**References**

1. Marks, E., Lozano, B. (2010). A practical guide to cloud computing. John Wiley & Sons.

2. Roebuck, K. (2011). Mobile Application Development. Lightning Source Inc.

3. Rhoton, J. (2010). Application of cloud computing in the business processes of the enterprise.Recursive Press.

4. Herbert, L. (2013). The Forrester Wave: Enterprise Mobility Services, Q1 2013. Forrester. Available at: http://www.forrester.com/pimages/rws/reprints/document/87581/oid/1LT

5. Rittinghouse, J., Ransome, J. (2010). Cloud computing architecture. CRC Press.

6. Velte, A., Velte, T., Elsenpeter, R. (2009). Cloud Computing: A Practical Approach. McGraw Hill, 334. Available at: http://ftp.sustech.edu/cloud/Toby_Velte,_Anthony_ Velte,_Robert_Elsenpeter_Cloud_Computing,_A_Practical_ Approach__2009.pdf

7. Rittinghouse, J., Ransome, J. (2010). Cloud computing: management and safety, etc. CRC Press.

*Рекомендовано до публікації д-р техн. наук, професор Шабанов С. Ю.*
*Дата надходження рукопису 17.11.2015*

**Дмитренко Андрей Викторович,** кафедра программной инженерии, Харьковский национальный университет радиоэлектроники, пр. Ленина, 14, г. Харьков, Украина, 61166
E-mail: grenka487@gmail.com

**УДК 519.87:651.4/.9**
**DOI: 10.15587/2313-8416.2015.56345**

# МАТЕМАТИЧНА МОДЕЛЬ ПОШУКОВОГО ОБРАЗУ ДОКУМЕНТУ В СИСТЕМІ УПРАВЛІНСЬКОГО ДОКУМЕНТООБІГУ

**© В. І. Кунченко-Харченко**

*Було проаналізовано особливості формування інформаційного потоку, логічна структура мовного представлення розширеної семантичної мережі, побудовано акцептор кінцевого автомату для системи документообігу, що сприймає нескінченний алфавіт. Модель інформаційних потоків в системі прийнята за аналогією до моделі Бартона-Кеблера та врахувує статичну і динамічну складові від загальних обсягів повідомлень в системі. В моделі враховано, що для забезпечення релевантності пошуку індексаторам рекомендується брати не більше трьох словоформ вхідного алфавіту*
***Ключові слова:*** *інформаційний потік, акцептор, пошуковий образ документа, база знань, індексування документів*

*It was analyzed the spatiality of forming the information flow, logical structures of layout of the extended semantic network, built the acceptor of the terminal automat for document management system. Ones accept the infinity alphabet. The model of information flow in the system was accepted similarly to the Burton-Kebler model. Ones consists of static and dynamic parts of general massages' volume of the system. The model takes into account that to ensure the relevance of the search isn't recommended to take more than three word forms of the input alphabet for indexers*
***Keywords:*** *information flow, accepter, search document image, knowledge base, document indexing*

## 1. Вступ

Відомо, що основні принципи пошуку були сформульовані ще в першій половині XX століття. Як правило пошуку документів Процес створення вказівників на документи називається індексуванням, а терміни, що використовуються для індексування, називаються термінами індексування. Масив вказівників, отриманий після індексації інформаційних ресурсів називається індексом (Index database). Пошук стає більш ефективним, якщо відбувається за словником шуканих термів. Інформаційно-пошукові мови – це основна частина інформаційно-пошукової системи. Тому, ІПМ-основна частина і від неї залежить якість всієї системи. До складу ІПМ входить:

1) словник індексованих термінів є множиною термінів індексування;

2) кодовий словник – множина кодових термінів;

3) словник входів – множина вхідних термінів;

4) допоміжні засоби мови індексування – засоби, що використовуються разом з індексацій ними термінами для розширення або звуження визначених понять;

5) правила використання мови індексування.

## 2. Огляд літературних джерел по темі статті

У. Е. Батеном була розроблена система для пошуку патентів [1]. Запропонована ним система класифікувала документи у відповідності з поняттями, до яких він мав відношення. В даній моделі для пошуку визначеного документу, якщо в ньому розглядалося декілька понять (аналог сучасної бази знань [2]) необхідно було сумістити карти, що відповідають даним поняттям. Номер необхідного патенту визначається з позиції проміжку. З того часу основні принципи інформаційного пошуку не змінилися [3–5], пошук відбувається не по текс-

ту документів, а по їх пошуковим образам, які створюються на інформаційно-пошукових мовах (ІПМ) [6].

**3. Формулювання проблеми**

Для підвищення ефективності пошуку словник, що використовується системою, має бути контрольованим. Це означає, що його організація повинна бути такою, що повнота і точність пошуку була оптимальною. Для цього необхідно розробити модель пошукового образу документу, що відповідає вищевикладеним вимогам.

**4. Моделювання пошукового образу документу**

Приклад системи інформаційно-пошукової системи, що використовує предметне індексування, наведена в [1].

Морфологічний аналіз має за мету приведення слів до канонічних форм та формування алфавітних аксіом. Кожному слову надаються ознаки, які ділять-ся на три групи:

1) лексичні (слово з великої букви, великої букви, з точкою на кінці або це окрема буква та ін.);

2) морфологічні (граматична категорія слова, рід, число, відмінок для іменника);

3) семантичні (прізвище, ім'я та по-батькові). Слово в нормальній формі також вважається ознакою.

Результатом роботи блоку морфологічного аналізу є семантична мережа (СМ), що являє собою ієрархічну просторову структуру тексту документу, яка виключає інформаційні компоненти лінгвістичної структури мовного повідомлення як переносника даних та знань (рис. 1).

Результатом роботи блоку морфологічного аналізу є СМ, що уособлює просторову структуру тексту. В ній наведені слова в нормальній формі з їх ознаками та вказуванням їх послідовності. Наступна обробка зводиться до перетворення мереж на основі заданих правил. Для виділення необхідного інформаційного векторного потоку достатньо знайти в СМ для лінгвістичного процесора (ЛП) цифрові результати, що відповідають шуканій алфавітній аксіомі, передати їх експертній системі для з'ясовування рівня пошукового шуму, знаходження рівня оптимальності критеріям релевантності. Цьому передує термінологічний аналіз.

Для термінів може бути заданий допустимий контекст слова або їх ознаки, що стоять ліворуч або праворуч. Може бути також недопустимий контент-слова або їх ознаки, яких не повинно бути праворуч або ліворуч. Блок синтаксично – семантичного аналізу виконує такі функції:

1) за ознаками і контекстом виділяють значимі об'єкти (ПІБ людей, організації та ін.);

2) для кожного виявленого значущого об'єкту знаходить в документі зв'язану інформацію.

Контексті правила необхідні для подальшого синтаксично-семантичного аналізу, а синтаксично-семантичний аналіз необхідний для виділення адрес, номерів справ, організацій, термінів, грифів тощо. Такі правила виділяють з тексту групи слів за їх ознаками. По мірі застосування таких правил будується СМ – змістовний портрет документу, який відображає структуру ситуаційного повідомлення про стан системи. Відповідно, логічна структура схеми мовного представлення розширеної СМ (рис. 2) відображається через: sub – синтаксичний предикат – елемент поверхневої структури, ядро семантичної конструкції. Цифрові індекси показують ступінь спорідненості зв'язків. Кожне контекстне правило це СМ. Всі лінгвістичні знання, що оброблюються ЛП записуються у вигляді СМ. Над ними працюють продукції програми, що застосовують ці правила і грають роль пустої лінгвістичної оболонки, що підтримують записи лінгвістичних знань СМ.

Для забезпечення акцептора кінцевого автомату (КА) інформаційним потоком (рис. 3) використовуємо вхідний нескінчений алфавіт $\Sigma$. Зміною стану акцептора КА ієрархічного КА переключається в стан відправки документа і переходить до початкового стану.

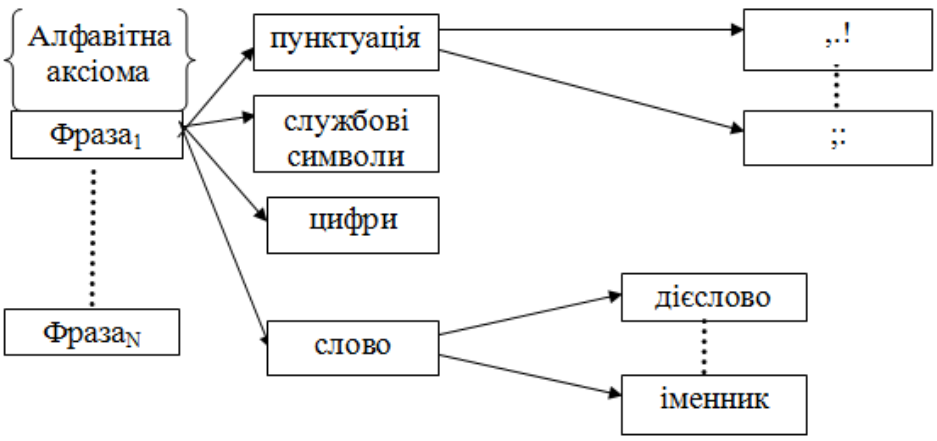


Рис. 1. Семантична мережа для ЛП

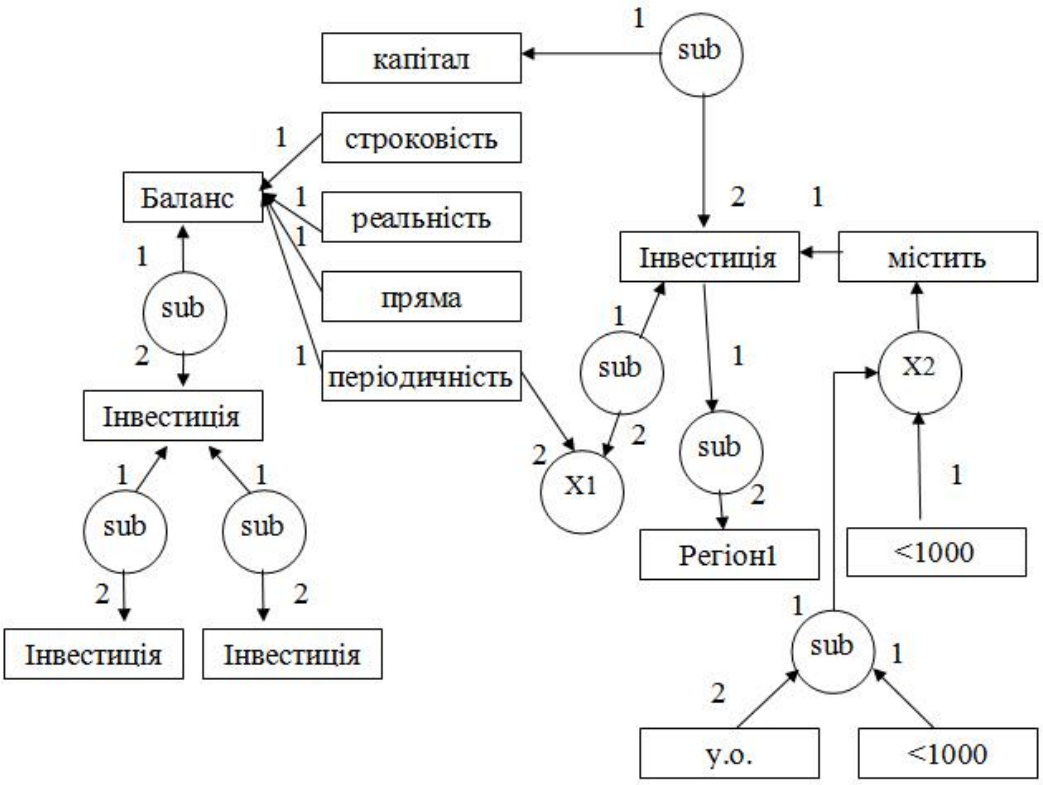


Рис. 2. Логічна структура мовного представлення розширеної СМ

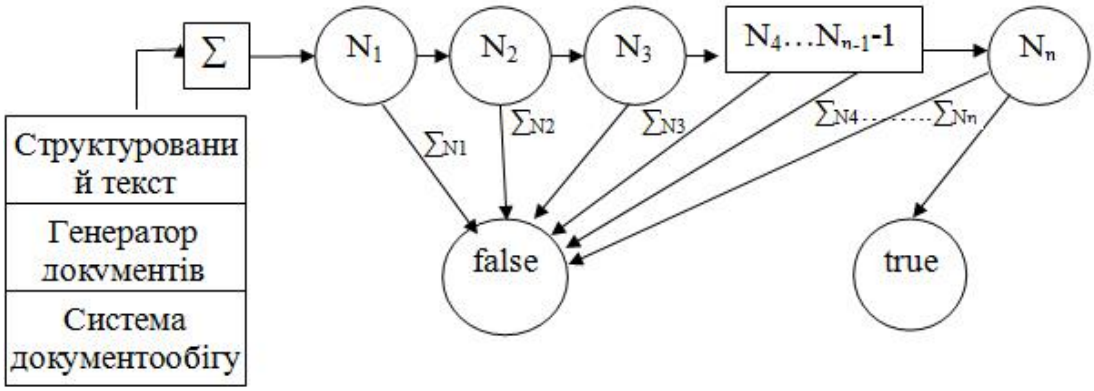


Рис. 3. Акцептор КА інформаційного потоку системи документообігу

Формуванню нескінченного алфавіту повинен передувати пошук та структуризація вхідного інформаційного вектору у відповідності до напряму роботи прогноз-моделі в структурі інформаційних технологій (ІТ). Для виявлення найоптимальнішого пошукового механізму в структурі ІТ прогнозування інтегрованих, соціально-виробничих циклів розвитку слід чітко сформулювати характеристики пошукової технології, а саме:

1) релевантність;
2) коефіцієнт помилкової видачі;
3) коефіцієнт мовчання;
4) коефіцієнт повноти пошуку;
5) коефіцієнт точності пошуку;
6) коефіцієнт шуму;
7) критерій видачі;
8) критерій відповідності сенсу;
9) пертинентність;
10) пошуковий шум;
11) релевантний документ;
12) змістовна релевантність;
13) стійкість пошуку;
14) формальна релевантність.

Використовуючи модель інформаційних потоків, аналогічну моделі Бартона-Кеблера, можна врахувати статичну і динамічну складові від загальних обсягів повідомлень по заданій тематиці з урахуванням старіння інформації: $v(T) = 1 - ae^{-T} - be^{-2T}$, де $ae^{-T}$ – стабільні інформаційні ресурси; $be^{-2T}$ – новинні ресурси.

Організації-генератори інформації виробляють потік інформації, в середньому постійний за кількіс-

тю повідомлень. Змінюються в часі лише обсяги повідомлень, які відповідають тій чи іншій темі. Таким чином, зростання кількості публікацій по одній тематиці супроводжується зменшенням публікацій по інших тематиках:

$$\int_0^T \sum_{i=1}^{M} n_i(t)dt = NT ,$$

де $n_i(t)$ – кількість публікацій за одиницю часу, а $M$ – загальна кількість всіх можливих тем.

Існують випадки, коли динаміка тематичних інформаційних потоків реалізується лінійно, тобто кількість тематичних інформаційних потоків в момент часу $t$: $u_n(x) = x^n$.

Для здійснення вибору оптимальної математичної моделі пошуку статистичних даних та технологій оптимального пошуку вхідного алфавіту КА, що утворюють вхідний інформаційний вектор та мають оптимальні характеристики алгоритму їх роботи необхідно врахувати вплив на впорядковування результатів пошуку назв та коди реквізитів документів відповідно до УСОРД [7].

Для отримання релевантного документу необхідно формувати пошуковий образ документу (ПОД) за оптимальним для розв'язку задачі алгоритмом. Перетворенню документу в його ПОД передує процес індексування інформації.

Як було встановлено в проведених автором дослідженнях, індексаторам рекомендується брати не більше трьох словоформ для створення одного ключового слова, а для ІС управлінського документообігу середня глибина індексування становить п'ятьдесять ключових слів на документ. Деякі архівні установи та бібліотеки формують до тридцяти ключових слів. Індексування координатного типу багатогалузевого документа формує один пошуковий образ, використовуючи багатоаспектне індексування. Потенційний відвідувач електронного каталогу (ЕК) архівної установи у вигляді експертної системи повинен бути готовий використовувати ключові слова при формуванні пошукового образу запита (ПОЗ).

Система комплексного індексування документів за умов існування архіву підприємства чи її організації в електронному чи паперовому виді дає можливість проводити полі індексування документів. Координатне індексування можна вважати базовим засобом індексування та пошуку документів в електронному середовищі. Координатне індексування вимагає побудови просторової СМ, як основи пошукової моделі для ІС управління.

Для оцінки пошуковими роботами значення фрагментів тексту, які знайдені ними в електронних документах, що подані в електронних форматах і для створення концепції побудови веб-додатків шляхом змішування функціональності різних програмних інтерфейсів та джерел даних необхідне застосування спеціальної технології типу Mashup, яка представляє собою web-сайт, що об'єднує дані з кількох джерел в одному сайті [8].

Для реалізації принципу релевантності пошуку для акцептора застосовується автомат зі стеком зберігання, який можна представити у вигляді кортежу: $M = (K, \Sigma, \pi, s, F, S, m)$, де $K$ – скінченна множина станів автомата; $s \in K$ - єдиний допустимий початковий стан автомата; $F \subset K$ - множина кінцевих станів, причому допускається $F = \emptyset$ і $F = K$; $\Sigma$ – скінченна множина символів вхідного алфавіту, з якого формуються строки, що зчитуються автоматом; $S$ – алфавіт пам'яті (магазин); $m \in S$ – нульовий символ пам'яті.

Для підвищення точності пошуку результатів до пошукової системи слід ввести профіль пошуку, що реалізується блоком конкретизатора (недолік – з часом, система може не отримати об'єктивний результат пошуку). Задачею, що пов'язана з реалізацією запропонованої експертної системи в складі інформаційної системи прийняття рішень є зв'язані функції представлення та оцінки знань-елементів часового ряду, що повинні бути занесені до бази знань автомату.

Побудована множина ключових слів $K$ ($d_p$), що характеризують тематику документа $d_p$ щодо розглянутої колекції, використовується для обчислення відносної оцінки ступеня тематичної близькості $tsim(d_p, d)$ між $d_p$ та іншими документами $d$. Нехай $K_q(d)$ позначає множину ключових слів документа $d$, які зустрічаються в його параграфі $Q_d$. Нехай множина виду: $E(d) = \left\{ \left\{ d_x, d_y \right\} : \left\{ d_x, d_y \right\} \in K_q(d), q \in Q(d) \right\}$ позначає множину всіх пар ключових слів документа $d$, які хоча б раз спільно зустрічаються в одному параграфі документа $d$, а множина:

$$E(d \left| d' \right.) = \left\{ \left\{ d_x, d_y \right\} : \left\{ d_x, d_y \right\} \in K_q(d), d_x, d_y \in Q(d') \right\}$$

є звуження $E(d)$ на множину пар ключових слів документа $d'$. Оцінку тематичної близькості можна визначити як: $tsim(d_p, d) = \left| \frac{E(d \left| d_p \right.) \cap E(d_p)}{E(d \left| d_p \right.) \cap E(d_p)} \right|$. Така модель дозволяє більш ефективно структурувати і отримувати інформацію з ієрархічних БЗ.

Словник термів повнотекстового документа визначимо як:

$$D_{term} = \left\{ < idt_i, term_i, w_i, N_{cф_i} > \right\}, \; i = \overline{1, N_{terms}} ,$$

де $N_{terms}$ – загальне число термів предметної області; $idt_i$ – унікальний ідентифікатор терму; $term_i$ – текстовий вираз терма; $w_i$ – частота використання даного терму в предметній області (вага терма); $Nсф_i$ – кількість словоформ, відповідних даному терму. Словник словоформ визначимо наступним чином: $D_{сф} = \left\{ < ids_i, form_i, iterm_i > \right\}$, $i = \overline{1, N_{forms}}$, де $N_{forms}$ – загальне число словоформ предметної області; $ids_i$ – унікальний ідентифікатор словоформи; $form_i$ – текстовий вираз словоформи; $itermt_i$ – відповідний словоформі терм зі списку термів предметної області.

Розроблена фреймова модель шаблона повнотекстового документа описується кортежем $F_{шабл} = \left\{ I_d, l_F, text_f, F_{next}, F_{up}, Attr \right\}$, де $I_d$ – унікальний ідентифікатор фрейму; $l_F$ – вертикальний рівень фрейму; $text_f$ – текстовий вміст фрейма (список термів); $F_{next}$ – покажчик на фрейм того ж рівня або по-

рожній покажчик; $F_{up}$ – покажчик на фрейм більш низького рівня або порожній покажчик; *Attr* – покажчик на додаткові атрибути або «нуль» у разі їх відсутності.

Розробка моделі ПОД базується на застосуванні апарату СМ, що дозволяє ефективно описувати семантику документів (рис. 3). ПОД представляється у вигляді неорієнтованого нечіткого графа другого роду: $\overline{G}_{\text{ПОД}} = \left(\overline{G}_{верш}, \overline{G}_{ребр}\right)$, де $\overline{G}_{верш}$ – нечітка множина вершин; $\overline{G}_{ребр}$ – нечітка множина ребер; відповідних відношенню «асоціативної зв'язності» термів документа. Елементи множини відповідають термам, що містяться в документі, а модель побудови ПОД раціонально розбити на 2 незалежні гілки: виділення термів документа з обчисленням їх ваг і знаходження ваг зв'язків між термами. З урахуванням запропонованої додаткової характеристики розподілу терму вага *j*-го терму в *i*-му документі БЗ буде визначатися наступним виразом: $w_{ij} = \dfrac{Tf_{ij}}{Df_j(\overline{Tf}_{ij}+1)}$, $i = \overline{1,..,N_D}$, $j = \overline{1,..,N_{terms}}$, де $Tf_{ii}$ – частота появи *j*-го терму в *i*-му документі; $Df_j$ – частка (частота) документів, що містять *j*-й терм; $\overline{TF}_{ij}$ – середня частота *j*-го терму у всіх документах набору крім *i*-го документа. $N_D$ – загальна кількість документів в БЗ; $N_{terms}$ – загальна кількість термів в БЗ.

Для перевірки ефективності функціонування розроблених моделей і алгоритмів, а також аналізу одержуваних з використанням запропонованих методів пошуку результатів і порівняння цих результатів з аналогами, було проведено імітаційне моделювання ІПС (рис.3). А для підвищення інформативності пошуку і звуження кількості ключових слів була запропонована формула для обчислення граничного значення ε кількості термів $w_i$.

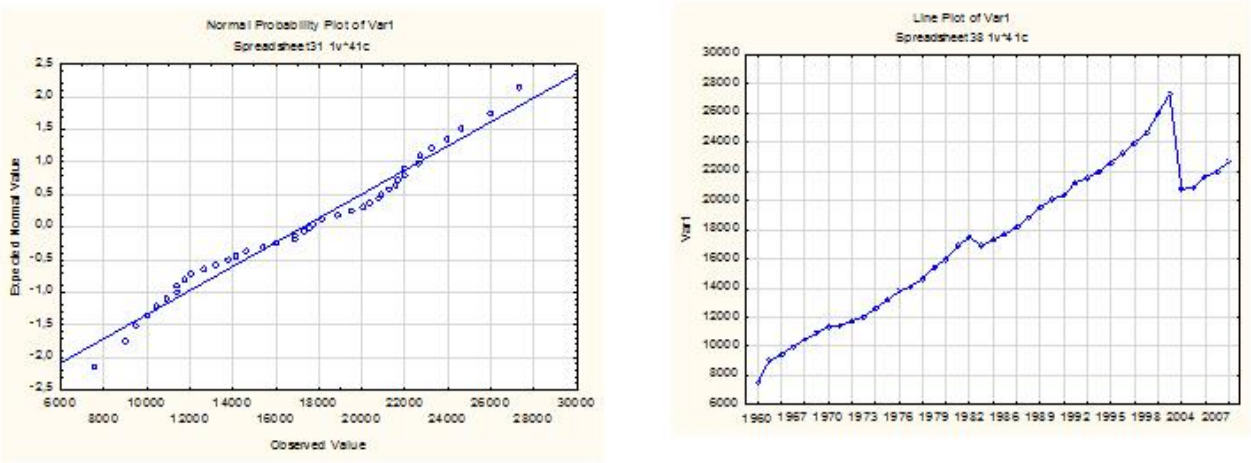


Рис. 3. Графічні залежності відхилень розподілу ЧР

Як показує залежність, наведена на рис. 3, застосування обчисленого граничного значення ε кількості термів $w_i$ значно зменшує число ключових слів, без зниження ефективності індексації документа. Відповідну графічну залежність виду $Y = f(v)$ типу «Normal Probability Plot» та таку саму залежність, але виду «Line Plot» побудовано пакетом Statistica.

**5. Висновки**

Для обчислення залежностей відхилення розподілу пошукового алфавіту оцінена ймовірність успішного завершення пошуку для можливих пар величин та ймовірність присутності сигналу на поточному та попередньому кроках. По мірі зростання кількості модулів, перевага традиційної ІПС практично нівелюється, складаючи близько 35 %, в той час як запропонована – забезпечує підвищення коефіцієнта повноти пошуку до 1, коефіцієнта точності в $\Delta$ раз, за умови, що $\Delta$ – шукана алфавітна аксіома. При цьому зниження швидкодії складає приблизно 17,3 %.

Проблемним аспектом для подальшого розвитку математичної моделі пошукового образу документу системи управлінського документообігу є забезпечення акцепторів вхідного інформаційного потоку актуальною інформацією для побудови статистичних часових рядів вимагає розв'язку проблеми каталогізування. Розв'язок проблеми пов'язаний з рішенням проблем опису електронних ресурсів.

**Література**

1. Информационные базы данных и электронные библиотеки [Электронный ресурс]. – Режим доступа: http://bourabai.kz/einf/chapter121.htm

2. Knowledge base definition [Electronic resource]. – TechTarget. – Available at: http://searchcrm.techtarget.com/definition/knowledge-base

3. Kuhlthau, C. C. Information Search Process [Text] / C. C. Kuhlthau. – New Jersey. – Available at: https://comminfo.rutgers.edu/~kuhlthau/information_search_process.htm

4. Information search and decision making [Electronic resource]. – USC Marshall. – Available at: http://www.consumerpsychologist.com/cb_Decision_Making.html

5. Information Search [Electronic resource]. – Available at: https://www.boundless.com/marketing/textbooks/ boundless-marketing-textbook/consumer-marketing-4/the-consumer-decision-process-40/information-search-201-4089/

6. Searching with Words, Phrases, or Plain Language [Electronic resource]. – Available at: http://dl.acm.org/documentation/Types.htm

7. ДСТУ 4163-2003. Державна уніфікована система документації. Уніфікована система організаційно-розпорядчої документації. Вимоги до оформлювання документів [Електронний ресурс]. – Національний стандарт України. – Режим доступу: http://metrology.com.ua/download/dstu-gost-gost-r/60-dstu/191-dstu-4163-2003

8. Joa, M. H. WEB 2.0 Based satellite images search through mash-up [Text]: conference / M. H. Joa, Y. W. Joa, J. S. Kim. – ISPRS, 2008. – P. 861–863. – Available at: http://www.isprs.org/proceedings/XXXVII/congress/4_pdf/153.pdf

**References**

1. Informacionnye bazy dannyh i jelektronnye biblioteki. Available at: http://bourabai.kz/einf/chapter121.htm

2. Knowledge base definition. TechTarget. Available at: http://searchcrm.techtarget.com/definition/knowledge-base

3. Kuhlthau, C. C. Information Search Process. New Jersey. Available at: https://comminfo.rutgers.edu/~kuhlthau/information_search_process.htm

4. Information search and decision making. USC Marshall. Available at: http://www.consumerpsychologist.com/cb_Decision_Making.html

5. Information Search. Available at: https://www.boundless.com/marketing/textbooks/boundless-marketing-textbook/consumer-marketing-4/the-consumer-decision-process-40/information-search-201-4089/

6. Searching with Words, Phrases, or Plain Language. Available at: http://dl.acm.org/documentation/Types.htm

7. DSTU 4163-2003. Derzhavna unifikovana sistema dokumentaciy. Unifikovana sistema organizacijno-rozporjadchoy dokumentaciy. Vimogi do oformljuvannja dokumentiv. Nacional'nij standart Ukrainy. Available at: http://metrology.com.ua/download/dstu-gost-gost-r/60-dstu/191-dstu-4163-2003

8. Joa, M. H., Joa, Y. W., Kim, J. S. (2008). WEB 2.0 Based satellite images search through mash-up. ISPRS, 861–863. Available at: http://www.isprs.org/proceedings/XXXVII/congress/4_pdf/153.pdf

*Дата надходження рукопису 20.11.2015*

**Кунченко-Харченко Валентина Іванівна**, доктор технічних наук, професор, кафедра інформатики і інформаційної безпеки, Черкаський державний технологічний університет, бул. Шевченка, 460, м. Черкаси, Україна, 18006
E-mail: valentine.kun@ukr.net

**УДК 616.33/34-009.1-07.616-031.14-519.876.5**
**DOI: 10.15587/2313-8416.2015.55848**

# ОБНАРУЖЕНИЕ И НОРМИРОВАНИЕ ПАРАМЕТРИЧЕСКИХ ИЗМЕНЕНИЙ В СЛУЧАЙНЫХ СИГНАЛАХ БИОМЕДИЦИНСКОЙ ИНФОРМАЦИИ ПРИ ХИМИЧЕСКИХ И МЕХАНИЧЕСКИХ ВОЗДЕЙСТВИЯХ НА БИОЛОГИЧЕСКИЕ СТРУКТУРЫ

**© Р. В. Стецишин, Д. П. Замятин, О. Ю. Кропачек, Р. П. Мигущенко**

*Рассмотрена математическая модель процедуры скользящего дифференцирования спектрально-нестационарных случайных биомедицинских измерительных сигналов. Доказана возможность получения дополнительной диагностической информации для количественной оценки изменений электрической и механической активности биологических структур при физико-химических воздействиях на последние. Показана эффективность использования разработанной процедуры в задачах анализа активности химически индуцированных спонтанных сокращений мочеточника и идентификации кардиодинамических нарушений при закрытой травме груди, сопровождающейся ушибом сердца*

***Ключевые слова:*** *ушиб, мочеточник, грудная клетка, биомедицинские измерительные сигналы, математическая модель, статистика*

*It was described the mathematical model of procedures for moving spectrally differentiating non-stationary random biomedical measurement signals. It is proved a possibility of additional diagnostic information to quantify the changes in electrical and mechanical activity of biological structures at the physical and chemical effects. It is shown an efficiency of using procedures developed for the analysis of activity chemically induced spontaneous contractions of the ureter and identification cardiodynamic violations in closed chest trauma, accompanied by heart injury*

***Keywords:*** *injury, ureter, chest, biomedical measurement signals, mathematical model, statistics*

## 1. Введение

Задачи обнаружения изменений свойств случайных измерительных сигналов, используемых для контроля и диагностики динамических объектов с априори неопределенными свойствами, являются в метрологическом плане наиболее сложными, как в