

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені І.І.МЕЧНИКОВА

(повне найменування вищого навчального закладу)

Факультет математики, фізики та інформаційних технологій

(повне найменування інституту, назва факультету (відділення))

Кафедра математичного забезпечення комп'ютерних систем

(повна назва кафедри (предметної, циклової комісії))

Дипломна робота

на здобуття ступеня вищої освіти «магістр»

на тему «Методи аналізу та визначення залежностей в предметних
областях наукових конференцій»

«Methods of analysis and definition of dependencies in subject
areas of scientific conferences»

Виконала: студентка денної форми навчання
спеціальності 123 – Комп'ютерна інженерія
(шифр і назва напрямку підготовки, спеціальності)

Левенець Юлія Олександрівна

(прізвище, ім'я, по-батькові)

Керівник д-р техн. наук, проф. Малахов Є.В.

(науковий ступінь, вчене звання, прізвище та ініціали, підпис)

Рецензент канд. фіз.-мат. наук, доц. Петрушина Т.І.

(науковий ступінь, вчене звання, прізвище та ініціали)

Рецензент канд. техн. наук, доц. Шапорін Р.О.

(науковий ступінь, вчене звання, прізвище та ініціали)

Рекомендовано до захисту:

Протокол засідання кафедри

№ _____ від «___» _____ р.

Завідувач кафедри

Є.В. Малахов

(підпис)

(прізвище, ініціали)

Захищено на засіданні ЕК № _____

протокол № _____ від «___» _____ р.

Оцінка _____

(за 4-х бальною шкалою, за шкалою ECTS, бал.)

Голова ЕК

О.О. Арсірій

(підпис)

(прізвище, ініціали)

АНОТАЦІЯ

У дипломній роботі створена інформаційна технологія для аналізу інформації про наукові конференції.

Метою роботи є підвищення ефективності управління науковими конференціями на основі методів аналізу інформації про конференцію, її учасника та статтю, яка була відправлена. Отже, на підставі отриманої інформації буде можливість дізнатися, які тематики наукових конференцій характерні для тих чи інших країн та університетів, зокрема: наскільки тематика наукової конференції кафедри математичного забезпечення комп'ютерних систем Одеського національного університету ім. І.І. Мечникова корелюється з тематиками наукових конференцій інших університетів. Крім того, буде можливість дізнатися, наскільки актуальні на сьогоднішній день статі того чи іншого учасника та чи варто запрошувати його для участі в науковій конференції кафедри.

В результаті аналізу та застосування методів Data Mining визначено неявні залежності між даними низки актуальних конференцій та забезпечено автоматизацію аналізу інформації про певну наукову конференцію з метою її подальшого розвитку.

Результати дипломної роботи доповідались на п'ятнадцятій всеукраїнській конференції студентів і молодих науковців «Інформатика, інформаційні системи та технології».

АННОТАЦИЯ

В дипломной работе создана информационная технология для анализа информации о научных конференциях.

Целью работы является повышение эффективности управления научными конференциями на основе методов анализа информации о конференции, её участнике и статье, которая была отправлена. Таким образом, на основании полученной информации будет возможность узнать, какие тематики научных конференций характерны для тех или иных стран и университетов, в частности: насколько тематика научной конференции кафедры математического обеспечения компьютерных систем Одесского национального университета им. И.И. Мечникова коррелируется с тематиками научных конференций других университетов. Кроме того, будет возможность узнать, насколько актуальны на сегодняшний день статьи того или иного участника и стоит ли приглашать его для участия в научной конференции кафедры.

В результате анализа и применения методов Data Mining определены неявные зависимости между данными ряда актуальных конференций и обеспечено автоматизацию анализа информации про определённую научную конференцию с целью её дальнейшего развития.

Результаты дипломной работы докладывались на пятнадцатой всеукраинской конференции студентов и молодых ученых «Информатика, информационные системы и технологии».

ABSTRACT

The information technology of data analysis on scientific conferences was created in the research.

The aim of the research is to increase the effectivity of scientific conferences management basing on the methods of conference data analysis, its participant and the sent article. Thus, having considered the given data one can distinguish the peculiar topics of scientific conferences for certain countries and institutions in particular: the degree of correlation of the topic of a scientific conference between the Department of Mathematical Support of Computer Systems of I.I. Mechnikov Odessa National University and others universities. Besides one will be able to discover the relevance of the participant's article for the present day and whether it is worth inviting him/her to participate in a scientific conference of the Department. As the result of the analysis and the application of the methods of Data Mining hidden dependencies between the data of a row of relevant conferences were determined and the automation of a particular scientific conference data analysis was elaborated aiming its further improvement and development.

The findings of the research were reported on the fifteenth Ukrainian conference of students and young scientist "Informatics, Information Systems and Technologies".

ЗМІСТ

	стр.
ВСТУП	6
1 ОГЛЯД ЗАДАЧ І МЕТОДІВ DATA MINING.....	8
1.1 Задачі Data Mining.....	8
1.2 Методи Data Mining	10
2 ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДЛЯ АНАЛІЗУ ІНФОРМАЦІЇ ПРО НАУКОВІ КОНФЕРЕНЦІЇ.....	18
2.1 Постановка завдання.....	18
2.2 Послідовність методів аналізу інформації про наукові конференції ...	18
2.2.1 Статистичний показник TF-IDF	19
2.2.2 Метод k-найближчих сусідів	20
2.2.3 Алгоритм Apriori.....	22
3 ПРОЕКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ НАУКОВИХ КОНФЕРЕНЦІЙ	26
3.1 Архітектура інформаційної системи наукових конференцій.....	26
3.2 Моделювання інформаційної моделі предметної області	29
4 РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ НАУКОВИХ КОНФЕРЕНЦІЙ	33
4.1 Обґрунтування вибору засобів розробки	33
4.2 Створення інформаційної системи на ASP.NET MVC.....	34
5 АНАЛІЗ ДАНИХ ПРО НАУКОВІ КОНФЕРЕНЦІЇ	45
ВИСНОВКИ.....	52
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	53
ДОДАТОК А СТВОРЕННЯ БАЗИ ДАНИХ	55
ДОДАТОК Б СТВОРЕННЯ НАБОРІВ АЛГОРИТМУ APRIORI.....	56
ДОДАТОК В РЕАЛІЗАЦІЯ СТАТИСТИЧНОГО ПОКАЗНИКА TF-IDF.....	58
ДОДАТОК Г РЕАЛІЗАЦІЯ АЛГОРИТМУ APRIORI.....	61

Розвиток методів запису і зберігання даних привів до бурхливого зростання обсягів інформації, що збирається та аналізується. Обсяги даних настільки значні, що людині просто не під силу проаналізувати їх самостійно, хоча необхідність проведення такого аналізу цілком очевидна, адже ці «сирі» дані містять знання, які можуть бути використані при прийнятті рішень. Для того, щоб провести автоматичний аналіз даних, використовується Data Mining.

Data Mining – це процес виявлення в «сирих» даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань, необхідних для прийняття рішень в різних сферах людської діяльності. Data Mining є одним з кроків Knowledge Discovery in Databases [1].

Інформація, знайдена в процесі застосування методів Data Mining, повинна бути нетривіальною і раніше невідомою, наприклад, середні продажі не є такими. Знання повинні описувати нові зв'язки між властивостями, передбачати значення одних ознак на основі інших і т.д. Знайдені знання повинні бути застосовні і на нових даних з деяким ступенем вірогідності. Корисність полягає в тому, що ці знання можуть приносити певну вигоду при їх застосуванні.

Алгоритми, що використовуються в Data Mining, вимагають великої кількості обчислень. Раніше це було стримуючим фактором широкого практичного застосування Data Mining, проте сьогоднішнє зростання продуктивності сучасних процесорів зняло гостроту цієї проблеми. Тепер за прийнятний час можна провести якісний аналіз сотень тисяч і мільйонів записів.

В дипломній роботі, метою якої є підвищення ефективності управління науковими конференціями, необхідно розв'язати наступні задачі:

- розглянути існуючі методи аналізу даних та оцінити їх застосовність до досягнення поставленої мети;

- удосконалити існуючі методи чи створити інформаційну технологію, що дозволить провести аналіз специфічних даних;
- спроектувати та створити інформаційну систему, яка реалізуватиме запропоновані підходи до аналізу даних;
- провести аналіз інформації про конференцію, її учасника і статтю, яка була відправлена для того, щоб визначити неявні залежності між ними та виробити певні рекомендації керівникові конференції.

ВИСНОВКИ

Кафедри ОНУ імені І.І. Мечникова проводять досить багато конференцій різного рівня. Отже, реалізація інформаційної технології для аналізу інформації про наукові конференції є актуальною задачею.

Після аналізу переваг і недоліків вже існуючих методів інтелектуального аналізу даних, було прийнято рішення розробити інформаційну технологію для аналізу інформації про наукові конференції, яка ґрунтується на послідовному застосуванні модифікованих методів інтелектуального аналізу даних, що надаватиме відчутні переваги в визначенні неявних залежностей.

В результаті аналізу та застосування методів Data Mining визначено неявні залежності між даними низки актуальних конференцій, зокрема:

- відповідними напрямками, які представлені на конференції;
- відповідними авторами, які приймають участь в конференції;
- відповідними містами, учасники з яких приймають участь в конференції;
- відповідними університетами, учасники з яких приймають участь в конференції.

Крім того, забезпечено автоматизацію аналізу інформації про певну наукову конференцію з метою її подальшого розвитку. Результати дослідження обговорювались на п'ятнадцятій всеукраїнській конференції студентів і молодих науковців «Інформатика, інформаційні системи та технології» [17].

Подальший розвиток системи можливий в напрямку формування підстав для аналітичних звітів та виробітку певних рекомендацій керівникові конференції.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Fayyad U., Advances in Knowledge Discovery and Data Mining / U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy – published by AAAI Press, 1996. – 560 p.
2. Дюк В.А., Data Mining. Учебный курс / В.А. Дюк, А.П. Самойленко – СПб.: Питер, 2001. – 368 с.
3. Reinschmidt J., Intelligent Miner for Data: Enhance Your Business Intelligence / J. Reinschmidt, H. Gottschalk, H. Kim, D. Zwietering – published by International Technical Support Organization, 1999. – 211 p.
4. Эдельштейн Г. Интеллектуальные средства анализа, интерпретации и представления данных в информационных хранилищах // ComputerWeek Москва. 1996. № 16. – С. 32-33.
5. An introduction to kernel-based learning algorithms / Muller K.-R., Mika S., Ratsch G. etc. // IEEE Transactions on Neural Networks. 2001. V. 12. № 2. – P. – 181-201.
6. Chikering D.M., Learning Bayesian networks: The combination of knowledge and statistical data / D. M. Chikering, D. Geiger, D. Heckerman // Machine Learning. 1995. V. 20. № 3. – P. 197-243.
7. Brand E., Naive-Bayes and Nearest Neighbor / E. Brand, R. Gerritsen // DBMS. 1998. № 7.
8. Hackerman D. Bayesian Networks for Data Mining // Data Mining and Knowledge Discovery. 1997. V. 1. № 1. P. – 79-119.
9. Brin S., Dynamic itemset counting and implication rules for market basket data / S. Brin, R. Motwani, J.D. Ullman, S. Tsur // International Conference on Management of Data. 1997. V. 26. № 2. – P. 255-264.
10. Багаторівнева архітектура [Електронний ресурс] – Режим доступу: <https://metanit.com/sharp/mvc5/23.5.php>. – 07.12.2018.
11. Іванов О.Є. Віртуальна реальність // Історія філософії. Енциклопедія. – Мінськ, 2002. – С. 183-186.

12. Керівництво з C # 7.0 і .NET Framework 4.7 [Електронний ресурс] – Режим доступу: <https://metanit.com/sharp/tutorial/>. – 07.12.2018.
13. Керівництво з ASP.NET MVC 5 [Електронний ресурс] – Режим доступу: <https://metanit.com/sharp/mvc5/>. – 07.12.2018.
14. Керівництво з Entity Framework [Електронний ресурс] – Режим доступу: <https://metanit.com/sharp/entityframework/>. – 07.12.2018.
15. Паттерн Репозиторій [Електронний ресурс] – Режим доступу: <https://metanit.com/sharp/articles/mvc/11.php>. – 07.12.2018.
16. Вивчаємо ASP.NET MVC 4 [Електронний ресурс] – Режим доступу: <https://metanit.com/sharp/mvc/>. – 07.12.2018.
17. Левенець Ю.О. Методи аналізу та визначення залежностей в предметних областях наукових конференцій / Ю.О. Левенець, Є.В. Малахов // Тези доповідей XV всеукраїнської конференції студентів і молодих науковців «Інформатика, інформаційні системи та технології». – Одеса, 2018. – С. 138-139.