Міністерство освіти і науки України

Тернопільський НАЦІОНАЛЬНИЙ технічний Університет

імені Івана Пулюя

ФАКУЛЬТЕТ КОМП’ЮТЕРНО-ІНФОРМАЦІЙНИХ СИСТЕМ ТА ПРОГРАМНОЇ ІНЖЕНЕРІЇ

КАФЕДРА КОМП’ЮТЕРНИХ НАУК

**шкЛЯРУК МАРІЯ ВОЛОДИМИРІВНА**

УДК 004.912

**КОМПАРАТИВНИЙ АНАЛІЗ АЛГОРИТМІВ ТА ЗАСОБІВ ВІДМІНЮВАННЯ ВЛАСНИХ ІМЕННИКІВ В УКРАЇНСЬКІЙ МОВІ**

8.05010101 «Інформаційні управляючі системи та технології»

**Автореферат**

дипломної роботи на здобуття освітнього ступеня «магістр»

Тернопіль

2017

|  |
| --- |
| Роботу виконано на кафедрі комп’ютерних наук Тернопільського національного технічного університету імені Івана Пулюя Міністерства освіти і науки України |
| **Керівник роботи:** | кандидат технічних наук, доцент кафедри комп’ютерних наук**Козак Руслан Орестович,**Тернопільський національний технічний університет імені Івана Пулюя,  |
| **Рецензент:** |  |

Захист відбудеться 21 лютого 2017 р. о 9.00 годині на засіданні екзаменаційної комісії №1 у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, навчальний корпус №1, ауд. 701.

**ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ**

**Актуальність теми роботи**. Комп’ютеризація та інформатизація всіх сфер людського життя вимагає нових підходів до обробки інформації та нових можливостей від комп’ютерних систем. Аналіз та розвиток засобів відмінювання власних іменників є перспективними завданнями в галузі штучного інтелекту. Вони вирішують завдання комп’ютерної лінгвістики, яка є напрямом штучного інтелекту, що останніми роками активно розвивається.

**Мета роботи:** порівняння результатів роботи розробленого та стороннього сервісу та проведення покращення роботи розробленого сервісу відмінювання.

**Об’єкт, методи та джерела дослідження.** Основним об’єктом дослідження є алгоритми та засоби відмінювання власних іменників в українській мові.

**Наукова новизна отриманих результатів:**

* досліджено існуючі підходи та способи відмінювання власних іменників в українській мові;
* досліджено можливості, функції та технічні характеристики програмного засобу «Морфер»;
* проаналізовано результати відмінювання тестового набору імен за допомогою сервісу відмінювання та веб-сервісу «Морфер»;
* запропоновано загальний алгоритм відмінювання власних іменників в українській мові;
* сформовано основні категорії власних іменників, алгоритмічне відмінювання яких є найскладнішим;
* розроблено структуру допоміжної для відмінювання бази даних складних випадків відмінювання.

**Практичне значення отриманих результатів.**

Загальний алгоритм відмінювання власних іменників є методологічною основною для розробки засобів відмінювання, застосування в процесі відмінювання додаткової бази даних дозволяє отримувати кращі результати відмінювання при відсутності додаткової інформації про іменники. Отримані результати дослідження призначені для використання в галузі комп’ютерної лінгвістики та системах документообігу.

**Апробація.** Окремі результати роботи доповідались на VIII Всеукраїнській студентській науково-технічній конференції «Природничі та гуманітарні науки. Актуальні питання», Тернопіль, ТНТУ, 23 – 24 квітня 2015 р.та IX Всеукраїнській студентській науково-технічній конференції «Природничі та гуманітарні науки. Актуальні питання», Тернопіль, ТНТУ, 20 – 21 квітня 2016 р.

**Структура роботи.** Робота складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається з вступу, 7 частин, висновків, переліку посилань та додатків. Обсяг роботи: розрахунково-пояснювальна записка – 115 арк. формату А4, графічна частина – 7 аркушів формату А1

**ОСНОВНИЙ ЗМІСТ РОБОТИ**

**У вступі** проведено огляд методів та принципів автоматичного опрацювання тексту та охарактеризовано основні завдання, які необхідно вирішити.

**В розділі «Аналіз існуючих підходів та способів відмінювання власних імнників»** проведено аналіз існуючих методів комп’ютерної лінгвістики для автоматичного опрацювання тексту, розглянуто підходи до відмінювання власних іменників, а саме словниковий та безсловниковий підходи, проаналізовано можливості частини програми «Морфер», яка призначена для відмінювання прізвищ, імен та по батькові української мови, та розробленого сервісу відмінювання.

**В розділі «Порівняння результатів роботи розробленого сервісу відмінювання та веб-сервісу “Морфер”»** проведено порівняльний аналіз результатів відмінювання тестового набору прізвищ, імен та по батькові програмою «Морфер» та сервісом відмінювання, проаналізовано роботу тестованих програм на виокремлених типових групах іменників і спеціальних випадках відмінювання, які не підпадають під загальні правила.

**В розділі «Покращення алгоритмів розробленого сервісу відмінювання»** описано методи та способи покращення роботи сервісу відмінювання та проведено його вдосконалення шляхом додання валідації даних на стороні бібліотеки відмінювання, додання можливості відмінювати подвійні імена та по батькові. Також у розроблений сервіс відмінювання додано базу даних складних випадків відмінювання, яка дозволяє отримувати правильні відмінки імен, для відмінювання яких за розробленими алгоритмами необхідна додаткова інформація.

**В спеціальній частині** проведено аналіз тестування програмного забезпечення як способу забезпечення якості, розглянуто види (функціональне, нефункціональне, пов’язане зі змінами) та рівні (компонентне або модульне, інтеграційне, системне, приймальне) тестування програмного забезпечення.

**В розділі «Обґрунтування економічної ефективності»** розраховано основні техніко–економічні показники дослідження, розраховано значення економічної ефективності та термін окупності.

**В розділі** «**Охорона праці та безпека в надзвичайних ситуаціях»** описано вимоги до робочого місця при роботі з ПК, заходи із забезпечення захисту від впливу іонізуючого випромінювання та заходи для підвищення стійкості роботи комп’ютеризованих систем в умовах дії ЕМІ ядерних вибухів.

**В розділі «Екологія»** проаналізовано методи дисперсного аналізу та кореляційного аналізу зв’язків в екології. Було розглянуто завдання, які згадані методи вирішують, основні поняття та їх значення в екології. Також було досліджено особливості кореляційного аналізу та основні етапи кореляційно-регресійного аналізу.

**У загальних висновках щодо дипломної роботи** описано результати компаративного аналізу функціоналу та можливостей веб-сервісу «Морфер» та розробленого сервісу відмінювання, заходи для покращення роботи розробленого сервісу відмінювання, техніко-економічні показники та їх порівняння з базовими.

В графічній частині приведено порівняння результатів відмінювання за допомогою аналізованих сервісів, загальний алгоритм відмінювання з доданими покращеннями, структуру бази даних складних випадків відмінювання.

**ВИСНОВКИ**

Результатом виконання дипломної роботи є проведений компаративний аналіз алгоритмів відмінювання власних іменників в українській мові та засобів відмінювання, які ці алгоритми реалізовують. Серед методів відмінювання було проаналізовано відмінності між формальним та неформальним опрацюванням, можливості та способи реалізації словникового та безсловникового підходів до відмінювання власних іменників. Серед засобів відмінювання було проаналізовано функціонал, можливості та проблеми у відмінюванні веб-сервісу «Морфер», який є частиною програмного продукту «Морфер», та розробленого сервісу відмінювання.

Проведено відмінювання створеного тестового набору українських прізвищ, імен та по батькові за допомогою двох сервісів відмінювання. Результати відмінювання проаналізовано на правильність та наявність помилок. Для виявлення проблем у відмінюванні результати проаналізовано як загалом, так і за окремими групами іменників: відмінками, групами, родами.

На основі аналізу результатів відмінювання було визначено та впроваджено заходи для покращення роботи розробленого сервісу відмінювання. Серед покращень було додано валідацію даних, удосконалено алгоритм відмінювання подвійних імен та прізвищ та додано базу даних складних випадків відмінювання. Наявність такої бази дозволить постійно збільшувати відсоток правильно провідміняних імен за рахунок додавання в базу імен, відмінювання яких за алгоритмами потребує додаткової інформації.

**СПИСОК ОПУБЛІКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ РОБОТИ**

1. Шклярук М. Модель реалізації алгоритму відмінювання власних іменників в українській мові [Текст] / Шклярук М., Карнаухов О., Козак Р.О. Тези доповіді на VIII Всеукраїнській студентській науково-технічній конференції «Природничі та гуманітарні науки. Актуальні питання». – Тернопіль, ТНТУ, 2015. – с. 113.
2. Шклярук М. Проектування лінгвістичних баз даних [Текст] / Шклярук М., Карнаухов О., Козак Р.О. Тези доповіді на IX Всеукраїнській студентській науково-технічній конференції «Природничі та гуманітарні науки. Актуальні питання». – Тернопіль, ТНТУ, 2016. – с. 126.

**АНОТАЦІЯ**

В дипломній роботі проведено компаративний аналіз алгоритмів та засобів відмінювання прізвищ, імен та по батькові в українській мові. Незважаючи на велику кількість досліджень методів та способів відмінювання власних назв в українській мові, готових для використання засобів відмінювання практично немає. З цього можна зробити декілька висновків. Першим висновком є те, що реалізація досліджуваних алгоритмів відмінювання є достатньо складною і потребує великої кількості як лінгвістичних знань, так і знань в галузі розробки програмного забезпечення. Також можна зробити висновок про те, що в Україні засоби автоматичного відмінювання є не настільки затребуваними, щоб дослідження в цій галузі виходили на достатньо високий рівень. Проте це є лише питанням часу, оскільки інформаційні технології розвиваються дуже швидко і можна припустити, що використання систем, де буде використовуватись автоматичне відмінювання, наприклад, систем електронного документообігу, почне швидко розповсюджуватись.

Серед готових рішень для відмінювання іменників української мови можна виділити програмний продукт «Морфер». Основними функціями даної програми є відмінювання слів та словосполучень російської мови, а відмінювання прізвищ, імен та по батькові української мови є додатковими можливостями.

В роботі було проведено порівняння теоретичних методів та способів відмінювання власних іменників в українській мові, а також практичні результати відмінювання великого тестового набору даних за допомогою програми «Морфер» та розробленого сервісу відмінювання.

На основі результатів відмінювання зроблено висновки про точність і правильність роботи обох програм, а також проведено вдосконалення розробленого сервісу відмінювання.

**Ключові слова**: ЛІНГВІСТИКА, КОМПАРАТИВНИЙ АНАЛІЗ, ВІДМІНЮВАННЯ ІМЕННИКІВ, АЛГОРИТМ, ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ПРОГРАМНА БІБЛІОТЕКА, ІНТЕРНЕТ, ІНТЕРФЕЙС, ПОРІВНЯННЯ РЕЗУЛЬТАТІВ.

**ANNOTATION**

Own noun declension task belongs to the problems of automatic text processing. It's being solved by methods of formal processing. There are two basic principles of producing morphological characteristics of a word by its structure in modern automatic morphological analysis systems. There are obtaining grammatical information from the word by his grapheme analysis and presentation of grammatical information in the dictionary of fundamentals and dictionary of inflections (parts of declined words that express grammatical meaning and perform the inflectional function). The first principle is based on grammatical features contained in the final combination of letters, and the second - in grammar information from the dictionary of fundamentals and dictionary of inflections.

The thesis considers issues of declension algorithms and tools comparative analysis in Ukrainian.

In the thesis the existing noun declension approaches and methods in Ukrainian have been researched, features, functions and technical specifications of the "Morfer" software have been examined, the declension algorithm that used in the developed declension service has been described, results of declination of names test set using the declension service and "Morfer" web service have been analyzed, the developed declension service through algorithms improving and a database of complex declension cases adding has been improved.

**Key words:** COMPUTATIONAL LINGUISTICS, COMPARATIVE ANALYSIS, NOUN DECLENSION, ALGORITHM, INFORMATIONAL TECHNOLOGIES, SOFTWARE LIBRARY, INTERNET, INTERFACE, RESULTS COMPARISON.