

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:

Д. С. Абрамов

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	8
Заключение	12
Список использованных источников	13

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*:

- *ГраммКИиГвРЯ-2024эл*

⇒ URL*:

[https://spravochnick.ru/yazykoznanie_i_filologiya/grammaticheskie_kategorii_imeni_i_glagoly_v_russkom_yazyke/]

- *ЛексиПСuТиФАОП-2015эл*

⇒ URL*:

[<https://studfile.net/preview/1633396/page:2/>]

- *Катен-2015эл*

⇒ URL*:

[<https://ru.wikipedia.org/wiki/>]

- *ГраммС-2011эл*

⇒ URL*:

[<https://ru.wikipedia.org/wiki/>]

- *КоммаС-2011эл*

⇒ URL*:

[<https://ru.wikipedia.org/wiki/>]

- *АнцедГ-2011эл*

⇒ URL*:

[[https://wikipedia.org/wiki/Antecedent_\(grammar\)](https://wikipedia.org/wiki/Antecedent_(grammar))]

- *ФормаСуАЕЯ-2008см*

⇒ URL*:

[<https://cyberleninka.ru/article/n/formalnaya-semantika-i-analiz-estestvennogo-yazyka/viewer>]

- *ТокенЛА-2019эл*

⇒ URL*:

[<https://ru.wikipedia.org/wiki/>]

- *МетодРС-2006эл*

⇒ URL*:

[<https://ru.wikipedia.org/wiki/>]

- *ГраммМ-2011эл*

⇒ URL*:

[https://wikipedia.org/wiki/Montague_grammar]

2 ФОРМАЛИЗОВАННЫЕ ФРАГМЕНТЫ ТЕОРИИ ИНТЕЛЛЕКТУАЛЬНЫХ КОМПЬЮТЕРНЫХ СИСТЕМ И ТЕХНОЛОГИЙ ИХ РАЗРАБОТКИ

грамматическая категория

- ∈ *одушевлённость*
- := [семантический категориальный признак имени, в основе которого лежит противопоставление «живого» «неживому»]
- ⇐ *семейство подмножеств**:
 - {• *ролевое отношение*
 - }
- ∈ *одушевление*
- ∈ *частичное одушевление*
- ∈ *неодушевление*
 - ⇒ *пояснение**:
[В основе одушевлённости лежит наивное знание о классификации существительных по степени живучести: от "живых"(человек, кошка, муха) через ряд промежуточных случаев (берёза, класс, подосиновик) к полностью "неодушевлённым"(камень, стул, бесконечность)]
- ⇒ *библиографическая ссылка**:
ГраммКИиГвРЯ-2024эл

лексические отношения

- ∈ *синонимы*
 - := [слова одной части речи, различные по звучанию и написанию, но имеющие одинаковое или очень близкое лексическое значение]
- ⇒ *разбиение**:
 - {• *семантические синонимы*
 - ⇒ *пояснение**:
[Синонимы, различающиеся оттенками в значениях, называются семантическими]
 - *семантико-стилистические синонимы*
 - ⇒ *пояснение**:
[Синонимы, которые отличаются и по смыслу, и своей стилистической окраской, называются семантико-стилистическими]
- }
- ∈ *антонимы*
 - := [слова одной части речи, различные по звучанию и написанию, имеющие прямо противоположные лексические значения]
- ∈ *омонимы*
 - := [слова одной части речи, которые совпадают в звучании и написании, но различны по значению]
- ∈ *паронимы*
 - := [Однокоренные слова, близкие по звучанию, но не совпадающие в значениях]
- ⇒ *библиографическая ссылка**:
ЛексиПСиТиФАОП-2015эл

катена

:= [Любой элемент (слово или морф) или любая комбинация элементов, неразрывная в вертикальном измерении (по оси Y)]

⇒ *примечание**:

[В терминах теории графов, любое синтаксическое дерево или подграф дерева является катеной. В этом отношении, каждая составляющая — катена, но существует множество катен, не являющихся составляющими. Таким образом, составляющая — это подвид катены.]

⇒ *библиографическая ссылка**:

Катен-2015эл

рекурсивность

:= [способность включать в себя составляющие той же фразовой категории]

⇒ *примечание**:

[Некоторые фразовые категории, в частности именная группа и предложение, обладают свойством рекурсивности]

⇒ *библиографическая ссылка**:

ГраммС-2011эл

командование составляющих

:= [отношение между отдельными узлами (составляющими) в деревьях синтаксического разбора]

⇒ *формальное определение**:

[Составляющая командует составляющей, если составляющая, которая непосредственно доминирует (включает в себя) над, доминирует также и над (не обязательно непосредственно) и ни, ни не доминируют друг над другом (иначе говоря, всякая составляющая командует своей сестрой и всеми составляющими, которые включает в себя сестра)]

⇒ *библиографическая ссылка**:

КоммаС-2011эл

анцедент

:= [Одно или несколько слов, которые устанавливают значение местоимения или другой проформы]

⇒ *пояснение**:

[Например, в предложении "Вилли сказал, что он любит шоколад" он - проформа, а Вилли - анцедент]

⇒ *библиографическая ссылка**:

АнцедГ-2011эл

пропозиция

:= [семантическая единица, содержащая всю необходимую информацию для формирования формулы исчисления предикатов, отражающей глубинную семантику предложения]

⇒ *разбиение**:

{• *пропозиция действия*

⇒ *пояснение**:

[пропозиция, ядром которой является действие, сообщает об определённом событии и состоит из семантических единиц, относящихся к этому событию]

- *пропозиция состояния*
 ⇒ *пояснение**:
 [пропозиция, ядром которой является отношение, существующее
 между двумя семантическими единицами]
 }
 ⇒ *библиографическая ссылка**:
 ФормаСиАЕЯ-2008ст

токен

- := [объект, создающийся из лексемы в процессе лексического анализа («токенизации»)]
- ⇒ *библиографическая ссылка**:
 ТокенЛА-2019эл

метод рекурсивного спуска

- := [алгоритм нисходящего синтаксического анализа, реализуемый путём взаимного вызова процедур, где каждая процедура соответствует одному из правил контекстно-свободной грамматики. Применения правил последовательно, слева направо поглощают токены, полученные от лексического анализатора. Это один из самых простых алгоритмов синтаксического анализа, подходящий для полностью ручной реализации.]
- ⇒ *задачи**:
 - *предсказывающий парсер*
 ⇒ *пояснение**:
 [Для парсеров этого типа нужна подходящая контекстно-свободная грамматика, конкретно — LL(k) грамматика, позволяющая по очередному токenu или токенам однозначно выбрать (предсказать) один из альтернативных вариантов раскрытия каждого нетерминала. Такой парсер работает за линейное время. Вариантом является LL-парсер — реализация предсказывающего парсера с автоматическим построением «таблицы предсказания», определяющей по заданному нетерминалу и очередному токenu подходящее правило для раскрытия нетерминала.]
 - *парсер с возвратом*
 ⇒ *пояснение**:
 [Вместо предсказания парсер просто пытается применить все альтернативные варианты правил по порядку, пока одна из попыток не увенчается успехом. Такой парсер может потребовать экспоненциального времени работы, и не всегда гарантирует завершение, в зависимости от грамматики. Уязвим для левой рекурсии]
- ⇒ *библиографическая ссылка**:
 МетодРС-2006эл

грамматика Монтегю

- := [подход к формализации семантики естественного языка]
- ⇒ *пояснение**:
 [Грамматика Монтегю основана на математической логике, особенно на логике предикатов высшего порядка и лямбда-исчислении, и использует понятия интенсивной логики через модели Крипке]
- ⇒ *библиографическая ссылка**:
 ГраммМ-2011эл

3 ФОРМАЛЬНАЯ СЕМАНТИЧЕСКАЯ СПЕЦИФИКАЦИЯ БИБЛИОГРАФИЧЕСКИХ ИСТОЧНИКОВ

ГраммКиИГвРЯ-2024эл

- ⇒ *тип источника**:
[электронный ресурс]
- ⇒ *ключевой знак**:
- *грамматическая категория*
 - *надеж*
 - *род*
 - *одушевлённость*
 - *время*
 - *вид*
 - *лицо*
 - *наклонение*
 - *залог*
- ⇒ *цитата**:
[грамматическая категория является важнейшим понятием грамматики]

ЛексиПСуТиФАОП-2015эл

- ⇒ *тип источника**:
[электронный ресурс]
- ⇒ *ключевой знак**:
- *синонимы*
 - *антонимы*
 - *паронимы*
 - *омонимы*
- ⇒ *цитата**:
[Парадигматические отношения – это такие отношения слов, когда они рассматриваются как одновременно сосуществующие в языке, в словаре, в его запасе. В парадигматике слова связаны друг с другом отношениями сходства и различия. Так, различают отношения: а) сходства в плане содержания - синонимы; б) сходства в плане выражения - омонимы; в) неполного (частичного) сходства - паронимы; г) включения – лексико-семантические, или тематические поля; д) противопоставления – антонимы]

Катен-2015эл

- ⇒ *тип источника**:
[электронный ресурс]
- ⇒ *ключевой знак**:
- *грамматика зависимостей*
 - *составляющая*
 - *компонента*
 - *строка*
- ⇒ *цитата**:
[Любой отдельный элемент (слово или морф) или комбинация элементов, связанные в вертикальном измерении — это катена. Структура предложения рассматривается

как существующая в двух измерениях. Комбинации, образованные вдоль горизонтального измерения (по порядку предшествования), называются строками, в то время как комбинации, образованные вдоль вертикального измерения (по порядку доминанции) — катены. В терминах декартовой системы координат, строки существуют вдоль оси X, а катены — вдоль оси Y]

ГраммС-2011эл

⇒ *тип источника**:

[электронный ресурс]

⇒ *ключевой знак**:

- *лингвистическая грамматика*
- *составляющая*
- *генеративная лингвистика Н. Хомского*
- *система составляющих*
- *фразовые категории*

⇒ *цитата**:

[Грамматика составляющих (метод составляющих; англ. constituency grammar, phrase structure grammar) основана на постулате, согласно которому всякая сложная грамматическая единица складывается из двух более простых и не пересекающихся единиц, называемых её непосредственными составляющими (англ. immediate constituent)]

⇒ *цитата**:

[Представление синтаксической структуры предложения в виде иерархии непосредственных составляющих используется в различных вариантах в формальных моделях языка, в частности в генеративной лингвистике Н. Хомского]

КоммаС-2011эл

⇒ *тип источника**:

[электронный ресурс]

⇒ *ключевой знак**:

- *Асимметрическое командование*
- *составляющая*
- *генеративная лингвистика Н. Хомского*

⇒ *цитата**:

[Отношение командования тесно связано с грамматиками составляющих в традиции Хомского (теория связывания, минималистская программа). Оно не подходит или не эффективно в деревьях разбора других синтаксических теорий. Например, командование в его обычном понимании едва ли может быть применено к деревьям грамматик зависимостей]

⇒ *цитата**:

[Понятие командования составляющих используется в теории связывания для формулировки ограничения на употребление анафорических местоимений: анафоров (возвратных и взаимных местоимений) и прономиналов (прочих анафорических местоимений). Так, условие, состоящее в том, чтобы анафор был связан в пределах некоторой составляющей, означает, что в пределах данной составляющей должно присутствовать выражение-антецедент, командующее анафором и кореферентное ему]

АнцедГ-2011эл

- ⇒ *тип источника**:
[электронный ресурс]
- ⇒ *ключевой знак**:
 - *проформа*
 - *неопределённый анцедент*
 - *постанцедент*
 - *подразумеваемый анцедент*
- ⇒ *цитата**:
[Почти любая синтаксическая категория может служить антецедентом для проформы]

ФормаСиАЕЯ-2008ст

- ⇒ *тип источника**:
[статья]
- ⇒ *автор**:
 - *Г.Ф. Дюбко*
 - *Д.В. Преснякова*
- ⇒ *ключевой знак**:
 - *естественный язык*
 - *семантическая структура*
 - *семантика*
- ⇒ *аннотация**:
[В предлагаемой вниманию читателя работе решается задача преобразования естественно-языковых текстов в формулы исчисления предикатов. При этом лексическая семантика интегрируется с формальной семантикой. преобразования осуществляются на основе грамматического разбора и семантической структуры текста, которые позволяют классифицировать отношения между составляющими текста]
- ⇒ *цитата**:
[Одной из задач обработки естественных языков средствами информатики является переход от суждений, выраженных ЕЯ-формами, к суждениям в определённой символической системе]
- ⇒ *цитата**:
[Представление смысла ЕЯ-форм сводится к представлению их формулами логики предикатов, независимо от того, какой способ получения логических следствий выбран в конкретной реализации. Поэтому необходимо сформулировать принципы и методы перевода ЕЯ-форм и формулы логики предикатов]

ТокенЛА-2019эл

- ⇒ *тип источника**:
[электронный ресурс]
- ⇒ *ключевой знак**:
 - *прикладное программирование*
 - *лексема*
 - *шаблон токена*
- ⇒ *цитата**:
[В прикладном программировании понятие токена и его лексема могут не различаться]

МетодРС-2006эл

⇒ *тип источника**:

[электронный ресурс]

⇒ *ключевой знак**:

- *нисходящий семантический анализ*
- *контекстно-свободная грамматика*
- *токен*

⇒ *цитата**:

[метод рекурсивного спуска - это один из самых простых алгоритмов синтаксического анализа, подходящий для полностью ручной реализации]

ГраммМ-2011эл

⇒ *тип источника**:

[электронный ресурс]

⇒ *ключевой знак**:

- *естественный язык*
- *логика предикатов высшего порядка*
- *лямбда-исчисление*

⇒ *цитата**:

[На мой взгляд, нет важной теоретической разницы между естественными языками и искусственными языками логиков; более того, я считаю возможным понять синтаксис и семантику обоих видов языков в рамках единой естественной и математически точной теории. По этому вопросу я отличаюсь от ряда философов, но, полагаю, согласен с Хомским и его единомышленниками]

⇒ *автор цитаты**:

[Ричард Монтегю]

ЗАКЛЮЧЕНИЕ

В результате работы было формализовано 10 понятий, связанных с темой "Формализация синтаксиса естественных языков". Также была сделана формальная семантическая спецификация источников информации, использовавшихся при выполнении работы. В результате были повышены навыки работы с информацией, формализации понятий, работы с LaTeX, в частности с макросами "scn-latex" были приобретены новые знания в рассматриваемой области.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Анцедент(грамматика). — 2011. <https://ru.wikipedia.org/wiki/>.
- [2] Г.Ф.Дюбко, Д.В.Преснякова. Формальная семантика и анализ естественного языка / Д.В.Преснякова Г.Ф.Дюбко. — 2008. — С. 48–53.
- [3] Грамматика Монтегю. — 2011. <https://ru.wikipedia.org/wiki/>.
- [4] Грамматика составляющих. — 2011. <https://ru.wikipedia.org/wiki/>.
- [5] Грамматические категории имени и глаголы в русском языке. — 2024. https://spravochnik.ru/yazykoznanie_i_filologiya/grammaticheskie_kategorii_imeni_i_glagoly_v_russkom_yazyke/.
- [6] Катена. — 2015. <https://ru.wikipedia.org/wiki/>.
- [7] Командование составляющих. — 2011. <https://ru.wikipedia.org/wiki/>.
- [8] Лексическая парадигматика. Синонимы, их типы и функции. Антонимы. Омонимы. Паронимы. — 2015. <https://studfile.net/preview/1633396/page:2/>.
- [9] Метод рекурсивного спуска. — 2006. <https://ru.wikipedia.org/wiki/>.
- [10] Токен (лексический анализ). — 2019. <https://ru.wikipedia.org/wiki/>.