

## ПОРІВНЯЛЬНИЙ АНАЛІЗ РЕЧЕНЬ ПРИРОДНОЇ МОВИ ЗА ЗМІСТОМ

\*Національний авіаційний університет, Київ, Україна

**Анотація.** У статті запропоновано метод порівняльного аналізу речень природної мови довільної складності за змістом, що базується на побудові їх логіко-лінгвістичних моделей. У представленому алгоритмі зосереджено увагу на етапі побудови інваріантних логіко-лінгвістичних моделей та їх безпосередньому впливі на розрахунок відсотку збігів.

**Ключові слова:** логіко-лінгвістична модель, порівняльний аналіз, природна мова, текстові збіги, інваріантні моделі.

**Аннотация.** В статье предложен метод сравнительного анализа предложений естественного языка произвольной сложности по содержанию, который основывается на построении их логико-лингвистических моделей. В представленном алгоритме внимание сосредоточено на этапе построения инвариантных логико-лингвистических моделей и их непосредственном влиянии на расчет процента совпадений.

**Ключевые слова:** логико-лингвистическая модель, сравнительный анализ, естественный язык, текстовые совпадения, инвариантные модели.

**Abstract.** The article proposes a method of comparative analysis of natural language sentences of arbitrary complexity, which is based on the construction of logic and linguistic models. In the presented algorithm the attention is focused on the stage of construction of invariant logic and linguistic models and their direct impact on the calculation of the percentage of coincidences.

**Keywords:** logic and linguistic model, comparative analysis, natural language, text coincidences, invariant models.

### 1. Вступ

Постійне зростання кількості електронних документів спричиняє дублювання матеріалів в інформаційних мережах. У зв'язку з цим, у всіх сферах соціального середовища виникає проблема появи ідентичних за змістом копій документів. Так, в освіті – це плагіат курсових та дипломних робіт, у політичній сфері – створення суперечливих законопроектів, в інформаційних технологіях – відсутність можливості здійснення коректної пошукової оптимізації, в економіці – створення однотипних проектів та ін. Все це загрожує дезінформацією, виникненням логічних протиріч і неточністю прийняття рішень.

Найбільш наближеним підходом до пошуку логічних зв'язків у текстовій інформації на сьогодні є створення семантичних мереж у вигляді графів зв'язності, де ключову роль відіграє відстань між словами, а схожість текстів оцінюється як середнє арифметичне цих відстаней [1]. Але зміст природномовних текстів неможливо виявити лише за допомогою статистичних закономірностей. Дана проблема вимагає об'єднання зусиль спеціалістів із області лінгвістики, інформаційних технологій, математики та філософії.

### 2. Постановка задачі

Для відслідковування істинності тих чи інших пропозицій у сучасних системах аналізу текстових даних використовується стандартний або дещо модифікований алгоритм:

- кешування логічного виводу;
- надання засобів формування конструктивних припущень та аналізу їх корисності;
- аналіз та усунення можливих протиріч у моделях [2].

Такий підхід був запропонований Дойлом та Мак-Аллестером для побудови системи відслідковування істинності пропозицій для знаходження єдиного вирішення проблеми,

що задовольняє певним обмеженням. Проте такий алгоритм не дозволяє виконати порівняння альтернативних варіантів вирішення проблеми.

Саме формалізація альтернативних варіантів запису текстової інформації у вигляді інваріантних форм логіко-лінгвістичних моделей лежить в основі матеріалів дослідження. Адже, ані такі методи пошуку текстових збігів, як метод шинглів [3] та методи пошуку підрядків у рядочках [4], що лежать в основі систем аналітичної обробки текстової інформації, ані створення семантичних мереж з веденням бази синонімів [5], досі не забезпечили якісної обробки електронних документів за змістом.

Над аналізом структури текстів та дослідженням закономірностей організації у них змістовних зв'язків працюють такі лінгвісти та філологи, як Філіппов К.А. [6], Головкіна С.Х. [7], Мельчук І.А. [8] та ін., проте вони не використовують у своїй роботі формальний апарат.

Тому актуальним залишається питання розробки єдиної теоретичної бази, що дозволила б аналізувати текстову інформацію на різних рівнях (прості, складні речення, абзаци, складні синтаксичні конструкції) за єдиним принципом.

### 3. Метод порівняльного аналізу речень природної мови довільної складності за змістом

Для розуміння змісту складних текстових конструкцій важливо зрозуміти, як згруповані між собою прості речення, що входять до їх складу [9]. Тому в основі методу пошуку текстових збігів у реченнях природної мови довільної складності лежать аналіз логіко-лінгвістичних моделей, правила побудови складних синтаксичних конструкцій та способи організації граматичного зв'язку між предикативними частинами.

Нехай два складних речення  $S_1$  та  $S_2$  описуються логіко-лінгвістичними моделями:

$$L(S_1) = \bigwedge_{\mu_1=1}^{v(S_1)} L_{\mu_1}(S_1), \quad (1)$$

$$L(S_2) = \bigwedge_{\mu_2=1}^{v(S_2)} L_{\mu_2}(S_2), \quad (2)$$

де  $L_{\mu_1}(S_1)$  – простий предикат, що описує частину речення  $S_1$ , яка відображає закінчений зміст;

$L_{\mu_2}(S_2)$  – простий предикат, що описує частину речення  $S_2$ , яка відображає закінчений зміст;

$\mu_1 = \overline{1, v(S_1)}, v(S_1)$  – кількість частин речення  $S_1$ , які мають закінчений зміст;

$\mu_2 = \overline{1, v(S_2)}, v(S_2)$  – кількість частин речення  $S_2$ , які мають закінчений зміст.

Алгоритм методу порівняльного аналізу речень природної мови довільної складності складається з таких етапів.

1. Побудова вектора характеристик речення. На цьому етапі для кожного складного речення природної мови  $S_1$  та  $S_2$  ідентифікуються параметри заданих моделей (1) та (2). Для кожної логіко-лінгвістичної моделі встановлюються значення кількості атомарних предикатів  $v(S_1)$  та  $v(S_2)$ , фіксуються логічні операції, використані у моделях (1) та (2). Внаслідок цього формується вектор характеристик  $U(S)$  складного речення:

$$U(S) = \{v(S), O(S), W(S)\},$$

де  $v(S)$  – кількість атомарних предикатів, що описують частини речення  $S$  та відображають закінчений зміст;

$O(S)$  – кортеж логічних операцій речення  $S$  :

$$O(S) = [o_e(S) \mid e = \overline{1, h}],$$

де  $h$  – загальна кількість логічних операцій, наявних у реченні  $S$  ;

$W(S)$  – вид окремої форми уніфікованої логіко-лінгвістичної моделі.

Таким чином, після виконання першого кроку алгоритму порівняльного аналізу речень природної мови довільної складності логіко-лінгвістичні моделі (1) та (2) будуть мати такі вектори характеристик:

$$U(S_1) = \{v(S_1), O(S_1), W(S_1)\}$$

та

$$U(S_2) = \{v(S_2), O(S_2), W(S_2)\}.$$

2. Ідентифікація елементів логіко-лінгвістичних моделей. Для кожного атомарного предиката  $L_{\mu_1}(S_1)$  та  $L_{\mu_2}(S_2)$  фіксуються предикатні змінні та константи, внаслідок чого кожній компоненті логіко-лінгвістичної моделі ставиться у відповідність конкретне значення концепту з моделі (1) та (2) відповідно [10]. Результатом виконання даного етапу є дві таблиці відношень (табл. 1, табл. 2).

Таблиця 1. Таблиця відношень складного речення  $S_1$

$\mu / \text{№}$	$P$	$x$	$c(x)$	$y$	$c(y)$	$z$	$c(z)$	$c(p)$
1	$p_1$	$x_1$	$c_1(x_1)$	$y_1$	$c_1(y_1)$	$z_1$	$c_1(z_1)$	$c_1(p_1)$
...	...	...	...	...	...	...	...	...
$\mu_1$	$p_{\mu_1}$	$x_{\mu_1}$	$c_{1\mu_1}(x_{\mu_1})$	$y_{\mu_1}$	$c_{\mu_1}(y_{\mu_1})$	$z_{\mu_1}$	$c_{\mu_1}(z_{\mu_1})$	$c_{\mu_1}(p_{\mu_1})$
...	...	...	...	...	...	...	...	...
$v(S_1)$	$p_{v(S_1)}$	$x_{v(S_1)}$	$c_{v(S_1)}(x_{v(S_1)})$	$y_{v(S_1)}$	$c_{v(S_1)}(y_{v(S_1)})$	$z_{v(S_1)}$	$c_{v(S_1)}(z_{v(S_1)})$	$c_{v(S_1)}(p_{v(S_1)})$

Таблиця 2. Таблиця відношень складного речення  $S_2$

$\mu / \text{№}$	$P$	$x$	$c(x)$	$y$	$c(y)$	$z$	$c(z)$	$c(p)$
1	$p_1$	$x_1$	$c_1(x_1)$	$y_1$	$c_1(y_1)$	$z_1$	$c_1(z_1)$	$c_1(p_1)$
...	...	...	...	...	...	...	...	...
$\mu_2$	$p_{\mu_2}$	$x_{\mu_2}$	$c_{1\mu_2}(x_{\mu_2})$	$y_{\mu_2}$	$c_{\mu_2}(y_{\mu_2})$	$z_{\mu_2}$	$c_{\mu_2}(z_{\mu_2})$	$c_{\mu_2}(p_{\mu_2})$
...	...	...	...	...	...	...	...	...
$v(S_2)$	$p_{v(S_2)}$	$x_{v(S_2)}$	$c_{v(S_2)}(x_{v(S_2)})$	$y_{v(S_2)}$	$c_{v(S_2)}(y_{v(S_2)})$	$z_{v(S_2)}$	$c_{v(S_2)}(z_{v(S_2)})$	$c_{v(S_2)}(p_{v(S_2)})$

3. Формування інваріантних логіко-лінгвістичних моделей. Кожному реченню при-  
таманий предикат (присудок), його визначення – логічний суб'єкт, що задає граматико-  
змістовну перспективу речення. Інваріантні форми логіко-лінгвістичних моделей опису-  
ють речення природної мови, що мають одне й те саме значення, але мають різну глибинну  
структуру.

Залежно від типу логіко-лінгвістичної моделі, до якого належать задані формули (1)  
та (2), формуються інваріантні форми, що відповідають тотожним за змістом реченням  
природної мови:

$$(L(S_1) \rightarrow Q(S_1)) \& (Q(S_1) \rightarrow L(S_1))$$

та

$$(L(S_2) \rightarrow Q(S_2)) \& (Q(S_2) \rightarrow L(S_2)),$$

де  $Q(S_1) \subseteq Q1$ , де  $Q1$  – множина можливих інваріантних форм логіко-лінгвістичної моделі  $L(S_1)$ ;

де  $Q(S_2) \subseteq Q2$ , де  $Q2$  – множина можливих інваріантних форм логіко-лінгвістичної моделі  $L(S_2)$ .

Отже, для кожного речення природної мови  $S_1$  та  $S_2$  формуються інваріантні логіко-лінгвістичні моделі:

$$Q(S_1) = \bigwedge_{\varphi_1=1}^{\vartheta(S_1)} Q_{\varphi_1}(S_1), \quad (3)$$

$$Q(S_2) = \bigwedge_{\varphi_2=1}^{\vartheta(S_2)} Q_{\varphi_2}(S_2), \quad (4)$$

де  $Q_{\varphi_1}(S_1)$  – простий предикат, що описує інваріантну до  $L(S_1)$  частину речення  $S_1$ , яка відображає закінчений зміст;

$Q_{\varphi_2}(S_2)$  – простий предикат, що описує інваріантну до  $L(S_2)$  частину речення  $S_2$ , яка відображає закінчений зміст;

$\varphi_1 = \overline{1, \vartheta(S_1)}$ ,  $\vartheta(S_1)$  – кількість інваріантних до  $L(S_1)$  простих предикатів, які мають закінчений зміст і описують речення  $S_1$ .

$\varphi_2 = \overline{1, \vartheta(S_2)}$ ,  $\vartheta(S_2)$  – кількість інваріантних до  $L(S_2)$  простих предикатів, які мають закінчений зміст і описують речення  $S_2$ .

$\vartheta(S_1)$  та  $\vartheta(S_2)$  – величини, що варіюються залежно від кількості спрацьованих правил.

Таким чином, після виконання етапу формування інваріантних логіко-лінгвістичних моделей для кожного речення природної мови табл. 1 та табл. 2 доповняться і будуть містити всі можливі варіанти відношень, що будуть відповідати синонімічно тотожним конструкціям.

Таблиця 3. Таблиця відношень складного речення  $S_1$  з урахуванням інваріантних моделей

$\mu / \text{№}$	$P$	$x$	$c(x)$	$y$	$c(y)$	$z$	$c(z)$	$c(p)$
1	$p_1$	$x_1$	$c_1(x_1)$	$y_1$	$c_1(y_1)$	$z_1$	$c_1(z_1)$	$c_1(p_1)$
...	...	...	...	...	...	...	...	...
$\mu_1$	$p_{\mu_1}$	$x_{\mu_1}$	$c_{1\mu_1}(x_{\mu_1})$	$y_{\mu_1}$	$c_{\mu_1}(y_{\mu_1})$	$z_{\mu_1}$	$c_{\mu_1}(z_{\mu_1})$	$c_{\mu_1}(p_{\mu_1})$
...	...	...	...	...	...	...	...	...
$v(S_1)$	$P_{v(S_1)}$	$x_{v(S_1)}$	$c_{v(S_1)}(x_{v(S_1)})$	$y_{v(S_1)}$	$c_{v(S_1)}(y_{v(S_1)})$	$z_{v(S_1)}$	$c_{v(S_1)}(z_{v(S_1)})$	$c_{v(S_1)}(p_{v(S_1)})$
...	...	...	...	...	...	...	...	...
$\Phi_1$	$P_{\varphi_1}$	$x_{\varphi_1}$	$c_{1\varphi_1}(x_{\varphi_1})$	$y_{\varphi_1}$	$c_{\varphi_1}(y_{\varphi_1})$	$z_{\varphi_1}$	$c_{\varphi_1}(z_{\varphi_1})$	$c_{\varphi_1}(p_{\varphi_1})$
...	...	...	...	...	...	...	...	...
$\vartheta(S_1)$	$P_{\vartheta(S_1)}$	$x_{\vartheta(S_1)}$	$c_{\vartheta(S_1)}(x_{\vartheta(S_1)})$	$y_{\vartheta(S_1)}$	$c_{\vartheta(S_1)}(y_{\vartheta(S_1)})$	$z_{\vartheta(S_1)}$	$c_{\vartheta(S_1)}(z_{\vartheta(S_1)})$	$c_{\vartheta(S_1)}(p_{\vartheta(S_1)})$

Таблиця 4. Таблиця відношень складного речення  $S_2$  з урахуванням інваріантних моделей

$\mu / \text{№}$	$P$	$x$	$c(x)$	$y$	$c(y)$	$z$	$c(z)$	$c(p)$
1	$p_1$	$x_1$	$c_1(x_1)$	$y_1$	$c_1(y_1)$	$z_1$	$c_1(z_1)$	$c_1(p_1)$
...	...	...	...	...	...	...	...	...
$\mu_2$	$p_{\mu_2}$	$x_{\mu_2}$	$c_{1\mu_2}(x_{\mu_2})$	$y_{\mu_2}$	$c_{\mu_2}(y_{\mu_2})$	$z_{\mu_2}$	$c_{\mu_2}(z_{\mu_2})$	$c_{\mu_2}(p_{\mu_2})$

...	...	...	...	...	...	...	...	...
$v(S_2)$	$p_{v(S_2)}$	$x_{v(S_2)}$	$c_{v(S_2)}(x_{v(S_2)})$	$y_{v(S_2)}$	$c_{v(S_2)}(y_{v(S_2)})$	$z_{v(S_2)}$	$c_{v(S_2)}(z_{v(S_2)})$	$c_{v(S_2)}(p_{v(S_2)})$
$\mu / \text{№}$	$p$	$x$	$c(x)$	$y$	$c(y)$	$z$	$c(z)$	$c(p)$
...	...	...	...	...	...	...	...	...
$\Phi_2$	$p_{\Phi_2}$	$x_{\Phi_2}$	$c_{1\Phi_2}(x_{\Phi_2})$	$y_{\Phi_2}$	$c_{\Phi_2}(y_{\Phi_2})$	$z_{\Phi_2}$	$c_{\Phi_2}(z_{\Phi_2})$	$c_{\Phi_2}(p_{\Phi_2})$
...	...	...	...	...	...	...	...	...
$\vartheta(S_2)$	$p_{\vartheta(S_2)}$	$x_{\vartheta(S_2)}$	$c_{\vartheta(S_2)}(x_{\vartheta(S_2)})$	$y_{\vartheta(S_2)}$	$c_{\vartheta(S_2)}(y_{\vartheta(S_2)})$	$z_{\vartheta(S_2)}$	$c_{\vartheta(S_2)}(z_{\vartheta(S_2)})$	$c_{\vartheta(S_2)}(p_{\vartheta(S_2)})$

4. Порівняння векторів характеристик. На даному етапі відбувається порівняння векторів характеристик двох заданих речень  $S_1$  та  $S_2$ :

$$U(S_1) = \{v(S_1), O(S_1), W(S_1)\},$$

$$U(S_2) = \{v(S_2), O(S_2), W(S_2)\}.$$

1) Зафіксувати значення  $W(S_1)$  та  $W(S_2)$  видів окремої форми уніфікованої логіко-лінгвістичної моделі.

2) Якщо кількість атомарних предикатів, що описують частини речення, які відображають закінчений зміст, для обох речень однакова  $v(S_1) = v(S_2)$ , то перейти до п. 3), інакше перейти до п. 5.

3) Порівняти кортежі логічних операцій, використані в реченнях  $S_1$  та  $S_2$ :

$$O(S_1) = [o_{e_1}(S_1) \mid e_1 = \overline{1, h_1}],$$

$$O(S_2) = [o_{e_2}(S_2) \mid e_2 = \overline{1, h_2}],$$

– якщо елементи кортежу логічних операцій  $o_1(S_1), \dots, o_{e_1}(S_1), \dots, o_{h_1}(S_1)$  повністю збігаються з елементами кортежу логічних операцій  $o_1(S_2), \dots, o_{e_2}(S_2), \dots, o_{h_2}(S_2)$ , тобто  $O(S_1) = O(S_2)$ , і кількість елементів першого і другого кортежів співпадає,  $h_1 = h_2$  (причому обов'язковим є врахування порядку елементів у кортежі), то виконати п. 5 і присвоїти  $\Delta h = 0$ ;

– якщо кількість елементів у кортежі логічних операцій одного з речень відрізняється від кількості елементів у кортежі логічних операцій другого, і при цьому елементи кортежу першого речення включають у себе елементи кортежу другого речення  $O(S_2)$ , тобто  $O(S_2) \subseteq O(S_1)$  при  $h_1 > h_2$ , або навпаки, елементи кортежу логічних операцій першого речення входять до множини логічних операцій речення  $S_2$ , тобто  $O(S_1) \subseteq O(S_2)$  при  $h_1 < h_2$ , то перейти до п. 5 і присвоїти:

а)  $\Delta h = h_1 - h_2$  при  $h_1 > h_2$ ;

б)  $\Delta h = h_2 - h_1$  при  $h_1 < h_2$ ;

– якщо кількість елементів у кортежах логічних операцій однакова  $h_1 = h_2$  і елементи кортежів  $O(S_1)$  та  $O(S_2)$  не повністю співпадають, проте  $(O(S_2) \subseteq O(S_1)) \& (O(S_2) \subseteq O(S_1))$ , то перейти до п. 5 і присвоїти  $\Delta h = \emptyset$ ;

– інакше, коли елементи кортежів логічних операцій речень  $S_1$  та  $S_2$  не співпадають,  $O(S_1) \neq O(S_2)$ , виконати п. 5 і присвоїти  $\Delta h = \otimes$ .

**5. Порівняння логіко-лінгвістичних моделей.** Даний етап передбачає перевірку ряду умов, зокрема, умов протиріччя та синонімії двох речень природної мови, можливість побудови імплікативних та пресуппозиційних зв'язків між атомарними предикатами, а також умови організації граматичного зв'язку між предикатними частинами речень. Внаслідок чого встановлюється істинність або хибність факту виявлення тотожних за змістом речень природної мови довільної складності.

1) Здійснити порівняння кожного простого предиката  $L_{\mu_1}(S_1)$  логіко-лінгвістичної моделі (1) та кожної інваріантної форми  $Q_{\varphi_1}(S_1)$  логіко-лінгвістичної моделі (3) з кожним простим предикатом  $L_{\mu_2}(S_2)$  логіко-лінгвістичної моделі (2) та інваріантними формами  $Q_{\varphi_2}(S_2)$  логіко-лінгвістичної моделі (4) по чергово методом порівняння простих речень за змістом [11].

2) Нехай змінна  $\theta(S)$  – загальна кількість збігів простих предикатів речення природної мови  $S_1$  з простими предикатами речення природної мови  $S_2$ .

3) Нехай змінна  $\tau(S)$  – загальна кількість збігів простих інваріантних предикатів речень природної мови  $S_1$  та  $S_2$ .

4) Перевірити істинність умов та зафіксувати загальну кількість збігів:  
– якщо істинною є умова

$$(L_{\mu_1}(S_1) \equiv L_{\mu_2}(S_2)),$$

то  $\theta(S)$  збільшується на одиницю, тобто  $\theta(S) = \theta(S) + 1$ ;

– якщо істинною є умова

$$(Q_{\varphi_1}(S_1) \equiv L_{\mu_2}(S_2)) \vee (L_{\mu_1}(S_1) \equiv Q_{\varphi_2}(S_2)) \vee (Q_{\varphi_1}(S_1) \equiv Q_{\varphi_2}(S_2)),$$

то  $\tau(S)$  збільшується на одиницю, тобто  $\tau(S) = \tau(S) + 1$ .

1) Загальна кількість порівнянь  $\omega(S)$  обчислюється як  $\omega(S) = (v(S_1) + \vartheta(S_1))(v(S_2) + \vartheta(S_2))$ , тобто дорівнює добутку суми кількості простих та простих інваріантних предикатів речення  $S_1$  і суми кількості простих та простих інваріантних предикатів речення  $S_2$ .

2) Обчислити відсоток збігів  $\eta(S)$  у реченнях природної мови  $S_1$  та  $S_2$ :

– якщо  $W(S_1) = W(S_2)$  і  $\Delta h = 0$ , то

$$\eta(S) = \frac{\theta(S) + \tau(S)}{v(S_1)} \cdot 100\% = \frac{\theta(S) + \tau(S)}{v(S_2)} \cdot 100\% ;$$

– якщо  $W(S_1) = W(S_2)$  і ( $\Delta h = h_1 - h_2$  або  $\Delta h = \emptyset$ ), то

$$\eta(S) = \frac{\theta(S) + \tau(S)}{v(S_2)} \cdot 100\% ;$$

– якщо  $W(S_1) = W(S_2)$  і ( $\Delta h = h_2 - h_1$  або  $\Delta h = \emptyset$ ), то

$$\eta(S) = \frac{\theta(S) + \tau(S)}{v(S_1)} \cdot 100\% ;$$

– якщо  $W(S_1) \neq W(S_2)$  або  $\Delta h = \otimes$ , то

$$\eta(S) = \frac{\tau(S)}{v(S_1)} \cdot 100\% , \text{ при } h_1 < h_2 ,$$

$$\eta(S) = \frac{\tau(S)}{v(S_2)} \cdot 100\% , \text{ при } h_2 < h_1 .$$

Отже, метод порівняльного аналізу речень природної мови довільної складності дає можливість визначити ступінь схожості речень за змістом. Це стає можливим завдяки формуванню інваріантних логіко-лінгвістичних моделей, що базуються на правилах та моделях побудови складних синонімічних конструкцій природної мови.

#### 4. Висновки

Для розуміння змісту складних конструкцій важливо зрозуміти, як згруповані між собою прості речення, що входять до їх складу [12]. Тому, на відміну від методу шинглів, що лежить в основі більшості сучасних систем порівняльного аналізу текстових документів, запропонований у статті метод базується на правилах побудови складних речень природної мови, виявленні механізмів утворення тотожних синтаксичних конструкцій і залежностях відсотку збігів від істинності логіко-лінгвістичних моделей та їх інваріантних форм. Метод порівняльного аналізу речень природної мови довільної складності є посередньою ланкою між порівнянням простих речень природної мови та порівнянням цілих текстових документів за змістом. Тому запропонований метод одночасно використовує для своєї роботи метод порівняння простих речень природної мови та являється основою для методу порівняльного аналізу текстових моделей.

#### СПИСОК ЛІТЕРАТУРИ

1. Тарануха В.Ю. Интеллектуальная обработка текстов / Тарануха В.Ю. – Київ, 2014. – 80 с.
2. Джексон П. Введение в экспертные системы / Джексон П. – Санкт-Петербург: Вильямс, 2001. – 624 с.
3. Вавіленкова А.І. Логіко-лінгвістичні моделі речень як засіб порівняння текстових документів за змістом / А.І. Вавіленкова // Математичні машини та системи. – 2012. – № 1. – С. 166 – 173.
4. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа / Ландэ Д.В. – М.: ООО «Вильямс», 2005. – 272 с.
5. Jurafsky D. Natural Language Processing [Електронний ресурс] / D. Jurafsky, C. Manning. – Stanford University, 2012. – Режим доступу: <https://www.coursera.org/course/nlp>.
6. Филиппов К.А. Лингвистика текста. Курс лекций / Филиппов К.А. – Спб.: Издательство С.-Петербургского университета, 2008. – 336 с.
7. Головкина С.Х. Лингвистический анализ текста / С.Х. Головкина, С.Н. Смольников. – Вологда: Издательский центр ВИРО, 2006. – 124 с.
8. Мельчук И.А. Опыт теории лингвистических моделей «СМЫСЛ-ТЕСТ» / Мельчук И.А. – М.: Школа «Языки русской культуры», 1999. – 346 с.
9. Vavilenkova A. A self-system to identify conceptual relationships in text / A. Vavilenkova // Proceedings of the National Aviation University. – 2015. – № 1 (62). – P. 63 – 69.
10. Вавіленкова А.І. Методологічні основи автоматичного аналізу логіко-лінгвістичних моделей текстових документів / А.І. Вавіленкова // Математичні машини та системи. – 2015. – № 1. – С. 65 – 71.
11. Вавіленкова А.І. Аналіз методів пошуку синонімів в електронних документах / А.І. Вавіленкова // Вісник Чернігівського державного технологічного університету: зб. наук. праць. – (Серія «Технічні науки»). – 2014. – № 2 (73). – С. 119 – 128.
12. Марчук Ю.Н. Компьютерная лингвистика / Марчук Ю.Н. – М.: АСТ: Восток – Запад, 2007. – 317 с.

*Стаття надійшла до редакції 21.04.2015*