ATAPY
SOFTWARE

"ATAPY reached 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 profound knowledge of linguistics."

Anna Zhavoronkova
Project Manager,
ABBYY Software House

ATAPY Software Participates in Development of International Computer Dictionaries

Electronic dictionaries and translation systems are an area of great practical importance in the ever-globalizing world. ABBYY Software House, a 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. In version 8.0, support of 3 more languages were planned for adding; to introduce those to Lingvo, it was required to digitize world's latest best-of-breed dictionaries reflecting the modern state of the new 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 digital conversion of two dictionaries from the list picked out by the Linguistics Department. The 3-volume 1750-page Leping German-Russian Dictionary and the 830-page Narumov Spanish-Russian Dictionary were to be recognized and proofread for automatic conversion into the ABBYY Lingvo database.

Highest possible text recognition accuracy was obviously a must. A single mistake could break the words' alphabetical order and tear the word away from its paradigm. If the number of such mistakes were above even a very modest threshold, the dictionary would have become unsearchable. Adequate interpretation of special dictionary marks was no less vital for the project. They were used as field delimiters in the automatic database conversion process and had to be recognized 100% accurately. Special marks appeared either as text characteristics (bold/italics), or 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 italization would break the article's structure. This is why the project required both intelligent programming and highly qualified manual effort - a true challenge for any contractor in the media service area.

The dictionaries were scanned and automatically recognized with ABBYY FineReader specially tuned-up 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 allowed to detect certain unexpected cases, such as typos in the source dictionary text, which have been corrected according to the ABBYY's guidelines. In its effort to automate the proofreading work to the maximum possible extent, ATAPY developed and customized a number of in-house utilities. One of them was Glyphica, a tool for quick input of characters that cannot be found on the keyboard. For Leping Dictionary ATAPY developed a custom converter with built-in spellchecking and punctuation checking utilities which allowed to weed out mistakes unspotted during the previous stages and finally convert the material into the Lingvo vocabulary database.

ABBYY
SOFTWARE HOUSE

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

ATAPY Software

630090, Engineernaya Street, 16
Novosibirsk, Russia
Tel. +7 383 36 39 69 9 ▪ Fax +7 383 36 39 69 8
www.atapy.com ▪ office@atapy.com

ABBYY

Otradnaya str. 2b/6, 127273
Moscow, Russia
Tel. +7 (495) 783 3700 ▪ Fax +7 095 783 2663
www.abbyy.com ▪ office@abbyy.com