

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

**ГАВАН РОМАН РУСЛАНОВИЧ**

УДК 004.89

**МЕТОДИ І ЗАСОБИ ВИЯВЛЕННЯ ТОНАЛЬНОСТІ ТЕКСТОВОЇ ІНФОРМАЦІЇ  
ДЛЯ ВДОСКОНАЛЕННЯ КОМП'ЮТЕРНИХ СИСТЕМ**

123 «Комп'ютерна інженерія»

**Автореферат**  
дипломної роботи на здобуття освітнього ступеня «магістр»

Тернопіль 2019

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 інформатики і математичного моделювання  
**Баран Ігор Олегович,**  
Тернопільський національний технічний університет імені Івана Пулюя

Захист відбудеться 23 грудня 2019 р. о 09<sup>00</sup> годині на засіданні екзаменаційної комісії №37 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.** Сучасний рівень розвитку інформаційних технологій зумовлює високу конкуренцію на ринку ІТ продуктів і послуг, а також змушує фірми-розробники автоматизувати практично усі бізнес-процеси. Наявність великої кількості даних та широке застосування методів інтелектуального аналізу текстової та графічної інформації, наявність апаратних і програмних засобів зберігання та опрацювання «великих даних» дає змогу майже в автоматичному режимі забезпечувати процеси комунікації між замовниками та розробниками, визначати переваги і недоліки наданих послуг чи проданих товарів, порівнювати конкурентоздатність продуктів, підвищити супровід комп'ютерних систем та їх вдосконалення.

Одним із найбільш перспективних напрямів, що сьогодні розвивається, є інтелектуальний аналіз відгуків користувачів про придбані товари, про використання тих чи інших апаратних засобів, програмного забезпечення і комп'ютерних систем в цілому. Розвиток цього напрямку зумовлений орієнтацією фірм-розробників на задоволеність користувачів і визначення конкурентних переваг чи недоліків в порівнянні з іншими продуктами. Визначення тональності відгуків користувачів про комп'ютерні системи та їх складові дає змогу визначити недоліки, які в подальшому будуть враховані при оновленні компонентів комп'ютерної системи. При цьому використовується методи семантичного та синтаксичного аналізу тексту, нейронні мережі, методи класифікації та кластеризації даних, що з певною точністю дозволяють визначити приналежність висловленої користувачем думки до одного з класів: позитивний, негативний чи нейтральний.

Дослідженню методів аналізу тональності і 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 аналізу тональності тексту, розроблені фірмами Google, Yahoo та Yandex.

На сьогодні розроблено багато методів для вирішення задач щодо визначення тональності тексту. Однак, ефективність їх застосування залежить від природи вхідних даних, області та способів використання. Обґрунтування і розробка методу визначення тональності текстової інформації при вдосконаленні комп'ютерних систем є однією з нових та актуальних задач і вимагає додаткового дослідження існуючих методів і засобів.

**Метою дипломної роботи** є дослідження методів і засобів визначення тональності текстової інформації для вдосконалення комп'ютерних систем.

Для досягнення цієї мети у дипломній роботі магістра були поставлені і вирішені наступні **задачі**:

- аналіз наукових праць і практик реалізації методів і засобів виявлення та аналізу тональності текстової інформації;
- аналіз особливостей опрацювання відгуків користувачів при вдосконаленні комп'ютерних систем;

- обґрунтування інтелектуальних методів аналізу та визначення тональності у відгуках користувачів про комп'ютерні системи;
- розробка інтелектуального методу виявлення тональності відгуків користувачів;
- розробка програмного засобу виявлення та оцінювання тональності текстової інформації при супроводі комп'ютерних систем.

**Об'єктом дослідження** є процес виявлення тональності текстової інформації.

**Предметом дослідження** методи і засоби виявлення та аналізу тональності текстової інформації.

**Методи дослідження.** Для вирішення, поставлених у роботі магістра задач, використано наступні методи:

- аналіз та узагальнення – при проведенні аналізу існуючих методів і засобів виявлення тональності текстової інформації;
- теорії імовірності та математичної статистики, машинного навчання – при розробці методу виявлення та аналізу тональності відгуків користувачів комп'ютерних систем;
- проектування та програмування – при розробці програмного засобу збору, аналізу та виявлення тональності текстової інформації при супроводі комп'ютерних систем;
- експеримент і тестування – для апробації розробленого методу та програмного засобу.

**Наукова новизна одержаних результатів:**

- уперше запропоновано архітектурне рішення та спроектовано базу даних програмного засобу, що оперує поняттями лінгвістично-інженерного підходу та машинного навчання і дає змогу автоматизувати процес оцінювання тональності текстової інформації та приймати рішення щодо вдосконалення властивостей комп'ютерних систем;
- набули подальшого розвитку методи і засоби аналізу текстової інформації при вдосконаленні комп'ютерних систем, в основі якого лежить поєднання лінгвістично-інженерного підходу та машинного навчання у контексті застосування шаблонів синтаксичних конструкцій, методів машинного навчання з учителем, що дає змогу виявляти та оцінювати аспекти комп'ютерних систем із заданою точністю і повнотою.

**Практична цінність результатів дослідження.** Запропонований підхід виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем впроваджено у вигляді програмного інструменту, який дає змогу виявляти аспекти у відгуках користувачів, групувати їх та кількісно виражати ступінь задоволеності користувачів щодо реалізованих у комп'ютерній системі властивостей.

**Публікації.** Результати дослідження апробовано на VIII міжнародній науково-технічній конференції молодих учених і студентів «Актуальні задачі сучасних технологій» (27-28 листопада 2019 р.) Тернопільського національного технічного університету імені Івана Пулюя та на VII науково-технічній конференції Тернопільського національного технічного університету імені Івана Пулюя

«Інформаційні моделі, системи та технології» (11-12 грудня 2019 року) у вигляді тез конференцій.

**Структура роботи.** Робота складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається із вступу, 6 розділів, висновків, переліку посилань та додатків. Обсяг роботи: розрахунково-пояснювальна записка – 125 арк. формату А4, графічна частина – 12 аркушів формату А1.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

**У вступі** обґрунтовано актуальність задач розробки та застосування методів і засобів виявлення, аналізу та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем, сформульовано мету, задачі і методи дослідження, наведено наукову новизну і практичне значення одержаних результатів.

**У першому розділі «Аналіз сучасного стану досліджень в області аналізу тональності текстової інформації при вдосконаленні комп'ютерних систем»** досліджено джерела одержання інформації щодо властивостей комп'ютерних систем та виявлено, що для їх вдосконалення можуть використовуватись відгуки користувачів, які наявні у відкритому доступі та інциденти, що зберігаються у спеціалізованих комп'ютерних системах підтримки користувачів. Проведено аналіз методів комп'ютерної лінгвістики та прикладні аспекти їх застосування, що дало змогу визначити можливі шляхи впровадження методів машинного навчання, препроцесингу текстової інформації на різних рівнях представлення тексту, а також застосування лінгвістично-інженерного підходу при побудові правил для опису шаблонів за частинами мовами.

**У другому розділі «Побудова моделей та обґрунтування методів і метрик оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем»** формалізовано представлення відгуків користувачів і задач, які при цьому необхідно розв'язати, у вигляді квантилю (сутність-аспект-тональність-користувач-час), що дало змогу врахувати позитивні і негативні властивості, реалізовані у комп'ютерних системах та в подальшому відобразити їх при вдосконаленні систем. Обґрунтовано необхідність застосування TF-IDF методів для виявлення та групування аспектів комп'ютерних систем, що позначаються на враженні кінцевих користувачів, обґрунтовано необхідність використання методів машинного навчання в комплексі з частотними методами синтаксичного аналізу при вирішенні задач вдосконалення комп'ютерних систем. Обґрунтовано застосування лінгвістично-інженерного підходу при оцінюванні тональності текстової інформації щодо вдосконалення комп'ютерних систем, що дає змогу врахувати якість аспектів на основі слів, що їх оцінюють та відстані між аспектами і такими словами. Встановлено та обґрунтовано доцільність застосування F-міри, що враховує повноту і точність оцінювання тональності відгуків користувачів і дає змогу виявити найбільш проблемні властивості комп'ютерної системи.

У третьому розділі «Система виявлення та аналізу тональності текстової інформації при вдосконаленні комп'ютерних систем» на основі UML діаграм, визначено функціональні вимоги та ролі користувачів при автоматизації процесу виявлення та оцінювання тональності текстової інформації, спроектовано схему бази даних для зберігання та маніпулювання даними щодо оцінювання тональності текстової інформації (відгуків користувачів), що дало змогу забезпечити гнучкість та масштабованість процесу вдосконалення комп'ютерних систем, а також спроектовано архітектуру системи виявлення та оцінювання тональності відгуків користувачів про комп'ютерну систему або її компоненти, що дало змогу реалізувати програмні модулі керування параметрами алгоритмів і моделей машинного навчання.

У четвертому розділі «Обґрунтування економічної ефективності» обчислено показники економічної ефективності науково-дослідної роботи щодо дослідження методів і засобів виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем. Встановлено, що собівартість розробленого методу і засобу становить 37571,03 грн., а термін окупності 1,87 року. Тому впровадження такої розробки є економічно обґрунтованою.

У п'ятому розділі «Охорона праці та безпека в надзвичайних ситуаціях» проведено аналіз вимог та норм охорони праці і техніки безпеки при експлуатації комп'ютерної техніки. Визначено особливості організації робочих місць, вимог з електробезпеки, природного та штучного освітлення для ефективної і безпечної роботи аналітиків з оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем.

Окрім цього, проведено аналіз способів запобігання забрудненню повітря, виробничих приміщень НХР, допустимі значення для основних речовин-забруднювачів; їх характеристика та засоби захисту, а також методи захисту від дії ЕМІ, що базується на врахуванні його можливого негативного впливу.

У шостому розділі «Екологія» проаналізовано питання зниження енергоємності та енергозбереження, а також вимог до мікроклімату, вмісту аеронів і шкідливих хімічних речовин у повітрі приміщень експлуатації моніторів і ПЕОМ, шляхи дотримання цих вимог.

У загальних висновках до дипломної роботи магістра наведено основні результати виконання розділів дипломної роботи магістра, їх наукове та практичне значення щодо методів і засобів виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем.

У додатках до пояснювальної записки наведено копії матеріалів конференцій, у яких опубліковано основні результати дипломної роботи магістра.

У графічній частині до дипломної роботи магістра наведено одержані наукові та практичні результати щодо методів і засобів виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем.

## ВИСНОВКИ

У дипломній роботі магістра досліджено методи і засоби виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем, у результаті якого одержано наступні наукові і практичні результати.

Проведено аналіз відкритих та спеціалізованих джерел одержання інформації для вдосконалення комп'ютерних систем у результаті якого визначено шляхи збору та аналізу текстових даних, їх інтерпретації та подальшого прийняття рішень щодо модернізації складових компонентів або системи в цілому.

Проаналізовано сучасний стан розвитку методів і засобів комп'ютерної лінгвістики на різних рівнях представлення текстової інформації та обґрунтовано необхідність застосування комплексного підходу щодо виявлення та аналізу тональності текстової інформації на синтаксичному, лексичному та семантичному рівнях.

Досліджено прикладні аспекти застосування комп'ютерної лінгвістики, що дало змогу визначити та обґрунтувати методи добування та аналізу відгуків користувачів про властивості комп'ютерних систем і її поведінки на основі аспектного підходу.

Формалізовано представлення відгуків користувачів і задач, які при цьому необхідно розв'язати, у вигляді квантилю (сутність-аспект-тональність-користувач-час), що дало змогу врахувати позитивні і негативні властивості, реалізовані у комп'ютерних системах та в подальшому відобразити їх при вдосконаленні систем.

Обґрунтовано необхідність застосування частотних методів при виявленні аспектів комп'ютерних систем, що позначаються на враженні кінцевих користувачів і дають змогу виявити найбільш важливі властивості комп'ютерних систем.

Проведено аналіз методів добування аспектів на основі підходу побудови правил, машинного навчання і тематичного моделювання, у результаті якого обґрунтовано необхідність використання методів машинного навчання в комплексі з частотними методами синтаксичного аналізу при вирішенні задач вдосконалення комп'ютерних систем.

Обґрунтовано застосування лінгвістично-інженерного підходу при оцінюванні тональності текстової інформації щодо вдосконалення комп'ютерних систем, що дає змогу врахувати якість аспектів на основі слів, що їх оцінюють та відстані між аспектами і такими словами.

Проаналізовано методи і метрики оцінювання тональності текстової інформації і встановлено доцільність застосування F-міри, що враховує повноту і точність оцінювання тональності відгуків користувачів і дає змогу виявити найбільш проблемні властивості комп'ютерної системи.

При розробці системи виявлення та аналізу тональності текстової інформації для вдосконалення комп'ютерних систем побудовано діаграми варіантів використання, що дало змогу визначити функціональні вимоги до системи та ролі користувачів.

Розроблено схему бази даних для зберігання та маніпулювання даними щодо оцінювання тональності текстової інформації (відгуків користувачів), що дало змогу забезпечити гнучкість та масштабованість процесу вдосконалення комп'ютерних систем.

Спроектовано архітектуру системи виявлення та оцінювання тональності відгуків користувачів про комп'ютерну систему або її компоненти, що дало змогу реалізувати програмні модулі керування параметрами алгоритмів і моделей машинного навчання.

Обґрунтовано економічну доцільність щодо впровадження запропонованих методів і засобу виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем шляхом обчислення показників економічної ефективності, зокрема собівартості, ціни і терміну окупності, який становить 1,87 року.

Проаналізовано вимоги з охорони праці та безпеки в надзвичайних ситуаціях, що дало змогу визначити шляхи мінімізації негативного впливу комп'ютерної техніки на користувачів програмного засобу виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем.

Проаналізовано методи і заходи щодо зниження енергоємності та енергозбереження, а також проведено аналіз вимог до мікроклімату у приміщеннях, де експлуатуються ПК.

#### **СПИСОК ОПУБЛКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ РОБОТИ**

1. Гаван Р.Р., Яцишин В.В. Підходи до вдосконалення комп'ютерних систем на основі аналізу тональності відгуків користувачів. Матеріали VIII Міжнародної науково-технічної конференції молодих учених та студентів «Актуальні задачі сучасних технологій» (27 - 28 листопада 2019 р.). Тернопіль. 2019. с. 16.

2. Яцишин В.В., Гаван Р.Р. Особливості аналізу тональності тексту на основі аспектів при вдосконаленні комп'ютерних систем. Матеріали VII науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя «Інформаційні моделі, системи та технології» (11-12 грудня 2019 року). Тернопіль, ТНТУ. 2019. с. 114.



## АНОТАЦІЯ

### **Гаван Р.Р. Методи і засоби виявлення тональності текстової інформації для вдосконалення комп'ютерних систем**

Дипломна робота на здобуття освітнього ступеня магістра 123 – Комп'ютерна інженерія. – Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль 2019.

У дипломній роботі досліджено методи і засоби виявлення та оцінювання тональності текстової інформації при вдосконаленні комп'ютерних систем на основі методів машинного навчання та методів лінгвістично-інженерного підходу, розроблено засіб автоматизації при виявленні та оцінюванні тональності на основі аспектного підходу.

Формалізовано представлення відгуків користувачів і задач, які при цьому необхідно розв'язати, у вигляді квантилю (сутність-аспект-тональність-користувач-час), що дало змогу врахувати позитивні і негативні властивості, реалізовані у комп'ютерних системах та в подальшому відобразити їх при вдосконаленні систем. Обгрунтовано необхідність застосування частотних методів при виявленні аспектів комп'ютерних систем, що позначаються на враженні кінцевих користувачів і дають змогу виявити найбільш важливі властивості комп'ютерних систем. Обгрунтовано застосування лінгвістично-інженерного підходу при оцінюванні тональності текстової інформації щодо вдосконалення комп'ютерних систем, що дає змогу врахувати якість аспектів на основі слів, що їх оцінюють та відстані між аспектами і такими словами.

Розроблено програмний засіб автоматизації процесу виявлення та оцінювання тональності текстової інформації. Спроектовано схему бази даних для зберігання та маніпулювання даними щодо оцінювання тональності текстової інформації (відгуків користувачів), що дало змогу забезпечити гнучкість та масштабованість процесу вдосконалення комп'ютерних систем. Спроектовано архітектуру системи виявлення та оцінювання тональності відгуків користувачів про комп'ютерну систему або її компоненти, що дало змогу реалізувати програмні модулі керування параметрами алгоритмів і моделей машинного навчання.

**Ключові слова:** ТОНАЛЬНІСТЬ, ТЕКСТ, ВИЯВЛЕННЯ, КОМП'ЮТЕРНА СИСТЕМА, ВДОСКОНАЛЕННЯ, МЕТОД, ЗАСІБ.

## ABSTRACT

**Havan R.R.** Methods and tools of text information tonality detection aimed at computer systems improvement

The diploma paper for obtaining the Master's degree 123 – Computer engineering – Ternopil Ivan Puluj National Technical University, Ternopil, 2019.

The diploma thesis investigates the methods and means of detecting and evaluating the tonality of textual information in the improvement of computer systems based on machine learning methods and methods of linguistic-engineering approach, the automation tool for the detection and evaluation of tonality based on the aspectual approach.

The presentation of user feedback and tasks that need to be solved was formalized in the form of a quantile (essence-aspect-tone-user-time), which allowed to take into account the positive and negative properties realized in computer systems and to further reflect them when improving systems. The necessity of applying frequency methods to identify aspects of computer systems that affect the end-user experience and reveal the most important properties of computer systems is justified. The use of a linguistic-engineering approach in evaluating the tonality of textual information for the improvement of computer systems is justified.

A software tool has been developed to automate the process of detecting and evaluating the tonality of textual information. A database schema has been designed to store and manipulate the tonality of textual information (user feedback) to allow for the flexibility and scalability of the process of improving computer systems. The architecture of the system for detecting and evaluating the tone of user feedback about the computer system or its components was designed, which made it possible to implement software modules for controlling the parameters of algorithms and machine learning models.

**Keywords:** SENTIMENT, TEXT, EXTRACTION, COMPUTER SYSTEM, IMPROVING, METHOD, TOOL.