
Electronic Dictionary and Dictionary Writing System: how this duo works for dictionary users' needs (ABBYY Lingvo and ABBYY Lingvo Content case)

Anna Rylova
ABBYY

The main idea we present in this paper is that using special markup in dictionary writing system and having appropriate functionality in electronic dictionary software we can achieve new results in satisfying most important needs of dictionary user. We describe the core functionality of the ABBYY Lingvo Content dictionary writing system and some features of ABBYY Lingvo electronic dictionary software that presents the dictionaries made in DWS. Then we show how the dictionary data can be used in text translation scenario and how DWS and electronic dictionary work together to meet the user's needs in translation and text analyzing.

Besides the core functionality ABBYY Lingvo Content DWS includes:

- *embedded in DWS interface user-friendly entry filtration system. Lexicographer doesn't need to know any special query language – just tick boxes in filtration window tabs;*
- *also embedded in interface tool for dictionary comparison and merge;*
- *visual markup of changes – you can always compare any two versions of dictionary entry and see what was added, deleted, changed or restored to their earlier versions.;*
- *possibility of working with many dictionaries (2 and more) in one window, editing their entries simultaneously.*

ABBYY Lingvo electronic dictionary has been developed since 1989 and nowadays it is used by 7 million users worldwide. One of the basic dictionary user's need is to find the appropriate translation for the word he met in a text (text reception) or translate a word from their mother tongue to a foreign language. Here we will describe how the electronic dictionary software works to satisfy the text reception need. One of the most challenging task for a dictionary producer – to help the dictionary reader find a good translation and all the relevant information about the word. This task can be done well if a lexicographer puts relevant markup for a dictionary entry in DWS and electronic dictionary has a proper functionality to process this markup and a good interface to show the result of this processing to dictionary user.

1. Introduction

ABBYY Lingvo Content is a dictionary writing system intended for compiling dictionaries, glossaries, encyclopedias, and other types of reference materials. It is part of the Lingvo platform, which, apart from the DWS, includes a number of content conversion and dictionary publishing tools, enabling the publication of dictionaries in electronic format (viewable in ABBYY Lingvo, electronic dictionary software for desktop computers and mobile devices), on paper, and online. Online dictionaries can be accessed via ABBYY Lingvo Server. The company has already created, on its own or in collaboration, a total of 40 dictionaries, which have been published both electronically and as paper volumes.

The ABBYY Lingvo Content dictionary writing system has been in development since 2003 and is now used by ABBYY's in-house lexicographers and by some publishing houses, teams of compilers, and individual authors.

2. Core ABBYY Lingvo Content Functionality

2.1. Basic Features

The basic ABBYY Lingvo Content features, intended to facilitate the dictionary compilation process and to automate the usual lexicographic chores, have been designed based on a careful study of the practical needs of lexicographers and editors, as well as on ABBYY's own experience in creating dictionaries and encyclopedias. These features include:

- managing the structure of the entries and automatic renumbering of entry elements, such as senses and homonyms
- automatic cross-references update
- spell-checking the text of the entries and validating their structure.

2.2. Some Special Features

One of the basic principles we put in the development of ABBYY Lingvo Content DWS is that lexicographer shouldn't need any special computer needs to work with the system and be as unassisted in using it as it is possible. So we can say that some special features of the system include:

- embedded in DWS interface user-friendly entry filtration system. The lexicographer doesn't need to know any special query language – just tick boxes in filtration window tabs;
- also embedded in interface tool for dictionary comparison and merge;
- visual markup of changes – you can always compare any two versions of dictionary entry and see what was added (highlighted with green colour), deleted (highlighted red) or changed (highlighted blue);
- possibility of working with many dictionaries (2 and more) in one window, editing their entries simultaneously.

2.3. Filtering

When working on a dictionary, a lexicographer or the head of a group of lexicographers often needs to examine a selection of entries that meet certain criteria. For example, one may wish to see the list of all phrasal verbs in the dictionary, or all entries that contain an idiom with a certain word, or all entries that contain a certain number of senses, or the entries marked by a lexicographer for future revision. One can easily imagine numerous other scenarios when a selection of entries may be needed. ABBYY Lingvo Content provides lexicographers and editors with a filtering feature which enables them to retrieve data quickly and easily without the use of a specialized query language. Instead, they can simply select the required filtering criteria in the filtering dialog box. The entries obtained in this manner can then be saved, either as a batch or as a separate dictionary, and a user may be assigned to edit them.

2.4. Multiuser Work

Based on a client /server architecture, ABBYY Lingvo Content supports multi-user concurrent access and is suitable for large dictionary-making projects. The lexicographers may be physically located anywhere in the world and work on the same dictionary together. The entries that are being edited or have been assigned to specific lexicographers will be marked accordingly in the word list, which is visible to the entire team. The lexicographers and editors may work on a dictionary in online mode, in which case all new texts are immediately sent to the central server, or offline, in which case all texts are created and stored locally and then uploaded to the server.

ABBYY Lingvo Content logs all changes made to the dictionary entries. Users of the system can easily find out which lexicographers worked on which entries during a given period of time, or refine the search criteria to see which entries have been changed, deleted or added, or in which entries the headword has been edited, or which entries have been restored to their earlier versions.

Version history is available for each entry: the segments of an entry that have been edited, deleted or added are highlighted in different colours. It is also possible to roll-back an entry to an earlier version.

All content created by the lexicographers is securely stored on the server with regular back-ups. It is possible to view the current version of the dictionary at any moment.

2.5. Workflow

A status can be specified for any entry. Each dictionary has its own set of statuses, which indicate the progress of the work. For example, a dictionary may have three entry statuses: 1) "entry has been created by lexicographer", 2) "entry has been reviewed by editor", 3) "entry has been proofread", 4) "entry is ready to be published". It is always possible to find out how many entries have a certain

status. For example, if 95% of entries are “ready to be published”, this means that the dictionary can soon be released on paper, on CD-ROM or other electronic media, or can be made available online.

2.6. Consistency of Content

For each dictionary, its overall structure and the structure of its entries may be specified. The structure of entries generally determines the order of entry sections and their “nesting”. The specified fields will be displayed on a toolbar. For the sake of convenience, the toolbar displays only those fields which are allowed in the entry section indicated by the cursor. Thus, the lexicographer only needs to click on an allowed field on the toolbar, without the need to open and scroll large lists.

It is also possible to specify a list of labels to be used in the dictionary. Then the system will either validate a label as it is typed in by the lexicographer, or prompt the lexicographer to select the appropriate label from the general list. Editing a label or its wording in the general list will consistently change this label throughout the dictionary.

One of the important features is automatic cross-reference update. If any word sense was moved in the entry (for example, from position 1a) to position 3b)) all the references to this entry will stay valid and the numbering in reference name will be automatically updated. If the entry or a word sense was deleted the system will issue a warning and show all the entries that are linked with the deleted entry. Then lexicographer can delete the references manually or automatically.

2.7. Merging and Comparing Dictionaries

When working on dictionaries for the same language combination, their word lists may be compared. The comparison tool has an intuitive visual interface. A lexicographer may expand a general dictionary by comparing it against specialist dictionaries. The result of such comparison will be a selection of entries not found in the general dictionary, which can either be edited and then added into the general dictionary or added into the general dictionary in its entirety. For each new entry thus obtained, its original source will be indicated.

ABBYY Lingvo Content allows merging dictionaries and selections of entries. The user can view several dictionaries or selections of entries in one window: the user sees the combined word list with the source dictionary or selection of entries indicated next to each item. Using this viewing mode, lexicographers can not only add new entries to their dictionaries but also create and edit entries for several dictionaries simultaneously (for example, lexicographers can work on a comprehensive and pocket edition at the same time).

2.8. Publishing Dictionaries

Dictionaries created with ABBYY Lingvo Content can be easily published on paper, on electronic media or on the Web. It only takes several minutes to publish a dictionary electronically: all the dictionary data are exported into a format that can be read by the Lingvo dictionary viewer, which is available for desktop computers and mobile devices. ABBYY has been working on Lingvo since 1989, and it currently has 7 million users worldwide.

If a dictionary is to be printed on paper, it is exported from ABBYY Lingvo Content into a publishing system via XML, RTF or DOCX file generated from the system.

To publish a dictionary on the Web, ABBYY Lingvo Server may be used, which is a Web service that enables searches across various types of reference sources, including dictionaries and encyclopedias. ABBYY Lingvo Server can be accessed over the Internet or over intranet networks.

3. Using Dictionaries in ABBYY Lingvo Electronic Dictionary Software

One of the basic dictionary user’s need is to find the appropriate translation for the word he met in a text (text reception) or translate a word from their mother tongue to a foreign language. Here we will describe how the electronic dictionary software works to satisfy the text reception need. When the

user see some new or unknown word in a text he will try to look it up in a dictionary and find an appropriate translation from the big dictionary entry with many translations, examples, synonyms and other information that is usually included in dictionaries. One of the most challenging task for a dictionary producer – to help the dictionary reader find a good translation and all the relevant information about the word. This task can be done well if a lexicographer puts relevant markup for a dictionary entry in a DWS and the electronic dictionary¹ has a proper functionality to process this markup and a good interface to show the result of this processing to dictionary user.

ABBYY Lingvo electronic dictionary software helps the user in text translation and analysis because it has pop up translation tool. When the user meets an unknown word in a text (even if the word is not in the initial form) he points to the word with mouse cursor and see the short pop up translation of the word, taken from the electronic dictionary. If the user clicks on a translation, he will see the full dictionary entry. This works well with the words in 40 languages and with German compound words – a pop up translation is shown for each part of the word. Short translation function helps users to save their time while reading and translating texts. This is the result of integration of ABBYY morphology engine for 40 languages with dictionary writing system and electronic dictionary software.

4. Future Prospects

We see the future of dictionary writing and dictionary using in sophisticated mark up of dictionary entry and processing of this information in electronic dictionary software. One of our goals is to develop in ABBYY Lingvo Content dictionary writing system tools for syntactic and semantic mark up of a dictionary entry. For example, lexicographer will refer the headwords and definitions to definite semantic fields and describe their basic syntactic patterns and contexts. That will help the user of ABBYY Lingvo electronic dictionary get an appropriate translation of a word in a text, because the electronic dictionary will analyze the context, basic semantics and grammar patterns and give the user only one definition from a big dictionary entry, and this will be the exact definition the user is looking for. That is one of the prospects of using electronic dictionaries that is closely allied with dictionary writing systems development.

¹ Here we discuss the case of usage electronic dictionary only, not a paper dictionary.

References

- Atkins, B.T.S.; Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Grefenstette, G. (1998). 'The future of linguistics and lexicographers: will there be lexicographers in the Year 3000?' In Fontenelle et al. (eds.). *Proceedings of the Eighth EURALEX Congress*, 25-41. Liege: University of Liege.
- Hockey, S.M. (2000). 'Dictionaries and lexical databases'. In *Electronic Texts in the Humanities: Principles and Practice*, 146-171. Oxford & New York: Oxford University Press.
- Heid, U. (2009) 'Aspects of Lexical Description for Electronic Dictionaries'. In Proceedings of *eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve.
- Rundell, M. (2009). 'The Road to Automated Lexicography: First Banish the Drudgery... the the Drudges?' In Proceedings of *eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve.
- Spohr, D. (2009). 'Towards a Multifunctional Electronic Dictionary Using a Metamodel of User Needs'. In Proceedings of *eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve.
- Tarp, S. (2008). *Lexicography in the Borderland between Knowledge and Non-Knowledge*. *Lexicographica. Series Maior, 134*. Tübingen: Niemeyer.

<http://www.abbyy.com/>

<http://www.lingvo.com/>