

словосполучень необхідно створити комп'ютерні перекладні словники з розвинутою системою багаторівневої спеціалізації. В таких словниках одночасно з базовими словами повинні бути гнізда, побудовані на цьому слові. Для цього у словосполученнях потрібно зберігати природний порядок слів, у якому вони вживаються. Наприклад, у словнику з радіоелектроніки базове слово "корекція" матиме гніздо, в якому в алфавітному порядку будуть наведені словосполучення "корекція спотворень; корекція високих частот; корекція зі стиранням помилок" і т.п. Для базового слова "кореляційний" - аналогічно: "кореляційний аналіз; кореляційний декодер; кореляційний інтеграл" і т.п.

За такими принципами побудований двомовний перекладний словник з радіоелектроніки [2], його російсько-українська частина.

Таким чином, запропонований алгоритм комп'ютерного перекладу з використанням нових словників словосполучень з розвинутою системою багаторівневої спеціалізації повинен забезпечити високий рівень автоматичного перекладу спеціалізованих текстів.

1. *Webster's Encyclopedic Unabridged Dictionary of the English Language. Gramercy Books, N.Y., New Jersey. 1989, 1854 pgs.* 2. *Російсько-український та українсько-російський словник з радіоелектроніки. Б.Рицар, К.Семенистий, І.Кочан. Логос, Львів, 1995, 608с.*

## УКЛАДАННЯ СЛОВНИКІВ ЗА ТЕХНОЛОГІЄЮ ЛЕКСИКОГРАФІЧНОЇ БАЗИ ДАНИХ

© Ігор Кульчицький, Василь Костирко, 2000

Львівська комерційна академія

В даній роботі розглянуто застосування технології укладання словників з використанням баз даних на персональних комп'ютерах [1]. Основу цієї технології складає лексикографічна база даних (ЛБД), що розглядається як спеціалізована надбудова над стандартною системою керування базами даних. Укладання певного словника за цією технологією має такі чотири етапи:

- побудова абстрактної моделі цього словника;
- генерація структури та прикладного програмного забезпечення лексикографічної бази даних;
- наповнення лексикографічної бази даних;
- конвертування лексикографічної бази даних до належного кінцевого вигляду.

Побудова абстрактної моделі словника базується на таких його визначальних рисах:

• **Членоване розміщення матеріалу.** Основною композиційною і комунікативною одиницею словника є незалежний відрізок тексту, що називається **словниковою статтею**. Їх сукупність складає **основу** словника.

• **Ліва та права частини.** Кожну словникову статтю можна розділити на ліву та праву частини. Ліва частина — це будь-яка одиниця мови, що є об'єктом лексикографування і вноситься у словник. Назвемо її лексикографічною одиницею, а сукупність всіх лексикографічних одиниць словника його реестром. Праву частину словникової статті творять співвіднесені з лексикографічною одиницею лексикографічні параметри. Лексикографічний параметр — це деякий квант лінгвістичної інформації, що має самостійний інтерес для користувача і є словниковим відображенням структурних рис мови (наприклад, орфографічний параметр, графічна довжина слова, наголос, вимова, поділ на склади, частина мови, переклади, глумачення тощо).

• **Системність.** У словнику реєстрові одиниці пов'язані між собою численними семантично-структурними зв'язками і в своїй сукупності творять певну систему, що відповідає задуму та меті укладачів. Словникові статті всякого словника мають в його межах однакову схему опису однотипних лексикографічних одиниць.

• **Впорядкованість.** Лексикографічні одиниці у кожному словнику обов'язково подають впорядкованими за якимось критерієм (найчастіше – за абеткою).

• **Допоміжний матеріал.** Будь-який словник окрім словникових статей, хоча б у мінімальній кількості, містить додаткову інформацію, що допомагає ним користуватись: передмову, критерії впорядкування та опис структури словникових статей, список скорочень, граматичні таблиці, джерела укладання тощо.

Таким чином абстрактна модель словника має три структурних рівні:

- макрорівень, який визначає словник як самостійну систему з її внутрішніми зв'язками та багатоплановою організацією;
- медіорівень, на якому лексикографічні одиниці об'єднуються у поля та групи на основі семантичних, тематичних та асоціативних ознак;
- мікрорівень, де визначаються параметри словникової статті: її структура, форми та способи розкриття семантики лексикографічних одиниць, ієрархія їхніх значень тощо.

Генерація структури та прикладного програмного забезпечення ЛБД передбачає породження порожніх структур для збереження інформації про словникові об'єкти і взаємозв'язки між ними та підключення відповідних програмних модулів, які дозволяють реалізувати у ЛБД абстрактну модель конкретного словника. До таких модулів належать:

• Модуль лексикографічних одиниць, який забезпечує їх додавання, редагування, нищення та побудову реєстру словника.

• Модуль міжмовних зв'язків, який забезпечує взаємозв'язки між перекладами лексикографічної одиниці різними мовами.

• Модуль омонімії, який забезпечує утворення, редагування та нищення оморядів лексикографічних одиниць.

• Модуль полісемії, який забезпечує утворення, редагування та нищення лексикографічних одиниць–полісемів.

• Модуль синонімії, який підтримує утворення, редагування та нищення синонімічних рядів лексикографічних одиниць.

• Модулі лексикографічних параметрів. Для кожного лексикографічного параметра підключається окремий модуль, який дозволяє його внесення у ЛБД, редагування та нищення. Кількість та конкретний набір цих параметрів визначається абстрактною моделлю словника.

• Модуль підтримки хронології, який забезпечує коректну обробку історичних дат, що характеризуються своєю різномірністю (різні стилі), неповною визначеністю та часто мають описувальний характер.

• Модулі конвертації вмісту ЛБД до кінцевих словникових форм. Найпоширенішими на сьогоднішній день кінцевими формами словника є книжна та електронна (інформаційно-довідкова система).

• Модулі сортування лексикографічних одиниць, що дозволяють впорядковувати реєстр словника за різними критеріями.

- Модуль представлення, який дозволяє відобразити на екрані комп'ютера вміст ЛБД різними способами.
- Модулі мовних середовищ, які дозволяють оперувати інформацією різними мовами (наприклад, у перекладних словниках).
- Модулі-утиліти, що виконують службові функції, такі як підтримка цілісності ЛБД, контроль за структурною коректністю інформації, архівування та стиснення даних, чорновий друк тощо.

На базі описаної технології авторами разом з укладачами було створено сімейство словників різних типів і обсягів реєстру [2-9]. Різноманітність типів словників та участь у їх укладанні фахівців різних галузей дозволила належно розвинути цю технологію і демонструє її індустріальний характер. Розглянемо особливості реалізації характерних типів словників.

Перекладні словники [4]. Генерувалися модулі: лексикографічних одиниць, міжмовних зв'язків, омонімії, полісемії, синонімії, конвертації до книжкової форми, впорядкування лексикографічних одиниць, представлення, мовних середовищ, утиліти. Використано модулі лексикографічних параметрів: орфографічний, наголосу, граматичної характеристики, стилістичної характеристики, перекладів іншими мовами.

Глумачні словники [4, 8]. Генерувалися модулі: лексикографічних одиниць, омонімії, полісемії, синонімії, конвертації до книжкової форми, впорядкування лексикографічних одиниць, представлення, утиліти. Використано модулі лексикографічних параметрів: орфографічний, наголосу, граматичної характеристики, стилістичної характеристики, тлумачення.

Таксономічні словники [7]. Генерувалися модулі: лексикографічних одиниць, міжмовних зв'язків, полісемії, синонімії, конвертації до книжкової форми, впорядкування лексикографічних одиниць, представлення, мовних середовищ, утиліти. Використано модулі лексикографічних параметрів: орфографічний, наголосу, перекладів іншими мовами.

Біографічні словники [3, 5, 6, 9]. Генерувалися модулі: лексикографічних одиниць, полісемії, синонімії, конвертації до книжкової форми, впорядкування лексикографічних одиниць, представлення, утиліти. Використано модулі лексикографічних параметрів: орфографічний, наголосу, хронологічний, енциклопедичної статті.

Орфографічні словники [2]. Генерувалися модулі: лексикографічних одиниць, омонімії, синонімії, конвертації до книжкової форми, впорядкування лексикографічних одиниць, представлення, утиліти. Використано лексикографічні параметри: орфографічний, наголосу, граматичної характеристики.

На основі вищесказаного можна зробити такі висновки:

1. Практично обов'язковими модулями для словників будь-якого типу є модулі:
  - лексикографічних одиниць,
  - полісемії,
  - синонімії,
  - конвертації до книжкової форми,
  - впорядкування лексикографічних одиниць,
  - представлення,
  - утиліт.
2. Модуль омонімії є обов'язковим для всіх словників філологічного типу.

3. Модулі міжмовних зв'язків та мовних середовищ використовуються для перекладних та таксонімічних словників.
4. Склад та кількість лексикографічних параметрів, в основному, визначають тип та сферу застосування словника.
5. Найбільш типовими для словників є наступні лексикографічні параметри:
  - орфографічний — побуквенна передача лексикографічної одиниці, обов'язковий для словників всіх типів;
  - наголос — обов'язковий для словників всіх типів;
  - граматична характеристика — обов'язкова для всіх філологічних словників;
  - стилістична характеристика — використовує більшість філологічних словників;
  - переклад іншою мовою — перекладні та таксонімічні словники;
  - тлумачення;
  - енциклопедична стаття
  - хронологічний параметр — типовий для енциклопедичних словників.

1. В. Костирко, І. Кульчицький. Використання термінологічних словникових баз даних та пакетів прикладних програм в сучасній лексикографії. // Вісник: Проблеми української термінології. Матеріали 5-ї міжнародної наукової конференції. — Львів: Львівська політехніка, 1998, № 336. — с.105-108. 2. Орфографічний словник українських медичних термінів. — Львів: Видавнича спілка "Словник" ЛДМУ, 1993. — 473с. 3. Ярославська гімназія 1940-1944 рр. Книга Пам'яті з нагоди 50 річчя останньої матури. Упорядник І. Гнаткевич. — Львів: 1994. — 176с. 4. Українсько-латинсько-англійський медичний тлумачний словник. У 2-х томах. — Львів: Видавнича спілка "Словник" ЛДМУ, 1995. — Т. 1 - 551 с., т. 2 - 786с. 5. Українська державна чоловіча гімназія у Перемишлі. 1895-1995. Упорядник І. Гнаткевич. — Дрогобич: Видавнича фірма "Відродження", 1995. — 304 с. 6. Український Інститут для дівчат у Перемишлі. 1895 - 1995: Ювілейна Книга Пам'яті до 100-річчя заснування./ Упорядник І.Гнаткевич. Дрогобич: Видавнича фірма "Відродження", 1995. — 168 с. 7. Зведений список мохів України. Упорядники: І.С.Данилків, П.Р.Третяк, Л.М.Петрова, І.М.Кульчицький. — Львів: Українське ботанічне товариство, 1995. — 35с. 8. Демська О.Кульчицький І.М.Словник омонімів української мови. — Львів: Видавництво "Фенікс", 1996. — 224с. 9. Партизанськими дорогами з командиром "Залізнякам": Упорядник, вступне слово Ігор Гнаткевич. — Дрогобич: Видавнича фірма "Відродження", 1997. — 359с.

## ВІДОБРАЖЕННЯ ХРОНОЛОГІЇ У БАЗАХ ДАНИХ

© Ігор Кульчицький, Анатолій Костенко, 2000

Львівська комерційна академія

На сьогоднішній день для уніфікації терміносистем з дотриманням вимог цілісності, експлікативності, повноти та несуперечливості [1], укладання термінологічних словників, які найповніше відбивають суть понять, повністю відповідають нормам і правилам сучасної української літературної мови, забезпечують погодженість термінів та задовольняють вимогам як фахівців у певній галузі, так і мовознавців [2], широко використовують лексикографічні бази даних (ЛБД) [3-5].

Важливим питанням функціонування ЛБД є відображення та обробка хронологічних даних, позаяк існуючі системи керування базами даних (СКБД) не мають стандартних засобів маніпуляції історичними датами. Вкажемо основні проблеми комп'ютерної обробки датування історичних подій:

1. У історичних документах події датовано за різними стилями, календарями та системами. Ці дати необхідно приводити до єдиної системи.
2. Стандартні засоби СКБД для відображення хронології надають такі одиниці: секунда, хвилина, година, день, місяць, рік. В той же час традиційно історія оперує іншими часовими категоріями.