

СИСТЕМА КОНТЕНТ-МОНІТОРИНГУ НОВИНИХ ІНТЕРНЕТ-РЕСУРСІВ

© Берко А.Ю., Кісь Я.П., Суховерський В.І., 2011

Проаналізовано основні методи та засоби системи контент-моніторингу. На основі цих методів і засобів виконано дослідження з метою виявлення найвживаніших категорій / слів у провідних періодичних виданнях України.

Ключові слова: контент, контент-аналіз, контент-моніторинг, таблиці прив'язки, таблиці взаємозв'язку, дайджести.

The article discusses the basic methods and content-monitoring system. The research, based on these methods and tools to identify the most frequently *zustrichayemyh* categories / keywords in leading journals press in Ukraine.

Key words: content, content analysis, content monitoring, reference tables, table relationships, digests.

Вступ

Однією з головних особливостей нашого часу є постійне зростання темпів виробництва інформації. Крім збільшення обсягів інформації до масштабів, які унеможливають її безпосередню обробку, виникла низка специфічних проблем, пов'язаних із швидким розвитком інформаційних технологій.

Саме інформація, яка вже сьогодні циркулює в мережі Інтернет, є тим невичерпним джерелом інформації та знань (цілком зрозуміло, що йдеться про інформацію, а не про “спам”), що по своїй суті, за наявності потужних пошукових систем, є найпривабливішим з погляду пошуку інформації, знань. Необхідно також враховувати, що вже давно поетапно вирішується питання щодо приведення до електронного вигляду всього того досвіду, який накопичило людство за часи свого існування у документальному вигляді.

Зважаючи на це, вже нині виникає цікава ситуація – з одного боку, є доволі потужний інформаційний масив (ресурси Інтернету) для прийняття рішень у різноманітних сферах життя держави, суспільства та окремої особи, а з іншого – брак необхідної для прийняття рішень інформації саме через її динаміку, обсяги, джерел та неструктурованість [4].

Охоплення та узагальнення великих динамічних інформаційних потоків, які безперервно генеруються в засобах масової інформації, вимагає якісно нових підходів.

Саме тому останнім часом використовують системи моніторингу ресурсів, тісно пов'язані з контент-аналізом. Цей перспективний напрям одержав назву контент-моніторинг. Під контент-моніторингом найчастіше розуміють змістовний аналіз інформаційних потоків з метою отримання необхідних якісних і кількісних зрізів, який ведеться постійно протягом не визначеного заздалегідь проміжку часу [3].

Аналіз останніх досліджень і публікацій

Контент – це набір інформації, об'єкт системи електронної контент-комерції, який неподільний у часі, є основним чинником функціонування системи електронної контент-комерції та існує лише в електронному вигляді.

Різновид контенту – це інформаційний товар, інформаційний вміст веб-сайта, інтернет-видавництва, маркетингових досліджень, консалтингових послуг, який є об'єктом бізнес-процесів систем електронної контент-комерції [1].

Контент-моніторинг – це змістовий аналіз інформаційних потоків з метою отримання необхідних якісних і кількісних зрізів, який, на відміну від контент-аналізу, здійснюється безперервно в часі.

Саме безперервна аналітична обробка повідомлень є найхарактернішою рисою цього підходу, який дає змогу видобути факти з текстів, виявляти нові поняття, формувати різноманітні статистичні звіти. Названі завдання сьогодні охоплюються двома основними технологіями – витягом фактографічної інформації з текстів (Information Extraction) та глибинним аналізом текстів (Text Mining).[2]

В автоматизованій технології контент-моніторингу є кілька важливих особливостей:

- використання ключового фрагмента публікації як одиниці формування текстового інформаційного масиву;
- формування банку ключових фрагментів публікацій є об'єднанням двох взаємозалежних автоматизованих процесів: аналітико-синтетичної переробки та багаторівневої процедури контент-аналізу текстів публікацій;
- індексація ключових фрагментів публікацій відбувається за допомогою багатофасетної класифікації.

Унікальність запропонованої технології полягає в об'єднанні змістових і кількісних методів контент-аналізу. Послідовність етапів змістового аналізу проблеми, яку досліджує конкретна інформаційна система, умовно можна поділити на змістовий (якісний) аналіз сукупності публікацій і формалізований (кількісний) аналіз інформаційних масивів: індексного, бібліографічного та масиву текстів ключових фрагментів публікацій [6].

Сучасний рівень контент-моніторингу охоплює також завдання виявлення взаємозв'язків категорій, що видобуваються з документів, групування цих категорій, візуалізації. У цьому випадку на допомогу приходять методи кластерного аналізу, що дають змогу на основі виявлення латентних ознак формувати компактні групи категорій, виявляти головні з них, візуалізувати взаємозв'язки. Названі завдання сьогодні частково вирішують провідні контент-провайдери у всьому світі. Так, в 2006 р. компанія «Яндекс» у межах свого новинного сервісу надала доступ до довідкової інформації про людей, згадуваних у ЗМІ, автоматично вилучаючи факти з текстів і групуючи їх у прес-портрети [2].

У компанії «Інтегрум-Техно» розроблена автоматично поповнювана база даних, що містить інформацію про людей і організації. Основний принцип виділення фактів полягає в такому: у реченні виділяються лексичні одиниці, які вказують на те, що в цьому місці може бути група «посади» або «компанії», потім навколо цих слів за допомогою граматик будуються певні іменні групи, в яких вершинами є знайдені слова.

Система контент-моніторингу забезпечує автоматизований збір інформації з web-сайтів у режимі реального часу, її структурування, групування за семантичними ознаками, а також тематичний виборчий розподіл та надання доступу до інформаційних баз даних у пошукових режимах. [2]

До теоретичних та методологічних передумов появи систем контент-моніторингу належать: розвиток теоретичних основ контент-аналізу, теорії "розкопок тексту" (Data Mining); методи математичної лінгвістики, теорія кластерного аналізу. Нині в інформаційному центрі " ELVISTI " на основі технології, що забезпечує одержання й обробку вхідного потоку інформації з web-сайтів Інтернету, розроблено систему контент-моніторингу InfoStream. Ця система призначена для вирішення завдань формування тематичних інформаційних каналів, дайджестів, таблиць взаємозв'язків категорій (щодо зустрічальності в мережевих публікаціях) і гістограм розподілу вагових значень окремих категорій. Розроблено основні компоненти технології, що забезпечують: автоматичне сканування доступних інформаційних ресурсів Інтернету; нормалізацію інформації, приведення її до єдиного формату, автоматичну класифікацію інформації, побудову тематичних інформаційних каналів; автоматичне створення оперативних і ретроспективних баз даних та доступ до цих баз даних у діалоговому режимі; автоматична побудова дайджестів, що відображають найвагоміші події; побудову таблиць взаємозв'язків і гістограм розподілу вагових значень категорій, зумовлених складними логічними виразами мови запитів [3].

Головні переваги системи InfoStream порівняно з традиційними мережевими інформаційно-пошуковими системами такі:

- Оперативність – бази даних системи поповнюються кожні 15 хвилин, джерела скануються у мережі в міру їх оновлення, в той час як період індексації традиційних інформаційно-пошукових систем може вимірюватися тижнями.
- Доступність ретроспективного фонду – навіть якщо інформація видалена з web-сайта джерела, вона збережена в інформаційному сховищі.
- Наявність аналітичного інструментарію – користувач може в режимі реального часу не тільки отримувати результати пошуку, але і формувати дайджести, будувати сюжетні ланцюжки, аналізувати взаємозв'язок рубрик, динаміку категорій тощо.
- Можливість селекції "дублів" – система здійснює автоматичне маркування ідентичної за змістом новинної інформації.

Порівняно зі звичайними новинними web-сайтами система InfoStream забезпечує такі переваги:

- Охоплення джерел – користувач має доступ до новин, що цікавлять його, одночасно з великої кількості web-сайтів, зокрема і тих вибраних, які він звик переглядати щодня.
- Принцип "одних рук" – користувач системи моніторингу має доступ до інформації з багатьох web-сайтів з одного інтуїтивного інтерфейсу.
- Пошукові можливості – новинні web-сайти, на відміну від системи InfoStream, не завжди мають розвинені пошукові можливості.
- Доступність ретроспективного фонду [7].

Останнім часом в Інтернеті з'явилися у вільному доступі й так звані інтегратори новин. Порівняно з цими службами система InfoStream володіє такими перевагами:

- Безперервно розвивається список джерел – система InfoStream сканує понад 4000 джерел – всі основні інформаційні сайти України та СНД, а також провідні закордонні інтернет-ресурси.
- Облік місцевих ресурсів – навіть найбільші платні зарубіжні інтегратори новин лише частково охоплюють, наприклад, українські web-ресурси. У систему InfoStream можливе долучення нових джерел за заявками користувачів.
- Розширений доступ до інформації – користувачам системи InfoStream доступні не тільки заголовки або анонси, але і повні тексти новинних повідомлень, посилання на подібні документи тощо.
- Аналітичний інструментарій.
- Розвинена служба підтримки абонентів – користувачі системи за необхідності можуть звертатися з питаннями до служби підтримки та отримувати вичерпні відповіді та консультації. [7]

Система контент-моніторингу ґрунтується на статистичних методах контент-аналізу, які останнім часом швидко розвиваються в усьому світі. Нині розроблена в ІЦ "ELVISTI" система контент-моніторингу забезпечує автоматизовану підготовку декількох видів звітів, до складу яких входять таблиці взаємозв'язків категорій і гістограми розподілу категорій. Ці звіти дають змогу аналізувати відображення процесів реального життя в кількісних зрізах, а також виявляти тенденції до взаємозв'язків окремих категорій [5].

Побудова таблиць взаємозв'язків і гістограм розподілу вагових значень категорій основана на мовних засобах інформаційно-пошукової системи (ІПС), а також методах статистики і кластерного аналізу. Таблиця взаємозв'язків категорій, яка будується як статистичний звіт, що відображає близькість окремих категорій реального світу, – це симетрична матриця $A = (a_{ij})$, де a_{ij} – коефіцієнти взаємозв'язку пари категорій i та j [3].

Формулювання цілей статті

Цілі статті такі:

- 1) наочне відображення алгоритмів роботи контент-моніторингу в результаті проведення дослідження з метою виявлення найпоширеніших категорій / слів у провідних періодичних виданнях України.
- 2) аналіз методів та засобів контент-моніторингу.

Постановка проблеми

Основною проблемою, яку розглянуто в статті, є проблема різкого збільшення обсягів інформації через слабку її структурованість, появу паразитної інформації (одержуваної як додатки), багаторазове дублювання інформації. Внаслідок цього інформаційно-пошукові системи поступово почали втрачати актуальність. Причина цього криється не стільки у фізичних обсягах інформаційних потоків, скільки в їхній динаміці, тобто в постійному систематичному оновленні інформації, яка далеко не завжди має очевидну регулярність.

Спосіб вирішення проблеми – це систематичне відстеження тенденцій і процесів у постійно оновлюваному інформаційному середовищі.

Засіб вирішення проблеми – впровадження системи моніторингу контенту в електронній комерції для розв’язання задач збору, систематизації та узагальнення інформації, яке необхідне для прийняття відповідального рішення, потребує створення своєрідного інтелектуального посередника між корпоративною інформаційною системою та Інтернетом. Саме система контент-моніторингу, що є таким посередником, виконує основну “чорнову” роботу зі збору інформації з мережі Інтернет та забезпечує створення документального сховища, яке відповідає інформаційним потребам споживача. Система контент-моніторингу надає ефективний доступ до потоків електронної інформації з Інтернету, чим має покращувати результати роботи користувачів, забезпечуючи такі показники, як оперативність, повнота та відповідність тематичним запитам. Система контент-моніторингу забезпечує постійне поповнення сховища оперативними повідомленнями, ефективний одночасний доступ до своїх баз даних багатьом користувачам, зручні інструменти та засоби пошуку потрібних новин.

Система контент-моніторингу новинних інтернет-ресурсів

Система контент-моніторингу відповідає потребам надійного приймання інформації, що надходить в електронному вигляді, її структурування, класифікації та надання до неї доступу в корпоративній мережі у режимі повнотекстового пошуку. У нашому випадку система ґрунтується на технології InfoStream®, елементи якої, програмні компоненти, підсистеми та інформаційне наповнення можна адаптувати до визначеної користувачами проблематики. Система реалізує повномасштабне інформаційне сховище, що враховує особливості цієї проблематики, накопичує та надійно зберігає інформацію для використання в аналітичній роботі.

Система контент-моніторингу містить такі автоматизовані підсистеми (рис. 1) [4]:

- збору та обробки інформації з мережі Інтернет;
- забезпечення доступу до повнотекстових баз даних;
- аналізу та узагальнення інформації.

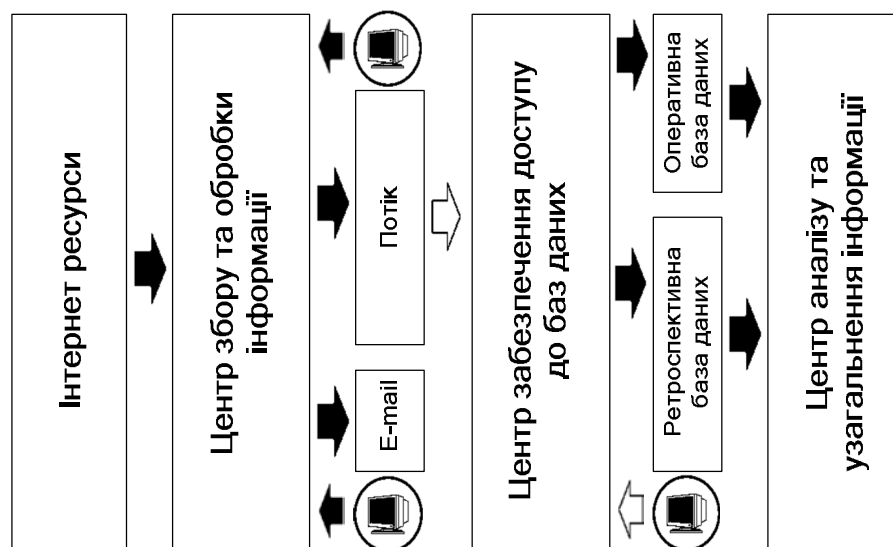


Рис. 1. Основні технологічні процеси системи контент-моніторингу

Відповідно до свого основного призначення перший центр забезпечує:

- збір інформації з різноманітних веб-сайтів та її форматування;
- виявлення ключових слів, категорій;
- автоматичну рубрикацію;
- виявлення змістового дублювання документів;
- вибіркове поширення інформації.

Головне завдання другого центру – формування баз даних та забезпечення доступу до неї користувачів, а саме:

- формування оперативних та ретроспективних баз даних;
- ротація баз даних;
- персоналізація роботи користувачів, збереження їхніх персональних запитів та джерел, ведення статистики роботи;
- забезпечення пошуку в базах даних;
- генерація вихідних форм;
- інформаційна взаємодія з базами даних інших підсистем.

Підсистема аналізу та узагальнення інформації забезпечує:

- формування інформаційних портретів;
- формування дайджестів;
- виявлення тематичних сюжетів;
- побудову таблиць взаємозв'язку категорій;
- розрахунок рейтингів категорій.

У цьому дослідженні розглянемо таблицю взаємозв'язків категорій та гістограми розподілу категорій, використовуючи категорії ціннісного словника Лассуелла. Ціннісний словник Лассуелла складається з восьми базових оціночних категорій:

1. Влада – можливість впливати на інших.
2. Справедливість – моральні цінності.
3. Повага – статус, слава, визнання, престиж.
4. Любов – любов і дружба.
5. Багатство – можливість володіння.
6. Благополуччя – здоров'я і безпека.
7. Просвіта – знання, розуміння і наявність інформації про міжособистісні або міжкультурні взаємини.

8. Кваліфікація – знання.

Проаналізуємо частоти зустрічальності цих категорій в текстах статей видань "Аргументи і факти", "Голос України", "Дзеркало тижня", щоб оцінити іміджеві характеристики й семантичну спрямованість тексту, а також здійснити мотиваційний аналіз.

Перш ніж почати побудову таблиць взаємозв'язків та гістограм розподілу категорій, необхідно сформулювати таблицю прив'язки ознак діяльності. Таблиця прив'язки сприймається як основа для пошуку, тобто складається зі слів або їхніх правильних скорочень і операторів. Наприклад, після введення слова «завод», можна знайти документи, що містять словоформи: "заводу", "заводити", "заводи" тощо. Ця таблиця прив'язки забезпечує додаткові можливості і оператори для пошуку, а саме:

- не відрізняє прописних і малих літер;
- для пошуку за словосполученнями використовується спеціальний оператор "~". Цей оператор забезпечує добір документів, до яких входять слова, зв'язані ним;
- можна задати відстані між словами в запиті: /0/ – сусідні слова (якщо іншого не вимагається), /1/ – не більше від 1 слова в тексті документа між словами з запиту; /2/ – не більше від двох слів і т.д.;

- у запитах можуть використовувати логічні операції: NOT – логічне НІ ("!" або "^"), AND – логічне І ("&" або "+"), OR – логічне АБО ("|" або ";").

Побудуємо таблицю прив'язки для нашого дослідження.

Таблиця 1

Таблиця прив'язки категорій до їхніх характеристик

Влада	(влада власть)&(викон исполн законодав законодат судов державн государств)
Справедливість	(справедлив)&(злочин преступл покаран наказан прац труд винагород вознаград)
Статус	(статус)&(добр хорош відмін отличн позитив положительн нормальн)
Слава	(слав популярн)&(позитив положит негативн отрицательн нормальн)
Визнання	(визнан признан)&(добр хорош відмін отличн позитив положительн нормальн)
Престиж	(престиж чарівн обаян очаров ілюз илюз)
Любов	(любов)&(платонічн платонич нерозділ неразделён безумовн безусловн)
Дружба	(дружб)&(безкорисл бескорыст шир искрен віддан предан)
Багатство	(багатств богатств)&(матеріал материальн духовн велике больш мал)
Здоров'я	(Здоров сил)&(добр хорош відмін отличн нормальн)
Безпека	(безпек безопасн)&(інформац информац гірнич гирнич судов державн государств)
Знання	(знан)&(наукв научн повсякден обыден інтуїтивн интуитивн релігійн религиозн)
Кваліфікація	(кваліфік квалифик)&(добр хорош відмін отличн вищ выш)

Розглянемо формальне визначення таблиці взаємозв'язків категорій TVP' , побудованої за допомогою першого алгоритму. Позначимо p_j – поняття ($j = 1, \dots, M$); D_i – документ ($i = 1, \dots, N$); e_{ij} – ознака відповідності поняття документам:

$$p_j \in D_i \Rightarrow e_{ji} = 1, \text{ інакше } e_{ji} = 0. \quad (1)$$

Можна визначити рівень зв'язку категорій p_j і p_k :

$$v'_{jk} = \sum_{i=1}^N e_{ji} e_{ki}. \quad (2)$$

Ввівши позначення: $E = \parallel e_{ij} \parallel j = 1, \dots, M; i = 1, \dots, N$, отримуємо:

$$TVP' = \parallel v'_{jk} \parallel j, k = 1, \dots, M. \quad (3)$$

Для відображення таблиці взаємозв'язків категорій окремі елементи матриці E подаються різними відтінками сірого кольору (залежно від значень коефіцієнтів взаємозв'язку e_{ij}), як це наведено на рис.2. Коефіцієнт e_{ij} відповідає кількості документів вхідного інформаційного потоку, що містить категорії (терміни або словосполучення, представлені мовою запитів, що відповідають категорії) i , а коефіцієнт e_{ij} – кількості документів у вхідному потоці, які одночасно відповідають категоріям i та j . При цьому для впорядкування категорій з метою виявлення блоків – множин найбільш взаємозалежних категорій – застосовується алгоритм кластерного аналізу [3].

Таблицю взаємозв'язків категорій, зображену на рис. 2, побудовано за допомогою 13 категорій. Принцип побудови таблиці такий: спочатку вибирають відповідні категорії, які визначаються як найтипівіші представники. Після цього кожній з наведених категорій присвоюється відповідний номер, за яким буде здійснюватися побудова таблиці. Потім визначаємо кількість документів для кожного з відповідних взаємозв'язків категорій. Після цього на основі отриманих результатів визначаємо відповідний контраст кольору залежно від співвідношення кількості документів між взаємозалежними категоріями та кількості документів категорії, яка є найтипівішим представником.

На рис.3 подано гістограму розподілу категорій, отриману на основі аналізу характеристик.

На цьому рисунку можна побачити категорії, які найчастіше вживають у текстах статей газет "Аргументи і факти", "Голос України", "Дзеркало тижня". Розподіл категорій ціннісного словника Лассуелла, розглянутого на рис. 3, свідчить про значне домінування категорій „Здоров'я”, „Влада”, „Любов”, „Безпека”, що дає нам змогу зробити певні висновки. Наприклад, те, що категорія

„Здоров’я” домінує у згаданих вище газетах, свідчить, що в нашій державі не все гаразд із медициною і тому необхідно приділяти більше уваги саме цій галузі.

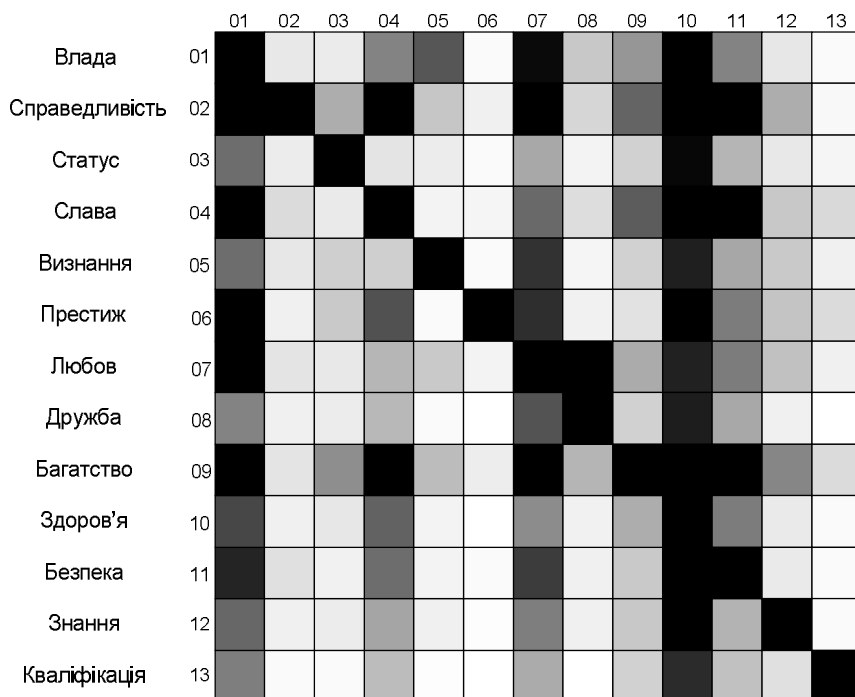


Рис. 2. Таблиця взаємозв’язків категорій

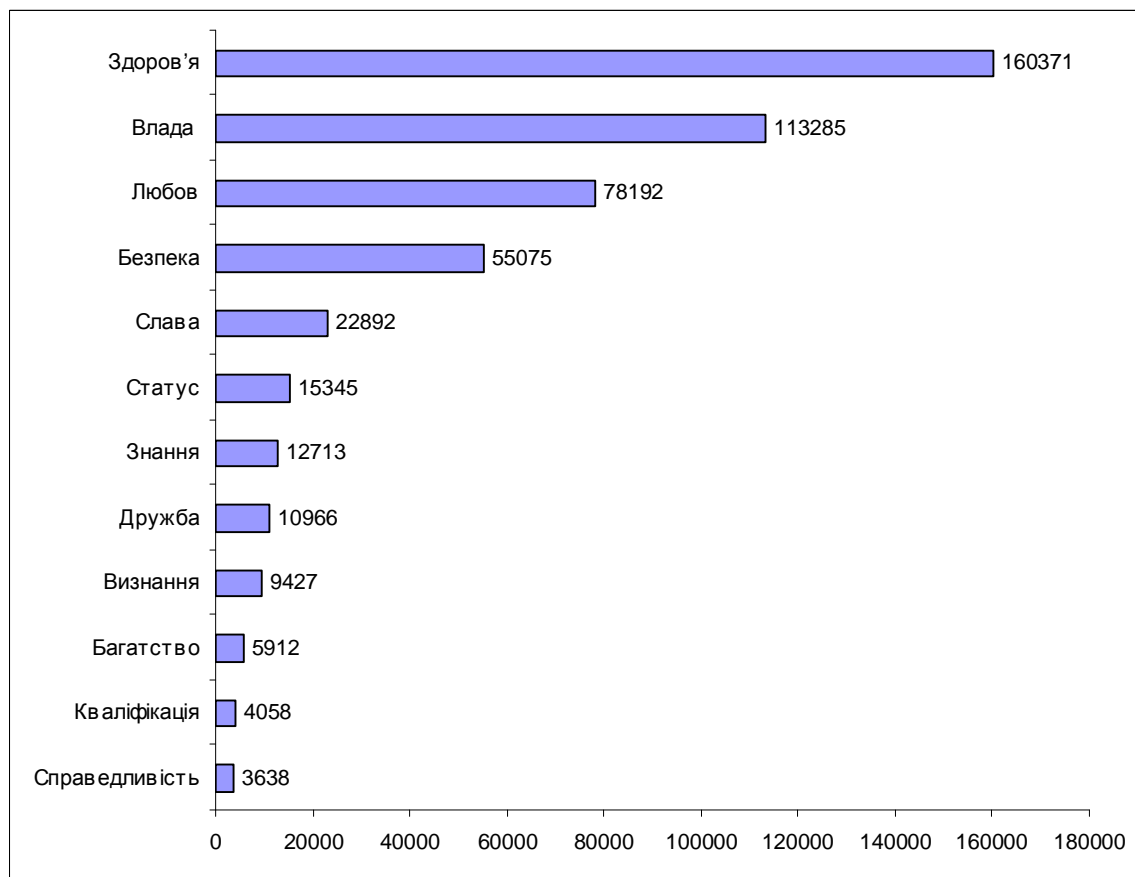


Рис. 3. Гістограма розподілу категорій

Важливо відзначити також те, що категорія "Влада" за кількістю зафіксованих випадків не надто поступається першій категорії, що означає пильну увагу преси України до проблем реалізації влади в Україні, а саме політичних конфліктів і домовленостей. Також важливе місце на цій гістограмі займають категорії „Любов” і „Безпека”, всі наступні категорії не є домінуючими у виданнях цих газет.

Ще одна базова процедура системи контент-моніторингу – це автоматичне формування дайджестів. Для дайджесту відбирають документи, в яких найбільш явно відбиті тенденції усього вхідного потоку. Такі дайджести, отже, повинні найбільшою мірою відповідати інформаційним потребам користувача, за запитом якого формується цей вхідний інформаційний потік. Для формування дайджестів використовують статистичні алгоритми, основані на «ваговому» частотному підході.

На першому етапі формування дайджесту на підставі слів, що входять у вхідний інформаційний потік, будують словник системи. Кожному зі слів приписується деякий ваговий коефіцієнт, який визначається як результат обліку декількох складових: частоти зустрічальності, тематичного словника, визначеного тематикою вхідного інформаційного потоку, «плюс-словника», що містить найважливішу лексику загального призначення. Для формування дайджестів спочатку виділяють всі слова вхідного потоку документів і з них будують послідовний словниковий масив. Потім виконується нормалізація слів за допомогою засобів автоматичного морфологічного аналізу. Після цього всі слова словникового масиву сортують і їм приписуються ваги – частоти зустрічальності [3].

Наступним кроком є видалення зі словникового масиву незначущих слів. Для цього використовуються спеціальні програмні засоби, основані на використанні "стоп-словника". Відповідно до критерію на формування вхідного потоку відбувається вибір тематичного словника, а відтак коригування словникового масиву з урахуванням тематичного словника – тобто коректування ваги окремих слів, що входять в тематичний словник. За таким самим алгоритмом словниковий масив коригують з урахуванням "плюс-словника".

Останній крок формування словника системи полягає у виборі N найвагомійших слів зі словникового масиву. Вибір вихідних документів для побудови дайджесту здійснюється також з урахуванням їх «терезів». Вага кожного документа визначається як нормована по довжині документа сума ваг окремих слів, що входять у цей документ. Етап вибору документів для дайджесту складається з таких кроків, як визначення ваги кожного документа, сортування вхідного потоку документів за вагами, визначення смислових дублів документів за статистичним критерієм, відкидання документів, непридатних для побудови дайджестів (за допомогою таблиць «стоп-адрес» і «недопустимих типів документів»), а також статистично визначених дублів. Останній крок етапу вибору документів для формування дайджесту полягає у виборі M найвагомійших документів з відсортованого і відфільтрованого на попередніх кроках масиву [3].

Висновок

Очевидно, треба визнати, що початкові парадигми пошукових систем і контент-аналізу, сформовані кілька років тому, вже не відповідають реальній ситуації. Один з підходів до вирішення завдання вилучення фактів з текстових документів та виявлення їх взаємозв'язків базується на технології контент-моніторингу, який можна розглядати як безперервний у часі змістовий аналіз інформаційних потоків з метою отримання необхідних якісних і кількісних інформаційних зрізів.

Порівняно з традиційними підходами впровадження технології контент-моніторингу забезпечує такі переваги:

- одержання оперативних кількісних і якісних аналітичних зрізів; своєчасне надання необхідної профільної інформації;
- забезпечення цілеспрямованої роботи працівників, усунення факторів відволікання уваги, притаманних Інтернету;
- захист даних, конфіденційність.

Завдяки таким характеристикам використовуваних інструментальних засобів роботи з інформаційними потоками, як оперативність, повнота і релевантність, а також наявності єдиного захищеного інтерфейсу технологія контент-моніторингу може сприяти значному підвищенню ефективності та якості інформаційно-аналітичної роботи.

1. Берко А. Ю. та ін. Системи електронної контент-комерції: моногр. / А. Ю. Берко, В. А. Висоцька, В. В. Пасічник. – Львів: Видавництво Львівської політехніки, 2009. – 612 с. 2. Додонов А. Г. Виявлення категорій і їх взаємозв'язків у рамках технології контент-моніторингу / Додонов А. Г., Ланде Д.В. // Вісник державної служби України. — 2006. — № 4. — С. 45–52 с. 3. Григорьев А., Ландэ Д.. Контент-мониторинг сетевых информационных потоков [Электронный ресурс] / Официальный ежемесячный www-регистр бизнес-ресурсов Украины и зарубежья — Режим доступа: <http://infostream.com.ua/infostream/publ/cbr/index.shtml> — Загол.с екрана. 4. Ланде Д.В. Програмно-апаратна система інформаційної підтримки прийняття рішень: наук.-метод. посіб. / Д.В. Ланде, В.М.Фурашев, О.М. Григор'єв. – Київ, 2006. – 48 с. 5. Леліков Г.І. Моніторинг діяльності органів виконавчої влади із застосуванням комп'ютерної системи контент-аналізу електронних ЗМІ / Г.І. Леліков, В.М. Сороко, О.М. Григор'єв, Д.В. Ланде // Вісник державної служби України. — 2002. — № 2. — С. 72–78 с. 6. Федорчук А. Г. Контент-мониторинг информационных потоков [Электронный ресурс] / Б-ки Нац. акад. наук: пробл. функционирования, тенденции развития. — К., 2005. — Вып. 3. — Режим доступа: <http://www.nbu.gov.ua/articles/2005/05fagmir.html>. — Загол. с экрана. 7. Григорьев А.Н. InfoStream. Мониторинг новостей из Интернет: технология, система, сервис: научно-методическое пособие / Григорьев А.Н., Ландэ Д.В., Бороденков С.А., Мазуркевич Р.В., Пацьора В.Н. – К., ООО “Старт-98”, 2007. – 40 с.

УДК 004.89

Є.В. Буров

Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

ІНТЕЛЕКТУАЛЬНА СИСТЕМА АВТОМАТИЗОВАНОГО ТЕСТУВАННЯ ПРОГРАМНОГО ПРОДУКТУ З ВИКОРИСТАННЯМ АЛГОРИТМІЧНИХ МОДЕЛЕЙ

© Буров Є.В., 2011

Розглянуто використання алгоритмічних моделей знань для автоматизації тестування програмних продуктів. Запропоновано архітектуру системи тестування, ієрархію моделей. Продемонстровано взаємодію різних типів моделей для розв'язування задачі автоматизованого тестування

Ключові слова: автоматизоване тестування, модель, керовані моделями системи.

Paper describes the use of algorithmic knowledge models for overnight testing automation. The architecture of intellectual testing system is described. Algorithmic models formalization is proposed. Knowledge model interaction and execution is presented for a real world test automation system.

Key words: model-driven systems, automated testing, knowledge model.

Постановка проблеми у загальному вигляді

Однією з найважливіших проблем галузі програмного забезпечення є високий рівень складності програмних систем і пов'язані з нею проблеми складності та високої вартості адміністрування, розроблення та модифікації, значний рівень дефектів у таких системах. За