

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. Р. Гетман

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 | 11 |
| Заключение | 15 |
| Список использованных источников | 16 |

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*:

- *Бородин Д.С. Системы ИИЗ ОЕЯ-2018*
⇒ URL*:
[<https://www.elibrary.ru/item.asp?id=34887564>]
- *Полякова М.Ю.. Разработка ПСАЭТИС-2011*
⇒ URL*:
[<https://cyberleninka.ru/article/n/razrabotka-podhoda-k-sozdaniyu-algoritma-sintaksicheskogo-analiza-estestvenno-yazykovogo-teksta-informatsionno-poiskovyh-sistem>]
- *Посевкин Р.В.. Естественный ПИ-2016*
⇒ URL*:
[<https://cyberleninka.ru/article/n/estestvenno-yazykovoy-polzovatelskiy-interfeys-dialogovoy-sistemy>]
- *Медведева И.В. ЛингвОДС-2010*
⇒ URL*:
[<https://elibrary.ru/item.asp?id=16287941>]
- *Луканин А.В. АвтомОЕЯ-2011*
⇒ URL*:
[<http://artyom.ice-lc.com/science/alukaninnlproc.pdf>]
- *Хлопенкова А.Ю.. Метод ОЕЯВГП-2019*
⇒ URL*:
[<https://cyberleninka.ru/article/n/metody-obrabotki-estestvennogo-yazyka-v-virtualnyh-golosovyh-pomoschnikah>]
- *Усталов Д.А. СеманСОЕС-2017*
⇒ URL*:
[<https://elibrary.ru/item.asp?id=29716048>]
- *Судаков Б.Н.. Синтаксический синтез естественного языка в экспертных системах*
⇒ URL*:
[<https://cyberleninka.ru/article/n/sintaksicheskii-sintez-estestvenno-yazykovykh-tekstov-v-ekspertnykh-sistemah>]

2 ФОРМАЛИЗОВАННЫЕ ФРАГМЕНТЫ ТЕОРИИ ИНТЕЛЛЕКТУАЛЬНЫХ КОМПЬЮТЕРНЫХ СИСТЕМ И ТЕХНОЛОГИЙ ИХ РАЗРАБОТКИ

вопросно-ответная система

:= [question-answering system]

:= [система, предназначенная для автоматического ответа на вопросы, заданные на естественном языке]

⇒ *примечание**:

[В вопросно-ответных системах ответ генерируется на языке запроса, в то время как в системах общения с базами данных ответ может формироваться из обычного набора данных, хранящихся в базе.]

Э Система *START*

Э *Женя Густман*

Э *Простая вопросно-ответная система на основе семантического анализатора русского языка*

Э *RAZOOM*

Э *ITFRU*

⇒ *библиографическая ссылка**:

- *Бородин Д.С. Системы ИИЗОЕЯ-2018 ст*

информационно-поисковая система

:= [information search engine]

:= [ИПС]

:= [программа для поиска, хранения и предоставления пользователю информации из различных источников на основе его запросов.]

⇒ *примечание**:

[Одним из основных элементов ИПС является лингвистический процессор (ЛП), выполняющий роль посредника между пользователем и базой данных, в которой хранится интересующая его информация.]

Э *Google*

Э *Yandex*

Э *Bing*

Э *Yahoo*

Э *Baidu*

⇒ *библиографическая ссылка**:

- *Полякова М.Ю.. Разраб. ПСА С А Е Т И С - 2011 ст*

лингвистический процессор

:= [linguistic processor]

:= [ЛП]

:= [программа, выполняющая роль посредника между пользователем и базой данных, в которой хранится интересующая его информация]

⇒ *разбиение**:

- *блок морфологического анализа*
 - *блок синтаксического анализа*
 - *блок семантического анализа*
- }

⇒ *примечание**:

- [Синтаксический анализ, является главным блоком, определяющим качество работы ЛП в целом.]
- ⇒ *библиографическая ссылка**:
- *Полякова М.Ю..РазраПСАЕТИС-2011см*

синтаксический анализ

- := [syntactic analysis]
- := [процесс определения грамматической структуры предложения путем выявления синтаксических отношений между его элементами.]
- ⇒ *задача**:
- [Используя морфологическую информацию о словоформах, построить синтаксическую структуру входного предложения (осуществить разбор предложения).]
- ⇒ *примечание**:
- [Синтаксический анализ может рассматриваться как процесс поиска дерева синтаксического анализа.]
- ⇒ *примечание**:
- [К началу синтаксического анализа весь текст представляется в виде последовательности характеристик к словоформам, так что алгоритм синтаксического анализа имеет дело не со словоформами, а лишь с соответствующими характеристиками.]
- ⇒ *программа синтаксического анализа**:
- {
 - *сегментация предложения*
 - *установление связей между словами*}
- ⇒ *разбиение**:
- {
 - *нисходящий синтаксический анализ*
⇒ *примечание**:
[Начать с вершины и искать дерево, листьями которого являются соответствующие слова.]
 - *восходящий синтаксический анализ*
⇒ *примечание**:
[Начать со слов и выполнять поиск дерева начиная с корня.]}
- ⇒ *библиографическая ссылка**:
- *Полякова М.Ю..РазраПСАЕТИС-2011см*

ограниченный естественный язык

- := [limited natural language]
- := [подмножество естественного языка, на котором текст воспринимается носителем естественного языка без дополнительных усилий.]
- ⇒ *примечание**:
- [При этом не требуется изучение ограниченной версии языка для составления текстов. Данный язык обладает сокращенным набором лексики и грамматики, что позволяет сократить время анализа естественно-языковых элементов в диалоговой системе, а также избежать неоднозначностей на лингвистическом уровне.]
- ⇒ *библиографическая ссылка**:
- *Посевкин Р.В..ЕстесПИ-2016см*

диалоговая система

- := [dialog system]
- := [человеко-машинная система “пользователь – ЭВМ”, обеспечивающая эффективное решение некоторого класса задач в режиме оперативного обмена информацией между партнерами]
- := [совокупность технических, программных и информационных средств, обеспечивающих заданному кругу пользователей возможность решения некоторого множества задач в интерактивном режиме взаимодействия с ЭВМ в соответствии со сценарием диалога]
- ⇒ *библиографическая ссылка**:
 - *Медведева И.В. ЛингвОДС-2010ст*

лингвистический автомат

- := [linguistic automaton]
- := [ЛА]
- := [сбалансированный комплекс аппаратных, программных, лингвистических, а иногда и лингводидактических средств, взаимодействующих с мощной базой лингвистических данных и знаний]
- ⇒ *примечание**:

[Принятый модулем ввода и коррекции текст поступает на вход лексикоморфологического анализатора, который с помощью соответствующего подблока БДЗ создает подстрочник. Подстрочник поступает в подблоки синтаксического и семантического анализа. Там он разбивается на именные и глагольные группы, каждая из которых анализируется и обрабатывается с использованием графов переходов. Полученный результат передаётся в подблок порождения выходного текста, а затем выдаётся потребителю.]
- ⊕ { свойства
 - *полифункциональность*
 - ⇒ *примечание**:

[способность осуществлять разные виды обработки больших потоков текстовой информации]
 - *минимизация информационных потерь*
 - *живучесть*
 - ⇒ *примечание**:

[способность ЛА сохранять свои наиболее существенные свойства в результате воздействия на ЛА таких катастрофических факторов, как выход из строя внешних устройств или участков оперативной памяти, искажение отдельных фрагментов текста и т.п.]
 - *способность к дальнейшему развитию и усовершенствованию*
 - *возможность подключения к различным каналам связи*
- ⇒ *разбиение**:
 - { • *база данных и знаний*
 - *блок ввода, распознавания и первичной обработки печатного текста*
 - *блок распознавания устной речи, преобразующий аудио-сигнал в текст*
 - *блок индексирования и аннотирования текста*
 - *блок машинного перевода*
 - *блок вывода печатного текста*
 - *блок, озвучивающий печатный текст*
 - *средства человеко-машинной синергетики и блок управления*

- ⇒ }
библиографическая ссылка*:
• Луканин А.В. АвтомОЕЯ-2011 *уп*

система автоматического синтеза речи

- := [automatic speech synthesis system]
- := [АСР]
- := [Text-to-Speech]
- := [TTS]
- := [компьютеризированная система, которая может произносить любой текст вне зависимости от того, введён ли он в компьютер оператором или отсканирован и распознан системой оптического распознавания символов]

⇒ сферы применения*:

- { • средства телекоммуникаций
⇒ примечание*:

[Системы АСР можно использовать для получения текстовой информации по телефону. В качестве текстов могут быть как простые сообщения, такие как важные культурные события (в кинотеатрах, театрах и т.д.), так и огромные базы данных, в которых трудно хранить записанную речь.]

- обучение языку
⇒ примечание*:

[На сегодняшний день системы АСР практически не применяются в этой области, так как их низкое качество не удовлетворяет требованиям таких задач.]

- помощь инвалидам
⇒ примечание*:

[С помощью специально спроектированных клавиатур и программ быстрого составления предложений речь может быть синтезирована за несколько секунд, позволив немым общаться с другими людьми.]

- говорящие книги и игрушки
- голосовой контроль
⇒ примечание*:

[В некоторых случаях произнесённая информация более эффективна, чем текстовые сообщения. Она обращает на себя больше внимания, в то время как можно получать другую информацию из визуальных источников. Данная идея привела к встраиванию речевых синтезаторов в системы контроля и измерения.]

- мультимедиа
- человеко-машинное общение

- ⇒ }
библиографическая ссылка*:
• Луканин А.В. АвтомОЕЯ-2011 *уп*

голосовой ассистент

- := [voice assistant]
- := [программа, способная распознавать голосовые команды пользователя и выполнять соответствующие задачи или предоставлять информацию]

⇒ примечание*:

[Голосовые помощники построены на искусственном интеллекте (ИИ), технологиях машинного обучения и распознавания голоса. Поскольку конечный пользователь взаимодействует с цифровым помощником, в программировании ИИ используются сложные алгоритмы, позволяющие учиться на вводе данных и лучше предсказывать потребности пользователя.]

⇒ *Alexa*

⇒ *разработчик**:

[Amazon]

⇒ *примечание**:

[Alexa построена на основе обработки естественного языка (NLP –Natural Language Processing) - процедур преобразования речи в отдельные слова, звуки и идеи. Интерпретация звуков требует больших вычислительных ресурсов поэтому записанная речь отправляется на серверы Amazon для более эффективного анализа]

⇒ *Cortana*

⇒ *разработчик**:

[Microsoft]

⇒ *примечание**:

[Возможности обработки естественного языка в Cortana получены от Tellme Networks и связаны с базой данных семантического поиска под названием Satori.]

⇒ *Алиса*

⇒ *разработчик**:

[Яндекс]

⇒ *примечание**:

[Алиса базируется на технологии, именуемой SpeechKit. SpeechKit представляет собой клиентскую библиотеку, через которую звук поступает на сервер, обрабатывается и возвращается в интерфейс Алисы в виде текста и звука]

⇒ *библиографическая ссылка**:

- *Хлопенкова А.Ю..МетодОЕЯВГП-2019ст*

семантическая сеть

:= [semantic network]

:= [размеченный ориентированный граф, вершины которого соответствуют некоторым сущностям (понятиям, событиям, характеристикам или значениям), а ребра выражают отношения между этими сущностями.]

⇒ *примечание**:

[Семантические сети не накладывают ограничений на структуру знаний или конкретную предметную область до тех пор, пока эти знания можно представить в виде ориентированного графа.]

⇒ *примечание**:

[Основная сложность при построении семантических сетей связана с большим объемом работы, которую приходится выполнять для их создания. Применение автоматизированных подходов сокращает этот процесс, но повышает требования к контролю качества.]

⇒ *WordNet*

⇒ *ConceptNet*

⇒ *Freebase*

⇒ *DBpedia*

- ∃ *OSTIS*
⇒ библиографическая ссылка*:
• Усталов Д.А. СеманСОЕС-2017ст

3 ФОРМАЛЬНАЯ СЕМАНТИЧЕСКАЯ СПЕЦИФИКАЦИЯ БИБЛИОГРАФИЧЕСКИХ ИСТОЧНИКОВ

Луканин А.В. АвтомОЕЯ-2011уп

⇒ *тип источника**:

[учебное пособие]

⇒ *автор**:

- А.В. Луканин

⇒ *аннотация**:

[В учебном пособии рассматриваются вопросы автоматизации анализа и синтеза текста и речи на естественном языке с использованием различных формальных моделей: от задач компьютерной морфологии, дешифрования, аннотирования, до распознавания и синтеза звучащей речи и автоматического перевода.]

⇒ *оглавление**:

[

- Системы автоматической обработки текста и лингвистический автомат
- Фреймы в системах автоматической обработки текстов
- Подъязыки в системах машинного перевода
- Теория конечных автоматов
- Системы аннотирования и реферирования
- Криптография
- Системы автоматического синтеза речи
- Системы автоматического распознавания речи
- Искусственные нейронные сети

]

⇒ *цитата**:

[Новые технологии позволяют использовать в полной мере возможности, предоставляемые такими считывающими устройствами, как карманный и главный сканеры и программы проверки орфографии, а также адаптером для связи с компьютером.]

Полякова М.Ю..РазраПСАСАЕТИС-2011ст

⇒ *тип источника**:

[статья]

⇒ *автор**:

- М.Ю. Полякова
- Б.Н. Судаков

⇒ *ключевой знак**:

- синтаксический анализ
- семантический анализ
- естественнoязыковoй текст

⇒ *аннотация**:

[Рассмотрены существующие методы синтаксического анализа естественнoязыковoго текста и выделены основные преимущества и недостатки. Разработан усовершенствованный алгоритм синтаксического анализа. Показано, что параллельное использование синтаксического и семантического анализа позволяет сократить временные затраты на обработку естественнoязыковoго текста.]

⇒ *цитата**:

[Программа синтаксического анализа, как правило, состоит из двух компонентов: сегментации предложения и установления связей между словами. Компоненты работают параллельно или последовательно, в зависимости от архитектуры синтаксического модуля.]

Посевкин Р.В..ЕстесПИИ-2016ст

⇒ *тип источника**:

[статья]

⇒ *автор**:

- Р.В. Посевкин
- И.А. Бессмертный

⇒ *ключевой знак**:

- диалоговая система
- обработка естественного языка
- пользовательский интерфейс
- БД

⇒ *аннотация**:

[В статье рассмотрены методы компьютерной лингвистики и обработки естественного языка. Представлено полное описание всех стадий обработки естественного языка, таких как морфологический, синтаксический и семантический анализ.]

⇒ *цитата**:

[Ограниченный естественный язык – это подмножество естественного языка, на котором текст воспринимается носителем естественного языка без дополнительных усилий.]

Хлопенкова А.Ю..МетодОЕЯВГП-2019ст

⇒ *тип источника**:

[статья]

⇒ *автор**:

- А.Ю. Хлопенкова
- Ю.С. Белов

⇒ *ключевой знак**:

- голосовой помощник
- Alexa
- Cortana
- Алиса
- граф знаний
- SpeechKit

⇒ *аннотация**:

[Данная статья посвящена обзору существующих голосовых помощников. Дается общее определение голосовым помощникам. Рассматривается программная реализация каждой из систем.]

⇒ *цитата**:

[Обработка сигналов является одной из наиболее важных задач в области звукозаписи на больших дистанциях. Основная задача - улучшить целевой сигнал, что означает возможность идентифицировать окружающий шум, такой как телевизор, и минимизировать его]

Усталов Д.А.СеманСОЕС-2017ст

⇒ *тип источника**:

[статья]

⇒ *автор**:

- Д.А. Усталов

⇒ *ключевой знак**:

- семантические технологии
- машинное обучение
- распознавание речи
- Академия ОС

⇒ *аннотация**:

[Машинное обучение вторглось в области, где до недавнего времени господствовали лишь семантические сети. Однако в задачах обработки естественного языка, построения инвентаря значений слов и связывания языковых ресурсов семантические сети по-прежнему позволяют достигать высокой точности.]

⇒ *цитата**:

[Достаточно важной проблемой является техническая и лицензионная доступность семантических сетей.]

Медведева И.В.ЛингвОДС-2010ст

⇒ *тип источника**:

[статья]

⇒ *автор**:

- Медведева И.В.

⇒ *ключевой знак**:

- мультимодальные диалоговые системы
- симуляция диалоговых систем
- речевые диалоговые системы
- взаимодействие "человек - машина"

⇒ *аннотация**:

[Современные диалоговые системы представляют собой мультимодальные системы с использованием не только голосовой информации, но и визуальной. Целью данной статьи является описание диалоговых систем, при этом особое внимание уделяется лингвистическому аспекту проблемы.]

⇒ *цитата**:

[Взаимодействие человека с машиной в век развития высоких информационных технологий является популярным и очень актуальным предметом исследований. В XXI в. очень сложно представить себе жизнь без компьютера. Практически каждый день человек взаимодействует с той или иной информационной системой. Внедрение интеллектуальных информационных систем, в частности систем речевого управления, например, диалоговых систем, отвечает тенденции современного общества к автоматизации как можно большего числа процессов.]

Бородин Д.С.СистеИИЗОЕЯ-2018ст

⇒ *тип источника**:

[статья]

⇒ *автор**:

- Бородин Д.С.

⇒ *ключевой знак**:

- *системы искусственного интеллекта*
- *вопросно-ответные системы*
- *обработка естественного языка*
- *транслятор естественно-языковых запросов*

⇒ *аннотация**:

[Данная работа посвящена рассмотрению интеллектуальных систем, позволяющих преобразовывать входную информацию на естественном языке. Приводится анализ рынка систем искусственного интеллекта, обосновывающий актуальность разработки подобных систем.]

⇒ *цитата**:

[Естественно-языковые системы обладают целым рядом преимуществ по сравнению с системами, в которых отсутствует поддержка интерфейсов на языке, понятном человеку – это и минимальная предварительная подготовка пользователя такой системы, и простота задания запросов на выборку данных на естественном языке, и большая скорость создания произвольного запроса за счет отсутствия стадии формального задания запроса.]

Судаков Б.Н..СинтаСЕТЭС-2012ст

⇒ *тип источника**:

[статья]

⇒ *автор**:

- *Судаков Б.Н.*
- *Филипенко А.Н.*

⇒ *ключевой знак**:

- *метод синтеза текстов*
- *ссылка*
- *диалог*

⇒ *аннотация**:

[В статье рассмотрены методы синтеза текстов, способы формирования синтаксической структуры ответов. Рассмотрены причины использования ссылок и деления диалогов на составные части. А также рассмотрены проблемы, которые встречаются при таких методах синтеза, и способы их устранения.]

⇒ *цитата**:

[Общение с помощью отдельных изолированных (несвязанных) предложений даже в ограниченной предметной области весьма неудобно. В то время как общение, допускающее связный текст, не имеет серьезных недостатков.]

ЗАКЛЮЧЕНИЕ

В рамках выполнения практического задания были формализованы понятия из научных материалов на тему естественно-языковых интерфейсов. Отточены навыки формализации понятий, библиографических источников, используя язык разметки LaTeX и ssp-библиотеку для него.

Также была формализована часть 4 "Онтологические модели интерфейсов интеллектуальных компьютерных систем нового поколения" из монографии OSTIS и создан подробное руководство по формализации понятий для будущих поколений.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] А.В., Луканин. Автоматическая обработка естественного языка / Луканин А.В. — Издательский центр ЮУрГУ, 2011. — С. 70.
- [2] А.Ю., Хлопенкова. Методы обработки естественного языка в виртуальных голосовых помощниках / Хлопенкова А.Ю., Белов Ю.С. — E-Scio, 2019.
- [3] Б.Н., Судаков. Синтаксический синтез естественно-языковых текстов в экспертных системах / Судаков Б.Н., Филипенко А.Н. — Вестник Национального технического университета Харьковский политехнический институт. Серия: Информатика и моделирование, 2012.
- [4] Д.С., Бородин. Системы искусственного интеллекта в задачах обработки естественного языка / Бородин Д.С. — Цифровое будущее инновационной экономики России, 2018.
- [5] И.В., Медведева. Лингвистическое обеспечение диалоговых систем / Медведева И.В. — Вестник Московского государственного лингвистического университета, 2010.
- [6] М.Ю., Полякова. Разработка подхода к созданию алгоритма синтаксического анализа естественно-языкового текста информационно-поисковых систем / Полякова М.Ю., Судаков Б.Н. — Вестник Национального технического университета Харьковский политехнический институт. Серия: Информатика и моделирование, 2011.
- [7] Посевкин, Р.В. Естественно-языковой пользовательский интерфейс диалоговой системы / Р.В. Посевкин, И.А. Бессмертный. — 2016.
- [8] Усталов, Д.А. Семантические сети и обработка естественного языка / Д.А. Усталов. — 2017.