Intelligent Enterprise

Tatiana Parfentieva

Deputy Director of Linguistic Products Department

One of the largest linguistic research centres

and one of the world's leaders in OCR & Data Capture

- About 560 employees in R&D

- More than 1000 developers already using ABBYY technologies in their solutions

- More than 350 linguists, programmers, and analysts involved in ABBYY's linguistic project

- **Morphology**
  40 languages

- **Universal Semantic Hierarchy**
  70,000 language-independent semantic classes
  More than 94,000 concepts for English
  More than 82,000 concepts for Russian
  French, and German are in the pipeline

- **Syntax**
  Complete syntax model
  Semantics is taken into account

- **Statistically-based Verification and Adjustment**
  Parallel corpora:
  500M sentences for EN<->RUS
  100M sentences for EN<->GE and EN<->FR

Challenge

Large volumes of data require fast and accurate search

- To find information or documents, the user has to:
  - Use a sequence of queries
  - Manually sift through the search results

- All this requires a lot of time, though the results could be rather disappointing
  - Low recall and precision

LT challenges

- Too many ways to name the same concept

- Too many concepts can be named with the same word

- It's easy to lose important relations between words

- Valuable information available in a limited set of languages is locked by them

Innovation

- **Goals**
  - Reduced complexity of queries thanks to natural language parsing
  - High recall and precision
  - Multilingual search

- **Joint projects**
  - Enrich existing search systems (partners)
    using comprehensive and precise semantic and syntactic text analysis (ABBYY)
    combined with taxonomies and ontologies (partners)

  - Create multilingual search applications (partners)
    using language-independent semantic concepts and relations (ABBYY)