

УДК 004.738.1:004.91]:001.102

Б.П. Топорівський*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»***АНАЛІЗ РЕКОМЕНДАЦІЙ ВИКОРИСТАННЯ WEB MINING ДЛЯ РОЗРОБКИ ТЕХНОЛОГІЇ ДОБУВАННЯ ЗНАНЬ З ІНФОРМАЦІЙНОЇ СИСТЕМИ ОРГАНІЗАЦІЇ**

У даній роботі наводиться аналіз моделей, методів та алгоритмів пошуку інформації у неструктурованих джерелах, в інформаційному просторі організації, та рекомендацій, що до їх застосування. У якості методології пошуку було обрано web mining, адже, наприклад, метод повного перебору вимагає великих ресурсів обчислювальних систем і з кожним разом потребує все більше часу, через неймовірне зростання кількості інформації в Інтернеті, та зростання об'єму даних організації. Наводиться аналіз підходів та поглядів ряду вчених до методології web mining, рекомендацій, що до удосконалення існуючих алгоритмів, а також ставляться задачі подальших досліджень, у висновках.

Ключові слова: тематичний пошук, неструктурована інформація, машапи, web mining, релевантність, інформаційний простір організації.

Форм. 2. Рис.: 3. Літ.: 7.

Б.П. Топоровский*Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского»***АНАЛИЗ РЕКОМЕНДАЦИЙ ИСПОЛЬЗОВАНИЯ WEB MINING ДЛЯ РАЗРАБОТКИ ТЕХНОЛОГИИ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ИЗ ИНФОРМАЦИОННОЙ СИСТЕМЫ ОРГАНИЗАЦИИ**

В данной работе приводится анализ моделей, методов и алгоритмов поиска информации в неструктурированных источниках, в информационном пространстве организации, и рекомендаций, к их применению. В качестве методологии поиска было выбрано web mining, ведь, например, метод полного перебора требует больших ресурсов вычислительных систем и с каждым разом требует все больше времени, через невероятный рост количества информации в Интернете, и рост объема данных организации. Приводится анализ подходов и взглядов ряда ученых методологии web mining, рекомендаций, к совершенствованию существующих алгоритмов, а также ставятся задачи дальнейших исследований в выводах.

Ключевые слова: тематический поиск, неструктурированная информация, машапов, web mining, релевантность, информационное пространство организации.

В.Р. Toporivskyi*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"***ANALYSIS OF RECOMMENDATIONS OF APPLYING WEB MINING FOR DEVELOPING TECHNOLOGY TO MINE KNOWLEDGE FROM THE ORGANIZATIONAL INFORMATION SYSTEM**

The research paper deals with the analysis of models, methods and algorithms of the information search in unstructured sources, in the information space of the organization, and the recommendations for their application. Web mining was chosen as a search methodology, because, for example, the method of full search requires huge computing resources, each time requiring more time due to the incredible increase in the volume of information on the Internet, and the growth of the organization's data volume. An analysis of the approaches to and views of a number of scholars on the web mining methodology, recommendations for improving existing algorithms are provided, as well as the tasks of further research are set in the conclusions.

Keywords: thematic search, unstructured information, mash-ups, web mining, relevance, information space of the organization.

Вступ. У наш час пошук релевантної до запиту інформації має три основні проблеми: формування запиту пошуку (терміни та їх зв'язки у запиту); оцінювання результатів на релевантність (міра, одиниці вимірювання); прискорення виконання операцій пошуку (пошук у індексованих кластерах).

Існуючі рішення, що існують для вирішення цих проблем є наступними: пошукові системи з прихованою індексацією; оптимізація контенту веб-ресурсів організації для пошукових систем; структуризація шаблонів побудови веб-ресурсів.

Аналіз останніх досліджень.

В існуючих технологіях, наведені рішення реалізуються з використанням одного з існуючих методів: видобуток на базі покращення процедур аналізу веб-контенту, видобуток на базі

покращення процедур аналізу структур веб-ресурсу, що стосується інформаційного простору організації; видобуток на базі покращення процедур аналізу самої процедури пошуку.

У даній проблематиці актуальними є роботи Г.Г. Асєєва, де він аналізує стислу модель алгоритму web mining, а також описує математичну модель машинного навчання інформаційних систем для роботи з web mining [1-3]. С.В. Царькова [7], де він описує на які особливості роботи користувача з сайтом, в першу чергу, повинен звертати увагу засіб, на якому реалізований web mining. Ю.О. Серова [5] у яких він дає розгорнуту характеристику алгоритму web mining. Петренко А.І. [4] де, крім алгоритму web mining описаний метод машап, побудований на аналізі інформації з кількох джерел одночасно.

Постановка задачі досліджень. Пропонується організувати процедури пошуку, які персоналізовані за темою з використанням інтелектуального аналізу двох складових:

- структур веб-ресурсу (шаблони тематично згрупованої інформації)
- порядку роботи з гіперпосиланнями до веб-ресурсів (послідовальність головної - підпорядкований).

Всі розрахунки оцінок релевантності засновуються на "важливості" (логічні змінні).

Технологія процедур пошуку базується на технологіях web mining.

Мета статті. У даній роботі пропонується проведення аналізу методів та моделей пошуку тематичної інформації, що базуються на використанні методів web mining, адже метою подальших робіт роботи є розробка інформаційної технології добування знань інтегрованою інформаційної системи організації, на основі методів web mining.

Виклад основного матеріалу. Автор робіт [1-3] Г.Г. Асєєв пропонує, розглядаючи основні етапи отримання результату в web mining, враховувати специфіку інформаційного простору організації, за допомогою наступної моделі:

1. Накопичення спеціальних даних (логи, спеціально зібрана статистика користувача та ін.).
2. Попередня обробка даних. Оскільки зібрані дані характеризуються значним ступенем зашумленості, вони потребують очищення.

Дана процедура також необхідна, у зв'язку з тим, що у таких даних часто зустрічаються посилання, на фішингові сайти, при переході на які, ці дані можуть бути перехоплені або знищені, що особливо небезпечно для інформаційного простору організації.

На думку Г.Г. Асєєва, найважливішими критеріями при виборі місця очищення є стійкість і ефективність методу[1-3].

3. Обробка і аналіз даних. Застосування конкретних методів (з урахуванням поставленої задачі) для отримання практичних результатів, аналіз адекватності та застосовності отриманих результатів.

4. Інтерпретація отриманих результатів і їх перенесення на основу бізнес-завдання, в залежності від характеру поставленої задачі та термінів її виконання.

Всі вищеописані дії можуть, як вважає Г.Г. Асєєв, відбуватися протягом одного сеансу роботи співробітника організації з веб-додатком з «підлаштуванням» цього додатка під запити користувача, або аналіз може виконуватися періодично, на основі дій великої кількості співробітників організації (аналіз ефективності різних блоків сайту)[1-3].

На думку С.В. Царькова [7], при використанні web mining, особливу увагу потрібно приділяти безпеці інформаційного простору організації. Необхідні для аналізу дані знаходяться в логах серверів і cookie-файлах. При завантаженні веб-сторінки браузер також буде відвантажувати всі вставлені в неї об'єкти, наприклад графічні файли. У зв'язку з цим виникає проблема з тим, що сервер додає в журнал запису про кожен такий запит. Звідси впливає необхідність попередньої обробки даних. Після того як виділені окремі перегляди сторінок співробітниками організації, їх об'єднують в сесію.

Як тільки дані були очищені і підготовлені для аналізу, необхідно задатися наступними питаннями:

1. Яка сторінка є загальною точкою входу для співробітників? Заходять співробітники на сайт через спеціально передбачену розробниками сторінку або ж вони відразу потрапляють на інші сторінки?

2. У якому порядку проглядалися сторінки? Чи відповідає цей порядок тому, який очікують від користувачів розробники?

3. Які інші веб-портали направляють користувачів на досліджуваний сайт? Які сайти відвідує найбільше і найменше число користувачів?

4. Як багато сторінок зазвичай переглядає користувач? Якщо користувачі йдуть з сайту після перегляду невеликого числа сторінок, то, можливо, варто зробити дизайн більш дружнім.

5. Як довго відвідувачі перебувають на сайті? Якщо цей час менше того, який очікують розробники, то варто знайти причини.

6. Яка сторінка є найбільш частою точкою догляду користувачів з сайту? Чому саме з цієї країни відвідувачі залишають сайт? Чи є вона спеціально передбаченою для цього або є якісь причини, які відлякують користувача з сайту? [7]

З роботи Ю.О. Серова [5] випливає, що web mining – це застосування алгоритмів і моделей притаманних для data mining, щоб оптимізувати автоматичний пошук і отримання (extraction) інформації з інформаційного простору організації. Ця галузь має досить великий науковий і практичний інтерес, через стрімке зростання кількості інформації, яка з'являється в Інтернет і великою великим інтересом в електронній комерції (e-commerce).

На думку Ю.О. Серова [5], web mining складається з наступних складових:

1. Пошук ресурсів – це знаходження відповідних веб-документів у інформаційному просторі організації. Знаходження ресурсів – це процес знаходження і вибору даних з текстових джерел, які містяться у інформаційному просторі організації: таких як тематичні новини, сайти новин, текстове наповнення гіпертекстового контенту з встановленням тегів, а також вибірка веб-ресурсів самим співробітником організації. До ресурсів також належать текстові ресурси прямого ефіру, джерела для науковців, бази даних бібліотек тощо.

2. Фільтрацію (Selection) інформації і попередню її обробку – автор [5] розуміє, як автоматичну вибірку і попередню обробку інформації із знайдених веб-ресурсів організації.

3. Узагальнення (Generalization) він розуміє, як автоматичне знаходження загальних закономірностей (pattern), як на окремих Веб-сайтах, так і на множині сайтів.

4. Аналіз, на думку автора – це перевірка і / або інтерпретація знайдених закономірностей. Web mining тісно пов'язаний з машинним навчанням і аналізом даних. Також Web mining часто асоціюється з пошуком інформації (Information Retrieval) та видобуванням інформації (Information Extraction), хоча насправді це не те саме.

Порівнюючи погляди двох авторів на алгоритм web mining, можна зробити висновок, що Г.Г. Асеев дає більш стисло характеристику алгоритму Web mining, що орієнтована на спеціалістів, які займаються в області бізнес-проектів, що пов'язані з обробкою великих тестових даних та / або роботою з хмарними обчисленнями.

Пояснення ж Ю.О. Серова, відповідно до даного алгоритму, більш зрозумілі початківцям, що хочуть займатися дослідженнями в області пошуку неструктурованої текстової інформації в інформаційному просторі організації. Пояснюючи кроки даного алгоритму, автор вводить ряд додаткових термінів, що дають більший рівень розуміння процесу пошуку текстової інформації за допомогою web mining.

У роботі [4] Петренко А.І. описує web mining наступним чином – web mining можна перекласти як «здобич даних у Web». Web здатний визначати індивідуальні особливості кожного користувача сайту організації, спостерігаючи за динамікою його користування інформаційним простором, що додає досить потужні бонуси на ринку електронних послуг. Системи web mining здатні знаходити багато закономірностей, наприклад, який з відвідувач є потенційним клієнтом інтернет-магазину, клієнти якої направленості приносить найбільший прибуток, які інтереси покупця або групи покупців.

Виходячи з роботи [4] Петренко А.І., технологія web mining має достатній набір алгоритмів та моделей здатних, на основі оформлення сайту організації, знайти нові, раніше невикористані елементи, щоб далі застосовувати їх при створенні інтерфейсу. Іншими словами, технологія web mining заснована на технології data mining для аналізу неструктурованої, неоднорідної, розподіленої і значної за об'ємом інформації, що міститься на Web-вузлах.

При розгортанні web mining, для впорядкування інформаційного простору, перед співробітниками IT-відділу організації виникає дві проблематики: перша – накопичення інформації, другий – розробка моделей та методів ідентифікації користувача. У підсумку накопичення певної кількості персоніфікованих ретроспективних даних про відповідного відвідувача, система формує набір його вподобань і може надавати йому, наприклад, банери з рекламою продукції, що його цікавить. На основі інформації про всіх відвідувачів сайту web-система може виявити групи відвідувачів і також рекомендувати їм товари або ж пропонувати товари в розсилках [4].

Автор [4] також відмічає, що в останні роки з'явилися Web-додатки типу Машап (від англ. mash-up — «змішувати»), у яких збираються дані з декількох джерел, серед документів організації. Будуються вони на основі комбінування діяльності різних програмних інтерфейсів і джерел даних.

Машапи вже застосовуються як:

- сервіси агрегування (інформацію з різних джерел інформаційного простору організації розміщують в одному місці);
- збирачі даних (із даних з різних джерел створюють новий сервіс (тобто агрегування));
- контролери змісту (відслідковують, фільтрують, аналізують та дозволяють пошук сервісів);
- сервісні збирачі. [4]

Webmining і пошук інформації (Information Retrieval). Пошук інформації (Information Retrieval – IR), за Ю.О. Серовим [5] – це автоматичне знаходження усіх релевантних документів інформаційного простору організації і водночас мінімізація не релевантних документів серед знайдених, а також рангування знайдених документів за мірою релевантності. Первинна мета IR – індексування текстів і пошук важливих документів.

Сучасні дослідження також відносять до IR моделювання, класифікацію документів організації, інтерфейси співробітників та клієнтів організації, візуалізацію даних, фільтрування тощо. Якщо вважати, що web mining – класифікація веб-документів з подальшою індексацією, то тоді web mining є частиною процесу IR. У будь-якому випадку задачі індексування використовують методи data mining.

Така думка Ю.О. Серова є досить критично, адже web mining лише у даному випадку можна вважати частиною Information Retrieval. У web mining є й інші задачі, що пов'язані з відслідковуванням поведінки користувача, а не тільки пошуку та індексації документів. Тому твердження автора, яке він вважає абсолютним, діє лише тоді, коли web mining застосовується лише зі сторони веб-контенту, без прив'язки до дій користувача, а тим більше оцінки діяльності певних груп користувачів, за спільними інтересами та уподобаннями.

Категорії Web mining. На думку Ю.О. Серова [5], web mining складається з 3 категорій, відносно до частин Веб, з якими можна проводити дослідження: дослідження веб-контенту інформаційного простору організації (web content mining), дослідження структури веб (web structure mining), дослідження діяльності відвідувачів (користувачів інформаційного простору організації) (Web usage mining).

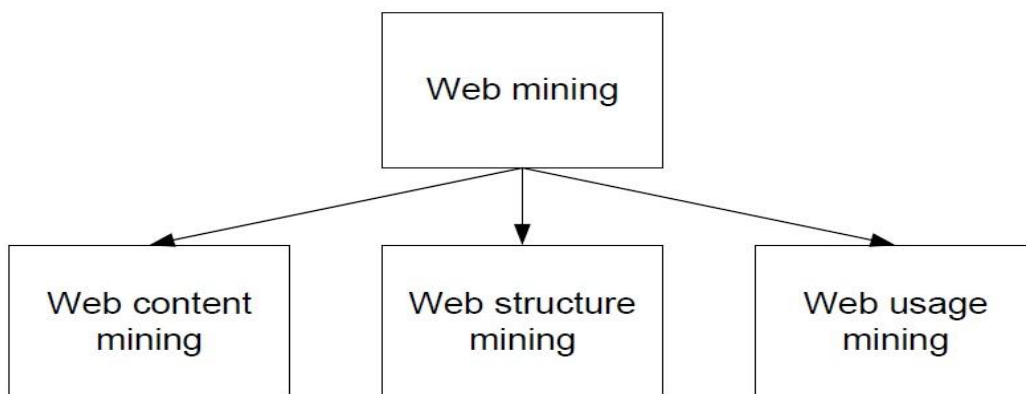


Рис. 1. Категорії web mining [8]

Дослідження вмісту Веб (Web content mining)

Загальний опис Web content mining. На думку автора [5], web content mining займається пошуком актуальної інформації у веб-документах і даних інформаційного простору організації, які містяться у веб, для того, щоб надалі видавати клієнту чи співробітнику той результат, який відповідає його уподобанням.

Web structure mining. Займається проблемою аналізу інтерфейсу сайту організації. Відслідковується, які елементи сайту подобаються клієнтам чи співробітникам організації краще, а які ні. І навіть, якщо після того web content mining підібрав, для клієнта те, що йому подобається,

він може не звернути на це увагу через «недружній інтерфейс» того чи іншого сайту, тому між ними потрібна синхронізація, ще перед відправку «на гору» (відносно схеми рис.1), результатів роботи, на відміну від аналогічної схеми у роботі [5], де такої синхронізації немає.

Точно так само, як робота web usage mining, взагалі, без синхронізації з результат атами двох попередніх, не можлива. Адже, якщо результати будуть аналізуватися лише у центральному вузлі «web mining», то кожен з 3 складових, видасть лише результати своєї діяльності, синхронізувати які між собою можна буде лише за мітками часу. А це може дати лише чіткі відповіді на запитання «коли?» і «де?», але далеко не завжди на питання «чому?».

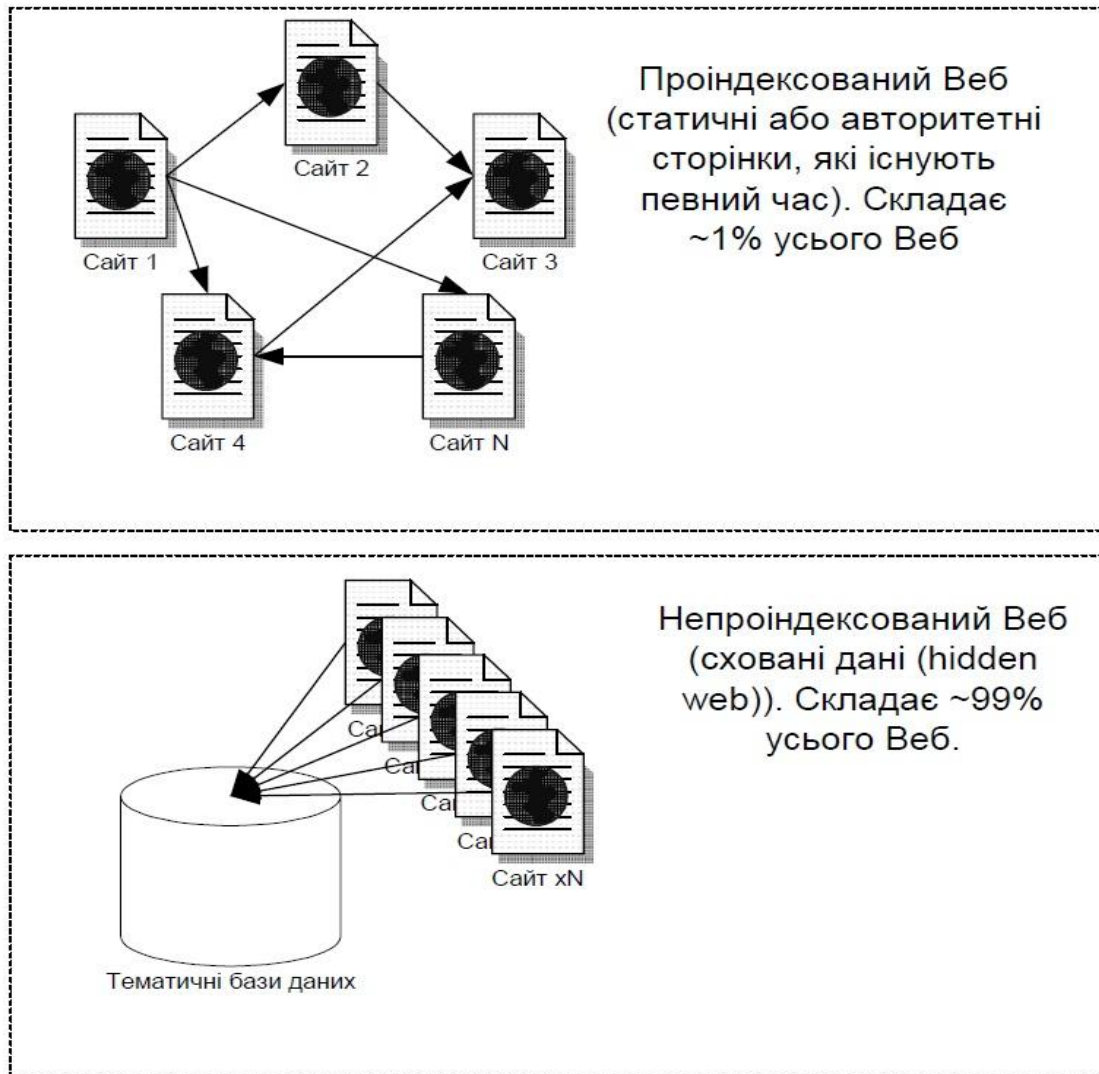


Рис. 2. Проблема схованих даних у WWW [5]

На рис. 2 показано, що значно збільшилася б ефективність пошуку, якщо б сайти не індексувалися б лише у базі даних інформаційного простору організації, а кожен сайт також містив би мітки на хоча б на декілька сайтів, що схожі за тематикою, то процес пошуку був би значно ефективнішим

Також Ю.О. Серов [5] виділяє, що у веб багато “прихованих” даних (hiddenweb), що не підлягають індексації. Це дані, котрі згруповані у базах даних за інтересами і генеруються в динаміці, згідно з тим чи іншим запитом до СУБД (наприклад, бази даних географічних карт, прогнозів погоди, телефонні довідники тощо) або є приватними.

Прихованими є також дані на авторитетних сайтах, які ще не були проіндексовані (форуми, чати, сайти інформаційних агентств). Об’єм схованих даних становить приблизно 7500 терабайт інформації.

Тому основне завдання web content mining – видобування знань з усіх даних, які містяться у веб.

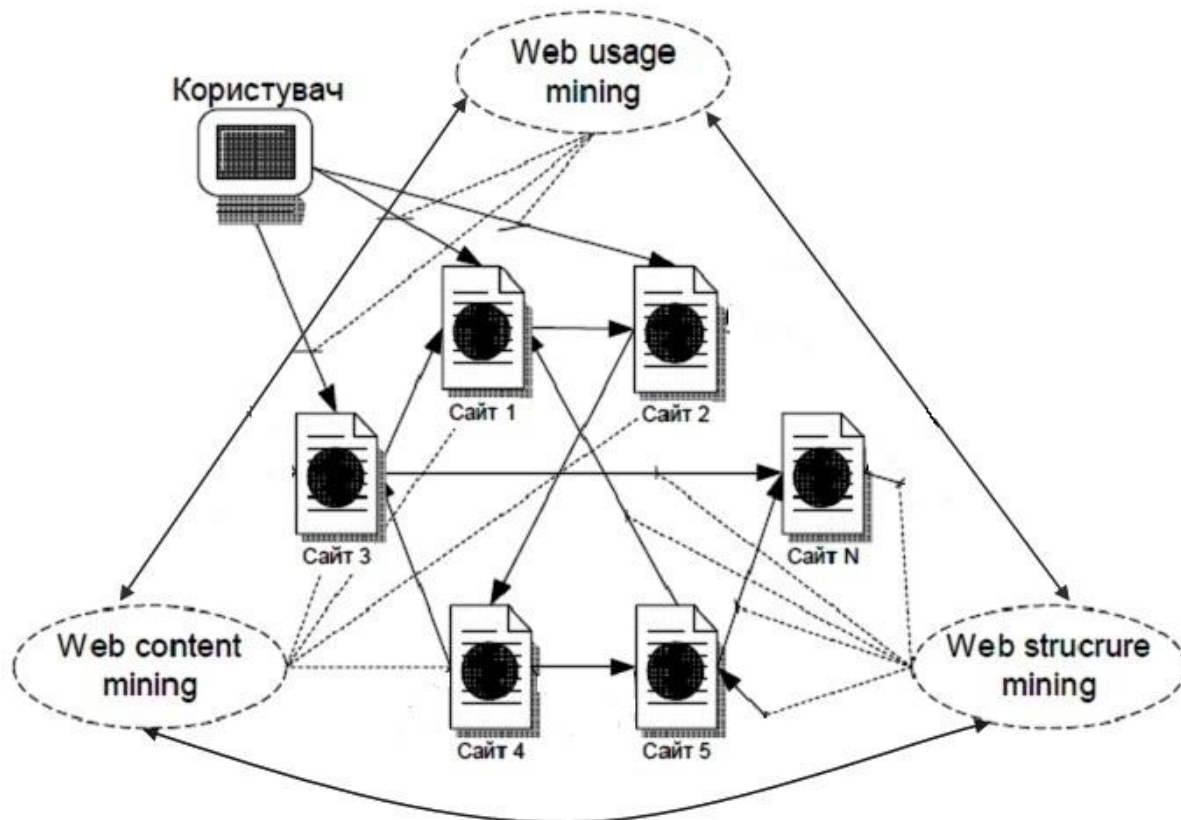


Рис. 3. Поділ web mining за частинами Веб, які досліджуються

На рис. 3 подано схему, що показує роботу схеми на рис.1 в динаміці, при взаємодії сайтів з співробітником. Тут також відображений зв'язок між відвідуваними сайтами та співробітником організації, але всі 3 компоненти web mining працюють синхронно, у порівнянні з роботу [5], де вони працюють окремо один від одного.

Методи Web Mining. На думку Г.Г. Асеева [2-3], галузі обробки веб-даних інформаційного простору організації використовується досить велика кількість різних моделей і алгоритмів обробки даних. Слід зазначити, що такі методи повинні забезпечити високий рівень стійкості, оскільки дані досить суттєво зашумлені, можуть вимірюватися в різних системах (дискретні, безперервні, текстові, дати і т.д.) і з досить великим діапазоном значень.

Г.Г. Асеев [2-3] наводить приклад, у якому методи класифікації інформаційного простору використовуються для виділення груп співробітників організації і надання кожній з них особистого контенту, що цікавить саме її. Ряд моделей та алгоритмів пошуку асоціацій використовуються для виявлення найбільш відповідних патернів в поведінці користувачів з метою визначення найбільш важливих сторінок і ресурсів сайту, скорочення шляху до них і видалення непотрібного «інформаційного сміття» зі сторінок.

Іншими словами, Г.Г. Асеев, відповідно до ієрархії компонентів web mining Ю.О. Серова, тут описує webusagemining, що базується на методах та алгоритмах відслідковування поведінки співробітника або групи співробітників організації за службовими інтересами.

Методи та алгоритми обробки текстових документів з інформаційного простору організації (Textmining) застосовуються для отримання документів, що пройшли відбір по групам за змістом і знаходження ідентичних за змістом сторінок. Наприклад, це застосовується у системах штучного інтелекту для відбору новин, налаштовані на отримання всіх посилань з усіх новинних сайтів і т.д.

Класифікація з навчанням. Нехай є набір об'єктів, кожен з яких належить одному з m класів. Як приклад Г.Г. Асеев [2-3] приводить базу клієнтів банку (як приклад користувачів інформаційного простору організації), які можуть бути віднесені до класів сумлінних або недобросовісних позичальників, або безліч солдат на фотознімку, які можна розділити на 2 групи, за характерними ознаками армій сторін конфлікту. Задачею класифікації з навчанням є складання алгоритмів, за якими для будь-якого об'єкта можна з великою імовірністю визначити клас, до якого відноситься той чи інший об'єкт. Так само і для відвідувачів сайту, які розподіляються на групи, наприклад, інтернет-магазину теж реалізується алгоритм, за яким покупцеві, при

авторизації, з початку пропонуються товари, які вони найбільше купують, або якими найбільше цікавляться.

У своїх роботах Г.Г. Асєєв [2-3] пропонує наступну модель визначення приналежності до того чи іншого класу: нехай x_1, \dots, x_k – атрибути об'єкта (вікова категорія користувача, стать, група улюблених товарів тощо), m – число класів (групи товарів, послуг, музичних композицій, книжок тощо), які об'єднані під ті чи інші атрибути користувачів x_1, \dots, x_k або їх поєднання. В результаті класифікації повинна бути отримана деяка функція

$$f(x_1, \dots, x_k),$$

значення якої належить $\{1, \dots, m\}$, і надає унікальний ідентифікатор класу, до якого відноситься об'єкт з атрибутами x_1, \dots, x_k .

На вибір дослідника представляється певна кількість об'єктів, у яких ідентифікатор, що задає приналежність до певного класу, вже відомий. Даний набір об'єктів може бути використаний для навчання моделі, тобто підбору унікальних ознак для моделі, і для випробувань побудованої моделі класифікації.

Класифікація з навчанням виділяє такі стадії:

Підготовка даних. Наявний набір об'єктів з відомими ідентифікаторами класів розбивається на 2 частини: навчальну вибірку і тестову вибірку.

Бажано, щоб це розбиття було вироблено випадковим чином. Найчастіше навчальна вибірка має розмір більше, ніж тестова.

Навчання моделі. Параметри моделі класифікації підбираються на основі навчальної вибірки таким чином, щоб домогтися максимальної відповідності умовам між передбаченими і фактичними ідентифікаторами класів.

Тестування моделі. Отримана в результаті навчання модель досліджуються на відповідність.

Для чого проводяться обчислення відсотку невірних результатів класифікації об'єктів з тестової вибірки.

Класифікація з навчанням має безліч випадків практичного застосування, наприклад, в таких областях, як кредитування, медична діагностика, прогноз доходів, маркетинг та інші задачі, що стосуються обробки даних в інформаційному просторі організації [2-3].

Матричний латентно-семантичний аналіз. Цей метод базується на кластерному аналізі LSA/LSI, що побудований на сингулярному розкладанні матриць. Представимо матрицю A , як рядки, кожен елемент яких, відповідає тому чи іншому документу, а стовбці набору вагових значень термів із множини m [6].

Сингулярне розкладання матриці, що має ранг r та розмірністю $m \times n$, є наступний вираз:

$$A = S \times U \times V,$$

де U та V – ортогональні матриці, що мають розмірності $m \times r$ і $r \times n$, а S – матриця діагональних елементів, усі з яких є позитивними, вони є сингулярними значеннями матриці A . Матриця S є квадратною, на відміну від матриці A . Приведений приклад розбиття матриці має властивість, що матриця S , якщо її залишити лише k найбільш сингулярних значень, а в матрицях U та V тільки ті колонки, що відповідають цим значенням, то матриця A буде найоптимальнішою апроксимацією по Фробеніусу, та буде мати ранг k , який не більше r [6].

Згідно з методом LSA, лише k максимальних сингулярних значень матриці A , створюють k -мірний факторний простір, відповідно до якого, проектуються як документи (матриця V), так і терміни (матриця U). У підсумковому факторному просторі документи і терміни групуються в масиви (кластери), що мають деякий загальний сенс, не заданий в явному вигляді, тобто латентний [6].

Вибір оптимального значення k для LSA – це проблема наступних досліджень. В ідеалі, значення k має бути досить велике для відображення усієї реально існуючої структури даних, але в той же час досить мале, щоб не враховувати випадкових залежностей [6].

Висновки та задачі подальших досліджень.

1. Проаналізовано існуючі рекомендації, що до вилучення знань з текстів, що знаходяться в межах інформаційного простору організації, на основі web mining.

2. Розглянуто та проаналізовано погляди науковців, що до алгоритму роботи web mining та їх оцінку можливостей даної технології, з погляду на пошук та аналіз текстових даних в інформаційному просторі організації.

3. Проаналізовано математичну модель класифікації з навчанням та латентно-семантичного аналізу, що описують метод web mining.

Великий об'єм інформації що циркулює в інтегрованих інформаційних системах організації, а також її низька структурованість та семантична неузгодженість зумовлюють необхідність розвитку технологій автоматичного вилучення інформації з тексту. Перспективним підходом у вирішенні даної проблеми може бути використання методів та алгоритмів інтелектуального аналізу даних. А саме технології Web Mining, що застосовує технологію Data Mining для аналізу неструктурованої, неоднорідною, розподіленої і значної за обсягом інформації, що міститься на Web-вузлах.

Пропонується вдосконалити роботу в інформаційному просторі організації та спроектувати підсистему інформаційного пошуку та видобування знань застосувавши категорій Web Mining:

- Попередня обробка даних для Web Mining.
- Виявлення шаблонів і відкриття знань з використанням асоціативних правил, тимчасових послідовностей, класифікації та кластеризації;
- Аналіз отриманого знання.

На жаль, на сьогодні жодна з інтелектуальних систем класифікації знань не володіє методами, що можуть сказати до якого виду відносяться ті чи інші дані, тому для формування знань та навичок, у процесі використання того чи іншого матеріалу інформаційного простору організації, не підходять, розпізнаючи лише ті чи інші ключові слова та / або словосполучення, не беручи до уваги їх сенс (дата, номер телефону, число, прізвища, імена і по батькові людей).

Тому у подальших дослідженнях ставляться наступні задачі, що потребують вирішення, а саме:

- Дослідити можливість використання методів web mining в моделях добування знань з текстових потоків.
- Вдосконалити інформаційну технологію формування інформаційного простору організації та спроектувати підсистему інформаційного пошуку та видобування знань.
- Результати дослідження пропонується реалізувати в вигляді веб-додатку, що здійснює моніторинг інформації в середовищі інформаційного простору організації.

1. Асеев Г.Г. Методы интеллектуальной предобработки данных в электронных хранилищах / Г.Г. Асеев // Радиоелектроніка. Інформатика. Управління. – 2010. - №2(23). – С. 106-111.

2. Асеев Г.Г. Методы интеллектуального анализа данных в электронных хранилищах: генетические алгоритмы / Г.Г. Асеев // Радиоелектроніка. Інформатика. Управління. – 2011. - №2(25). – С. 82-86.

3. Асеев Г.Г. Нейросетевой анализ непрерывных потоков данных из электронных хранилищ / Г.Г. Асеев // Системи обробки інформації. – Х.: ХУ ПС, 2013. – Вип. 4(111). – С. 52-56.

4. Петренко А.І. Grid та інтелектуальна обробка даних data mining / А.І. Петренко // Системні дослідження та інформаційні технології, 2008. – № 4. – С. 97-104.

5. Серов О.Ю. Технології пошуку та видобування даних у www (аналіз проблеми) / О.Ю. Серов // Lviv Polytechnic National University Institutional Repository. – С. 276-286. [Електронний ресурс] Режим доступу: <http://ena.lp.edu.ua/bitstream/ntb/9419/1/27.pdf>

6. Токман М.В. Моделі і методи поліпшення релевантності пошуку текстових документів / М.В. Токман, В.В. Сокол, Н.С. Лесна // Системи обробки інформації, 2012, випуск 5 (103). – С. 109 – 113

7. Царьков С.В. Web Mining: анализ использования веб-ресурсов, обработка веб-лога [Електронний ресурс] Режим доступу: <https://basegroup.ru/community/articles/web-usage-mining-part1>

Стаття надійшла до редакції 31.01.2018