

**ЗАСТОСУВАННЯ КЛАСТЕРНОГО АНАЛІЗУ
В РОБОТІ ПОШУКОВИХ СИСТЕМ ТОРГОВЕЛЬНИХ МЕРЕЖ**

Анотація

Розглянуто теоретичні та практичні навички роботи інформаційно-пошуковими системами. Наведено організацію пошуку даних в інформаційних ресурсах інформаційних систем. Подано матеріали щодо організації електронного обігу інформації та взаємодії інформаційної системи з користувачем. Наведено приклади організації інформаційних потоків та використання інформаційних систем в організаціях. Сформульовано основні підходи до функціонування організацій не з погляду реалізації окремих функцій, а з позиції виконання цілісних бізнес-процесів. Розкрито ефективність автоматизації процесів, підвищення контрольованості та передбачено усі можливі дії, покращити процес взаємодії між користувачем та додатком. Розроблено методику до організації інформаційних потоків у пошукових системах для автоматизації роботи, що проводиться при організації бізнес-процесів. Запропоновано матеріал для вивчення та ознайомлення з методикою застосування кластерного аналізу у пошукових системах торговельних мереж.

Ключові слова: інформація, інформаційний потік, моніторинг, технології агрегування та узагальнення інформації, глобальні мережі, аналітична обробка інформації, торговельне підприємство.

Н.И. Бойко, к.э.н.,
Львовская коммерческая академия, г. Львов

**ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА
В РАБОТЕ ПОИСКОВЫХ СИСТЕМ ТОРГОВЫХ СЕТЕЙ**

Аннотация

Рассмотрены теоретические и практические навыки работы информационно-поисковыми системами. Приведены организации поиска данных в информационных ресурсах информационных систем. Представлены материалы по организации электронного обращения информации и взаимодействия информационной системы с пользователем. Приведены примеры организации информационных потоков и использования информационных систем в организациях. Сформулированы основные подходы к функционированию организаций не с точки зрения реализаций отдельных функций, а с позиции выполнения целостных бизнес-процессов. Раскрыто эффективность автоматизации процессов, повышение контролируемости и предусмотрено все возможные действия, улучшить процесс взаимодействия между пользователем и приложением. Разработана методика к организации информационных потоков в поисковых системах для автоматизации работы, проводимой при организации бизнес-процессов. Предложено материал для изучения и ознакомления с методикой применения кластерного анализа в поисковых системах торговых сетей.

Ключевые слова: информация, информационный поток, мониторинг, технологии агрегирования и обобщения информации, глобальные сети, аналитическая обработка информации, торговое предприятие.

Nataliya I. Boiko, Candidate of Economics,
Lviv Academy of Commerce, Lviv

**USAGE OF CLUSTER ANALYSIS IN THE SEARCH ENGINES OF THE
RETAILERS**

Annotation

The theoretical and practical skills of information retrieval systems are examined. The organization of data in search of information resources information systems are shown. Materials of the electronic

information flows and interaction with the information system users are presented. Examples of information flow and use of information systems in organizations are given. The basic approach to the operation of organizations not in terms of implementation of some functions, but from the implementation of entire business processes is formulated. The effectiveness of process automation is shown, controllability increase and all possible steps to improve the process of interaction between the user and the application are provide. The method to organize information flows in the search engines for automation of work is carried out in the organization of business processes. A material for study and review of the application of the method of cluster analysis in search engines retailers is proposed.

Keywords: information, information flow, monitoring technology aggregation and synthesis of information, global networks, analytical data processing, trading company.

Постановка проблеми. На сучасному етапі фахівцям у сфері інформаційних технологій необхідно мати уявлення, теоретичні та практичні навички роботи з інформаційно-пошуковими системами торговельних мереж.

Саме теоретичним питанням організації пошуку в інформаційних ресурсах мережі Інтернет, навігації в цій мережі присвячена дана стаття. Ми розглядатимемо питання, що відносяться до інформаційної структури Web-простору, повнотекстових інформаційно-пошукових систем, їх алгоритмічному та лінгвістичному забезпеченні, можливостям ранжування, аналітичного узагальнення результатів пошуку, загальним закономірностям сучасного інформаційного простору.

Аналіз останніх досліджень і публікацій. Особливістю досліджуваної проблематики є її недостатня представленість у друкованих та, особливо, в електронних публікаціях. У працях О. М. Томашевського, Д. В. Ланде та ін. розроблено нову концепцію комплексного аналізу ІТ-галузі економіки, починаючи від базових термінів, таких як «технологія» та «інформація», до задач оптимізації створення інформаційного продукту в економічній системі.

Метою даного дослідження є розвиток теоретичних і технологічних засад, методів і використання кластерного аналізу для агрегування та узагальнення інформаційних потоків великого обсягу у комп'ютерних мережах торговельних підприємств.

Виклад основного матеріалу. Активне використання інформаційно-комунікаційних технологій (ІКТ) є не тільки одним з основних факторів, що визначають економічне зростання торговельних підприємств [11].

У міру впровадження ІКТ в організаційно-господарську діяльність стає все складніше самостійно забезпечувати її функціонування та вдосконалення. Внаслідок цього отримала поширення і розвиток практика передачі підприємствами частини функцій з обслуговування власної інформаційної інфраструктури зовнішнім підрядникам.

Традиційно подібна інформація аналізується засобами OLAP. Незважаючи на наявність у назві слова "аналіз", фактично дані програмні засоби просто надають можливість зручного відображення інформації в тих чи інших розрізах, а аналіз, тобто пошук закономірностей і формування висновків на цій основі, проводить людина [7; 11].

МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ

На відміну від традиційних засобів OLAP, програмне забезпечення KEIS є інтелектуальним (KEIS – інтелектуальний аналіз даних (кластеризатор), призначений для аналізу інформації (різноманітних статистик, показників, відомостей), що зберігається в базах даних підприємств, текстових документах і т.д.), і дозволяє робити деякі висновки, що надаються користувачеві на основі інформації, яка аналізується.

KEIS обробляє великі обсяги інформації (сотні тисяч записів) і формує групи подібних за певними ознаками даних – кластери. Елементи всередині цих груп мають більше «подібностей» між собою, аніж з елементами сусідніх кластерів.

Кластерний аналіз дозволяє відкрити в даних раніше невідомі закономірності, які важко досліджувати іншими способами, і представити їх у зручній для користувача формі. Кластеризацію слід використовувати для аналізу не тільки категоріальних, але і текстових (категоріальних), і змішаних (категоріально-числових) атрибутів, що дуже часто зустрічається на практиці та викликає помітні труднощі у більшості відомих алгоритмів [4].

Розроблені способи нормування метрик для категоріальних атрибутів дозволили одноманітно аналізувати числові та категоріальні записи. При цьому розроблені алгоритми обробки, які надходять в систему записів, дозволяють здійснити кластеризацію за один прохід. При цьому не використовуються ітераційні та рекурсивні процеси, що скорочує час обчислень і знижує вимоги до продуктивності комп'ютерного апаратного забезпечення [1–3].

Сьогодні досягнуто певних успіхів у розв'язанні проблеми старіння інформації в рамках моделі Бартон-Кеблера, водночас питання динаміки інформаційних потоків залишаються майже не дослідженими. Динаміка мережевих інформаційних потоків обумовлена багатьма факторами, більшість з яких не піддаються точному аналізу. Однак можна припустити, що загальний характер тимчасової залежності числа тематичних публікацій у базі даних визначаються закономірностями, які цілком допускають побудову математичних моделей.

Організації-генератори нової інформації в основному працюють в стаціонарному режимі. Це означає, що кожна організація-генератор виробляє потік інформації, в середньому постійний за кількістю повідомлень. Змінюються в часі лише обсяги повідомлень, які відповідають тій чи іншій темі. Іншими словами, зростання кількості публікацій по одній темі супроводжується зменшенням публікацій на інші теми, так що для кожного проміжку часу T справедливо [5]:

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

де $n_i(t)$ – кількість публікацій в одиницю часу,

M – загальна кількість всіх можливих тем.

При моделюванні основний інтерес представляє вивчення динаміки окремого тематичного потоку, який описується щільністю $n_i(t)$.

Тематичній рубриці відповідає її інформаційний портрет:

$$P_i = \{v_{ij}\}, \quad (j = 1, \dots, K), \quad (2)$$

де v_{ij} – ваговий коефіцієнт, що відповідає j -му терму,

K – кількість термів у словнику системи.

Зважування потоку документів у просторі інформаційного портрета $M = \{m_{ij}\}$ ($i = 1, \dots, N; j = 1, \dots, K$) – матриця відповідності потоку документів D інформаційному портрету I [5; 6].

$$D = \{d_i\} \quad \{i = 1, K\} \quad (3),$$

d_i – визначається як $TF * IDF$.

Близькість D і P_i – $sim(D, P_i)$ – скалярний добуток K -мірних векторів.

Алгоритм зважування можна інтерпретувати таким виглядом:

```

while поки не вичерпався список актуальних на момент  $t$  документів do
  for кожного документа do
    while не вичерпався список інформаційних портретів
      for кожного інформаційного портрету do
        Визначення ваги документа
        If вага більша за порогове значення
          then do приписування документа рубрики, яка
            відповідає інформаційному портрету
          end if
        end for
      end while
    end for
  end while

```

Кінцевий підрахунок ваги потоку документів

Метод кластерного аналізу LSI (латентне семантичне індексування), базується на сингулярному розкладанні матриць (SVD). Сингулярним розкладанням матриці A називається її розкладання вигляду $A = USV^T$, де U і V – ортогональні матриці, а S – діагональна матриця, елементи якої $s_{ij} = 0$, якщо i не дорівнює j , а $s_{ii} > = 0$. У розглянутому прикладі (таблиць

МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ

взаємозв'язків) матриця $A = M^T M$ – квадратна, однак метод LSI застосовується і до прямокутних матриць, але в цих випадках розмірність матриці S відповідає рангу матриці A [7; 10].

За методом LSI у розгляд беруться k найбільших сингулярних значень, а кожному такому сингулярному значенню матриці A відповідає кластер взаємопов'язаних документів. A апроксимується матрицею $A_k = \sum u_i s_{ij} v_i^T$.

Метод LSI застосовний і до ранжування видачі інформаційно-пошукових систем, заснованому на цитуванні. Це алгоритм HITS (Hyperlink Induced Topic Search) – один з двох найпопулярніших на сьогодні в сфері інформаційного пошуку.

Зважаючи на свою обчислювальну трудомісткість (що дорівнює $O(N^2)$, N – розмірність A), цей метод LSI застосовується тільки для відносно невеликих матриць.

Механізм аналізу даних і прогнозування (рис. 1) надає користувачам торговельних мереж можливість здійснювати пошук неочевидних закономірностей у даних, накопичених в інформаційній базі. Цей механізм дозволяє: 1) здійснювати пошук закономірностей у вихідних даних інформаційної бази; 2) управляти параметрами під час виконання аналізу як програмно, так й інтерактивно; 3) здійснювати програмний доступ до результату аналізу; 4) автоматично виводити результат аналізу в табличний документ; 5) створювати моделі прогнозу, що дозволяють автоматично прогнозувати наступні події або значення деяких параметрів нових об'єктів [4].

Механізм аналізу даних – це набір взаємодіючих один з одним об'єктів вбудованої мови, що дозволяє розробнику використовувати його складові частини в довільній комбінації у будь-якому прикладному рішенні. Вбудовані об'єкти дозволяють легко організувати інтерактивне налаштування параметрів аналізу користувачем, а також дозволяють виводити результат аналізу в зручній для відображення формі у табличний документ.

Механізм дозволяє працювати як з даними, отриманими з інформаційної бази, так і з даними, отриманими зі зовнішнього джерела, попередньо завантаженими в таблицю значень або табличний документ [8]:

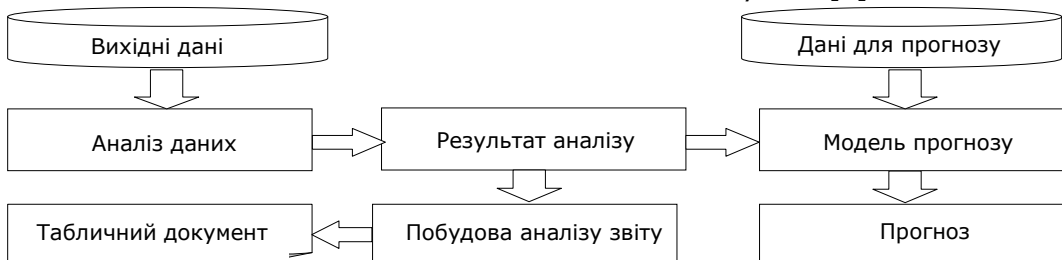


Рис. 1. Механізм аналізу даних і прогнозування в інформаційній базі

Застосовуючи до вихідних даних один з видів аналізу, можна отримати результат аналізу – певну модель поведінки даних. Він може бути відображений у підсумковому документі, або збережений для подальшого використання. Останнє полягає в тому, що на його основі може бути створена модель прогнозу, яка дозволить передбачити поведінку нових даних відповідно до наявної моделі.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямку. Технологічне інформаційне середовище руйнує сталу ієрархію управління, створюючи на її місці більш гнучкі вільні структури. Автоматизовані інформаційні системи та нові технології дають можливість оптимізувати і раціоналізувати управлінські функції, відкривають нові шляхи побудови ефективної пошукової системи торговельних мереж.

Тому впровадження нових інформаційних технологій в економічну діяльність має на меті створити не тільки автоматизацію методів опрацювання економічної інформації, але і організацію інформаційно-комунікативного процесу на якісно новому рівні.

Запропонований підхід, звичайно ж, не можна вважати остаточним рішенням поставленого завдання. Отримані практичні результати показали свою ефективність як істотне доповнення до пошукових режимів. При цьому найважливішим, мабуть, є те, що користувач прив'язується не до нових повідомлень, а до нових подій, які виникають в реальному режимі часу.

Список використаних джерел:

1. Gianna M., Del Corso, Antonio, Gulli, Francesco, Romani (2005). Ranking a stream of news. *Proceedings of the 14th international conference on World Wide Web* [International World Wide Web Conference]. Chiba, Japan, p. 97–106.
2. Арнольд В. И. Аналитика и прогнозирование: математический аспект / В. И. Арнольд // Научно-техническая информация. – Сер. 1. – 2003. – Вып. 3. – С. 1–10.
3. Брайчевский С. М. Современные информационные потоки: актуальная проблематика / С. М. Брайчевский, Д. В. Ландэ // Научно-техническая информация. – Сер. 1. – 2005. – №11. – С. 21–33.
4. Вольтерра В. Математическая теория борьбы за существование / В. Вольтерра. – М.: Наука, 1976.
5. Горькова В. И. Информетрия (количественные методы в научно-технической информации) / В. И. Горькова // Итоги науки и техники. – Сер. Информатика. – Т.10. – М.: ВИНТИ, 1988.
6. Ефимов А. Н. Информация: ценность, старение, рассеяние / А. Н. Ефимов. – М., 1978.
7. Иванов С. А. Статистический анализ документальных информационных потоков / С. А. Иванов, Н. В. Круковская // Научно-техническая информация. – Сер. 2. – Вып. 2. – 2004. – С. 11–14.
8. Ландэ Д. В. Некоторые методы анализа новостных информационных потоков / Д. В. Ландэ // Научные труды Донецкого национального технического университета. – Серия: Информатика, кибернетика и вычислительная техника (ИКВТ-2005). – Вып. 93. – Донецк: ДонНТУ, 2005. – С.277–287.
9. Ландэ Д. В. Основы интеграции информационных потоков / Д. В. Ландэ. – К.: Инжиниринг, 2006. – 240 с.
10. Мотылев В. М. Старение научно-технической литературы / В. М. Мотылев. – Л.: Наука, 1986.
11. Томашевський О. М. Інформаційні технології та моделювання бізнес-процесів : навч. посіб. / О. М. Томашевський, Г. Г. Цигелик, М. Б. Вітер, В. І. Дудук. – К.: Центр учбової літератури, 2012. – 296 с.

References:

1. Gianna M., Del Corso, Antonio, Gullí, Francesco, Romani (2005). Ranking a stream of news. *Proceedings of the 14th international conference on World Wide Web* [International World Wide Web Conference]. Chiba, Japan, p. 97–106.
2. Arnold, V.I. (2003). Analytics and prediction: mathematical aspects. *Nauchno-tehnicheskaya informatsiya [Scientific and Technical Information]*, ser. 1, no. 3. pp. 1–10 (in Russ.).
3. Braychevskiy, S.M., Lande, D.V. (2005). Modern Clearing streams: a crucial perspective. *Nauchno-tehnicheskaya informatsiya [Scientific and Technical Information]*, ser. 1. no. 11, pp. 21–33 (in Russ.).
4. Volterra, V. (1976). *Matematicheskaya teoriya borby za suschestvovanie* [The mathematical theory of the struggle for existence]. Nauka, Moscow (in Russ.).
5. Gorkova, V.I. (1988). Informetriya (quantitative methods in scientific and technical information). *Itogi nauki i tehniki [Results of Science and Technology]*, ser. Informatika [Information], vol. 10, VINITI, Moscow (in Russ.).
6. Efimov, A.N. (1978). *Informatsiya: tsennost, starenie, rasseyanie* [Information: value, aging, scattering], Moscow (in Russ.).
7. Ivanov, S.A., Krukovskaya N.V. (2004). Statistical analysis of documentary information flows. *Nauchno-tehnicheskaya informatsiya [Scientific and Technical Information]*, ser. 2, no. 2, pp. 11–14 (in Russ.).
8. Lande, D.V. (2005). Some methods for analyzing information flow of news. *Nauchnyie trudyi Donetskogo natsionalnogo tehnikeskogo universiteta. Seriya: Informatika, kibernetika i vychislitel'naya tehnika (IKVT-2005) [Transactions of Donetsk National Technical University. Series: Informatics, Cybernetics and Computer Science (IKVT 2005)]*, no. 93, DonNTU, Donetsk, pp. 277–287 (in Russ.).
9. Lande, D.V. (2006). *Osnovy integratsii informatsionnykh potokov* [Fundamentals of integration of information flows]. Kyiv, p. 240 (in Russ.).
10. Motylev, V. M. (1986). *Starenie nauchno-tehnicheskoy literatury* [Aging science and technology literature]. Nauka, Leningrad (in Russ.).
11. Tomashevskiy, O.M., Tsyheliy, H.H., Viter, M.B., Duduk, V.I. (2012). *Informatsiini tekhnologii ta modelivannia biznes-protsesiv* [Information technology and business process modeling]. Kyiv, p. 296 (in Ukr.).

УДК 530.16:004

М.К. Вишневська, к.т.н.,
Національна металургійна академія України,
м. Дніпропетровськ

**ПОБУДОВА МОДЕЛІ ПРОЕКТНО-ОРІЄНТОВАНОГО ЦЕНТРУ
ІННОВАЦІЙНОГО РОЗВИТКУ РЕГІОНУ**

Анотація

У роботі проаналізовано сучасні форми організаційних структур управління, переваги та недоліки бюрократичних та адаптивних організаційних структур. Зроблено висновок, що для підвищення ефективності функціонування сучасних організацій доцільним є впровадження гібридних різновидів організаційних структур управління, що містять у собі підсистеми різного типу. Розглянуто функціональну модель та модель роботи об'єкта дослідження, на базі яких розроблено структурну модель центру інноваційного розвитку. Остання більше придатна для виконання завдань державних програм інноваційного розвитку регіонів. Результати отримано на основі системного підходу із застосуванням методики функціонально-вартісного аналізу, зокрема проведення функціонально-ідеального моделювання. Теоретико-методологічною основою статті виступили концептуальні положення та наукові розробки вітчизняних і зарубіжних вчених з питань управління проектами та менеджменту.

Ключові слова: організаційна структура управління, проект, інноваційна програма розвитку, регіон, центр інноваційного розвитку регіону, функціонально-вартісний аналіз, структурна модель, функціональна модель.