

# Text Analyzer



## What is Text Analyzer?

*Text Analyzer*, *TextAn*, is an innovative software component that allows multilingual communication between users and systems using unstructured text.

## How to get Text Analyzer?

If you are interested about how *TextAn* could integrate into your system please contact us at [prodaja@capabilis.hr](mailto:prodaja@capabilis.hr)

# Text Analyzer

## *Who is Text Analyzer meant for?*

---

*TextAn* software component is meant for everyone who is developing software applications or services that rely on any form of textual communication with the end-user (e.g. e-mail, SMS, instant messaging ...).

## *What does Text Analyzer do?*

---

*TextAn* recognizes words and phrases from free text in digital form. Furthermore, it is able to discover word meaning regarding to context in which the discovered words or phrases were written, giving service providers the ability to tackle real life situations originating from real users.

*TextAn* handles unstructured digital text (e.g. e-mail, SMS, instant messaging, output of speech-to-text conversion ...), analyzes it, discovers its meaning and transforms it in structured text suitable for communication with system, application, service or, for example, human administrator. This functionality enables the user to issue a service request or an application command by using free text, the same form, as he/she would send toward a real person.

*TextAn* can also be used as unified access point for various services and applications, where requests for specific service/application are routed based on the free text entered by the user.

*TextAn* can be a valuable addition to lawful interception, parental control and similar solutions, since it enables control of SMS, e-mail and MMS communications, as well as additional component for speech-to-text solutions.

# Text Analyzer

*TextAn* can be used for data mining purposes on both dynamic and static content (e.g. forums, social networks...).

Taking into account the fact that *TextAn* removes the need for formalisms and makes service access more leisure, it enables greater penetration of ICT services based on text communication. In general, *TextAn* can be used whenever there is a need for recognizing words/phrases and their meaning, on both unstructured and structured digital texts.

## *What makes Text Analyzer different from similar solutions?*

---

*TextAn* eliminates common drawbacks in existing similar solutions such as:

- Many current applications and services rely on text communication (SMS, e-mail...) due to its simplicity, availability and pricing
- Many current services require technical and grammatical literacy from the end user, which may not always be the case
- Current text communication shows significant patterns: common typos, common grammatical errors, missing uppercase letters and punctuation, numbers and abbreviations used frequently and similar.
- In many cases application and services that rely on text communication require formal and structured text syntax
- Application and services with stronger formalism/structure requirements have fewer acceptances from end users than services with lower requirements.
- Additional problems may arise from services and applications that are rarely used so possible customers are not sure how to access them which, in turn, results in users cancelling or not using the service at all.

# Text Analyzer

- Native application solutions are not always applicable and will rarely focus on all market requirements.

## *What are the main advantages of Text Analyzer?*

---

- Free-form text input without worrying about syntax or grammatical errors.
- Handles dialects, numbers, abbreviations and similar.
- Free-form text input is suitable for less technical oriented users.
- Free-form text input is suitable for services that are rarely used.
- Communication with the user via e-mail, SMS, MMS or HTTP (and WAP).
- Communication with the user in his/her language.
- Recognizing unstructured text and transforming it into a service request/command.

## *What are the elements of Text Analyzer?*

---

*TextAn* is a commercial software component suitable for integration into new or existing ICT services that handle digital content. It contains the following modules:

**Communicator** – interface for communication with the environment. The environment is dependent on the application/service in question and its architecture. This module communicates with other *TextAn* modules through HTTP/HTTPS.

**Extractor** – module for extracting non-textual characters that do not affect the word/phrase meaning. The module is language independent.

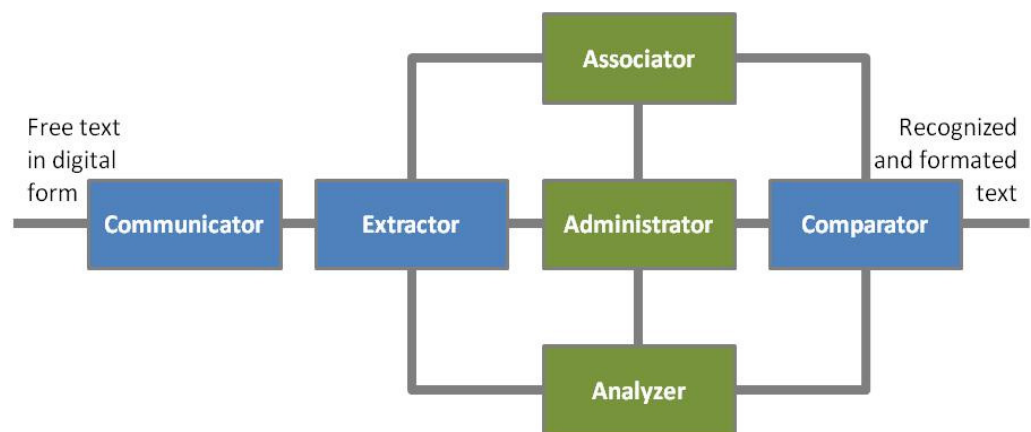
**Analyzer** – module that performs grouping and routing of digital content based on previously discovered meaning of detected character arrays. The module is implemented using ternary tree, whose dimension is specifically set for the given

# Text Analyzer

service/application. Full module functionality is achieved when learning over the whole known dataset.

**Administrator** – module that contains control and reporting functionalities. Some of the functionalities are modeling the learning dataset, running learning functions, overview of communication channels, resolving conflicts and insight into statistical data.

**Associator** – module for associative meaning detection from character arrays in text. It is implemented by using artificial neural networks and its full functionality is achieved when learning over the representative dataset. The embedded generalization ability enables meaning detection on a previously unknown data. This ability can be further fine tuned to reflect the needs of the service/application in question. Analysis results show the meaning of each detected and recognized character array in the original text.



**Comparator** – Module for assessing the analysis accuracy and routing analysis results. If correctly recognized, the results are forwarded to the application/service, while any suspected errors are weighted with certainty factor and forwarded to administrator and, if necessary, the application/service. Administrator can, based on the error notification, perform manual or automatic update of the modules in order to resolve the error in recognition.

# Text Analyzer

## **Capabilis d.o.o.**

---

Capabilis d.o.o. was established in 2011 after a successful completion of a technology project co-financed by the Croatian Institute of Technology.

The company was registered for providing services in field of information and communication technology.

Company's research and development efforts aim to create new ideas, applications and services that bring new value and deal with real problems in the field of IT market.

If you have an issue in the field of electronic communications such as:

- Questions that should be answered?
- Ideas you want to realize?
- Software problems you have to deal with?

**Then Capabilis is your best partner!**

### **Contact**

**Capabilis d.o.o.**  
Dragutina Golika 63  
10000 Zagreb  
Croatia  
[www.capabilis.hr](http://www.capabilis.hr)  
[info@capabilis.hr](mailto:info@capabilis.hr)