the record, attached documents or text extracts, etc.

Fields can be imported using the buttons above the table, depending on the sources which have been set up.

Name	Label	Type	Indexed	Stored	Multivalued	Default	List Mode	Doc Mode	Highlight	Named entities	Link	Facet Mode
DBName	DataBase	string	true	true								
TableName	Table	string	true	true			true	true				
root		string	true	false								
version		long	true	true								
a	tempo	text_general	true	false								
createdAt	Created At	date	true	true			true	true				
cs_UKValue		string	true	true								
cs_contents		text_general	true	false	true							
cs_oper	Oper	string	true	true		stream						list
cs_timestamp		tdate	true	true		NOW						
cs_type		string	true	true								
cs_uid	Id	string	true	true			true	true				
dateJour	Day	date	true	true				true				
favouriteCount	Favourite Count	int	true	true				true				
geo	Geography	string	true	true	true			true				
hashTag	Hashtag	string	true	true	true		true	true			true	treeMap
hour	Hour	string	true	true				true				bar
id	Identifier	string	true	true				true				
lang	Language	string	true	true			true	true				pie
mediaPageUrl	Media Page URL	string	true	true	true			true				
mediaType	Media Type	string	true	true	true			true				

Fields tab

For each field, various options can be customized. Most of these are Boolean: *true* or *false*. For the others the option has a list of values.

Predefined values must be kept.

The first fields are in the schema.xml file. The others are stored in a JavaScript file.

Adding fields

This table can be added to manually or by using the buttons above it, depending on the sources which have been set up:

Buttons for adding fields to the index

Some index fields are common to several configurations.

The default fields are as follows:

Fields common to several configurations:

- cs_type
- cs_lang
- Id

Default fields to be kept:

- DBName
- TableName
- Title
- cs_timestamp
- cs_contents

The fields which are displayed in red are the ones which were already in the list; the ones in blue are new and are at the end.

Import

Import all fields from the table declared in the [Data sources] tab.

Standard fields

Add the fields needed for each index (id, cs_lang, cs_type, cs_content_, etc.).

File system

Add the fields to be used when indexing a file system.

Named entities

Add fields corresponding to named entities, such as names of organizations, personalities, locations (geographic) or nationalities, or dates. Indexing can also identify titles (Mr., Mrs., Dr., Excellency, Sir, etc.) or structured character strings, such as an email address.

Add/Delete

The last two buttons allow you to add fields to, or delete fields from, the list freely:

- [Add]: add a row to the table (free field entry)
- [Delete]: delete a row

Description of table columns

Name	Label	Type	Indexed	Stored	Multivalued	Default	List Mode	Doc Mode	Highlight	Named entities	Link	Facet Mode
DBName	DataBase	string	true	true								
TableName	Table	string	true	true			true	true				
root		string	true	false								
version		long	true	true								
a	tempo	text_general	true	false								
createdAt	Created At	date	true	true			true	true				
cs_UKValue		string	true	true								
cs_contents		text_general	true	false	true							
cs_oper	Oper	string	true	true		stream						list
cs_timestamp		tdate	true	true		NOW						
cs_type		string	true	true								
cs_uid	Id	string	true	true			true	true				
dateJour	Day	date	true	true				true				
favouriteCount	Favourite Count	int	true	true				true				
geo	Geography	string	true	true	true			true				
hashTag	Hashtag	string	true	true	true		true	true			true	treeMap
hour	Hour	string	true	true				true				bar
id	Identifier	string	true	true				true				
lang	Language	string	true	true			true	true				pie

[Fields] tab

Name

Name of the field to index

Label

Label of the field to index

Type

Type of indexing for the field The main types are:

boolean: two possible values: *true* and *false*

date: date format (e.g. 2013-12-01T12:30:45Z)

double: double precision decimal value

float: approximate value, which may contain decimals

int: numeric value - integer

long: large numeric value (too long to be an integer)

string: character string : indexing of the whole exact value in the field

tdate : t try : more rapid indexing and searching

text_en: English text

norm-string: character string : indexing of the whole exact value in the field, without taking case or accents into account

text: text format: indexing "splits" the field into words

text_fr: French text

text_ar: Arabic text

phonetic: phonetic: used for proper names, the phonetizer can work in FR/EN/AR

path: field type to use for the facet path.

Information
text_fr, text_en and text_ar are derived from the text type, with more advanced analysis specific to each language.

Indexed
This option indicate whether the field must be indexed or not, to be available for searching.

Values: *true* | *false*

Stored
This option indicates whether the value of the field must be stored in the index in order to be returned in the results list (the field value appears in the results without returning to the data source (database, file system, etc.)).

Values: *true* | *false*

Multivalued
This option indicates if the field is multivalued.

Values: *true* | *false*

true: each value in the field is indexed individually.

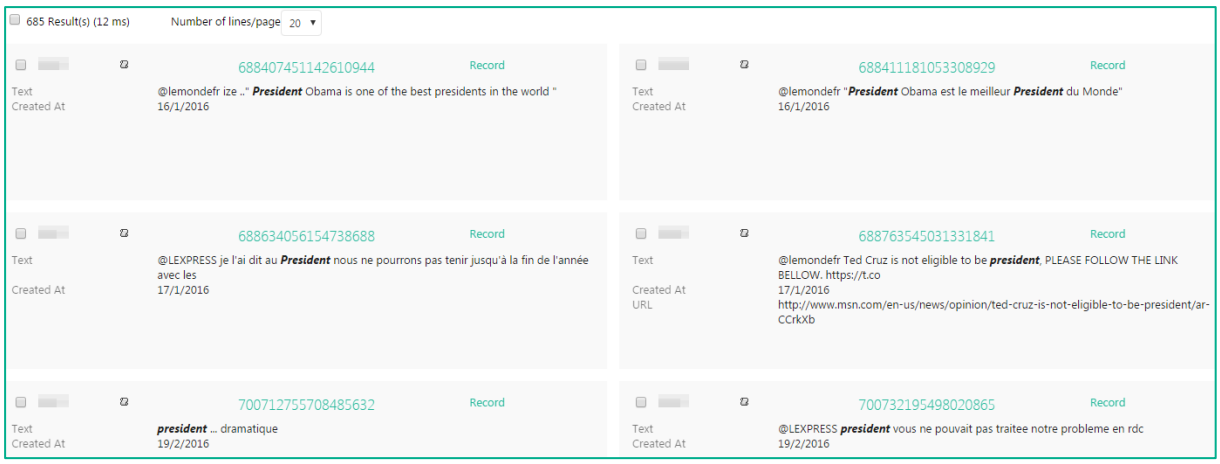
Default
This option indicates how indexing will handle the field when it is empty: it may use a default value.

For example, for a date field, today's date (*NOW* - current date and time) may be given; for a numeric field, the value *0*may be given.

List mode
This option indicates whether the field value is displayed in the results list or not.

Values: *true* | *false*

If List mode = *true*, then Stored = *true*.



Results list, with field linking to the record

The first field which links to the record is declared in the *General parameters* block of the [Application params] tab.

Doc. mode

This option indicates whether the field is displayed in each record.

Values: *true* | *false*

If Mode Doc. = *true*, then Stored = *true*.

688407451142610944 More like this

Text	@lemondefr ize .." President Obama is one of the best presidents in the world "
Created At	16/01/2016
Day	16/01/2016
Hour	18
Favourite Count	0
Retweet Count	0
User ID	1403483858
User Name	Princess Ameera
Screen Name	Princesseess
Followers Count	2399
Friends Count	1799
Listed Count	72
Statuses Count	23690
Mention ID	24744541
Mention Name	Le Monde
Mention Text	lemondefr

Record example

Highlight

This option indicates whether search terms (as well as their synonyms and translations) are highlighted or not.

Values: *true* | *false*

Named entities

This option makes it possible to extract structured data from unstructured data.

These data, known as "named entities" and present in the current field are indexed in specific index fields, added in the [Named entities] tab: email addresses, geographical names, names of personalities, etc.

Values: *true* | *false*

Link

This option indicates if the field value must appear as a hyperlink, in the results list and in the record.

The value of this column indicates how this hyperlink should be prepared.

Values: *true* (default) | *esfct.addFilePathLink* | *esfct.addHttpLink*

- *true*: the link allows you to run a new search using the value of the current field as the criterion.
- *esfct.addFilePathLink*: the link allows you to access a file (example: URI addresses).
- *esfct.addHttpLink*: the link gives access to a web address (example: a character string, like "www.everteam.com").

These last two options correspond to scripts in the .js file for the index collection.

Other options can be customized by the administrator following the same procedure, provided that the naming convention is respected: esfct.xxx.

Facet mode

This option is for indicating how to display data in facets.

Values: *bar* | *list* | *pie* | *palette* | *colorMap* | *TreeMap* | *path*

bar: values displayed in a histogram. Labels are shown under the bars. The number of occurrences is shown when the cursor is on the bar.

list: values in a list, followed by the number of occurrences.

pie: values represented in a circle (pie chart), in proportion to the number of occurrences. The legend gives the labels and the number of occurrences.

palette: values displayed in areas of the same size. The number of occurrences is displayed in each area.

colorMap: values displayed in areas sized in proportion to the number of occurrences.

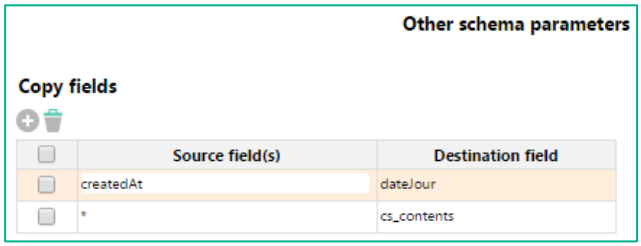
treeMap: same type of presentation as *colorMap*, based on any value.

path: values displayed in a tree. The number of occurrences is given in brackets.

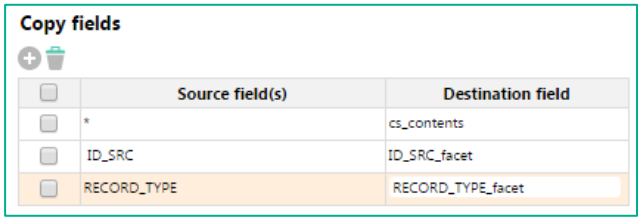
4.2.3. Other schema parameters

In the [Other schema] tab, fields are copied and indexed in the search field given (e.g. cs_contents).

In the example below, all fields which are copied and indexed in the "cs_contents" field.



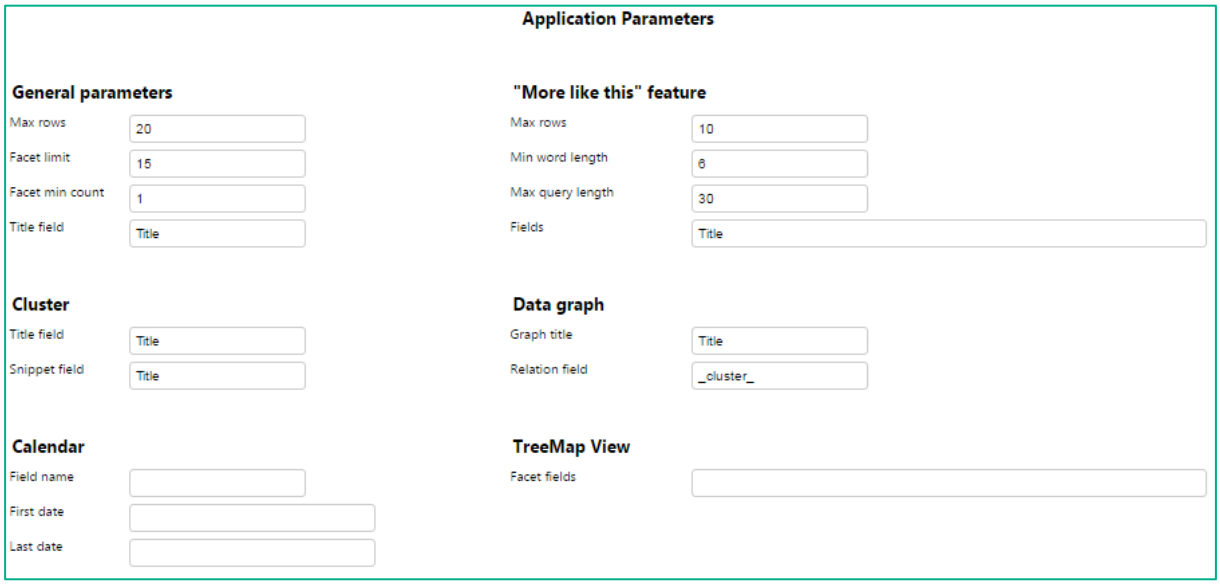
Other schema parameters / Example 1



Other schema parameters / Example 2

4.2.4. Application parameters

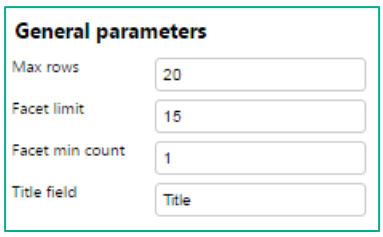
The [Application params] tab contains several blocks for defining how results are displayed in the search interface (used when generating the JavaScript file (.js) for the index).



Application params tab (extract)

General parameters

Display of the results list and faceted graphs.



General parameters

Max rows: default number of matches per page in the results list.

Facet limit: maximum number of values shown for facets (maximum number of sectors). This option is limited to *10* for a pie (pie chart); only a value set to less than *10*will be used when displaying the facet.

Facet min count: minimum number of occurrences for a facet value to be displayed.

Title field: name of the field used to make the link with the record in the results list

More like this

To display similar items, everteam.analytics searches for them using terms which it takes from the selected item.

More like this

Max rows: number of matches per page in the results list.

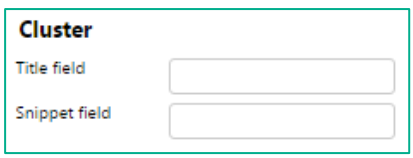
Min word length: minimum size of a word for it to be used as a criterion, as a number of characters.

Max query length: Maximum size of the query, as a number of words.

Fields: list of the indexes in which search term will be looked for.

Cluster

Result displayed grouped into clusters, accessible in the search interface using the [List] > [Cluster] menu.



Cluster

Title field: field used for cluster titles.

Snippet field: field (usually text type) containing the text stream to analyze in order to group results.

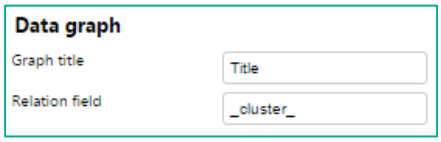
Example: after doing a search, everteam.analytics automatically groups results using the contents of the ACCORD field (summary text) in the matches and deduces the title of each group from terms in common found in the TITRE field.

Example of cluster

Note: the fields mentioned here must be stored (Stored = *true* in the [Fields] tab) and have an indexing mode defined (e.g.: *text_fr*).

Data graph

Graphical display of groups, accessible in the search interface using the [List] > [Data graph] menu.

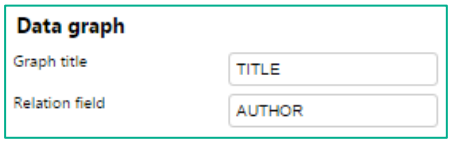


Data graph

Graph title: field giving the title of the graph.

Relation field: grouping field. The *_cluster_* entry refers to the grouping field defined in the Cluster block (described above).

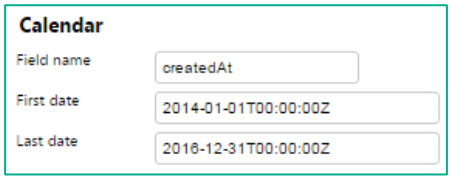
Another field may be defined for the graphical link, such as TITLE or AUTHOR:



Graphical link on TITLE or AUTHOR

Calendar

Results displayed in a calendar, accessible in the search interface using the [List] > [Calendar] menu.



Calendar

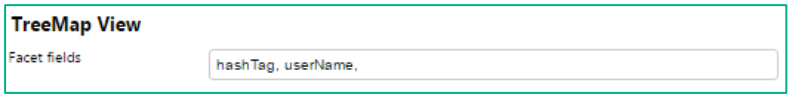
Field name: field on which calendar display will be based

Start date: date the calendar starts

End date: date the calendar ends

TreeMap view

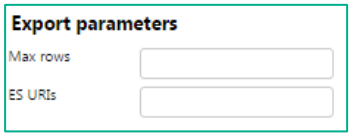
Results displayed in a TreeMap, accessible in the search interface using the [List] > [TreeMap] menu.



TreeMap view

Facet fields: list of fields which will be displayed in the TreeMap.

Export parameters



Export parameters

Max rows: number of rows to export even if there are a large number of results.

ES URIs: everteam URIs taken into account in the export (e.g. ESDB, http, etc.)

Sort criteria

Sort criteria

Default sort fields: list of fields for the default sort (the separator is the comma ","). It is recommended to keep this field empty in order to have the sort by relevance.

Allow field selection: check this box to show the list of possible sort fields in the [Sort] menu in the search interface.

Predefined sort fields: this table allows you to define sorts on several fields. These labels will appear in the [Sort] menu of the search interface. The sort fields are separated by a comma ",".

Actions for selected items

This block allows an administrator with advanced knowledge of everteam and JavaScript to configure actions to carry out on the results list or on a selection, accessible in the [Actions] menu in the search interface.

Actions for selected items

Title: title of the action which will appear in the [Actions] menu in the search interface.

Type: there are two possible types:

- *javascript*: the action is a js function. It can be declared in the .esfct.js file for the index collection.
- *url*: the action is an everteam URL.

Action: this field contains the name of a function (if the type is j*avascript*), or an everteam URL (if not).

Target: display in a new window (*New Window*) or in a dialog box in the same interface (*Dialog Box*).

Declaring facets for ranges of values and groups of queries

This block lets you display data (ranges of values) by facet, on fields containing a large number of values (queries). These facets are indicated in the [Fields] tab.

The idea is to limit the values displayed either by:

- Ranges of values: define the range of values analyzed.
- Queries and query sets: the ranges displayed can each be defined by a specific query (so there is no fixed range). These queries are grouped under the same label, called a query set.

Range of values

Define ranges of values analyzed and the size of the segments in the ranges.

Range of values

Field name	Facet title	Start	End	Gap	Mode	Function (value)
createdAt	Date	NOW/DAY-10DAY	NOW/DAY	+1DAY	bar	function (value) { return " " + value.substr(8, 2) + "/" + value.substr(5, 2); }
userStatusesCount	Statuses Count	0	100000	10000	bar	

Ranges of values

Field name: name of a field in the index used.

Facet title: title of the facet.

Start: start of the range (first value).

End: end of the range (last value).

Gap: size of a segment in the range (step).

Mode: facet display mode.

Function (value): function which decides the Title for each facet value. This may be:

- static text
- a character string returned by a function
- the entire value (if the field is empty)

Example: the standard index field, cs_timestamp which contains the dates the index was updated for each record in the table.

Range of values

Field name	Facet title	Start	End	Gap	Mode	Function (value)
cs_timestamp	Date	2010-10-01T00:00:00Z	NOW/DAY+1MONTH	+1MONTH	bar	function (value) { return "Month " + value.substr(5, 2) + "/" + value.substr(0, 4); }

Example of a range of values facet

The query exposes dates from 1st October 2010 to next month (today plus 1 month), month by month. "January 01/2011").

Start: in ISO format: year-month-dayThours :minutes :secondsZ
(The final Z is for the Zero meridian for universal time).

End: the end date for the range is today's date plus one month (*NOW* = current date and time; *DAY* = the day of the month (0-31) ; *+1MONTH* = the next month).

The segments (Gap) in the date range are for 1 month.

The graph (Mode) is a histogram (*bar*).

The function defines the title of each facet value:
function (value) {return "Month" + value.substr(5,2) + "/" + value.substr(0,4);}

Example 2:

For a field containing prices, it would be possible to define a range of prices (e.g. from 100 € to 1000 €), with prices grouped by 100 €. This would give 10 bars in a histogram.

Query sets and queries

The ranges displayed may each be defined by a specific query, so that ranges are not necessarily identical. These queries are grouped under the title of the graph or list.

You define a query set before detailing the queries for each range. The query set is defined by its name, the display title and the facet type. Query groups can use queries on character strings, dates, numbers or any other data type.

Query sets			*Queries*	
Name	**Title**	**Mode**	**Title**	**Text**
userFollowersCount	Followers	bar	< 10K	userFollowersCount:[0 TO 10000]
			10K-100K	userFollowersCount:[10000 TO 100000]
			100K-1M	userFollowersCount:[100000 TO 1000000]
			> 1M	userFollowersCount:[1000000 TO *]

Set of queries on followers

Query sets			*Queries*	
Name	**Title**	**Mode**	**Title**	**Text**
dates	Dates	pie	Recent	IT_CREATED_dt:[NOW-1YEAR/DAY TO *]]
			Mid-Term	IT_CREATED_dt:[NOW-3YEARS/DAY TO NOW-1YEAR/DAY*]
			Old	IT_CREATED_dt:[* TO NOW-3YEARS/DAY]

Set of queries on date ranges: "dates"

For example, for the selected query set, "dates", a list of queries is displayed, defining three date ranges: old, current, recent.

The Queries/Text column contains the query for filtering on a DATE field, for recent items (this year), current (between 1 and 3 years) and old (more than 3 years old).

Query sets			Queries	
Name	Title	Mode	Title	Text
languages	Language	bar	English	cs_lang:EN
			other	cs_lang:(FR | | AR)

Set of queries on languages

The syntax used in queries on ranges is that recommended for everteam.analytics.

Square brackets mean that boundaries are included; braces would mean that they are excluded.

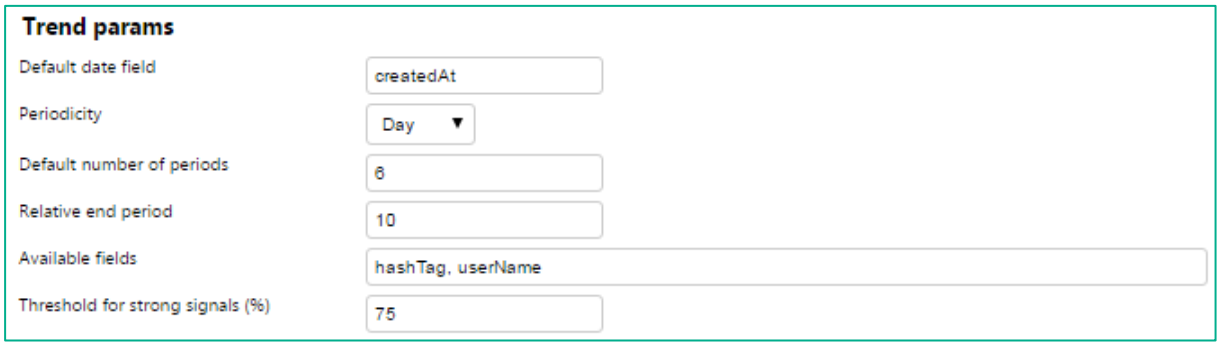
Old: *date:[* TO NOW-3YEARS/DAY]*

Medium: *date:[NOW-3YEARS/DAY TO NOW-1YEAR/DAY]*

Recent: *date:[NOW-1YEAR/DAY to *]*

Trend analysis

This block of parameters makes it possible to analyze trends in a preconfigured period, which can be changed in the search interface, in the [List] > [Trends] menu.



Trend analysis

Date field: *date* field to use for calculating trends.

Periodicity: type of period to use for analyzing changes: *Day*, *Month*, *Year*.

Default number of periods: default number of period to analyze (this can be changed in the search interface, but is always limited to a maximum of 25 periods).

Relative end period: end of the period being analyzed relative to today's date (example: 10 -> today's date minus 10 days).

Available fields: fields to analyze to calculate trends

Threshold for strong signals (%): threshold separating strong signals from weak signals, as a percentage.

Pivots

This function makes it possible to declare two criteria and to produce a double entry table to show how the first criterion is distributed relative to the second.

In the search interface, these tables are obtained via the [List] menu.

Pivots

Max. number of values per facet: 15

Min count/facet: 0

Pivot name	Title	Left facet	Top facet
lang/month	Language / Month	fields.cs_lang	ranges.cs_timestamp
hashtag/person	Hashtag / Person	fields.hashTag	fields.userName
hashtag/followers	Hashtag / Followers	fields.hashTag	querySets.userFollowersCount
statues/hashtags	Statuses / Hashtags	ranges.userStatusesCount	fields.hashTag
followers/hashtags	Followers / Hashtags	querySets.userFollowersCount	fields.hashTag

Cross tabs

The criteria used respect the following constraints:

Max Number of Values/Facet: maximum number of values shown for a facet (maximum number of sectors).

Min count/facet: minimum number of occurrences for a facet value to be displayed.

Table name: system name (unique).

Title: name visible in the [List] menu in the search interface.

Left facet: corresponds to the rows in the generated table

Top facet: corresponds to the columns in the generated table

Example: a table has been set up based on hashtags used by different Twitter users.

Pivot name	Title	Left facet	Top facet
lang/month	Language / Month	fields.cs_lang	ranges.cs_timestamp
hashtag/person	Hashtag / Person	fields.hashTag	fields.userName

Cross tabs on hashtags by user

In the search interface, the [Hashtag/Person] menu will appear in the [List] menu. The table looks like this:

1135138 Result(s) (52 ms) Number of lines/page 20

Left Facet Hashtag Top Facet User Name Display Display Chart

Le Sphinx	20 Minutes	Les Echos	Le Monde	Marguerite Perle	Wintsor	Libération	L'Express	Paris	Le Journal du Geek	Mediapart	OGÉ Stéphanie	Cédric Le Clech	Cyrille Trignac	La relou	Le Sphinx	Total	Chart
Bruxelles	37	75	190	91			49	61					45			30527	
Zapping		789			608											10869	
PanamaPapers			208	71				67					32			10137	
LoiTravail	53	30	118	39		57		48		97				29		9636	
Brussels			69	19				20					20			5610	
NuitDebout	45		42	27		15		25		65		12				5172	
abonnés			153	103				105					15	24		4018	
panamapapers			25							12			11		6	3044	
MayThe4thBeWithYou									23							2976	
StarWarsDay			3						5							2423	
EgyptAir	8	14	38	14			7	11								1763	
Syrie5ans			71	11		23		15					7			1681	
ONPC										17						1480	
Cannes2016	46	9	70	24		10		22					10			1341	
VivreAvec			64	26				21					8			1152	
Others	*14665*	*13506*	*8899*	*8159*	*5197*	*5646*	*5303*	*3990*	*3553*	*2828*	*2893*	*2453*	*1552*	*1514*	*1301*	***81459***	
Total	**14854**	**14423**	**9950**	**8584**	**5805**	**5751**	**5359**	**4385**	**3581**	**3019**	**2893**	**2465**	**1700**	**1567**	**1307**	**85643**	

Table display

Static filters

The results list is displayed by a dynamic calculation using a single criterion.

Each interval of the filter is defined.

Static filters

Field name	Title	Widget Type	Min	Max
cs_timestamp	Date	time		

Query title	Criteria
Last 3 Months	[NOW-3MONTH TO NOW]
Last Month	[NOW-1MONTH TO NOW]
Last Week	[NOW-7DAYS TO NOW]
Last Day	[NOW-1DAY TO NOW]

Static filters

Field name: field to be used by this filter.

Title: keyword containing the title of the filter, which will appear in the search interface.

Widget type: *time* to build a filter with several time ranges, *range* to build a filter with a lower and upper value. How the table is displayed on the right depends on this choice.

Min: minimum value (range type).

Max: maximum value (range type).

Query title: keyword containing the title of the query, which will appear in the search interface.

Criteria: query using everteam.analytics query syntax.

Results may overlap and a record can meet several criteria.

Filter by date

Example: the last three months, last month, last week, last day.

Static filters

Field name	Title	Widget Type	Min	Max
cs_timestamp	Date	time		

Query title	Criteria
Last 3 Months	[NOW-3MONTH TO NOW]
Last Month	[NOW-1MONTH TO NOW]
Last Week	[NOW-7DAYS TO NOW]
Last Day	[NOW-1DAY TO NOW]

Filter on a date

Displaying a filter on a date in the search interface

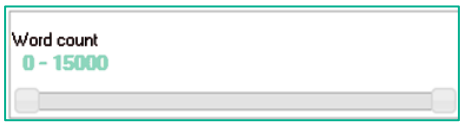
Last month selected (extract)

To come back to the previous display, you must simply delete the corresponding filter in the dynamic filter block.

Filter on a numeric field (range)

This filter allows items to be analyzed based on a numeric field (examples: price spread, range of property sizes in a development, etc.).

Example with TypeWidget = *range*, Min = *0* et Max = *15000* :



Displaying a static filter

Another example would be a filter by file size.

Motion chart

This is a chart showing how data has evolved over several periods.

Motion Charts

Name	Title	x Axis Field	y Axis Field	Main Facet	Date Field	Nb Of Periods	Periodicity
motion1	Motion Chart	userFollowersCount	userFriendsCount	userName	createdAt	10	DAY

Motion Chart

Dynamic tree

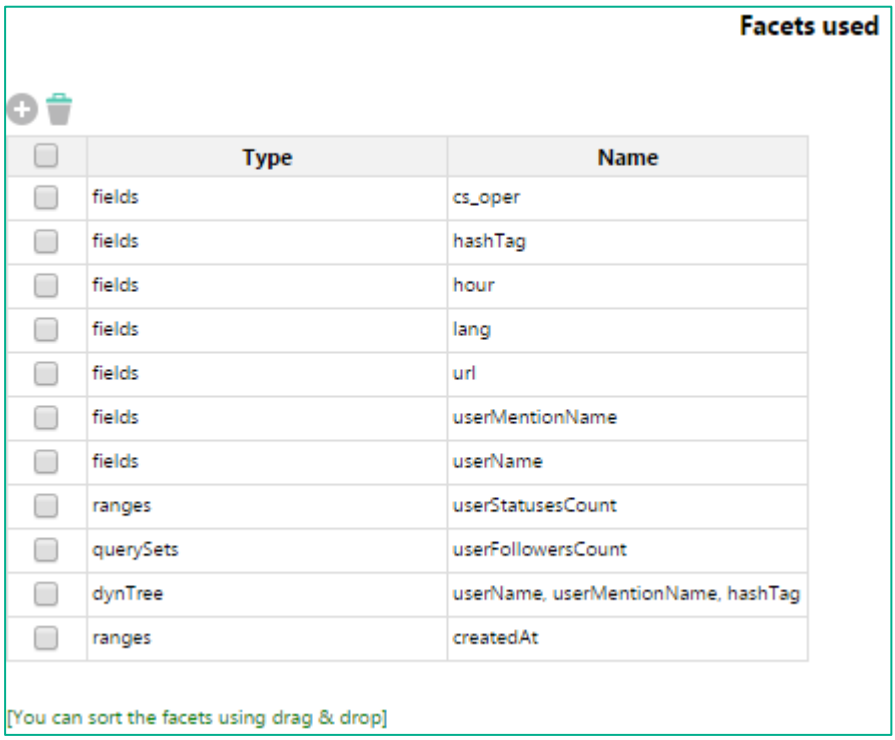
This shows hierarchies with three levels (facet/sub-facet/sub-sub-facet).

Dynamic tree

4.2.5. Facets used

List of facets used in the application, including:

- facets identified in the [Fields] tab, *Facet mode* column,
- facets declared in the [Application params] tab.



Facets used

Remember that there are three types of facet:

- fields: these facets, entered in the table in the [Fields] tab, are enabled or disabled in the [Facets used] tab.
- ranges: for a given field, dynamic ranges are defined for displaying the facet. This type is for any date or numeric field (price, ages, etc.). This is configured in the [Application params] tab, Range of values block.
- querySets: these facets are the result of a query set. They are also configured in the [Application params] tab, Query set and Queries block.

A click on the [Add missing facets] tab adds facets to the list, which are checked in the [Fields] tab or declared in the [Application params] tab but are missing from this table.

In the example below, three filters have been added:

The same field may be affected by several facets (e.g. Ranges of values and Query set).