



## Agri-food Information Search System for the University of Kaposvar

### Preliminary requirements

During the project we want to develop an application:

- that is for students and teachers of agricultural sciences
- and provides fast and structured access to the global daily information on the agri-food sector in one place
- primarily based on English language sources.

Our idea is that University of Kaposvar and KCA (Croatian partner university) together:

- gather possible sources (we think about 50 free professional portals in English, plus in Hungarian market information may be available eg. agricultural research institute average buying prices)
- searchable text information
- search terms
- develops the concept of display structure

The service provider then creates the application that provides:

- Generate a browsable interface ((which is similar to news portals) on a daily basis about the latest news from the industry from the predefined sources, search terms, and structure.
- ***Periodically download the content from information portals (eg, demand forecasting, price information, daily commodity exchanges, international logistical information, weather) and makes it searchable***

- It also includes an ad-hoc search application, in which the user enters the search term himself, but also receives the results in a structured way. For example, if I type "agroforestry technologies", I can group the results:
  - o by Publication Time,
  - o by country names in text 2 (country grouping),
  - o based on similarity between two texts (the same news appears in several places, linked to each other in a visual way)
  
- A mobile application

The system administrator of the two Universities should be prepared for adding, and replacing information sources.

The application should be supported for 5 years.

We would like to have a beta version of the test, the experience of which would be incorporated into the final product.

## Our proposed solution

As a solution to this requirement we would deliver a customized Precognox TAS Enterprise Search solution, which is an intelligent, advanced search system that, with its components, serves the text search needs of large organizations and enterprises. These provide a backend software for collecting and searching the necessary information for this project.

*The portal like frontend would be developed by a different entity in a joint project with whom we would collaborate during their development.*

Our solution consists the following Precognox TAS products:

**TAS – Enterprise Search Engine** – is a widely customizable Solr / Elastic-based enterprise search engine with massive data searching capability (access rights to your data). TAS Search Engine enables the user to accomplish searches in the data collected by TAS Data Collector. It is a perfect combination for when you

not just need the data, but you also want your dataset to be effectively searchable. TAS Search Engine is also capable of finding named entities (i.e. company names or dates) in various formats.

**TAS – Search Log Analyzer** – once you have your structured database and it is searchable, you may be keen on getting information about the launched searches. As an example, TAS – Log Analyzer lets you know which keywords are used frequently or without any match. These and similar information can be used to continuously improve your search system. Analyzing the search logs TAS Log Analyzer is the perfect tool to help you discover how users use our enterprise search engine.

**TAS – Thesaurus Manager** – is a synonym-builder module that enables the building of more intelligent enterprise search engines with the TAS Search Engine platform. Searches launched with the combination of TAS Search Engine and Thesaurus Manager lead to more meaningful and relevant matches. TAS Thesaurus Manager allows you to define and control conceptual relationships (e.g: synonyms, broader or narrower term) between search terms.

**TAS – Data Collector** – is able to collect Internet data content in a structured format so as to make this content available for information systems or for further processing and analysis. TAS Data Collector is a very effective tool if there are web-based contents amongst the information sources of the organization. This solution provides such output format results that is ready for further utilization (e.g: for Business Intelligence tools).

### **Development and system integration process**

With our system analyst, with the active involvement of the client's experts, system design through the mapping of current workflows, customized solutions will be introduced by customizing these software products. System design requires personal consultation.

Based on the final requirements list and functional specification accepted by the customer, the exact system architecture is designed.

The configured software components are continuously tested by our QA team in a constantly expanding test environment.

Our offer does not consist the development of the responsive web page (nor the mobile app) which should be created by another party. Our offer does not consist a mobile app either. We do not offer a solution to the similarity based document grouping.

### **Expected Schedule for a Search Deployment Project**

The chosen search solution can be delivered to the customer in **6 months** after the order and final specification is accepted.

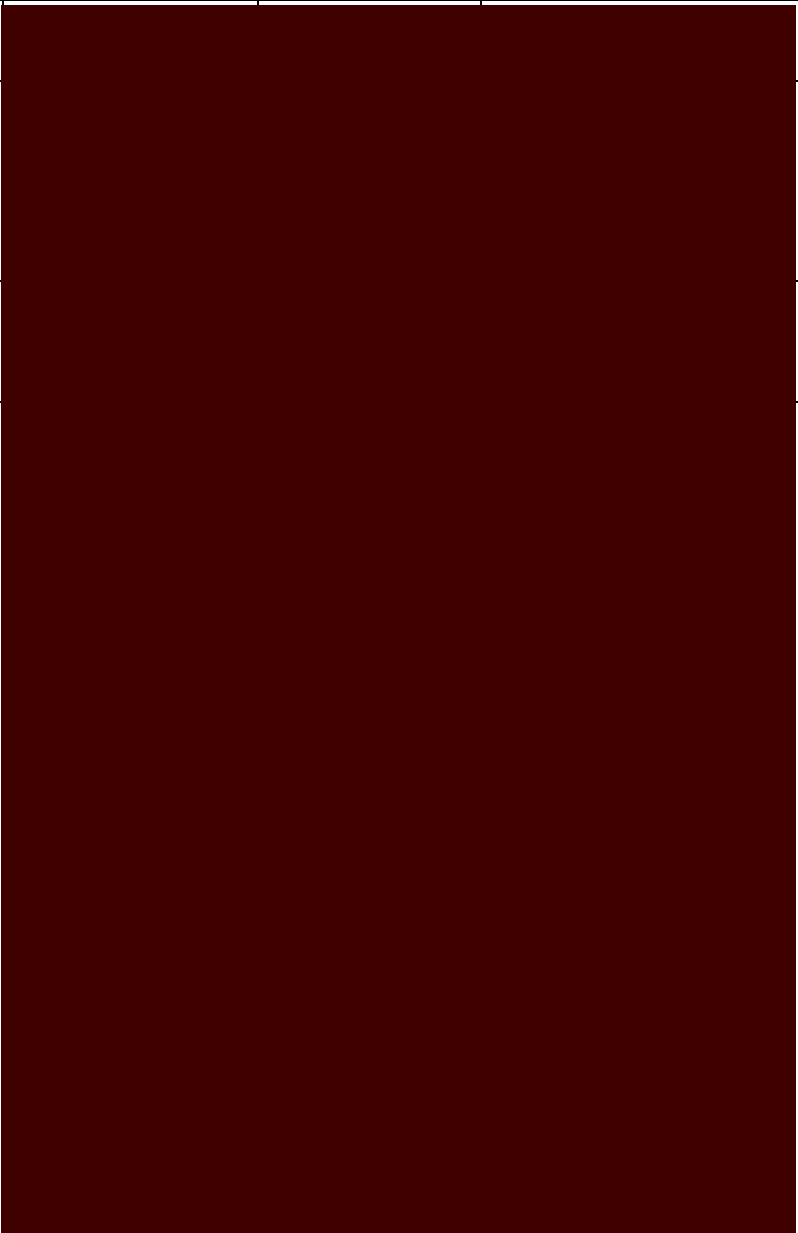
During that period we provide support for the another Development team who is going to develop the web interface / homepage for the project.

A detailed technical description of the software products we deliver is provided in the annexes.

We deliver an on premise version which is administered and supported primarily by the system administrators of the Universities, and we provide 3<sup>rd</sup> level support from Monday to Friday between 9-16 PM for 5 years. *The maximum number of support hours monthly provided in the contract is 2 hours.*

### **Offer**

	Product / service	
1.	TAS Enterprise Search engine, on premise version	
2.	Precognox TAS – Search Log Analyzer –	

	integrated to the search engine	
3.	Precognox TAS - Thesaurus Manager integrated to the search engine	
4.	Precognox TAS - Data Collector	
5.	Education and training of two system administrators  Training to the client how to configure and update sources for the TAS Data Collector.	



## Annexes

Precognox price list.....	8
Precognox TAS Thesaurus Manager .....	8
Precognox TAS Search Log Analyzer .....	14
Precognox TAS Enterprise Search .....	19
Precognox TAS – Data Collector .....	24

Kaposvár, 2019. 04. 29.

**Endre Jofoldi**

General Manager



**Endre Jofoldi**  
General Manager

[endre.jofoldi@precognox.com](mailto:endre.jofoldi@precognox.com)



**Karoly Kasa**  
CTO

[karoly.kasa@precognox.com](mailto:karoly.kasa@precognox.com)



**precognox**

[info@precognox.com](mailto:info@precognox.com) | [precognox.com](http://precognox.com)



# Precognox price list

<b>IV</b>	Manual Tester / Test Analyst, Annotator, Programmer Trainee (less than 1 year experience)
<b>III</b>	Java Developer, Frontend / Javascript developer, Sitebuilder, Test Engineer, Project Manager, Config manager (for our Data Collector product), Data Acquisition and Quality, Junior Data Analyst
<b>II</b>	Senior Java developer, QA Lead, Lead Annotator, Linguist, Senior Front End / Javascript Developer, Operations expert (DevOps), Business Analyst, Test Manager, Senior Test Engineer, Senior Project Manager
<b>I</b>	Search Expert, NLP Expert, Data Analyst, Test Expert, Senior Business Analyst, Solution Architect

**Valid from January, 2021**



# Precognox TAS Thesaurus Manager



## The goal of the tool

TAS Thesaurus manager is a synonym-builder module, Searches launched with the combination of TAS Search Engine and Thesaurus manager leads to more meaningful and relevant matches.

Thesaurus Manager allows you to define and control conceptual relationships between search terms.

## Requirements

- The system needs to be able to create and modify relationships between pairs of words (relationships: synonym, correct form, spelling, narrower group, broader group, stop list)
- It should perform - in multiple steps and considered to access rights - spell-based administration, such as annotation, verification, finalization, and so on.
- Provide bulk imports (CSV)
- Provide bulk export (CSV, standard OWL)
- Available through API for other IT systems

## Implementation

TAS Thesaurus Manager works within the Precognox TAS platform.

## Features

- Compatibility with TAS Search on Precognox TAS via Thesaurus Service.
- (Read more Technical Description - Precognox TAS Search)
- Accessibility on widespread browsers, intranet / internet
- User Access Restriction (Authentication / Authorization)
- Multilingual subtitle support (automatically detected based on browser language setting)

## Word pairs

- must include a basic expression (base expression, base word, first word)

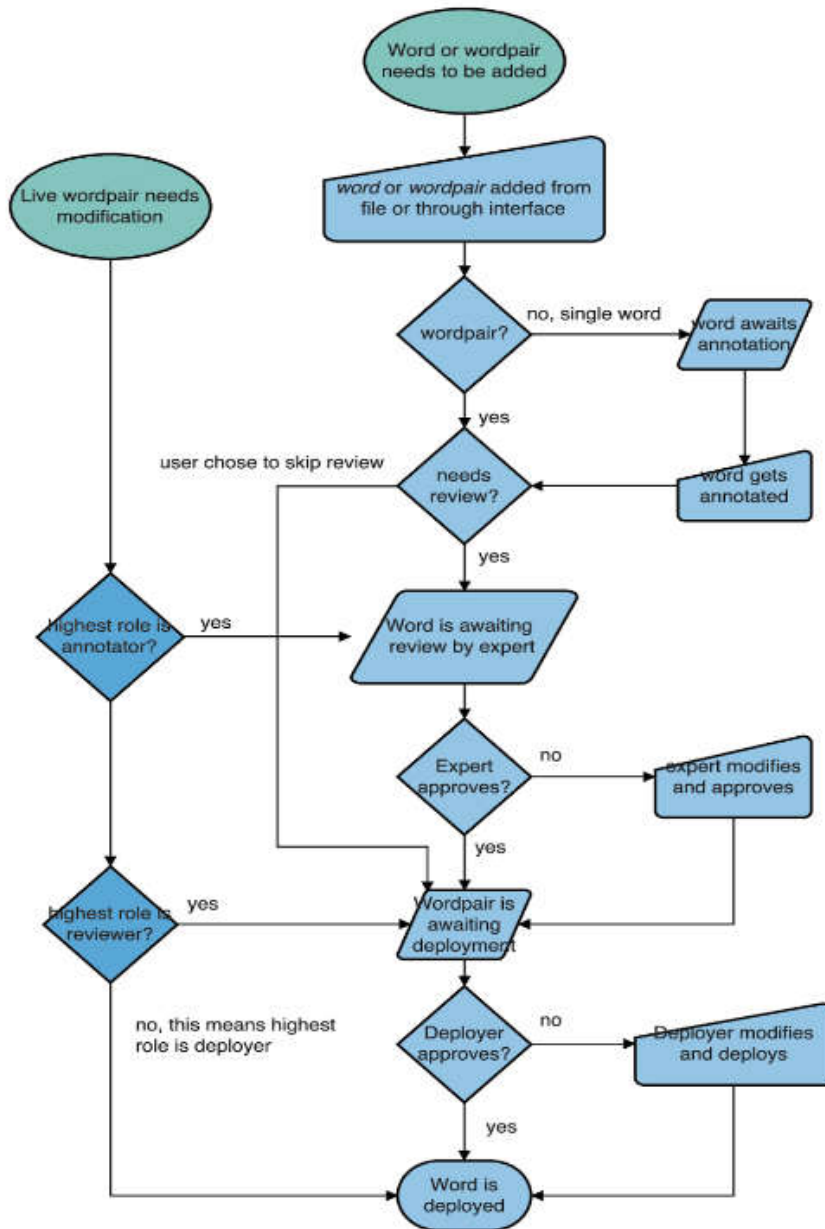
- can include a related expression (related word, related word, second keyword)
- has the specification opportunity of the relationship between the two terms (synonym, stoplist, narrower term of, broad form of, typo of)
- has category definition option to help annotators and reviewers decide which areas of the word pair are related to

## Workflow Roles

- Annotator: basic role, can handle word pairs added to annotation
- Reviewer: Has annotator rights, could already accept previous annotations
- Deployer: has both annotator and reviewer role, can install the already adopted spellings for sharp use (eg search)

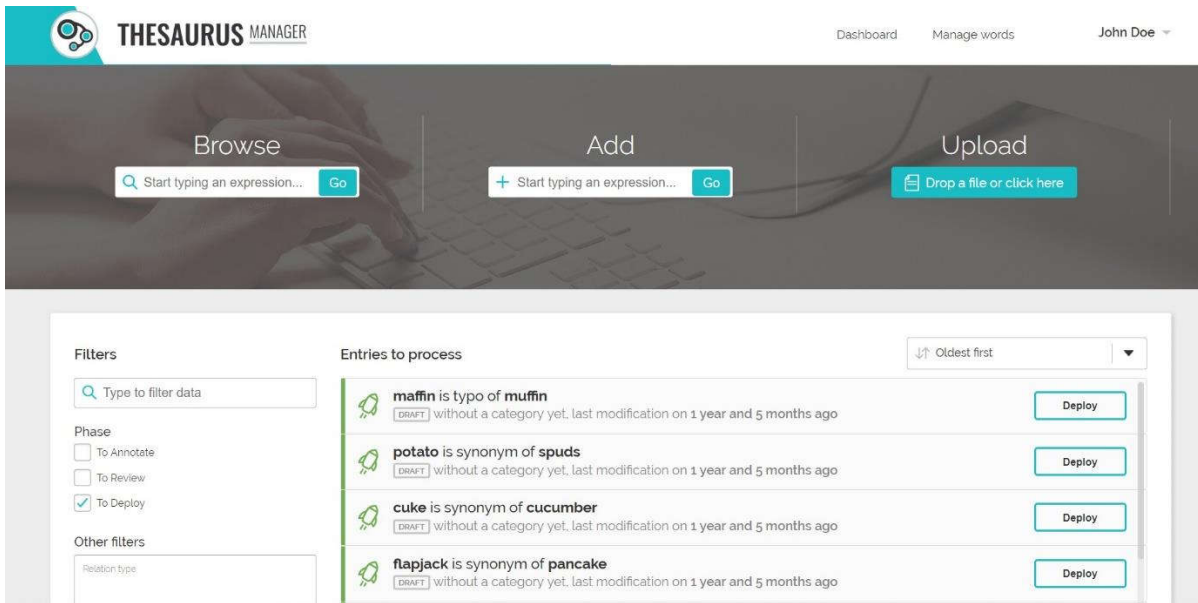
## Workflow

- Any user can add a new spell by specifying the base term for annotation, with as little information as possible to save the word pairs
- The related phrase could be already added during insertion or annotation (this is true for the following fields)
- Then the specification of the relationship between the two terms (or a category to which the pair belongs) is possible
- After insertion, the pair of words can be annotated (fields not filled in are specified by the annotator)
- The process must be accepted by a reviewer
- Approved spellers can be deployed by a Deployer



## The Thesaurus Manager GUI

The appearance of Thesaurus Manager is consistent with the corporate identity of TAS Platform. GUI is user friendly and can be handled intuitively. The interface is widely customizable, it is possible to individually modify the elements according to customer needs.



Technical description (technical needs, integration with other softwares, used open source softwares)

*Initial resource requirements (On Premise)*

x86\_64 CPU at least 4 core

at least 16GB RAM

35GB disk (it may grow as the amount of logs increase)

64-bit Linux, Windows, or macOS - 64-bit JDK 1.8 or above

*Availability and platform support*

*For development*

Cloud API - On Premise API - Java SDK is available

*Integration with other products*

Precognox TAS platform

# Precognox TAS Search Log Analyzer



## The goal of the tool

The basic idea of developing TAS Search Log Analyzer was to create an analyzing tool that provides the user information about your search log and search history. It gives the user actionable insight with special emphasis of search expressions, their frequency and efficiency.

## Implementation

TAS Search Log Analyzer works within the Precognox TAS platform.

## Requirements

- Compatibility with the TAS Search and Thesaurus Manager within the Precognox TAS platform.
- Availability to popular browsers, intranet / internet interface
- User Access Restriction (Authentication / Authorization)
- Support for multilingual subtitles (automatically detected based on browser language setting)
- Available through API for other IT systems
- Storing search logs for multiple search instances at any time
- Search terms and other search event features, e.g. date, user ID, data filtering by search mode
- Filtering and listing the matching search results in a configurable order
- User Access Restriction (authentication / authorization)

## Features

Through using Search Log Analyzer it is possible to analyze the launched searches and hereby it enables the user to utilize the TAS Search more effectively. Search Log Analyzer gives the opportunity to examine the top searches, the frequent searches or to compare the top searches in a given period, therefore the user can get a deeper insight in search trends.

The userfriendly UI provides the user many possibilities: sorting by date, number of searches or visualize the results or comparisons.

The dashboard indicates what and how (compound words) the users search in TAS Search UI. It shows the followings:

- search expressions: list of expressions searches launched with
- number of searches in (reference) date range: list of expressions are in descending order by number of searches
- filtering options: Raw / Expressions / Words, trends, date range, number of searches, change the order of results (sorting by parameter)

The screenshot displays the LOG ANALYSER interface. On the left, there are filter options for 'Raw', 'Expressions', and 'Words', with 'Expressions' selected. Below this are sections for 'Search query', 'Reference date range' (04/12/2018 - 04/18/2018), 'Date range' (04/12/2018 - 04/18/2018), 'Number of searches' (min 100, max 2000), and 'Collection filter' (Collection A, Query type 1-5). The main table shows the following data:

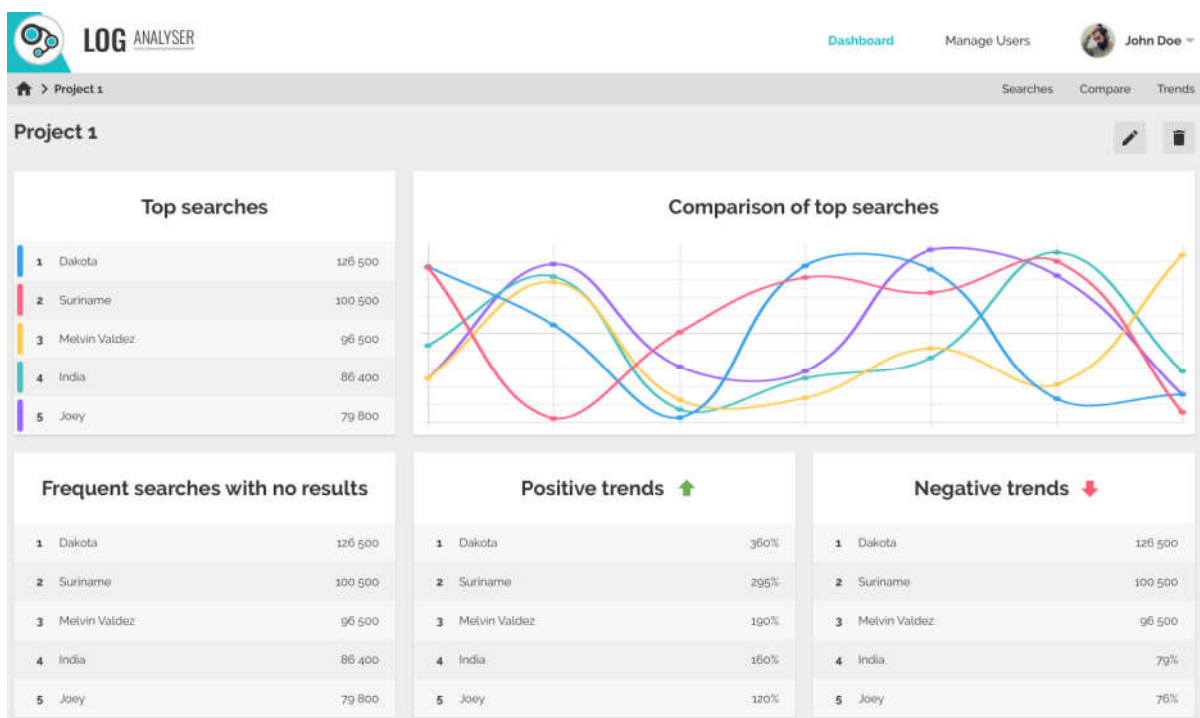
Expression	Number of searches in reference date range	Change	Number of searches in date range
Samson	5000	27.2% ↑	1000
Honey	5000	27.2% ↑	1000
Batman	5000	27.2% ↑	1000
Ginger	5000	27.2% ↑	1000
Fluffy pumkin	5000	27.2% ↑	1000
Precious	5000	27.2% ↑	1000
Marley	5000	27.2% ↑	1000
Jasper	5000	27.2% ↑	1000
Maggie	5000	27.2% ↑	1000
Harriett Reid	5000	27.2% ↑	1000
Sam	5000	27.2% ↑	1000
Sophie	5000	27.2% ↑	1000
Daisy	5000	27.2% ↑	1000
Connor Williams	5000	27.2% ↑	1000
Willie James	5000	27.2% ↑	1000

- top searches: The list of expressions are in descending order by number of searches (in the last 30 days). The list shows the expression and the number of searches for expression.
- frequent searches with no result: The list of expressions which has 0 search results are in descending order by number of searches (in the last 30 days). The list shows the expression and the number of searches for expression.
- positive and negative search trends: In the trends the base of the comparison is the last 30 days, and the 30 days before. In case of positive trends the expressions are listed in descending order by percent of change. In case of negative trends the expressions are listed in ascending order by percent of change. The list shows the expression and the percent of change for expression. The sign of the values are are not displayed as a



minus/plus sign but with down/up arrows and colors. The biggest value is the biggest positive value and the smallest value is the biggest negative value. If the system does not have 60 days history, the history should be split into 2 pieces, and it should be compared.

- If the number of searches was 0 in the Reference date range, then we show “X new searches” instead of % value in the “Change” column (since % of change can not be calculated in this case).
- the comparison of top searches: Comparison of the Top 5 searches in “Top searches” box. Static graph with 30 days of data with one day resolution.



Technical description (technical needs, integration with other softwares, used open source softwares)

*Initial resource requirements (On Premise)*

- x86\_64 CPU at least 4 core
- at least 16GB RAM
- 35GB disk (it may grow as the amount of logs increase)
- 64-bit Linux, Windows, or macOS - 64-bit JDK 1.8 or above

*Availability and platform support*

*For development*

- Cloud API - On Premise API - Java SDK is available

*Integration with other products*

- Precognox TAS platform
- Tableau
- RapidMiner
- Power BI

## Precognox TAS Enterprise Search



## The goal of the tool

The basic idea of developing TAS Search was to help identifying and enabling all content across the enterprise to be indexed, searched, and displayed to authorized users.

## Implementation

TAS Search is a Solr Cluster based enterprise search engine within the Precognox TAS Platform.

## Features

### TAS Search

- identifies and displays (highlights) the results of the search.
- provides linguistically relevant answers. Not only character-based matching is given but the linguistic features of the text are also considered whether the language is Hungarian, English or another one.
- helps in search through using industry specific synonym dictionary, therefore the search phrase can be found in multiple forms (ie. bike - bicycle), which is supported by the TAS Thesaurus Manager.
- produces very fast responses.

## Searching process details

The user interface is widely customizable – based on the requirements of the customer and could be used by multiple users simultaneously.

During the usage of the UI the user can choose between simple search and advanced search functions. Simple search is for a quick lookup, while advanced search is optimized for more sophisticated and detailed searches.

### Operators you can use in a search:

- Boolean operators: for intersections, for unions and to exclude specific documents
- Grouping terms: to form sub-queries
- Wildcard searches: single or multiple character wildcard searches

- Fuzzy searches: fuzzy searches discover terms that are similar to a specified term without necessarily being an exact match
- Proximity searches: a proximity search looks for terms that are within a specific distance from one another.
- Range searches: ranges can be specified for date, numeric or string fields.

We provide a Help documentation with details and examples to understand how to use these operators.

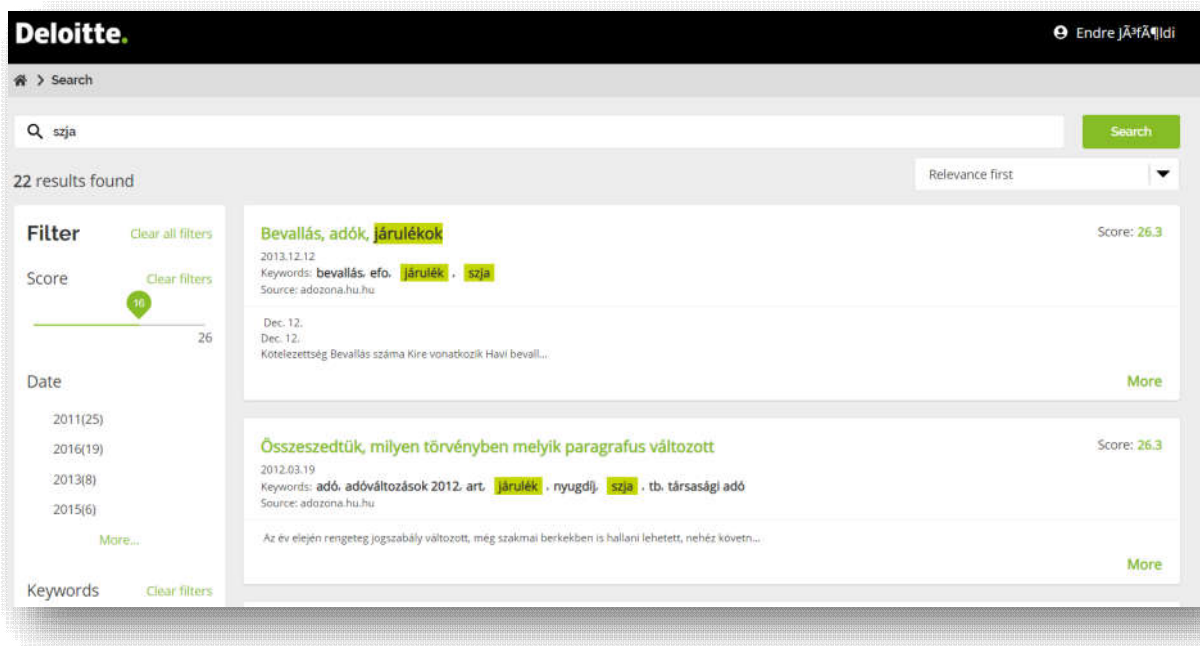
The user can see all the available details of a search result. For example:

- Title
- Source of the result
- Creation date
- Modification date
- Relevant parts of the document with the searched term highlighted
- More detailed preview of a document
- Or any other details which are available in the searched documents

## The look of TAS Search

TAS Search GUI can be created within the confines of TAS Platform (TAS Cloud service) or On Premise (locally installed). The look of the GUI can be made using the corporate identity of TAS Platform or we can build a unique search GUI. The visualization of search results and other part of the search user interface are also configurable. The particular solution depends on the customer's needs.

We have made a DEMO UI where you can get insight of how the searching user interface looks like and how it works. The given example, "Recipe Search" is able to let you launch searches among recipes (by an ingredient or food name) and displays you the results – your search expressions may be Hungarian or English.



Enterprise Search GUI customized to customer's design

## Technical features

- Simultaneous searches in great volume are feasible even with different access-level
- Integration with Basis Technology industry leading Rosette Name Matching in 18 languages (English, German, French, Spanish, Portuguese, Russian, Hungarian, Greek, Italian, Japanese, Pashto, Persian, Arabic, Chinese (Traditional and Simplified), Urdu, Thai, Korean)
- Integration with Rosette Entity Extractor to identify the entities (names, date, phone numbers)
- Automatic indexing structured and unstructured data
- Indexing solutions to standard formats: RTF, DOC, DOCX, XLS, XLSX, CSV, TXT, HTML, XML, PDF, PS
- Special indexing possibility for further file formats
- Availability to other data processing systems via API
- Document access restriction with LDAP-compatible access management

## Additional features

- We offer industry leading solution for person identification, as we can match names with misspellings, aliases, nicknames, initials, names in

different languages (the list of available language is in the Technical features)

- We can recognize named entities in the texts, such as persons, locations, organizations.
- Several after-search filtering possibilities are at hand. These are either based on patterns found in the results (such as frequent words) or information stored in specific fields (ie. date, source, author etc).
- The search box can have an in-built keyword suggestion, based on either your search history and/or the topic you are looking for
- You have full control over access rights to your data. You can assign different levels of access rights to the users of your search system. The user can be informed if there is / are other relevant search results but the appropriate information is not available with the specified access right.

Technical description (technical needs, integration with other softwares, used open source softwares)

#### *Initial resource requirements (On Premise)*

- server
- Docker compatible Linux
- 16 GB RAM
- 20 GB disc
- x86\_64 CPU at least 4 core

#### *Supported languages*

- All languages are available

#### *Integration with other services*

- TAS Platform, TAS Thesaurus Manager, TAS Search Log Analyzer

#### *Availability*

- TAS Cloud (authentication / authorization)
- TAS On Premise (authentication / authorization)
- Standalone On Premise (without authentication)



## Precognox TAS – Data Collector



## The goal of the tool

Data Collector—as part of the TAS platform—is able to collect Internet data content in a structured format so as to make this content available for information systems or for further processing and analysis.

We have developed Data Collector to provide advanced and flexible solutions working with large datasets as a way to ensure advantages for your business.

## Details of the data collection process

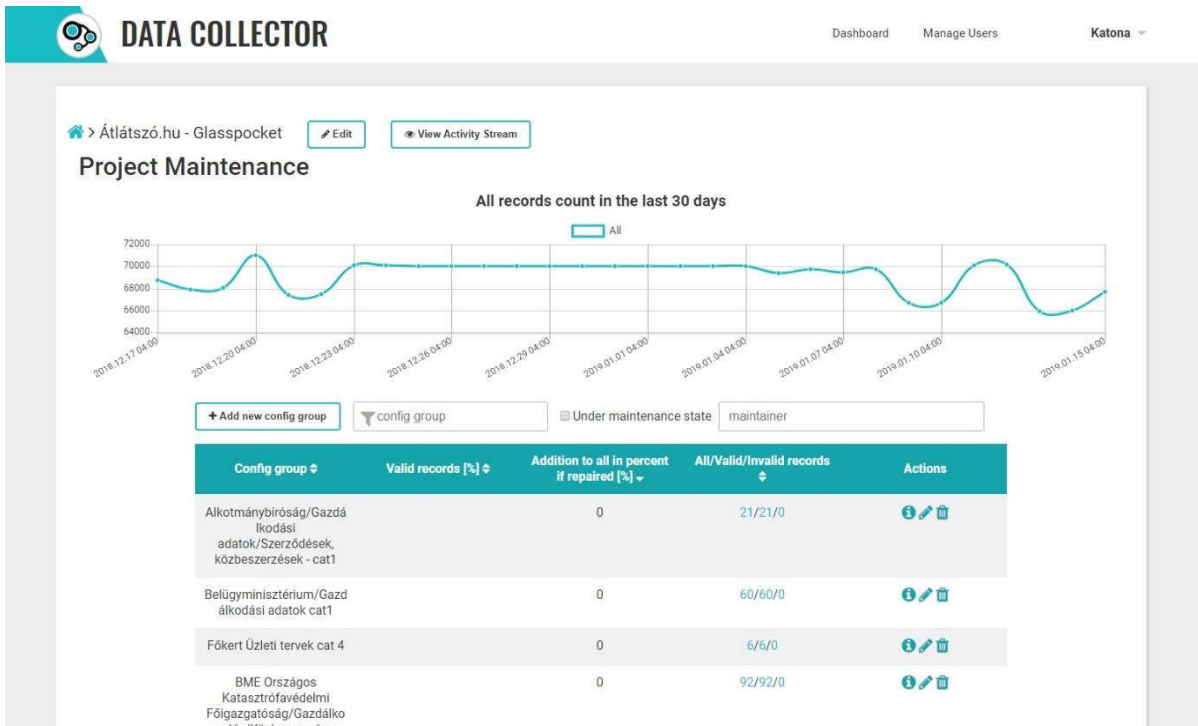
The data contained by the specified sites or documents is collected with respect to the following details:

- Data Collector is able to extract the visible data, metadata (tags, picture description) or pagination from a website.
- Sites, subpages, login-required pages, even hierarchical sites or pages with a slideshow component or with multilingual content also cause no problem for Data Collector.
- When data is recognized as hidden, we offer a screenshot solution (the original exact look of the data).
- In some cases it is forbidden—by robots.txt—to collect data. We respect this; however, this data is also possible to collect.
- We can extract texts from a lot of different documents and image formats (PDF, spreadsheet, diagram or image file formats).

We are prepared to produce and deliver any required output format, even ones that require software development.

## The look of TAS Data Collector

TAS Data Collector GUI provides you with the opportunity to constantly monitor the data flow.



Example of maintenance (monitoring) by recurring data collection

Technical description (technical needs, integration with other softwares, used open source softwares)

*Initial resource requirements (On Premise)*

- x86\_64 CPU at least 4 core
- at least 16GB RAM
- 35GB disk (it may grow as the amount of logs increase)
- 64-bit Linux, Windows, or macOS - 64-bit JDK 1.8 or above

*Availability and platform support*

*For development*

Cloud API - On Premise API - Java SDK is available

*Integration with other products*

- Precognox TAS platform
- Tableau
- RapidMiner
- Power BI

*Output format*  
Default: JSON