

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ
ФАКУЛЬТЕТ КОМП'ЮТЕРНО-ІНФОРМАЦІЙНИХ СИСТЕМ
І ПРОГРАМНОЇ ІНЖЕНЕРІЇ

ПОДОЛЯН ДМИТРО ОЛЕГОВИЧ

УДК 004.622

**МЕТОДИ І ЗАСОБИ АНАЛІЗУ ТА КЛАСИФІКАЦІЇ ТЕКСТОВОЇ
ІНФОРМАЦІЇ ДЛЯ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ОБСЛУГОВУВАННЯ
КОМП'ЮТЕРНИХ СИСТЕМ**

123 «Комп'ютерна інженерія»

Автореферат

дипломної роботи на здобуття освітнього ступеня «магістр»

Тернопіль 2018

Роботу виконано на кафедрі комп'ютерних систем та мереж Тернопільського національного технічного університету імені Івана Пулюя Міністерства освіти і науки України

Керівник роботи: кандидат технічних наук, доцент кафедри комп'ютерних систем та мереж
Яцишин Василь Володимирович,
Тернопільський національний технічний університет імені Івана Пулюя,

Рецензент: кандидат технічних наук, доцент кафедри автоматизації технологічних процесів та виробництв
Коноваленко Ігор Володимирович,
Тернопільський національний технічний університет імені Івана Пулюя

Захист відбудеться 27 грудня 2018 р. о 9⁰⁰ годині на засіданні екзаменаційної комісії №34 у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, навчальний корпус №1, ауд. 603

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми роботи. Складність сучасних комп'ютерних та інформаційних систем, велика кількість різнотипних і складноструктурованих даних вимагає від фірм-розробників створення служб технічної підтримки для забезпечення ефективності функціонування програмно-апаратних комплексів, аналітичних сервісів та задоволення потреб кінцевих споживачів послуг.

Кожен розробник в сфері ІТ технологій для формування і підтримки своєї репутації на ринку, повинен забезпечити обслуговування проданих товарів чи наданих послуг, а для цього йому необхідно організувати службу підтримки. Основними функціями таких служб є допомога користувачам або замовникам, аналіз та формалізація звернень і керування процесом їх виконання. На даний час служби підтримки користувачів організовані по-різному: починаючи від винайму офісу для прийому клієнтів, закінчуючи цілодобовими Call центрами та інтелектуальними системами підтримки користувачів. Однак такі організації вимагають значних затрат людських ресурсів, а відповідно і фінансових. Тому актуальною задачею у сфері підтримки користувачів, є розробка засобу автоматизації для опрацювання звернень користувачів із автоматичною класифікацією повідомлень щодо можливого інциденту, що виник при експлуатації комп'ютерних систем.

На сьогодні розроблено багато методів для вирішення задач автоматичної класифікації текстової інформації. Однак, ефективність їх застосування залежить від природи вхідних даних, області та способів використання. Обґрунтування і розробка методу класифікації повідомлень користувачів для забезпечення ефективності підтримки комп'ютерних систем є однією з нових задач і вимагає додаткового дослідження існуючих методів і засобів класифікації текстів.

Дослідженню методів класифікації текстової інформації присвячено ряд наукових робіт як українських, так і закордонних вчених, зокрема Дж. Мак-Кіна, Г. Болла, Д. Холла, Г. Ланса, У. Уільямсона, Н. Джардайна, Є.М. Бравермана, А.А. Дорофеюка, І.Б. Мучника, М.Г. Загоруйко та ін. Сьогодні ефективно застосовуються засоби класифікації текстової інформації розробленими компаніями Google, Yahoo та Yandex.

Однак, хоч і розроблено та апробовано ряд методів і засобів автоматичної класифікації текстової інформації, все ж актуальним залишається задача класифікації звернень користувачів комп'ютерних систем у вигляді текстових повідомлень для підвищення ефективності підтримки комп'ютерних систем.

Метою роботи є дослідження методів і засобів класифікації текстової інформації для підвищення ефективності підтримки комп'ютерних систем.

Для досягнення вказаної мети в роботі поставлено наступні **задачі**:

- аналіз наукових праць і практик реалізації методів і засобів класифікації текстової інформації;
- аналіз особливостей функціонування систем технічної підтримки користувачів при використанні комп'ютерних систем;
- обґрунтування методів класифікації текстової інформації для підвищення ефективності процесу підтримки комп'ютерних систем;

- розробка методу класифікації текстових повідомлень щодо підтримки комп'ютерних систем;
- розробка програмного засобу для підтримки користувачів комп'ютерних систем з інтеграцією функції автоматичної класифікації звернень користувачів.

Об'єкт дослідження: процес класифікації текстової інформації.

Предмет дослідження: методи і засоби класифікації текстової інформації в системах технічної підтримки користувачів комп'ютерних систем.

Методи дослідження. Для вирішення поставлених задач використано наступні методи: аналіз та узагальнення – при проведенні аналізу існуючих методів і засобів класифікації текстової інформації; теорії імовірності та математичної статистики, машинного навчання – при розробці методу класифікації повідомлень користувачів комп'ютерних систем; проектування та програмування – при розробці програмного засобу технічної підтримки користувачів комп'ютерних систем та розгортанні системи; експеримент і тестування – для апробації розробленого методу та програмного засобу.

Наукова новизна одержаних результатів:

- уперше запропоновано метод автоматичної класифікації текстової інформації в системах підтримки користувачів комп'ютерних систем на основі методів машинного навчання, зокрема нейронних мереж (з та без вчителя), що дало змогу скоротити часові і фінансові витрати, а також підвищити ефективність реагування на звернення користувачів комп'ютерних систем.

- набув подальшого розвитку метод проектування систем технічної підтримки користувачів комп'ютерних систем шляхом інтеграції методу автоматичної класифікації звернень, що дало змогу оптимізувати функціональність засобу автоматизації і скоротити людські ресурси служби технічної підтримки.

Практичне значення одержаних результатів. Впровадження методу класифікації текстової інформації і розробленого засобу підтримки комп'ютерних систем дає змогу підвищити ефективність обслуговування користувачів та функціонування самих систем.

Апробація. Результати дослідження апробовано на міжнародній науково-технічній конференції «Фундаментальні та прикладні проблеми сучасних технологій» до 100 річчя з дня заснування НАН України та на вшанування пам'яті Івана Пулюя (100 річчя з дня смерті) 22-24 травня 2018 р. та VII міжнародній науково - технічній конференції молодих учених і студентів «Актуальні задачі сучасних технологій» (28-29 листопада 2018 р.) Тернопільського національного технічного університету імені Івана Пулюя у вигляді тез конференцій.

Структура роботи. Робота складається з пояснювальної записки та графічної частини. Пояснювальна записка складається із вступу, 6 розділів, висновків, списку використаної літератури та додатків. Обсяг роботи: розрахунково-пояснювальна записка – 135 арк. формату А4, графічна частина – 10 аркушів формату А1.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність дослідження методів і засобів аналізу і класифікації текстової інформації для підвищення ефективності обслуговування комп'ютерних систем, визначено мету і задачі дипломної роботи магістра, наведено наукову новизну і практичну цінність одержаних результатів.

У першому розділі «Роль та місце методів і засобів класифікації текстової інформації при функціонуванні систем підтримки користувачів комп'ютерних систем» проведено аналіз наукових публікацій і практик організації систем технічної підтримки комп'ютерних систем і встановлено, що найбільш ефективною є трьохрівнева організація таких систем, однак у ній відсутня можливість автоматичної класифікації вхідних повідомлень для скорочення часу обслуговування звернень користувачів. Формалізовано постановку задачі класифікації текстової інформації при розробці систем підтримки користувачів комп'ютерних систем, що дало змогу в подальшому врахувати особливості автоматичної класифікації повідомлення до конкретного збою або несправності компонента комп'ютерної системи.

У другому розділі «Розробка методу класифікації текстової інформації для підвищення ефективності підтримки комп'ютерних систем» проведено аналіз існуючих підходів і методів класифікації текстової інформації і встановлено, що найбільш ефективним для задач автоматичної класифікації звернень користувачів комп'ютерних систем щодо інцидентів є підхід з використанням нейронних мережах, що дає змогу забезпечити необхідний рівень точності щодо приналежності повідомлення користувача до певного класу несправностей комп'ютерних систем, в порівнянні з іншими методами. Запропоновано метод автоматичної класифікації текстової інформації в системах підтримки користувачів комп'ютерних систем на основі методів машинного навчання, зокрема нейронних мереж (з та без вчителя), що дало змогу скоротити часові і фінансові витрати, а також підвищити ефективність реагування на звернення користувачів комп'ютерних систем.

У третьому розділі «Засіб автоматизації процесу класифікації текстової інформації для підвищення ефективності обслуговування комп'ютерних систем» на основі реляційного підходу спроектовано базу даних системи підтримки користувачів комп'ютерних систем з врахуванням особливостей методу автоматичної класифікації звернень щодо інцидентів, спроектовано клієнт-серверну архітектуру та реалізовано алгоритми навчання нейромережі засобу технічної підтримки користувачів комп'ютерних систем, що дало змогу визначити компоненти системи і типи зв'язків між ними.

У четвертому розділі «Обґрунтування економічної ефективності» на основі проведених розрахунків встановлено економічну доцільність дослідження методів і засобів класифікації текстової інформації для підвищення ефективності підтримки комп'ютерних систем.

У п'ятому розділі «Охорона праці та безпека в надзвичайних ситуаціях» проведено аналіз вимог з охорони праці і техніки безпеки при використанні засобу класифікації текстової інформації підтримки комп'ютерних систем, забезпечення

стійкості роботи вибухопожежонебезпечного об'єкту та проведення аварійно-відновлювальних робіт на комп'ютерних та електричних мережах.

У шостому розділі «Екологія» проаналізовано методи екологічної статистики і досліджено статистичне оцінювання техногенних впливів на навколишнє середовище.

У загальних висновках до дипломної роботи магістра наведено результати виконання розділів дипломної роботи магістра щодо застосування методів аналізу і класифікації текстової інформації для підвищення ефективності обслуговування комп'ютерних систем.

У додатках до пояснювальної записки наведено копії матеріалів щодо апробації одержаних у дипломній роботі магістра результатів.

У графічній частині до дипломної роботи магістра наведено основні теоретичні і практичні результати дослідження щодо застосування методів класифікації текстової інформації для підвищення ефективності обслуговування комп'ютерних систем.

ВИСНОВКИ

Основні наукові та практичні результати роботи полягають у наступному:

Проаналізовано сучасні підходи до проектування та реалізації систем технічної підтримки користувачів при експлуатації комп'ютерних систем у результаті якого виявлено основні їх недоліки та запропоновано шляхи оптимізації підтримки комп'ютерних систем на основі автоматичної класифікації повідомлень користувачів про необхідність усунення інцидентів.

Формалізовано постановку задачі класифікації текстової інформації при розробці систем підтримки користувачів комп'ютерних систем, що дало змогу в подальшому врахувати особливості автоматичної класифікації повідомлення до конкретного збою або несправності компонента комп'ютерної системи

Проведено аналіз існуючих підходів і методів класифікації текстової інформації і встановлено, що найбільш ефективним для задач автоматичної класифікації звернень користувачів комп'ютерних систем щодо інцидентів є підхід з використанням нейронних мережах, що дає змогу забезпечити необхідний рівень точності щодо приналежності повідомлення користувача до певного класу несправностей комп'ютерних систем, в порівнянні з іншими методами.

Уперше запропоновано метод автоматичної класифікації текстової інформації в системах підтримки користувачів комп'ютерних систем на основі методів машинного навчання, зокрема нейронних мереж (з учителем та без учителя), що дало змогу скоротити часові і фінансові витрати, а також підвищити ефективність реагування на звернення користувачів комп'ютерних систем.

Набув подальшого розвитку метод проектування систем технічної підтримки користувачів комп'ютерних систем шляхом інтеграції методу автоматичної класифікації звернень, що дало змогу оптимізувати функціональність засобу автоматизації і скоротити людські ресурси служби технічної підтримки.

На основі реляційного підходу спроектовано базу даних системи підтримки користувачів комп'ютерних систем з врахуванням особливостей методу

автоматичної класифікації звернень щодо інцидентів, що дало змогу забезпечити можливість зберігання та маніпулювання даними, а також накопичення даних для підвищення точності класифікації звернень користувачів.

Спроековано клієнт-серверну архітектуру та реалізовано алгоритми навчання нейромережі засобу технічної підтримки користувачів комп'ютерних систем, що дало змогу визначити компоненти системи і типи зв'язків між ними.

Розроблено інструкції з розгортання і налаштування системи підтримки користувачів комп'ютерних систем на основі IIS, що дало змогу масштабувати систему при зростанні навантаження на сервіс.

Обґрунтовано економічну ефективність від впровадження методу і засобу класифікації текстової інформації в системах підтримки користувачів комп'ютерних систем, зокрема термін окупності методу і засобу становить близько 2 років при собівартості робіт 58691,01 грн.

Проведено аналіз вимог з охорони праці і техніки безпеки при використанні засобу класифікації текстової інформації підтримки комп'ютерних систем, забезпечення стійкості роботи вибухопожежонебезпечного об'єкту та проведення аварійно-відновлювальних робіт на комп'ютерних та електричних мережах.

Проаналізовано методи екологічної статистики і досліджено статистичне оцінювання техногенних впливів на навколишнє середовище.

СПИСОК ОПУБЛІКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ РОБОТИ

1. Давидов А.О. Методи колаборативної фільтрації при побудові систем кроспродаж/ А.О. Давидов, В.В. Яцишин, Д.О. Подолян// – Матеріали міжнародної науково-технічної конференції «Фундаментальні та прикладні проблеми сучасних технологій» до 100 річчя з дня заснування НАН України та на вшанування пам'яті Івана Пулюя (100 річчя з дня смерті) 22-24 травня 2018 р. – Тернопіль, ТНТУ – 2018 – с. 243 – 244

2. Подолян Д.О. Класифікація та препроцесинг текстових даних/ Д.О. Подолян, В.В. Яцишин, А.О. Давидов // – Матеріали VII міжнародної науково - технічної конференції молодих учених і студентів «Актуальні задачі сучасних технологій» (28-29 листопада 2018 р.) Тернопільського національного технічного університету імені Івана Пулюя у вигляді тез конференцій. – Тернопіль, ТНТУ – 2018 – с. 201

АНОТАЦІЯ

Подолян Д.О. Методи і засоби аналізу та класифікації текстової інформації для підвищення ефективності обслуговування комп'ютерних систем.

Дипломна робота на здобуття освітнього ступеня магістра 123 – Комп'ютерна інженерія. – Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль 2018.

У дипломній роботі магістра проаналізовано сучасні підходи до проектування та реалізації систем технічної підтримки користувачів при експлуатації комп'ютерних систем у результаті якого виявлено основні їх недоліки та запропоновано шляхи оптимізації підтримки комп'ютерних систем на основі автоматичної класифікації повідомлень користувачів про необхідність усунення інцидентів.

Запропоновано метод автоматичної класифікації текстової інформації в системах підтримки користувачів комп'ютерних систем на основі методів машинного навчання, зокрема нейронних мереж (з та без вчителя), що дало змогу скоротити часові і фінансові витрати, а також підвищити ефективність реагування на звернення користувачів комп'ютерних систем.

Спроектовано клієнт-серверну архітектуру та реалізовано алгоритми навчання нейромережі засобу технічної підтримки користувачів комп'ютерних систем, що дало змогу визначити компоненти системи і типи зв'язків між ними. Розроблено інструкції з розгортання і налаштування системи підтримки користувачів комп'ютерних систем на основі ІІS, що дало змогу масштабувати систему при зростанні навантаження на сервіс.

Ключові слова: МЕТОД, ТЕКСТ, КЛАСИФІКАЦІЯ, КОМП'ЮТЕРНА СИСТЕМА, ЕФЕКТИВНІСТЬ.

ABSTRACT

Podolian D.O. Methods and tools of text information analysis and classification to raise the efficiency of computer systems service.

The diploma paper for obtaining the Master's degree 123 – Computer engineering – Ternopil Ivan Puluj National Technical University, Ternopil 2018.

The master's thesis analyzes the modern approaches to the design and implementation of technical support systems for the use of computer systems as a result of which their main shortcomings were identified and ways of optimizing the support of computer systems based on the automatic classification of users' messages about the need to eliminate incidents were proposed.

The method of automatic classification of text information in computer systems support systems based on the methods of machine learning, in particular neural networks (with and without the teacher) was proposed, which allowed to reduce time and financial expenses, as well as to increase the efficiency of responding to computer users' systems.

The client-server architecture was designed and the algorithms of training the neural network for technical support of users of computer systems were implemented, which made it possible to determine the components of the system and the types of links between them. Developed instructions for deploying and configuring IIS-based user support for computer systems, which allowed scaling the system with increasing service load.

Keywords: METHOD, TEXT, CLASSIFICATION, COMPUTER SYSTEM, EFFICIENCY.