



Test Bench

User Guide

Cogito Test Bench
User Guide



Authors	Cogito Labs
Document version	Version 1.0
Published date (MM/DD/YYYY)	31/07/2015
Document visibility	CONFIDENTIAL

Cogito and Sensigrafo are registered trademark of Expert System S.p.A.

All other trademarks, service marks and company names in this document or website are properties of their respective owners.

Notice: This document is for informational purposes only, and does not set forth any warranty, expressed or implied, concerning any product, component or service offered or to be offered by Expert System

Table of Contents

Table of Contents	iii
Before You Start	1
Cogito Overview	2
What is Cogito?	2
About Semantic Technology	2
What Does Cogito Do?	2
Test Bench Overview	3
What Is Test Bench?	3
Accessing Test Bench	4
Opening And Closing Test Bench	5
How to Log into Test Bench for the First Time	5
How to Open Test Bench	5
How to Return to Cogito Studio	5
Working With Libraries	6
What Is a Test Bench Library?	6
How to Add a Library in Test Bench	6
What Is Document Target Association?	7
How to Open a Library	7
How to Delete a Library	8
How to Upload a Set of Documents into a Library	8
Documents Targeting	9
What Is the Document Target?	9
What Is the Taxonomy Tree?	9
Understanding the Information Represented in the Taxonomy Tree	10
How to Add a Target to a Library Documents in Test Bench	10
How to Modify a Document Target in Test Bench	11
How to Remove Selected Targets	11
How to Remove a Target from a Document	12
Documents Targeting in Analytics	13
How to Open Analytics from a Test Bench Analysis	13
How to Assign a Target to a Document	13
Analyzing Documents	13
How to Analyze a Set of Documents	14
How to Stop a Test Bench Analysis	14
How to Export Your Analysis Results	14
Testing and Validating Rules	15
What Is a Linguistic Resource?	15
What Is the Linguistic Resource Option?	15
What Are the Automatically Generated Rules?	15
How to Generate Automatic Rules	16
How to Deploy the Updated Linguistic Rules	16
Test Bench Key Concepts	17
Analyzing Your Results	17

What Are the Positives and Negatives Predictive Values?	17
What Is a True Positive?	18
What Is a False Positive?	18
What Is a True Negative?	18
What Is a False Negative?	18
What Are the Precision and Recall Values?	19
What Is the F-measure?	20
What Are the Micro-Average and the Macro-Average?	21
What Is the Micro-Average?	21
What Is the Macro-Average?	22
Viewing the Test Bench Analysis Results	23
Viewing the Analysis Results	23
How to View the Analysis Results	23
How to View the Analysis Results for a Category	24
What Is the Analysis Dashboard?	25
Which Data Is Displayed in the Information Section?	25
Which Data Is Displayed in the Documents Section?	26
Which Data Are Displayed in the Micro/Macro-Averaged Sections?	26
Understanding the Statistical Information for a Single Analysis on the Dashboard	27
How to Compare Two Analyses	27
How to View Documents Comparison for a Category of Two Different Analyses	28
Searching in Test Bench	29
Standard Search and Advanced Search	29
What Is the Standard Search?	29
How to Search for Documents	29
How to Search Inside a Document	30
Searching Using Analytics	31
Accessing Analytics	31
How to Access Analytics from a Browser Window	31
How to Access Analytics in Test Bench	32
Which Are the Analytics Search and Display Tools?	33
Searching in analytics	34
What Are the Semantic Item Types?	34
How to Change the Semantic Item Type	34
What Is a Boolean Operator?	35
How to Search in Analytics	35
Editing a Search in Analytics	36
Which Are the Analytics Editing Tools?	36
Searching Sensigrafo from Analytics	37
Which Data Is Displayed in Sensigrafo?	37
How to Search Using the Sensigrafo Search	39
How to Navigate in Sensigrafo	40
How to Search Using Sensigrafo Expanded Concepts	40
Searching with the Analytics Advanced Search Options	41
What Are the Analytics Search Options?	41
How to Display the Advanced Search Options	41
How to Change Search Scope – Context	42
How to Change the Documents Display Order – Order by	42

How to Filter Documents Based on Target and Category Availability – Target Filters	42
How to Search in a Document Section – Sections.....	43
What Are the Visual Search Refinement Elements?.....	44
Using the Search Refinement Views.....	45
Refining Your Search from the Documents View	45
What Is the Document Preview?.....	45
Which are the Document Actions?.....	46
What Is the Full-Text Document Window?	46
How to Refine a Search from the Document View?.....	47
Refining Your Search from the Analysis List View.....	48
What Is the Frequency Item Type?.....	49
What Is the Analysis List Search Box?	49
How to Search Using an Analysis List	49
What Is the Analysis List Menu?.....	50
What Is the Exclusion List?	51
How to Add an Item to the Exclusion List	51
Refining Your Search from the Sentence View	52
How to Switch to Sentence View	52
How to See the Analysis List Related to a Selected Sentence.....	52
Refining Your Search from the Sao-Analysis View	53
What Is Displayed in the SAO List View?	54
What Is Displayed in the SAO Triples View?.....	54
How to Refine a Search Using the SAO Lists	54
How to Export the SAO List	54
Refining Your Search from the Bar Chart View	55
How to Open the Bar Chart View	55
Refining Your Search from the Pie Chart View.....	56
What Information Is Displayed in the Pie Chart when Opening Analytics?	56
How to Open the Pie Chart View.....	56
Refining Your Search from the Tag Cloud View	58
How to Open the Tag Cloud View	58
How to Select a Category to Display in Tag Cloud.....	58
Index	60
Contacts.....	63

Before You Start

Introduction

Test Bench supports rules development in Cogito Studio.

This user guide explains how to use Test Bench for testing, viewing and validating Cogito rules.

Audience

This guide is intended for Cogito Studio users.

The document starts with an overview of the Cogito family of products, and continues to show how to use Test Bench.

Prerequisites

You need Cogito Studio advanced knowledge to benefit from this user guide.

Objectives

After you have read this guide, you will be able to:

- Create a library
- Upload a set of documents into a library
- Target documents within a library
- Create an analysis
- View existing rules
- Interpret the analysis results in terms of rules
- Generate automatic rules
- Test rules comparing different analyses
- Export automatic rules for further inspection
- Search for documents and items inside a documents
- Search for items and documents inside a library or an analysis

Cogito Overview

WHAT IS COGITO?

Cogito is a family of products that applies semantic technology to enterprise information to optimize searches and automate document categorization.

You can integrate Cogito with enterprise applications such as SharePoint to extend their categorization and extraction capabilities. If needed, using Cogito Studio Development Environment, you can write your own categorization and extraction custom rules.

About Semantic Technology

Semantic technology is based on linguistic rules to assign meaning to content.

Semantic software applies this technology to automatically attach - through built-in rules- meaning to content. It analyzes text, resolving language ambiguities and contextualizing information to create semantic models. Cogito uses semantic networks, which are a representation of knowledge and relationships between concepts (syncons).

Semantic technology addresses the need to manage the exponential growth and the complexity of today's information. It finds its applications in text mining, big data management, intelligence and all the areas that require data categorization and strategic decision making.

The advantage of semantic technology over traditional technologies is in that it provides a higher quality in terms of precision and recall of processed information.

What Does Cogito Do?

Cogito relies on deep semantic analysis and the rich semantic network to categorize and tag automatically all of the existing documents in your organization.

Cogito analyzes text in input against its reference system to retrieve, disambiguate, categorize and extract content. This analysis helps you to discover hidden relationships, trends, patterns and events.

Cogito capabilities are reflected in the vertical solutions Cogito Categorizer and Cogito Discover.

Cogito Categorizer

Cogito Categorizer classifies incoming documents content based on pre-built or customized taxonomies.

Cogito Discover

Cogito Discover based on pre-built or custom entity definitions automatically tags, extracts and normalizes the most relevant data from both structured and unstructured content, contextualizing the data and turning it into meaningful knowledge.

Test Bench Overview

Test Bench helps you to optimize Cogito Studio semantic technology.

You create categorization and extraction rules for your project in Cogito Studio; rules that you can review and fine-tune with the Test Bench tool.

With Test Bench, you can also generate automatic rules that you can import into Cogito Studio to inspect them and modify them.

Test Bench allows you to:

- Work on a chosen set of documents through the Test Bench libraries
- Manage documents desired target categories
- Analyze a set of documents
- View the precision and recall values for each set of documents analysis
- Enhance documents precision and recall by means of modifying the current rules

WHAT IS TEST BENCH?

Test Bench is a tool designed to allow you to view, test and modify categorization and extraction rules on a set of documents.

The rules that you are going to modify, have been created in Cogito Studio.

Test Bench only allows you to modify existing Cogito Studio rules or to generate automatic rules based on your current set of rules.

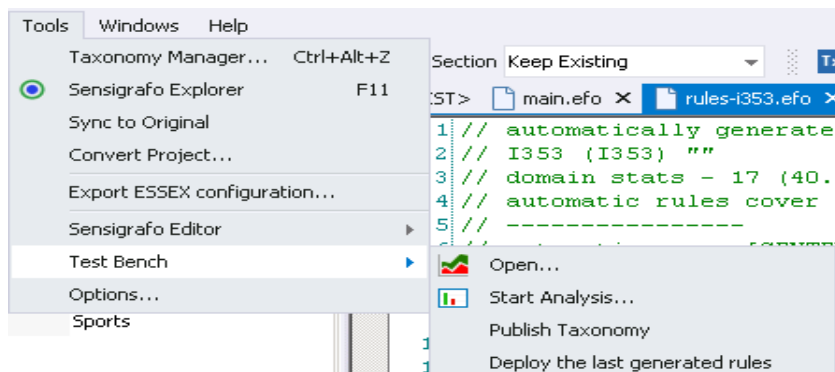
You can organize the documents that you want to examine in Test Bench libraries, allowing you to target a specific set of documents for analysis.

Accessing Test Bench

You access Test Bench from Cogito Studio through the Tools menu, where you select the Test Bench menu.

From the Test Bench menu, you can:

- Open Test Bench
- Jump to the analysis dialog box and start analyzing a library
- Deploy the latest Test Bench generated rules without opening Test Bench
- Export your taxonomy tree (Publish taxonomy).



Commented [A1]: As per conversation with GS, perhaps we could:
a) Change the label name to Export taxonomy tree
b) Put some kind of visual feedback, because at the moment when you click on Publish taxonomy, you only get a message that the taxonomy was published successfully. Where, which one etc does not say

OPENING AND CLOSING TEST BENCH

How to Log into Test Bench for the First Time

1. In Cogito Studio, click on Tools.
2. Click on Test Bench.
A dialog box asking for your credentials is displayed.
3. Enter your user name and password.
4. Enter the Team server name.
5. Select the Use the same address for all the Cogito Studio Services check box.
Select this option, if the Test Bench server is installed on the same server where the Repository component is.
OR
Enter the Test Bench server IP address.
You enter the Test Bench server name if Test Bench and Cogito Studio are on separate servers.
6. Click on Test Connection to ensure that the connection is successful.

 If the connection is not successful, contact the server administrator who installed Cogito Studio Services.

How to Open Test Bench

1. In Cogito Studio select the project whose rules you want to view and validate.

File-> Open project-> Select project

2. In Cogito Studio, click on Tools.
3. Click on Test Bench and select Open.
Test Bench main window is now open.

How to Return to Cogito Studio

1. In Test Bench, click on the Exit icon .
2. Click on Exit.
You are now returned to the Cogito Studio main window.

Working With Libraries

When you have identified a Cogito Studio project, or an area of a project that requires testing, you can create a library in Test Bench.

A library allows you to work on a restricted set of documents.

You can create as many libraries as you require.

In each library, you upload the set of documents that you want to analyze.

Using a project-specific library allows you to pinpoint issues with the existing rules and to fine-tune them.

What Is a Test Bench Library?

A Test Bench library is a set of documents that you upload in Test Bench for testing purposes.

You can create as many libraries as your Cogito Studio project requires.

When you create a library, it is displayed in the Test Bench library pane.

From the Library pane, you can:

- Add libraries
- Remove libraries
- Upload documents into a selected library
- Analyze the selected library documents
- Generate automatic rules for the selected library

How to Add a Library in Test Bench

1. In Test Bench, click on the  icon.

The New library dialog box is displayed.

2. Enter the new library name.

3. Click on Ok.

The Upload documents dialog box for the library is displayed.

4. Browse to the folder where your documents are stored.

5. Click on Next.

6. Select the Target association

Select this option if your documents are organized in sub-folders whose name matches your taxonomy.

7. Click on Start to upload the documents into the library.

For more information, see What Is Document Target Association?

What Is Document Target Association?

During library creation, the target association is a functionality that allows you to define the target for your documents during documents upload.

You can associate a set of documents either to the source folder name or the document folder name.

Note that if you choose the document folder name, the folder name must correspond to a category in the taxonomy tree.

If you choose the option **Use the documents folder name**, and the folder name is not in the taxonomy, the Target categories summary displays the message **Not in current taxonomy**.

Last modified: Monday, December 08, 2014 2:47:40 PM

Size: 3.09 KB [3,150 characters]



You can view the target association when you open a document.

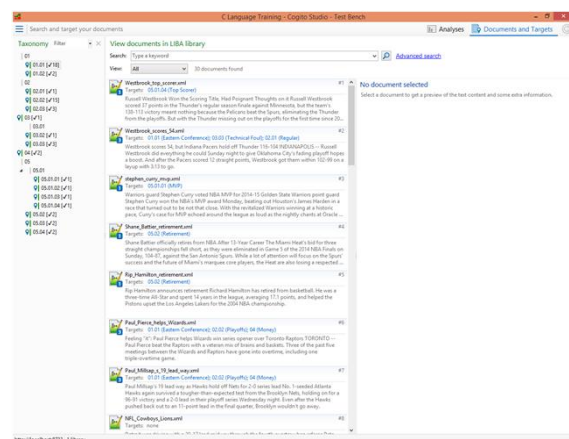
How to Open a Library

1. In Test Bench, double-click on the library that you want to open.

The current taxonomy and the list of documents in the library are displayed.

You can add a filter to choose the documents that you want to be displayed.

 The target set shows the target distribution of documents, whilst the analysis correlation shows the variation in target distribution after analyzing the set of documents.



How to Delete a Library

1. In Test Bench, select the library that you want to delete.
2. Click on the  icon.
The Delete dialog box is displayed asking you to confirm deletion.
3. Click on Yes to delete the library.
OR
Right-click on the selected library.
4. Click on Delete.
The Delete dialog box is displayed asking you to confirm deletion.

How to Upload a Set of Documents into a Library

1. Click on the upload documents icon .
The Upload documents dialog box is displayed.
2. Click on Browse to navigate to the documents folder that you want to upload.
In the folder selection dialog box, click on Select Folder.

Commented [A2]: It would be useful for the user to have a list of accepted file formats and max upload size



*You can select more than one folder using the **shift-key***

3. From the Upload dialog box, click on Next to confirm that you want to upload the folder into the library.
From the Upload documents dialog box, you can also remove one or more folders.
4. Choose the target **association**.
You can choose between Use source folder name and Use documents folder name.
For more information, see What Is Document Target Association?
5. In the Custom XML extraction pane, specify the XML sections that you want to extract.
This option will be used in the search, allowing you to choose the section of the document that you want to analyze.
6. Click on Start.
The documents are first uploaded and then indexed.

Commented [A3]: I would change the name of the labels



Indexing is the process, through which all the documents elements are parsed and stored to facilitate information retrieval. Elements are grouped into categories such as main lemmas, proper nouns, people, places URL, email and more.

DOCUMENTS TARGETING

What Is the Document Target?

A target is a classification label that you attach to a document for indexing purposes.

The target may correspond to a category in the taxonomy of the Cogito Studio project; it is a category where the document should be categorized into.

You allocate a target for a document or a set of documents.

When you give a target to a document, you define an additional rule that can help you to refine the precision and recall values.

Each document displays the target category or categories it belongs to, in the document information.

If the target does not match a category in the taxonomy, the document information displays the message **Not in current taxonomy**.

For more information, see:

[What Are the Precision and Recall Values?](#)

[What Are the Precision and Recall Values?](#)

What Is the Taxonomy Tree?

The taxonomy is the structure that lies beneath the categorization process.

The structure contains the categories against which documents are classified by means of categorization rules.

The taxonomy has a hierarchical structure reflecting the organization of the information within a knowledge domain; flat taxonomies (list of non interdependent items) are also possible.

Understanding the Information Represented in the Taxonomy Tree

The Documents and Targets view for a library or for an analysis provide additional pieces of information near each node of the taxonomy tree.

This information is given in the form of icons and numbers.

Each taxonomy node is indicated by a vertical bar; the color of the bar gives additional information:

- A gray bar  indicates that the category has no targets
- A yellow bar  indicates that the node has been modified at document level.
For example, when a user has made a change such as adding or removing a target.
- A green bar  indicates that one or more documents have this category as a target
- A pin  indicates that there are documents that have that category as a target.
If there is a pin, the vertical bar is green, whilst changes in the target are indicated by the bar turning yellow.
- A check with a number ³² shows the number of targeted documents with this category.
- A star with a number ³³ identifies the number of categorized documents with this category.

Commented [A4]: I would redo this section completely.
I would put icons in a table with a description

 The number of targeted documents can differ from the number of categorized documents inside the same taxonomy node.
The taxonomy tree displays a star just after the analysis has finished.

How to Add a Target to a Library Documents in Test Bench

1. Double-click to open the library with the documents whose target you want to set.

The Documents and Targets window is displayed.

 You can search for documents through the search box
OR

You can use the View filter and choose from:

- All
 - Without target
 - Target in selected categories
2. Select the document that you want to target.
You can select multiple documents using the ctrl-key.
 3. Drag and drop your selection into a category in the Taxonomy panel.

How to Modify a Document Target in Test Bench

1. Double-click to open the library with the documents whose target you want to modify.

The Documents and Targets window is displayed.



You can search for documents through the search box

OR

You can use the View filter and choose from:

- All
- Without target
- Target in selected categories

2. Select the document whose target you want to modify.

You can select multiple documents using the ctrl-key.

3. Drag and drop your selection into the required category, in the Taxonomy panel.

How to Remove Selected Targets

1. Double-click to open the library where you want to remove the targets.

The Documents and Targets window is displayed.



You can search for documents through the search box

OR

You can use the View filter and choose from:

- All
- Without target
- Target in selected categories

2. Remove selected targets.

3. Right-click on the target category that you want to remove.

4. Click on Remove from target set.

Commented [A5]: Perhaps we need a new style for bullets inside a note. Not frequent but useful

Commented [A6]: For Giovanni to review

How to Remove a Target from a Document

1. Double-click to open the library with the documents whose target you want to remove.

The Documents and Targets window is displayed.



You can search for documents through the search box

OR

You can use the View filter and choose from:

- All
- Without target
- Target in selected categories

2. Open the document.

Inside the document, the Target categories summary displays the current categories.

3. Right-click on the target category that you want to remove from the document.
4. Click on Remove from target set.



To verify that the document has been removed from the target category, look at the documents count for the category in the Taxonomy panel.

Commented [A7]: Perhaps we need a new style for bullets inside a note. Not frequent but useful

Documents Targeting in Analytics

When you open Analytics in a browser window, the taxonomy tree is not displayed.

To assign a target to your documents in the Analytics results page, you need to open a Test Bench analysis where the taxonomy is displayed.

With both the Analytics results page and the Test Bench analysis page open, you can drag and drop a document from the Analytics results page into the Test Bench analysis taxonomy.

For more information on searching in Analytics, see [Searching Using Analytics](#).

How to Open Analytics from a Test Bench Analysis

1. Right-click on an analysis.
2. Select Browse with Analytics.

A list of documents for the analysis is displayed in the Analytics browser.

How to Assign a Target to a Document



For this procedure, you need to open a Test Bench window and an Analytics window.

1. In Test Bench, double-click on an analysis.
2. Open the Analytics browser.

For more information on opening Analytics, see [How to Open Analytics from a Test Bench](#)

3. In the Analytics search results page, select the document to which you want to assign a target.
You can select multiple documents using the **ctrl** –key.
4. Drag and drop the selected document into a category of the Test Bench analysis that you have selected.

Analyzing Documents

Test Bench allows you to analyze sets of documents.

When you analyze a set of documents, you can choose between two options:

- Selecting one of the existing linguistic resources.

For more information, see

How to Analyze a Set of Documents

- Generating automatic rules specific to the analysis.

For more information, see [How to Generate Automatic Rules](#)

The analyses results display statistical information on your categorization rules, allowing you to identify target categories that require attention.

You can interrupt a library analysis at any time, and you can also export your analysis results.

How to Analyze a Set of Documents

1. Right-click on the library that you want to analyze.
2. Select Analyze documents.
The New analysis dialog box is displayed
3. Enter a name for the new set of rules.
Spaces and special characters are not allowed.
4. In the Note box, enter any comment that might be useful as a future reference.
5. Confirm or change the library where you want to run the analysis on from the drop down list.
6. Select the linguistic resource that you want to use for the document library analysis.
7. Click on Start.

How to Stop a Test Bench Analysis

1. Right-click on the library where the Test Bench analysis is running.
2. Click on Abort.

The Abort analysis dialog box is displayed.

Commented [A8]: In the UI I would name this dialog box, and call it Abort Analysis

How to Export Your Analysis Results

1. Right-click on the analysis whose results you want to export.
The Windows Explorer dialog box is displayed.
2. Enter the export file name.
3. Select where you want to save the export file.
4. Click on Save.

The file is saved as a .csv file.



The .csv file format allows you to import the analysis results data into Excel spreadsheets.

Testing and Validating Rules

WHAT IS A LINGUISTIC RESOURCE?

ESSEX is Expert System semantic engine that allows text analysis and text disambiguation.

You can deploy more than one ESSEX instance, each with specific extraction and categorization rules.

The linguistic resource represents the ESSEX instance that you want to use.

What Is the Linguistic Resource Option?

The linguistic resource exports the ESSEX service from Cogito Studio into Test Bench.

You can have more than one linguistic resource, depending on your projects and the number of rules updates that you have made.

The Linguistic resources tab allows you to choose the linguistic resource that you want to test your documents against.

For each linguistic resource, you can choose from:

- Download
This option downloads the ESSEX package.
- Publish
This option uses the ESSEX linguistic resource that was previously exported from Cogito Studio.
- Deploy rules to Cogito Studio
This option uploads the automatic generated rules to Cogito Studio.

Commented [A9]: Is this correct?

What Are the Automatically Generated Rules?

Test Bench allows you to analyze document libraries using the Generate automatic rules option.

This option is based on your current taxonomy, and provides a starting point for your categorization strategy.

Commented [A10]: I think this section needs expanding, need help!!!

You can view the new set of rules in Cogito Studio by selecting Deploy rules to Cogito Studio.

For more information, see:

[How to Generate Automatic Rules](#)

[How to Deploy the Updated Linguistic Rules](#)

How to Generate Automatic Rules

1. Select the library for which you want to generate rules.
2. Click on Generate automatic rules.

The Generate automatic rules dialog box is displayed.

3. Enter a name for the new set of rules.

In the Notes box, you can enter any comment that might be useful as a future reference.



Kickoff

Notes

Rules generation summary

Information 

4. From the drop-down list, confirm or change the library name where you want to apply the rules.

How to Deploy the Updated Linguistic Rules

1. Right-click on the analysis results whose rules you modified.

For more information, see [How to Generate Automatic Rules](#)

2. Select Deploy Rules to Cogito Studio.

The new rules are displayed in Cogito Studio in XML format, where you can view them and further modify them.



If Deploy Rules to Cogito Studio is grayed out, the rules are already deployed.

Test Bench Key Concepts

Test Bench uses standard statistical concepts to evaluate search results.

The concepts that Test Bench uses are:

- Predictive values known as false positive, true positive, false negative and true negative
- Precision and Recall
- Macro and micro average
- F-measure

ANALYZING YOUR RESULTS

When you analyze a set of documents, an array of statistical values is displayed in the Analysis dashboard. The values indicate the accuracy of your rules and categorization strategy.

For example, the precision or the recall values might increase, or decrease, suggesting that you may have to do some tweaking.

The analysis results change when you:

- Modify the target for a set of documents
- Generate new rules
- Use a different linguistic resource.

What Are the Positives and Negatives Predictive Values?

The Test Bench analysis shows how documents are categorized according to your rules.

Each analysis highlights the documents categorization, that is whether documents fall into the correct category.

Despite the best efforts, there will still be documents that are incorrectly categorized or that should fall into a category but they don't. In addition, after the analysis process has taken place, a document can still result as uncategorized.

If we focus on a specific category, we can check the documents that have been categorized in that category against their target.

In statistical terms, we refer to these predictive values as:

- True Positive
- False Positive
- False Negative
- True Negative

What Is a True Positive?

A document flagged as true positive is a document targeted within the examined category.

The document was expected in that category and shows up in the expected category at the end of the analysis process.

What Is a False Positive?

A document flagged as false positive is a document targeted with a different category than the examined one.

The document was not expected to be found in the examined category, but it shows up in the category at the end of the analysis process, even though it has a different target.

What Is a True Negative?

A document flagged as a true negative is a document which was not expected to be found in a category, and it is not found in that category – as expected.

What Is a False Negative?

A document flagged as false negative is a document targeted with the examined category, which does not appear in the category. The document was expected to be found in the examined category, but it doesn't show up in the category.

What Are the Precision and Recall Values?

When you analyze a set of documents, Test Bench displays statistical information about the documents categorization and retrieval; these values are precision and recall.

Precision and recall are the basic statistics measures for understanding relevancy in document discovery and evaluating search strategies.

Both precision and recall are calculated relying on the positive and negative predictive values.

EXAMPLE

For a given category C , let us define:

- **TP (C)** the number of true positives, that is the number of documents that are both categorized in C and that have C as a target.
- **FP (C)** the number of false positives, that is the number of documents that are categorized in C despite not having C as a target.
- **FN (C)** the number of false negatives, that is the number of documents that are not categorized in C despite having C as a target.

Precision for a given category is the portion of categorized documents that are targeted in the category; it identifies the number of correct results against the target, and can be expressed as:

$$\text{Precision}(C) = \frac{TP(C)}{TP(C) + FP(C)}$$

The recall is the portion of both correct and not correct documents that are categorized in the category. The recall identifies the number of results against the target, and can be expressed as:

$$\text{Recall}(C) = \frac{TP(C)}{TP(C) + FN(C)}$$

The recall refers to the quantity of documents that are categorized in a given category, whilst precision indicates the accuracy.

Precision and recall are inversely related values, that is, the higher the precision value is, the lower the number of categorized results and viceversa.

The aim is to achieve the highest possible value for both precision and for recall.

What Is the F-measure?

The F-measure combines the precision value and the recall value.

If accuracy is not the goal, the F-measure provides information on retrieval as a balance between precision and recall since it ignores the true negatives.

You cannot achieve a high F-measure, if either precision or recall have low values.

$$\text{F-measure}(C) = \frac{2}{\frac{1}{\text{Precision}(C)} + \frac{1}{\text{Recall}(C)}}$$

Or, analogously;

$$\text{F-measure}(C) = 2 \cdot \frac{\text{Precision}(C) \cdot \text{Recall}(C)}{\text{Precision}(C) + \text{Recall}(C)}$$

Commented [A11]: Cannot get rid of the line

What Are the Micro-Average and the Macro-Average?

Precision and recall are defined for a single category. The question is how to get aggregate information to evaluate them for the overall classification.

Two methods are available to calculate the overall Precision and Recall, depending on the method used for averaging them over the whole taxonomy:

The **micro-averaged precision and recall** are calculated as if there were only one single large category, thus summing all the TP, FP and FN for all categories and then calculating the overall precision and recall.

The **macro-averaged precision and recall** are calculated by taking the average values of precision and recall over all categories.

WHAT IS THE MICRO-AVERAGE?

The micro-average calculates the precision and recall values by summing all the true positives, false positives and false negatives for each individual category.

$$\text{Micro-Averaged Precision} = \frac{\sum_{i=1}^N TP(c_i)}{\sum_{i=1}^N TP(c_i) + \sum_{i=1}^N FP(c_i)}$$

$$\text{Micro-Averaged Recall} = \frac{\sum_{i=1}^N TP(c_i)}{\sum_{i=1}^N TP(c_i) + \sum_{i=1}^N FN(c_i)}$$

$$\text{Micro-Averaged F-measure} = \frac{2}{\frac{1}{\text{Micro-Averaged Precision}} + \frac{1}{\text{Micro-Averaged Recall}}}$$

WHAT IS THE MACRO-AVERAGE?

The macro-average is the arithmetic mean of precision and recall of each individual category.

$$\text{Macro-Averaged Precision} = \frac{\sum_{i=1}^N \text{Precision}(c_i)}{N}$$

$$\text{Macro-Averaged Recall} = \frac{\sum_{i=1}^N \text{Recall}(c_i)}{N}$$

$$\text{Macro-Averaged F-measure} = \frac{2}{\frac{1}{\text{Macro-Averaged Precision}} + \frac{1}{\text{Macro-Averaged Recall}}}$$

Viewing the Test Bench Analysis Results

In Test Bench, you can view the analysis results both in a table and on the dashboard.

The table is a clickable list of documents that helps you to view variations in documents categorization, at document level.

The dashboard allows you to view statistical information on the analysis results, at category level.

Viewing the Analysis Results

The analysis results consist of a clickable table that displays:

- The list of library documents that are contained in the current taxonomy.
The number of documents in each category.
- Statistical information such as Precision and Recall.

For more information, see [Analyzing Your Results](#).

When you click on a category in the table, the list of all documents for that category is displayed.

You can filter the documents by the positive and negative predictive values to view the documents categorization.

When you filter using the predictive values, you may obtain three types of documents:



True positive



False positive



False negative

This value shows that a document has not been categorized in a category, but it was expected to be.

For more information, see [What Are the Positives and Negatives Predictive Values?](#)

How to View the Analysis Results

1. Right-click on the analysis whose results you want to view.
2. Select View Results.

The results page displays statistical information for each category.

3. Click on a category that you want to explore.

The Documents and Targets view displays the list of documents for that category, and the taxonomy tree shows additional information for the category.



You can further refine the list of documents.

For more information, see [How to View the Analysis Results for a Category](#)

How to View the Analysis Results for a Category

1. In Test Bench double-click on the analysis whose predictive values you want to view.

The Results table is displayed showing a detailed list of statistical values by category.

2. Click on the category that you want to explore.

The Search and target your documents page is displayed. By default, this page shows all of the documents in the selected category.

3. You can choose from different views and filter by the predictive value.

The views that you can apply are:

- All
- False Negatives
- False Positives
- True Positives
- Targets
- Results

4. From the View drop-down list, select the value that you want to examine.

EXAMPLE

If you select True positives, the list displays all the documents that are true positives, that is documents, which correctly display in the assigned categorization.

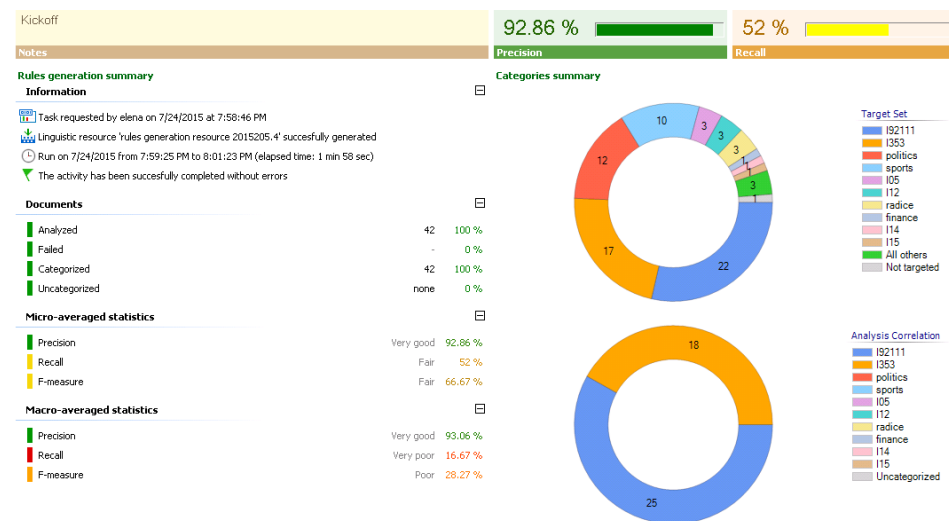
For more information, see **Error! Reference source not found.**

Commented [A12]: There are two pages with the same name Search and Target your documents. However, they are different pages and they should have a different name. (from one you can filter TP, FP etc, from the other not possible)

WHAT IS THE ANALYSIS DASHBOARD?

The Test Bench dashboard is a visual tool that provides a statistical representation for your analysis results.

When you select more than one analysis, the dashboard displays four sections of comparative statistical information that allows you to see variations for the selected analyses.



Which Data Is Displayed in the Information Section?

This section provides log information about the library that has been analyzed.

Collapse this section to view:

- Analysis requestor
- Time stamp
- Linguistic resource in use
- Analysis outcome

Which Data Is Displayed in the Documents Section?

The documents section provides global statistics on the number of documents that has been categorized.

Collapse this section to view:

- Total number of analyzed documents
- Number of categorized documents
- Number of not categorized documents
- Number of failed documents, that is documents that were not detected by the analysis.

A documents analysis may fail due to technical issues such as the documents size.

Which Data Are Displayed in the Micro/Macro-Averaged Sections?

These sections show the calculations for the Micro and Macro averaged data.

For more information on how the micro and macro average are calculated, see:

What Are the Micro-Average and the Macro-Average?

What Is the Micro-Average?

What Is the Macro-Average?

Understanding the Statistical Information for a Single Analysis on the Dashboard

When you compare two analyses, the Test Bench dashboard changes and shows different statistical information related to the comparison.

This information is named Comparison summary and is organized in four **clickable** sections:

- Information
- Documents
- Micro-averaged statistics
- Macro-averaged statistics

The statistical data is also plotted into graphical elements that compare the documents categorization before, and after new rules have been set.

The graphical elements are:

- Two bar indicators for the Precision and Recall values.
The bar indicators can change color based on the value that has been calculated during the analysis process.
- Two donut charts that show the retrieved target set and results, based on the targets.
- A horizontal stacked bar chart with three bars showing:
 - The target set – each color represents a target.
 - The most recent analysis in the selection.
 - The oldest analysis in the selection.When you move the pointer over each bar, you are able to see the number of documents for the pointed category.
 - Two trend charts represent the variation in micro and macro-averaged data (precision, recall and F-measure) between the two selected analysis.



*A horizontal stacked bar chart is a bar chart that places related values near to one another. This chart is used when a category naturally divides into **components**.*

How to Compare Two Analyses

1. Select the analyses that you want to compare using the ctrl-key.

Ensure that you select analyses that are based on the same linguistic resource; if you select analyses that are based on different linguistic resources, Test Bench is unable to compare them.

2. Select Compare results.

The results page displays statistical information for each category.



If you select more than two analyses, Test Bench Overall evaluation compares the most recent analysis with the oldest one, whilst the Quality trend graph displays all the analyses that you have selected.

Commented [A13]: I counted 4 sections and not three as in the revision

Commented [A14]: This is not clear, perhaps it should say sub-categories

Document Category	Icon	Oldest Analysis	Last Analysis	Description
Classification				
		True Positive	True Positive	Document remains good. The document was both categorized and targeted in the examined category; the last analysis shows no changes.
		False Negative	True Positive	Good document entered – it was not at first categorized in the examined category, but now it is and it is targeted with this category.
		False Positive	True Negative	Bad document gone from the examined category – it was categorized in the examined category but it was not targeted with it. Now it is no longer categorized in the examined category nor is targeted with it.
		False Positive	False Positive	Document remains bad – it was categorized in the examined category and not targeted with it and nothing changed in the last analysis.
Correctly categorized		True Negative	False Positive	Bad document entered – it was not categorized in the examined category and not targeted with it but now it is categorized in this category but yet not targeted with it.
		False Positive	True Negative	Good document gone – it was categorized in the examined category and targeted with it but now it is not categorized in the examined category anymore but it kept the same target.
Incorrectly categorized		N/A	N/A	It was categorized in the examined category and it stayed in the same category in the newest analysis. Since it has no target, we cannot identify the document as good or bad.
		N/A	N/A	It was not categorized in the examined category but now it is. Since it has no target, we cannot identify the document as good or bad.
No Target		N/A	N/A	It was categorized in the examined category but now it is not anymore. Since it has no target, we cannot identify the document as good or bad.

Commented [A15]: This is slightly confusing. Perhaps it could be rephrased. Too wordy

How to View Documents Comparison for a Category of Two Different Analyses

1. Select the analyses that you want to compare using the **ctrl**-key.
2. Right-click on one of the selected analyses.
The analysis menu is displayed.
3. Click on Compare results.
The Review or compare your analysis results pane is displayed.
4. Select the category that you want to explore.
The Search and Targets view displays the list of documents belonging to the new analysis for the selected category. An icon next to the category shows the document transition at each analysis.

Commented [A16]: What does it mean?

Commented [A17]: This topic seems unnecessary. Also it's not good practice to describe the UI. This could be explained in the previous procedure

Searching in Test Bench

STANDARD SEARCH AND ADVANCED SEARCH

Test Bench offers a standard search and an advanced search. The advanced search is also known as Analytics.

Both searches allow you to search through a library, and when you open a document, within the document.

The two searches are separate; you can use the advanced search without starting from the standard search.

For more information on the advanced search, see

Searching Using Analytics

What Is the Standard Search?

The Test Bench standard search is a functionality that allows you to search both through a library and inside a document.

You can search by:

- Category
You select a category from the taxonomy tree
- Term
You enter the term in the search box.

The library search results are a clickable list of documents.

To further refine your list of documents, you can search by term or by predictive value, or both.

In addition, when you click on a document in the list, the document opens, and a dedicated search box allows you to search inside the document.

How to Search for Documents

1. Click on the category where you want to search .
The list of all documents in that category is displayed.
2. In the Filter box, enter a term that you want to find in your documents.
3. Click on the View drop-down list and select a value.
A new subset of documents matching your criteria is displayed.



You can just either enter a term or a value. Both steps are not necessary.

How to Search Inside a Document

1. In your search results, click on the document that you want to explore.
The document opens in a separate pane.
2. In the Document **pane** search box, enter the term that you want to find in the document.
3. Click on the Search icon  .

Commented [A18]: Need info about Advanced Search since I cannot visualize the page: SEE DOC FROM ANDREA

Commented [A19]: Perhaps we need a name for this pane, for clarity

Commented [A20]: The Search does not seem to work very well

Searching Using Analytics

Analytics search granularity is achieved through search tools and additional search functionalities.

You can access Analytics either from a library or from an analysis.

When you access Analytics from a library, you can search all the library documents. Because library documents were indexed using a standard ESSEX, the documents are not categorized, and only the documents targets are displayed – if available.

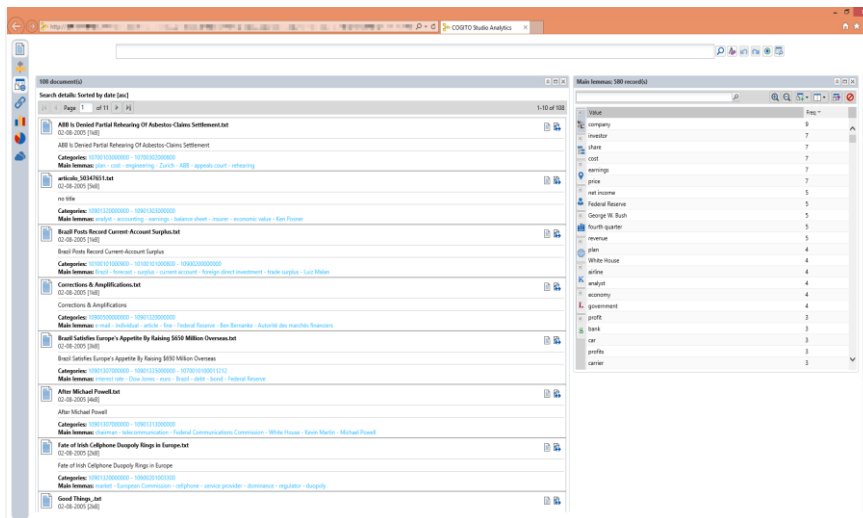
When you access Analytics from an analysis, you can search through all the analysis documents. In this instance, documents were indexed against a specific ESSEX – usually the ESSEX of the opened Cogito Studio project – and categorized. In addition, target details are provided – if available.

ACCESSING ANALYTICS

How to Access Analytics from a Browser Window

1. Right-click on a library or an analysis to show the contextual menu.
2. Select **Browse with Analytics**.

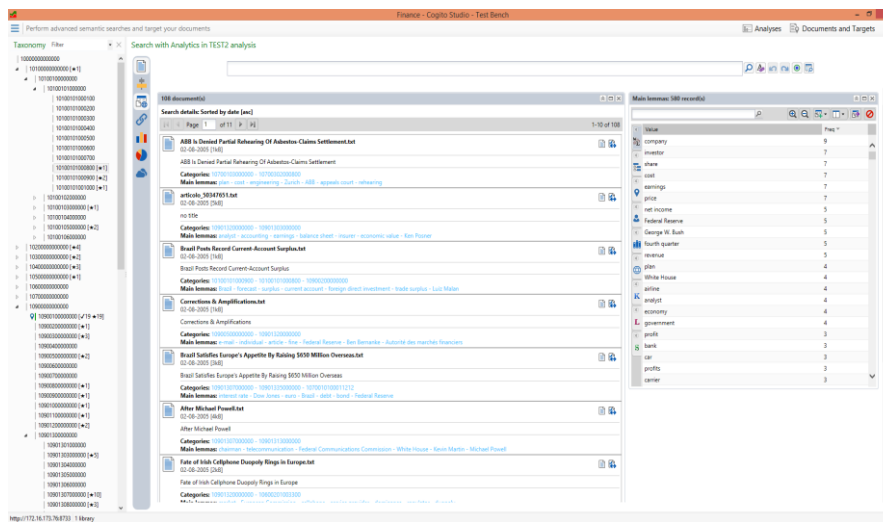
When loading the Analytics interface, a preferred browser window is displayed.



How to Access Analytics in Test Bench

1. Select a library.
 2. Click on Documents and Targets tab.
 3. Click on Search with Analytics.
- OR
- Double-click an analysis
4. Click on Documents and Targets tab.
 5. Click on Search with Analytics.

When you access Analytics from the Test Bench window, the taxonomy tree is displayed.



WHICH ARE THE ANALYTICS SEARCH AND DISPLAY TOOLS?

The search and display tools speed a library search, and offer versatility when viewing the search results.

They can be grouped in:

- Search box and Main Tools

Both the search box and the main tools buttons allow you to do a variety of search actions that may speed your search.

Each search functionality allows you to be specific with your search criteria, and achieve a higher degree of accuracy in your results.

- Analytics Panels Area

The Panels area is where the returned documents, analysis lists and charts are displayed, providing a detailed overview of your documents and trends.

- Analysis Panels

The Analysis panels display the search results in formats such as pie chart, tag cloud and so on, enriching the understanding of your results.

Commented [A21]: I prefer to use Search tools instead of Main components (it sounds like electronics)

SEARCHING IN ANALYTICS

The Analytics search allows you to sieve through your libraries and documents through the search tools.

In addition, you can enter Boolean operators in the search box, and define the semantic type for your term.

You can edit a search using the search editing tools.

For more information on the editing tools, see Which Are the Analytics Editing Tools?

What Are the Semantic Item Types?

The semantic item types are filters that allow you to specify the domain of the term that you are looking for.

For example, if you entered the term “trade”, are you looking for the lemma trade or for an organization called trade?

The semantic item types are displayed below the search box, after you do a search.

A drop-down list named after your search term is displayed. For each search term that you add, a search type filter is added.

The semantic type allows you to be specific in your search.

-  Keyword
-  Lemma
-  Multiple entries may appear on the list for all the found meanings in Sensigrafo.
-  Lemma
-  People
-  Organization
-  Place
-  Domain
-  Categories
-  Subject
-  Action (verbs only)
-  Object
-  Targets

How to Change the Semantic Item Type

1. Click on each drop-down list named after the term that you want to change the semantic type for.
The semantic items list is displayed under the search box.
2. Select the semantic type that you are looking for.

 You can type in the search field as many tokens as you desire, they are searched as a Boolean string.

Commented [A22]: Grammar type or item type?

Commented [A23]: Is there a maximum of search item types that can be displayed?

Commented [A24]: Is this in the right place?

The default Boolean operator is the AND operator, for example, if you enter diamonds pearls the engine will search for diamonds AND pearls.
To be more specific, you can use the Boolean operators.

What Is a Boolean Operator?

Boolean operators are simple words that connect search words together.

Boolean operators are commonly used in search engines to widen and refine the results, and more generally, to make searches more accurate and effective.

You can use brackets to define priority in Boolean expressions.

 When you use brackets in the Boolean search, you are required to put a single space after the opening bracket, and before the closing bracket.

Boolean Operator	Description
AND	This operator returns the documents that contain all the strings typed in the search box. This operator is applied by default.
OR	This operator returns the documents that contain either all strings entered in the search box, or at least one of them.
NOT	This operator returns the documents that contain the string preceding this operator but not the documents with the string following it. The NOT operator must be considered as an AND_NOT, so it must be used as a binary operator.

Commented [A25]: Is this correct?

How to Search in Analytics

1. Enter the term that you want to search in the search box.

When the system recognizes the term, it rewrites the term under the search box.

The color of the term stands for the semantic type.

 By default, the system recognizes search strings as lemmas and highlights them in pink. A string that is not recognized by the system, such as a foreign word, abbreviation and so on, is marked as a keyword and highlighted in blue.

2. Ensure that the system has recognized the semantic type of the search term.

The semantic item type is indicated by the colored search term under the search box.

3. Click on Search.

The system loads all the documents where the searched term for the selected semantic type is found. You can change the semantic item type. Any relates analysis is reloaded depending on this filter on documents.

Commented [A26]: This is not clear to me

EDITING A SEARCH IN ANALYTICS

Which Are the Analytics Editing Tools?

In Analytics, you search and edit your search using five search tools:

-  The Clear button clears the terms entered in the search box.
-  The Undo button removes the changes that you have made to the search.
-  The Redo button reinserts the changes that you have made to the **search**.
-  The Sensigrafo button allows you to look for a concept in Sensigrafo and add it to the search box.

For more information on searching Sensigrafo, see

Commented [A27]: Is this correct/clear?

Searching Sensigrafo from Analytics



The Advanced search options button.

For more information on the Advanced search options, see

Searching with the Analytics Advanced Search Options

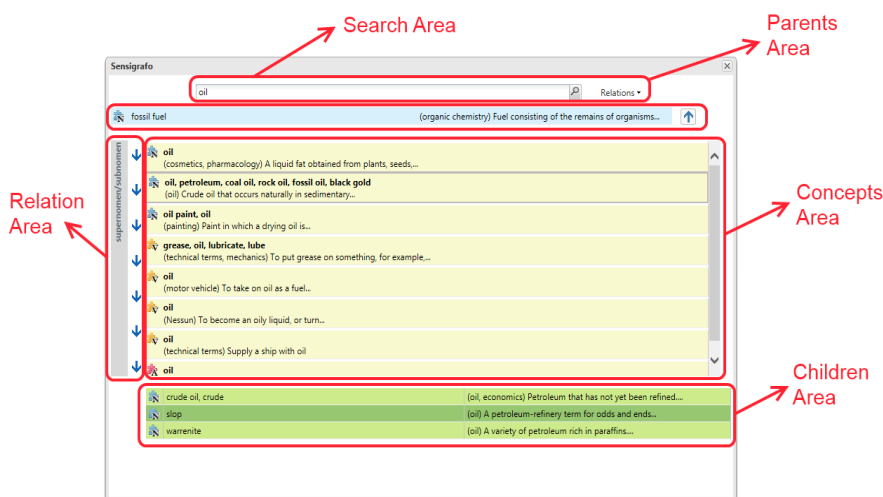
Searching Sensigrafo from Analytics

The Analytics search allows you to open and navigate Sensigrafo.

In particular, you can search the semantic network. The Sensigrafo search returns syncons and lists of lemmas, also showing how these elements are linked to other nodes within the network.

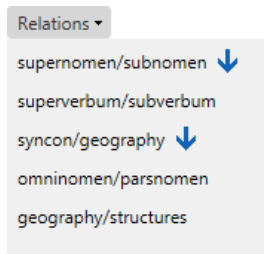
Which Data Is Displayed in Sensigrafo?

The Sensigrafo Navigator window consists of five main areas.



These areas are as follows:

- **SEARCH AREA**
This is where you insert your new items to search and specify a relation (supernomen/subnomen, superverbum/subverbum, syncon/geography, omninomen/parsnomen or geography/structures) to navigate semantic chain.
- **RELATION AREA**
This area indicates the current relations selected.



These are the four main relations that you can find in Sensigrafo.

- Supernomen/subnomen: (type of) it is a relationship between a specific concept (subnomen) and a more general one (supernomen). A supernomen is a node of the semantic net that has a general meaning compared to those that represent a specification of the same meaning (e.g. "plane" is supernomen to "seaplane"; in other words a "seaplane" is a kind of "plane");
- Superverbum/subverbum: (way of) it is a semantic relationship linking verb syncons together. This link is the equivalent for verbs compared to what supernomen/subnomen is for nouns (e.g. "sleep" is superverbum to "snooze", in other words "snoozing" is a particular way of "sleeping");
- Ominomen/parsnomen: (part of) it is a "part/all" semantic relationship. A parsnomen is a node of the semantic net that represents a part of something (ominomen) (e.g. "hand" is ominomen to "finger", in other words a "finger" is part of a "hand");
- Syncon/geography: it is a relationship between proper nouns expressing geographical inclusion. E.g. "Europe" includes "France", which in turns includes "Region Bretagne", which includes the city of "Rennes".
- Geography/Structures: it is a relationship among proper nouns that identifies geographical places in which structures (airports, buildings, railway stations) are placed (e.g. place "New York State" contains structure "New York JFK Airport"). This is a distinct link to "syncon/geography" as the latter both ends are geographical places (so long containment chains are possible), while "structures" are not.

- PARENTS AREA

This blue area represents the semantic parent (broader) of the selected concept. For each parent, we have a grammar type icon (noun, proper noun, verb or adjective), a list of lemmas and a short gloss.

After a search, you will see only the direct parent of the searched term. However, if you click on the chain button, you will be able to see the full parent chain.

The parent chain stops at the Sensigrafo root, or on the first step of the chain, where you can find more than one single parent node. Next to the full parent chain, the reset button allows you to return to the first parent.

Commented [A28]: This is not clear

Commented [A29]: Is this correct?

- CONCEPTS AREA

The yellow area shows, for each line, a concept, the domain, a short gloss, a list of lemmas, highlighting the grammar type.

 Noun

 Proper noun

 Verb

 Adjective

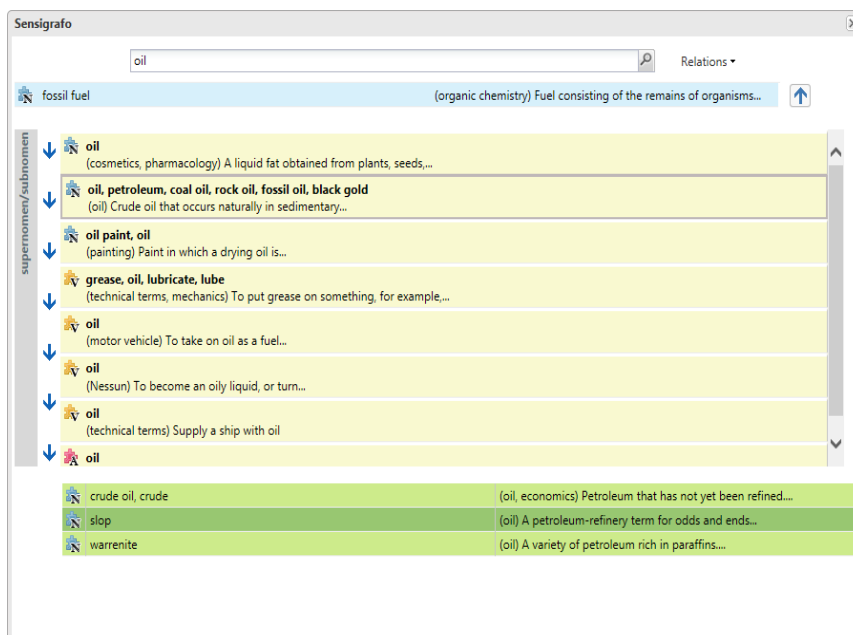
CHILDREN AREA

The green area represents the semantic children (narrower) of the selected concept. Each line is a concept, with its grammar type icon (noun, proper noun, verb or adjective), the list of lemmas, the domain (into brackets) and a short gloss.

Commented [A30]: ??????

How to Search Using the Sensigrafo Search

1. Click on the Sensigrafo icon  to recall the Sensigrafo search field.
The Sensigrafo search field is displayed.
2. In the Sensigrafo search field, enter the search term.
3. Click on Search.
The full Sensigrafo Navigator window is displayed.
4. Double-click on the concept that you want to research.
Analytics adds the concept to the search box and searches for documents.



How to Navigate in Sensigrafo

1. Double-click on a parent or a child to navigate the semantic network.

How to Search Using Sensigrafo Expanded Concepts

1. Search for a term in the Sensigrafo search field.

The search results are displayed in the Sensigrafo window.

 For more information on how to search Sensigrafo from Analytics, see *How to Search Using the Sensigrafo Search*.

2. Click on the arrow icon  next to the term that you want to add to the Analytics search.
3. Drag the term into the Analytics search box.

The term is displayed in the Analytics search box.

A message lets you know if the term has already been added.

 Ensure to drag the term into the Analytics search box and not the Sensigrafo search field.

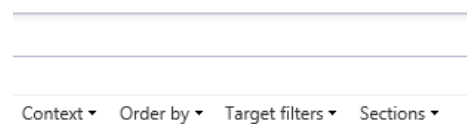


Searching with the Analytics Advanced Search Options

What Are the Analytics Search Options?

The Analytics advanced search options is a set of filters that allow you to narrow down and to organize your search results.

This set of options consists of four sub-menus that are displayed below the search box. You can choose one or more options to specify additional search criteria, and how to display your search results.



The advanced search options are:

- **Context**
This option determines the search scope.
Choose between Document or Sentence.
- **Order by**
This option lets you choose how to view your search results.
Choose by documents **relevance**, date or size.

 *Document relevance is the result of a query that applies a search algorithm to determine the best matches; the algorithm finds the most relevant results based on the search term in the search field.*

- **Target filters**
This option allows you to choose a filter for your search.
Choose between All to view all documents, and No target to view only documents that have no target.
- **Sections**
Cogito Studio uses XML sections to create custom document segmentation.
Document segmentation allows you to choose the XML sections to examine at search refinement.

 *You define the XML sections both in the config.xml file and during the documents upload to a library.*
For more information, see How to Upload a Set of Documents into a Library.

How to Display the Advanced Search Options

1. In the Analytics Search tools menu, click on the Show/hide toggle  to display the advanced search options.

OR

Click on  to hide the advanced search options.

Commented [A31]: Is this accurate?

Commented [A32]: Awaiting for info on this part from GS and AT

How to Change Search Scope – Context

1. Click on the Context drop-down list.

The drop-down list allows you to choose between Document and Sentence.

2. Select the search scope.

 *If you select Document, Analytics returns all the documents that contain the search terms.
If you select Sentence, Analytics returns only documents where the search terms are on the same sentence.*

How to Change the Documents Display Order – Order by

1. Click on the Order by drop-down list.
2. Select how you want the documents to be displayed.

The criteria that you can choose from are:

- Date ascending
- Date descending
- Size ascending
- Size descending
- Relevance

How to Filter Documents Based on Target and Category Availability – Target Filters

1. Click on the Target filters drop-down list.

You can choose from:

- All
If you select this option, Analytics displays all documents.
- No target
If you select this option, Analytics displays all documents without a target.
- No category
If you select this option, Analytics displays all the uncategorized documents.

 *Because Categories are not available in libraries, the No category option is only available when you search an analysis documents in Analytics.*

How to Search in a Document Section – Sections

1. Click on the Sections drop-down list.
2. Select the section where you want to look for the search term.

You can choose from:

- All sections
- <Section Name #1>; <Section Name #2> and so on.

All the available document sections are listed in the Sections drop-down list.

When you select a section, Analytics searches through the selected section.

WHAT ARE THE VISUAL SEARCH REFINEMENT ELEMENTS?

A visual search refinement element is a graphic representation of your search results that you can further refine.

Test Bench provides seven visual search refinement elements that allow different views of your search results.

Each view is displayed in a dedicated pane, where you can click on the results representation for further refinement.

You can close, enlarge, minimize, expand or collapse each pane through standard icons.



For more information, see [Using the Search Refinement Views](#)

The Test Bench visual search refinement elements are:



View Documents

This element displays the documents list as a result of your search.



View sentences centered on the first search element

This element displays all sentences that contain the first search term in your search.



Analysis list

This element displays a set of elements extracted from the documents in your search results.



View SAO analysis

This element displays the SAO analysis. This analysis extrapolates all the key elements having the searched term as a subject, the verbs associated to the searched term and the objects that refer to the searched term.



Bar chart

This element displays all the main lemmas contained in your search results.



Pie chart

This element displays all the documents with a target.



Tag Cloud

This element displays a panel that provides a visual representation of the most common key tags in your search.

Note: The visual elements rely on the same taxonomies that have been injected into your Cogito **Studio** project.

Commented [A33]: Is this correct?

USING THE SEARCH REFINEMENT VIEWS

With the search refinement views, you can view your results and further refine them by criteria such as target, main lemmas, sao-list, key words and more.

Each view contains a number of clickable parts that allow you to:

- Position your search
- View the available search options
- Understand the implications of your search results
- Identify trends
- Decide your search strategy.

Refining Your Search from the Documents View

The Document view is where you browse documents and collect useful information about them.

From the document view, you can gain a higher level of details about the documents and refine your next search clicking on the document target, categories and lemmas.

You can navigate through your search results either by using the next/previous pagination button, or by entering the page number.

From the Documents view, you can see document information and perform actions on the selected documents.

What Is the Document Preview?

The document preview box contains information about each document in your search results:

- Document name
- Document size
- Document **date**
- Document summary
- Targets
- Categories
- Lemmas

You can click on each item, with the exception of the document date, to view more information about a document and its content.

Commented [A34]: Is this the doc date or the doc upload?



Which are the Document Actions?

You can open the document, view all related documents and view the document sentence by sentence.

- Use the  icon to open the document.
- Use the  icon to view all related documents.
- Use the  icon to expand the document one sentence at the time.

By default Analytics shows only the sentences that match the search keywords.

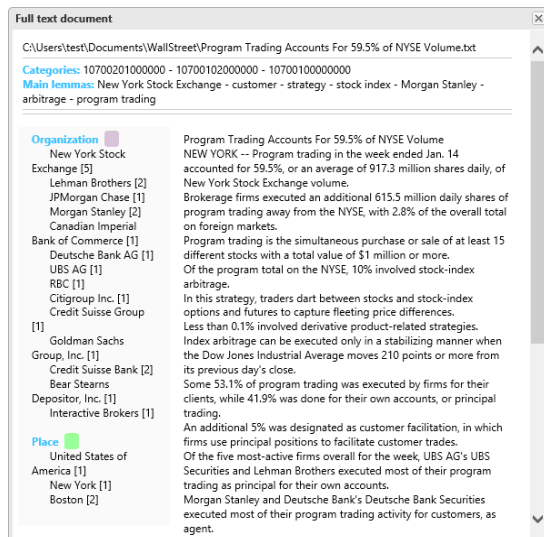
Use this icon to show more sentences.

What Is the Full-Text Document Window?

From the Document preview, you can open the full document in the Full text document window.

The full-text document window shows the full document and provides document information as follows:

- People – named entities of type Person that are retrieved from the document.
- Organization – named entities of type Organization retrieved from the document.
- Place – named entities of type Place retrieved from the document.
- Document filename, categories and main lemmas as highlighted in the Preview box.

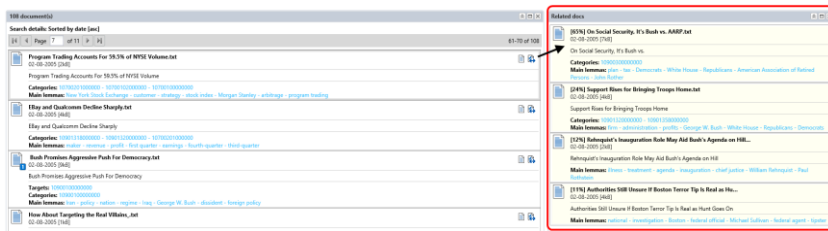


What Is the Related Documents View?

The Related documents view shows documents similar to the one that you have selected.

The similarity is assessed by comparing terms that are extracted from documents and their relations.

The Related documents list is similar to the Documents view and shows the similarity score next to the title of each related document.



How to Refine a Search from the Document View?

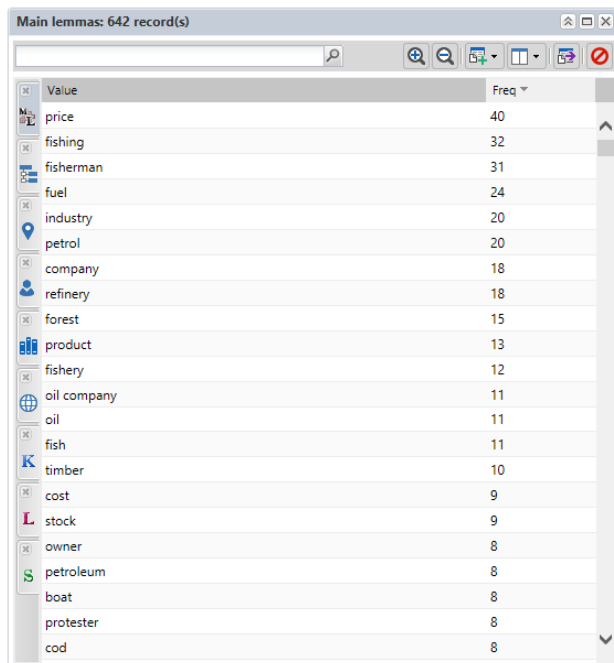
To refine your search, you can add a category, target and main lemmas into the Analytics search box.

1. Click on Categories, Targets or Main lemmas to add them to the search box.

You can add one or more items to the search box.

Refining Your Search from the Analysis List View

The analysis list shows the terms that are extracted from the documents in your search results.



The screenshot shows a window titled 'Main lemmas: 642 record(s)'. It contains a table with two columns: 'Value' and 'Freq'. The table lists various terms and their frequencies, organized by category. The categories are indicated by icons on the left side of the table.

Value	Freq
price	40
fishing	32
fisherman	31
fuel	24
industry	20
petrol	20
company	18
refinery	18
forest	15
product	13
fishery	12
oil company	11
oil	11
fish	11
timber	10
cost	9
stock	9
owner	8
petroleum	8
boat	8
protester	8
cod	8

The terms are organized by category as set in your taxonomy.

You can select the category that you want to investigate.

When you select the People category, the analysis list shows all the people in your search results, whilst the Organizations category displays the organizations in the search results.

For example, the term Dell can be found in documents that refer to the company Dell, or in documents that refer to Michael Dell, the CEO of Dell.

Viewing search results terms by category allows you to quickly find the information and identify the documents that are relevant.

You can refine a search and select the type of information that you want to display using:

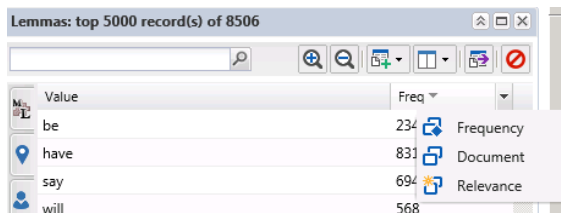
- The frequency item type
- The search box
- The Analysis list menu

What Is the Frequency Item Type?

For each category, you can view three frequency item types :

- The number of documents where a term is found .
- The term relevance
- The frequency of a term.

To change the frequency item type, click on the item drop-down list next to the Freq column header.



What Is the Analysis List Search Box?

The Analysis list search box is where you type a term to apply a filter to your current category-list.

You can apply a filter for each category in the analysis list.

How to Search Using an Analysis List

1. In the Analysis list, click on the **View more analysis** tab icon  to show more analyses lists.
A list of analyses organized by type is displayed.
2. Select an analysis type.
A list of terms regarding the analysis type that you have selected is displayed.
You can use the smaller search box to filter each list.
3. Double-click on an item in the list to add it to the Analytics search box.



The term is added using the AND Boolean operator.

For more information on Boolean operators, see [What Is a Boolean Operator?](#)

What Is the Analysis List Menu?

The analysis list menu allows you to choose the terms that you view, and how you view them.

The analysis list menu items are:



View all

This menu item displays all the terms that the search has produced for a given category.

By default, it shows the first 5000 results.



View top

This menu item displays the most frequently found terms in a search.



View more analysis categories

This menu item lists all the available analysis organized by semantic type.



Compare lists

This menu item allows you to select a category to compare with the currently selected category.

In compare lists, double-click on a term in the lists to add search terms to the Analytics search box.



Export current analysis

This menu item allows you to create and export the terms that are extracted from your search results into a csv file.

The content of the exported file depends on the category that you have chosen in the analysis list view.

Choose from:

- Open
Click on Open to open the file in the associated CSV viewer.
- Save
Click on Save to decide where to save the CSV.
- Cancel
Click on Cancel to abort the export process.



View exclusion list

This menu item shows all the terms in the exclusion list.

Commented [A35]: Se questo button è quantitativo (mostra solo i primi 1000 risultati) allora sarebbe meglio si chiamasse View less. SE è qualitativo, allora View top (implied top results) va bene così come va bene quello che ho scritto. Altrimenti va cambiato

What Is the Exclusion List?

You can refine your search results by excluding unwanted items that continue to appear in your analysis list.

To prevent an item to be displayed again in the list analysis, you can add the item to the Exclusion list. This item will be deleted.

How to Add an Item to the Exclusion List

1. In the Analysis list view page, right-click on the item that you want to remove from your search results.

The Add item to exclusion list dialog is displayed.

2. Click on the dialog box to add the item to the exclusion list.

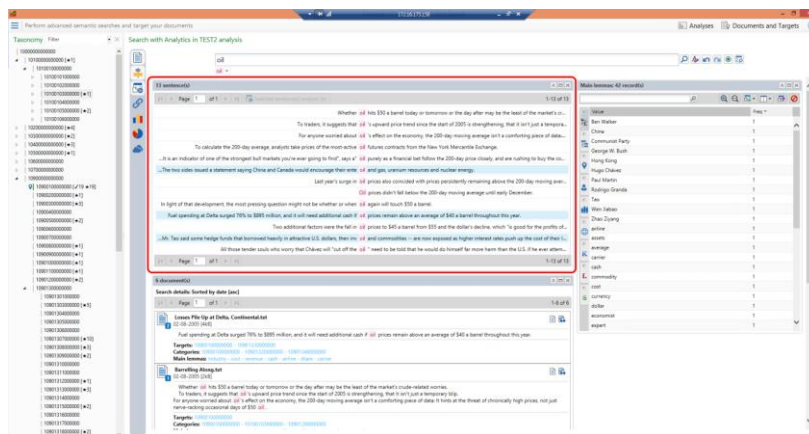
 *When you add an item to the exclusion list, it is not possible to move it back from the Exclusion list. The item will be lost.*

Refining Your Search from the Sentence View

You use this view to improve the results granularity.

You can look for a term in your search results list and view the number of occurrences, and the sentences where the term occurs.

This view highlights the term in each document, and displays a clickable summary of all the documents where the term is found.



How to Switch to Sentence View

1. Enter a term in the search box.
2. Select the Sentence view icon .

A list of clickable sentences centered on the searched term is displayed in the Analytics main area.

3. Click on each sentence to view the document.

The resulting documents are shown in the Document view.

How to See the Analysis List Related to a Selected Sentence

1. Enter a term in the Analytics search box.
2. Click on the Sentence view  icon.
3. When the Sentence view is loaded, select one or more sentences.

You can select multiple sentences using the **ctrl**-key.

4. Click on Selected Sentence(s) analysis list.

The analysis list is displayed for the selected sentences.

Refining Your Search from the Sao-Analysis View

SAO stands for Subject - Action – Object. The SAO list view consists of three clickable lists that identify the grammatical relationship between lemmas and other words in a document.

The three lists are:

- Subjects
- Actions
- Objects

SAO lists		
View SAO triples		
Subject: 664 record(s)	Action: 651 record(s)	Object: 634 record(s)
Value	Value	Value
Freq	Freq	Freq
price 22	be 42	price 21
company 18	say 38	oil 13
government 13	have 28	fuel 10
demand 13	make 22	production 10
oil 10	use 18	business 8
oil company 10	see 17	demand 8
cost 9	add 15	barrel 7
business 7	take 13	product 7
be 6	do 13	company 6
day 6	sell 13	be 6
biodiesel 6	produce 12	car 6
United States of America 6	run 12	week 5

The SAO view shows:

- Every sentence, where the lemmas produced by the search results are the subject.
- All the verbs (actions) that are associated to the subject-lemma.
- All the objects associated to the subject lemma.

Because the SAO list view highlights the relationship between the subject-lemma and associated verbs and objects, you can be very specific when you refine a search.

In the SAO view, you can toggle between SAO triples and SAO list.

You can export the SAO lists in .csv file.

What Is Displayed in the SAO List View?

The SAO list screen displays all the terms produced by your search that have no relationship with one another.

What Is Displayed in the SAO Triples View?

The SAO triples screen displays the SAO relationships for each lemma returned by your search.

How to Refine a Search Using the SAO Lists

1. From Analytics, click on the SAO list button .

The SAO list view is displayed.

 Each list displays a term and the frequency with which the term occurs in your search results.

2. In each list, type in the term that you are investigating.

The auto-complete shows a list of suggestions to choose from.

When you have entered a term in each SAO list search box, a list of documents containing the terms that you have entered, is displayed.

OR

In the SAO list that you want to investigate, double-click on an term to add it to the main search box.

 To remove a term from the SAO list search box, click on the remove icon .

How to Export the SAO List

1. In the SAO list, click on the export icon .

The SAO list is exported in .csv file **format**.

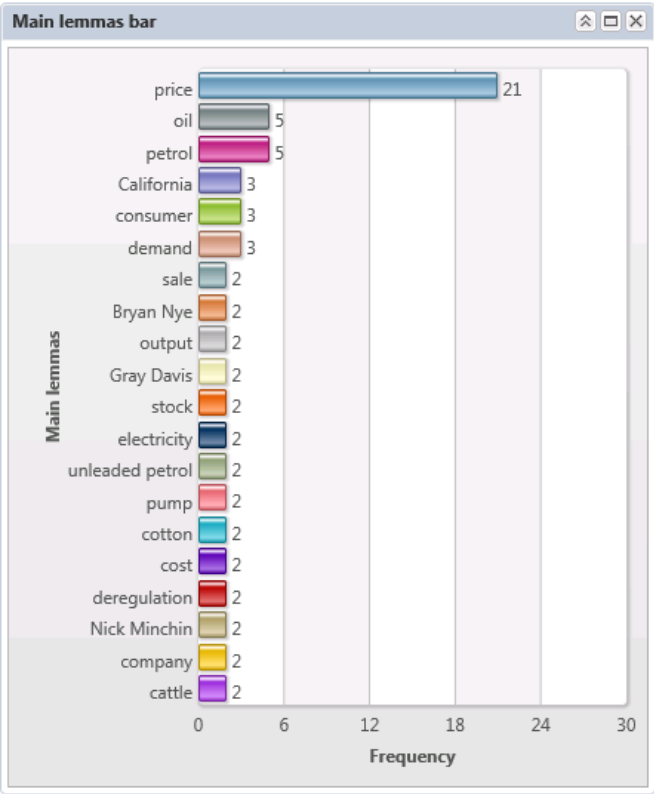
Commented [A36]: IT doesn't seem to ask where to save the file, give a name etc

Refining Your Search from the Bar Chart View

The bar chart view shows the 20 most frequent lemmas that your search has produced.

The chart is clickable to allow you to refine your search based on the found lemmas:

When you click on a bar representing a lemma, the lemma is added to the search box for your next search.



How to Open the Bar Chart View

1. From Analytics, click on .

The bar chart showing your search results is displayed.

Refining Your Search from the Pie Chart View

The pie chart view provides a quick overview of the categorization results that your search has produced.

The pie chart helps you to understand if categorized documents (ruled-based categorization) have a distribution that is similar to the targeted documents (manual categorization)

The chart is clickable to allow you to refine your search based on the categories found as a result of your previous search.

Each slice of the chart represents a category of your taxonomy.

What Information Is Displayed in the Pie Chart when Opening Analytics?

When you select the pie chart view in Analytics, you can view one or two pie charts.

You can access Analytics from:

- An analysis,
The Targeted docs and the Categorized docs pie charts are displayed.
- A library
The Categorized docs pie chart is displayed.

The Targeted docs pie chart shows your documents targets.

A slice of the Targeted docs pie chart represents one of your documents targets.

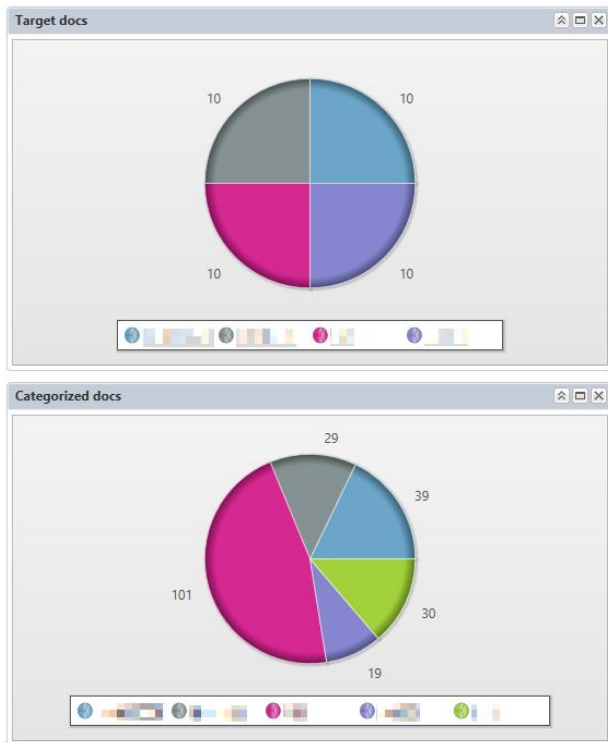
The Categorized docs pie chart shows your documents categories.

A slice of the Categorized docs pie chart represents one of your documents categories.

How to Open the Pie Chart View

1. From Analytics, click on .

The pie chart showing your search results is displayed.



Refining Your Search from the Tag Cloud View

The Tag Cloud view provides an overview of the documents tags that your search has returned.

The tag font size and color indicates the frequency and the relevance of a tag:

The bigger the font size the most frequent the tag is.

By default, the tag cloud view shows the main lemmas tags, but you can choose the category that you prefer clicking on the Tag cloud content drop-down list.

Each tag is clickable to allow you to refine your search.

When you click on a tag, the term is added to the search box for your next search.

You can choose the category that you want to display in the tag cloud window.

How to Open the Tag Cloud View

1. From Analytics, click on .

The tag cloud showing your search results is displayed.

How to Select a Category to Display in Tag Cloud

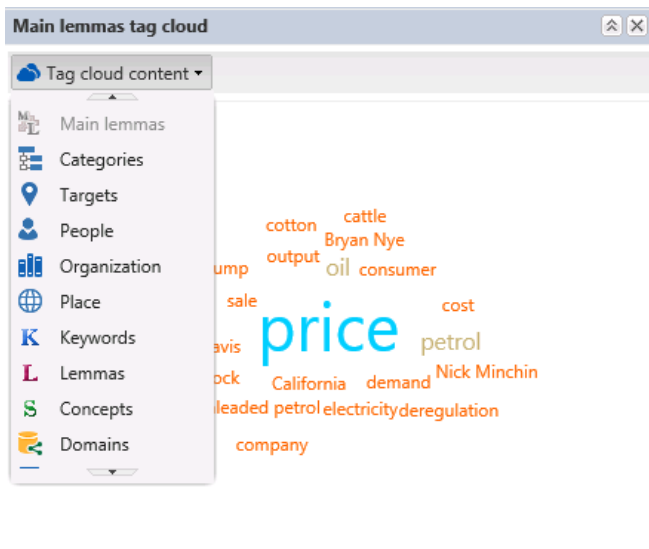
1. Click on the Tag cloud content drop-down list.

A list of categories based on your taxonomy is displayed.

2. Select the category that you want to display in the tag cloud window.

Commented [A37]: Is this correct?

Commented [A38]: How does it work with colours? And with colours in conjunction with the font size? Is there a colour-code that indicates the most important?



Index

Accessing Cogito TestBench; 4
 Advanced search options
 Analytics; 41
 Displaying; 41
 Analyses
 Comparing; 27
 Analyses comparison
 Table; 28
 Analysis
 Interrupting; 14
 Analysis dashboard; 25
 Analysis list
 Menu; 50
 Search refinement; 48
 Searching; 49
 Viewing; 52
 Analysis list search box; 49
 Analysis results
 csv file format; 14
 Exporting; 14
 Interpreting; 17
 Viewing; 23
 Analytics
 Accessing; 31
 Accessing from a browser; 31
 Accessing in Test Bench; 32
 Searching; 31
 Analytics search
 Editing; 36
 Analytics Sensigrafo search; 39
 Automatic rules; 15
 Automatic Rules
 Generating; 16
 Bar chart
 Search refinement; 55
 Bar chart view
 Opening; 55
 Boolean operator; 35
 Category selection
 for Tag cloud; 58
 Cogito Overview
 Cogito Caegorizer, Cogito Discover; 2
 Cogito TestBench; 3
 Contextual search; 41
 Data displayed; 26
 Micro and macro average; 26
 Document actions
 List of; 46
 Document section
 Searching; 43
 Document target
 Classification; 9
 Modifying; 11
 Document Target
 Association; 7
 Documents
 Analyzing; 13
 Searching for; 29
 Searching inside; 30
 Documents analysis
 Comparison; 28
 Documents order by
 Changing; 42
 Documents section; 26
 Documents view
 Search refining; 45
 Exclusion list
 Adding an item to; 51
 What is; 51
 False negative; 18
 False positive; 18
 First term
 Search refinement; 52
 First term search
 View activation; 52
 F-measure
 Harmonic mean; 20
 Frequency item type
 What is; 49
 Information summary; 25
 Key Concepts
 Test Bench; 17
 Libraries
 Working with; 6

Expert System

Library

- Adding; 6
- Deleting; 8
- Documents uploading; 8
- Opening; 7
- What is a library?; 6

Library documents

- Analyzing; 14
- Targetting; 10

Linguistic resource; 15

Linguistic resource option

- ESSEX service; 15

Macro average

- What is; 22

Macro-averaged

- Precision and Recall; 21

Micro and Macro average; 21

Micro average

- What is; 21

Micro-averaged

- Precision and Recall; 21

Order by search; 41

Pie chart

- Information; 56
- Opening; 56
- Search refinement; 56

Precision value; 19

Predictive values

- Positives and Negatives; 17
- Viewing; 24

Predictive Values; 17

Recall value; 19

Related documents

- Viewing; 47

Returning to Cogito Studio; 5

Rules

- Testing and validating; 15

SAO Analysis

- Search refinement; 53

SAO list

- Exporting; 54
- Using for search refinement; 54

SAO list view

- What is; 54

SAO triples view

- What is; 54

Expert System

Search and Display Tools

- Analytics; 33

Search context

- Changing; 42

Search editing tools

- Analytics; 36

Search refinement

- Views; 44; 45
- Visual elements; 44

Search Refinement

- Definition; 47

Searching

- Analytics; 34

Searching in Analytics; 35

Sections search; 41

Semantic technology; 2

Semantic type

- Changing; 34

Semantic item types; 34

Sensigrafo expanded concepts; 40

Sensigrafo navigation

- Analytics; 40

Sensigrafo searching

- Analytics; 37

Standard search; 29

Statistical information

- Understanding; 27

Tag cloud

- Opening; 58

Tag cloud view

- Search refinement; 58

Target

- Removing from a document; 12

Target filters; 42

Target filters search; 41

Targets

- Removing; 11

Taxonomy tree

- Understanding; 10
- What is; 9

TestBench

- Logging into; 5
- Opening; 5
- Searching with; 29

TestBench Analysis

- Interrupting; 13

TestBench analysis results
Viewing; 23
True negative; 18

True positive; 18
Updated linguistic rules
Deploying; 16

Contacts

Contact Information:

Mail: info@expertsystem.com

WebSite: <http://www.expertsystem.com/>

Regional Technical Support

EMEA and Asia Pacific: support@expertsystem.com

Americas: techsupport@expertsystem.com