

Альона Дорожинська

Український мовно-інформаційний фонд НАНУ

ФОРМАЛЬНА МОДЕЛЬ ТЛУМАЧНОГО ТРИМОВНОГО ТЕРМІНОЛОГІЧНОГО СЛОВНИКА

У статті окреслено коло завдань, підходів і етапів перетворення паперового словника в онлайн-продукт, засноване на дослідженні «Словника української біологічної термінології» [1]. Серед усього словникового розмаїття була обрана ця праця тому, що цей авторитетний академічний «Словник» є першою в україністиці лексикографічною працею нового покоління, що охоплює найуживанішу біологічну термінологію української, російської та англійської мов. У процесі роботи була побудована модель лексикографічної системи, яка покладена в основу XML.

Ключові слова: *комп'ютерна лексикографія, лексикографічна система, парсинг, XML, база даних, цифровий простір, веб-сайт.*

The article outlines the range of tasks, approaches and stages of transforming a paper dictionary into an online product, based on the study «Dictionary of Ukrainian biological terminology» [1]. Among all the dictionary diversity, this work was chosen because this authoritative academic Dictionary is the first lexicographical work of a new generation in Ukrainian studies, covering the most common biological terminology of Ukrainian, Russian and English. In the process, a model of the lexicographic system was built, which is the basis of XML.

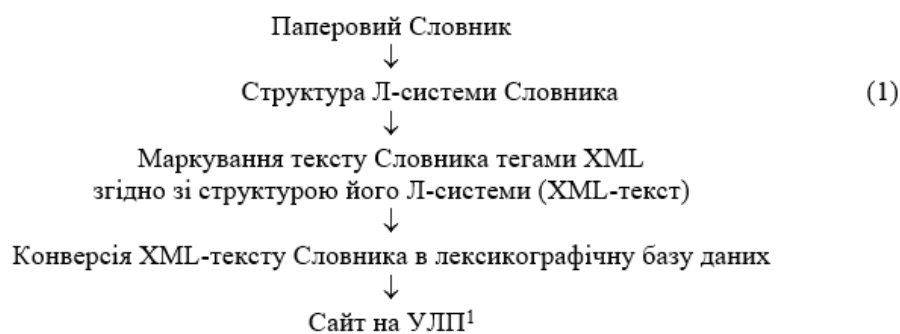
Keywords: *Computer lexicography, lexicographic system, parsing, XML, database, digital space, website.*

Вибуховий ріст інформації у вебi останнім часом привів до необхідності якісно нових засобів опрацювання даних [2]. Сьогодні створено досить багато інструментальних засобів для автоматизації окремих етапів термінологічної роботи, проте універсальних рішень основних задач цифрової термінографії досі не розроблено. Особливо це стосується проблематики, пов'язаної із цифровою рецепцією традиційної термінологічної спадщини, насамперед, багатомовної.

У цій роботі ми продемонструємо ефективний підхід для трансформації паперового словника в онлайн-продукт на прикладі перетворення «Словника української біологічної термінології» (СУБТ) [1] на онлайн-цифровий словник.

За словами авторів, створення цього Словника зумовлене, насамперед, необхідністю зібрати й систематизувати сучасну українську наукову біологічну термінологію. Потреба в такій праці назріла давно, але особливо гостро постає тепер, у період стандартизації та кодифікації наукової термінології, поповнення її термінами, що позначають нові реалії багатоаспектної біологічної науки. Запропонована термінографічна праця обіймає нормативну загальнонаукову та широковживану термінологію біологічних наук, зафіксовану в сучасних енциклопедичних, загальномовних та спеціальних словниках, у науковій, науково-популярній, навчальній та інформативно-реферативній літературі. Словник покликаний дати читачеві максимум відомостей щодо сучасного стану української національної біологічної терміносистеми, а також еквівалентних форм вираження відповідних понять російською та англійською мовами.

Запропонований нами підхід до цифрової трансформації передбачає наступні етапи:



1

¹ УЛП – Український лінгвістичний портал

Цей процес, на нашу думку, містить етапи, які можна застосувати й до інших словників, тому ми вважаємо, що наведена послідовність представляє ефективний і універсальний шлях переведення традиційних словників до цифрового формату. Розглянемо окремі кроки, що складають процес перетворення.

Структура лексикографічної системи (Л-системи) СУБТ

Слідуючи за теорією лексикографічних систем [3], структуру Л-системи СУБТ представимо у вигляді:

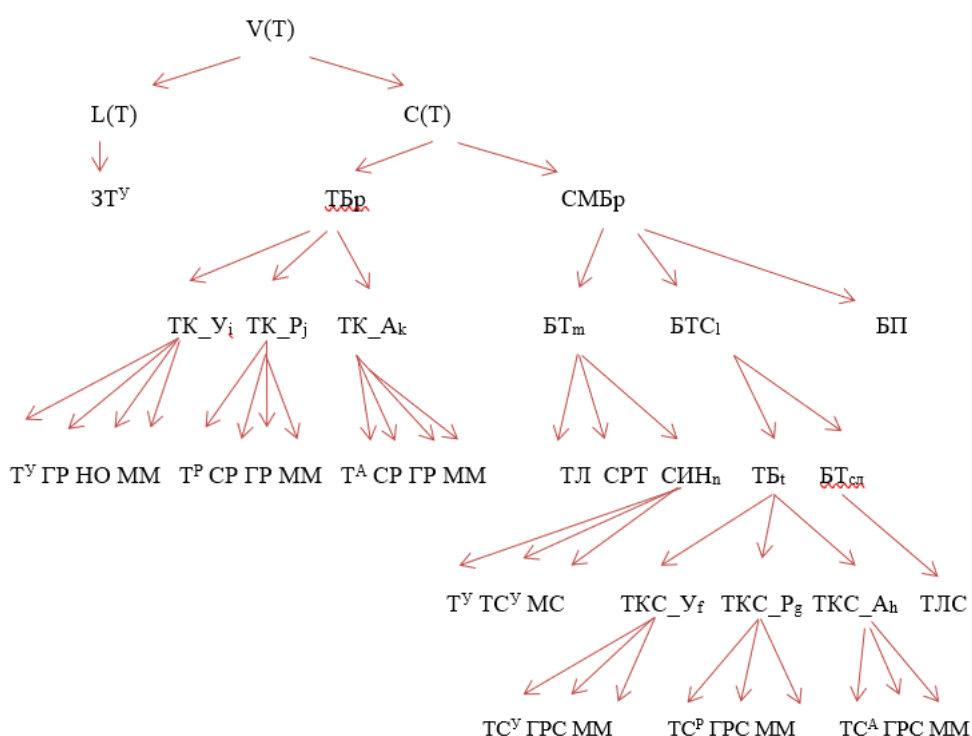


Схема 1. Загальна схема словникової статті СУБТ

У цій схемі вжито такі

позначення:

V(T) – словникова стаття

L(T) – ліва частина

C(T) – права частина

ZTY – заголовний термін

український

TBp – термінологічний блок

SMBr – семантичний блок

TK_Ui – термінологічний комплекс
укр.

TK_Pj – термінологічний комплекс
рос.

TK_Ak – термінологічний комплекс
англ.

TU – термін український

ГР – граматична ремарка

НО – номер омоніма	ТС ^А – термінологічне
ММ – маркер мови (укр)	словосполучення на
Т ^Р – термін російський	англійській мові
ГР – граматична ремарка	ММ – маркет мови (англ)
ММ – маркер мови (рос)	
СР – семантична ремарка	
Т ^А – термін англійський	
ГР – граматична ремарка	
ММ – маркер мови (англ)	
СР – семантична ремарка	
БТ _м – блок тлумачень	
БТС ₁ – блок термінол. словоспол	
БП – блок посилань	
ТЛ тлумачення	
СРТ – семантична ремарка до	
тлумачення	
СИН _н – синонімічний блок	
ТУ – термін український	
ТС ^У – термінологічне сл. українське	
МС – маркер синоніма (син.)	
ТБ _т – термінологічний блок	
словосполучень	
БТ _{сл} – блок тлумачень	
словосполучень	
ТЛС – тлумачення	
ТСК_У _f – термінологічний комплекс	
словосполучення укр.	
ТСК_Р _g – термінологічний комплекс	
словосполучення рос.	
ТСК_А _h – термінологічний комплекс	
словосполучення англ.	
ТС ^У – термінологічне	
словосполучення на	
українській мові	
ГРС – граматична ремарка	
словосполучення	
ММ – маркер мови (укр)	
ТС ^Р – термінологічне	
словосполучення на	
російській мові	
ММ – маркер мови (рос.)	

Маркування тексту СУБТ тегами XML згідно зі структурою його Л-системи

Наступний етап – автоматичне конвертування тексту словника в XML-документ, що надає можливість експлікувати всі визначені нами структурні елементи та зв'язки між ними. Для забезпечення автоматичного маркування тексту словника тегами XML було розроблено програму, яка виокремлює елементи структури тексту відповідно до будови Л-системи. Використовуються поліграфічні ознаки текстової ідентифікації Л-системи, а саме: межі словникової статті (абзаци), різні спеціальні символи, які відмежовують структурні елементи, позиційні характеристики, зміни мови, шрифтів, регістру літер та ін.

Надалі планується на основі XML-документа формування відповідної бази даних в автоматичному режимі та створення сайту. Ці етапи будуть розглянуті в наступній статті.

Пройшовши ряд етапів ми досягнули багато переваг:

1. Це дозволило нам подавати словник сучасним способом, а не прив'язувати його до старого формату.
2. Перетворення даних дозволило нам зафіксувати багато помилок та невідповідностей.
3. Належний XML полегшить реалізацію функціоналу пошуку і це також спростить створення прямих перехресних посилань на веб-сайті.
4. Якщо автори вирішать внести будь-які доповнення або зміни у майбутньому, вони можуть це зробити, використовуючи сучасну систему редагування словника замість текстового редактора.
5. Стає можливим зверстати книгу в InDesign та перевидавати.

Хоч ще і залишилися питання було продемонстровано, що можна оцифрувати паперовий словник та створити веб-сайт для термінологічної праці. Це є хороший спосіб перетворити та зберегти словники в XML та в Інтернеті.

Література

1. Словник української біологічної термінології. К.: КММ, 2012. 746 с.
2. Лінгвістично-інформаційні студії : праці Українського мовно-інформаційного фонду НАН України : у 5 т. / В. А. Широков та ін. Т. 1 : Наукова парадигма та основні мовно- інформаційні структури. Київ. Український мовно-інформаційний фонд НАН України. 2018. 271 с. DOI 10.33190/978-966-02-8683-2/8684-9.
https://movoznavstvo.org.ua/files/tom_1_B5_print.pdf
3. Лінгвістично-інформаційні студії: праці Українського мовно-інформаційного фонду НАН України : у 5 т. / В. А. Широков та ін. Т. 5 : Віртуалізація лінгвістичних технологій. Київ. Український мовно-інформаційний фонд НАН України. 2018. 239 с. DOI 10.33190/978-966-02-8683-2/8690-0
https://movoznavstvo.org.ua/files/Ling_inf_studio_TOM_5_umif_B5.pdf