

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:

М. А. Пушко

Студент группы  
321701

Проверил:

В. Н. Тищенко

Минск 2024

## СОДЕРЖАНИЕ

Введение . . . . .	3
1 Постановка задачи . . . . .	4
2 Формализация денотационной семантики естественных языков . .	5
3 Формальная семантическая спецификация библиографических ис- точников . . . . .	9
Заключение . . . . .	10
Список использованных источников . . . . .	11

## **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\*:

- Смирнов И.В..СеманАЕЯ-2013ст  
:= [Семантико-синтаксический анализ естественных языков]
- Джумабаева М.Ш.ИнфорТвОЛИ-2023ст  
:= [Информационные технологии в обработке лингвистической информации]

## 2 ФОРМАЛИЗАЦИЯ ДЕНОТАЦИОННОЙ СЕМАНТИКИ ЕСТЕСТВЕННЫХ ЯЗЫКОВ

### *синтаксический анализ*

**:=** [parsing]

**⇒** *определение\**:

[Синтаксический анализ - это одна из задач обработки естественного языка, которая заключается в анализе грамматической структуры предложения и определении взаимосвязей между его частями.]

**⇒** *разбиение\**:

**=** { • *глубокий синтаксический анализ*

**⇒** *разбиение\**:

*задачи*

**=** { • *построение полного синтаксического дерева предложения с максимальной связанностью с учетом дальних связей*  
• *определение грамматических функций слов предложения*  
}

• *поверхностный синтаксический анализ*

**⇒** *разбиение\**:

*задачи*

**=** { • *разделение предложения на рекурсивно невложенные синтаксические группы*  
**:=** [chunking]  
• *сегментацию*  
**:=** [выделение в предложении различных оборотов и простых предложений в составе сложного]  
• *построение поверхностного синтаксического дерева*  
}

}

**⇒** *библиографическая ссылка\**:

• *Смирнов И.В..СеманАЕЯ-2013ст*

### *обработка естественного языка*

**:=** [Natural language processing]

**:=** [NLP]

**⇒** *определение\**:

[Обработка естественного языка - это область, посвященная разработке алгоритмов и компьютерных программ, которые позволяют компьютерам «понимать» естественные языки, такие как английский, немецкий, французский и т.д.]

**⇒** *включение\**:

*методы*

**=** { • *статистический анализ*  
• *машинное обучение*  
• *глубокое обучение*  
}

**⇒** *разбиение\**:

*задачи для решения*

**=** { • *часторечной анализ тональности*  
• *часторечной анализ эмоций*

- *разбор сложных синтаксических конструкций*
- }
- ⇒ *разбиение\*:*  
*задачи*
- = { • *морфологический анализ*  
⇒ *пояснение\*:*  
[Разбор текста на отдельные слова и выделение их формы и значения, данная технология используется для анализа структуры слов в тексте.]
- *синтаксический анализ*
- *Анализ смысла слов и выражений в контексте текста*  
:= [семантический анализ]  
⇒ *определение\*:*  
[Семантический анализ - это анализ значения слова и выражения, то есть изучение того, какое значение они несут в конкретном контексте]
- *Извлечение именованных сущностей, таких как имена людей, организаций и географических мест*  
:= [Named Entity Recognition]  
:= [NER]  
⇒ *определение\*:*  
[NER - это задача обработки естественного языка, которая заключается в определении именованных сущностей в тексте, таких как имена личностей, компаний, организаций, географических мест и т.д.]
- *Определение тональности и эмоциональной окраски текста*  
:= [анализ тональности]  
⇒ *определение\*:*  
[Анализ тональности - это процесс определения эмоциональной окраски текста, который может быть положительным, отрицательным или нейтральным. ]  
⇒ *разбиение\*:*  
*широко используется*  
= { • *маркетинг*  
• *социальное исследование*  
• *область для изучения мнений людей*  
• *область для изучения настроений людей*  
}
- *Машинный перевод с одного языка на другой*  
⇒ *определение\*:*  
[Машинный перевод - это задача обработки естественного языка, которая заключается в автоматическом переводе текста с одного языка на другой язык.]  
⇒ *включение\*:*  
*основные подходы*  
= { • *статистический машинный перевод*  
⇒ *пояснение\*:*  
[Статистический машинный перевод - основан на идеи, что перевод одного предложения с одного

языка на другой язык должен быть похож на перевод других предложений с аналогичной грамматической структурой. Он использует большие объёмы параллельных корпусов текста, чтобы вычислить вероятности перевода различных слов и фраз.]

- *машинное обучение*  
⇒ *пояснение\**:

[Машинное обучение - использует нейронные сети для создания моделей машинного перевода на основе больших объёмов параллельных текстов;]

- *глубокий машинный перевод*  
⇒ *пояснение\**:

[Глубокий машинный перевод - это подход к машинному обучению, который использует глубокие нейронные сети для перевода текстов]

}

- *Генерация текста на основе заданных параметров и условий*  
:= [текстогенерация)]  
⇒ *определение\**:

[Текстогенерация - это процесс создания текста компьютером или программой.]

- *Классификация текстов по заданному критерию.*  
⇒ *определение\**:

[Классификация текстов - это задача обработки естественного языка, которая заключается в автоматической классификации текстов на основе заданного критерия, например, по тематике, жанру, настроению и т.д.]

- ⇒ *разбиение\**:  
*метод классификации*

- = { • *метод наивного Байеса*  
⇒ *пояснение\**:

[Метод наивного Байеса использует статистические методы на основе частотности слов и фраз в текстах]

- *метод максимальной энтропии*  
⇒ *пояснение\**:

[Метод максимальной энтропии использует теорию информации для поиска наилучшей модели классификации текстов на основе их свойств.]

- *метод опорного вектора*  
⇒ *пояснение\**:

[Метод опорного вектора использует машинное обучение для построения модели, которая может разделять текст на различные классы.]

}

- *Распознавание голоса и естественной речи.*  
⇒ *определение\**:

[Распознавание голоса и естественной речи - это задача обработки естественного языка, которая заключается в переводе устной речи

в текстовую форму и обратно.]

⇒ разбиение\*:

= { • распознавание речи

⇒ определение\*:

[Распознавание речи - процесс, который заключается в преобразовании звуков человеческой речи в признаки, которые могут быть интерпретированы компьютерной программой.]

• обработка естественного языка

⇒ определение\*:

[Обработка естественного языка - процесс, который заключается в интерпретации текста, полученного от распознавания речи]

}

}

⇒ библиографическая ссылка\*:

• Джумабаева М.Ш.ИнфорТвОЛИ-2023ст



### 3 ФОРМАЛЬНАЯ СЕМАНТИЧЕСКАЯ СПЕЦИФИКАЦИЯ БИБЛИОГРАФИЧЕСКИХ ИСТОЧНИКОВ

*Смирнов И.В..СеманАЕЯ-2013ст*

⇒ аннотация\*:

[Рассмотрены задачи семантико-синтаксического анализа текстов на естественных языках. Приведен обзор подходов и методов синтаксического и семантического анализа текстов. Сделаны выводы о применимости существующих подходов к разработке методов семантико-синтаксического анализа текстов.]

⇒ ключевые термины\*:

- синтаксический анализ
- семантический анализ
- формальная грамматика
- машинное обучение

*Джумабаева М.Ш.ИнфорТвОЛИ-2023ст*

⇒ аннотация\*:

[В статье рассмотрены информационные технологии в обработке лингвистической информации. Более того показан процесс о принципах компьютерной лингвистики, методах извлечения и обработки информации, а также об основных алгоритмах и методах анализа языка.]

⇒ ключевые термины\*:

- обработка естественного языка
- машинное обучение
- анализ текста
- машинный перевод
- распознавание речи
- генерация текста
- компьютерная лингвистика
- искусственный интеллект
- многопоточность
- большие данные
- тотальность текста

## **ЗАКЛЮЧЕНИЕ**

В рамках учебной практики были изучены навыки, которые необходимы для формализации научных текстов. Был изучен стандарт OSTIS для соблюдения синтаксических правил оформления формализованной теории и проведена работа по поиску необходимой литературы для выбранной темы. В ходе практической работы были дополнены уже формализованные понятия в монографии примечаниями и пояснениями. Были формализованы понятия, связанные с изучаемой дисциплиной и темой «Формализация денотационной семантики естественных языков». Все поставленные цели и задачи были выполнены.

## **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

[1] Джумабаева, Мардона Шакировна. Информационные технологии в обработке лингвистической информации / Мардона Шакировна Джумабаева, Ринат Фаритович Бурнашев. — ООО «Open science», 2023. — Vol. 4. — P. 643–653.

[2] Смирнов, Иван Валентинович. Семантико-синтаксический анализ естественных языков Часть I. Обзор методов синтаксического и семантического анализа текстов / Иван Валентинович Смирнов, Артем Олегович Шелманов. № 1. — Федеральное государственное учреждение "Федеральный исследовательский центр . . . , 2013. — P. 41–54.