# semantha®

## the semantic platform

AI with **natural language understanding** to speed up document processes.

**#BeADocumentHero**

hatched by
**THINGS THINKING**

# Find relevant content independent of the wording.

Our artificial intelligence semantha® understands natural language regardless of the choice of words. She is able to read large amounts of text and finds relevant content.

The car was very fast.

A vehicle drove with high speed.

# 1000 documents? No problem!

semantha® finds the content you are looking for in piles of unstructured data.

# Flexible output or transfer.

semantha® is a flexible platform and she can export all results in different formats (e.g., xlsx, docx, xml etc.). Various user interfaces and a standard REST interface are available for this purpose.

PDF

CSV

JSON

www.semantha.de

semantha®
the semantic platform

hatched by
THINGS
THINKING
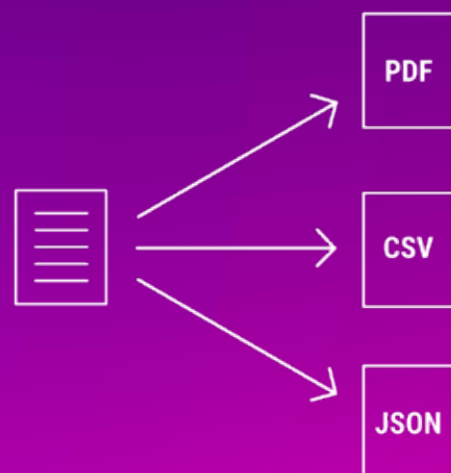
FORBES magazine describes us with the words „a brain with **artificial intelligence (AI)**". After more than 14 years of research in the field of language processing and AI, we founded the AI startup thingsTHINKING GmbH in Karlsruhe in 2017. We have won numerous awards and serve customers in the industry sectors car manufacturing, chemicals, pharmaceuticals, law, construction, auditing, and insurance. Our platform **semantha**® is able to compare documents at the level of meaning and does not stop at a word-by-word comparison.

**Marc takes the train to work.**

**To get to the office, Marc takes the railway.**

A conventional search would not find a match in this example. However, semantha® says that these two sentences are very similar in meaning. A human would reach the same conclusion.

semantha® uniquely combines different AI elements to understand language, which allows semantic analyses on unstructured data (text). She can read, understand and compare text regardless of their wording. The customer determines what semantha® pays attention to during the analysis. Document driven processes can be automated and / or accelerated. The added value is greatest where people have to read a lot of text to perform tasks or companies have to manually extract information from unstructured data.
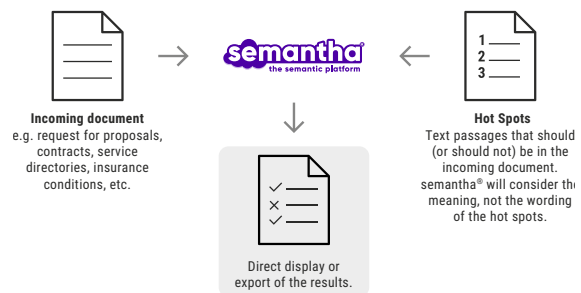
Currently, we target companies which have document-driven processes and want to improve them in terms of quality, repeatability, acceleration, and automation.

We offer a SaaS solution (Software as a Service) and start with a four week evaluation phase. Here, we work hand in hand with subject matter experts to test use cases and develop a common road map.
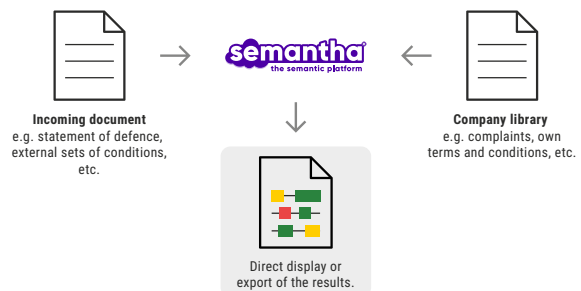
This is followed by an annual subscription and the customer can individually integrate the platform into their own IT infrastructure using semantha's REST API.
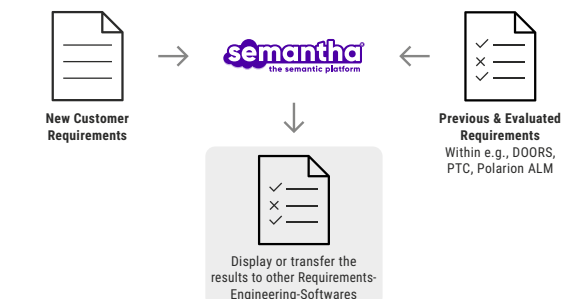
## Example Applications

**A** **Define content that is important and check incoming documents.**

**Incoming document**
e.g. request for proposals, contracts, service directories, insurance conditions, etc.

**Hot Spots**
Text passages that should (or should not) be in the incoming document. semantha® will consider the meaning, not the wording of the hot spots.

Direct display or export of the results.

**B** **Compare multiple documents with each other and find the matches immediately.**

**Incoming document**
e.g. statement of defence, external sets of conditions, etc.

**Company library**
e.g. complaints, own terms and conditions, etc.

Direct display or export of the results.

**C** **Check new incoming customer requirements with previous and already evaluated requirements**

**New Customer Requirements**

**Previous & Evaluated Requirements**
Within e.g., DOORS, PTC, Polarion ALM

Display or transfer the results to other Requirements-Engineering-Softwares

You want to discuss how semantha can improve your document-driven processes?
Just email us: **info@semantha.de**

hatched by
**THINGS THINKING**

## Detailed Information

semantha® processes unstructured documents in three consecutive steps:

- Step 1 determines semantic similarities between text passages.
- Step 2 extracts data points from documents and thereby structures them.
- Step 3 uses the data points to draw logical conclusions.

Various AI elements are used in combination for this purpose. We are convinced that many use cases cannot be solved with just one AI element, but that it is the combination of different AI ingredients that leads to success.

Alternative software solutions without AI can hardly or only with difficulty compare different formulations with each other and are often not able to compare text documents independently of concrete formulations.

In the following, we explain to which AI elements of the periodic table the processing mechanisms are assigned.

On the input side, semantha® must first read in a document and transfer it into an internal data structure. If the data is available in a semi-structured form, e.g. in a Word document, semantha® uses the existing document structure to divide chapter sections by headings, for example. If the input data is available in a less structured form, e.g. in a PDF document, semantha® not only accesses the textual content, but also visual properties (element Ir - Image Recognition). For example, tables have to be recognized and evaluated separately or texts in illustrations as well as in headers and footers have to be skipped (i.e. not assigned to the continuous text); at the same time, visual properties of the text are used to draw conclusions about the extended document structure (e.g. recognition of headings and marginalia). This kind of preparation is necessary because real-world use cases are not limited to „technically beautiful" file formats. Structured file formats, such as XML dialects, can also be used as data formats.

The next stage of processing is the determination of semantic similarities (Element Lu - Language Understanding). Text passages are compared with each other and checked for overlapping content. In this way, identical or closely related text passages can be identified in order to check a document for previously defined hot spots, for example. It is also possible to compare two documents (or two versions of a document) directly with each other.

At this stage, we like to emphasize two features of semantha: Firstly, semantha® has a pre-trained language comprehension. Therefore, there is no need for application- or customer-specific training and semantha can start right away with processing your documents. Since language understanding hinges on the language itself, we provide different language models for different languages (for example, English and German). Yet, these models are not static but can be modified to better suit the use case. The easiest way is tweaking semantha's configuration. If a domain uses a very technical language or heavy jargon, we can also tailor the language mo-

del or train a full language model specifically for a single use case. The second characteristic of semantha®'s text analyses is its independence from wording: the language model enables semantha® to capture texts at the level of meaning, which enables her to analyse documents at the level of meaning. The actual wording therefore plays a subordinate role which, for example, greatly simplifies the definition of the above mentioned hot spots.

If it is sufficient for a use case to identify the hot spots, they can be highlighted in the document and including the references can be noted in the document and reported back to the user. The data basis on which the matches are determined can be influenced by the user. semantha® also offers the possibility of providing feedback on individual results (e.g. a user note „This text passage fits better / best") (Element Lt - Knowledge Refinement). The adaptability of the language model described above is another feature of semantha® that implements the element Lt.

A simple use case is the review of contracts for no-go clauses, where semantha® identifies the hot spots (the corresponding text passage can be highlighted in red for the user).