

3. Горбач, А. Современная методика совершенствования технического обслуживания медицинского оборудования в практике лечебных учреждений [Текст] / А. Горбач // Медична техніка. — 2008. — № 3(4). — С. 95–99.
4. Багрецов, А. Эффективность медицинского оборудования [Текст] / А. Багрецов // Вопросы организации и информатизации здравоохранения. — 2010. — № 3. — С. 87–89.
5. Новокрещенова, И. Проблемы обеспечения работоспособности медицинского оборудования в условиях модернизации здравоохранения [Текст] / И. Новокрещенова // Бюллетень медицинских Интернет-конференций. — 2013. — Т. 3, № 12. — С. 1343–1347.
6. Абакумов, В. Г. Системы відображення в медицині [Текст] / В. Г. Абакумов, О. І. Рибін, Й. Святош. — К.: ВЕК+, 2000. — 320 с.
7. Кузовик, В. Методы моделирования технологии ремонта АО по техническому состоянию с использованием ЭВМ [Текст] / В. Кузовик, В. Ткаченко // ГУРАТ. — 1984. — № 1. — С. 28–30.
8. Кузовик, В. Идентификация параметров оценки технического состояния электрооборудования ВС на ЭВМ [Текст] / В. Кузовик, В. Ковальчук // Материалы всесоюзной научно-технической конференции «Совершенствование технологических процессов ремонта авиационной техники на заводах гражданской авиации», г. Москва, 14–17 сентября 1984 г. — С. 56.
9. Шеннон, Р. Имитационное моделирование систем [Текст] / Р. Шеннон. — М.: Мир, 1998. — 208 с.
10. Малиновский, А. В. Руководство по ремонту и техническому обслуживанию медицинской техники РМТ 59498076-03-2007 [Текст] / А. В. Малиновский. — СПб.: Медтехника, 2007. — Т. 3, Ч. 2. — 272 с.
11. Фатхутдинов, Р. М. Стратегический менеджмент [Текст] / Р. М. Фатхутдинов. — М.: Наука, 1995. — 273 с.
12. Патерсон, Дж. Теория сетей Петри и моделирование систем [Текст] / Дж. Патерсон. — М.: Мир, 2000. — 263 с.

СТРУКТУРА ФОРМИРОВАНИЯ И ПРЕОБРАЗОВАНИЯ ИНФОРМАЦИОННОГО РЕСУРСА В ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА РЕМОНТА МЕДИЦИНСКОГО ОБОРУДОВАНИЯ

Проведен анализ современных систем эксплуатации медицинского оборудования. Предложен метод оценки фактического технического состояния, основанный на использовании информационного ресурса по оценке функциональных режимов медицинского оборудования. Для практической реализации эффективного метода ремонта приведена структура формирования и преобразования информационного ресурса в информационной технологии технологического процесса ремонта медицинского оборудования.

Ключевые слова: медицинское оборудование, фактическое техническое состояние, информационная технология, компьютеризированная информационная система.

Кучеренко Валентина Леонідівна, кандидат технічних наук, доцент, кафедра біокібернетики та аерокосмічної медицини, Національний авіаційний університет, Київ, Україна, e-mail: bikam_nau@mail.ru.

Кучеренко Валентина Леонидовна, кандидат технических наук, доцент, кафедра биокібернетики та аерокосмічної медицини, Національний авіаційний університет, Київ, Україна.

Kucherenko Valentina, National Aviation University, Kyiv, Ukraine, e-mail: bikam_nau@mail.ru

УДК 004.93:159.95

DOI: 10.15587/2312-8372.2015.37274

Бісікало О. В., Яхимович О. В.

МЕТОД ВИЗНАЧЕННЯ КЛЮЧОВИХ СЛІВ АНГЛОМОВНОГО ТЕКСТУ НА ОСНОВІ DKPRO CORE

Розглянуто підходи до пошуку ключових слів тексту, що грають важливу роль в задачах комп'ютерної лінгвістики. Запропоновано новий метод визначення ключових слів, який базується на знаходженні зв'язків між словоформами англomовного тексту за допомогою інструментальних можливостей пакету DKPro Core. Проілюстрований аналізом прикладів застосування, метод спрямовано на розв'язання задач ефективної обробки текстових документів.

Ключові слова: метод, ключові слова, англійська мова, лінгвістичний пакет, DKPro Core, синтаксичний аналіз.

1. Вступ

Значна кількість доступних лінгвістичних систем, що орієнтовані на обробку природно-мовних текстів, пропонують функції автоматичного виділення ключових слів. Цей функціонал побудовано на певних методах визначення ключових слів, які діляться на лінгвістичні та статистичні. Лінгвістичні методи ґрунтуються на значеннях слів, зокрема використовують онтології та семантичні дані про слово. На жаль, ці методи ресурсомісні на ранніх етапах: розробка онтологій, наприклад, вельми трудомісткий процес [1]. З іншого боку, ста-

тистичні методи супроводжуються значними обсягами «вербального шуму», який суттєво впливає на якість визначення ключових слів. Тому найбільш перспективними для дослідження, на думку авторів, є гібридні методи, для яких швидкість статистичної обробки тексту підсилюється можливостями сучасних лінгвістичних пакетів.

2. Аналіз літературних даних та постановка проблеми

Як відомо, не всі слова в тексті рівнозначні. Є слова, які дозволяють представити текст у згорнутому вигляді,

при цьому таке подання досить точно відображає зміст вихідного тексту. В навчальному процесі результат визначення ключових слів використовується, наприклад, при конспектуванні лекції, книги або статті. Ключові слова — це ті слова, які можуть представити основний зміст тексту.

Основний зміст документа (тексту) може бути виражений за допомогою певних слів, узятих безпосередньо з цього тексту. Як правило, до кожного розгорнутого тексту можна скласти цілий набір ключових слів різного обсягу (від 5 до 15 слів). Але взагалі кількість ключових слів може варіюватися в широких межах. Відповідно до думки Л. В. Цукрового і А. С. Штерн, найоптимальнішим є невеликий набір ключових слів і словосполучень — 7–10 слів. В окремих випадках компресія може привести до виділення одного основного ключового слова, яке має найбільшу частоту.

Ключове слово — це таке слово в тексті, яке здатне в сукупності з іншими ключовими словами представляти текст. Набір ключових слів близький до анотації, плану і конспекту, які теж представляють документ з меншою деталізацією, але, на відміну від ключових слів, пов'язані у синтаксичній структурі.

Термін «ключові слова» значною мірою умовний: як ключові знаки в тексті можуть виступати не тільки слова, а й словосполучення і навіть речення.

Ключове (опорне) слово — це термін, що відноситься до основного змісту тексту і повторюється в ньому кілька разів (з урахуванням всіх можливих синонімів).

Ключове словосполучення — це поєднання слів, серед яких є одне або кілька ключових.

Ключовим реченням може вважатися речення, що містить два і більше ключових слова або ключових словосполучення.

Ключові слова мають ряд суттєвих ознак, зокрема:

- високий ступінь повторюваності даних слів у тексті, частотність їх вживання;

- здатність знака конденсувати, згорнути інформацію, виражену цілим текстом, об'єднувати «його основний зміст»; ця ознака особливо яскраво проявляється у ключових слів у позиції заголовку.

У всіх текстових документах, що створені людиною, можна виділити статистичні закономірності. Їх визначення базується на використанні законів Ципфа [2], який досліджував цікавий парадокс — у будь-якій мові є слова, що зустрічаються частіше, ніж інші, але не мають значення. З іншого боку, є слова, які зустрічаються рідше, але мають набагато більше смислове значення.

Найбільшою популярністю для визначення ключових слів користується модель TF*IDF [3]. Однак застосовуються й інші індексувальні функції, включаючи ймовірні способи індексування [4] і методики індексування структурованих документів [5]. Інші функції індексації можуть знадобитися в тих випадках, коли спочатку навчальну множину не дано і документну частоту не вдається порахувати. У цих випадках TF*IDF змінюють на емпіричні функції [6].

При використанні цього підходу не виключена ймовірність попадання в ключові слова випадкових спеціальних термінів, рідкісних слів і власних імен та іншого «шуму». Тому необхідно в попередню обробку тексту включати алгоритм, що підвищує якість відбору. Евристики такого відбору зазвичай залежать від конкретно взятого випадку.

Однак вибір ключових слів є дуже непростою операцією і вимагає зваженого підходу. Слід вибирати ті ключові слова, які найбільш точно відображають специфіку розглянутої теми. При цьому необхідно уникати випадкових і загальних фраз, не рекомендується повторювати кілька разів одні й ті ж ключові слова. Отже, процес пошуку ключових слів є аналітичним [7].

В процесі попереднього оброблення тексту проводиться видалення неінформативних частин. Стоп-слова — це слова, що не представляють цінності як потенційно ключові. Найчастіше це прийменники, сполучники, вигуки тощо.

Кандидати в ключові слова можна відбирати у вигляді N -грам, що не розділені знаками пунктуації (крім дефіса і лапок) і стоп-словами. Ключовими словами можуть бути як поодинокі слова, так і пари слів, трійки і т. д. Інший відомий підхід ґрунтується на аналізі зв'язків між словами в реченнях і в тексті, отриманих або за допомогою тих же N -грам (підхід менш трудомісткий), або на розібраному тексті. Якщо текст пройшов більш затратну обробку на етапах аналізу (морфологічний дасть базові словоформи, синтаксичний — зв'язки між словами, семантичний — смислову карту зв'язків), можна отримати більш точну інформацію про текст, наприклад, на підставі дерев синтаксичного розбору (за наявності синтаксичної розмітки).

Найважливішим етапом в задачі знаходження ключових фраз є розрахунок їх ваг інформативності, який дозволяє оцінити значимість фраз по відношенню один до одного в документі. Для кожної з відібраних ключових фраз розраховуються ознаки, які дозволяють судити про важливість кандидата для даного документа. Набір відібраних ключових фраз ранжується за значеннями чисельних ознак, наприклад, відповідно до їх частотності та ваги інформативності, розрахованими за однією з відомих методик. Після ранжування проводиться відбір кращих ключових фраз з цього списку або відбираються кандидати, що перевищують встановлений мінімальний поріг значення ознаки [1].

3. Об'єкт, мета та задачі дослідження

Мета роботи полягає у підвищенні точності визначення ключових слів з англійського тексту.

Задачі, що вирішуються у даній роботі: розробка методу визначення ключових слів, знаходження чисельних зв'язків між словами.

Об'єкт дослідження — процес обробки вербальної інформації для визначення ключових слів в тексті.

Предмет дослідження — методи знаходження ключових слів в тексті.

4. Інформаційна оцінка парсерингу тексту для задачі визначення ключових слів

Розглянемо задачу визначення ключових слів тексту як певну інформаційну технологію, що має на вході текст, а на виході — множину з l ключових слів $W^k = \{w_1^k, \dots, w_l^k\}$. Без применшення загальності будемо вважати, що текст T складається з m різних слів, а в окреме його j -те речення з k налічує n слів з m можливих, причому $m \gg n$ та $m \gg l$. Більшість відомих методів визначення ключових слів тексту беруть за основу частотний словник тексту, який фактично є списком або

упорядкованою множиною пар $D = \{ \langle w_i, f_i \rangle \}$, $i = \overline{1, m}$, де w_i — одне слово з m , а f_i — його частота ($f_i \geq f_{i+1}$, $i = \overline{1, m-1}$), що визначена для T . За певною фільтрацією окремих незначущих категорій слів ключовими вважають перші l слів зі списку D , тобто, дещо спрощено маємо $W^k = \{w_1, \dots, w_l\}$.

Проте результати парсерингу природних мов за допомогою сучасних лінгвістичних пакетів дозволяють на доступному програмному рівні [8] оперувати синтаксичними зв'язками між словами окремого речення. Окрім того, можливості цих пакетів дозволяють суттєво зменшити значення m шляхом об'єднання слів у словоформи, а останні — у леми та стеми. Отже, необхідно з'ясувати, які формальні переваги для визначення $W^k = \{w_1^k, \dots, w_l^k\}$ надасть нам програмно-лінгвістичне забезпечення процедури парсерингу всіх речень тексту T .

З інформаційної точки зору розуміння сенсу речення окремим суб'єктом супроводжується розпізнаванням: а) окремих слів, з яких воно складається та б) зв'язків між парами цих слів з відповідною побудовою дерева таких зв'язків [9]. Вважатимемо, що всі ці процеси відбуваються шляхом порівняльного аналізу та залучення інформації з деякої загальнолінгвістичної бази знань суб'єкта розуміння. Якщо кожен з цих етапів супроводжується збільшенням інформації, то приймаємо робочу гіпотезу:

— Рівень загального розуміння тексту T може змінюватися від мінімально можливого до максимального в залежності від обсягу та інших параметрів загальнолінгвістичної бази знань суб'єкта.

— Якість визначення $W^k = \{w_1^k, \dots, w_l^k\}$ пропорційна рівню загального розуміння тексту, що має підтверджуватися формальними ознаками.

Нехай будь-яке j -те речення з k складається з n різних слів, що не є досить жорстким обмеженням. Тоді зв'язне дерево парних залежностей такого речення налічує або $n-1$ гілок, якщо не брати до уваги зворотну залежність між підметом та присудком, або n — якщо брати. Відповідно загальна кількість слів цього речення для подальшого поглибленого аналізу збільшується або до $2 \times n - 2$ або до $2 \times n$. Проте таке збільшення відбувається нерівномірно — для всіх не термінальних (кінцевих) вузлів дерева частоти відповідних слів не змінюються, а для термінальних (проміжних) можуть зрости суттєво. В табл. 1 показані випадки зміни частот слів з урахуванням парних залежностей для різних типів речення.

Навіть такий елементарний аналіз на рівні одного речення показує, що збільшуються частоти саме тих слів, які потенційно можуть належати до множини ключових. Проведемо формальну оцінку такого збільшення для накладених обмежень щодо наявності виключно різних слів у реченні та не врахуванням зворотної залежності між підметом та присудком:

1. Мінімальне збільшення відсутнє за умови знаходження i -го слова з m серед не термінальних (кінцевих) вузлів дерева кожного речення, де це слово зустрічається, тобто $f_i^{\min} = 0$, $f_i^{\text{new}} = f_i$, $i = \overline{1, m}$.

2. Якщо i -те слово знаходиться у кожному з k речень тексту та, окрім того, відповідає у кожному реченні найбільш розгалуженому термінальному вузлу, то максимальне збільшення частоти складає:

$$f_i^{\max} = \sum_{j=1}^k (n_j - 2), \quad i = \overline{1, m}.$$

Відповідно:

$$f_i^{\text{new}} = f_i + f_i^{\max} = k + \sum_{j=1}^k (n_j - 2) = \sum_{j=1}^k (n_j - 1).$$

3. В загальному та більш реальному випадку $f_i = z | z \leq k$, тобто i -те слово знаходиться у z реченнях з k маємо:

$$f_i^{\text{new}} = z + \sum_{j=1}^z (n_j - 2) = \sum_{j=1}^z (n_j - 1),$$

як оцінку зверху збільшення частоти i -го слова.

Отже, експериментальні дослідження мають підтвердити справедливості отриманих формальних оцінок процесу визначення ключових слів тексту у встановлених межах.

5. Реалізація програмного експерименту в DKPro Core

Для експериментальної перевірки результатів теоретичного аналізу було розроблене програмне забезпечення на основі DKPro Core.

Таблиця 1

Аналіз збільшення частоти значимих слів унаслідок урахування парних залежностей для різних типів речення

№ з/п	Склад речення/кількість слів	Тип речення та граф його дерева залежностей	Частотна формула	Кінцева частота
1	Ab/2	Словосполучення (Коріння дерева)	$A + b$	2
2	Abc/3	Лінійна трійка (Бережи скарби природи)	$A + 2b + c$	4
3	Abcd/4	Лінійна четвірка (Отримав переклад слова дивного)	$A + 2b + 2c + d$	6
4	Abcde/5	Розгалуження (Пустий ліс нізвідки завершився проваллям)	$A + 2b + c + 3d + e$	8
5	Abcdef/6	Група підмета (Сині примружені очі коханого говорили багато)	$A + b + 4c + d + 2e + f$	10
6	Abcdef/6	Група присудка (Досвідчений кінь борозну швидко відчує нюхом)	$A + 2b + c + d + 4e + f$	10
7	Abcdef/6	Обидві групи (Старий дід Еол зобрав всіх вітрів)	$A + 3b + c + 2d + e + 2f$	10

DKPro Core — це набір програмних компонентів для обробки природної мови, що базується на Apache UIMA framework. Він був побудований з метою підвищення продуктивності дослідників, які працюють з автоматичним аналізом мови. Підхід DKPro Core полягає в тому, що дослідники повинні мати можливість зосередитися на своїх реальних наукових питаннях, а не на розробці технологій [10, 11].

Визначення ключових слів відбувається за кількома етапами:

- створення багаторівневої розмітки тексту;
- синтаксична розмітка, що враховує складні залежності між парами лем;
- заміна займенників в отриманих парах на відповідні до них іменники;
- розбиття пар на окремі слова і визначення кількості зв'язків;
- вибір перших n слів з найбільшою кількістю зв'язків, де n — кількість потрібних ключових слів.

Аналогами розробленої програми можуть бути сайти SEO оптимізації, де є функція визначення ключових слів. Для даного експерименту вибрані сервіси: advego.ru/text/seo/, rise-top.com/keywordstext.php та seotool.by/analiz/seo/key-wordstext.php.

Для проведення експерименту було взято текст тез «Variability management in software product lines using adaptive object and reflection» [12], де ключові слова задані авторами: Software product line, Variability, Adaptive object model, Reflection, Brazilian Satellites Launcher.

Результати знаходження ключових слів для власної розробки і аналогів наведено в табл. 2.

Таблиця 2

Результати пошуку ключових слів

Слова задані автором	seotool	rise-top	advego	Власна розробка
software	+	+	+	+
product	+	+	+	+
line	+	+	+	+
variability	+	+	+	+
adaptive	+	+	+	—
object	—	—	—	—
model	—	—	—	+
reflection	—	—	—	+
Brazilian	—	—	—	—
Satellites	—	—	—	—
Launcher	—	—	—	—

Як видно з результатів пошуку ключових слів (табл. 2) власна розробка знаходить найбільше слів заданих автором — 6 слів. Аналоги знаходять по 5 слів. Власна розробка, так само як і аналоги, знаходить перших 4 заданих автором слова, але не знаходить п'яте — adaptive, проте вона знаходить такі ключові слова, як model і reflection, чого не роблять аналоги.

Якщо ключові слова не задані автором, то результати роботи можна порівняти з частотним словником даного тексту.

Для тексту «Presidential Address to the Federal Assembly 2013» [13] перші 10 ключових слів за роз-

робкою автора, разом із стоп-словами: we, need, work, I, make, be, system, have, authorities, development, ask, people. Де *be* і *have* — стоп-слова. Перші 10 ключових слів із стоп-словами, для частотного словника: *that, is, we, will, I, our, be, are, must, it, have, not, all, their, work, Russia, need, should, also, Russian, system*. Стоп-слова для частотного словника: *that, is, will, be, are, it, have, not, their, should, also*. З результатів пошуку ключових слів видно, що при знаходженні однакою кількістю ключових слів власна розробка має 2 стоп-слова, а частотний словник 11. Однаковими ключовими словами є: we, I, work, need, system.

Для тексту «Address by President of the Russian Federation 2014» [14] перші 10 ключових слів за власною розробкою, разом із стоп-словами: right, people, Russia, have, be, work, I, do, provide, create, support, make, this, like. Де *have, be, do, this* — стоп-слова. Перші 10 ключових слів із стоп-словами, для частотного словника: *that, we, will, I, be, is, Russia, our, are, have, it, should, not, all, people, must, has, their, was, also, its, they, who, Russian, them, work, national, can, what, course*. Стоп-слова для частотного словника: *that, will, be, is, our, are, have, it, should, not, has, their, was, also, its, they, who, them, can, what*. З результатів видно, що при знаходженні однакою кількістю ключових слів власна розробка має 4 стоп-слова, а частотний словник 20. Однаковими ключовими словами є: I, Russia, people, Russian, work.

6. Обговорення результатів визначення ключових слів англomовного тексту

Як видно з результатів експерименту (табл. 1) розроблений метод знаходить більше ключових слів заданих автором, ніж аналоги. Окрім цього, запропонований метод без жодних додаткових фільтрів не менш ніж у 5 разів зменшує кількість стоп-слів серед першої десятки значимих (ключових) слів. Отримані результати можна використати для підвищення точності аналізу контенту сайту і підняття позиції сайту в результатах пошуку.

Якість отриманих результатів потенційно можна підвищити через окремий аналіз частин мови, оскільки ймовірність релевантності ключового слова, наприклад, іменника і прислівника буде відрізнятися. Окрім цього, варто оцінити збільшення частотних показників для ключових слів шляхом реалізації наявних в DKPro Core компонентів для визначення кореференційних зв'язків, а також визначити кількісні характеристики отриманих результатів.

Цікавим для семантичного аналізу результатом формального дослідження є власне склад перших п'ятирок ключових слів у зверненнях президента РФ до власного народу — у 2013 році йшлося про we, I, work, need, system; у 2014 році вже чуємо — I, Russia / Russian, people, work.

7. Висновки

Оскільки краща якість обробки тексту досягається лінгвістичними методами або ж при їх комбінації зі статистичними, систему автоматичного визначення ключових фраз з тексту природною мовою варто розробляти з використанням морфологічного словника (лексикону) і синтаксичних правил. Ці дані визначаються попередньо і зберігаються в базі даних. Текст підлягає

обробці аналізатором, який виробляє інформацію про розділення тексту на абзаци, речення та окремі слова, що необхідно для подальшого оброблення. Кожне слово, виділене аналізатором, піддається морфологічному аналізу з метою побудови морфологічної інтерпретації, визначення основи слова і формування леми. На основі наявної інтерпретації тексту виконується побудова та наповнення синтаксичних груп і виявлення відношень між ними.

В статті запропоновано метод визначення ключових слів, що базується на використанні додаткової інформації про складні залежності між членами англомовного речення. Для функціональної реалізації аналізатора тексту обрано популярний лінгвістичний пакет DKPro Core. Проведені експериментальні дослідження теоретичного обґрунтування методу підтвердили його якісні переваги у порівнянні з відомими аналогами.

Література

1. Ершов, Ю. С. Выделение ключевых слов в русскоязычных текстах [Текст] / Ю. С. Ершов // Молодежный научно-технический вестник. — М.: ФГБОУ ВПО «МГТУ им. Н. Э. Баумана», 2014. — Режим доступа: \www/URL: http://sntbul.bmstu.ru/file/out/730754. — 21.01.2015.
2. Андреев, А. М. Модели и методы автоматической классификации текстовых документов [Электронный ресурс] / А. М. Андреев, Д. В. Березкин, В. В. Сюзов, В. И. Шабанов // Вестник МГТУ им. Н. Э. Баумана. Сер. Приборостроение. — 2003. — № 4. — Режим доступа: \www/URL: http://vestnikprib.bmstu.ru/articles/397/html/files/assets/basic-html/page1.html. — 21.01.2015.
3. Joachims, T. Text categorization with Support Vector Machines: Learning with many relevant features [Text] / T. Joachims // Machine Learning: ECML-98 Lecture Notes in Computer Science. — 1998. — Vol. 1398. — P. 137–142. doi:10.1007/bfb0026683
4. Jensen, R. A Rough Set-Aided System for Sorting WWW Bookmarks [Electronic resource] / R. Jensen. — The University of Edinburgh, 2000. — Available at: \www/URL: http://users.aber.ac.uk/rkj/research/mscthis.pdf. — 21.01.2015.
5. Larkey, L. S. Combining classifiers in text categorization [Text] / L. S. Larkey, W. B. Croft // Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval — SIGIR '96. — ACM Press, 1996. — P. 289–297. doi:10.1145/243199.243276
6. Scott, S. Text Classification Using WordNet Hypernyms [Electronic resource] / S. Scott, S. Matwin. — University of Ottawa, 1998. — Available at: \www/URL: http://www.aclweb.org/anthology/W98-0706. — 21.01.2015.
7. Даркулова, К. Н. Необходимость выделения ключевых слов для свёртывания текста [Электронный ресурс] / К. Н. Даркулова, Г. Ергешова // VI Международная студенческая электронная научная конференция «Студенческий научный форум» 15 февраля — 31 марта 2014 года. Лингвистический анализ научного текста. — Южно-Казахстанский государственный университет им. Мухтара Ауэзова Шымкент, 2014. — Режим доступа: \www/URL: http://www.scienceforum.ru/2014/476/70. — 21.01.2015.
8. Бісикало, О. В. Концептуальна модель системи образного аналізу і синтезу природно-мовних конструкцій [Текст] / О. В. Бісикало // Математичні машини і системи. — 2013. — № 2. — С. 184–187. — ISSN 1028-9763.
9. Бісикало, О. В. Формальні методи образного аналізу та синтезу природно-мовних конструкцій [Текст]: монографія / О. В. Бісикало. — Вінниця: ВНТУ, 2013. — 316 с. — ISBN 978-966-641-528-1.
10. Natural Language Processing: Integration of Automatic and Manual Analysis [Electronic resource]. — Technischen Universität Darmstadt, 2014. — Available at: \www/URL: http://tuprints.ulb.tu-darmstadt.de/4151/1/rec-thesis-final.pdf. — 21.01.2015.
11. Gurevych, I. Darmstadt Knowledge Processing Repository Based on UIMA [Electronic resource] / I. Gurevych, M. Muhlhauser, Ch. Muller, J. Steimle, M. Weimer, T. Zesch. — February 9, 2007. — Available at: \www/URL: https://www.ukp.tu-darmstadt.de/fileadmin/user_upload/Group_UKP/publikationen/2007/gldv-uima-ukp.pdf. — 21.01.2015.
12. Burgareli, L. A. Variability management in software product lines using adaptive object and reflection [Electronic resource]: Thesis Abstracts / L. A. Burgareli // Journal of Aerospace Technology and Management. — Jul.-Dec. 2009. — V. 1, № 2. — Available at: \www/URL: http://www.jatm.com.br/papers/vol1_n2/JATMv1n2_thesis_abstracts.pdf
13. Address by President of the Russian Federation [Electronic resource]. — Available at: \www/URL: http://eng.kremlin.ru/transcripts/6402. — 21.01.2015.
14. Address by President of the Russian Federation [Electronic resource]. — Available at: \www/URL: http://eng.kremlin.ru/news/6889. — 21.01.2015.

МЕТОД ОПРЕДЕЛЕНИЯ КЛЮЧЕВЫХ СЛОВ В АНГЛОЯЗЫЧНЫХ ТЕКСТАХ НА ОСНОВЕ DKPRO CORE

Рассмотрены подходы к поиску ключевых слов текста, играющих важную роль в задачах компьютерной лингвистики. Предложен новый метод определения ключевых слов, основанный на нахождении связей между словоформами англоязычного текста с помощью инструментальных возможностей пакета DKPro Core. Метод, проиллюстрированный анализом примеров применения, направлен на решение задач эффективной обработки текстовых документов.

Ключевые слова: метод, ключевые слова, английский язык, лингвистический пакет, DKPro Core, синтаксический анализ.

Бісикало Олег Володимирович, доктор технічних наук, професор, директор ІнАЕКСУ, кафедра автоматики та інформаційно-виміральної техніки, Вінницький національний технічний університет, Україна, e-mail: obisikalo@gmail.com.

Яхимович Олександр Вікторович, кафедра автоматики та інформаційно-виміральної техніки, Вінницький національний технічний університет, Україна, e-mail: yahimovich.olexandr@gmail.com.

Бисикало Олег Владимирович, доктор технических наук, профессор, директор ИнАЕКСУ, кафедра автоматики и информационно-измерительной техники, Винницкий национальный технический университет, Украина.

Яхимович Александр Викторович, кафедра автоматики и информационно-измерительной техники, Винницкий национальный технический университет, Украина.

Bisikalo Oleg, Vinnytsia National Technical University, Ukraine, e-mail: obisikalo@gmail.com.

Yahimovich Alexander, Vinnytsia National Technical University, Ukraine, e-mail: yahimovich.olexandr@gmail.com