



"ATAPY attained 99.992% text accuracy in the German-Russian Dictionary (1 mistake per 8,760 symbols), and 99.997% quality for the Spanish-Russian Dictionary project (1 mistake per 31,500 symbols). They also corrected many mistakes in the source dictionary text, including typographical misprints and even mistakes in special marks that are almost impossible to detect without special programming tools and a profound knowledge of linguistics."

Anna Zhavoronkova
Project Manager,
ABBYY Software House



Development of International Computer Dictionaries for ABBYY

Electronic dictionaries and translation systems are an area of great practical importance in the ever-globalizing world. ABBYY Software House, the world leader in OCR/ICR and linguistic technologies, develops and sells Lingvo electronic dictionaries. For many years Lingvo has been known as the best English-Russian dictionary on the market. Version 8.0, supported three more languages. For the next version, it was decided to use the world's latest best-of-breed dictionaries that represent the modern state of supported languages.

ABBYY® Lingvo® The ABBYY Lingvo 8.0 product line includes ABBYY Lingvo 8.0 Multilingual Edition, ABBYY Lingvo 8.0 for Pocket PC, as well as an updated and expanded version of ABBYY Lingvo English-Russian Edition.

ABBYY Lingvo 8.0 Multilingual Edition supports eight translation directions: English-Russian, German-Russian, French-Russian, Italian-Russian, Russian-English, Russian-German, Russian-French, and Russian-Italian. This Edition of ABBYY Lingvo includes more than 40 dictionaries containing more than 2,400,000 entries.

ABBYY turned to ATAPY Software, its outsourcing partner in Novosibirsk, for digitization of two dictionaries from the list selected by the Linguistics Department. The 3-volume 1,750-page Leping German-Russian Dictionary and the 830-page Narumov Spanish-Russian Dictionary were chosen to be recognized and proofread for automatic conversion to the Lingvo database.

The highest possible text recognition accuracy was obviously a requirement. A single mistake could break the words alphabetical order and tear the word away from its paradigm. If the number of mistakes went beyond a very modest threshold, the dictionary would unsearchable. Proper interpretation of special dictionary marks was equally important for the project. They were used as field delimiters in the automatic database conversion process and had achieve 100% recognition. Special marks appeared either as text characteristics (bold/italics), as special symbols (brackets, asterisks), or as a combination of the two (e.g., italics brackets indicated a dictionary comment). Omitting a single bracket or missing italics would break the article's structure. That is why the project required both intelligent programming and a highly qualified manual effort - a true challenge for any contractor in the media service sphere.

The dictionaries were scanned and automatically recognized with an ABBYY FineReader OCR that was specially tuned for processing this material. Then a team of qualified operators proofread and cross-checked the results using the Double verification technique to ensure recognition accuracy. Double verification detected unexpected situations, such as typos in the source dictionary text, that were corrected according to the ABBYY's guidelines. In its effort to maximize automation of the proofreading work,

ATAPY developed and customized a number of in-house utilities such as Glyphica, a tool for quick input of characters that are not found on the keyboard. For Leping's Dictionary, ATAPY developed a custom converter with built-in spell- and punctuation-checking utilities that weeded out mistakes unnoticed in previous stages and finally converted the material into the Lingvo vocabulary database.



ABBYY Software House (www.abbyy.com) is based in Moscow, Russia. The company was founded in 1989. ABBYY has over 880 employees, with offices in Russia, USA, Ukraine, UK, Germany, Taiwan, Japan and Cyprus. ABBYY has developed software in the fields of artificial intelligence, document recognition, data capture and applied linguistics. ABBYY is most notable for their optical character recognition software, FineReader.

©2013-2016 ATAPY Software. All rights reserved.
ABBYY and ABBYY Lingvo are registered trademarks of ABBYY Software.
All the other trademarks are the property of their respective owners.



ATAPY Software

16 Engineernaya Street
630090 Novosibirsk, Russia
Tel. +7 383 36 39 69 9
www.atapy.com ■ office@atapy.com

ABBYY

2b/6 Otradnaya Street
127273 Moscow, Russia
Tel. +7 (495) 783 3700
www.abbyy.com ■ office@abbyy.com

