eLexicography, collaboration, bilingual corpora - all in one!

Vera Kuzmina,
ABBYY
ABBYY Lingvo.pro Dictionary Portal

- Single access point for:
  - Lookup and search in dictionary sources
  - Quires in bilingual and monolingual corpora
  - Collaboration with dictionary users
  - Text alignment
  - Term extraction
Language Research in ABBYY

- Typical tasks in lexicography and language research
  - Identify new words in different subject areas
  - Identify different meanings of a word
  - Compare usage of a word in different contexts
  - Find translation(s) for each meaning
  - Finding collocations
  - etc.

- Professional linguists and lexicographers
  - ABBYY Press Publishing House – over 25 in-house lexicographers
  - More than 100 linguists and lexicographers work in the R&D department
  - Authors and editors
ABBYY Lingvo.Pro Concept

Lingvo.Pro is a complete set of language data and technology tools that are used in the following cases:

- web-based linguists’ and lexicographers’ working environment
- language website with over 1,300,000 queries per day*
- solution for language resources management

* as of November 2011
ABBYY Lingvo.Pro Features

- **Language data**
  - Dictionaries (publishers, community)
  - Text corpora (bilingual and monolingual)
  - External sources (Web, Wikipedia, user content)

- **Tools**
  - Corpus query system
  - Alignment of bilingual texts
  - User suggestions and discussions
  - Terminology extraction (coming soon)
ABBYY Lingvo.Pro Content: Dictionaries

- More than 20 language directions
  - en<>ru, en<>de, en<>fr, en<>ch, de<>ru, fr<>ru, ...
  - other directions can be added

- Millions of entries
  - and growing ~ 1m entries a month

- Different viewing options
  - for occasional users
  - for lexicographers, linguists, translators
ABBYY Lingvo.Pro Content: Text Corpora

- Bilingual and monolingual
- More than 10 language directions
  - en<>ru, en<>de, en<>fr, en<>ch, de<>ru, fr<>ru, ...
  - other directions can be added
- Almost 1bn words
  - and growing 2m words a month
- Users can add their own corpora
Corpus Query System

- Simple search in dictionaries and corpora
- Bilingual search
- Advanced search
  - exact phrase, all words vs. any word, distance, metadata filtering, specify POS and grammar features, match case...
- Fuzzy search
- Semantic search
ABBYY Lingvo.Pro

- Online demo
Corpus Query System: Bilingual Search

construction

Aufbau

Examples found: 22

The latest neutrino telescope, Ice Cube, currently under construction in 2006, will employ 4200 detectors in 70 shafts sunk deep into the Antarctic ice.

Das neueste Neutrino Teleskop, das 2006 noch im Aufbau befindliche Ice Cube, wird 4200 Detektoren in 70 Schächten umfassen, die tief in das antarktische Eis getrieben werden.

As a rule the construction and maintenance of energy infrastructure should be subject to market principles.

Für den Aufbau und die Unterhaltung der Energieinfrastruktur sollten in der Regel marktwirtschaftliche Grundsätze gelten.

Information on the construction of any vector systems or sequences of genetic material used with the carrier DNA shall be available.

Informationen über den Aufbau von Vektorsystemen oder die mit der Träger-DNA verwendeten Sequenzen von genetischem Material haben vorzuliegen.

Welcomes the Commission communication, since the future development of mobile and wireless communications is the cornerstone of the construction of an information society geared to all citizens;

begrüßt die Mitteilung der Kommission, weil die zukünftige Entwicklung der Drahtlos- und Mobilkommunikation einen Grundpfeiler für den Aufbau einer Informationsgesellschaft zum Nutzen aller Bürger darstellt;
ABBYY Aligner

- ABBYY Aligner is a professional tool for aligning parallel texts and making bilingual corpora
  - high quality of alignment
  - automatic processing of large amounts of documents
  - easy to use
  - Integrated in ABBYY Online

ABBYY Aligner

Address bar and processed document language

Toolbar for opening existing project or for creating new ones, exporting results, working with segments etc.

Underlining of spelling mistakes

Source Text

Translated Text

Edited segment number and total number of rows in the document

Highlighting of aligned segments with low match probability or empty rows
Terminology Extraction (coming soon)

- Allows extracting potential monolingual or bilingual terms from text corpora
- Ideal for making an initial pool of term candidates for specialized dictionaries
Community Discussions

● **Now**
  - Forums at Lingvo.Ru ([http://forum.lingvo.ru](http://forum.lingvo.ru)) where users discuss language issues
  - Adding user translations on Lingvo.Pro

● **Coming soon**
  - Discussions of dictionary entries
  - Immediate community feedback on dictionary entries
  - “Q&A” service
Conclusion

Lingvo.Pro is an all-in-one workspace for language researchers that offers:

- Dictionary entries and translations from authors, publishers, and users
- Large monolingual and bilingual text corpora and a corpus query system
- Advanced linguistic tools
- Collaboration and discussions
Questions?

Contact information:

Vera Kuzmina  
Head of Strategic Partnership Development Group  
ABBYY Headquarters  
Phone: +7 495 783 37 00 ext. 2766

www.abbyy.com

Visit www.lingvo.pro!