

## ТЕХНОЛОГІЧНІ АСПЕКТИ РЕАЛІЗАЦІЇ ОНЛАЙНОВОЇ ЛЕКСИКОГРАФІЧНОЇ СИСТЕМИ „СЛОВНИКИ УКРАЇНИ”

Розкриваються основні технологічні аспекти реалізації онлайнкової лексикографічної системи „Словники України”. Описується архітектура системи, програмні засоби реалізації проекту, розглядаються внутрішньосистемні інформаційні потоки, розкривається внутрішня модель. Звертається увага на проблемні аспекти реалізації системи та ключові принципи організації інтерфейсу користувача.

### Вступ

Характерною рисою сучасності є активізація інформаційних процесів, потужним інструментом якої стали Інтернет-технології, що відкрили нові напрямки у розвитку і функціонуванні інформаційних систем. Серед останніх чільне місце займають системи спеціального типу – лексикографічні. Під поняттям лексикографічної системи (Л-системи) розуміємо абстрактний мовно-інформаційний об’єкт, орієнтований на реалізацію комплексного інформаційного опису лексики довільної мови або сукупності мов [1, 2]. Зважаючи на викладене реалізація подібних систем у глобальному мережевому середовищі стає все актуальнішим завданням. Роботи над створенням Л-систем у середовищі Інтернету проводяться в Українському мовно-інформаційному фонді (УМІФ) НАН України. Однією з розробок зазначеного напрямку є Л-система „Словники України он-лайн”, технологічним аспектам побудови якої присвячена ця стаття.

### Загальний опис системи

Однією з проблем прагматики існуючих словників (не беручи до уваги обсяг інформації) є розсіяння різних відомостей і функцій по багатьох різних джерелах. Вирішенням цієї проблеми стає поєднання словників, але не внаслідок механічної комбінації, а завдяки інтеграції їх до єдиної лексикографічної системи, яка створює своєрідне середовище, що представляє різноманітні функції для роботи користувача та розвинені пошукові можливості [3].

Описаний у статті онлайнвий програмний словниковий комплекс призначений відтворити в мережевому середовищі

функціонування інтегрованої лексикографічної системи „Словники України”, версія 1.03 [4], випущеної Українським мовно-інформаційним фондом НАН України у вигляді лазерного диску в 2003 році.

До складу Л-системи „Словники України” як агрегованого об’єкта ввійшли чотири словникові підсистеми: „Словозмінна парадигма”, „Фразеологія”, „Синонімія” та „Антонімія”.

Основу генерального реєстру (понад 186 тис. слів) системи "Словники України он-лайн" складає реєстр Орфографічного словника української мови [5], який практично повністю відтворено й значно розширено.

Підсистему "Словозмінна парадигма" створено на основі розробленої в УМІФ НАН України словозмінної класифікації української лексики, в якій виділено за певними формальними ознаками понад 2000 парадигматичних класів для всіх відмінюваних повнозначних частин мови [6,7]. Завдяки вказаній класифікації та реалізованій програмно процедурі парадигматизації – побудові повної словозмінної парадигми виходячи з канонічної (словникової) форми лексеми – побудовано повний перелік усіх граматичних форм для всіх лексичних одиниць, наведених у реєстрі. Це забезпечило можливість візуалізації словоформ в усіх граматичних значеннях. Загальна кількість словоформ для понад 186 тис. одиниць реєстру сягає близько 3,4 млн. Підсистема забезпечує відображення таблиці всіх словоформ для вибраної реєстрової одиниці з поданням їх граматичних параметрів.

Підсистема синонімії з достатньою повнотою відображає синонімічне багатство

української мови (на матеріалах літературної мови XIX-XX ст. та народнорозмовної лексики). Джерелом лінгвістичної інформації послуговував "Словник синонімів української мови" у 2-х томах [8]. Програмний комплекс забезпечує відтворення синонімічних рядів (близько 9200), що складаються зі слів або їхніх окремих значень, а також сталих словосполучень (номінативних та граматичних). Ядром кожного синонімічного ряду є його домінанта – лексична одиниця, що має найзагальніші для цього ряду семантичні особливості. Члени синонімічних рядів супроводжуються семантичними, граматичними та стилістичними характеристиками. Уживання синонімів ілюструється типовими для них контекстами – цитатним матеріалом із художньої, публіцистичної, наукової, іншої літератури та словосполученнями.

Базою підсистеми антонімії став Словник антонімів української мови [9], у якому подано 253 словникові статті, котрі представляють близько 2200 компонентів антонімічних пар.

Лінгвістичним джерелом наповнення лексикографічної бази підсистеми „Фразеологія” стали близько 56 тис. фразеологічних одиниць, представлених у Словнику фразеологізмів української мови, що вийшов у серії "Словники України" [10], де найповніше відображено загальноновживану фразеологію української мови і дано всебічну лексикографічну характеристику фразеологізмів.

Яскравою характеристикою масштабу Л-системи „Словники України” є обсяг роздруку його змісту на папері – він сягає близько 40 тис. сторінок, зрозуміло, без урахування внутрішніх зв’язків Л-системи, які забезпечують його пошукові можливості, принципово відрізняючи цей лінгвістично-програмний продукт від його традиційних аналогів. За нашими оцінками, Л-система „Словники України он-лайн” на даний момент представляє найбільший онлайн-новий український словниковий ресурс в Інтернеті.

### **Архітектура системи „Словники України он-лайн”**

Тим часом реалізація онлайн-ової версії Л-системи „Словники України” за сво-

їми системними підходами суттєво відрізняється від локального варіанту у вигляді лазерного диску. Вивчення моделі предметної галузі та аналіз варіантів реалізації поставлених завдань зумовили вибір багаторівневої схеми “клієнт-сервер” в якості базової архітектури для проектування і побудови системи. Подібна архітектура дозволяє досягти високого рівня масштабованості застосувань, значно спрощує його супровід та дає можливість багаторазового використання його компонентів [11]. Архітектура такого типу забезпечує зручний метод побудови застосувань типу “клієнт-сервер” з використанням стандартних інтерфейсів і методів. При цьому для кожного архітектурного рівня використовуються незалежні сервіси і процеси, завдяки чому забезпечується системотехнічна декомпозиція базових компонентів системи на відносно незалежні основні частини, що спрощує і значною мірою уніфікує розробку її програмних модулів. Поєднання клієнтських і серверних технологій забезпечує максимальну продуктивність роботи системи.

Першим рівнем вищезгаданої архітектури є “клієнт – термінал”, який формує запит й отримує та відображає результати його виконання. Наступний архітектурний рівень – „сервер застосувань” – приймає запити від клієнтів і виконує їх обробку. Для типових мережевих застосувань, побудованих за технологією Інтернет/Інтранет він є комбінацією спеціалізованого програмного забезпечення системи обробки запитів клієнта у комплексі з типовим HTTP-сервером. Останній рівень архітектури “клієнт-сервер” представлено системою керування базами даних.

### **Програмні засоби розробки та функціонування системи**

У якості Web-сервера було обрано Apache – один з найпоширеніших серверів Інтернету, популярність якого зумовлена його швидкодією, технологічним лідерством, розширеною підтримкою функцій серверної сторони, модульною архітектурою та вільним ліцензуванням.

В УМІФ НАН України Web-сервер (версія 2.0.48) функціонує під керуванням операційного середовища Red Hat Linux r.9.0, використання якого зумовлено двома

причинами. По-перше, система Red Hat Linux відноситься до програмного забезпечення з відкритим кодом і розповсюджується згідно з GNU Public License. По-друге, використання Unix/Linux платформи надає можливість забезпечення високого ступеня надійності та захищеності серверної частини програмних систем в мережі Інтернету.

Сервер застосувань розроблено на мові інтерпретації сценаріїв PHP4 в поєднанні з HTML та JavaScript сценаріями.

В результаті аналізу альтернативних варіантів вибору системи керування базами даних обрана SQL92-сумісна система PostgreSQL. Вагомими аргументами стали: модульна архітектура цього програмного забезпечення; підтримка великої кількості типів даних (у тому числі можливість розробки власних); доступність інтерфейсів програмування PostgreSQL для таких мов, як C, Perl, Python, Tcl, Java і PHP [12].

### Інформаційні потоки всередині системи

Окреслимо головні напрямки потоків даних між апаратно-програмними засобами

системи. На рис.1 цифрами позначено:

1. Транспортування даних з внутрішнього сервера до сервера баз даних онлайнового проекту. Л-система „Словники України он-лайн” не обмежена рамками реалізації в глобальній мережі, а являє собою лише один з варіантів представлення зовнішньої моделі лексикографічної системи. Технологічне ядро зосереджене у спеціальному програмному комплексі, що функціонує в локальній мережі Українського мовно-інформаційного фонду. Одним із технологічних завдань процедур зазначеного програмного засобу є формування даних для наповнення бази онлайнної лексикографічної системи. Оскільки програмні платформи функціонування вищезгаданої системи та онлайнного проекту є різними (Windows NT 4.0 та Red Hat Linux r.9.0.) , так само як і різні системи керування базами даних (Microsoft SQL Server 7.0 та PostgreSQL 7.3), важливою ланкою роботи є конверсія отриманих даних. Їх транспортування супроводжується зміною системи кодування та встановленням відповідностей між елементами вихідного потоку та полями бази да-

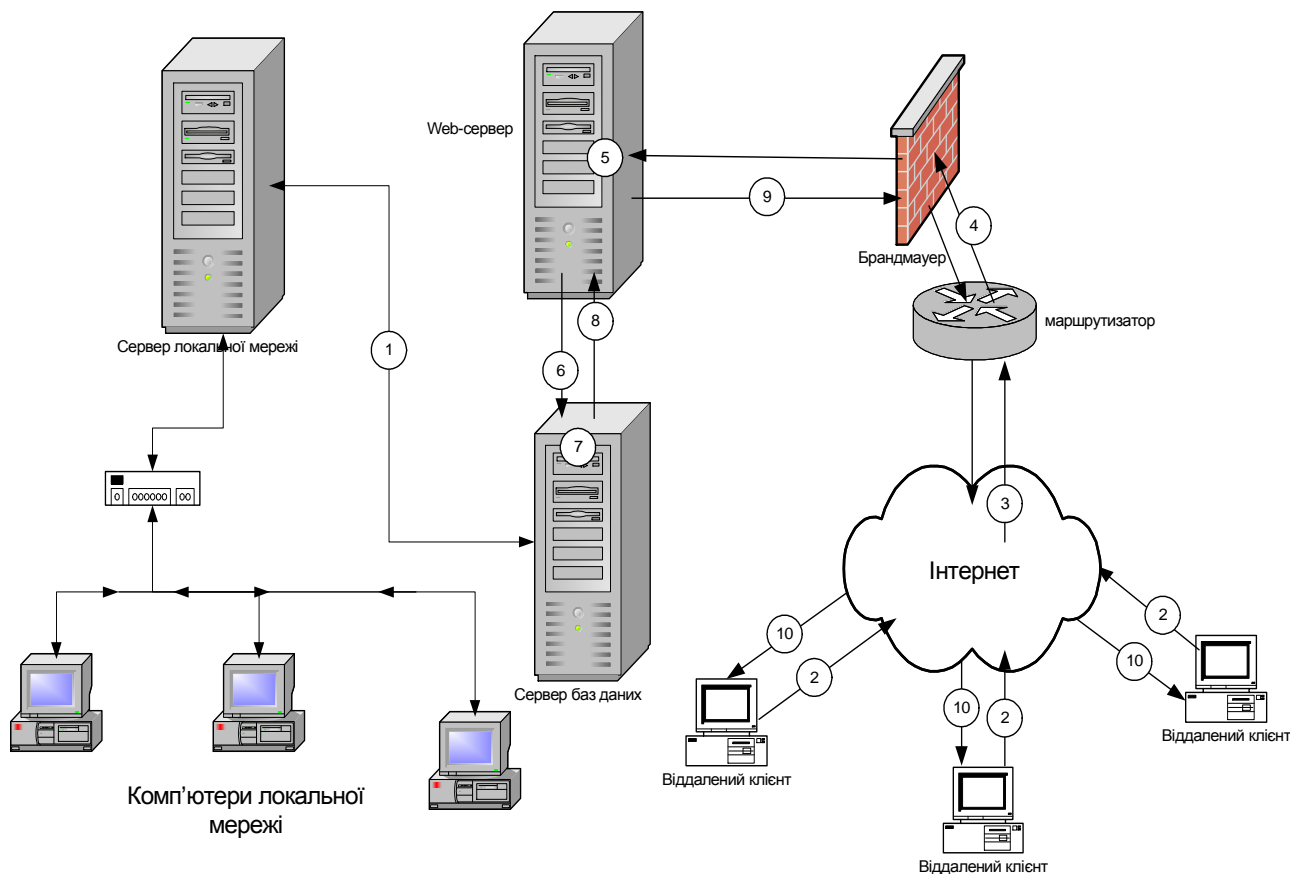


Рис.1. Внутрішньосистемні інформаційні потоки

них, структура якої розглядається нижче.

2. Вихідний запит користувача до сервера (передача URI, GET/POST-даних, Cookies та ін.).

3. Передача запиту через глобальне мережеве середовище.

4. Запит через маршрутизатор та брандмауер потрапляє до Web-сервера та ініціює створення примірника серверного процесу.

5. Apache виконує код PHP сценарію.

6. Сервер дає запит до СКБД на отримання даних.

7. PostgreSQL виконує частину програмного коду, віднесеного на БД, та синтезує необхідні інформаційні елементи.

8. Дані передаються до Web-сервера,

який генерує Web-сторінку.

9. Повертається динамічно створена сторінка.

10. Відбувається обробка результатів запиту браузером.

### Структура бази даних

База даних складається з ряду таблиць, поєднаних мережею логічних зв'язків (рис.2).

“Реєстрова таблиця парадигматичного словника” містить такі поля:

1. Унікальний ідентифікатор лексеми (в межах інтегрованої системи), завдяки якому кожне слово однозначно ідентифікується в межах реєстрового ряду будь-якого словника.

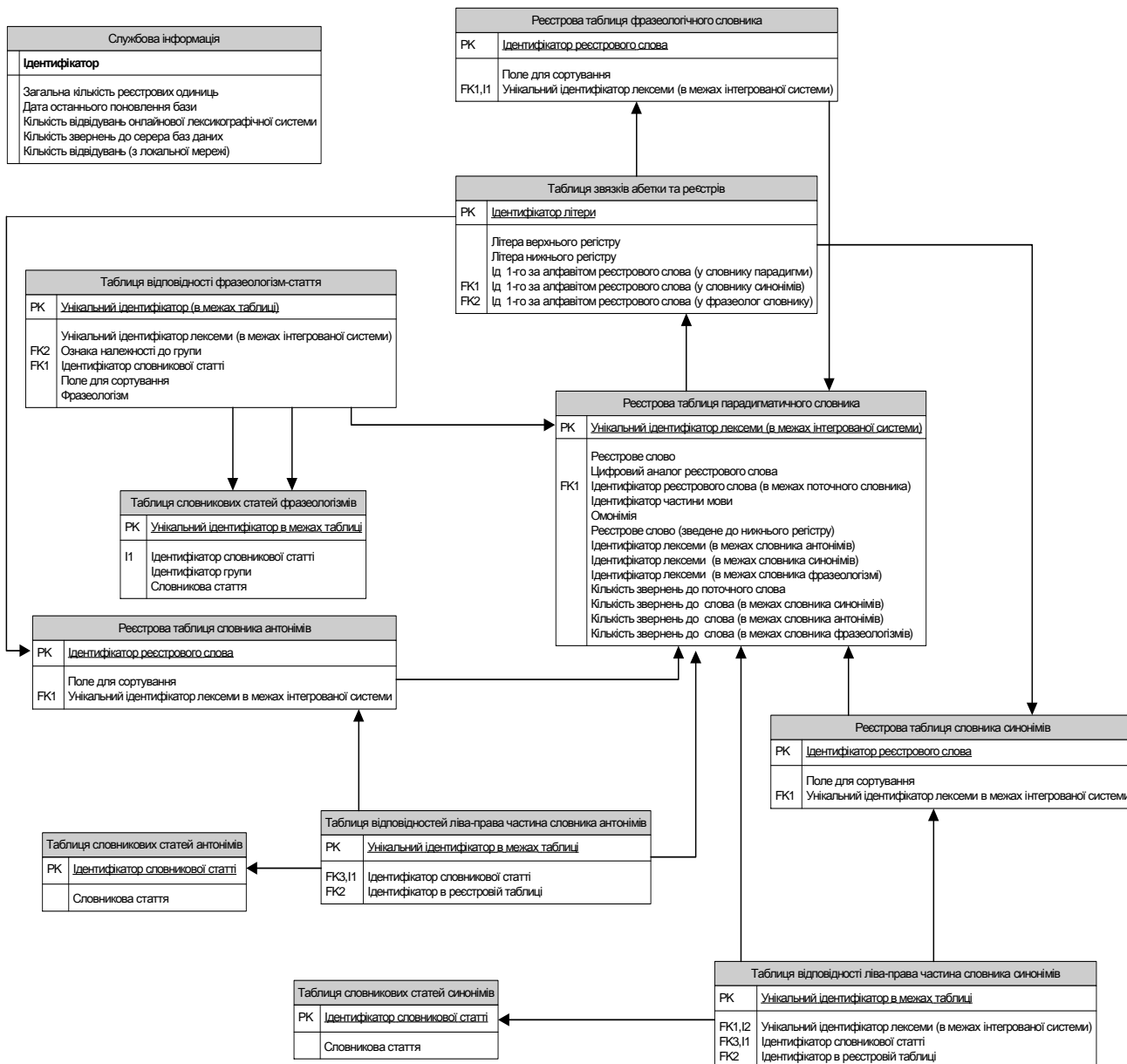


Рис.2. Схема зв'язків між таблицями

2. Реєстрове слово призначене безпосередньо для візуалізації.

3. Цифровий код реєстрового слова для сортування та пошуку генерується за принципом: літери українського алфавіту замінюються двоцифровим кодом (А – ‘01’; Б – ‘02’; В – ‘03’ і т. д.), а інші символи ігноруються.

4. Ідентифікатор реєстрового слова (в межах поточного словника). Масив ідентифікаторів є неперервною послідовністю чисел, яка відповідає відсортованому за алфавітом реєстру. Це дає змогу швидко визначити найближче оточення до заданого слова без послідовного перебору всього реєстрового ряду, що значно оптимізує запит до бази даних.

5. Код частини мови реєстрового слова.

6. Омонімія – номер слова в омонімічній групі.

7. Реєстрове слово зведене до літер нижнього регістру без апострофу та наголосів.

8.–10. Ідентифікатори лексем в межах словника антонімів, синонімів, фразеологізмів відповідно, які застосовуються для

інтеграції словників до єдиної лексикографічної системи.

11.–14. Кількість звернень до поточного слова в словнику антонімів, синонімів, фразеологізмів відповідно.

Приклад даних представлено в табл. 1-2.

„Реєстрова таблиця словника антонімів”, „Реєстрова таблиця словника синонімів”, „Реєстрова таблиця фразеологічного словника” мають подібну структуру. До їх складу входять:

1. Ідентифікатор реєстрового слова.

2. Поле для сортування (неперервний індекс).

3. Унікальний ідентифікатор лексеми в межах інтегрованої системи.

У табл.3 унаочнено представлення даних.

Таблиці відповідностей „ліва-права частини словника синонімів” та „ліва-права частини словника антонімів” (приклад даних представлено в табл. 4) в своїй структурі містять:

1. Унікальний ідентифікатор в межах таблиці.

2. Ідентифікатор словникової статті.

Таблиця 1

1	2	3	4	5	6	7
1189915	краків'янин	18111833031215012115	66014	7	NULL	краків'янин
25827	втішатися	032312290123112233	25732	36	NULL	втішатися
176922	французький	2521011827241031151114	174220	11	NULL	французький
120723	полягти	20191633042311	118896	35	1	полягти

Таблиця 2

8	9	10	11	12	13	14
NULL	NULL	NULL	684297	293850	3456	4375
70713	NULL	6526	707138	38974	43577	437
NULL	NULL	NULL	914756	2345213	45763	765874
5852	NULL	29529	808526	346326	4536	12343

Таблиця 3

1	2	3
161827	1725	159551
161830	1726	159554
161836	1727	159560
161924	1728	159650

Таблиця 4

1	2	3
40812	88023	5044
40425	88028	4988
39805	88029	4907
39852	88041	4912

3. Ідентифікатор в реєстровій таблиці.

Зазначеним таблицям підпорядковані „Таблиця словникових статей синонімів” (табл. 5) та „Таблиця словникових статей антонімів” (табл. 6); за своєю структурою вони становлять сукупність полів:

1. Ідентифікатор словникової статті.
2. Словникова стаття. Форма даних внесених в це поле має свою внутрішню розгалужену структуру, на деталях якої ми не зупиняємося.

„Таблиця відповідностей фразеологізм-стаття” (табл. 7):

1. Унікальний ідентифікатор (в межах таблиці).
2. Унікальний ідентифікатор лексеми (в межах інтегрованої системи).
3. Ознака належності до групи.
4. Ідентифікатор словникової статті.
5. Поле для сортування.
6. Фразеологізм.

Дочірньою до вищенаведеної є таблиця словникових статей фразеологізмів

Таблиця 5

1	2
246	<table border="1" width="100%" height="255"> <tr> <td width="50%" height="37" align="middle"> <b><font color="#000080" size="2" face="...
247	<table border="1" width="100%" height="255"> <tr> <td width="50%" height="37" align="middle"> <b><font color="#000080" size="2" face="...
248	<table border="1" width="100%" height="255"> <tr> <td width="50%" height="37" align="middle"> <b><font color="#000080" size="2" face="...
249	<table border="1" width="100%" height="255"> <tr> <td width="50%" height="37" align="middle"> <b><font color="#000080" size="2" face="...

Таблиця 6

1	2
184071	<P><B><a class="handT" onclick="href='./index.php?key_reestr=184071&dict=sinonim&<?php echo SID; ?>" target="_parent">ШЕЛЄСТІТИ</A></B> (<a class="handT" onclick="href='./index.php?key_reestr=14917 &dict=sinonim&<?php echo SID; ?>" target="_parent">видавати</A>, <a class="handT...
183964	<P><B><a class="handT" onclick="href='./index.php?key_reestr=183964&dict=sinonim&<?php echo SID; ?>" target="_parent">ШВОРА</A></B> (<a class="handT" onclick="href='./index.php?key_reestr=139540 &dict=sinonim&<?php echo SID; ?>" target="_parent">ремінь</A>, <a class="handT" onc...
157918	<P><B><a class="handT" onclick="href='./index.php?key_reestr=157918&dict=sinonim&<?php echo SID; ?>" target="_parent">СПОСТЕРЕЖЛИВИЙ </A></B> (<a class="handT" onclick="href='./index.php?key_reestr=53603 &dict=sinonim&<?php echo SID; ?>" onclick="href='./index.php?key_rees...
12344	<P><B><a class="handT" onclick="href='./index.php?key_reestr=123448&dict=sinonim&<?php echo SID; ?>" target="_parent">ТАРАДАЙКА </A></B> (<a class="handT" onclick="href='./index.php?key_reestr=53603 &dict=sinonim&<?php echo SID; ?>" onclick="href='./index.php?key_rees...

Таблиця 7

1	2	3	4	5	6
56545	188627	23	260	48917	тягти ярмо
56546	188627	33	31	56589	ярмо на шию
56547	188739	21	113	56593	ясна річ

(табл. 8), яка у структурному відношенні являє собою множину таких полів:

1. Унікальний ідентифікатор (у межах таблиці).
  2. Ідентифікатор словникової статті.
  3. Ідентифікатор групи.
  4. Словникова стаття.
- „Таблиця зв’язків абетки та реєстрів”

(табл.9):

1– 2. Літера відповідно верхнього та нижнього регістрів.

3–6. Ідентифікатори першого за алфавітом реєстрового слова відповідно у парадигматичному словнику, словнику синонімів, словнику антонімів та фразеологічному словнику.

„Службова інформація”:

1. Ідентифікатор.
2. Загальна кількість реєстрових одиниць.

3. Дата останнього очищення тимчасового сховища.

4. Кількість відвідувань онлайнової лексикографічної системи.

5. Кількість звернень до сервера баз даних.

6. Кількість відвідувань (з локальної мережі).

### Проблемні аспекти реалізації системи

Роботу сценарію "Словників України on-line" реалізовано поєднанням двох механізмів – це, по-перше, робота з БД та, по-друге, – використання файлової системи як засобу збереження великого обсягу інформації в певній ієрархічній структурі.

Спеціально згенеровані файли правих частин словникових статей (парадигматичного словника) зберігаються поза базою даних, описаною у попередньому підроз-

Таблиця 8

1	2	3	4
7916	28	123	<P><B><a class="handT" onclick="href='../index.php?key_reestr= 379354 &dict=fraseolog &<?php echo SID; ?>" target="_parent">ще</A> <A>й</A> <A>як</A>. </B><a class="handT" onclick="href='../index.php? key_reestr= 52660&dict=fraseolog&<?php echo SID; ?>" target="_parent"> Звичайно...
7917	29	42	<P><B><a class="handT" onclick="href='../index.php?key_reestr=187784 &dict= fraseolog&<?php echo SID; ?>" target= "_parent">язик</A> <A>як </A><a class="handT" onclick="href='../index.php?key_reestr=71863 &dict=fraseolog&<?php echo SID; ?>" target="_parent">лопатень</A>...
7918	30	643	<P><B><a class="handT" onclick="href='../index.php?key_reestr=187784 &dict=fraseolog& <?php echo SID; ?>" target="_parent">язик</A> <a class="handT" onclick= "href='../index.php?key_reestr=149820 &dict= fraseolog&<?php echo SID; ?>" target="_parent">свербить</A>...
7919	31	21	<P><B><a class="handT" onclick="href='../index.php?key_reestr=186727 &dict=fraseolog& <?php echo SID; ?>" target="_parent">що</A> <a class="handT" onclick= "href='../index.php?key_reestr=68065&dict= fraseolog&<?php echo SID; ?>" target= "_parent">куди</A> <A>й</A> ...

Таблиця 9

1	2	3	4	5	6
Б	б	4675	369	55	38
В	в	11457	1949	197	235
Г	г	26538	6777	416	977
Ґ	ґ	32329	8096	492	1196

ділі. Усі файли розділено за початковими літерами реєстрових слів, що їм відповідають. Назви ж файлів, для зручності роботи з ними, збігаються з унікальним кодом реєстрового слова. Зазначимо, що зберігаються вони у стані максимального стиснення з правами доступу “тільки зчитування” для сервера обробки сценарію Apache.

Те, що користувачів системи можуть бути одночасно десятки і сотні тисяч, зумовлює необхідність механізму, який би дозволив зберігати певні дані для кожного клієнта між запусками сценарію. Для Web-застосувань реалізація цього процесу є нетривіальною. Справді, доводиться зберігати всі дані окремо взятого клієнта в певному тимчасовому сховищі, причому після закінчення роботи користувача ці дані повинні знищуватися. Для цього, звичайно, можна було б використати функції серіалізації і файли, але ж потрібно прив'язати конкретного користувача до конкретних тимчасових даних з метою запобігання конфліктів між ними. Вирішенням цієї проблеми стало використання механізму сесій. У момент потрапляння клієнта до системи відбувається старт сесії. Користувачеві присвоюється унікальний ідентифікатор сесії, який дає змогу при кожному перезапуску сценарію однозначно визначити його ініціатора. Знаючи ідентифікатор сесії користувача, сервер застосувань встановлює, де зберігаються дані окремого користувача. В сесії зберігаються такі параметри, як номер обраної сторінки, код вибраного реєстрового слова, пошукове слово тощо. Механізм сесії також застосовується для формування правої частини словника. За переданим на сервер кодом реєстрового слова чи пошуковим словом PHP генерує шлях до файлу, де знаходиться відповідна інтерпретаційна частина. Тимчасово створені файли один раз на день знищуються залежно від терміну їх існування.

Одним із критеріїв ефективності роботи системи є час відповіді та реакції на дії користувача. Звичайно, не можна контролювати такі проблемні аспекти, як порушення в роботі мережі, затримка від значної фізичної віддаленості, ефективність протоколів передачі даних, швидкість роботи системи та локального браузера на клієнтській

стороні. Проте ми намагалися мінімізувати час взаємодії клієнт-сервера за рахунок зменшення обсягів даних, що передаються через мережеве середовище, завдяки оптимізації роботи серверних сценаріїв та часу виконання запитів до бази даних.

Одним із прикладів переваг інтеграції словників до єдиної лексикографічної системи є специфіка переходу між словниками. При виборі одного із слів з будь-якого реєстрового ряду ми вільно змінюємо тип словника, змінюючи при цьому словниковий масив і водночас залишаючись локалізованими на тій же реєстровій одиниці, або ж переходимо на найближчу за написанням лексему. Зазначимо, що для користувача завжди присутня ознака наявності цього слова в тому чи іншому словнику, що позбавляє необхідності додаткового відкриття транзакції зв'язку між клієнтською та серверною частинами. Отже завдяки глобальній систематизації даних отримуємо доступ до всіх наявних ресурсів з єдиного лексикографічного середовища.

Позитивним моментом є також і те, що множина входів до системи не обмежується реєстровим рядом, а охоплює і праві частини словникових статей. Кожне слово правої частини є активним – воно проіндексоване і стає додатковою точкою входу до відповідної словникової одиниці. Це свідчить про густу сітку зв'язків у системі, що відкриває великі функціональні можливості при досить простому й прозорому інтерфейсному відображенні та забезпечує високу ступінь інтерактивності.

### **Принципи організації інтерфейсу користувача**

При реалізації інтерфейсної частини оптимізація системи здійснювалась з дотриманням загальноприйнятих стандартів до створення Web-додатків та основних принципів Web-дизайну [13]. Зрозуміло, що основою будь-якої Web-сторінки є розмітка. В даному випадку – це традиційний HTML, визначений шаблоном DTD стандартної узагальненої мови розмітки SGML. При формуванні Web-сторінок використовувались теги HTML, які не виходять за межі стандартів, затверджених консорціумом W3C (World Wide Web Consortium). Техно-



логія HTML використовується перш за все для структурування документа, а не для його представлення. Вигляд сторінок керується елементами дизайну, що досягається використанням каскадних таблиць стилів (Cascading Style Sheets, CSS). Правила таблиць стилів застосовуються для визначення великої множини візуальних аспектів відображення об'єктів сторінки (колір, розмір, положення тощо). Така технологія відкриває широкі можливості зміни візуалізації елементів інтерфейсу та спрощує упорядкування й супровід документів. У системі використовуються глобальні стильові специфікації та передбачено альтернативні варіанти для CSS-несумісних браузерів.

Сторона клієнта в основному визначається браузером, який є інтерпретатором Web-сторінок, отриманих з Інтернет-середовища за допомогою мережевих протоколів. Тому важливим завданням є контроль сумісності з браузерами різних виробників та різних версій. Програмна система пройшла тестування з використанням таких найбільш розповсюджених браузерів, як Microsoft Internet Explorer (починаючи з версії 4.0), Netscape Navigator (6.x, 7.x), Mozilla (1.x) та Opera (версії 5.x та вище).

Ще однією вимогою до реалізації системи є збереження балансу форми представлення і функціонального навантаження. Важливо, щоб система надавала деякий спеціальний інтерфейс з браузером, це дозволило б отримувати всі необхідні дані і представляти їх користувачеві в зрозумілій і зручній формі. Система „Словники України он-лайн” забезпечує поєднання розвиненої множини функцій з досить зручним і несуперечливим інтерфейсом.

Зупинімося детальніше на елементах інтерфейсу (рис.3).

1. У лівому верхньому куті розміщено елемент вводу, який призначено для пошуку слів у реєстрі словника. Користувач має змогу швидко переміститися на слово, яке його цікавить, або ж, у разі його відсутності у множині реєстрового ряду, на групу найближчих за написанням лексем.
2. Додаткова навігація по реєстру здійснюється швидким переходом за абетковою літерою. Реалізовано ефект закла-

док, за допомогою яких відбувається позиціонування на перше слово, що розпочинається з обраної літери.

3. Ще один елемент навігації. Весь реєстровий ряд розбитий на так звані віртуальні групи (кількість одиниць в групі задається програмно). Це дозволяє легко продивитися найближче оточення, що чимось нагадує звичне перелистування паперового словника.
4. Активна частина реєстрового ряду. Обране слово помічено з використанням специфічного стильового відображення.
5. Ознака наявності слова в тому чи іншому словнику: а,с,ф – відповідно в словнику антонімів, синонімів та фразеологізмів.
6. Зміна типу словника. Активний словник помічено більш інтенсивним кольоровим забарвленням.
7. Область виводу словникової статті. Вона реалізована з використанням плаваючого фрейму. Такий підхід дав змогу позбутися проблеми позиціонування при різних величинах правої частини словникових статей. Ще більш виправданий вибір фреймової репрезентації у фразеологічному словнику, де ця область розбивається на дві частини (в першій відображається множина фразеологізмів, що відповідають обраному реєстровому слову, а в другій – інтерпретаційна частина відповідної фразеологічної групи). Саме таке візуальне представлення дозволяє відчувати ієрархію зв'язків даного словника.

Зазначимо, що описана вище система на сьогодні реалізована та функціонує в мережевому середовищі за адресою: [http://ulif.org.ua/ulp/dict\\_all/](http://ulif.org.ua/ulp/dict_all/) у складі Українського лінгвістичного порталу (<http://ulif.org.ua/>). Розвиток даного програмного комплексу передбачається за такими напрямками:

- розширення та удосконалення лінгвістичного наповнення лексикографічної бази;
- інтеграція з іншими Л-системами, зокрема з тлумачним, етимологічним, термінологічними та іншими словниками української мови;
- інтеграція з іншомовними словни-

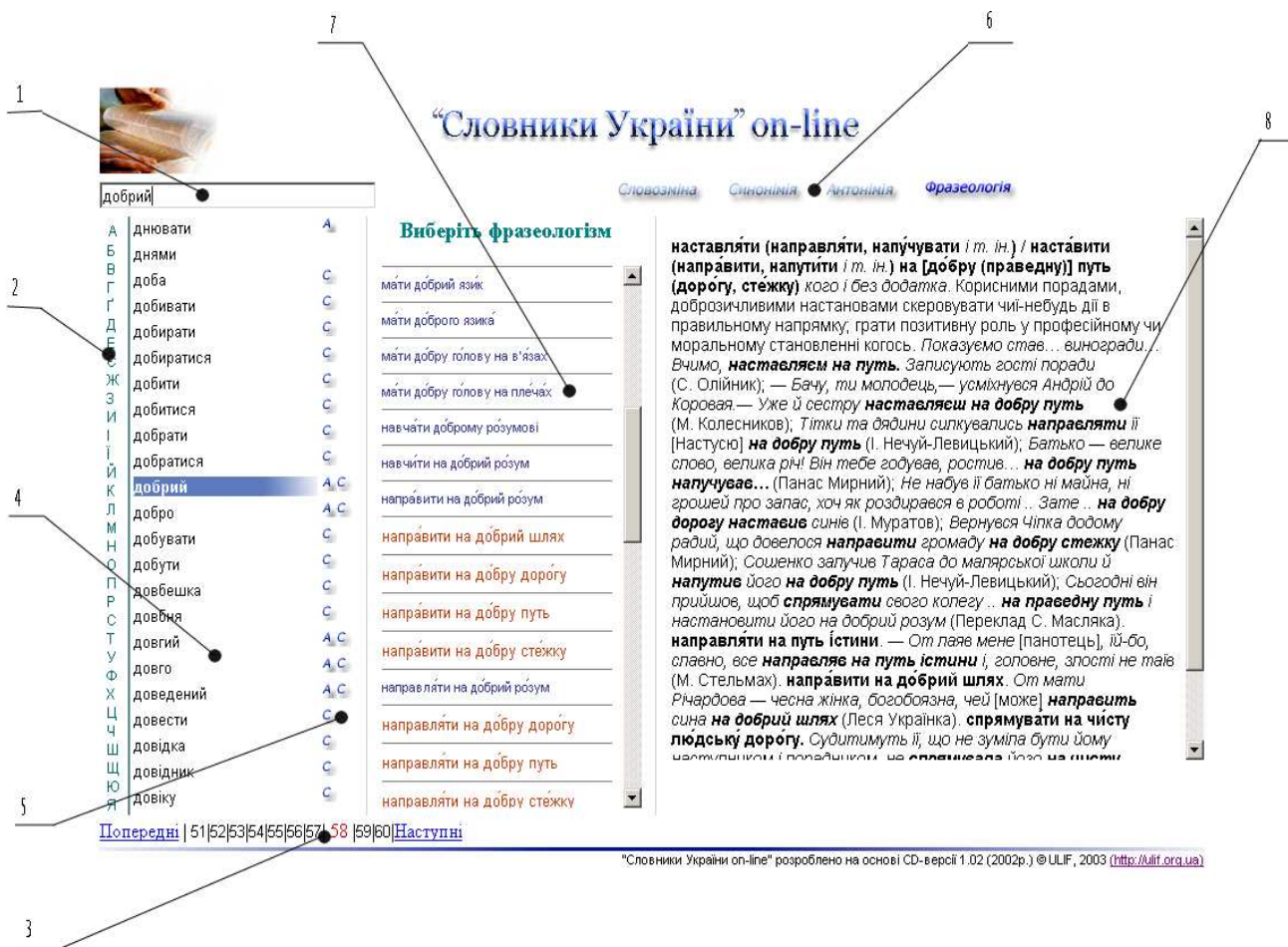


Рис.3. Загальний інтерфейс користувача

ками та створенням багатомовних онлайн-нових лексикографічних середовищ;

- реалізація додаткових функціональних можливостей (пошук в реєстрі з використанням процедури лематизації, пошук контекстів в словникових статтях, зв'язок ілюстративного матеріалу з їх вихідними бібліографічними описами та повними електронними текстами цих джерел);

- застосування у структурі онлайн-ового варіанту Українського лінгвістичного корпусу;

- інтеграція до онлайн-ових систем пошуку та опрацювання текстів;

- удосконалення інтерактивного спілкування з користувачами;

- можливість використання варіативних інтерфейсних схем та ін.

Особливо відзначимо можливість проектування з використанням наших напрацювань віртуальної лексикографічної лабораторії, яка повинна стати основою для створення системи Всеукраїнського лінгвістичного діалогу [14].

Автор висловлює подяку співробітникам Українського мовно-інформаційного фонду НАН України за допомогу та підтримку при створенні системи „Словники України он-лайн” та плідні обговорення при написанні цієї статті.

1. Широков В.А. Інформаційна теорія лексикографічних систем. – К.: Довіра, 1998. – 331 с.
2. Широков В.А. Феноменологія лексикографічних систем. – К.: Наук. думка, 2004. – 331 с.
3. Рабулець О.Г. Інтегровані лексикографічні системи: Автореф. дис... канд. техн. наук / НАН України; Український мовно-інформаційний фонд. – К., 2002. – 18 с.
4. Інтегрована лексикографічна система "Словники України". Електронне видання на лазерному диску. ISBN 966-507-149-1 / В.А. Широков, О.Г. Рабулець, І.В. Шевченко, О.М. Костишин, К.М. Якименко. – К., 2003.
5. Бурячок А.А. Орфографічний словник української мови: 4-те вид., доопрац. / – К.: Наук. думка, 2002. – 464 с.
6. Шевченко І.В. Алгоритмічна словозмінна класифікація української лексики // Мовознавство. – 1996. – №4-5. – С. 40-44.

7. Шевченко І.В. Моделі та алгоритмічно-програмне забезпечення лексикографічних систем: Дис... канд. техн. наук / НАН України; Український мовно-інформаційний фонд. – К.: 2000. – 167 с.
8. Словник синонімів української мови: У 2 т. / Под ред. А.А. Бурячка; НАН України; Інститут мовознавства ім. О.О. Потебні. – К.: Наук. думка, 1999. – Т. 1 : А - Н. – 1040 с.; Т. 2 : О - Я. – 960 с.
9. Полюга Л.М. Словник антонімів української мови: 2-ге вид., доп. і випр. / Под ред. Л.С. Паламарчук; НАН України; Інститут українознавства ім. І. Крип'якевича; Український мовно-інформаційний фонд. – К.: Довіра, 2001. – 276 с.
10. Словник фразеологізмів української мови: – К.: Наук. думка, 2003. – 1104 с.
11. Буч Г. Объектно-ориентированный анализ и проектирование с примерами приложений на С++. – 2-е изд. / Пер. с англ. – М.: "Изд-во Бинном"; СПб.: "Невский диалект", 2001. – 560 с.
12. Гешвинде Э., Шениг Г.-Ю. Разработка Web-приложений на PHP и PostgreSQL: Руководство разработчика и администратора / Пер. с англ. – СПб: ООО «ДиаСофтЮП», 2003. – 608 с.
13. Пауэлл Т. Web-дизайн. – 2-е изд., перераб. и доп. / Пер. с англ. – СПб.: БХВ–Петербург, 2004. – 1072 с.
14. Широков В.А. Всеукраїнський лінгвістичний діалог у контексті теорії лексикографічних систем // Мовознавство. – 2003. – №6. – С. 3-7.
15. Коннолли Т., Бегг К., Страчан А. Базы данных: проектирование, реализация и сопровождение. Теория и практика. – 2-е изд. / Пер. с англ. – М.: Изд. дом "Вильямс", 2000. – 1120 с.

Отримано 19.04.05

**Про автора**

*Сидорчук Надія Миколаївна*  
аспірантка

*Місце роботи автора:*

Український мовно-інформаційний фонд  
НАН України  
01601, Київ, вул. Володимирська, 54  
Тел.: (044) 259 4895  
E-mail: nadine@ulif.org.ua