SYSTEM FOR CREATION OF A DICTIONARY ICELANDIC-CZECH STUDENTS' DICTIONARY

Aleš Chejn
Department of Information and Communication Technologies, Pedagogical Faculty,
University of Ostrava, Fráni Šrámka 3, Ostrava-Mariánské Hory, Czech Republic
achejn@gmail.com

Abstract

This article concerns the creation of a lexicographic tool for compiling one-way bilingual dictionaries and encyclopaedias; the tool has served to create the Icelandic-Czech Students' Dictionary. The Introduction explains the aims of the project. The article briefly mentions the history of developing the application Dictionary System as well as the creation of the Icelandic-Czech Students' Dictionary. It also describes the programming methods, explains how the application works and presents its functions. Finally, the article introduces the practical use of the application and the Icelandic-Czech Students' Dictionary.

Keywords

Dictionary Writing System (DWS), Icelandic-Czech Students' Dictionary, web application

Introduction

In terms of education, the information technologies provide new possibilities of teaching and learning foreign languages. One of areas in which the information technologies are successfully applied, is lexicography.

The creation of a lexicographic tool for compiling dictionaries is closely associated with the Icelandic-Czech Students' Dictionary. At the time when the project of creating the dictionary was getting its shape (in 2001), there wasn't any suitable software allowing such a creation. At that time, there were several lexicographic tools that were expensive, e.g. TLex[1]. As the project of creating the Icelandic-Czech Students' Dictionary has been non-profit since the very beginning, such a tool couldn't be used. Other tools, e.g. Matapuna[2], were for free, nevertheless, they didn't meet requirements of complexity and data presentation or they didn't provide the possibility of publishing.

The lexicographic tool was required to:

1) fully support the creation of dictionaries with Icelandic as a first language;

2) guarantee a quality work environment for lexicographers;

---

[1] TLex. *TLex Lexicography and Terminology Software.*
[2] Matapuna. *Dictionary Writing System.*

3) provide a simple interface for editing dictionaries in any operating system;

4) have a secured system allowing none but a registered and logged-in user to edit the dictionary;

5) provide a possibility of publishing data either in a printed form, or through an offline dictionary program or online on the Internet;

6) be universal and support the creation of one-way bilingual dictionaries in any language combination;

7) be published under the free license with the source code.

The creation of the application Dictionary System (DS) and the creation of the Icelandic-Czech Students' Dictionary (ICSD) are interconnected. Therefore, the following part describes the history of creating both the application DS and the ICSD.

History of creating the application Dictionary System and the Icelandic-Czech Students' Dictionary

The project for the Icelandic-Czech Students' Dictionary began in 2001 upon an indirect suggestion of Doc. PhDr. *Helena Kadečková*, CSc. from the Charles University (CU). Having passed the optional subject „Icelandic language" at the CU, we created the first small dictionary, or more precisely, a list of words. That list contained Icelandic words and their Czech translations, information on morphology was added later as well as declension and conjugation endings, following the *Concise Icelandic-English Dictionary*[3]. At that time, the list contained about 1 200 words. It was published on static web pages and later, it was placed on the website of the CU. In 2006, the dictionary was extended by 1 500 new headwords and existing headwords were given another meanings, especially those of phrasal verbs. After passing a summer course of Icelandic language in Iceland in 2006, we converted the dictionary from the Excel program to the MySQL database and we started to create the application DS for supporting the creation of the ICSD. The content of the dictionary has been published under the free license on the website www.hvalur.org.

In 2007, during a scholarship at the University of Copenhagen, the first version of the application DS was created. The following year, a scholarship to study „Icelandic as a second language" at the University of Iceland was obtained. At that time, the project was joined by M.A. Renata Pešková Emilsson. She improved the dictionary by adding an extensive list of words related to everyday life in Iceland. Moreover, the dictionary was extended by new headwords following the *Icelandic-English Dictionary*[4]. Over the next three years, about 15 000 words were added to the dictionary. At the time, we set up new dictionary items: synonyms, examples, frequency.

In summer 2009, we created a script for generating headword declensions and conjugations. That function turned out to be very important as Icelandic language has retained a complex inflectional system with a huge amount of exceptions.

In 2009, we started to cooperate with *Biolib.cz*[5]. We were given a database containing pictures of zoological and botanical species. In the winter term, together with Amir Mulahumic and

---

[3] HÓLMARSSON, Sverrir; SANDERS, Christopher and TUCKER, John. *Íslensk-ensk orðabók / Concise Icelandic-English Dictionary*.

[4] HÓLMARSSON, Sverrir; SANDERS, Christopher and TUCKER, John. *Íslensk-ensk orðabók / Concise Icelandic-English Dictionary*.

[5] BioLib - Taxonomic tree of plants and animals with photos. *Biolib.cz.*

Dorota Nierychlewská-Chejn, we started to create a sound database of Icelandic words. Jón Gíslason recorded 22 000 words on a digital recorder. We created web pages called *Icelandic word sound database6*; those web pages were translated into 6 languages: English (Aleš Chejn), Icelandic (Jón Gíslason), Polish (Dorota Nierychlewská-Chejn), Serbo-Croatian (Amir Mulahumic), Slovak and French (Ján Zaťko). It is possible to search for an Icelandic word and listen to its pronunciation. Nowadays, the web pages provide also the IPA (International Phonetic Alphabet)[7] phonetic transcription of words. A compressed file containing all sound files is freely available to download and it is published under The General Public License GNU v.3.

In 2010, we added to the database the rules of Icelandic pronunciation included in *Handbók um íslenskan framburð[8]*. We created a script for generating a phonetic transcription of any Icelandic word. Together with Jón Gíslason, we keep improving the script, especially for compounds in which the rules of pronunciation are slightly different.

At the end of 2011, we merged the web pages of the dictionary with those of the application, we also unified and improved the design and added a drop-down menu function. The project was joined by Ján Zaťko, a student of Translation and Interpreting at Matej Bel University in Banská Bystrica. He translated the web pages and the Guide to the Dictionary into Slovak, French and English and the application DS into English.

In January 2012, we created web pages providing information on the application DS and we placed the complete application in three versions on the SourceForge.net portal[9]. The application is published under The General Public License GNU v.3.

Programming

The application Dictionary System is a web application. It is written in PHP using CSS and JavaScript. The application uses the relational database MySQL. Information included in database tables can be divided into tables containing system data, dictionary data, lists of abbreviations, information on the dictionary and information on photographs from Biolib.cz. Application uses the charset UTF-8[10], that guarantees correct data visualization and manipulation in all languages. The configuration file *connection.php* contains values for comparing two languages of the dictionary; by modifying these values, it is possible to achieve a correct comparison of any language combination.

Application Dictionary System

The application Dictionary System is a lexicographic tool designed for creating one-way bilingual dictionaries or encyclopaedias. It is so-called DWS (Dictionary Writing System[11]) or DPS (Dictionary Production / Publishing System) application. In the application, it is possible to add, edit and delete headwords and their meanings, to add photographs and sound files, to add, edit and delete declensions and conjugations, to control the process of creating a dictionary and to publish a dictionary in several versions. A new project – dictionary – can be either

---

[6] http://hljod.hvalur.org/index.php
[7] IPA. *Wikipedia: the free encyclopedia.*
[8] GÍSLASON, Indriði and ÞRÁINSSON, Höskuldur. *Handbók um íslenskan framburð.*
[9] The Dictionary System. *SourceForge - Download, Develop and Publish Free Open Source Software*
[10] UTF-8. *Wikipedia: the free encyclopedia.*
[11] http://en.wikipedia.org/wiki/Dictionary_writing_system

imported from a CSV file[12] or added word-by-word. The final result of the application is a multimedia online version of a dictionary, a printed version of a dictionary typeset in a quality form for a book format in LaTex (and then exported to PDF) and an offline version of a dictionary in DSL format (Abbyy Lingvo) for offline programs like GoldenDict. The application supports team work of lexicographers – it monitors activities, helps to coordinate objectives, records headword editing, etc. It provides an easy-to-use user environment for lexicographers, WYSIWYG[13] headword previewing and the access to the application through any browser.

Publishing the dictionary

The dictionary has three final versions. The first one is a printed version, the second one is an online version and the third one is an offline version designed for offline browsing in dictionary programs like GoldenDict. The printed version is designed either for book or normal printing. This version is limited by place (it is not possible to include complete declension and conjugation tables) and by medium (it is not possible to use sound files). Unlike the printed version, the online one allows to get complex information as it is not limited by place. Therefore, the online version contains more information (e.g. sound files with headword pronunciation, complete declension and conjugation tables, etc.) and it is possible to search for headwords by entering any word form. The offline version allows to browse the dictionary without Internet connection. Its advantage is high-speed headword searching (depending on the computer's speed), but it has some limitations in comparison to the online version. For example, it doesn't contain declension and conjugation tables and it is not possible to search for headwords by entering various word forms. However, unlike the printed version, it contains sound files and, through hyperlinks, it allows to move quickly among headwords.

Printed version

Fig. 1: Printed version sample

---

[12] http://en.wikipedia.org/wiki/Comma-separated_values
[13] WYSIWYG. *Wikipedia*: *the free encyclopedia.*

Online version

The headword part is arranged in Icelandic alphabetical order and headwords can be found through the search field. The screen consists of three columns: left, central and right. The following picture shows the arrangement of information in the online version:
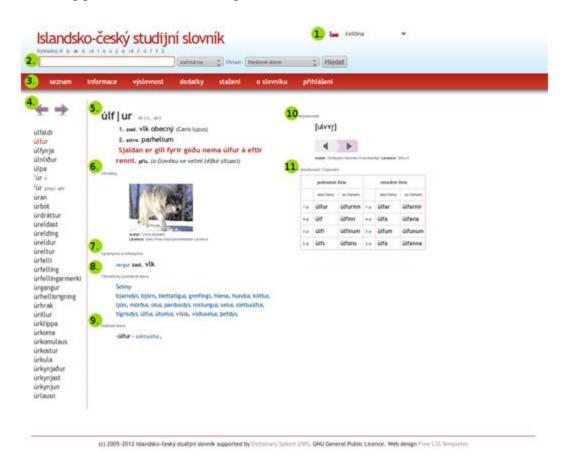

Fig. 2: Online version sample

Legend to the picture above:

1)	Application language selection
2)	Search field with options
3)	Application menu
4)	Headword list
5)	Headword
6)	Photograph
7)	Synonyms and antonyms
8)	Thematically related words
9)	Compounds
10)	Pronunciation – IPA phonetic transcription and sound record
11)	Declension and conjugation tables

Fig. 3: Entry arrangement in the online version

Offline version

The offline version has a structure similar to that of the printed one; in addition, it contains sound files. The structure of the dictionary is simple and contains only information on the dictionary, the authors and the license as well as the headword part arranged in Icelandic alphabetical order.

Monitoring

Nowadays, the application DS is used for creating the ICSD. It registers when and from which IP address a user visited the web page, for which headword and in which dictionary field he searched. The table of geographical location of the IP addresses allows to get automatically information from which country a user visited the dictionary. Recording all these data has two purposes: statistics of words and monitoring the dictionary visit rate. Statistical data like word frequency help to find out for which words users search most frequently; if such a word isn't in the dictionary, it can be added. Monitoring the visit rate also helps to secure the application against attacks of spambots.

Using the application and the dictionary

The ICSD can be used not only for searching for Icelandic words and their Czech equivalents. Users can also search for declensions and conjugations because each headword contains declension or conjugation tables. In addition, the dictionary allows listening to Icelandic headword pronunciation and it also provides information on how to pronounce a phonetic transcription of a headword. The dictionary contains simplified rules of Icelandic phonetics and it is possible to generate a phonetic transcription of any Icelandic word. The rules of Icelandic phonetics are illustrated by examples with sound files. All these options are universal and do not relate directly to Czech language; they can be used by people from around the world. This is one of the reasons why the application and the web pages are translated into six languages (Czech, Icelandic, Polish, English, Slovak and French). We also put on the domain www.hvalur.org educational games for practising Icelandic language, e.g. Scrabble with Czech and Icelandic words, Memory game with cards that can be printed after winning the game and other morphological and phonetic games. Users from various countries can use the dictionary as an educational aid for learning Icelandic.

The creation of the ICSD mainly involves Aleš Chejn, Jón Gíslason and Ján Zaťko. At the present time, Jón Gíslason is a lecturer of Icelandic language at the University of Iceland in Reykjavík. Jón recorded all sound files of headwords. In the application, he mostly revises the phonetic transcription of words. He finds dictionary editing intuitive and quick. He appreciates that he can listen to the pronunciation and see the phonetic transcription of words at the same time. The alphabetical list of headwords allows him to browse the dictionary easily. Ján Zaťko translated the web pages of the ICSD into Slovak, French and English. He also translated the Manual for the application DS into English. Ján uses the alphabetical list of headwords on the left side to navigate the dictionary. He greatly appreciates that the dictionary allows him to find a headword even if it is misspelt in the search field. He also uses the option „Headwords in example" to find other headwords. Ján points out that a new user may find searching for Czech words difficult.

Conclusion

Nowadays, the application Dictionary System supports the creation of the Icelandic-Czech Students' Dictionary and, since 2006, it has run on the domain www.hvalur.org. The dictionary contains the introductory, the headword and the final part. The introductory part includes information on the dictionary and information on the phonetic transcription of headwords, tables of Icelandic pronunciation and a list of abbreviations. The headword part contains more than 24 000 headwords, the IPA phonetic transcription and sound files with pronunciation recorded by a native speaker, grammatical information on the parts of speech and the word endings, syntactic information, word phrases, proverbs, examples and example translations, synonyms, antonyms, Latin names, field and language categories. The dictionary uses multimedia options of the application and provides photographs and sound files for headwords. The final part of the dictionary includes the bibliography. The Icelandic-Czech Students' Dictionary and the application Dictionary System are published under The General Public License GNU v.3 and they are freely available to download.

At the present time, together with Jón Gíslason and Ján Zaťko, we revise the dictionary in detail. We intend to check all headwords in the dictionary, to add missing meanings and examples, to choose photographs and to publish the dictionary at our own expense in 2014. We expect that the printed version of the dictionary will have about 600 pages. This version in PDF format will be freely available to download on the web pages www.hvalur.org. The long-term objective is to add new headwords to the dictionary and to create a handbook of Icelandic phonetics and morphology.

References

BioLib: Taxonomic tree of plants and animals with photos [online]. 2012-10-26 [cit. 2012-10-26]. Dostupné z: http://www.biolib.cz/cz/main/

INDRIĐI GÍSLASON, Höskuldur Þráinsson. *Handbók um íslenskan framburð*. 2. útg. Reykjavík: Rannsóknarstofnun, 2000. ISBN 99-798-4736-0.

SVERRIR HÓLMARSSON, Christopher Sanders a Svavar Sigmundsson RÁÐGJÖF. *Íslensk-ensk orðabók*. 1. útgáfa, 4. prentun. Reykjavík: Iðunn, 1989. ISBN 99-791-0049-4.

Matapuna: Dictionary Writing System [online]. 2012-10-26 [cit. 2012-10-26]. Dostupné z: http://www.matapuna.org/

Obecná veřejná licence GNU v.3. [online]. 2012-10-26 [cit. 2012-10-26]. Dostupné z: http://www.gnugpl.cz/v3/

The Dictionary System. *SourceForge: Download, Develop and Publish Free Open Source Software* [online]. 2012-10-26 [cit. 2012-10-26]. Dostupné z: http://sourceforge.net/projects/dict-system/

TLex: TLex Lexicography and Terminology Software [online]. 2012-10-26 [cit. 2012-10-26]. Dostupné z: http://tshwanedje.com/tshwanelex/

Wikipedia: the free encyclopedia [online]. 2012-10-26 [cit. 2012-10-26]. Dostupné z: http://cs.wikipedia.org