

Магістерська кваліфікаційна робота на тему:

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ІНТЕЛЕКТУАЛЬНОГО ЛІНГВІСТИЧНОГО ПОШУКУ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ ХЕМІНГА

Виконав студент гр. 1КН-15м

Семенюк П.М.

Науковий керівник: к.т.н., доц., Колесницький О.К.

АКТУАЛЬНІСТЬ ДОСЛІДЖЕННЯ

застосування для розв'язання задачі лінгвістичного пошуку сучасних методів штучного інтелекту забезпечує подальше впровадження інформаційних технологій у галузь пошуку текстової інформації, а також є актуальною темою дослідження нових методів використання штучного інтелекту

Технологія інтелектуального пошуку дозволяє:

- розширювати запит близькими за написанням словами,
- знаходити всі лексикографічно близькі слова, що відрізняються замінами, пропусками і вставками символів,
- вести пошук слів з друкарськими помилками, а також у випадках, коли виникають сумніви в правильному написанні - прізвища, назви організації і т.п.
- застосовуватись в системах перевірки орфографії

МЕТА ТА ЗАВДАННЯ ДОСЛІДЖЕННЯ

Метою дослідження магістерської кваліфікаційної роботи є підвищення швидкодії інтелектуального пошуку набраного слова у словнику (файл з правильними словами) з метою виправлення помилок за рахунок використання нейронної мережі Хемінга.

Для досягнення поставленої мети необхідно розв'язати такі завдання:

- провести аналіз предметної області інтелектуального лінгвістичного пошуку;
- розглянути існуючі методи вирішення задачі інтелектуального лінгвістичного пошуку та обґрунтувати вибір методу, який задовольняє мету даної магістерської кваліфікаційної роботи;
- сформулювати стадії інформаційної технології;
- розробити структуру використовуваної нейронної мережі;
- розробити алгоритм функціонування програмної реалізації інтелектуального лінгвістичного пошуку на основі нейронної мережі Хемінга;
- виконати програмну реалізацію інформаційної технології інтелектуального лінгвістичного пошуку на основі нейронної мережі Хемінга;
- провести тестування програмної реалізації інтелектуального лінгвістичного пошуку на основі нейронної мережі Хемінга та проаналізувати отримані результати;

ОБ'ЄКТ, ПРЕДМЕТ ТА МЕТОДИ ДОСЛІДЖЕННЯ

Об'єкт дослідження – це процес автоматизованого інтелектуального пошуку набраного слова у файлі.

Предмет дослідження – це методи та програмні засоби інтелектуального пошуку набраного слова у файлі на основі нейронної мережі Хемінга та їх швидкодія.

Методи дослідження

У роботі використані наступні методи наукових досліджень:

- системного аналізу для аналізу структури інформаційної системи,
- теорії нейронних мереж для реалізації інформаційної технології інтелектуального лінгвістичного пошуку,
- методи математичної статистики для розробки процесу пошуку та обрахунків результатів експериментів із програмним засобом,
- об'єктно-орієнтованого програмування для програмної реалізації .

НАУКОВА НОВИЗНА ОДЕРЖАНИХ РЕЗУЛЬТАТІВ

полягає в наступному:

знайшла подальшого розвитку інформаційна технологія інтелектуального лінгвістичного пошуку за рахунок використання штучної нейронної мережі Хемінга для визначення подібних слів, що дозволило підвищити швидкодію виконання пошукових запитів.

ПРАКТИЧНЕ ЗНАЧЕННЯ ОДЕРЖАНИХ РЕЗУЛЬТАТІВ

полягає у наступному:

- Розроблено алгоритм інтелектуального лінгвістичного пошуку на основі нейронної мережі Хемінга.
- Розроблено алгоритм функціонування нейронної мережі Хемінга для підвищення швидкодії виконання пошукових запитів.
- Розроблено програмний засіб для інтелектуального лінгвістичного пошуку на основі нейронної мережі Хемінга.

Розроблені алгоритми можуть бути впроваджені в початковий процес як лекція на тему «Нейромережевий метод інтелектуального лінгвістичного пошуку з використанням нейронної мережі Хемінга» дисципліни «Нейромережеві методи обчислювального інтелекту». Також розроблений в роботі програмний продукт використаний в програмному забезпеченні, що розробляється компанією ПрАт «Фарлеп-Інвест», про що свідчить **довідка про впровадження**.

АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

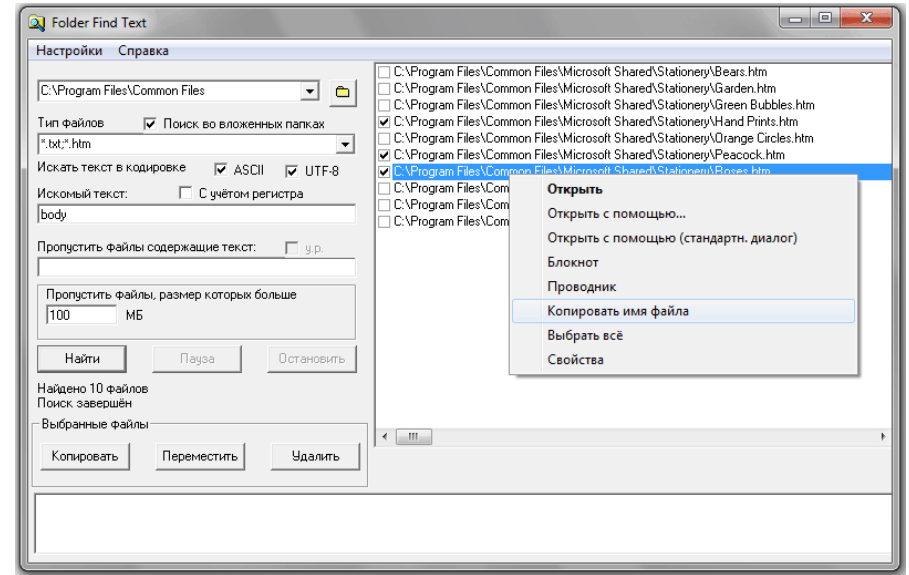
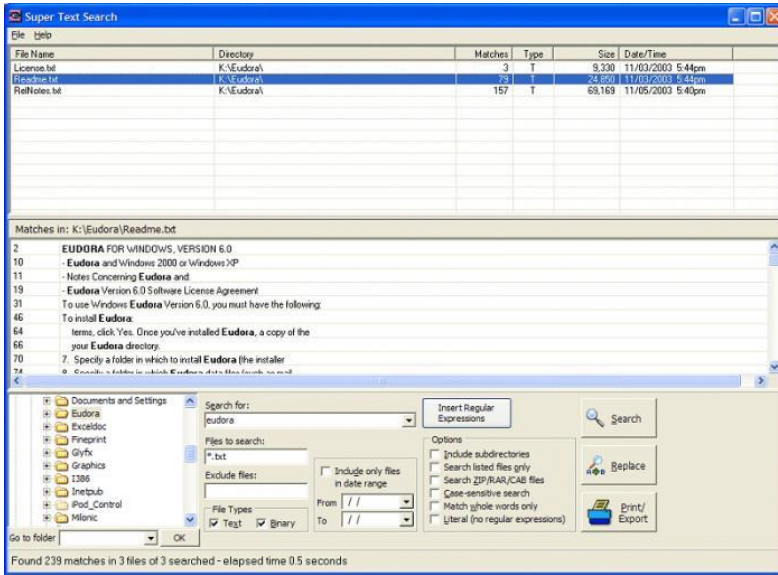
Відомі методи для реалізації пошуку на неточну відповідність:

- 1) перебір або сканування словника;
- 2) розширення вибірки або метод спел-чекера;
- 3) методи n-грам;
- 4) хешування;
- 5) триангуляційні дерева;
- 6) trie-дерева (промені).

Недолік відомих методів лінгвістичного пошуку - низька швидкодія

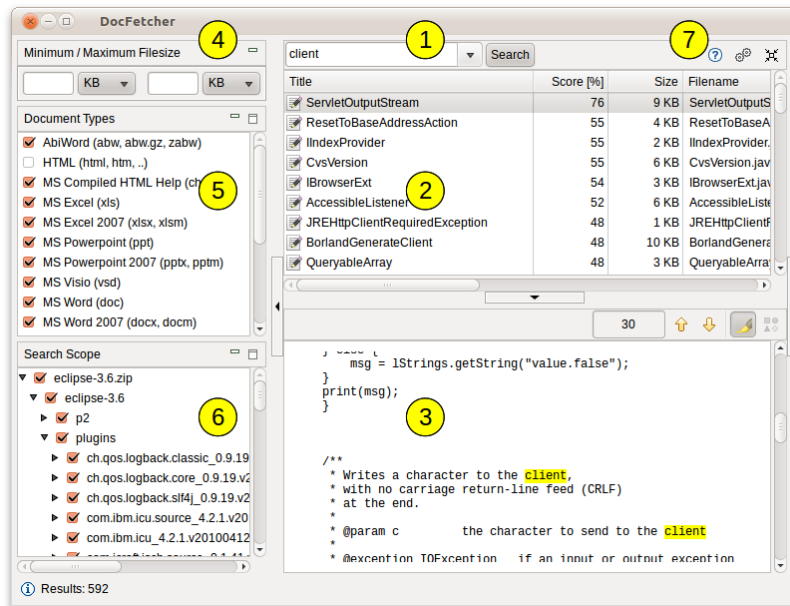
Для уникнення недоліків відомих методів у даній роботі пропонується більш перспективна технологія, а саме – технологія інтелектуального пошуку за подібністю на базі штучних нейронних мереж.

АНАЛІЗ ІСНУЮЧИХ ПРОГРАМНИХ ЗАСОБІВ ЛІНГВІСТИЧНОГО ПОШУКУ



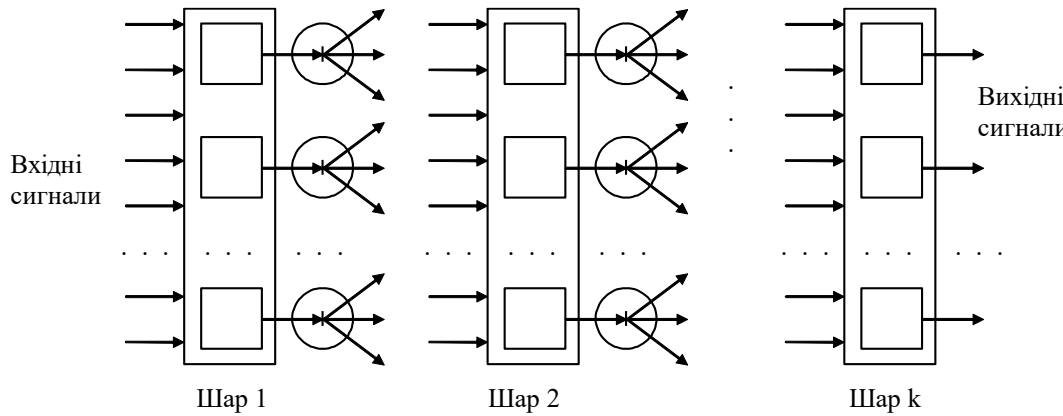
**Програма
Super Text
Search**

Програма
DocFetcher

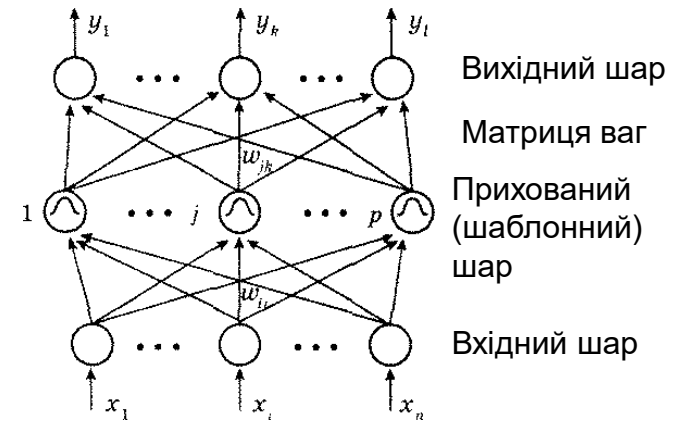


Програма
Folder Find Text

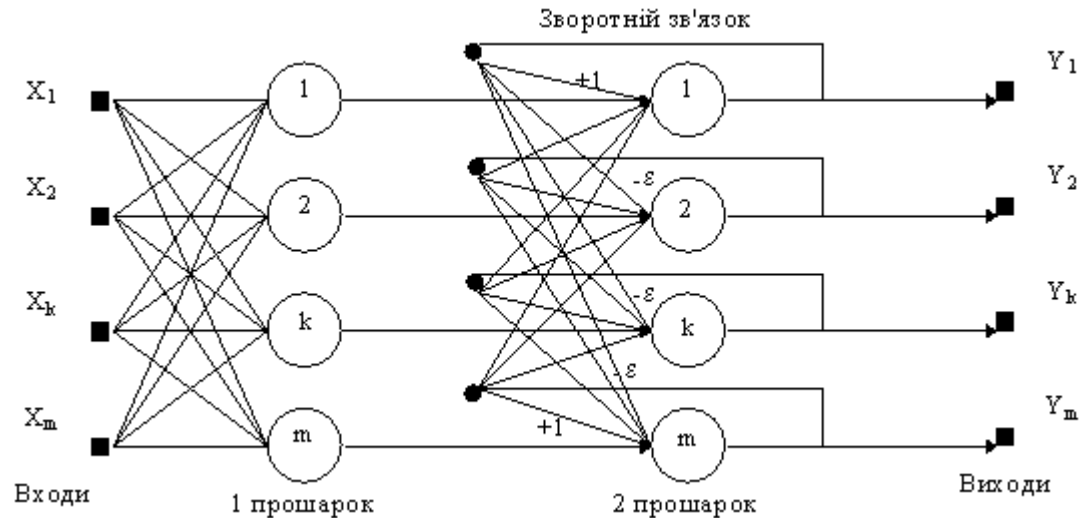
ОБГРУНТУВАННЯ ВИБОРУ НЕЙРОННОЇ МЕРЕЖІ ХЕМІНГА



Багатошаровий перцептрон



Радіально-базисна мережа



Мережа Хемінга

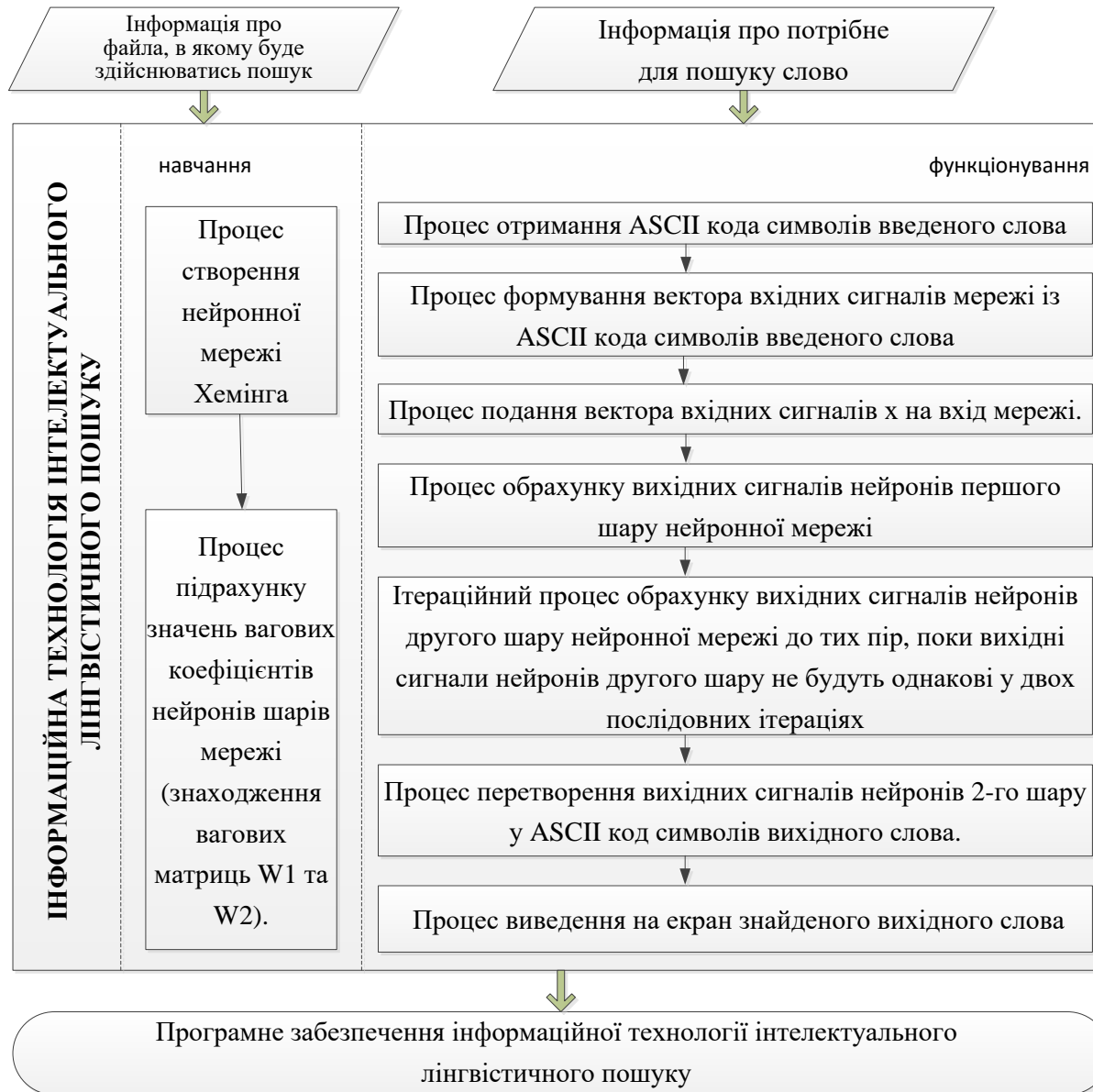
МАТЕМАТИЧНА МОДЕЛЬ НЕЙРОННОЇ МЕРЕЖІ ХЕМІНГА

<p>1. На стадії ініціалізації ваговим коефіцієнтам першого шару і порогу передатної функції присвоюються такі значення:</p> <p>Вагові коефіцієнти гальмуючих синапсів у другому шарі беруть рівними деякій величині $0 < v < 1/m$. Синапс нейрона, пов'язаний з його ж виходом має вагу +1.</p>	$W_{ik} = x_i^k / 2, i = 0 \dots n-1, k = 0 \dots m-1$ $b_k = n / 2, k = 0 \dots m-1$ <p>Тут x_i^k - i-ий елемент k-ого зразка.</p>
<p>2. На входи мережі подається невідомий вектор $x_1, x_2, x_n \dots$ Розраховуються стани нейронів першого шару (верхній індекс у дужках вказує номер шару):</p>	$y_j^{(1)} = S_j^{(1)} = \sum_{i=1}^n w_{ij} x_i + b_j, j = 0 \dots m-1$
<p>Після цього отримання значення ініціалізують значення виходів другого шару:</p>	$y_j^{(2)} = y_j^{(1)}, j = 0 \dots m-1$
<p>3. Обчислюються нові стани нейронів другого шару:</p>	$S_j^{(2)}(t+1) = y_j(t) - \varepsilon \sum_{k=1}^m y_k^{(2)}(t), k \neq j, j = 1 \dots m$
<p>і значення їх виходів:</p> <p>Передавальна функція f має вигляд порога, причому величина b повинна бути досить великою, щоб будь-які можливі значення аргументу не призводили до насичення.</p>	$y_j^{(1)}(t+1) = f[S_j^{(2)}(t+1)]$
<p>4. Перевіряється, чи змінилися виходи нейронів другого шару за останню ітерацію. Якщо так - перейти до кроку 3. Інакше - кінець.</p>	

АЛГОРИТМ РОБОТИ ПРОГРАМИ ІНТЕЛЕКТУАЛЬНОГО ЛІНГВІСТИЧНОГО ПОШУКУ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ ХЕМІНГА



СТРУКТУРА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ІНТЕЛЕКТУАЛЬНОГО ЛІНГВІСТИЧНОГО ПОШУКУ

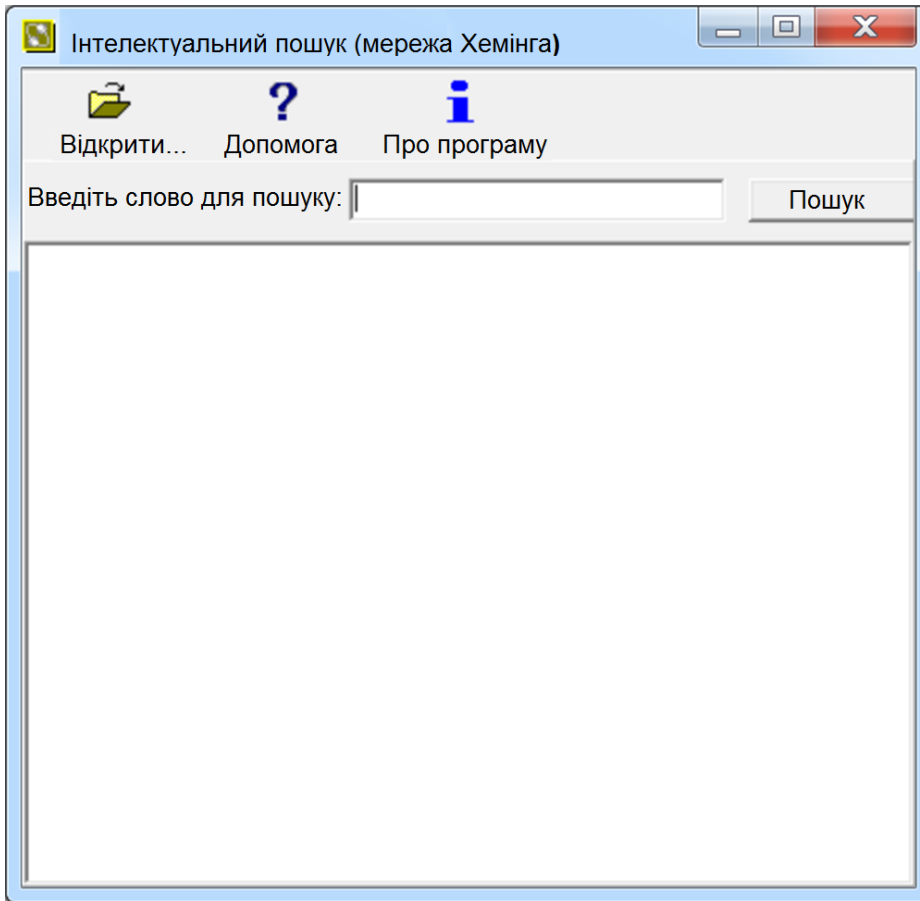


ПРОГРАМНА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОГО ЛІНГВІСТИЧНОГО ПОШУКУ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ ХЕМІНГА

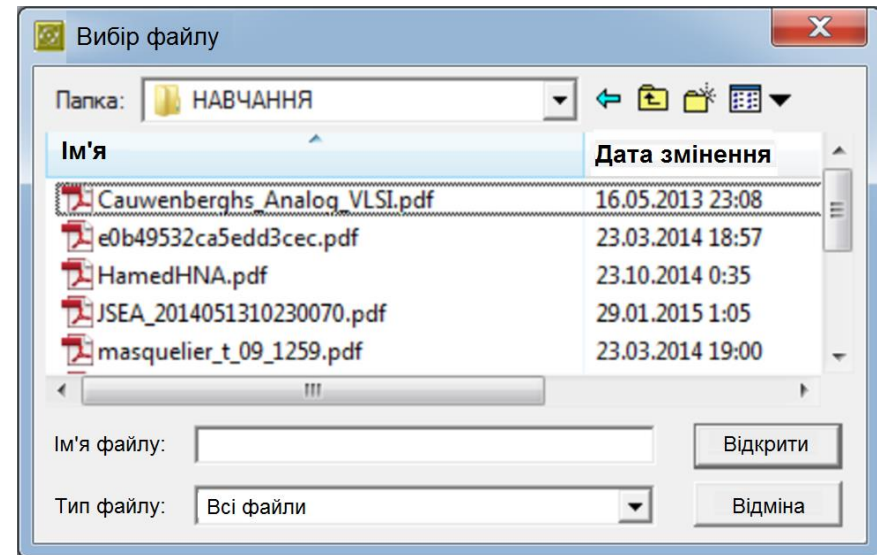
- мова програмування C++
- Середовище розробки - Microsoft Visual Studio 2012,

надане кафедрою КН у рамках
програми Dream Spark

РОБОЧІ ВІКНА ПРОГРАМИ ЛІНГВІСТИЧНОГО ПОШУКУ

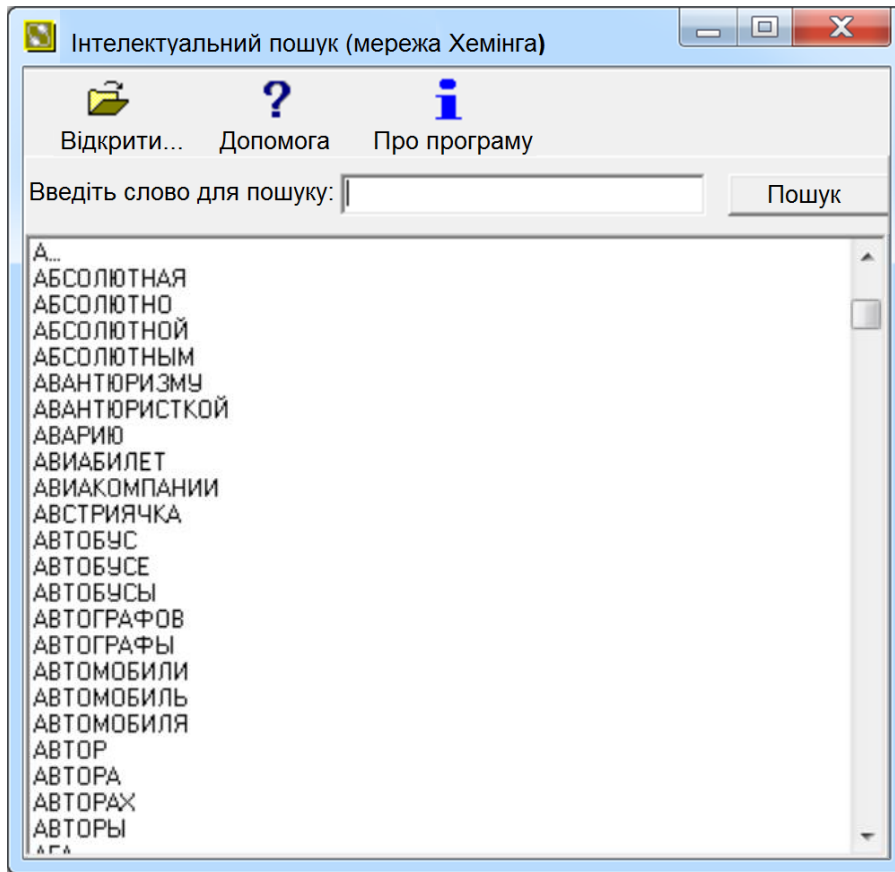


Стартове вікно програми

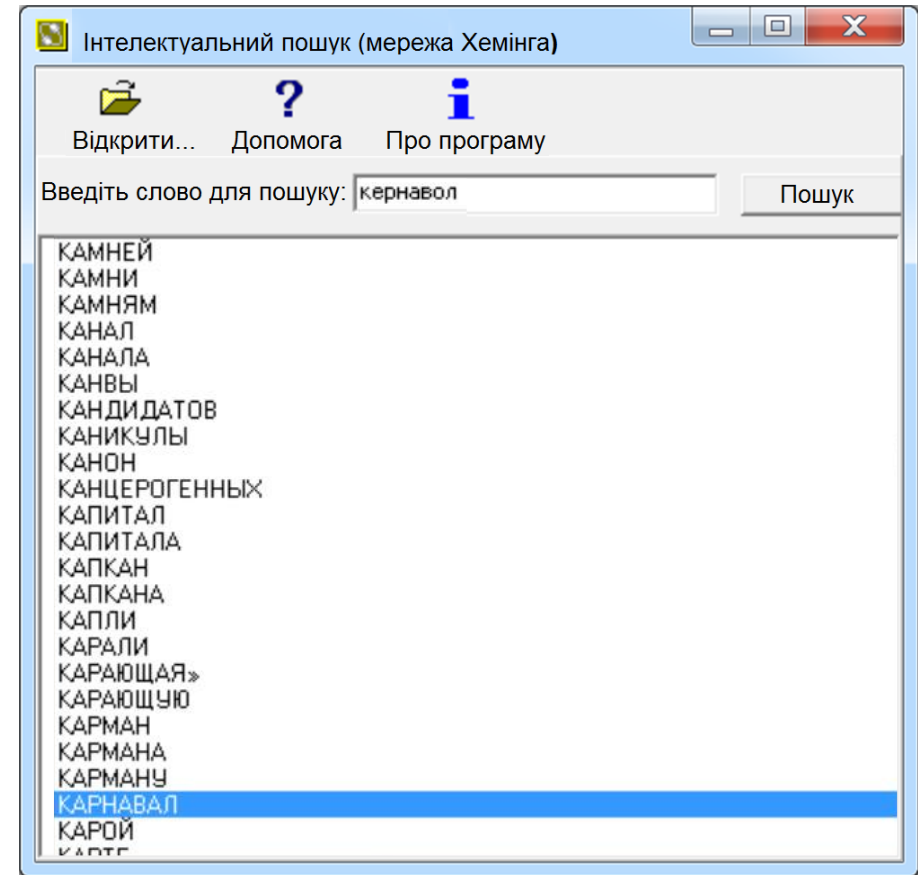


Вікно вибору файлу для пошуку

Результати роботи програми інтелектуального лінгвістичного пошуку на основі нейронної мережі Хемінга



Вікно з відображенням словнику у робочому полі



Вікно з відображенням результату пошуку

АНАЛІЗ ШВИДКОДІЇ РОЗРОБЛЕНОЇ ПРОГРАМИ

Таблиця 3.1 – Показники швидкодії розробленої програми та аналога

	Витрати часу (сек)			
	На створення словника	На пошук слова		
		min	max	середн
Розроблена програма	14,1	30,5	32,1	31,3
Аналог (Super Text Search)	14,5	27,8	55,6	41,7

Із табл. видно, що розроблена програма в середньому витрачає на 33,2%
((41,7-31,3/31,3)x100%=33,2%))
менше часу на пошук слова у словнику, ніж програма-аналог.

ЕКОНОМІЧНА ЧАСТИНА

Було виконано оцінювання комерційного потенціалу розробки інформаційної технології інтелектуального лінгвістичного пошуку на основі нейронної мережі Хемінга.

- Загальні витрати на виконання науково-дослідної, дослідно-конструкторської та конструкторсько-технологічної роботи складають 32224,57 грн.
- Приведена вартість всіх чистих прибутків, що їх отримає підприємство від реалізації результатів наукової розробки 295550,55 грн.
- Щорічна ефективність вкладених в наукову розробку інвестицій - 109%, що вище за мінімальну бар'єрну ставку дисконтування, яка складає 32%.
- Термін окупності вкладених інвестицій - 0,92 року.

АПРОБАЦІЯ РЕЗУЛЬТАТІВ РОБОТИ ТА ПУБЛІКАЦІЇ

Апробація результатів роботи.

Результати роботи були апробовані на
-- науково-практичній конференції «Сучасні тенденції розвитку системного програмування» (25-26 листопада 2016 р.), м. Київ [2],
-- XLV регіональній науково-технічній конференції професорсько-викладацького складу, співробітників та студентів ВНТУ з участю працівників науково-дослідних організацій та інженерно-технічних працівників підприємств м. Вінниці та області (2016) [1]

Публікації. За результатами магістерської кваліфікаційної роботи опубліковано 2 тез доповідей конференції [1, 2].

Висновки

В роботі створено інформаційну технологію інтелектуального лінгвістичного пошуку на основі нейронної мережі Хемінга та її програмну реалізацію, що здатна шукати близькі по написанню слова. Поставлена мета роботи досягнута, оскільки середня швидкодія інтелектуального пошуку набраного слова у словнику підвищена на 33,2%. Крім цього, у розробленій програмі час пошуку слова мало відрізняється для різних слів, а у програмі-аналозі має широкий діапазон розкиду.

Дякую за увагу.