

## КОМП'ЮТЕРНА ЛІНГВІСТИКА

**„СЛОВАРЬ УКРАЇНСЬКОЇ МОВИ” ЗА РЕД. Б. ГРІНЧЕНКА  
(1907-1909 РР.): ЕЛЕКТРОННА ВЕРСІЯ****© Валентина Балог, 2007**

к. філол. н., Інститут української мови НАН України (Київ)

**© Олександр Балог, 2007**

Інститут української мови НАН України (Київ)

УДК 161.2.81'374.72'22

*У статті обґрунтовано й описано обрану методику та процедуру організації лексикографічного матеріалу „Словаря української мови” за ред. Б. Грінченка для функціонування його в електронному вигляді: створення бази даних на основі відсканованого тексту словника та інтерфейсу за допомогою мови програмування Delphi 8.*

Створення електронного варіанта «Словаря української мови» у 4-х томах за ред. Б. Грінченка – це проект, упровадження якого здійснюють співробітники групи Національного корпусу відділу лексикології, лексикографії та Національного корпусу української мови Інституту української мови НАН України.

Швидкий розвиток інформаційних технологій, розширення електронних бібліотек та зростання кількості літератури на електронних носіях, зокрема й словників – це лише невелика частина умов, що спровокувала цю роботу. «Словарь української мови» попри свою давність (а словники належать до категорії праць, що мають термін давності) не втратив свого функціонального призначення до сьогодні у світлі суспільно-історичних та, зокрема, лінгво-історичних процесів в Україні.

Базовим для електронного варіанта словника стало останнє його паперове видання 1996 р. з додатком О. О. Тараненка, ухвалене вченою радою Інституту української мови НАН України, надруковане з видання 1907–1909 рр. фотоспособом. Вибір видання ґрунтувався на якості друку матеріалу з метою полегшення процесу сканування та розпізнавання тексту комп'ютерними програмами.

Загалом створення електронного словника складається з трьох етапів<sup>1</sup>:

---

<sup>1</sup> Описання етапів подано в спрощеному вигляді, оскільки для створення повнофункціонального електронного словника та запису його на електронний носій інформації повинні бути написані відповідні DLL-бібліотеки та Ini-файли з метою

- створення бази даних словника;
- створення інтерфейсу для роботи з цією базою даних за допомогою мови програмування високого рівня;
- запис одержаного електронного словника на диск.

Створення бази даних відбувається шляхом:

- сканування паперового варіанту словника;
- вичитування та виправлення відсканованого тексту в Microsoft Word;
- імпорту виправленого тексту з Microsoft Word в Microsoft Excel;
- розбиття тексту в електронній таблиці Microsoft Excel на відповідні поля;
- імпорту отриманих таблиць в СУБД Microsoft Access.

Створення інтерфейсу електронного словника передбачає:

- зв'язок з базою даних через стандарт ODBC;
- безпосереднє створення інтерфейсу.

#### **Створення бази даних.**

У процесі виконанні поставленого завдання були запропоновані дві системи управління базами даних: сервер InterBase та СУБД Microsoft Access.

Промисловий сервер бази даних InterBase призначений для розв'язання широкого кола задач. Він поєднує в собі високу надійність та простоту установки. Проте введення даних необхідно робити вручну, використовуючи запити SQL, що є досить тривалим процесом. Ураховуючи цей недолік, а також кількість дефініцій в розроблюваному електронному словникові (близько 75 тис.), у цьому завданні ми надали перевагу системі управління базами даних (СУБД) Microsoft Access, яка дозволяє з більшою зручністю та економією часу створювати бази даних та програми для роботи з даними, проводити управління та контроль за ними.

У процесі створення бази даних в СУБД Microsoft Access було виконано такі завдання:

- створено електронну таблицю словника, імпортовано дані з Microsoft Word в Microsoft Excel;
- в одержаній електронній таблиці розділено відповідно на два стовпці заголовне слово та словникову статтю за допомогою функцій, які підтримує Microsoft Excel;
- створено третій стовпець, в кожному рядку якого вказується літера, на яку починається слово в цьому рядку;
- імпортовано дані з електронної таблиці Microsoft Excel в СУБД Microsoft Access;

---

забезпечення повноцінного його функціонування незалежно від комп'ютера, на якому він запускається.

- визначено первинні ключі в кожній з утворених таблиць в СУБД Microsoft Access;
- зв'язано утворені таблиці в СУБД Microsoft Access відповідно до структури словника.

Оскільки словник повинен містити п'ять частин, то в результаті виконаної роботи було створено п'ятнадцять таблиць, по три відповідно для кожної частини:

- „Alphabet\_x”,
- „Word\_x”,
- „Article\_x”,

де x – це номер відповідної частини електронного словника.

Кожна з таблиць містить певні поля, деякі з яких можуть візуально відображатися на формі інтерфейсу створюваного електронного словника.

Первинні ключі:

- в таблиці Alphabet\_x – поле „Алфавіт” символного типу;
- в таблиці Word\_x – поле „№\_Слова” числового типу;
- в таблиці Article\_x – поле „№\_Статті” числового типу.

Зв'язок між таблицями:

- зв'язок „Alphabet\_x - Word\_x” – один до багатьох,
- зв'язок „Word\_x - Article\_x” – один до одного.

Такий зв'язок між таблицями є стандартним при створенні різноманітних електронних словників, зокрема словника Lingvo x.x.

#### **Створення інтерфейсу електронного словника.**

Для створення інтерфейсу цього електронного словника було обрано мову високого рівня Delphi 8. Вибір цієї мови був обумовлений такими факторами:

- система програмування ® Delphi™ For Microsoft ® .NET Framework – складний програмний продукт, який надає програмісту всі необхідні засоби для створення програм будь-якої складності на призначення; характерною особливістю системи є підтримка нової технології .NET;
- в мові високого рівня Delphi 8 наявний широкий спектр різноманітних функцій для роботи з базами даних.

При створенні інтерфейсу електронного словника перш за все необхідно налаштувати зв'язок з наявною базою даних через ODBC. За допомогою цього стандарту створюється псевдонім бази даних (*Database Alias*), за яким відбувається звернення до неї. Сама база даних при цьому зберігається або в каталозі на дискові, або знаходиться на віддаленому серверові.

ODBC – це стандарт, який описує систему інтерфейсів, за допомогою яких прикладні програми можуть звертатися до баз даних та обробляти їх незалежно від вибраної СУБД. ODBC надає інтерфейси для доступу до реляційних баз даних та до баз даних з табличною організацією. Широке поширення цей стандарт отримав завдяки підтримці Microsoft.

Для доступу до даних в системі програмування був створений окремий модуль даних – невізуальний контейнер, де містяться компоненти доступу до даних, а також всі необхідні компоненти, такі як:

- TDatabase – для зв'язку з базою даних;
- TTable – для зв'язку з таблицями бази даних, підключення даних;
- TDataSource – для передачі даних у візуальні компоненти; відстежування синхронних змін їхнього стану (якщо змінюється стан зв'язаного з ними набору даних, відбувається повернення даних в набір);
- TDataSet – клас, на базі якого побудовані компоненти, в яких зберігаються набори даних.

У цьому разі набори даних – це масиви записів, які отримані з бази даних.

За допомогою інспектора об'єктів були виконані всі необхідні налаштування для зв'язку з базою даних.

Для управління даними в системі програмування було створено п'ять форм, кожна з яких відповідає окремій частині електронного словника. На кожній з цих форм були розміщені такі компоненти як:

- IDBGrid – для візуального відображення алфавіту та слів;
- TDBMemo – для візуального відображення словникових статей;
- TMainManu;
- TButton;
- TEdit.

Для створення швидкого пошуку слова в електронному словнику в створеному модулі даних був розміщений компонент TQuery (черга), та в редакторі SQL, властивості SQL, компонента TQuery, був написаний відповідний SQL-запит.

Також був написаний відповідний програмний код та SQL-запити для введення або видалення записів в базі даних.

Таким чином, після остаточного вичитування текстового матеріалу та експлуатації в тестовому режимі, електронна версія «Словаря української мови» за ред. Б. Грінченка може бути записана на диск та запропонована користувачеві як готовий продукт, що дасть можливість користуватися словником і зберегти цю важливу для української мови працю в якісно новому форматі.

#### Література

1. Сорокин А. Delphi. Разработка баз данных. – СПб.: Питер, 2005.