Sk Skolkovo Sk IT ABBYY INFOPOISK LTD. ABBYY®

Product features

The revolutionary features of ABBYY Compreno lay in the fundamental approach. ABBYY began research in this area in 1995, i.e. 17 years ago. At present more than 350 people are working on this project. ABBYY's investment in the project for spring of 2012 reached over 65 million USD.

The technology consists of three main elements.

1 The first element is a universal semantic hierarchy. It is based on an idea that people, even speaking different languages, think with identical notions because they live in similar conditions. The structure of the notions is universal. It can be presented in the form of a tree, where the thick branches are general notions (for example, "tableware") and the thin branches are more specific conceptual notions (like "knife", "spoon", "fork", etc.). Such hierarchy designed as a tree can be used for transition from one language to another.

2 The second element is syntax describing how the notions are interconnected inside one or several sentences.

3 The third important component is statistics that enables improving the quality of syntactic or semantic analysis.

Currently there is no such a universal technology in the world that allows solving so many applied problems related to languages and text analysis. The universal notion system includes over 129,000 lexical units of Russian language and 130,000 lexical units of English language. The description of such scope is unique and has no analogues.

Product features

Compreno technology is a universal technological platform for various linguistic applications:

1. Translation from one language into another.
2. Extraction of facts and links between the objects of search or monitoring (fact extractor). For example, it might be necessary to understand all facts related to a certain personality, all events that occurred at certain moments/periods of time or at any place, etc.
3. Smart search: search by meaning rather than by keywords, as well as the possibility to obtain answers to the questions asked in natural language (for example: "What does Ivan Ivanovich Bobrov possess?").
4. Multilingual search, i.e. when in reply to a question in one language the answers in all the languages supported by the system are found.

5. Classification of documents. For example, classification of the letters that get to the company electronic mailbox, or sorting the enquiries sent to the technical support service.
6. Comparison and search of differences in the documents in order to prevent plagiarism and control the document versions (for example, versions of agreements).
7. Analysis of the tonality of texts (for example, as a result of press monitoring).