

АКТУАЛЬНІ МЕХАНІЗМИ ДОСЛІДЖЕННЯ ТА ДОБУВАННЯ ІНФОРМАЦІЇ З WEB-ДОКУМЕНТІВ ТА СЕРВІСІВ

Інформаційно-комп'ютерні технології являють собою процеси, методи пошуку, збору, зберігання, обробки, надання, поширення інформації та способи здійснення таких процесів і методів. Наповнення Інтернет простору є динамічним та у більшості випадків неструктурованим, що є глобальною проблемою серед користувачів-пошуковців. Веб-інтелектуальний аналіз – є провідною сферою використання методів інтелектуального аналізу даних для автоматичного виявлення і вилучення інформації зі Всесвітньої павутини. Аналіз веб-структури допомагає користувачам отримувати відповідні документи, аналізуючи структуру посилань в Інтернеті.

У рамках даного наукового дослідження пропонується інноваційний підхід до процесу добування інформації з Web сутність якого базується на згаданих алгоритмах. Модернізований алгоритм дозволяє здійснювати запити на інформацію, відповідно до гіперпосилання (аналогічно HITS) та потребує мінімального запиту та часу обробки (аналогічно PageRank).

На сторінках даної роботи було проведено дослідження сучасних алгоритмів добування інформації зі Всесвітньої павутини. Розкрито три напрямки Web mining по використанню веб-даних, що застосовуються в якості вхідних даних процесу інтелектуального аналізу даних, а саме, видобуток веб-контента, видобуток веб-структур і аналіз використання веб-ресурсів. Аналіз гіперпосилань і описані алгоритми пов'язані з інтелектуальним аналізом веб-структури. Незважаючи на те, що існує три області веб-інтелектуального аналізу даних, відмінності між ними зужуються, оскільки всі вони взаємозв'язані.

Запропонований інноваційний підхід до процесу добування інформації з Web базується на алгоритмах ранжування та застосування гіперпосилань у пошуку. У межах запропонованої моделі здійснюється розрахунок, у автономному режимі, набору векторів, кожен з яких відповідає певній темі, з метою створення для кожної окремої сторінки набору оцінок важливості за конкретними темами. Таким чином, знижується кількість часу на обробку запиту та розширюються можливості пошуку у рамках більшої області.

Перспективи подальших розробок у даному напрямку ґрунтуються на вдосконаленні запропонованої моделі добування інформації з Web документів та сервісів шляхом впровадження механізму кластеризації та аналізу веб-сторінок на рівні інформаційних блоків

Ключові слова: Інтелектуальний аналіз даних, терміни, ранжування, інформаційні технології, Web-документи, гіперпосилання, добування інформації.

B.P. TOPORIVSKYI, O.O. GAGARIN

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

UP-TO-DATE MECHANISMS OF RESEARCH AND EXTRACTION OF INFORMATION FROM WEB-DOCUMENTS AND SERVICES

Information and computer technologies are processes, methods for the search, collection, storage, processing, provision, dissemination of information and methods for implementing such processes and methods. Internet space filling is dynamic and in most cases unstructured, which is a global problem among search engine users. Web-based intelligence is a leading area in the use of data mining methods to automatically detect and retrieve information from the World Wide Web. Web structure analysis helps users get relevant documents by analyzing the structure of links on the Internet.

In the framework of this scientific research an innovative approach to the process of obtaining information from the Web is proposed, which is based on the mentioned algorithms. The upgraded algorithm allows you to query information in accordance with a hyperlink (similar to HITS) and needs a minimal request and processing time (similar to PageRank).

In the framework of this work, a study of modern algorithms for obtaining information from the World Wide Web was conducted. There are three areas of Web mining for using web data that is used as input to the data mining process, namely, extraction of web content, extraction of web structures and analysis of the use of web resources. Analysis of hyperlinks and described algorithms are related to the intellectual analysis of the web structure. Despite the fact that there are three areas of web-based data mining, the differences between them are narrowed, because they are all interconnected.

The proposed innovative approach to the process of obtaining information from the Web is based on ranking algorithms and the use of hyperlinks in the search. Within the framework of the proposed model, an autonomous calculation, a set of vectors, each corresponding to a specific topic, is calculated, with the aim of creating for each individual page a set of values estimates for specific topics. This reduces the time it takes to process the request and expands the scope of search within a larger area.

Prospects for further developments in this area are based on the improvement of the proposed model for obtaining information from Web documents and services by introducing a mechanism for clustering and analyzing web pages at the level of information blocks.

Key words: Data mining, terms, ranking, information technology, Web documents, hyperlinks, information extraction.

Вступ та постановка проблеми дослідження. XXI ст. відзнаменувалося стрімким розвитком інформаційних технологій, які заповнили усі сфери життя людини. Інформаційно-комп'ютерні технології являють собою процеси, методи пошуку, збору, зберігання, обробки, надання, поширення інформації та способи здійснення таких процесів і методів [1]. Наповнення Інтернет простору є динамічним та у більшості випадків неструктурованим, що є глобальною проблемою серед користувачів-пошуковців.

Веб-інтелектуальний аналіз – є провідною сферою використання методів інтелектуального аналізу даних для автоматичного виявлення і вилучення інформації зі Всесвітньої павутини. Аналіз веб-структури

допомагає користувачам отримувати відповідні документи, аналізуючи структуру посилань в Інтернеті.

Мета роботи. Розкрити основні механізми дослідження та добування інформації з Web-документів та сервісів. Охарактеризувати актуальні алгоритми добування та аналізу інформації зі всесвітньої павутини. Запропонувати пріоритетні напрямки модифікації алгоритму ранжування посилання та розкрити алгоритм HTS у якості основи для розробки інноваційного алгоритму з мінімальними витратами часу на обробку запиту.

Аналіз останніх досліджень і публікацій. На сьогодні, питанням дослідження та добування інформації з Веб займалося чимало, як зарубіжних так і вітчизняних вчених, до їх числа варто віднести: В. Ф. Ситника, М. Т. Краснюка [1], Ф. Барсеґяна, М. Купріянова, В. Степаненко, І. Холоду [2] та В. Дюка [3].

Dr. G. N. Purohit [4] пропонує використовувати набір веб-сторінок, які містять якийсь термін в упередженому наборі для впливу на обчислення PageRank, з метою повернення термінів, для яких ця сторінка має високу репутацію. N. Duhan, A.K. Sharma та K.K. Bhatia [5] запропонували підхід до підвищення рейтингу пошуку шляхом створення вектора PageRank для кожного можливого терміну запиту з вигідними результатами. Однак цей підхід вимагає значного часу обробки та зберігання і не може бути легко розширений для використання контексту користувача та запиту. А.В. Ліфференко, С.С. Бежитський [6] пропонують використання конкретних оцінок PageRank для підвищення вертикального пошуку.

Однак, незважаючи на масштабність наукових досліджень за темою даної роботи, питання інтелектуального аналізу даних залишається відкритим та потребує детального опрацювання.

Виклад основного матеріалу дослідження. Основними завданнями інтелектуального аналізу даних (Web mining) є:

- пошук ресурсів: завдання вилучення передбачуваних веб-документів;
- вибір інформації та попередня обробка: автоматичний вибір та попередня обробка конкретної інформації з добутих веб-ресурсів;
- узагальнення: автоматичне виявлення загальних шаблонів на окремих веб-сайтах, а також на декількох сайтах;
- аналіз: перевірка та інтерпретація моделей.

На сьогодні відокремлюють три напрямки Web mining по використанню веб-даних, що застосовуються в якості вхідних даних процесу інтелектуального аналізу даних, а саме, видобуток веб-контенту, видобуток веб-структур і аналіз використання веб-ресурсів. Аналіз веб-контенту пов'язаний з отриманням інформації з www у більш структурованій формі і з індексацією інформації. Аналіз використання веб-ресурсів – це процес ідентифікації шаблонів перегляду шляхом аналізу навігаційної поведінки користувача. Аналіз веб-структури полягає в тому, щоб знайти модель, яка лежить в основі структури посилань веб-сторінок, каталогізувати їх і генерувати таку інформацію, як подібність і взаємозв'язок між ними, використовуючи переваги їх топології гіперпосилань. Аналіз гіперпосилань і описані алгоритми пов'язані з інтелектуальним аналізом веб-структури. Незважаючи на те, що існує три області веб-інтелектуального аналізу даних, відмінності між ними звужуються, оскільки всі вони взаємозв'язані.

Структура веб-даних: традиційна система пошуку інформації в основному орієнтована на інформацію, що міститься в тексті веб-документів. Метод веб-інтелектуального аналізу надає додаткову інформацію через гіперпосилання, до яких підключені різні документи.

Веб можна розглядати як орієнтований граф, в вузлах якого знаходяться документи або сторінки, а ребра – гіперпосилання між ними. Цей спрямований граф, називають Web-графом. Граф QW складається з двох множин MV і MR . Набір MV – це кінцевий, не порожній набір вершин.

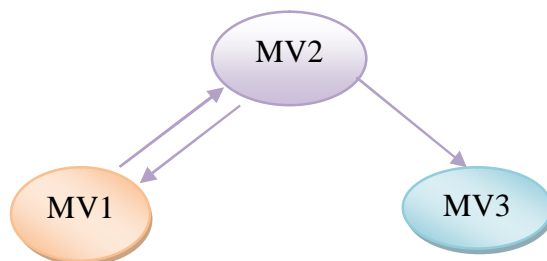


Рис. 1 Дирекційний граф (QW)
* розробка автора на основі [5]

Набір MR – це набір пар вершин; ці пари називаються ребрами. Позначення MV (QW) і $MR(QW)$ являють собою набори вершин і ребер, відповідно графа QW . Для подання графа можна також записати $QW = (MV, MR)$. Граф на рис. 1 – це спрямований граф з 3 вершинами і 3 ребрами.

На орієнтованому графі з n вершинами максимальна кількість ребер дорівнює $n(n-1)$. При

використанні 3 вершин максимальне число ребер може становити $3(3-1) = 6$.

Граф наведений на рис. 1 є сильно зв'язний так як відсутній шлях від $MV1$ до $MV3$. Загалом Web мережу можна уявити як великий графік, що містить кілька сотень мільйонів або мільярдів вузлів або вершин і кілька мільярдів дуг або ребер.

Аналіз гіперпосилань: багато веб-сторінок не містять слів, що описують їх основне призначення (наприклад, рідко портал пошукової системи включає слово "пошук" на своїй домашній сторінці), і існують веб-сторінки, які містять дуже мало тексту (наприклад, зображення, музика, відео ресурси), що робить методи пошуку на основі тексту не здійсненними. Тоді актуальним виступає використання гіперпосилання, яке включає в себе характеристику інформації наведеної на сторінці.

Багато дослідників [7,8] вивчали це питання та запропонували рішення та шляхи вирішення вказаної проблеми пошуку, індексації або запиту до мережі Інтернет з урахуванням її структури, а також метайнформації, яку включено до гіперпосилання, і навколишнього тексту. На основі аналізу зв'язків пропонується ряд алгоритмів [4,5,9], основні з них алгоритм ранжування посилання та алгоритм HITS з усіма модифікаціями.

Алгоритм ранжування посилання (рис. 2). Це алгоритм який виявляє важливість веб-сторінки за посиланням на неї, тобто проводиться ранжування посилання на веб-сторінку.

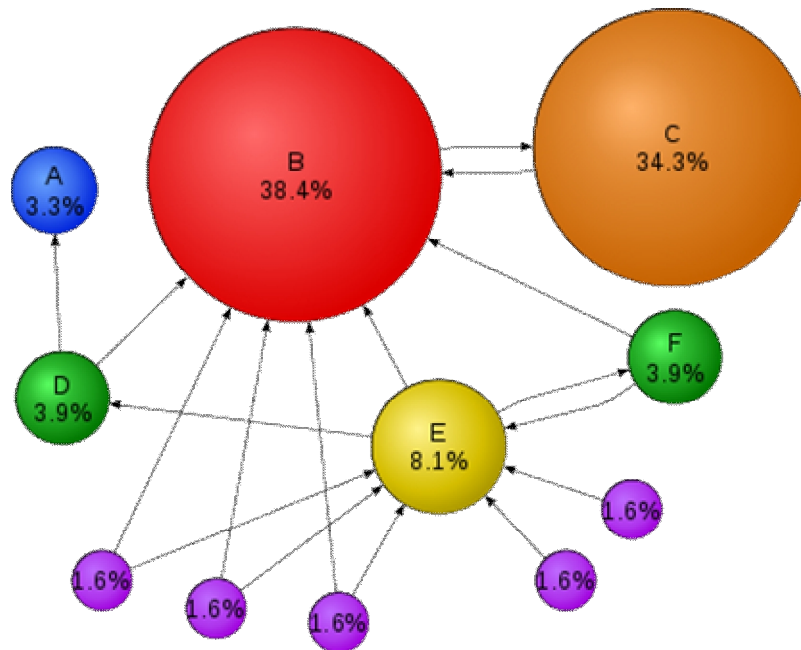
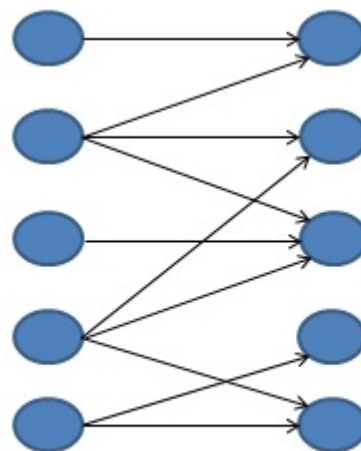


Рис. 2 Алгоритм ранжування посилання (PageRank) [8]

Алгоритм HITS (англ. Hyperlink Induced Topic Search) (рис. 3), запропонований у 1999 році Джоном Клейнбергом [10]. Діє аналогічно алгоритму ранжування посилання, проте за відмінністю, що полягає у аналізі вмісту інформації, яка є у гіперпосиланні.



Хаб-документи Авторитетні документи

Рис. 3 Алгоритм HITS [10]

У рамках даного наукового дослідження пропонується інноваційний підхід до процесу добування інформації з Web сутність якого базується на згаданих алгоритмах. Модернізований алгоритм дозволяє здійснювати запити на інформацію, відповідно до гіперпосилання (аналогічно HITS) та потребує мінімального запиту та часу обробки (аналогічно PageRank).

У межах запропонованої моделі здійснюється розрахунок, у автономному режимі, набору векторів, кожен з яких відповідає певній темі, з метою створення для кожної окремої сторінки набору оцінок важливості за конкретними темами.

Ідея упередженості обчислення PageRank вперше була запропонована у [5] з метою персоналізації, але так і не була повністю досліджена. Цей процес упередженості передбачає введення штучних посилань на веб-граф під час автономного обчислення рангу.

Наукова праця [11] показує, що властивості веб-графу чутливі до теми сторінки. Зокрема, що сторінки, як правило, вказують на інші сторінки, які знаходяться на тій же "широкій" темі. Хоча ця властивість допомагає пояснити, чому незалежна від запиту оцінка PageRank може бути корисна для ранжирування, автор також припускає, що є можливість поліпшити продуктивність обчислень на основі посилань, беручи до уваги теми сторінки. Модернізувавши PageRank, тобто увівши показник чутливості, уникаємо проблем, взаємопов'язаних сторінок отримуючи місце для запитів, для яких немає особливих прав. Сторінки, які вважаються важливими в деяких предметних областях, не можуть вважатися важливими в інших, незалежно від того, які ключові слова можуть відобразитися на сторінці або в тексті прив'язки, що посилається на сторінку. Підхід HITS, запропонований [12], має аналогічну мотивацію, але призначений для поліпшення результатів популярних запитів. Алгоритм HITS формує запит на конкретний результат шляхом виявлення і індексування сторінок, які з'являються, за певними ключовими словами, що лежать в основі їх гіпер-посилань. Однак умови запиту, для яких не знайдено експерта, не будуть оброблені алгоритмом HITS, що є певним недоліком у застосуванні.

Підхід запропонований, у рамках даного дослідження є новим у використанні невеликого числа репрезентативних базисних тем, взятих з відкритої директорії, в поєднанні з багатовимірним наївним байєсовським класифікатором для класифікації контексту запиту.

На сторінках даної роботи розглядається два сценарії. По-перше, припустимо, що користувач з певною інформацією повинен видавати запит пошуковій системі звичайним способом, ввівши запит у поле пошуку. У цьому випадку, визначаємо теми, найбільш тісно пов'язані із запитом, і використовувати відповідну тему-чутливого вектору PageRank для ранжирування документів, що задовольняють запиту. Це гарантує, що оцінки "важливості" відображають уподобання структури для посилань на сторінки, які мають певний вплив на запит. Як і в звичайному PageRank, тема чутливої оцінки може бути використана як частина функції оцінки, яка приймає до уваги інші важливі показники. У другому випадку передбачається, що користувач переглядає документ (наприклад, переглядає веб-сайт або читає електронну пошту) і вибирає термін з документа для якого він хотів би отримати більше інформації. Це поняття пошуку в контексті обговорюється Ланде та ін [8]. Обравши відповідну тему чутливого вектору залежно від контексту запиту, є надія, забезпечити більш точні результати пошуку. Варто зазначити, що навіть коли запит видається звичайним способом, без виділення терміна, історія запиту являє собою форму контексту запиту. Ще одне джерело контексту виходить від користувача, який відправив запит.

Структура алгоритму полягає в наступному: нехай S_p ступінь результату сторінки p , і нехай $Q(q)$ являє собою значущість сторінки q . Зв'язок (p, v) присвоює $Q(q) / S_p$ одиниць рангу сторінці v .

Якщо n – кількість сторінок, варто надати всім сторінкам початкове значення $1/n$. Нехай W_v представляє набір сторінок, що вказують на v . Тоді при кожній ітерації, ряди поширюються наступним чином:

$$\forall_v Q^{(i+1)}(v) = \sum_{p \in W_v} Q^{(i)}(p) / S_p \quad (1)$$

Ітерації продовжуються до тих пір, поки $Q(q)$ не стабілізується на певному проміжку. Остаточний вектор $Q(q)$ містить вектор мережі. Цей вектор обчислюється тільки один раз після кожного обходу веб-сторінки; значення можуть використовуватися для впливу на ранжирування результатів пошуку [1].

Процес також може бути виражений у вигляді наступного розрахунку власних векторів, що забезпечує розуміння значення $Q(q)$.

Нехай Y – квадратна, стохастична матриця, яка відноситься до спрямованого графу мережі H , і передбачається, що всі вершини H мають хоча б одне ребро. Якщо між сторінками i та j є ребро (посилання), то елемент матриці має значення y_{ij} / S_j . Нехай усі інші записи мають значення 0. Одна ітерація попереднього обчислення фіксованої точки відповідає матрично-векторному множенню:

$$\dot{Q} = Y \cdot \dot{Q} \quad (2)$$

Тому що Y – це стохастична матриця переходу графа H , $Q(q)$ можна розглядати як стаціонарний розподіл ймовірностей для Марківського ланцюга.

Один з нюансів полягає в тому, що конвергенція $Q(q)$ гарантується тільки в тому випадку, якщо Y є непорушним (тобто H сильно пов'язаний) і аперіодичний [5]. Останнє гарантується у випадку з Веб, тоді як попереднє вірно тільки при додаванні фактора загасання $1-a$ в ранжируванні. На основі вищевикладеного є можливість знайти матрицю Y' , в яку додаються транзитивні ребра ймовірності a/S між кожною парою вершин H . Нехай N - кількість вузлів (тобто сторінок) у веб-графіку. Нехай l n -мірний вектор-стовпець, який представляє собою рівномірний розподіл ймовірності по всім вузлам:

$$l = \left[\frac{1}{n} \right]_{n \cdot 1} \quad (3)$$

Нехай d це n -мірний вектор-стовпець ідентифікації вузлів з нульовим вхідним блоком:

$$d_i = \begin{cases} 1 & \text{якщо } \partial(j) = 0 \\ 0 & \text{у протилежному випадку} \end{cases} \quad (4)$$

Тоді матриця Y' має вигляд:

$$K = l \cdot d^T \quad (5)$$

$$G = l \cdot [1]_{1 \cdot n} = \left[\frac{1}{n} \right]_{n \cdot n} \quad (6)$$

$$Y' = (1-a)(Y+K) + aG \quad (7)$$

Ця модифікація покращує якість алгоритму ранжування посилання шляхом введення коефіцієнта розпаду $1-a$, який обмежує ефект рангових поглиначів [6], в доповнення до гарантованого зближення з унікальним ранговим вектором. Підставляючи Y' замість Y в рівнянні 2, є можливість окреслити модифікований алгоритм добування інформації з Web:

$$Q(q) = (1-a)(Y+K) \cdot Q + al \quad (8)$$

Ключ до створення тематично чутливого алгоритму полягає у ітерації обчислень для збільшення ефекту певних категорій сторінок за допомогою нерівномірного вектора персоналізації $n \cdot 1$ для q . Щоб переконатися, що нова матриця не є похідною, недоступні ненульові вузли, повинні бути видалені. Ця зміна не є імплементаційною.

З точки зору моделі випадкового блукання вектор персоналізації представляє собою додавання повного набору ребер переходу, де ймовірність переходу з ребра (p, v) дається у вигляді aq_v .

Висновки з даного дослідження і перспективи подальших розробок у даному напрямку. У рамках даної роботи було проведено дослідження сучасних алгоритмів добування інформації зі Всесвітньої павутини. Розкрито три напрямки Web mining по використанню веб-даних, що застосовуються в якості вхідних даних процесу інтелектуального аналізу даних, а саме, видобуток веб-контента, видобуток веб-структур і аналіз використання веб-ресурсів. Аналіз гіперпосилань і описані алгоритми пов'язані з інтелектуальним аналізом веб-структури. Незважаючи на те, що існує три області веб-інтелектуального аналізу даних, відмінності між ними звужуються, оскільки всі вони взаємозв'язані.

Запропонований інноваційний підхід до процесу добування інформації з Web базується на алгоритмах ранжування та застосування гіперпосилань у пошуку. У межах запропонованої моделі здійснюється розрахунок, у автономному режимі, набору векторів, кожен з яких відповідає певній темі, з метою створення для кожної окремої сторінки набору оцінок важливості за конкретними темами. Таким чином, знижується кількість часу на обробку запиту та розширюються можливості пошуку у рамках більшої області.

Перспективи подальших розробок у даному напрямку ґрунтуються на вдосконаленні запропонованої моделі добування інформації з Web документів та сервісів шляхом впровадження механізму кластеризації та аналізу веб-сторінок на рівні інформаційних блоків.

Література

1. Ситник В. Ф. Інтелектуальний аналіз даних (дейтамайнінг) : навч. посіб. / В. Ф. Ситник, М. Т. Краснюк. - К.: КНЕУ, 2007. - 376 с.
2. Барсегян Ф. А. Методы и модели анализа данных OLAP и Data Mining / Ф. Барсегян, М. Куприянов, В. Степаненко, И. Холод. – СПб.: БХВ. – 2008. – 267 с.
3. Дюк В. "Data Mining" : учебный курс / В. Дюк, А. Самойленко. – СПб. : Питер, 2001. – 368 с.
4. Dr. G. N. Purohit Page Ranking Algorithms for Web Mining International Journal of Computer

5. N. Duhan, A.K. Sharma and K.K. Bhatia, Page Ranking Algorithms: A Survey, Proceedings of the IEEE International Conference on Advance Computing, 2009.

6. Лиференко А.В. О задачах применения web-mining/ А.В. Лиференко, С.С. Бежитский // Актуальные проблемы авиации и космонавтики. – 2015. – №11. [Электронный ресурс]– Режим доступа: <http://cyberleninka.ru/article/n/o-zadachah-primeneniya-web-mining> (дата обращения: 20.09.2017).

7. Рассел С. Искусственный интеллект: современный подход / С.Рассел, П.Норвинг – М.: Издательский дом "Вильямс", 2006. – 1408 с.

8. Ландэ Д.В. Интернетика. Навигация в сложных сетях: модели и алгоритмы / Д.В. Ландэ, А.А. Снарский, И.В. Безсуднов. М.: – Либроком, 2009. – 264 с.

9. Шелудкова Е.А. Извлечение знаний из WEB MINING / Е.А. Шелудкова // Научное сообщество студентов XXI столетия. Технические науки: сб. ст. по мат. XLVIII междунар. студ. науч.-практ. конф. № 11(47). URL: [https://sibac.info/archive/technic/11\(47\).pdf](https://sibac.info/archive/technic/11(47).pdf) (дата обращения: 24.09.2017)

10. Алгоритм HITS [Электронный ресурс] – Режим доступа : http://http-wikipediya.ru/wiki/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC_HITS#CITEREFHITS2008

11. Борисова А.Н. Технология Web Mining / А.Н. Борисова, Т.Г. Долгова // Актуальные проблемы авиации и космонавтики. – 2015.–№11. [Электронный ресурс] – Режим доступа: <http://cyberleninka.ru/article/n/tehnologiya-web-mining> (Дата обращения: 20.09.2017).

12. Web Mining: интеллектуальный анализ данных в сети Internet. – [Электронный ресурс]. – Режим доступа: <https://sites.google.com/site/upravlenieznaniami/tehnologii-upravlenia-z...> (дата обращения: 20.09.2017).

References

1. Sytnyk V. F. Intelktual'nyy analiz danykh (deytamaynih) : navch. posib. / V. F. Sytnyk, M. T. Krasnyuk. - K.: KNEU, 2007. - 376 s.

2. Barsegian F. A. Metody i modeli analiza danyh OLAP i Data Mining / F. Barsegian, M. Kuprijanov, V. Stepanenko, I. Holod. – SPb.: BHV. – 2008. – 267 s.

3. Djuk V. "Data Mining" : uchebnyj kurs / V. Djuk, A. Samojlenko. – SPb. : Piter, 2001. – 368 s.

4. Dr. G. N. Purohit Page Ranking Algorithms for Web Mining International Journal of Computer Applications (0975 – 8887) Volume 13– No.5, January 2011. – <http://www.ijcaonline.org/volume13/number5/pxc3872448.pdf>

5. N. Duhan, A.K. Sharma and K.K. Bhatia, Page Ranking Algorithms: A Survey, Proceedings of the IEEE International Conference on Advance Computing, 2009.

6. Liferenko A.V. O zadachah primeneniya web-mining/ A.V. Liferenko, S.S. Bezhitskij // Aktual'nye problemy aviacii i kosmonavtiki. – 2015. – №11. [Jelektronnyj resurs]– Rezhim dostupa: <http://cyberleninka.ru/article/n/o-zadachah-primeneniya-web-mining> (data obrashhenija: 20.09.2017).

7. Rassel S. Iskusstvennyj intellekt: sovremennyj podhod / S.Rassel, P.Norving – M.: Izdatel'skij dom "Vil'jams", 2006. – 1408 s.

8. Landje D.V. Internetika. Navigacija v slozhnyh setjah: modeli i algoritmy / D.V. Landje, A.A. Snarskij, I.V. Bezsudnov. M.: – Librokom, 2009. – 264 s.

9. Sheludkova E.A. Izvlechenie znaniy iz WEB MINING / E.A. Sheludkova // Nauchnoe soobshhestvo studentov XXI stoletija. Tehnicheskie nauki: sb. st. po mat. XLVIII mezhdunar. stud. nach.-prakt. konf. № 11(47). URL: [https://sibac.info/archive/technic/11\(47\).pdf](https://sibac.info/archive/technic/11(47).pdf) (data obrashhenija: 24.09.2017)

10. Algoritm HITS [Jelektronnyj resurs] – Rezhim dostupa : http://http-wikipediya.ru/wiki/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC_HITS#CITEREFHITS2008

11. Borisova A.N. Tehnologija Web Mining / A.N. Borisova, T.G. Dolgova // Aktual'nye problemy aviacii i kosmonavtiki. – 2015.–№11. [Jelektronnyj resurs] – Rezhim dostupa: <http://cyberleninka.ru/article/n/tehnologiya-web-mining> (Data obrashhenija: 20.09.2017).

12. Web Mining: intellektual'nyj analiz danyh v seti Internet. – [Jelektronnyj resurs]. – Rezhim dostupa: <https://sites.google.com/site/upravlenieznaniami/tehnologii-upravlenia-z...> (data obrashhenija: 20.09.2017).

Рецензія/Peer review : 16.06.2017 р.

Надрукована/Printed :15.10.2017 р.

Стаття рецензована редакційною колегією