COGNITIVE STUDIES | ÉTUDES COGNITIVES, 16: 90–99
Warsaw 2016

DOI: 10.11649/cs.2016.008

VOLODYMYR SHYROKOV[A], **IRYNA OSTAPOVA**[B],
MAKSYM NADUTENKO[C], & **YULIA VERBYNENKO**[D]

Ukrainian Lingua-Information Fund of NAS of Ukraine, Ukraine
[A]vshirokov48@mail.ru ; [B]irinaostapova@gmail.com ;
[C]hostmaster@ulif.org.ua; [D]yulia_verbinenko@yahoo.co.uk

ONTOLOGIZED LEXICOGRAPHICAL SYSTEMS IN MODERN TERMINOGRAPHY

Abstract

The principles of the construction and use of the Virtual Terminographical Laboratory “Welding” are described. VTL “Welding” provides support for the compiling of electronic terminological multimedia dictionaries in the field of welding and allied processes.
Keywords: virtual lexicographical laboratory; welding terminology; glossary of terms; lexicographical system

1 Introduction

Nowadays, terminological support is a necessary and important direction in the development of any subject area. It has a special importance in the context of the global communication of professionals and product end-users. The necessity to harmonize terms and their meanings in different languages, to create tools of interlingual adaptation, and to develop tools for the handling of so-called controlled languages creates new, challenging requirements for the terminographic field.

Due to this, there is a growing need to develop models which are able to provide some unified means of compilation, modification and versatile usage by multilingual dictionary systems in different subject areas. These models also need to enable users to combine linguistic and encyclopedic descriptions of the terms of different subject areas and to engage with multimedia information. They should also provide online interaction between geographically widespread specialists, in order to support a continuous terminographic process and the rapid tracking of changes occurring in the subject field.

Some of these aspects, concerning the subject area “Welding”, are going to be examined in this article. This subject area was chosen because the authors, in collaboration with the E. O. Paton Electric Welding Institute of the National Academy of Sciences of Ukraine, one of the world’s leading centers of welding science and technology, have created several electronic dictionaries on this subject over several years (Широков et al., 2008, 2010, 2013).

2 The Conceptual Modelling of the Terminographical System "Welding"

Problem formulation The conceptual ideas that will be described in this article were implemented in the most recent dictionary (Широков et al., 2013). This publication saw the implementation of new ideas in subject area linguistic ontology, and its implementation in the form of a virtual terminographical laboratory, which will be described below.

As was mentioned earlier, the E. O. Paton Electric Welding Institute of the NAS of Ukraine was the end-user, and it stated the following requirements:

1. The final product had to be made in two forms: an electronic dictionary on CD-ROM and an online system on the Ukrainian Linguistic Portal (`http://lcorp.ulif.org.ua/`). The functionality of both forms had to be identical.
2. The dictionary had to be compiled in three languages: Ukrainian, English and Russian. Also, the order of languages in the dictionary and the language of interface were not to be fixed and set by the user.
3. The interpretative zone in the dictionary was to be divided into two components: a) terminological definitions of register units; b) multimedia illustrations (drawings, photographs, and video information).
4. Taxonomic classification (so-called linear ontology) was to be implemented in the dictionary, and it had to cover the basic subareas of welding science and technology.
5. The dictionary was to contain a basic core of terminology of about 1,200 register units, with full terminographic development, and over 10,000 units of peripheral terms, which only have equivalents in three languages.
6. The dictionary had to be an open system. There had to be a possibility to continuously modify and update it.

3 The Structure of the Ontologized Lexicographical System "Welding"

The set of requirements described above determined a conceptual model of the dictionary as an ontologized lexicographical system, and its implementation as a virtual terminographical laboratory. In fact, the so-called model of linguistic ontology, according to the (Широков & Потапова, 2012), was applied.

The linguistic ontology of a subject area differs from other linguistic information systems in that it has a lingual level as well as an extra-linguistic level — a level which consists of the meta-descriptions of concepts and their relations. The possibility of outlining extra linguistic data in the system makes it possible to display it in several language systems simultaneously. Dividing a linguistic ontology system into two levels makes it possible to create a model for two or more languages which share a common core of concepts and relations between them.

Let us consider the conceptual description of the subject area as a lexicographical system. The concepts and denotations of lexicographical theory are discussed in: (Широков, 2011, 2009a, 2009b, 2009c, 1998; Широков, В. А., Сидорчук, Н. М., & Остапова, И. В., 2012).

Let us denote the domain, "welding", as system D, and $I^Q(D)$ as a class of elementary information units in this system.

$V(I^Q(D))$ is a set of descriptions of units that belong to the class $I^Q(D)$. As a result of subject S, perception of the elementary information units, we obtain the following:

$$S : I^Q(D) \to V(I^Q(D)) \tag{1}$$

The concepts and terms of the subject area "welding" are the basic information units.

Let us consider the representation of "form – meaning", where complex $I^Q(D)$ is its medium, and let us metalize it with a process of recursive reduction $RR \downarrow [V(I^Q(D))]$. Taking into account the above, the structure of description $V(I^Q(D))$ will be:

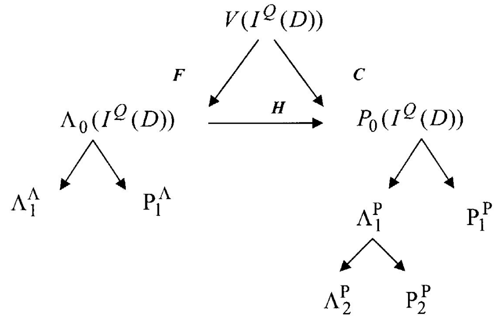


Figure 1: Diagram of the description structure of the subject area "welding" lexicographic system

According to the theory of L-systems, one can distinguish from the structure of the initial basic L-system a set of informational and linguistic substructures, which are regarded as separate L-systems. The basis of these systems is the lexicographic parameterization of an ontological term (Широков & Потапова, 2012).

$\Lambda_0(I^Q(D))$ corresponds to the part of descriptions $V(I^Q(D))$ which represents form $I^Q(D)$.

$P_0(I^Q(D))$ corresponds to the part of descriptions $V(I^Q(D))$ which represents content $I^Q(D)$.

For the ontology of the application domain, $\Lambda_0(I^Q(D))$ contains certain concept designations and its meta-description.

$P_0(I^Q(D))$ — is a set of connections with other concepts ($\langle t^R \rangle$ in an ontological model), through which its meaning is revealed. In addition, the multiplicity of connections between concepts and additional information represented by set $\langle t^I \rangle$ could be included here.

In turn, in the structure of L-systems at zero level $\Lambda_0(I^Q(D))$ and $P_0(I^Q(D))$ could be distinguished as subsystems of the next level. L-system $\Lambda_0(I^Q(D))$ of concepts' meta descriptions may be represented as a combination of two subsystems of the next level and another lexicographical effect.

Λ_1^Λ — a set of concepts' indexes (in fact, it is an index of the main register in the database) that represents a formal part.

P_1^Λ — a set of pointers to the ontological class of a concept which characterizes its content in a certain way. This subsystem can be disclosed further: an index of an ontological class and its definition. A set $\langle OntC \rangle$ in the ontological model contains this information.

In the other "branch" of Diagram 1 the process of reduction is similar.

$P_0(I^Q(D))$ — a complex of concept connections, through which meaning is represented, and can be interpreted as a combination of two subsystems: form Λ_1^P (indexes of connection type) and content P_1^P (a set of parameters that describe connections between terms).

L-system Λ_1^P of the type of connection pointers, in turn, may be disclosed as a combination of the formal part of the connections' descriptions, which consists of indexes Λ_2^P, and a content description through the specification of ontology classes P_2^P. This system, as well as P_1^Λ, may be disclosed through another level by the description of ontological classes (this stage was not shown in Diagram 1 so as not to over-complicate it).

In order to model a multilingual ontology of the application domain, let us look at how the terminological concept $\Lambda_0(I^Q(D))$ and its relations $P_0(I^Q(D))$ are reflected in the languages which are represented in the system.

The upper part of Diagram 2 (level 1) is a meta-description of ontology, which was disclosed in detail in the previous diagram. Level 2 is the language level. The lowest row contains elements of

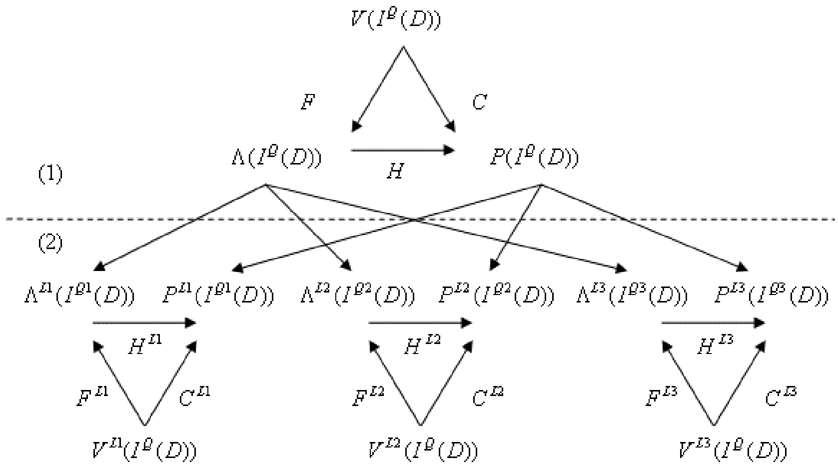


Figure 2: Diagram showing the representation of "welding" ontology on its lexicographical system

type $V^{L1}(I^Q(D))$ — they are sets of descriptions of elementary information units in the language $L1(L2, L3)$ where $L1, L2, L3$ — languages indexes. The middle row consists of elements of type $\Lambda^{L1}(I^{Q1}(D))$ and $P^{L1}(I^{Q1}(D))$; on the one hand they set a reflection of the ontological extralinguistic complexes $\Lambda(I^Q(D))$ and $P(I^Q(D))$ in language $L1$, and on the other hand, $\Lambda^{L1}(I^{Q1}(D))$ — is a set of application domain concepts of language $L1$, and a set of its descriptions in grammatical categories of language $L1$. $P^{L1}(I^{Q1}(D))$ — is a set of text definitions of application domain concepts in the language $L1$, where connections between terms are reflected.

Let us bring together the lexicographic model of the system, represented in Diagrams 1 and 2, and a conceptual model of the ontology. It has two levels:

Level 1 corresponds to the ontology of application domain concepts, represented as an L-system:

$Ob(LS) = \{T, OntC, Pers, Br, R\}$, where T — a set of terminological concepts, R — a set of types of connections between concepts, $OntC$ — a set of lexical and ontological classes. $Pers$ — a set of bibliographical parameters and Br — a set of sections of application domain;

$RelOb(LS) = \{TT, TPers, TBr\}$, where TT — a set of connections between concepts (in other words, between elements of T-set), $P(I^Q(D))$ corresponds to it in Diagram 1. $TPers$ — a set of connections between a set of concepts and a set of bibliographical data, TBr — a set of connections between a set of concepts and a set of application domain sections.

The second level is a reflection of the formal ontology on the system's languages:

$Ob(LS^{Li}) = \{T^{Li}, TCont^{Li}, TDef^{Li}, DefSource^{Li}\}$ where T^{Li} is a set of Li language's terms, for $\forall t \in \langle T^{Li}\rangle t(t^{Li}, \langle T^{Li}, \langle T^{GLi}\rangle, tDef^{Li})$, where t^{Li} — is an orthographic standard for a term in language Li, T^{GLi} — a set of grammatical characteristics of term t in the language Li. This corresponds to a member $\Lambda^{Li}(I^{Qi}(D))$ from Diagram 2. Text definition of the term — $tDef^{Li}$ and $\langle TCont^{Li}\rangle$ — a set of term t^{Li} contexts in the language Li corresponds to members of $P^{Li}(I^{Qi}(D))$ type from Diagram 2. $DefSource^{Li}$ — a set of bibliographic descriptions of text definition sources.

$RelOb^{(LS^{Li})} = \{TTCont^{Li}\}$ — this is a set of relations between a set of terms and a set of contextual examples (from the corpus of subject area texts); it corresponds to the complex $P^{Li}(I^{Qi}(D))$ in diagram 2.

Integrity limitations:

1. No element of set T^{Li} at level 2 can exist without connection to the corresponding element of set T at the level 1;

2. Several elements of set T^{Li} at level 2 in each language could correspond to one element of set T at level 1;

3. Connections between objects (terms) from different languages (translation) are made only through an element of set T of level 1.

Thus, level 1 of the conceptual model contains formal data of ontology, while level 2 contains linguistic information about the subject area. This distinction enables the compilation of a multi-language system, which would be difficult to do when modelling a subject area through ontology. Interaction between the linguistic data complexes of the different languages occurs only through the formal ontology of level 1.

The process of abstracting a dictionary (lexicographic) structure is a kind of reconstruction of the lexicographical effect which led to the creation of the structure. Due to the fact that in linguistic ontology we can observe two levels of lexicographical effects, the structure of a dictionary (ontological) entry for a concept T, which is expressed by the term T^{L1} in language $L1$ with grammatical parameters G^{L1}, is as follows:

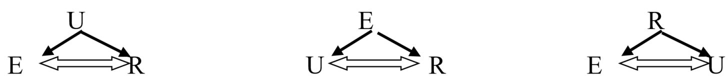
$$T(T^{L1}, G^{L1}) = C^{O} + C^{L1} \ ; \tag{2}$$
$$C^{O} = T^{OntC} + TT + TPers + TBr \ ;$$
$$C^{L1} = S^{L1} + T^{L2} + T^{L3} + TDef^{L1} + TCont^{L1}, \text{where } T^{L2} \supseteq S^{L2} \text{ and } T^{L3} \supseteq S^{L3} \ .$$

C^{O} — a set of ontological parameters, C^{L1} — a set of linguistic parameters in the language $L1$. T^{OntC} — an ontological class of concepts. S^{L1} — synonyms to T^{L1}. The sections of translated equivalents T^{L2} and T^{L3} in languages $L2$ and $L3$ contain whole synonymous sets S^{L2} and S^{L3}, respectively.

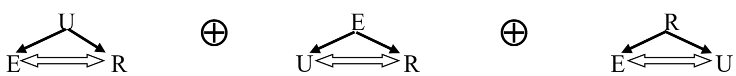
Dictionary Database and User Interface A lexicographical database is defined as the implementation of a specific L-system by means of a computer-based relational data model.

During the compilation of the lexicographical database "Welding" we developed a general scheme for compiling and representing the internal structure of terminological dictionary entries. It was designed to meet the following conditions: the possibility of random order of translation in a dictionary entry; the independence of the entry structure from the language of a source word; the possibility to increase the number of translation languages.

We built the structure of the lexicographical system on the basis of the conditions listed above. This required building an integrated lexicographic system, rather than an integrated one. This was done as follows. First, we defined three basic lexicographical systems, whose input languages are Ukrainian (U), Russian (R) and English (E). They are denoted as follows:



The next step is to integrate these three basic L-systems, i.e. to compile an integrated L-system:



In the last diagram symbol ⊕ denotes an operation of lexicographical systems integration. As we see, an L-system that is compiled in this way is symmetrical, i.e. it is invariant to any rearrangement of the languages (U; E; R). From this it follows that in the compiled L-system, any language can be the input language.

Terminology units were chosen as the main structural element of the dictionary entry. They are the structural part of a dictionary entry that contains the term or terminological phrase, its grammatical parameters, possible synonyms, and phonetic or morphological variants that correspond to a certain terminological concept in one of the languages. They are denoted as follows: $\mathbf{UT_i}$ —

terminology unit of i^{th} concept in Ukrainian; $\mathbf{RT_i}$ — terminology unit of i^{th} concept in Russian; $\mathbf{ET_i}$ — in English; $\mathbf{C_i}$ — definition of i^{th} terminological concept; $\mathbf{Il_i}$ — a set of media illustrations of i^{th} concept (audio, video, picture in raster or vector format); $\mathbf{S}_i$ — a set of structural elements that correspond to the i^{th} terminological concept.

Some examples of entries from the digital dictionary "Welding" can be seen below. The user could select the language of the register and the category which he/she wants to work with.

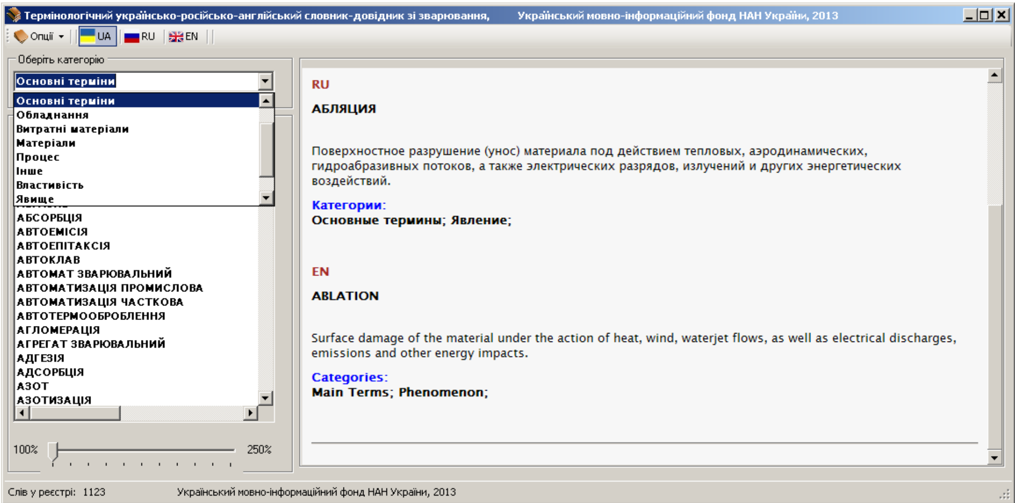


Figure 3: An example of a dictionary entry. Access via Ukrainian terms register

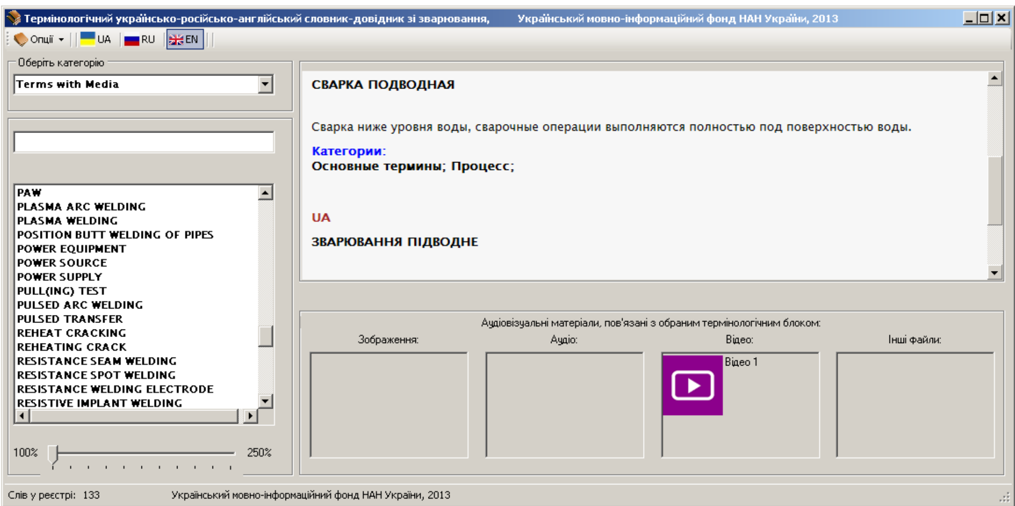


Figure 4: An example of a dictionary entry with media illustration (one video file)

The Virtual Terminographical Laboratory The proposed conceptual model was the basis of the software which was developed for the experimental zone of the Virtual Terminographical Laboratory "Welding", access to which is provided via the Ukrainian Linguistic Portal (`http://lcorp.ulif.org.ua`). According to the general scheme of the VTL, the "Welding" software operates on service-oriented technology and consists of the following parts: 1) data repository —

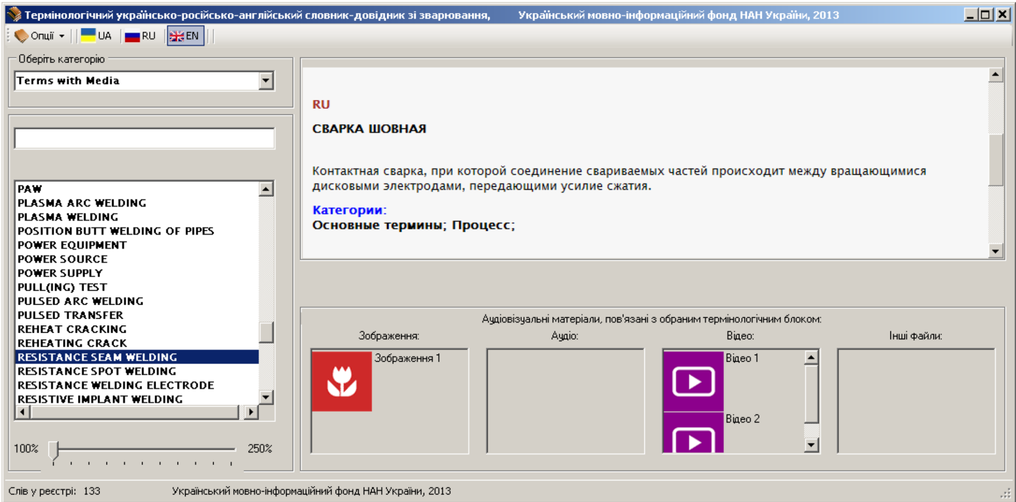


Figure 5: An example of a dictionary entry with media illustrations (one image file and two video files)

a database in SQL Server 2008 R2; 2) web services with APIs to access and manipulate data; 3) client-side applications that provide functionality and graphical interface circuits of automatized working places.

For interaction between different levels of the Terminographical Laboratory "Welding" we use Windows Communication Foundation — a service-oriented system of messaging and data transferring, which provides interaction between software components via a simplified model of a unified cross-platform interaction. For the effective functioning of the system, we use powerful security and data integrity tools, as the lab is focused on the collaboration of a large number of users.

In accordance with the structure of dictionary entries and the requirements of multilingual dictionaries, we developed an internal form of information representation in the digital environment. Based on this, we compiled a lexicographical database (LDB) "Welding", which consists of tables, connected by a network of logical links.

Such an internal structure of a multilingual dictionary enables the build-up of external interfaces, in accordance with the requirements listed above.

The external interfaces of the VTL "Welding" (see Fig. 3, 4, 5) feature, on the left-hand side, three registers of terms and terminological expressions listed in alphabetical order in Russian, Ukrainian and English respectively. The register is divided into 30 units per page. Navigation through the pages is conducted with the help of a tool placed directly under the register. A search box that is located above the register is used to search for terms. On the right-hand side there is a dictionary entry that has been dynamically compiled from the elements of the lexicographical database. The user can print the dictionary entry he/she needs after reviewing it.

The digital dictionary is equipped with editing tools; a terminological unit is an editing unit. Users have the right to introduce new terminological concepts to the digital dictionary. In this version of the dictionary, an order of terminological units is strictly specified. In future versions there will be the possibility to enter the editing subsystem via any language. Editing of the entries is performed at the level of the terminological concepts, and for this purpose some standard forms were developed.

For example, during the edition of nouns in Ukrainian or Russian, the editor can choose the gender, number, declension, phonetic and morphological variants of a term. Fields of types and aspects are available for verbs. When editing a term in English, the editor may indicate its American variant and mark plurality. During the process of editing the elements of terminological concept, users also have the possibility to view a box where the term is represented completely. If there are more translation languages, the editor has to develop parameter forms for all the terms in all the languages to be added. There is also the possibility to add comments to the blocks of terminological concepts, as well as to the separate terms.

With such a system it is necessary to create different levels of access rights, from users whose rights are restricted to viewing the dictionary entries and comment units of terminological concepts from a particular part of the register, to users who have editorial and administrative access.

Experts from the E. O. Paton Electric Welding Institute have access to the Virtual Terminographical Laboratory "Welding" (Ukrainian Linguistic Portal `http://lcorp.ulif.org.ua/WeldingOntology/`) that, via the Internet, provides constant interaction between colleagues from Ukraine and other countries.

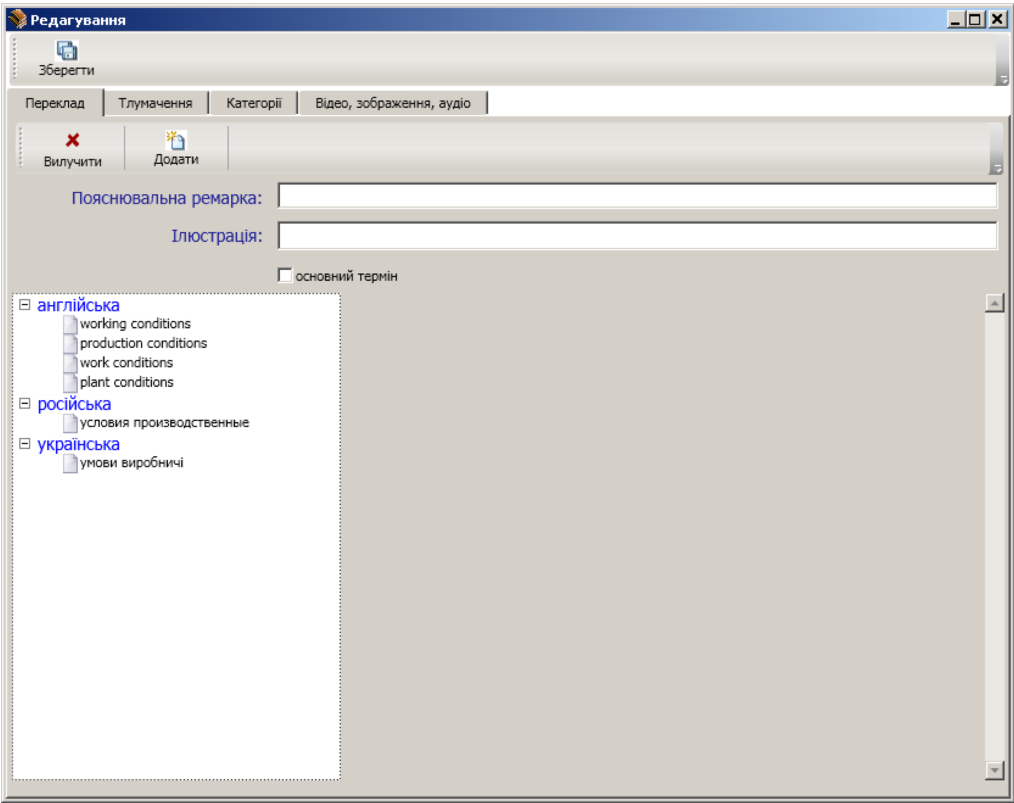


Figure 6: Window for dictionary entry editing

4 Conclusion

The results of the compilation of the ontologized terminological L-system "Welding" are the following:

— A general scheme of the dictionary entries' internal structure representation in the lexicographical terminological database was developed. Algorithms for generating the dictionary entries' elementary units, depending on the end-user's needs and material availability, were developed. The internal structure of terms representation was designed to meet the following conditions: the possibility of a random order of translation in the dictionary entry; the independence of the entry structure from the language of a word's source; the possibility to increase the number of translation languages;

— The lexicographical database "Ukrainian-Russian-English Dictionary on Welding", which is automated, processed and structured according to a standard model, was created;

— Software, based on service-oriented technology, was developed for the experimental zone of the Virtual Terminographical Laboratory "Welding";

— An experimental simulation of the Virtual Terminographical Laboratory "Welding" was based on the corpus of "Ukrainian-Russian-English Dictionary of Welding", whose register consists of about 12,000 terms;

— Experts from the E. O. Paton Electric Welding Institute of the National Academy of Sciences of Ukraine have access to the Virtual Terminographical Laboratory "Welding" (Ukrainian Linguistic Portal `http://lcorp.ulif.org.ua/WeldingOntology/`) that, via the Internet, provides constant interaction between colleagues from Ukraine and other countries.

References

Shirokov, V. (2009a). Theory of lexicographic systems (Pt. 1). In R. Garabík (Ed.), *«Metalanguage and encoding scheme design for digital lexicography»: Proceedings of Mondilex Third Open Workshop* (pp. 151–167). Bratislava: Tribun EU.

Shirokov, V. (2009b). Theory of lexicographic systems (Pt. 2). In V. Koseska-Toszewa, L. Dimitrova, & R. Roszko (Eds.), *«Representing semantics in digital lexicography»: Proceedings of Mondilex Fourth Open Workshop* (pp. 89–105). Warszawa: Institute of Slavic Studies, Polish Academy of Sciences.

Shirokov, V. (2009c). Theory of lexicographic systems (Pt. 3). In T. Erjavec (Ed.), *«Research infrastructure in digital lexicography»: Proceedings of Mondilex Fifth Open Workshop* (pp. 98–119). Ljubljana.

Широков, В. А. et al. (2008). *Українсько-російський, Російсько-український словник зі зварювання.* Київ: Довіра.

Широков, В. А. et al. (2010). *Українсько-російсько-англійський словник зі зварювання.* Київ: Довіра.

Широков, В. А. et al. (2013). *Термінологічний українсько-російсько-англійський словник-довідник зі зварювання.* Київ: НАН України, Ін-т електрозварювання ім. Є. О. Патона, Укр. мов.-інформ. фонд.

Широков, В. А. & Потапова, Е. В. (2012). Онтология предметной области как лексикографическая система особого типа. *Казанская наука*, (12), 209–213.

Широков, В. А. (1998). *Інформаційна теорія лексикографічних систем.* Київ: Довіра.

Широков, В. А. (2011). *Комп'ютерна лексикографія.* Київ: Наукова думка.

Широков, В. А., Сидорчук, Н. М., & Остапова, И. В. (2012). Лексикографические системы для разработки многоязычных компьютерных терминологических словарей. In *Прикладная лингвистика в науке и образовании: Сб. тр. [VI междунар. конф., 5–7 апреля 2012, Санкт-Петербург]* (pp. 303–306). Санкт-Петербург: ООО "Книжный дом".

References (Transliteration)

Shirokov, V. (2009a). Theory of lexicographic systems (Pt. 1). In R. Garabík (Ed.), *«Metalanguage and encoding scheme design for digital lexicography»: Proceedings of Mondilex Third Open Workshop* (pp. 151–167). Bratislava: Tribun EU.

Shirokov, V. (2009b). Theory of lexicographic systems (Pt. 2). In V. Koseska-Toszewa, L. Dimitrova, & R. Roszko (Eds.), *«Representing semantics in digital lexicography»: Proceedings of Mondilex Fourth Open Workshop* (pp. 89–105). Warszawa: Institute of Slavic Studies, Polish Academy of Sciences.

Shirokov, V. (2009c). Theory of lexicographic systems (Pt. 3). In T. Erjavec (Ed.), *«Research infrastructure in digital lexicography»: Proceedings of Mondilex Fifth Open Workshop* (pp. 98–119). Ljubljana.

Shirokov, V. A. & Potapova, E. V. (2012). Ontologiia predmetnoĭ oblasti kak leksikograficheskaia sistema osobogo tipa. *Kazanskaia nauka*, (12), 209–213.

Shirokov, V. A., Sidorchuk, N. M., & Ostapova, I. V. (2012). Leksikograficheskie sistemy dlia razrabotki mnogoiazychnykh komp'iuternykh terminologicheskikh slovareĭ. In *Prikladnaia lingvitika v nauke i obrazovanii: Sb. tr. [VI mezhdunar. konf., 5–7 aprelia 2012, Sankt-Peterburg]* (pp. 303–306). Sankt-Peterburg: OOO "Knizhnyĭ dom".

Shyrokov, V. A. (1998). *Informatsiĭna teoriia leksykohrafichnykh system.* Kyïv: Dovira.

Shyrokov, V. A. (2011). *Komp'iuterna leksykohrafiia.* Kyïv: Naukova dumka.

Shyrokov, V. A. et al. (2008). *Ukraïns'ko-rosiĭs'kyĭ, Rosiĭs'ko-ukraïns'kyĭ slovnyk zi zvariuvannia.* Kyïv: Dovira.

Shyrokov, V. A. et al. (2010). *Ukraïns'ko-rosiĭs'ko-anhliĭs'kyĭ slovnyk zi zvariuvannia.* Kyïv: Dovira.

Shyrokov, V. A. et al. (2013). *Terminolohichnyĭ ukraïns'ko-rosiĭs'ko-anhliĭs'kyĭ slovnyk-dovidnyk zi zvariuvannia.* Kyïv: NAN Ukraïny, In-t elektrozvariuvannia im. IE. O. Patona, Ukr. mov.-inform. fond.

Acknowledgment

This work was supported by a core funding for statutory activities from the Ministry of Education and Science of Ukraine.

The authors declare that they have no competing interests.

The authors' contribution was as follows: concept of the study and the writing: Volodymyr Shyrokov, Iryna Ostapova, Maksym Nadutenko, Yulia Verbynenko.

Publisher: Institute of Slavic Studies, Polish Academy of Sciences, University of Silesia & The Slavic Foundation