

Министерство образования Республики Беларусь

Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.891.2, 004.42

Филиппова
Татьяна Леонидовна

**МОДЕЛЬ И АЛГОРИТМЫ СЕМАНТИЧЕСКОЙ ОБРАБОТКИ
ТЕКСТОВ**

АВТОРЕФЕРАТ

на соискание степени магистра инженерных наук
по специальности 1-40 80 02 Системный анализ, управление и обработка
информации

Научный руководитель

Герман Олег Витольдович

кандидат технических

наук, доцент

Минск 2022 г.

КРАТКОЕ ВВЕДЕНИЕ

Технологии обработки текстов являются одними из наиболее распространенных технологий обработки информации (TextMining). Текст – любая последовательность символов, к которым относятся буквы, пробел, знаки препинания, цифры, знаки арифметических операций и т.п. Текст можно создать карандашом, пером, авторучкой, на пишущей машинке, наконец, на компьютере.

Появление сети Интернет и бурный рост доступной текстовой информации значительно ускорило развитие TextMining–научной области, существующей уже много десятков лет и известной как обработка естественного языка (NaturalLanguageProcessing) и компьютерная лингвистика (ComputationalLinguistics).

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования.Целью данной магистерской диссертации является разработка анализатора запросов на естественном языке и их перевод в SQL запрос.

Актуальность разработки обусловлена большим спросом на такие системы, т.к. в ИТ-индустрии сформировался класс программных систем, в которых человеку проще сформулировать свою цель на естественном человеческом языке (ЕЯ), а не в терминах традиционных формальных компьютерных интерфейсов. Примерами таких областей служат СУБД, электронные секретари, справочные системы, системы управления умными домами и проч.

Новизна полученных результатов. Состоит в обосновании концепции использования семантической сети для работы с текстом на естественном языке и реализация соответственной семантической сети.

Положения, выносимые на защиту.В рамках данной работы были получены следующие практические и научные результаты:

Представлена и реализована концепция семантической обработки текста по схеме: семантическая сеть – xml-документ – база данных – язык SQL.

Рассмотрены возможности языков Python и C# для реализации парадигмы семантической обработки текста.

Апробация результатов диссертации. Промежуточные результаты данной магистерской работы были представлены на ССXLIV Международной научно-практической конференции «Молодой исследователь: вызовы и перспективы» в секции «Информационные технологии.».

Опубликованность результатов исследования. По теме диссертации опубликована статья: Филиппова Т.Л. Использование XML-документов как базы знаний при обработке текстов на естественном языке / Т.Л. Филиппова // Молодой исследователь: вызовы и перспективы: сб. ст. по материалам ССXLIV Международной научно-практической конференции «Молодой исследователь: вызовы и перспективы». – № 2(244). – М., Изд. «Интернаука», 2022.

Структура и объём диссертации. Данная магистерская работа обладает следующей структурой:

1. Исследование программных средств для обработки естественного языка;
2. Использование семантической Сети для обработки текстов на естественном языке;
3. Интеллектуальная обработка текста на python и C#;

Полный объём диссертации: 51 страница.

Количество изображений: 7

Количество таблиц: 0

Количество использованных библиографических источников: 17

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Глава 1. Исследование программных средств для обработки естественного языка. В данной главе производится анализ систем программных средств для обработки естественного языка. Проведен сравнительный анализ ЕЯ-интерфейсов и традиционных интерфейсов к сид. Сделан обзор структуры ЕЯ-интерфейсов и изучены подходы к анализу предложения на естественном языке.

Глава 2. Использование семантической Сети для обработки текстов на естественном языке. В этой главе представлен краткий разбор понятия семантическая сеть и история его возникновения. Описано использование цепочки семантическая сеть – xml – база данных – sql, а также показан один из вариантов реализации такого подхода.

Глава 3. Интеллектуальная обработка текста на python и C#. В главе описана программная реализация интеллектуальной обработки текстов на таких языках программирования как python и C#, а также даны общие сведения об языке python и его использовании в разных областях.

ЗАКЛЮЧЕНИЕ

В рамках данной работы были получены следующие практические и научные результаты:

1. Проведён обзор существующих программных средств для обработки естественного языка.
2. Проведён разбор понятия семантическая сеть и изучена его история.
3. Рассмотрены возможности языков Python и C# для реализации парадигмы семантической обработки текста
4. Проведен сравнительный анализ интерфейсов, использующих естественные языки.

В процессе работы был получен комплекс программного обеспечения семантической работы с текстом на естественном языке, обоснованно его использование.

Научная новизна диссертационной работы:

- 1 Представлена концепция семантической обработки текста по схеме: семантическая сеть – xml-документ – база данных – язык SQL.
- 2 Показан возможность практической реализации концепции

Результатом данной диссертации является программно-алгоритмическое обеспечение для семантической работы с текстом, основанное на цепочке семантическая сеть – xml – база данных – sql.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

Филиппова Т.Л. Использование XML-документов как базы знаний при обработке текстов на естественном языке / Т.Л. Филиппова // Молодой исследователь: вызовы и перспективы: сб. ст. по материалам ССXLIV Международной научно-практической конференции «Молодой исследователь: вызовы и перспективы». – № 2(244). – М., Изд. «Интернаука», 2022.

Библиотека БГУИР