

TECHNICAL SCIENCES AND TECHNOLOGIES

5. *Олифер В. Г.* Компьютерные сети. Принципы, технологии, протоколы : учебник / В. Г. Олифер, Н. А. Олифер. – СПб. : Питер, 2010. – 943 с.

6. *Оценка эффективности информационных систем: Часть 3. Пример расчета эффективности использования ИС методом ТЕІ [Электронный ресурс].* – Режим доступа : http://www.ibm.com/developerworks/ru/library/l-otcenka_efektivnosti_3/index.html.

7. *Критеріальне оцінювання ефективності інформаційних пристроїв та систем / М. А. Філінюк, В. О. Багацький, Л. Б. Ліщинська, О. В. Войцеховська.* – Вінниця : ВНТУ, 2014. – 143 с.

8. *Маслов В. П.* Інформаційні системи і технології в економіці : навчальний посібник / В. П. Маслов. – К. : Слово, 2006. – 264 с.

Сахно Євген Юрійович – доктор технічних наук, професор, завідувач кафедри управління якістю та проектами, Чернігівський національний технологічний університет (вул. Белова, 4, м. Чернігів, 14000, Україна).

Сахно Евгений Юрьевич – доктор технических наук, профессор, заведующий кафедрой управления качеством и проектами, Черниговский национальный технологический университет (ул. Белова, 4, г. Чернигов, 14000, Украина).

Sakhno Yevhen – Doctor of Technical Sciences, Professor, head of Department of Quality Management and Projects, Chernihiv National University of Technology (4 Bielova Str., 14000 Chernihiv, Ukraine).

E-mail: kafurap@ukr.net

Маргасов Дмитро Валерійович – асистент кафедри управління якістю та проектами, Чернігівський національний технологічний університет (вул. Белова, 4, м. Чернігів, 14000, Україна).

Маргасов Дмитрий Валерьевич – ассистент кафедры управление качеством и проектами, Черниговский национальный технологический университет (ул. Белова, 4, г. Чернигов, 14000, Украина).

Marhasov Dmytro – assistant of the Department of Quality Management and Projects, Chernihiv National University of Technology (4 Bielova Str., 14000 Chernihiv, Ukraine).

E-mail: mn123@i.ua

Корнієць Катерина Євгенівна – студент факультету проектного менеджменту та управління якістю, Чернігівський національний технологічний університет (вул. Белова, 4, м. Чернігів, 14000, Україна).

Корниец Екатерина Евгеньевна – студент факультета проектного менеджмента и управления качеством, Черниговский национальный технологический университет (ул. Белова, 4, г. Чернигов, 14000, Украина).

Korniiets Kateryna – student of the faculty of Project Management and Quality Management, Chernihiv National University of Technology (4 Bielova Str., 14000 Chernihiv, Ukraine).

E-mail: kate.sahno@yandex.ua

УДК 510.635:004.891(045)

Анастасія Вавіленкова

СТРУКТУРА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПОРІВНЯЛЬНОГО АНАЛІЗУ ТЕКСТОВИХ ДОКУМЕНТІВ

Анастасія Вавіленкова

СТРУКТУРА ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ СРАВНИТЕЛЬНОГО АНАЛИЗА ТЕКСТОВЫХ ДОКУМЕНТОВ

Anastasiia Vavilenkova

THE STRUCTURE OF INFORMATION TECHNOLOGY OF TEXT DOCUMENTS COMPARATIVE ANALYSIS

Здійснено аналіз наявних програмних продуктів пошуку плагіату серед електронних текстових документів, що дозволило виділити основні недоліки функціонування систем інтелектуального аналізу текстової інформації. На усунення цих недоліків спрямовані матеріали дослідження, де запропоновано структуру інформаційної технології порівняльного аналізу текстових документів. Її основними блоками є: інтерфейс користувача, компонента побудови логіко-лінгвістичної моделі природномовного текстового документа, лінгвістичний процесор для побудови лінгвістичних складових текстових документів, лінгвістичний процесор системи автоматизованого формування логіко-лінгвістичних моделей речень природної мови, компонента порівняння лінгвістичних складових, компонента порівняння семантико-синтаксичних складових. У статті аргументовано роль кожної із компонент у вилученні змістовної інформації з текстів, що порівнюються.

Ключові слова: природна мова, логіко-лінгвістична модель, змістовний аналіз, лінгвістичний процесор, база знань.

Рис.: 4. Бібл.: 14.

Осуществлен анализ существующих программных продуктов поиска плагиата среди электронных текстовых документов, что позволило выделить основные недостатки функционирования систем интеллектуального анализа текстовой информации. На устранение этих недостатков направлены материалы исследований, где предложено

структуру информационной технологии сравнительного анализа текстовых документов. Ее основные блоки: интерфейс пользователя, компонента построения логико-лингвистической модели документа, написанного на естественном языке, лингвистический процессор для построения лингвистических составляющих текстовых документов, лингвистический процессор системы автоматизированного формирования логико-лингвистических моделей предложений естественного языка, компонента сравнения лингвистических составляющих, компонента сравнения семантико-синтаксических составляющих. В статье аргументирована роль каждой из компонент в извлечении содержания из текстов, которые сравниваются.

Ключевые слова: естественный язык, логико-лингвистическая модель, содержательный анализ, лингвистический процессор, база знаний.

Рис.: 4. Библ.: 14.

The article analyzes the existing software of plagiarism search in electronic text documents, allowing to identify the main shortcomings of intellectual analysis of textual information. The research is aimed on reducing these shortcomings using a proposed structure of information technology of text documents comparative analysis. Its major units include a user interface, module of constructing a logic and linguistic model of natural language text documents, linguistic processor for constructing linguistic components of text documents, linguistic processor of the system of automated formation of logic and linguistic models of natural language sentences, module of linguistic components comparison, module of semantic-syntactic components comparison. The paper argued the role of each unit in extracting context information from the compared texts.

Key words: natural language, logic and linguistic model, context analysis, linguistic processor, knowledge base.

Fig.: 4. Bibl.: 14.

Постановка проблеми. На сьогодні мережа Інтернет переповнена програмними продуктами різноманітної якості, спрямованими на пошук плагіату електронних текстових документів. Парадоксом є те, що пропорційно кількості програм виявлення дублікатів зростає кількість програм по рейтингу текстових документів. Основною проблемою виникнення такого явища є брак якісної теоретичної основи, що забезпечила б глибокий аналіз текстової інформації. Все це передбачає наявність якісно нових методів та алгоритмів автоматичного лінгвістичного аналізу електронних документів [1]. Адже наявні програмні продукти дозволяють здійснювати пошук, класифікацію, класифікацію текстів, проводити автоматичний морфологічний та синтаксичний аналіз, проте не встановлюють логічні зв'язки між частинами текстових документів, не визначають її структуру та не вирішують проблему екстракції знань [2].

Таким чином, актуальною залишається проблема створення систем автоматичного змістовного аналізу електронних текстових документів, що в свою чергу передбачає розроблення методики їх функціонування, проектування структури та розроблення формального апарату для кожного блока такої системи.

Аналіз останніх досліджень і публікацій. Аналіз останніх досліджень у сфері розроблення інтелектуальних систем оброблення текстової інформації (NLP) дав змогу виділити їх основні недоліки:

- відсутність бази знань у лінгвістичному процесорі (робота відбувається з базами даних) [3];
- функціонування на основі методу шинглів (у разі зміни кількості слів у шинглі системи постійно видають різний відсотковий показник збігів) [4];
- використання статистичних закономірностей для виявлення ключових слів;
- відсутність змістовного аналізу [5].

Роботу інтелектуальних систем оброблення текстової інформації може значно поліпшити використання лінгвістичних корпусів [6] або тезаурусів мови. Останні публікації у сфері інформаційних технологій [7–8] показали, що дослідження у галузі оброблення текстової інформації спрямовані на вдосконалення методів кластеризації та залучення до процесу аналізу нейронних мереж.

Дослідження ж комп'ютерних лінгвістів [9–11] направлені на поглиблення й уточнення можливих варіантів формування логічних зв'язків за рахунок аналізу текстів різноманітної тематики.

Виділення не вирішених раніше частин загальної проблеми. Створення системи автоматичного змістовного аналізу електронних текстових документів вимагає об'єднання зусиль як комп'ютерних лінгвістів та спеціалістів у галузі структурної лінг-

вістики, так і науковців у сфері інформаційних технологій. Основною проблемою на шляху створення такої системи є відсутність комплексної методики формалізованого представлення текстової інформації. Саме на її опис і застосування для побудови нової інформаційної технології порівняльного аналізу текстових документів спрямовані результати дослідження автора.

Мета статті. Метою статті є опис структури інформаційної технології порівняльного аналізу текстових документів як верифікації методики пошуку текстових збігів на основі побудови логіко-лінгвістичних моделей речень природної мови.

Виклад основного матеріалу. Інформаційна технологія порівняльного аналізу текстових документів являє собою сукупність методів, процесів та програмно-технічних засобів, об'єднаних у технологічний ланцюжок, який забезпечує введення, оброблення, зберігання, розподілення та відображення [12] текстової інформації, що представлена у вигляді природномовних електронних документів, для отримання нового показника якості у вигляді відсотку збігів.

Технічними засобами для створення інформаційної технології є апаратне, програмне та математичне забезпечення, що дає можливість здійснювати переробку первинної інформації, представленої у вигляді природномовних електронних документів, в інформацію нової якості. На рис. 1 представлено основні компоненти інформаційної технології порівняльного аналізу текстових документів.

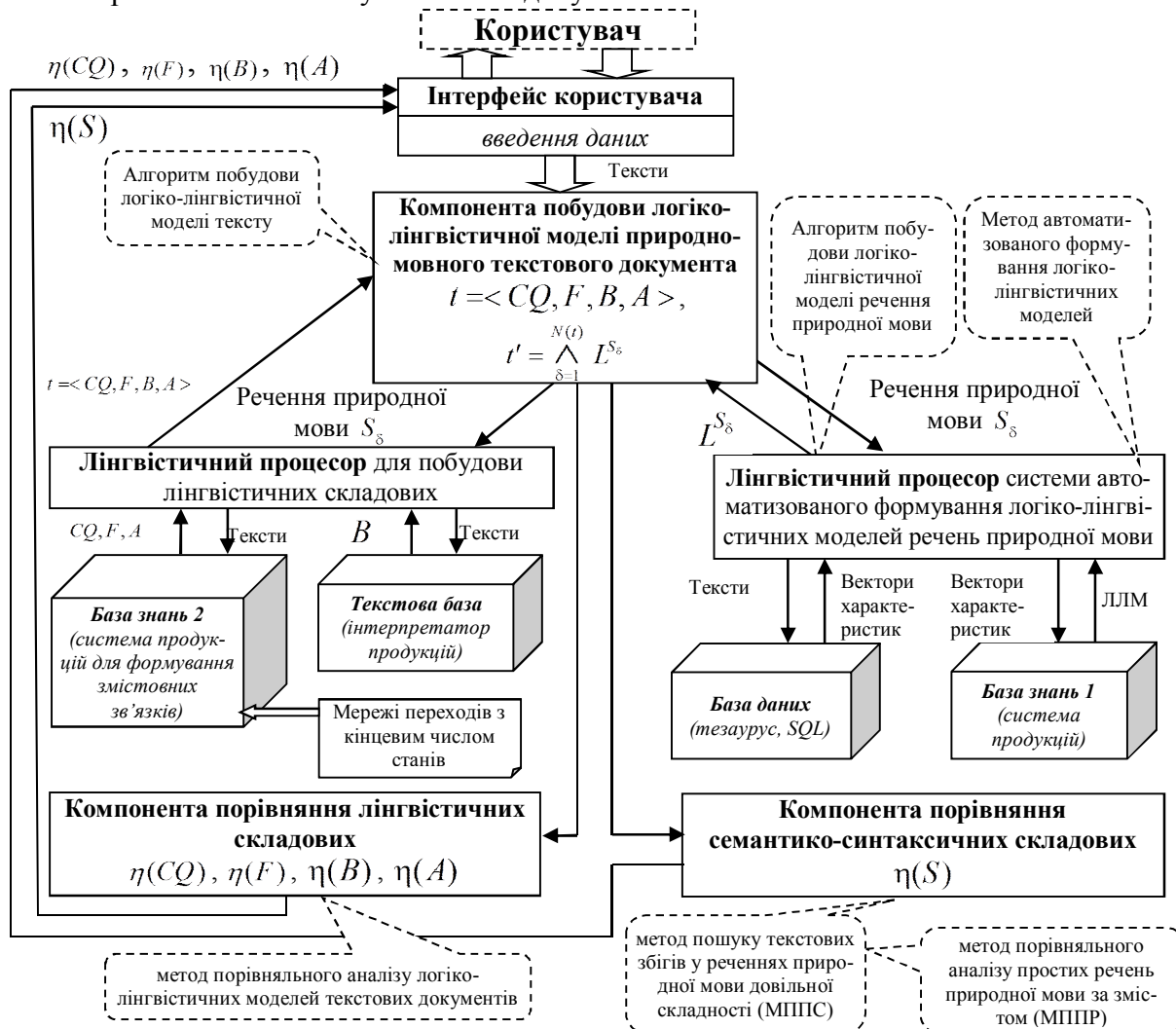


Рис. 1. Структурно-аналітична схема інформаційної технології порівняльного аналізу текстових документів

Інтерфейс користувача (рис. 2) являє собою програмний модуль, який відповідає етапу введення текстової інформації.

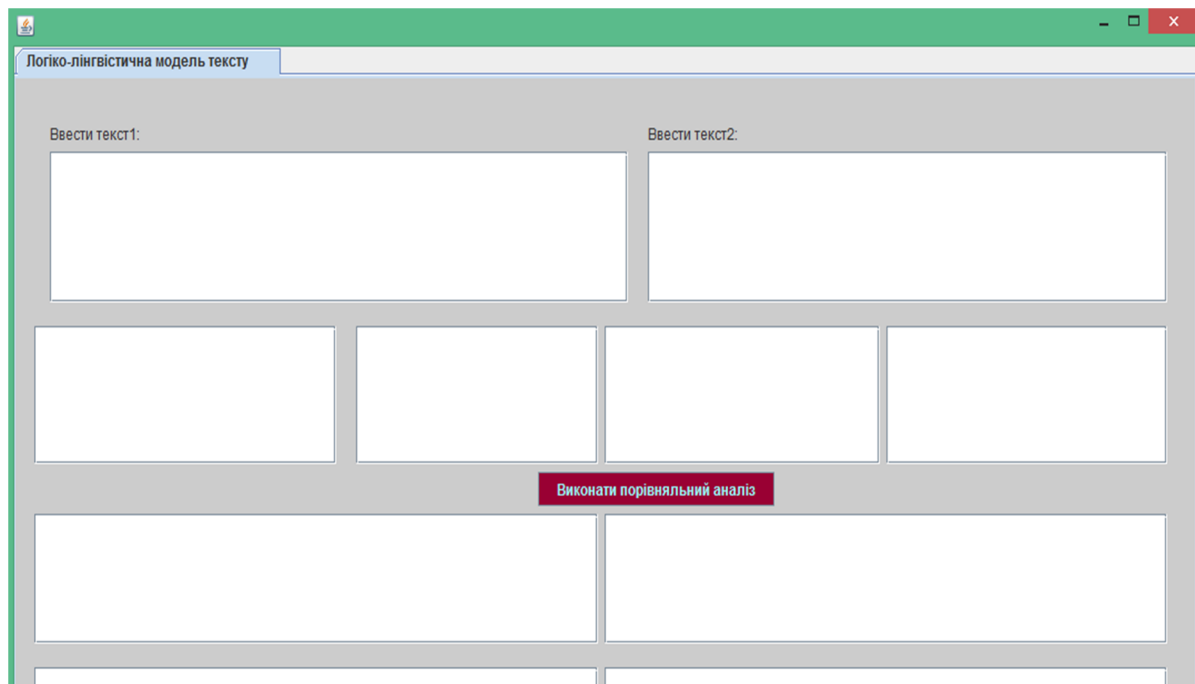


Рис. 2. Інтерфейс користувача інформаційної технології порівняльного аналізу текстових документів

У полях введення обираються два тексти, порівняльний аналіз яких необхідно здійснити. На тексти, які вводяться, накладаються певні обмеження:

- 1) система обробляє речення української мови;
- 2) речення повинно бути грамотно написано;
- 3) слова у реченні розділені пробілами;
- 4) кінцем речення вважаються такі знаки пунктуації, як «.», «!», «?»;
- 5) знаки пунктуації необхідно вказувати безпосередньо після слова, далі пробіл;
- 6) тире і дефіс позначаються в реченні одним і тим самим символом «-», після тире ставиться пробіл, після дефіса – одразу записується слово.

Компонента побудови логіко-лінгвістичної моделі природномовного текстового документа відповідає першому етапу методу порівняльного аналізу логіко-лінгвістичних моделей текстових документів – побудові таких моделей [13]. Ця компонента входить до блока оброблення інформації і включає в себе роботу двох лінгвістичних процесорів: один з яких працює на основі методу автоматизованого формування логіко-лінгвістичних моделей текстової інформації, а другий – на основі системи продукцій для формування змістовних зв'язків у тексті.

Формалізація інтелектуальної діяльності людини полягає у побудові функціональної моделі природної мови, тому логіко-лінгвістична модель виступає як транслятор, що встановлює відповідність між текстом та його змістом [14].

Компонента системи, що реалізує створення формальної лінгвістичної моделі та здатна працювати з природною мовою у всьому її об'ємі, є **лінгвістичним процесором**. Дві основні функції лінгвістичного процесору полягають у вилученні змісту із заданої текстової інформації та вираженні отриманого змісту мовою логіки предикатів. Зміст висловлювання – це вся семантико-прагматична інформація, яку користувач передає на вхід системи. Внутрішнє представлення змісту містить сутності проблемної частини (слова), що потрапляють до системи з цим висловлюванням, властивості та відношення, що відповідають цим сутностям. Тобто лінгвістичний процесор здійснює аналіз

TECHNICAL SCIENCES AND TECHNOLOGIES

введеної текстової інформації згідно з алгоритмом, описаним у методі автоматизованого формування логіко-лінгвістичних моделей тексту, виокремлюючи сутності, властивості та концептуальні відносини між ними, таким чином готуючи матеріал для виведення у вигляді логіко-лінгвістичної моделі.

Лінгвістичний процесор являє собою багаторівневий перетворювач, що містить три рівні представлення тексту: морфологічний, синтаксичний та семантичний (рис. 3). Кожному рівню відповідає один з етапів виконання методу автоматизованого формування логіко-лінгвістичної моделі відповідно: ідентифікація вхідного тексту, концептуалізація та синтаксичний аналіз.

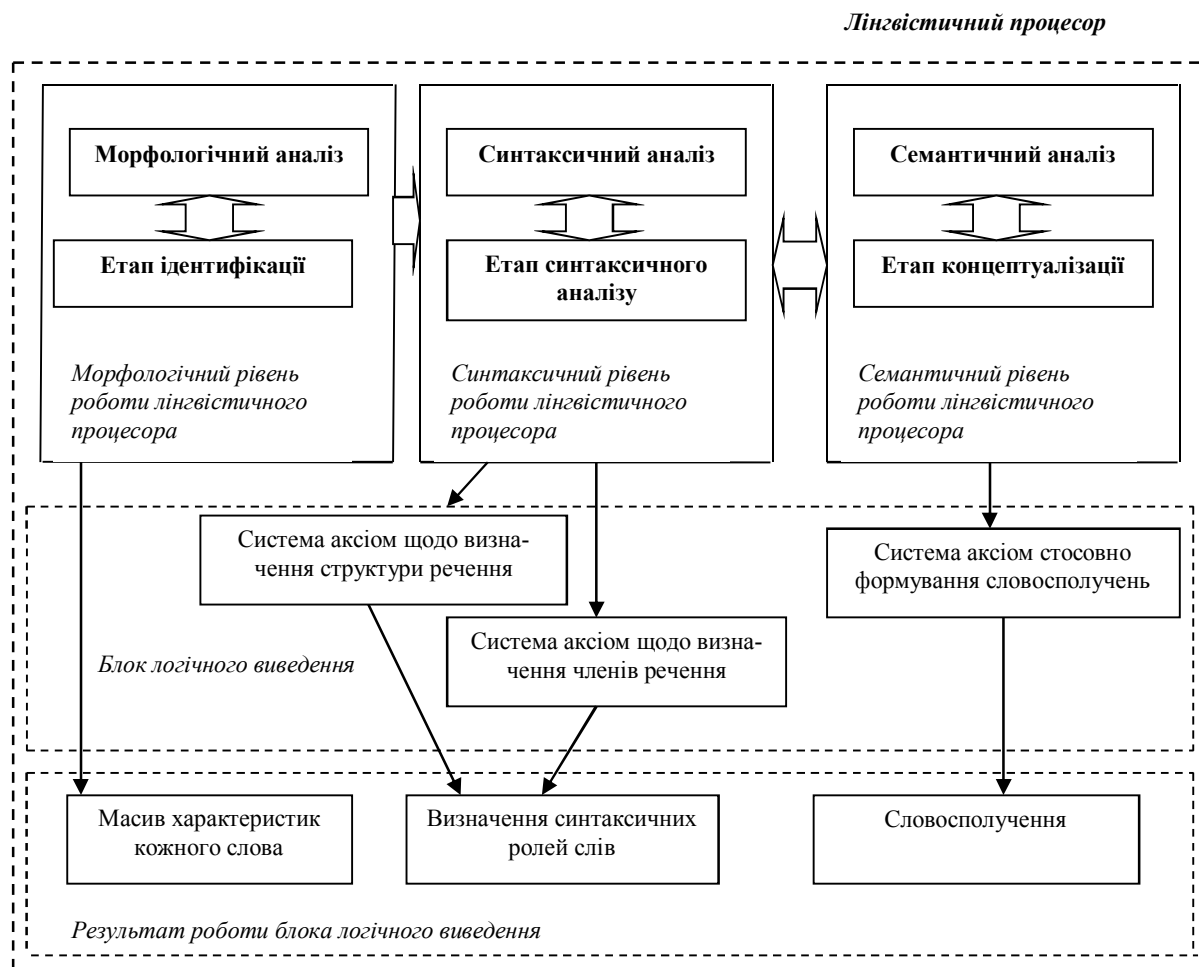


Рис. 3. Структура лінгвістичного процесора для побудови лінгвістичних складових

Предметною областю інформаційної технології порівняльного аналізу текстових документів є вся природна, флективна мова. Тому ця система не ґрунтується на стандартних шаблонах, в які включені основні слова, що стосуються певної предметної області (наприклад, медицини, діагностування, продажу тощо), і користувач не повинен обмежувати себе у використанні певної термінології.

База даних, що використовується в інформаційній технології порівняльного аналізу електронних текстових документів, складається з таблиць, кожна з яких відповідає за певну частину мови, конкретне поле кожної з таблиць характеризується граматичними параметрами, такими як відмінок, рід, число, час, особа і т. ін. Це дає змогу під час звернення до бази даних надавати конкретних значень елементам формальної системи і внаслідок цього формувати масиви характеристик для кожного конкретного слова, що подається на вхід інтелектуальної системи.

Компонента побудови логіко-лінгвістичної моделі текстового документа використовує для своєї роботи дві *бази знань*. *Перша база знань* являє собою систему продукційних правил для визначення логічних зв'язків між словами речень природної мови, це правила утворення словосполучень та привила визначення синтаксичних ролей. *Друга база знань* містить систему продукцій для пошуку логічних зв'язків між фрагментами тексту. Продукції побудовані на основі досліджених закономірностей використання у текстах засобів когезії та на базі основних принципів синтезу логіко-лінгвістичних моделей речень природної мови [1].

Компонента порівняння семантико-синтаксичних складових функціонує на основі методу порівняльного аналізу простих речень природної мови за змістом та методу пошуку текстових збігів у реченнях природної мови довільної складності [13]. За методом пошуку текстових збігів у реченнях природної мови довільної складності почергово порівнюються логіко-лінгвістичні моделі $L^{S_{\delta_1}}$ та $L^{S_{\delta_2}}$, $\delta_1 = \overline{1, N(t_1)}$, $\delta_2 = \overline{1, N(t_2)}$. Це дає можливість визначити ступінь схожості речень, що входять до текстів. Основою порівняльного аналізу є формування інваріантних логіко-лінгвістичних моделей, які ґрунтуються на правилах та моделях побудови складних синонімічних конструкцій природної мови. Інваріантні форми логіко-лінгвістичних моделей безпосередньо впливають на розрахунок відсотку збігів у реченнях текстів. Після виконання цього етапу буде знайдено відсоток збігів $\eta(S)$ для електронних текстових документів t_1 і t_2 .

Компонента порівняння лінгвістичних складових почергово порівнює параметри лінгвістичних складових логіко-лінгвістичних моделей текстів, що розглядаються, після чого визначаються чотири відсотки збігу: $\eta(CQ)$ – за стилістикою, $\eta(F)$ – за структурою складних синтаксичних частин, $\eta(B)$ – за схожістю ключових слів та словосполучень, $\eta(A)$ – за тематикою. Визначення відсотку збігів за різними компонентами логіко-лінгвістичної моделі електронного текстового документа дає можливість проаналізувати характер збігів, а також визначити засоби утворення текстових дублікатів. Результати функціонування інформаційної технології порівняльного аналізу текстових документів продемонстровано на рис. 4.

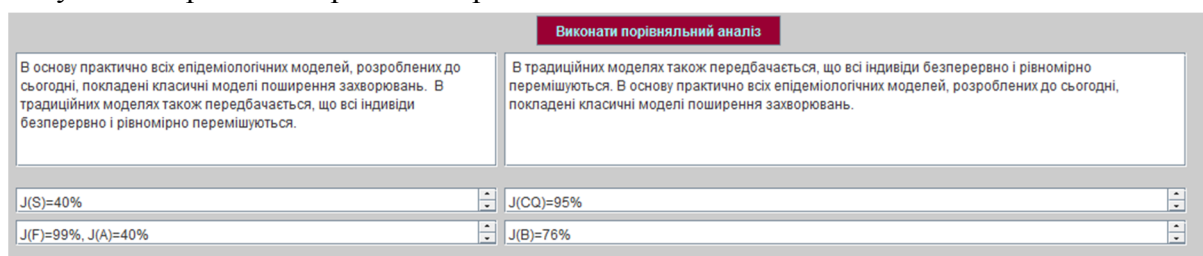


Рис. 4. Результати функціонування інформаційної технології порівняльного аналізу текстових документів

Висновки і пропозиції. Інформаційна технологія порівняльного аналізу текстових документів спроектована на основі використання нової методики, що включає в себе: алгоритм побудови логіко-лінгвістичної моделі речення природної мови, метод автоматизованого формування логіко-лінгвістичних моделей текстової інформації, алгоритм побудови логіко-лінгвістичної моделі текстового документа, метод порівняльного аналізу простих речень природної мови, метод порівняння речень природної мови довільної складності, метод порівняльного аналізу логіко-лінгвістичних моделей текстових документів.

Система, створена на основі розробленої інформаційної технології, здійснює змістовний аналіз електронних документів і є верифікацією методики пошуку текстових збігів на основі побудови логіко-лінгвістичних моделей речень природної мови.

Список використаних джерел

1. *Вавиленкова А. И.* Основные принципы синтеза логико-лингвистических моделей / А. И. Вавиленкова // Кибернетика и системный анализ. – 2015. – Т. 51, № 5. – С. 176–185.
2. *Методы и модели анализа данных: OLAP и Data Mining* / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб. : БХВ-Петербург, 2007. – 384 с.
3. *Онлайн-сервис проверки текста на уникальность*, 2016 [Электронный ресурс]. – Режим доступа : www.text.ru.
4. *Advego Plagiatus* – проверка уникальности текста, 2008–2016 [Электронный ресурс]. – Режим доступа : <http://advego.ru/plagiatus/top/>.
5. *Dan Jurafsky*, Professor of Linguistics, Christopher Manning, Associate Professor of Computer Science Natural Language Processing, Stanford University, 2012, <https://www.coursera.org/course/nlp>.
6. *Словники України он-лайн, 2001–2014* [Електронний ресурс]. – Режим доступу : <http://lcorp.ulif.org.ua/dictua/>.
7. *Рассел С.* Искусственный интеллект: современный подход (AIMA-2) / С. Рассел, П. Норвиг. – 2-е изд. – М. : Вильямс, 2015. – 1408 с.
8. *Кормен Т.* Алгоритмы: построение и анализ / Т. Кормен, И. Чарльз, Лейзерсон, Рональд Л. Ривест, Клиффорд Штайн. – 3-е изд. – М. : Вильямс, 2013. – 1328 с.
9. *Белоусов К. И.* Теория и методология полиструктурного синтеза текста: монография / К. И. Белоусов. – 2-е изд., стер. – М. : ФЛИНТА, 2012. – 214 с.
10. *Корпусна лінгвістика* / [Широков В. А., Бугаков О. В., Грязнухіна Т. О. та ін.]. – К. : Довіра, 2005. – 471 с.
11. *Никитин М. В.* Курс лингвистической семантики / М. В. Никитин. – СПб. : Из-во РГПУ им. Герцена, 2007. – 819 с.
12. *Марчук Ю. Н.* Компьютерная лингвистика / Ю. Н. Марчук. – М. : АСТ: Восток – Запад, 2007. – 317 с.
13. *Вавиленкова А. И.* Порівняльний аналіз речень природної мови за змістом / А. И. Вавиленкова // Математичні машини та системи. – 2015. – № 2. – С. 97–103.
14. *Vavilenkova A. I.* The formal model of knowledge retrieval and processing / A. I. Vavilenkova // Black Sea Scientific journal of academic research. – 2014. – V. 16. – P. 115–119.

Вавиленкова Анастасія Ігорівна – кандидат технічних наук, доцент, доцент кафедри комп'ютеризованих систем управління, Національний авіаційний університет (просп. Космонавта Комарова, 1, м. Київ, 03068, Україна).

Вавиленкова Анастасія Ігорівна – кандидат технических наук, доцент, доцент кафедры компьютеризированных систем управления, Национальный авиационный университет (просп. Космонавта Комарова, 1, г. Киев, 03068, Украина).

Vavilenkova Anastasiia – PhD in Technical Sciences, Associate Professor, Associate Professor of Computerized Control Systems Department, National Aviation University (1 Kosmonavta Komarova Av., 03068 Kyiv, Ukraine).

E-mail: a_vavilenkova@mail.ru

УДК 621.3.08

Володимир Войтенко, Оксана Федорова, Роман Єршов

**ЕЛЕКТРОННА СИСТЕМА РЕЄСТРАЦІЇ ТА ОБРОБЛЕННЯ
БІОЕЛЕКТРИЧНИХ СИГНАЛІВ**

Владимир Войтенко, Оксана Федорова, Роман Ершов

**ЭЛЕКТРОННАЯ СИСТЕМА РЕГИСТРАЦИИ И ОБРАБОТКИ
БИОЭЛЕКТРИЧЕСКИХ СИГНАЛОВ**

Volodymyr Voitenko, Oksana Fedorova, Roman Yershov

**ELECTRONIC SYSTEM FOR BIOELECTRICAL SIGNAL REGISTRATION
AND PROCESSING**

Методи запису біопотенціалів, що генерує мозок, серце або м'язи, мають два основних аспекти застосування: медичний та дослідницький. Медична діагностика в умовах стаціонару найчастіше спирається на такі методи неінвазивних досліджень, в яких використовується громіздка та вартісна апаратура. Водночас актуальною про-