Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

(повна назва)

Кафедра Програмної інженерії (повна назва)

**АТЕСТАЦІЙНА РОБОТА**

**Пояснювальна записка**

другий (магістерський)

(рівень вищої освіти)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(позначення документа)

Дослідження проблеми омонімії при побудові гіперланцюгів лексичних одиниць природної мови

(тема)

Виконав: студент 2 курсу, групи ПЗСм-16-2

спеціальності 121 – Інженерія програмного забезпечення

(код і повна назва спеціальності)

спеціалізації Програмне забезпечення систем

(повна назва спеціалізації)

Тищенко О. О.

(прізвище, ініціали)

Керівник проф. Четвериков Г.Г.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри \_\_\_\_\_\_\_\_\_\_\_ \_\_\_Дудар З.В.\_\_\_\_\_\_\_\_

(підпис) (прізвище, ініціали)

2018 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

Кафедра Програмної інженерії

Рівень вищої освіти\_\_\_\_\_\_\_\_\_\_\_другий (магістерський)

Спеціальність 121 – Інженерія програмного забезпечення

(код і повна назва)

Спеціалізація Програмне забезпечення систем

(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис)

«\_\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 20 \_\_\_ р.

**ЗАВДАННЯ**

НА АТЕСТАЦІЙНУ РОБОТУ

студентові Тищенко Олені Олександрівні

(прізвище, ім’я, по батькові)

1. Тема проекту (роботи) *«Дослідження проблеми омонімії при побудові гіперланцюгів лексичних одиниць природної мови»* затверджена наказом університету №451 Ст від «16» квітня 2018 р.
2. Термін здачі студентом закінченої роботи «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_ 2018 р.
3. Вихідні дані до проекту (роботи): *В програмній системі побудови гіперланцюгів передбачити: збір інформації про зв’язки між словами, можливість редагування зібраної інформації, механізм виводу гіперланцюгів користувачу, вибір коректного омонімічного терміна при тлумаченні. Використовувати середовище об'єктно-орієнтованого програмування MS Visual Studio та мову C#. ОС Windows, СУБД SQL CE, технологію WCF.*
4. Зміст пояснювальної записки *(перелік питань, що їх належить розробити): мета роботи, аналіз стану досліджень обробки природної мови, дослідження та моделювання елементів лексикографічних систем, архітектура програмної системи та структури даних, опис розробленої програмної системи,висновки.*
5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень): *Мета завдання, актуальність роботи, постановка задачі, приклади побудови гіперланцюгів лексичних одиниць, метод побудови гіперланцюгів лексичних одиниць, варіант побудови гіперланцюгів для слова «Орел», лексикографічна система тлумачного словника української мови, архітектура програмної системи, структура бази даних, алгоритм автоматизації пошуку зв’язків, інтерфейс програмної системи, результати виводу гіперланцюгів, подальший розвиток, висновки.*

**КАЛЕНДАРНИЙ ПЛАН**

|  |  |  |  |
| --- | --- | --- | --- |
| № | Назва етапів дипломного проекту (роботи) | Термін виконання етапів проекту (роботи) | Примітка |
| 1 | Об'єктний аналіз поставленої задачі | 25-02-2018 | виконано |
| 2 | Розробка моделі взаємодії даних | 10-03-2018 | виконано |
| 3 | Розробка структури зберігання даних | 15-04-2018 | виконано |
| 4 | Створення коду програми | 08-05-2018 | виконано |
| 5 | Тестування і налагодження програми | 20-05-2018 | виконано |
| 6 | Підготовка пояснювальної записки. | 01-06-2018 | виконано |
| 7 | Підготовка презентації та доповіді | 04-06-2018 | виконано |
| 8 | Попередній захист |  |  |
| 9 | Нормоконтроль, рецензування |  |  |
| 10 | Занесення диплома в електронний архів |  |  |
| 11 | Допуск до захисту у зав. кафедри |  |  |

Дата видачі завдання «\_\_\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2018 р.

Студент Тищенко О.О.

(підпис)

Kepiвник роботи \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ проф. Четвериков Г.Г.

(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Пояснювальна записка до атестаційної роботи магістра **містить**: 62 с., 20 рис., 23 джер., 2 додатки.

ГІПЕРЛАНЦЮГ, ОМОНІМІЯ, ЛЕКСИКОГРАФІЯ, ТЛУМАЧНИЙ СЛОВНИК, УМІФ, ВЕБ-СЕРВІС.

Об’єктом дослідження є процес удосконалення програмної системи для побудови гіперланцюгів лексичних одиниць електронного тлумачного словника з метою запобігання виникненню неоднозначностей у тлумаченні через виникнення омонімів.

Метою роботи вдосконалення програмної системи, яке забезпечить мінімізацію випадків неправильного тлумачення слів через проблему омонімії, що допоможе подальшій обробці інформації лексикографами.

Методи розробки базуються на використанні теорії класифікації мовної інформації, системно-лінгвістичному аналізі, комп’ютерному моделюванні мовних процесів, базуючись на математичному апараті теорії інтелекту.

У результаті роботи здійснена програмна реалізація системи побудови гіперланцюгів для виявлення семантичних зв’язків між словами електронного тлумачного словника української мови.

HYPERCHAIN, HOMONYMY, LEXICOGRAPHY, EXPLANATORY DICTIONARY, ULIF, WEB-SERVICE.

The object of research is the process of improving the software system for building hyperlinks of lexical units of the electronic explanatory dictionary in order to prevent the emergence of ambiguities in interpretation due to the emergence of homonyms.

The purpose of the work is to improve the software system, which will ensure minimization of cases of misinterpretation of words due to the problem of homonymy, which will help the further processing of information by lexicographers.

Methods of development are based on the use of the theory of classification of language information, system-linguistic analysis, computer modeling of speech processes, based on the mathematical apparatus of the theory of intelligence.

As a result of the work, a program implementation of the hyperlink construction system was implemented to identify the semantic links between the words of the electronic explanatory dictionary of the Ukrainian language.

ЗМІСТ

Вступ 5

1 Аналіз стану досліджень обробки природної мови 7

1.1 Сучасний стан розвитку систем обробки природної мови 8

1.2 Морфологічний аналіз в системах обробки природної мови 14

1.3 Засоби моделювання фрагментів природної мови 17

1.4 Постановка задач дослідження 19

2 Дослідження та моделювання елементів лексикографічних систем 20

2.1 Дослідження лексикографічних середовищ 20

2.2 Аналіз загальних положень теорії лексикографічних систем 21

2.3 Аналіз лексикографічної системи тлумачного словника 23

2.4 Аналіз ланцюгів лексичних одиниць 33

2.5 Метод побудови ланцюгів лексичних одиниць 36

3 Архітектура програмної системи та структури даних 40

3.1 Загальна архітектура проекту 40

3.2 Опис модулів програмної системи 41

3.3 Структура бази даних 42

3.4 Взаємодія компонентів програмної системи 44

3.5 Алгоритми та структури даних 45

3.5.1 Алгоритм вибору омонімічних термінів 49

3.5.2 Приклад роботи алгоритма вибору омонімічних термінів 51

4 Опис програмної системи 53

4.1 Загальні відомості 53

4.2 Установка, виклик та видалення програмної системи 53

4.3 Інтерфейс користувача 54

4.3.1 Вікно авторизації 54

4.3.2 Головне вікно 55

4.3.3 Вікно побудови гіперланцюгів 58

Висновки 60

Перелік посилань 61

Додаток А 64

Додаток Б 71