

Система управления для двуязычных словарей

Dictionary management system for bilingual dictionaries

Loopmann A. (andres.loopmann@eki.ee), **Langemets M.** (margit.langemets@eki.ee), **Viks Ü.** (ylle.viks@eki.ee)
Eesti Keele Instituut, Tallinn, Eesti

Целью проекта является создание системы для составления и редактирования двуязычных словарей, позволяя пользователем выбирать дизайн (структура и стиль) и элементы исходной части (эстонская). Пользователь должен получить возможность заниматься с содержанием словаря, в то время как общие показатели будут охраняться системой. Это позволяет уменьшить время компиляции словаря, а также управлять групповые проекты. Во вводной части изложена цель проекта и настоящая ситуация, во второй части главные характеристики и пользовательские качества системы “EELex“, в третьей части кратко описана публичная версия “Estonian-X Dictionary“.

1. The aim of the project is to create the Dictionary writing system for compiling and editing Bilingual dictionaries, with the options of user defined design (structure and style) and selection of the size and elements of the source part (Estonian). The user should gain the opportunity to deal with the contents of the dictionary, while the overall performance will be guarded by the System. This allows users to reduce dictionary compilation time and increase the consistency of their products, as well as manage team projects.

2. The System for bilingual dictionaries constitutes one part of the dictionary writing system called EELex, created at the Institute of the Estonian Language (presented at DWS 2006 workshop, Turin, see Langemets et al. 2006). The project "Lexicographer's Workbench" has been supported by national program for Estonian Language Technology (2006–2010). Some of the basic characteristics of the EELex: web-based, full Unicode support, XML databases, XSD schemas, XSL transformations for presenting different views (XML view, Edit view, Layout view), click-to-edit, i. e. clicking on any field within the Layout preview will immediately link to the appropriate area(s) for editing that field, etc. The tools for lexicographers enable to compile, edit and layout dictionaries, both, starting from scratch, or (depending on the contract) obtaining list of headwords (3 different scopes) or the whole dictionary database of the source language (Estonian) to start working with.

3. One of the most essential applications of the System is the Estonian–X Dictionary (henceforth, EXD), which itself is edited and updated using the System, and else, is the core basis for new bilingual dictionaries. EXD database contains the material of different bilingual dictionaries of different sizes, with the Estonian part of the voluminous Estonian–Russian dictionary (80,000 headwords) serving as the central database. The material for the medium size dictionary (40,000 headwords) has been pre-edited following the line of structural units (composed words as headwords or examples, the system of the cross-references has been updated, as well as the presenting of homonyms, unified lists of usage and domain labels, etc.). The function of an entry is to supply enough information to allow the user and the system to identify a distinct sense of a word or phrase in the source language. Example (1) presents the noun article (*kuld* 'gold'), example (2) presents the lists of compound words (*Ls:*), classified by form and sense, belonging to the noun. (The abbreviation TE stands for the translation equivalent.) The <morphological description standard> belongs to the type of "inflected form based entries", presenting all basic forms, part of speech, and the inflectional type number (there are 3 types available in the System).

- (1) **kuld** <k'uld kulla k'ulda k'ulda, k'ulda[de k'ulda[sid_ &_ k'uld/i S 22]>
1 (*teatud väärismetall 'a yellow precious metal'*) |TE| ♦ **puhas kuld** |TE|; **hambakuld** |TE|; **kullast ehted** |TE|; **kulda pesema** |TE|; **lõpetasin keskkooli kullaga kõnek** |TE|
2 (*värvuselt ja läikelt kulla sarnane 'colour resembling gold'*) |TE| ♦ **päikesekuld** |TE|; **sügiskuld** |TE|; **kased puistavad juba kulda** |TE|
3 (*midagi väärtuslikku ja head 'sth highly respected'*) |TE| ♦ **tema nõuanded on kulda väärt** |TE|; **nendel sõnadel on kulla kaal v hind** |TE|
- (2) ■ **Ls: *kuld+*** (kullast, kullatud 'made of gold') ♦ **kuldbrokaat** tekst |TE|; **kuldmedal** |TE|; **kuldmünt** |TE|; **kuldsõrmus** |TE|; **kuldvillak müt** |TE|
■ **Ls: *kuld+*** (kulla värvi 'lustrous yellow') ♦ **kuldblond** |TE|; **kuldjuukseline** |TE|; **kuldkollane** |TE|; **kuldpõrnikas** zool (*Cetonia aurata*) |TE|
■ **Ls: *kulla+*** ♦ **kullaauk** kõnek, piltl |TE|; **kullafond** (1) maj |TE|, (2) piltl |TE|; **kullakang** |TE|; **kullaketraja** |TE|; **kullaliiv** |TE|; **kullaläige** |TE|; **kullamäed** piltl |TE|; **kullaotsija** |TE|; **kullapalavik** piltl |TE|; **kullaproov** |TE|; **kullasoon** |TE|; **kullastandard** maj |TE|; **kullatera** |TE|

The in-house EXD system, although permanently improved and updated, has already been used for compiling of the Estonian–Russian School Dictionary (to appear in 2009), as well as the Estonian–Ukrainian and the Estonian–Udmurt dictionaries.

4. The public version of the EXD system is intended for various users (lexicographers, freelancers, students, learners, etc.) willing to produce its own dictionary. There has been developed a standard schema as a starting point to fulfill the basic needs of the user. The multifunctional user interface enables to select appropriate parameters for an application:
- (a) Selection of the target language (the list is short so far, but will definitely grow in the future), linked with language specific elements (e.g. gender marking); during the editing process the switching between languages functions automatically;
- (b) Designing and altering of the layout view: users may predetermine the style of the element and the markers of the boundaries of the elements (groups), as well as recurrent elements (groups);
- (c) The morphological interface of EXD offers different standards of morphological description see example (1), depending on scope, aim and target group of the user dictionary. Estonian inflectional morphology is problematic for non-Estonians, as it is characterized by the great number of inflected forms and the extensive variation of its morphological units. With help of the morphological interface the inflectional information of the selected type will be added automatically. The interface applies the rule-based morphological system (Viks 2000), so it will be possible to generate necessary forms for words not included in EXD database, as well.
5. The first public version (1.0) of the Dictionary management system for creating of bilingual dictionaries will be released in the beginning of 2009 (<http://exsa.eki.ee>).

Introduction to EELex

EELex is a fully web-based dictionary management system. After entering EELex and selecting one of the entries (“füüsika” – “physics” in this case), we will have the following picture:

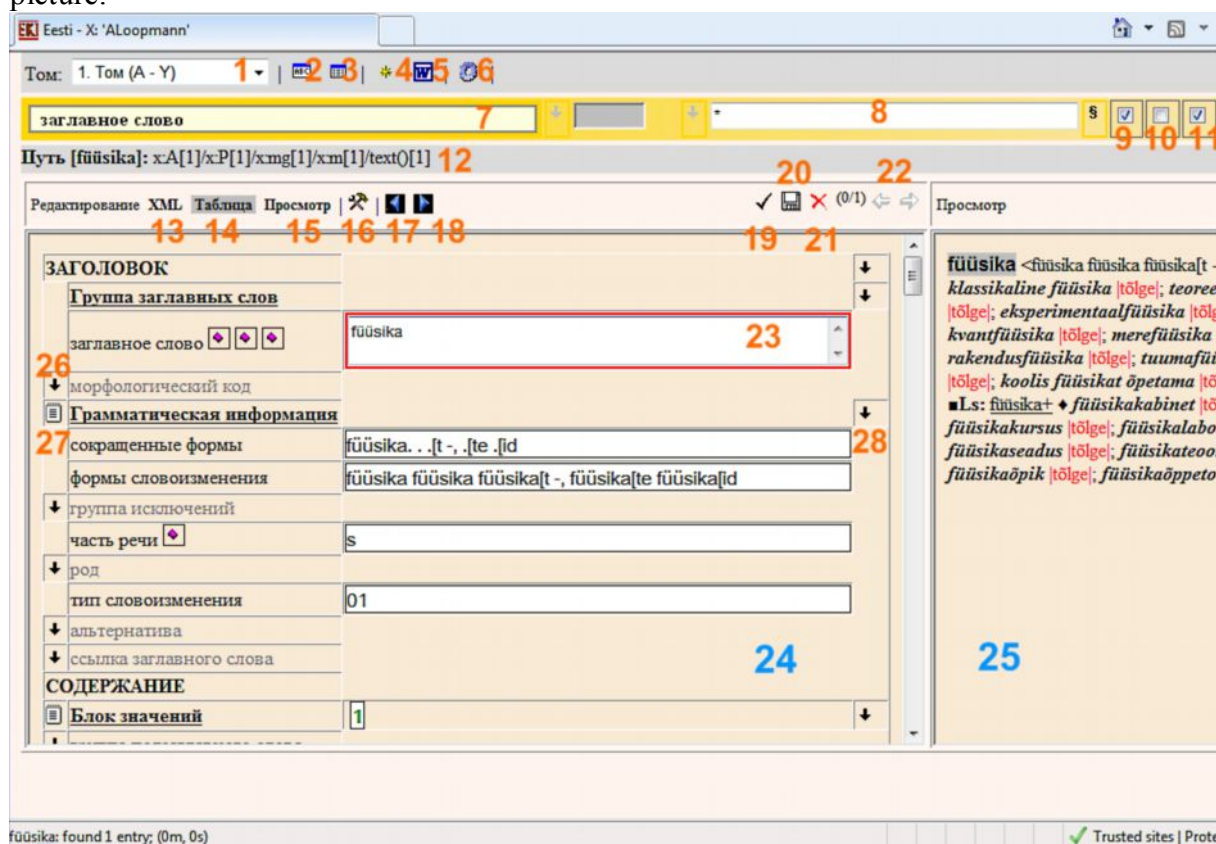


Fig 1. Entry screen

On editing area (24 on Fig 1) headword text “füüsika” is selected for editing (23). One can edit text in table mode (clicking 14), XML mode (clicking 13) and in layout mode (clicking 15). In our example:

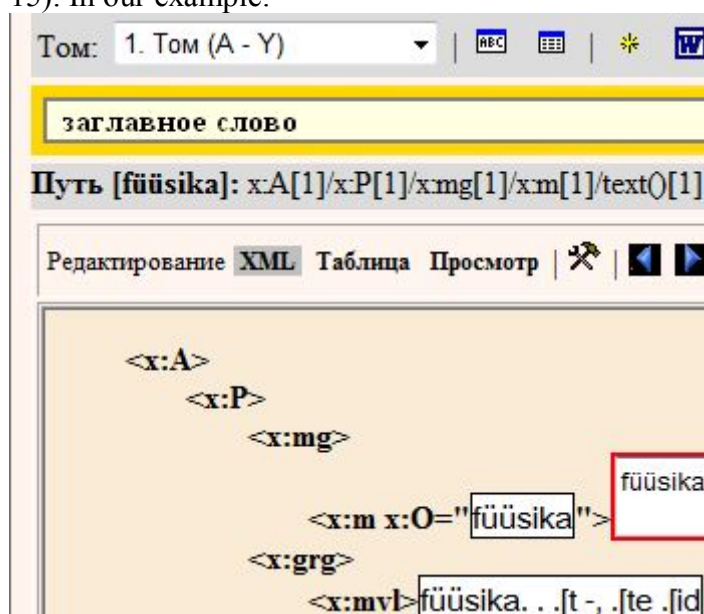


Fig 2. Editing in XML mode

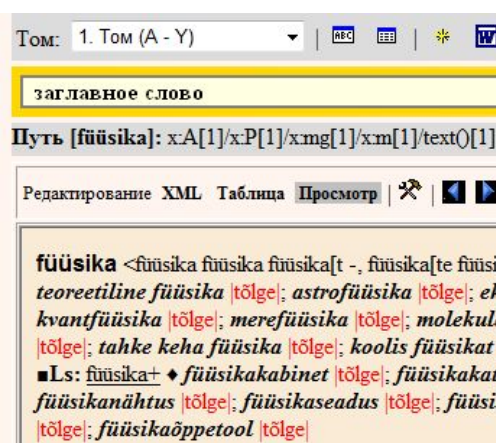


Fig 3. Editing in layout mode

Purpose of all buttons, checkboxes and areas shown in Fig 1 are described in the following table:

1	“Select volume”
2	“Select dictionary entry” (default)
3	“Select query results”
4	“Add new dictionary entry”
5	“Layout dictionary in MS Word”
6	“Dictionary tools” (activity log, import, export, EELex configuration, usage and domain labels management etc)
7	Query element (and attribute) selection
8	Query text
9	Search case insensitive/case sensitive
10	Search exactly/non-letters excluded
11	Search globally/locally
12	Selected element’s entry path
13	“Edit in XML mode”
14	“Edit in table mode”
15	“Edit in layout mode”
16	“Entry tools” (mark entry as complete, finalize entry etc)
17	“Previous entry”
18	“Next entry”
19	“Validate entry”
20	“Save entry”
21	“Delete entry”
22	“Undo/Redo”
23	Selected text for editing
24	Edit area
25	View/Layout area
26	“Create group”
27	“Add group”
28	“Add elements for this group”

Table 1. Legend

In every dictionary, structure of the entry is set in dictionary’s XSD schema. Structure of the entry we can see when we are selecting element, for which we want to query dictionary against (button 7 on Fig 1):

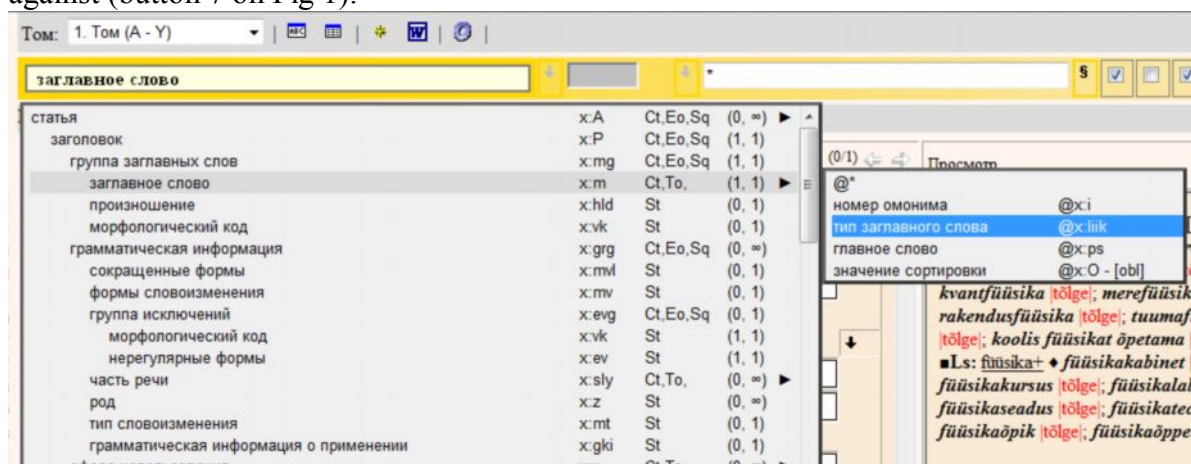


Fig 4. Query element selection

The content of the context menus (right clicking an element name) is also context sensitive: only operations, allowed by XSD schema, are displayed:

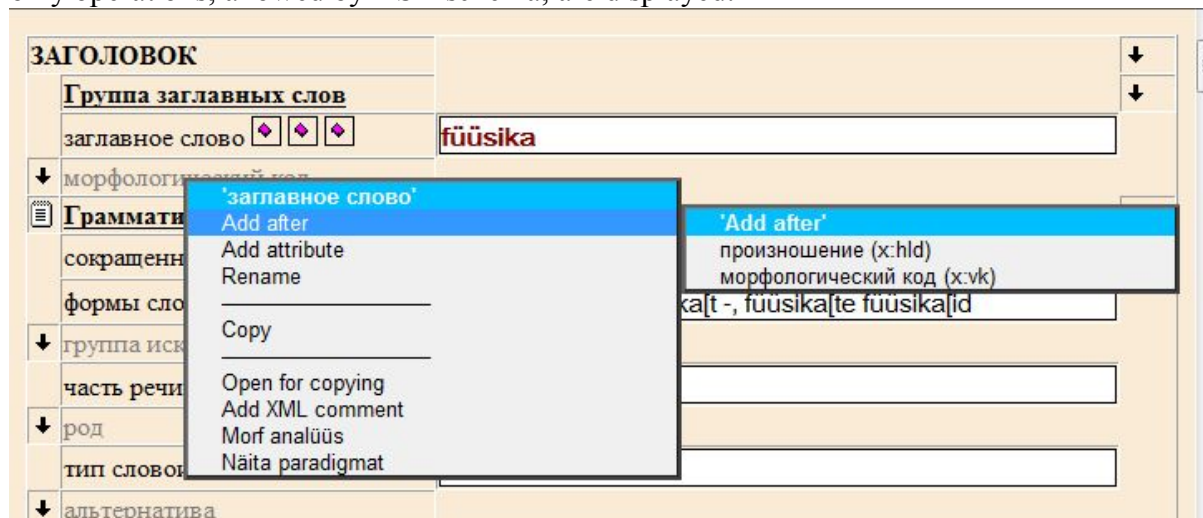


Fig 5. Context menu

Finally: entry is saved back to the dictionary, when it is successfully validated against the schema.

In view area (25 on Fig 1) the entry is shown in layout view. Same layout is used, when we are printing contents of the dictionary to the MS Word (button 5 on Fig 1).

There is a multilevel rights system in EELex: editors, editor-in-chiefs, exporters, administrators etc. Import supports (suitably formatted) RTF documents and one can export dictionaries in native XML format.

Besides to dictionary source and target languages, one can choose between three working languages in EELex: Estonian, English and Russian.

For inflectional forms one can use the rule-based morphological system (Viks 2000): system generates (synthesizes) all the needed inflectional forms for the current headword.

One can check the spelling of the edited text (23 on Fig 1, for example), if the corresponding target MS Word spelling dictionary is installed and for multilingual dictionary, there is possible to automatically switch between input languages (the system switches the input language when the language code of the current textbox changes).

Introduction to Estonian-X Dictionary

Estonian-X Dictionary (EXD) is a specific EELex application, designed for compiling and editing bilingual dictionaries. Estonian part of the dictionary was collected from several bilingual and Estonian monolingual dictionaries. Public version of the EXD is intended for various users and groups, who want to produce their own dictionary. Public version contains also ca 70 demo entries.

References

1. Langemets, Margit; Loopmann, Andres; Viks, Ülle (2006). The IEL Dictionary Management System of Estonian. // in de Schryver, Gilles-Maurice (ed.) 2006. *DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writing Systems*. Turin: Turin University, 11–16.

2. Viks, Ülle (2000). Tools for the Generation of Morphological Entries in Dictionaries. // in *Second International Conference on Language Resources and Evaluation. Proceedings Volume I*. Athens, Greece, 31 May – 2 June 2000, 383–388.