

Alona DOROZHYNska¹

Supervisor: Volodymyr SHIROKOV²

ORGANIZACJA DOSTĘPU DO DANYCH W FACHOWYM SŁOWNIKU WIELOJĘZycznym

Streszczenie: W dzisiejszych czasach, sociologizacja jakiegokolwiek produktu zawierającego spisy czegoś w ujęciu leksykograficznym jest zdeterminowana głównie poprzez adaptację tego dokumentu do funkcjonowania/użytkowania go/ w mediach cyfrowych (w cyfrowym środowisku). Badania przeprowadzono dla publikacji "Dictionary of Ukrainian Biological Terminology" [1], czyli w odniesieniu do słownika terminologii biologicznej w języku ukraińskim. Ta publikacja była pierwszym dziełem z leksykograficznym ujęciem tej tematyki na Ukrainie. Publikacja zawiera najczęściej używane terminy z zakresu biologii w językach ukraińskim, rosyjskim oraz angielskim, a także zawiera wyjaśnienia zestawionych terminów.

Słowa kluczowe: komputerowy porządek leksykograficzny, system leksykograficzny, analiza składniowa, język XML, baza danych, przestrzeń cyfrowa, strona w sieci.

ORGANIZATION OF ACCESS TO THE DATA OF THE MULTILINGUAL TERMINOLOGY DICTIONARY

Summary: In modern conditions, the sociologization of any lexicographic product is largely determined by its adaptation to functioning in the digital environment. The research is carried out for the "Dictionary of Ukrainian Biological Terminology", [1] which is the first lexicographic work in Ukrainian studies, covering the most used biological terminology in Ukrainian, Russian and English languages and containing the interpretation of the terms.

Keywords: computer lexicography, lexicographic system, parsing, XML, database, digital space, website.

1. Introduction

In modern conditions, the sociologization of any lexicographic product is largely determined by its adaptation to functioning in the digital environment. This is true not

¹ Ukrainian Lingua-Information Fund of NAS of Ukraine, junior researcher, alonychkatkachyk@gmail.com

² Ukrainian Lingua-Information Fund of NAS of Ukraine, Academician of NAS of Ukraine, vshirokov48@gmail.com

only for the mass consumer, but also for professional communities. The digital environment provides a qualitatively new level of service for research work with linguistic data presented in dictionary form.

But the digital transformation of lexicographic works requires some general theoretical scheme for describing and presenting the widest class of lexicographic objects. Our developments are based on the theory of lexicographic systems [2]. The history of the formulation of the theory of lexicographic systems began with the study of a significant number of structures of actually existing traditional dictionaries, their generalization and the construction of appropriate formal models.

The research is carried out for the "Dictionary of Ukrainian Biological Terminology", [1] which is the first lexicographic work in Ukrainian studies, covering the most used biological terminology in Ukrainian, Russian and English languages and containing the interpretation of terms (both word terms and phrase terms).

Our approach consists in the step-by-step transformation of the dictionary text into a website. Basic technological stages (Fig. 2).

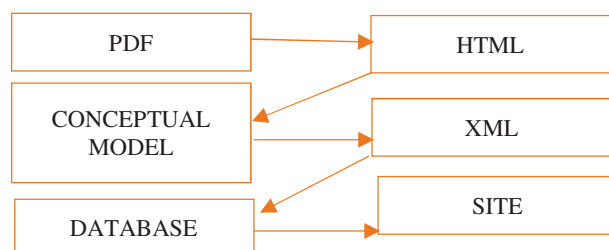


Figure 1. Basic technological stages

This process, in our opinion, contains steps that can be applied to other dictionaries, so we believe that this sequence is an effective and universal way to translate traditional dictionaries into digital format.

2. The steps

Consider the individual steps that make up the transformation process.

2.1. Paper book (PDF)

If from the very beginning the text of the Dictionary is accessible in electronic format, then work with it. In particular, the text of the DUBT [1] we had from the beginning in the form of a PDF file. For convenience, this file has been converted to Word doc in order to do some text conversion. Among these transformations we note the following: the disclosure of abbreviations of a certain type (for example: п. заґруднінна → пульсація заґруднінна and the like.; replacement of stressed letters with a combination of two characters: «літера#». The following replacements were made: á—a#, é—e#, ý—и#, í—i#, ó—o#, ý—y#, í—i#, я—я#, é—e#, ю—ю#, Ъ—ъ# (for Cyrillic); ý—y# (for Latin). All dictionary articles were processed in this way.

2.2. Construction of a conceptual model of the lexicographic system

Following the theory of lexicographic systems [2], the structure of the L-system of SMS is presented in the form:

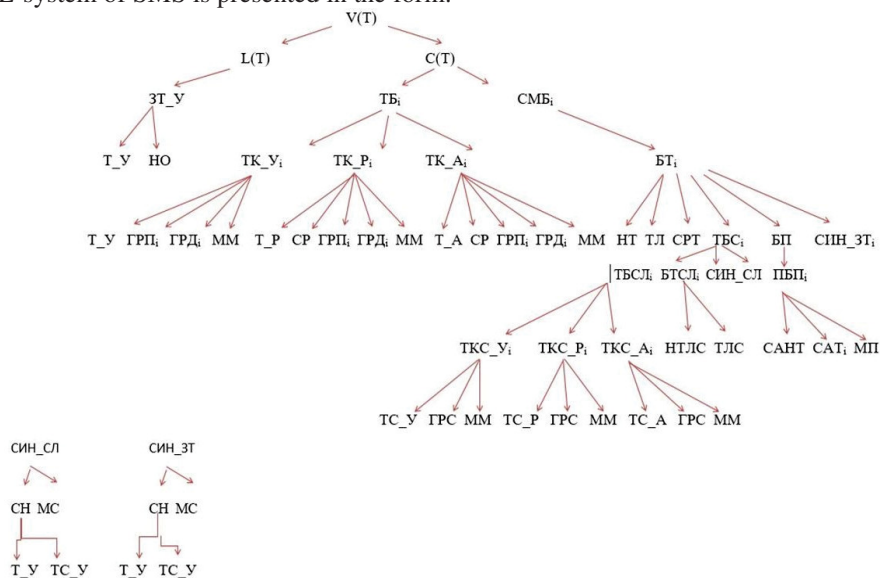


Figure 2. General scheme of the dictionary article of DUBT

The following notation is used in this scheme:

- CC — dictionary article
- 3T_Y — the title term is Ukrainian
- TB_i** — **terminological block**
- TK_Y_i** — **terminological complex ukr.**
- T_Y — Ukrainian term
- ГP — grammatical trailer
- HO — homonym number
- MM — language marker (ukr)
- TK_P_i** — **terminological complex rus.**
- T_P — Russian term
- ГP — grammatical trailer
- MM — language marker (rus)
- CP — semantic trailer
- TK_A_i** — **terminological complex eng.**
- T_A — English term
- ГP — grammatical trailer
- MM — language marker (eng)
- CP — semantic trailer
- CMB_i** — **semantic block**
- BT_i** — **interpretation block**
- HT — **interpretation number**
- CИH_i** — **synonymous block**
- CH — synonym
- MC — synonym marker (син.)

- БП** — **link block**
- ПБП_i** — **subblock of links**
- TL — **interpretation**
- CPT — semantic trailer to TL
- TBC_i** — **block of terminol. phrases**
- TBCЛ_i** — **terminol. block of phrases**
- TCK_Y_i — terminological complex of phrases ukr.
- TC_Y — terminological phrase ukr.
- ГPC — grammatical remark of the phrase
- MM — language marker (укр)
- TCK_P_i — terminological complex of phrases rus.
- TC_P — terminological phrase rus.
- MM — language marker (rus.)
- TCK_A_i — terminological complex of phrases eng.
- TC_A — terminological phrase eng.
- MM — language marker (eng)
- BTCL_i** — **block of interp. of phrases**
- TЛC — interpretation
- HTЛC — number TЛC
- CAHT — **sender**
- CAT_i — **recipient**

2.3 Marking text with XML tags according to its structure L-systems

The XML document provides an opportunity to explain all the structural elements we have defined and the relationships between them. To ensure automatic marking of the dictionary text with XML tags, a program was developed that separates the elements of the text structure in accordance with the structure of the L-system and HTML markup.

```

<CC> Словникова стаття
      <ЗТ_У номер омоніма=i>заголовний термін
український</ЗТ_У>
      <ТВ номер=i> Термінологічний блок
        <ТК_У номер=i> укр. термінологічний комплекс
        <Т_У> Термін український</Т_У>
        <ГРД номер=i> Граматична ремарка до</ГРД>
        <ГРП номер=i> Граматична ремарка до</ГРП>
        <ММ> укр. </ММ>
        </ТК_У>
        <ТК_Р номер=i> рос. терм. комплекс
        <Т_Р> Російський термін</Т_Р>
        <СР> Семантична ремарка</СР>
        <ГРД номер=i> Граматична ремарка до</ГРД>
        <ГРП номер=i> Граматична ремарка до</ГРП>
        <ММ> рос. </ММ>
        </ТК_Р>
        <ТК_А номер=i> англ. терм. комплекс
        <Т_А> Термін англійський</Т_А>
        <СР> Семантична ремарка</СР>
        <ГРД номер=i> Граматична ремарка до</ГРД>
        <ГРП номер=i> Граматична ремарка до</ГРП>
        <ММ> англ.</ММ>
        </ТК_А>
      </ТВ>
      <СМВ номер=i>
        <БТ номер=i> Блок тлумачення
          <ТЛ> Тлумачення </ТЛ>
          <НТ> Тлумачення </НТ>
          <СРТ> Семантична ремарка </СРТ>
          <СИН_ЗТ номер=i> Синонімічний блок
            <Т_У> термін</Т_У>
            <Т_С_У> термін</Т_С_У>
            <МС> Син. </МС>
          </СИН_ЗТ>
        <БТС номер=i> Блок термінологічних сл.
          <ТБСЛ номер=i> Терм. блок сл.
          <ТКС_У номер=i> Укр. терм. ком. сл.
          <Т_С_У> Термінологічне сл.</Т_С_У>
          <ГРС> Граматична ремарка</ГРС>
          <ММ> Маркер мови</ММ>
          </ТКС_У>
          <ТКС_Р номер=i> Рос. тер. ком. сл.
          <Т_С_Р> Термінологічне сл.</Т_С_Р>
          <ГРС> Граматична ремарка</ГРС>
          <ММ> Маркер мови</ММ>
          </ТКС_Р>

```

```

<ТКС_А номер =i> Англ. тер. ком. сл.
<ТС_А> Терм. словосполучення</ТС_А>
<ГРС> Граматична ремарка</ГРС>
<ММ> Маркер мови</ММ>
</ТКС_А>
</ТВСЛ>
<ВТСЛ номер =i> Блок тлумачення сл.
<ТЛС> Тлумачення до сл.</ТЛС>
<НТЛС> Номер тлумачення до сл.</НТЛС>
</ВТСЛ>
<СИН_СЛ номер=i> Синонімічний блок
    <Т_У> термін</Т_У>
    <ТС_У> термін</ТС_У>
    <МС> Син. </МС>
    </СИН_СЛ >
</ВТС>
<БП номер = i> Блок посилань
    <ПБП номер = i> Підблок посилань
        <САНТ> адресант</САНТ>
        <САТ номер=i> адресат /САТ>
        <МП> маркер посилань <МП>
    </ПБП>
</БП>
</ВТ>
</СМБ>
</СС>

```

Example

СС [словникова стаття]: **а#апа** (рос. аа#па, англ. аара) 1. Безлісий, дуже обводнений, грядково-мочарний вид болота на півночі Євразії.

According to the structure of the dictionary article, we will highlight all the structural elements.

ТБ [термінологічний блок]: **а#апа** (рос. аа#па, англ. аара)

ПТБ_У [підблок український]: **а#апа**

ТК_У [термінологічний комплекс український]: **а#апа**

ЗТ [заголовний термін]: **а#апа**

ПТБ_Р [підблок російський]: рос. аа#па

ТК_Р [термінологічний комплекс російський]: рос. аа#па

ММ [маркер мови]: рос.

Т_Р [термін російський]: аа#па

ПТБ_А [підблок англійський]: англ. аара

ТК_А [термінологічний комплекс англійський]: англ. аара

ММ [маркер мови]: англ.

Т_А [термін англійський]: аара

СМБ [семантичний блок]: 1. Безлісий, дуже обводнений, грядково-мочарний вид болота на півночі Євразії.

БТ [блок тлумачень]: 1. Безлісий, дуже обводнений, грядково-мочарний вид болота на півночі Євразії.

НТ [номер тлумачення]: 1

ТЛ [тлумачення]: Безлісий, дуже обводнений, грядково-мочарний вид болота на півночі Євразії.

A dictionary article is represented by one terminological block and one semantic block.

The terminological block consists of three sub-blocks, which are expressed by complexes. In this example, sub-blocks are equal in value to complexes. The Ukrainian terminological complex is represented by a pronominal term, the Russian terminological complex consists of a term and a language marker, and the English terminological complex consists of a term and a language marker. A semantic block is represented by a block of interpretations, which consists of an interpretation number and an interpretation. It can be shown schematically as follows:

```

<CC>
  <текст_СС><В>а#апа</В> (<I>рос.</I> аа#па, <I>англ.</I>
аара) 1. Безлісий, дуже обводнений, грядково-мочарний вид болота
на півночі Євразії.</текст_СС>
  <ЗТ> а#апа </ЗТ>
  <ТВ номер="1">
  <ТК_У номер="1">
    <Т_У> а#апа </Т_У>
    <ММ>укр.</ММ>
  </ТК_У>
  <ТК_Р номер="1">
    <Т_Р> аа#па </Т_Р>
    <ММ>рос.</ММ>
  </ТК_Р>
  <ТК_А номер="1">
    <Т_А> аара </Т_А>
    <ММ>англ.</ММ>
  </ТК_А>
  </ТВ>
  <СМБ номер="1">
  <БТ номер="1">
    <НТ>1</НТ> <ТЛ>. Безлісий, дуже обводнений,
грядково-мочарний вид болота на півночі Євразії.</ТЛ>
  </БТ>
  </СМБ>
</СС>

```

The database and site software are being tested now.

3. Discussion

Having passed a number of stages, we have achieved many advantages. The main advantage is the digital preservation of paper texts (using the developed technologies repeatedly) and the possibility of republishing them.

REFERENCES

1. Dictionary of Ukrainian biological terminology. - К.: KMM, 2012. – 746 p.
2. SHIROKOV V. A. et al.: Linguistic and information studies: works of the Ukrainian language and information fund of the NAS of Ukraine: in 5 volumes, 1(2018), Kyiv 2018. 271 p.