

Министерство образования Республики Беларусь

Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет Информационных технологий и управления
Кафедра Интеллектуальных информационных технологий

ОТЧЁТ
по ознакомительной практике

Выполнил:

Н. В. Клишевич

Студент группы
321703

Проверил:

В. Н. Тищенко

Минск 2024

СОДЕРЖАНИЕ

Введение	3
1 Постановка задачи	4
2 Формализованные фрагменты теории интеллектуальных компьютер- ных систем и технологий их разработки	5
3 Формальная семантическая спецификация библиографических ис- точников	15
Заключение	18
Список использованных источников	19

ВВЕДЕНИЕ

Цель:

Закрепить практические навыки формализации информации в интеллектуальных системах с использованием семантических сетей.

Задачи:

- Построение формализованных фрагментов теории интеллектуальных компьютерных систем и технологий их разработки.
- Построение формальной семантической спецификации библиографических источников, соответствующих указанным выше фрагментам.
- Оформление конкретных предложений по развитию текущей версии Стандарта интеллектуальных компьютерных систем и технологий их разработки.

1 ПОСТАНОВКА ЗАДАЧИ

Часть 2 Учебной дисциплины "Представление и обработка информации в интеллектуальных системах"

⇒ библиографическая ссылка*:

- *Обработка естественных языков*
⇒ URL*:
[<http://repo.ssau.ru/handle/Perspektivnye-informacionnye-tehnologii/OBRABOTKA-ESTESTVENNYH-YaZYKOV-72702>]
- *Автоматическая обработка текстов на естественном языке и компьютерная лингвистика*
⇒ URL*:
[<http://dwl.kiev.ua/art/comp-ling/comp-ling.pdf>]
- *Машинный перевод*
⇒ URL*:
[<https://elib.bsu.by/bitstream/123456789/181916/4/>]
- *Нейросетевые методы в обработке естественного языка*
⇒ URL*:
[<https://medialex.brsu.by/NLP-BOOK/633658.pdf>]
- *Классификация задач компьютерной лингвистики для создания инструмента по управлению потоками задач*
⇒ URL*:
[<https://www.cs.vsu.ru/ipmt-conf/conf/2023/works/2248.dokl.pdf>]
- *Википедия*
⇒ URL*:
[<https://ru.wikipedia.org>]
- ...

2 ФОРМАЛИЗОВАННЫЕ ФРАГМЕНТЫ ТЕОРИИ ИНТЕЛЛЕКТУАЛЬНЫХ КОМПЬЮТЕРНЫХ СИСТЕМ И ТЕХНОЛОГИЙ ИХ РАЗРАБОТКИ

Обработка естественных языков

:= [наука о проектировании методов и алгоритмов, которые принимают или порождают неструктурированные данные естественного языка]

⇒ *прикладные задачи обработки естественных языков**:

{ • *машинный перевод*

⇒ *пояснение**:

[В наши дни существует большое количество систем автоматического перевода с разным качеством, применяющих сложные технологии.]

• *информационный поиск*

⇒ *пояснение**:

[Наиболее часто используемый функционал поисковых систем.]

• *реферирование*

⇒ *пояснение**:

[Для автоматического реферирования в настоящее время применяется отбор наиболее важных предложений текста на основе лингвистических и структурных особенностей текста и используется статистика слов и словосочетаний.]

• *аннотирование*

⇒ *пояснение**:

[Для составления аннотации используют перечень ключевых тем с применением лингвистических и статистических критериев.]

• *классификация*

⇒ *пояснение**:

[При классификации каждый документ относят к заранее определенному классу с известными параметрами.]

• *кластеризация*

⇒ *пояснение**:

[При кластеризации документы разбиваются на кластеры(близкие по тематике документы).]

• *рубрицирование*

⇒ *пояснение**:

[Отнесение документа к определенной тематической рубрике.]

• *формирование ответов на вопросы*

⇒ *пояснение**:

[Решается путем поиска текстов, потенциально содержащих ответ на задаваемый вопрос.]

• *анализ тотальности текстов и выделения мнений*

⇒ *пояснение**:

[Широкое применение получил в коммерческих целях и в вопросах анализа общественного мнения.]

• *поддержка диалога на естественном языке*

⇒ *пояснение**:

- *редактирование текстов*
⇒ *пояснение**:
[Как правило, применяется в специализированных базах знаний.]
 - *обучение естественному языку*
⇒ *пояснение**:
[Разработаны программы обучения морфологии, лексики, словари и т.д.]
 - *автоматическая генерация текстов*
⇒ *пояснение**:
[Спецификой этого направления является автоматический перевод на несколько языков исходя из специфики документа.]
 - *распознавание и синтез речи*
⇒ *пояснение**:
[Возникающие неизбежно при этом ошибки автоматически исправляются на основе морфологических моделей и словарей.]
- ⇒ *автор**:
Йоав Гольдберг
- ⇒ *автор**:
А.И.Соловьев
- ⇒ *библиографическая ссылка**:
- *Нейросетевые методы в обработке естественного языка*
 - *Обработка естественных языков*

Машинный перевод

- := [в узком смысле - процесс перевода некоторого текста с одного естественного языка на другой, реализуемый компьютером полностью или почти полностью. В широком смысле - область научных исследований, находящаяся на стыке лингвистики, математики, кибернетики и имеющая целью построение систем, реализующих машинный перевод в узком смысле.]
- ⊂ *прикладные задачи обработки естественных языков*
- ⇒ *преимущества**:
- {• *Высокая скорость перевода*
⇒ *пояснение**:
[Использование системы машинного перевода позволяет значительно сократить время, требуемое для перевода текстов.]
 - *Низкая стоимость перевода*
 - *Конфиденциальность*
⇒ *пояснение**:
[Многие пользователи регулярно используют системы МП для перевода личных писем, ведь далеко не каждый человек готов отдать постороннему переводчику личную переписку или доверить перевод финансовых документов.]
 - *Универсальность*
⇒ *пояснение**:
[Программа-переводчик справится с переводом текстов из самых разных областей: для правильного перевода терминов достаточно подключить необходимые настройки.]

- *Перевод в режиме онлайн и перевод содержания Интернет-страниц*
- }
 - ⇒ *автор*:*
В.В.Воронович
 - ⇒ *библиографический источник*:*
Машинный перевод

Функциональные стили речи

- := [исторически сложившаяся система речевых средств, используемых в той или иной сфере человеческого общения; разновидность литературного языка, выполняющая определённую функцию в общении]
- ⇒ *обобщенная декомпозиция*:*
основные составляющие
 - = {
 - *литературно-художественный*
 - ⇒ *пояснение*:*
[Неоднороден с точки зрения своей функциональности, в нем реализуется практически вся палитра функций языка.]
 - *новостной*
 - *научный*
 - ⇒ *пояснение*:*
[Множество научных текстов неоднородно. С одной стороны – эта неоднородность определяется тем, что при общем доминировании информативной функции языка в текстах смешанного научного стиля по-разному реализуется взаимодействие информативной функции и функции воздействия на адресата: например, в научной публицистике или учебной литературе. С другой стороны, это связано с неоднородностью самих предметных областей. Во множестве научных языков сосуществует множество языков, различающихся именно в соответствии с предметной областью: языки математики, физики, техники, лингвистики, философии и т.д. .]
 - *официально-деловой*
 - ⇒ *пояснение*:*
[Безусловен приоритет именно информационной составляющей.]
- ⇒ *автор*:*
 - *Е.И. Большакова*
 - *Э.С. Клышинский*
 - *Д.В. Ландэ*
 - *А.А. Носков*
 - *О.В. Пескова*
 - *Е.В. Ягунова*
 - ⇒ *библиографический источник*:*
АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ И КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Информационный поток

:= [множество текстов, выступающих как единый объект: адресатов интересует смысл, заключенный сразу в сотнях и даже тысячах текстов.]

⇒ пример':

коллекция текстов

⇒ автор*:

- Е.И. Большакова
- Э.С. Клышинский
- Д.В. Ландэ
- А.А. Носков
- О.В. Пескова
- Е.В. Ягунова

⇒ библиографический источник*:

АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТОВ НА ЕСТЕСТВЕННОМ
ЯЗЫКЕ И КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Лингвистические информационные ресурсы

:= [терминология, переводные словари, корпус текстов (устн, письм) и т.д.]

:= [это множество определенным образом оформленных и организованных речевых и языковых данных, находящихся на машинных носителях и используемых в различных сферах деятельности человека.]

⇒ обобщенная декомпозиция*:

основные составляющие

= { • Корпус текстов
⇒ пояснение*:

[Совокупность текстов, считающаяся достаточной и представительной для данного языка, диалекта или любого иного подмножества языка, предназначенная для различных типов анализа.]

- Одноязычные и многоязычные словари

⇒ пояснение*:

[Справочные книги, содержащие морфемы, слова, словосочетания, расположенные в определенном порядке с приписанной информацией о них. Используются в информационно-поисковых системах, системах машинного перевода, проверки орфографии, индексирования.]

- Терминологические словари и базы данных
- Массивы устной речи

}

⇒ автор*:

- Е.И. Большакова
- Э.С. Клышинский
- Д.В. Ландэ
- А.А. Носков
- О.В. Пескова
- Е.В. Ягунова

⇒ библиографический источник*:

АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТОВ НА ЕСТЕСТВЕННОМ
ЯЗЫКЕ И КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Система анализа текста

:= [процесс извлечения высококачественной информации из текста.]

⇒ разбиение*:

{• Графематический анализ

⇒ пояснение*:

[Обеспечивает выделение синтаксических или структурных единиц из входного текста, который может представлять собой линейную структуру, содержащую единый фрагмент текста. Однако в более общем случае текст может состоять из многих структурных единиц: основного текста, заголовков, вставок, врезок, комментариев и т.д. При машинном переводе ставится задача сохранить подобную структуру текста. Однако в случае диалоговых систем обычно используется первый вариант (без вставок). Но и в этом случае графематический анализ должен выделять синтаксические единицы: абзацы, предложения, отдельные слова и знаки препинания. В ряде случаев здесь же проводится предморфологический анализ – объединение неразрывных неизменяемых словосочетаний в одну единицу: «_ , , _»]

• **Морфологический анализ**

⇒ пояснение*:

[Обеспечивает определение нормальной формы, от которой была образована данная словоформа, и набора параметров, приписанных данной словоформе. Это делается для того, чтобы ориентироваться в дальнейшем только на нормальную форму, а не на все словоформы, использовать параметры, например, для проверки согласования слов.]

• **Предсинтаксический анализ**

⇒ пояснение*:

[Отвечает за две противоположные задачи: объединение отдельных лексических единиц в одну синтаксическую или, наоборот, ее разделение на несколько. В одну синтаксическую единицу объединяются изменяемые неразрывные словосочетания (например, «бить баклуши»). Делением слов особенно необходимо заниматься, например, в немецком языке, где несколько произвольных связанных между собой слов могут объединяться в одно сложное «на лету», а помещать в морфологический анализ все подобные сочетания не представляется возможным. Еще одной задачей предсинтаксического анализа является проведение синтаксической сегментации. Её задачей является разметка линейного текста на фрагменты, привязанные правилам следующего этапа – синтаксического анализа, который является задачей с экспоненциальным ростом сложности. В связи с этим любая помощь при его проведении может привести к существенному ускорению его работы.]

• **Синтаксический анализ**

⇒ пояснение*:

[Самая сложная часть анализа текста. Здесь необходимо определить роли слов и их связи между собой. Результатом этого этапа является набор деревьев, показывающих такие связи. Выполнение задачи осложняется огромным количеством альтернативных вариантов,

возникающих в ходе разбора, связанных как с многозначностью входных данных (одна и та же словоформа может быть получена от различных нормальных форм), так и неоднозначностью самих правил разбора.]

- *Постсинтаксический анализ*

⇒ *пояснение**:

[Служит двум целям. С одной стороны нам необходимо уточнить смысл, заложенный в слова и выраженный при помощи различных средств языка: предлогов, префиксов или аффиксов, создающих ту или иную словоформу. С другой стороны, одна и та же мысль может быть выражена различными конструкциями языка. В случае с многоязыковой диалоговой системой, одну и ту же мысль можно выразить различными синтаксическими конструкциями. В связи с этим дерево необходимо нормализовать, т.е. конструкция, выражающая некоторое действие различным образом для различных языков или ситуаций, должна быть сведена к одному и тому же нормализованному дереву. Кроме того, на этом же этапе может проводиться обработка разрывных изменяемых словосочетаний, в которых слова словосочетания могут изменяться и могут быть разделены другими словами («белый офицер» vs «белый корниловский офицер»).]

- *Семантический анализ*

⇒ *пояснение**:

[Проводит анализ текста «по смыслу». С одной стороны, семантический анализ уточняет связи, которые не смог уточнить постсинтаксический анализ, так как многие роли выражаются не только при помощи средств языка, но и с учетом значения слова. С другой стороны, семантический анализ позволяет отфильтровать некоторые значения слов или даже целые варианты разбора как «семантически несвязные».]

⇒ }
*автор**:

- *Е.И. Большакова*
- *Э.С. Клышинский*
- *Д.В. Ландэ*
- *А.А. Носков*
- *О.В. Пескова*
- *Е.В. Ягунова*

⇒ *библиографический источник**:

*АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТОВ НА ЕСТЕСТВЕННОМ
ЯЗЫКЕ И КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА*

Морфологический параметр

:= [это пара <имя параметра, значение параметра>.]

⇒ *пояснение**:

[Параметры равны между собой, если равны их имена и значения. Параметры равны по имени, если совпадают их имена. В ряде случаев значение параметра определить невозможно или в этом нет необходимости.]

⇒ *разбиение**:

- { • *Имя параметра*

- ⇒ *пояснение**:
[Именем параметра может служить род, число, время, склонение, краткость формы прилагательного и другие признаки слов, принятые в данном языке]
- *Значение параметра*
⇒ *пояснение**:
[Конкретное значение, которое может принимать данный признак.]
- }
- ⇒ *автор**:
 - *Е.И. Большакова*
 - *Э.С. Клышинский*
 - *Д.В. Ландэ*
 - *А.А. Носков*
 - *О.В. Пескова*
 - *Е.В. Ягунова*
- ⇒ *библиографический источник**:
АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ И КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Расстояние Левенштейна

- := [минимальное количество ошибок, исправление которых приводит одно слово к другому]
- := [метрика, измеряющая по модулю разность между двумя последовательностями символов. Она определяется как минимальное количество односимвольных операций (а именно вставки, удаления, замены), необходимых для превращения одной последовательности символов в другую.]
- ⇒ *применение**:
 - {• для исправления ошибок в слове (в поисковых системах, базах данных, при вводе текста, при автоматическом распознавании отсканированного текста или речи).
 - для сравнения текстовых файлов утилитой diff и ей подобными. Здесь роль «символов» играют строки, а роль «строк» — файлы.
 - в биоинформатике для сравнения генов, хромосом и белков.
- }
- ⇒ *недостатки**:
 - {• При перестановке местами слов или частей слов получаются сравнительно большие расстояния.
 - Расстояния между совершенно разными короткими словами оказываются небольшими, в то время как расстояния между очень похожими длинными словами оказываются значительными.
- }
- ⇒ *автор**:
 - *Е.И. Большакова*
 - *Э.С. Клышинский*
 - *Д.В. Ландэ*
 - *А.А. Носков*
 - *О.В. Пескова*
 - *Е.В. Ягунова*
- ⇒ *библиографическая ссылка**:

- АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ И КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА
- <https://ru.wikipedia.org>

Синтаксическая сегментация

- ⊂ сегментация
- := [процесс разделения письменного текста на значимые единицы, такие как слова, предложения или темы]
- ⇒ *пояснение**:
- [Этот термин применим как к мыслительным процессам, используемым людьми при чтении текста, так и к искусственным процессам, реализованным в компьютерах, которые являются предметом обработки естественного языка. Проблема нетривиальна, потому что, хотя в некоторых письменных языках есть явные маркеры границ слов, такие как пробелы между словами в письменном английском и отличительные начальные, средние и конечные формы букв в арабском, такие сигналы иногда неоднозначны и присутствуют не во всех письменных языках.]
- ⇒ *задачи**:
- уменьшение количества омонимов, соответствующих каждой словоформе.
 - выделение синтаксических конструкций
- }
- ⇒ *проблемы**:
- Сегментация слов

⇒ *пояснение**:

[проблема разделения строки письменного языка на составляющие ее слова]
 - Намеренная сегментация

⇒ *пояснение**:

[проблема разделения написанных слов на ключевые фразы (2 или более групп слов).]
 - Сегментация предложений

⇒ *пояснение**:

[проблема разделения строки письменного языка на составляющие ее предложения.]
 - Сегментация по темам

⇒ *пояснение**:

[Тематический анализ состоит из двух основных задач: идентификации темы и сегментации текста. В то время как первый представляет собой простую классификацию конкретного текста, последний случай подразумевает, что документ может содержать несколько тем, и задача компьютерной сегментации текста может заключаться в автоматическом обнаружении этих тем и соответствующей сегментации текста. Границы тем могут быть очевидны из названий разделов и абзацев. В других случаях необходимо использовать методы, аналогичные тем, которые используются в классификации документов.]
- }
- ⇒ *автор**:

П. А. Чересов

⇒ библиографический источник*:

Классификация задач компьютерной лингвистики для создания инструмента по управлению потоками задач

Формальная семантика

:= [дисциплина, изучающая семантику (интерпретации) формальных и естественных языков путём их формального описания в математических терминах.]

⊂ семантика

⇒ основные современные подходы*:

= { • Теоретико-модельная семантика
⇒ пояснение*:

[Архетип семантики теории истинности Альфреда Тарского, основанной на его Т-схеме, является одной из ключевых концепций модельной теории. Это один из наиболее распространённых подходов. Основная его идея в том, что смысл различных частей утверждения задаются всевозможными способами рекурсивного задания группы функций интерпретации, отображающих предложения на некоторые заранее заданные математические множества. Так, интерпретация логики предикатов первого порядка задаётся отображением термов в универсум, и отображение предикатов в значения истинности «истина» и «ложь». На модельно-теоретической семантике основан подход в теории смысла под названием семантика условной Истины, который впервые был предложен Дональдом Девидсоном. Семантика Крипке по сути вносит некоторые дополнения к семантике Тарского.]

- Теоретико-доказательная семантика
⇒ пояснение*:

[Связывает смысл утверждений с ролями, которые они играют в рассуждении. Герхард Генцен, Даг Prawitz (швед. Dag Prawitz) и Майкл Даммет считаются основателями этого подхода. На него сильно повлияла поздняя философия Людвига Витгенштейна, особенно его афоризм «смысл — это применение».]

- Семантика значений истинности
⇒ пояснение*:

[(также известная как подстановочная квантификация) была предложена Рут Маркус для модальных логик в начале 1960-х и затем развита в трудах Дана (Michael Dunn), Белнапа (англ. Nuel Belnap) и Леблана (Hugues Leblanc) в качестве стандартной логики первого порядка. Джеймс Гарсон (англ. James Garson) получил некоторые результаты в областях адекватности интенциональных логик, снабжённых такой семантикой. Условия истинности квантифицированных формул задаются исключительно в терминах истинности, без использования множеств (отсюда и название).]

- Теоретико-игровая семантика

⇒ *пояснение**:

[Недавно была возрождена Яакко Хинтиккой для логик (конечной) частично покрытой квантификации, которые изначально исследовались Леоном Хенкиным.]

- *Вероятностная семантика*

⇒ *пояснение**:

[Обобщение семантики значений истинности, созданное Филдом (Hartry Field).]

}

⇒ *библиографический источник**:
<https://ru.wikipedia.org>

3 ФОРМАЛЬНАЯ СЕМАНТИЧЕСКАЯ СПЕЦИФИКАЦИЯ БИБЛИОГРАФИЧЕСКИХ ИСТОЧНИКОВ

Нейросетевые методы в обработке естественного языка

⇒ *тип источника**:

[книга]

⇒ *автор**:

Йоав Гольдберг

⇒ *ключевой знак**:

- *естественный язык*
- *нейронные сети*
- *обработка естественного языка*

⇒ *аннотация**:

[Это классическое руководство посвящено применению нейросетевых моделей к обработке данных естественного языка (Natural Language Processing – NLP). Рассматриваются основы машинного обучения с учителем на лингвистических данных и применение векторных, а не символических представлений слов. Обсуждается абстракция графа вычислений, которая позволяет легко определять и обучать произвольные нейронные сети и лежит в основе современных программных нейросетевых библиотек. Также даются обзорные сведения специализированных нейросетевых архитектур, включая одномерные сверточные сети, рекуррентные нейронные сети, модели условной генерации и модели с механизмом внимания. Издание предназначено студентам вузов, а также специалистам в области машинного перевода и нейронных сетей. Предполагается знание теории вероятностей, алгебры и математического анализа, а также базовое владение алгоритмами и структурами данных]

⇒ *цитата**:

[Помимо проблем, связанных с обработкой неоднозначных и вариативных входных данных в системе с плохо определенными и отсутствующими наборами правил, у естественного языка есть и дополнительные свойства, которые еще больше затрудняют разработку вычислительных подходов на основе машинного обучения: дискретность, композиционность и разреженность.]

Обработка естественные языков

⇒ *тип источника**:

[статья]

⇒ *автор**:

А.И.Соловьев

⇒ *цитата**:

[В настоящее время с бурным ростом информации и интернета значительное развитие получила автоматическая обработка текстов.]

Машинный перевод

⇒ *тип источника**:

[лекция]

⇒ *автор**:

В.В.Воронович

⇒ *цитата**:

[Термин машинный перевод (МП) понимается по крайней мере в двух смыслах. Машинный перевод в узком смысле – это процесс перевода некоторого текста с одного естественного языка на другой, реализуемый компьютером полностью или почти полностью. В ходе данного процесса на вход машины подается текст, словесная часть которого не сопровождается никакими дополнительными указаниями, а на выходе получается текст на другом языке, являющийся переводом входного, причем преобразование входного текста в выходной происходит без вмешательства человека (иногда допускается постредактирование). Машинный перевод в широком смысле – это область научных исследований, находящаяся на стыке лингвистики, математики, кибернетики, и имеющая целью построение систем, реализующих машинный перевод в узком смысле.]

Автоматическая обработка текстов на естественном языке и компьютерная лингвистика

⇒ *тип источника**:

[учебное пособие]

⇒ *автор**:

- *Е.И. Большакова*
- *Э.С. Клышинский*
- *Д.В. Ландэ*
- *А.А. Носков*
- *О.В. Пескова*
- *Е.В. Ягунова*

⇒ *аннотация**:

[В учебном пособии рассматриваются базовые вопросы компьютерной лингвистики: от теории лингвистического и математического моделирования до вариантов технологических решений. Дается лингвистическая интерпретация основных лингвистических объектов и единиц анализа. Приведены сведения, необходимые для создания отдельных подсистем, отвечающих за анализ текстов на естественном языке. Рассматриваются вопросы построения систем классификации и кластеризации текстовых данных, основы фрактальной теории текстовой информации. Предназначено для студентов и аспирантов высших учебных заведений, работающих в области обработки текстов на естественном языке.]

⇒ *цитата**:

[Говоря о том, что язык — знаковая система, имеют в виду, что основной элемент такой системы — знак. Знак служит средством отражения того или иного элемента действительности. Благодаря наличию в языке данного знака этот элемент не только получает представительство в системе знаний о мире, присущей носителю языка, — возникает возможность передать эти знания другому.]

Классификация задач компьютерной лингвистики для создания инструмента по управлению потоками задач

⇒ *тип источника**:

[статья]

⇒ *автор**:

П.А. Чересов

⇒ *ключевой знак**:

- *компьютерная лингвистика*
- *анализ текста*
- *прикладные задачи компьютерной лингвистики*

- *задачи этапов анализа текста*

⇒ *аннотация**:

[В статье рассматриваются задачи компьютерной лингвистики и их разделение на базовые и прикладные. Распределение проводится с целью создания инструмента по управлению потоками задач, связанных с компьютерной лингвистикой. Он позволит снизить нагрузку на оборудование при наличии большого количества задач различной сложности.]

⇒ *цитата**:

[Каждый из этапов анализа текста включает в себя набор специфических задач, например, выделение отдельных слов для графематического анализа, или разрешение морфологической омонимии для морфологического анализа. В рамках работы такие задачи называются базовыми. Практически любая прикладная задача КЛ реализуется с их применением.]

ЗАКЛЮЧЕНИЕ

В рамках учебной практики были повышены навыки формализации научных статей и текстов. Были изучены и формализованы понятия связанные с изучаемой дисциплиной и темой «Языковые средства формального описания синтаксиса и денотационной семантики естественных языков в ostis-системах», при помощи пакета макросов «scn-latex».

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] А.И.Соловьев,. Обработка естественных языков / А.И.Соловьев. — РИТ, 2018. — 551 с.
- [2] В.В.Воронович,. Машинный перевод / В.В.Воронович. — Белорусский государственный университет, 2017. — С. 57.
- [3] Гольдберг, Йоав. Нейросетевые методы в обработке естественного языка / Йоав Гольдберг. — ДМК Пресс, 2019. — С. 281.
- [4] Е.И. Большакова Э.С. Клышинский, Д.В. Ландэ А.А. Носков О.В. Пескова Е.В. Ягунова. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / Д.В. Ландэ А.А. Носков О.В. Пескова Е.В. Ягунова Е.И. Большакова, Э.С. Клышинский. — МИЭМ, 2011. — С. 271.
- [5] Чересов, П. А. Классификация задач компьютерной лингвистики для создания инструмента по управлению потоками задач / П. А. Чересов. — Московский Авиационный Институт. — С. 8.