

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Астраханский государственный университет»
(Астраханский государственный университет)

кафедра философии

РЕФЕРАТ

**для сдачи кандидатского экзамена
по истории и философии науки**

**на тему: «История научных достижений в области семантического
анализа текстовой информации»**

Выполнил:

Зорин Кирилл Андреевич,
*Аспирант кафедры информационной безопасности
и цифровых технологий*

Астрахань – 2021 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
1. Первые упоминания семантического анализа	6
2. Первые научные работы	7
3. Первые компьютерные системы	9
4. Современные системы системного анализа.....	13
5. Семантический анализ текстовой информации в повседневной жизни	18
ЗАКЛЮЧЕНИЕ	20
Список используемой литературы	21

ВВЕДЕНИЕ

Философия науки оформилась в отдельную научную дисциплину относительно недавно – в XX веке – и сегодня это одна из наиболее динамично развивающихся философских дисциплин, отвечающая на запросы современного наукоёмкого производства и сциентизированного общества. [9]

Помимо знаний о структуре языка, семантика тесно связана с философией, психологией и другими науками, так как неизбежно затрагивает вопросы о происхождении значений слов, их отношении к бытию и мышлению. [6]

Семантический анализ – совокупность методов, приемов и инструментов, направленных на выявление смысловой нагрузки, представленной в виде текстовой информации. Семантический анализ является трудной математической задачей, так как компьютеры не умеют правильно отличать разницу в контексте текста переданную с помощью разных алфавитов. Разрабатываемые методы и алгоритмы должны позволять компьютерам «понимать» и интерпретировать предложения, абзацы, документы, анализировать их грамматическую структуру, и выявлять отношения между предложениями и словами в различных контекстах.

Большой скачок в развитии компьютерных технологий коснулся множества областей, в том числе и в развитии научных достижений, связанных с семантическим анализом текстовой информации.

История семантического анализа берет свое начало в XVII веке, когда философы Готфрид Вильгельм Лейбница и Рене Декарт выдвинули предположения о кодах, которые связывали бы слова между языками. Однако все эти предположения оставались теоретическими, и ни одно не привело к созданию реальной машины.

Начало применения семантического анализа текста в компьютерной науке приходится на конец XX века. С появлением глобальной сети – Интернет задача определения семантической направленности текста становилась все более актуальной, поэтому основной период достижений в этой области приходится на начало XXI века.

Лексическая семантика позволяет понимать компьютеру отношения между лексическими элементами:

Гипонимы – лексические элементы общего лексического элемента – гипероним;

Мерономия – логическое расположение текста и слов, обозначающее расположение текста и слов, обозначающее составную часть чего-либо;

Полисемия – отношение между значениями слов или фраз, которые могут отличаться друг от друга, но иметь общее основное значение;

Синонимы – слова, которые имеют тоже или почти такое же смысловое значение что и другое слово;

Омонимы – слова, которые звучат и пишутся одинаково, однако имеют абсолютно разное значение;

Антонимы – два слова, которые имеют противоположное значение;

Методы семантического анализа должны учитывать все вышеперечисленные определения. Поэтому очень часто в настоящее время применяются алгоритмы машинного обучения. В настоящее время существует много методов выявления смысла из текстовой информации:

И.А. Мельчук [11] ввел понятие лексической функции, разработал понятия синтаксических и семантических валентностей и рассмотрел их в контексте толково-комбинаторного словаря, который представляет собой языковую модель[8].

В.Ш. Рубашкин и Д.Г. Лахути [10] ввели иерархию синтаксических связей для более эффективной работы семантического анализатора. Самыми важными являются обязательные ролевые связи, далее идут связи кореференции, потом факультативные ролевые связи и только потом предметно-ассоциативные. [8]

Известный лингвист Е.В. Падучева [13] предлагает рассматривать тематические классы слов, в частности глаголов, поскольку они несут основную смысловую нагрузку: глаголы восприятия, глаголы знания, глаголы эмоций, глаголы принятия решения, речевых действий, движения, глаголы звука, бытийные глаголы и др.[8]

Статей, учебников по описанию и реализации методов семантического анализа в компьютерной лингвистике огромное количество. Практически каждый журнал, посвященный достижениям информационных технологий содержит статьи о новых методах семантического анализа текста.

Область применения методов семантического анализа в последние годы потерпела резкий сдвиг с точки зрения предлагаемых методов и решений для данной задачи. Вместе с разработанными методами появились новые программные продукты, начиная от разговорных систем и заканчивая новыми решениями, использующими естественный как способ описания моделей в разных областях науки. По мере решения задачи семантического анализа растет и количество областей, в которых можно примерить данные методы.

Одной из самых важных подзадач семантического анализа является задача обработки естественного языка (NLP).

Решение данной задачи помогает использовать инструменты машинного обучения в привычных для нас вещах таких как: чат – боты, боты-техподдержки, поисковые системы. Также инструменты семантического анализа могут помочь компаниям извлекать необходимую информацию из поступающих данных из разных каналов, таких как электронная почта, звонки в техническую поддержку, отзывов клиентов на сайтах.

1. Первые упоминания семантического анализа

История развития семантического анализа текста начинается в XVII веке, когда два философа: Готфрид Вильгельм Лейбниц и Рене Декарт выдвинули предположения о кодах, которые связывали бы слова между языками. Готфрид Вильгельм Лейбниц – немецкий философ, логик, математик, физик. В 1678 году Готфрид Лейбниц представил эссе, в котором он представил созданный им философский язык, которому дал название – «lingua universalis». Данный язык был принят в качестве универсального языка для расчета. В результате данной работы, Лейбниц разработал двоичное исчисление. Одними из его важнейших достижений для компьютерной науки является создание комбинаторики как науки, основы математической логики, а также описание современной двоичной системы используя только цифры 0 и 1. Пример двоичной системы представлен на рисунке 1.

Один	—1
Два	—10 (одна единица второго разряда)
Три	—11
Четыре	—100 (одна единица третьего разряда)
Пять	—101
Шесть	—110
Семь	—111
Восемь	—1000 (одна единица четвертого разряда)
Девять	—1001
10	—1010
11	—1011
12	—1100
13	—1101
14	—1110
15	—1111
16	—10000 и т. д.

Рисунок 1. Двоичная система

В дальнейшем его достижения повлияют на развитие всей информатики, в том числе и семантического анализа в компьютерной науке. Однако на момент XVII века все предположения Лейбница и Декарта так и остались теоретическими, и так и не привели к созданию реальной машины.

2. Первые научные работы

В начале XX века в середине 30-х годов Жорж Арцруни подал заявку на получение первых патентов на автоматические машины для перевода на двуязычный словарь с использованием бумажной ленты (Рисунок 2). Бумажная лента – один из вариантов хранения данных, представляет собой длинную полосу бумаги, в которой пробиты отверстия.

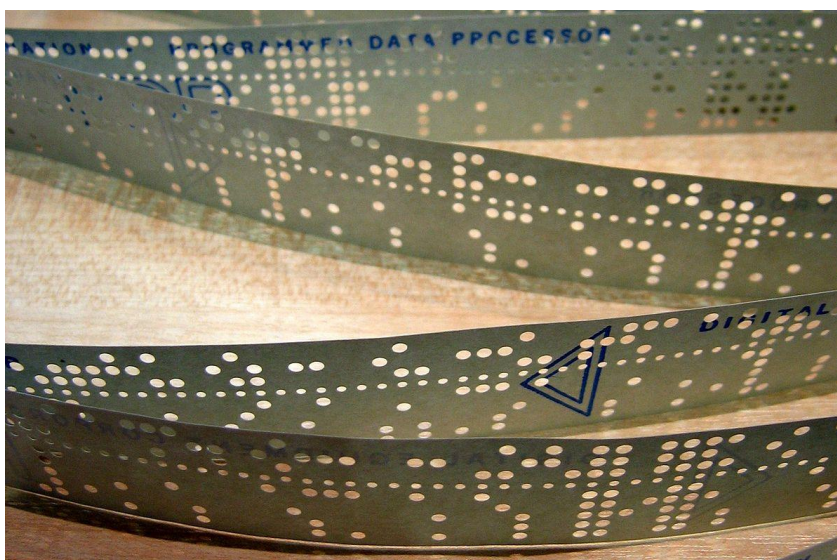


Рисунок 2. Перфолента

Алан Тьюринг – английский математик, логик, человек оказавший огромное влияние на развитие информатики, логики, сформулировавший концепции алгоритмов и вычислений с помощью машины Тьюринга. В 1950 году Алан Тьюринг опубликовал знаменитую статью «Вычислительные машины и интеллект». В качестве «интеллекта» Тьюринг предложил то, что называется тестом Тьюринга. Тестом Тьюринга называется имитация игры, где происходит проверка способности машины к интеллектуальному поведению, эквивалентно или неотличима от человека. Тьюринг решил выяснить может ли машина мыслить. Первоначальная игра в имитацию, которую описал Тьюринг, представляет собой простую групповую игру с участием трех игроков. Игрок «А» - человек, игрок «В» - машина, и игрок «С» (который играет роль экзаменатора) человек. В игре игрок «С» не может видеть игрока – он сидит в отдельной комнате и общается как с игроком «А», так и «В». Он может общаться с ними только посредством письменных заметок или любой другой формы, которая не раскрывает никаких подробностей о них. Задавая вопросы игроку «А» и игроку «В», игрок «С» пытается определить, кто из них машина, а кто человек. Если экзаменатор не понимает кто из них человек, а кто машина – значит выиграла машина.

Переписка должна производиться через контролируемые промежутки времени, чтобы судья не мог делать заключения, исходя из скорости ответов. Во времена Тьюринга компьютеры реагировали медленнее человека. Сейчас это правило тоже необходимо, потому что теперь они реагируют гораздо быстрее, чем человек. В настоящее время обратным тестом Тьюринга считается CAPTCHA. CAPTCHA – как правило используется на сайтах. Прежде чем дать пользователю выполнить какое-либо действие, ему предлагают буквенно-цифровые символы, на искаженном изображении и просят ввести их. Так как сложного и точного приложения для распознавания текста в искаженном изображении не существует, только человек может понять, что изображено, соответственно система делает вывод – что это человек.

В 1957 году была опубликована работа «Синтаксические структуры» Ноама Хомского. Эта краткая монография объемом около 100 страниц была признана одним из самых значительных исследований 20 века. В данной публикации была описана система синтаксических структур, основанная на правилах. Хомский выступал за независимость (изучение структур предложения) синтаксиса от семантики (изучение значения). Хомский утверждает, что язык — это набор предложений, каждое из которых имеет конечную длину. Лингвист должен отделить «грамматические последовательности» или предложения языка от «неграмматических последовательностей». В третьей главе, озаглавленной «Элементарная лингвистическая теория», Хомский пытается определить, какой тип устройства или модели адекватно описывает данный набор «грамматических» предложений. Хомский предполагает, что это устройство должно быть конечным, а не бесконечным. Затем он рассматривает грамматику с конечным числом состояний, теоретическую модель коммуникации, которая рассматривает язык как Цепь Маркова. Цепь Маркова – является вероятностной моделью, описывающей последовательность возможных событий, в которых вероятность каждого события зависит только от состояния, достигнутого в предыдущем случае. Затем в четвертой главе, озаглавленной «Структура фразы», он обсуждает грамматику структуры фраз, модель, основанную на непосредственном анализе составляющих. В пятой главе, озаглавленной «Ограничения описания фразовой структуры», он утверждает, что показывает, что обе эти модели неадекватны для целей лингвистического описания. В качестве решения он вводит трансформацию, «более мощную модель ... которая может исправить эти недостатки».

3. Первые компьютерные системы

В 1954 году состоялся Джорджтаунский эксперимент – демонстрация машинного перевода от компании IBM, который был разработан совместно с Джорджтаунским университетом и IBM. Данный эксперимент включал в себя полностью автоматический перевод более шестидесяти русских предложений на английский язык. Задуманная и реализованная в первую очередь для привлечения государственного и общественного интереса и финансирования путем демонстрации возможностей машинного перевода, это не было полнофункциональной системой: в ее словаре было всего шесть грамматических правил и 250 лексических элементов (основ и окончаний). Пример перевода представлен на рисунке 3.

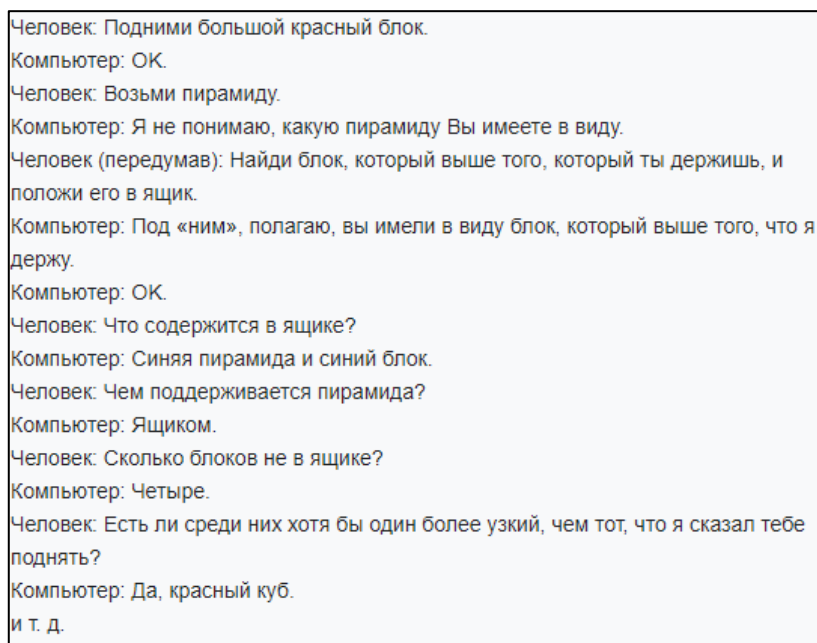
Russian (Romanized)	English translation
Mi pyeryedayem mislyi posryedstvom ryechyi.	We transmit thoughts by means of speech.
Vyelyichyina ugla opryedyelayetsya otnoshyenyiyem dlyini dugi k radiusu.	Magnitude of angle is determined by the relation of length of arc to radius.
Myezhdunarodnoye ponyimaniye yavlyayetsya vazhnim faktorom v ryeshyenyiyi polyityichyeskix voprosov.	International understanding constitutes an important factor in decision of political questions.

Рисунок 3. Пример перевода

Авторы утверждали, что в течение трех-пяти лет машинный перевод станет решенной проблемой. Однако реальный прогресс был намного медленнее, и после отчета ALPAC (Комитет из семи ученых во главе с Джоном Р. Пирсом, созданный в 1964 году правительством США для оценки прогресса компьютерной лингвистики в целом и машинного перевода в 1966 году) в котором было обнаружено, что десятилетние исследования не оправдали ожиданий, финансирование машинного перевода резко сократилось. Небольшие дальнейшие исследования в области машинного перевода проводились до конца 1980-х годов, когда были разработаны первые системы статистического машинного перевода. Статистический машинный перевод – это парадигма машинного перевода, в которой переводы генерируются на основе статистических моделей, параметры которых выводятся из анализа двуязычных текстовых корпусов. Статистический подход контрастирует с основанными на правилах подходами к машинному переводу, а также с машинным переводом на основе примеров. Первые идеи статистического машинного перевода были введены Уорреном Уивером в 1949 году, включая идеи применения Клода Шеннона «теории информации». Статистический машинный перевод был вновь введен в конце 1980 - х и начале 1990 - х годов

исследователями IBM что способствовало значительному всплеску интереса к машинному переводу в последние годы. До появления нейронного машинного перевода это был наиболее широко изученный метод машинного перевода.

В 1960 – х годах была разработана программа SHRDLU. Данная программа считается ранней компьютерной программой для понимания естественного языка, разработанной Терри Виноградом из Массачусетского технологического института. В программе пользователь ведет диалог с компьютером, перемещает объекты, присваивает имена коллекциям и запрашивает состояние. SHRDLU была настолько проста, что весь набор объектов и мест можно было описать, включив всего 50 слов: существительные, такие как «блок» или «конус», глаголы, такие как «место на» или «переехать», и такие прилагательные, как «большой» или «синий». Возможные комбинации этих базовых языковых строительных блоков были довольно просты, и программа довольно хорошо понимала, что имел в виду пользователь. Пример работы программы изображен на рисунке 4.



Человек: Подними большой красный блок.
Компьютер: ОК.
Человек: Возьми пирамиду.
Компьютер: Я не понимаю, какую пирамиду Вы имеете в виду.
Человек (передумав): Найди блок, который выше того, который ты держишь, и положи его в ящик.
Компьютер: Под «ним», полагаю, вы имели в виду блок, который выше того, что я держу.
Компьютер: ОК.
Человек: Что содержится в ящике?
Компьютер: Синяя пирамида и синий блок.
Человек: Чем поддерживается пирамида?
Компьютер: Ящиком.
Человек: Сколько блоков не в ящике?
Компьютер: Четыре.
Человек: Есть ли среди них хотя бы один более узкий, чем тот, что я сказал тебе поднять?
Компьютер: Да, красный куб.
и т. д.

Рисунок 4. Пример работы программы SHRDLU

SHRDLU считался чрезвычайно успешной демонстрацией искусственного интеллекта (ИИ). Это привело других исследователей искусственного интеллекта к чрезмерному оптимизму, который вскоре был утрачен, когда более поздние системы пытались справиться с ситуациями с более реалистичным уровнем сложности.

С 1964 по 1966 год в лаборатории искусственного интеллекта MIT была создана программа ELIZA для демонстрации примеров общения между людьми и машинами.

Данная разработка моделировала разговор с психотерапевтом. Создатель ELIZA, Вайценбаум, рассматривал программу как метод, демонстрирующий связь между человеком и машиной, но был удивлен количеством людей, которые приписывали компьютерной программе человеческие чувства. Многие ученые полагали, что программа сможет положительно повлиять на жизнь многих людей, особенно тех, кто страдает психологическими проблемами, и что она может помочь врачам, работающим над лечением таких пациентов.

Помимо программы ELIZA существовала и другая программа под названием PARRY – чат-бот (виртуальный собеседник). Программа была написана в 1972 году психиатром Кеннетом Колбай. В отличие от ELIZA, PARRY пыталась симулировать человека с параноидальной шизофренией. Программа реализовала грубую модель поведения человека с параноидальной шизофренией, основанную на концепциях, и убеждениях (принять, отвергнуть, нейтрально). Она также воплощала диалоговую стратегию и, как таковая, была гораздо более серьезной и продвинутой программой, чем ELIZA. Она была описана как «ELIZA с отношением». PARRY была протестирована в начале 1970-х годов с использованием разных разновидностей теста Тьюринга. Группа опытных психиатров проанализировала комбинацию реальных пациентов и компьютеров, работающих с PARRY через терминалы. Другой группе из 33 психиатров показали стенограммы разговоров. Затем обеим группам было предложено определить, какие из «пациентов» были людьми, а какие – компьютерными программами. Психиатры смогли сделать правильную идентификацию только в 48 процентах случаев.

В 1978 году была разработана одна из первых систем обработки естественного языка баз данных – LIFER. Она была разработана как интерфейс на естественном языке к базе данных о кораблях ВМС США. Эта система, как описано в статье, использовала семантическую грамматику для анализа вопросов и запросов к распределенной базе данных. Система LIFER / LADDER могла поддерживать только простые запросы к одной таблице или запросы к нескольким таблицам с простыми условиями соединения.

Примеры таких запросов:

1. Какая длина, ширина и осадка у Kitty Hawk?
2. Когда Ривз достигнет степени готовности C2?
3. Какой ближайший корабль к Неаполю с врачом на борту?
4. Какие суда перевозят грузы для Соединенных Штатов? Куда они идут?
5. Распечатать текущие позиции и состояние готовности американских крейсеров?

В 1982 году британский программист Ролло Карпентер создал чат-бота – Jabberwacky. Его заявленная цель - «имитировать естественный человеческий чат в интересной, развлекательной и юмористической манере» [2]. Это ранняя попытка создания искусственного интеллекта посредством человеческого взаимодействия. Заявленная цель проекта - создание искусственного интеллекта, способного пройти тест Тьюринга. Он предназначен для имитации человеческого взаимодействия и общения с пользователями. Он не предназначен для выполнения каких-либо других функций.

В отличие от более традиционных программ ИИ, технология обучения предназначена для развлечения, а не для использования в системах компьютерной поддержки или корпоративном секторе. Конечная цель состоит в том, чтобы программа перешла от текстовой системы к полностью голосовой системе, обучаясь непосредственно по звуку и другим сенсорным данным. Её создатель считает, что систему можно интегрировать в предметы по всему дому, такие как роботы или говорящие домашние животные, чтобы они были полезными и развлекательными, составляя компанию людям.

4. Современные системы системного анализа

Вторым известным продуктом Ролло Карпентера было web-приложение Cleverbot. Он использует алгоритм искусственного интеллекта (ИИ) для общения с людьми. Ему предшествовал Jabberwacky, проект чат-ботов, который начался в 1986 году и вышел в Интернет в 1997 году. В первое десятилетие своего существования Cleverbot провел несколько тысяч разговоров с Карпентером и его сотрудниками. С момента запуска в сети количество разговоров превысило 150 миллионов. Помимо веб-приложения, Cleverbot также доступен для iOS, Android и Windows Phone.

В отличие от другого виртуального собеседника, ответы Cleverbot в предварительно не запрограммированы. Вместо этого он учится на человеческом вводе: люди вводят текст в поле под логотипом Cleverbot, и система находит все ключевые слова или точную фразу, соответствующие введенным данным. После поиска в сохраненных разговорах он реагирует на ввод, обнаруживая, как человек отреагировал на ответ, когда Cleverbot частично или полностью спросили его.

Cleverbot принял участие в формальном тесте Тьюринга на фестивале Techniche 2011 в Индийском технологическом институте Гувахати 3 сентября 2011 года. Из 1334 поданных голосов, Cleverbot был признан человеком на 59,3% по сравнению с оценкой в 63,3%, полученной человеком. Балл 50,05% или выше часто считается проходным. Программное обеспечение, запущенное для мероприятия, должно было обрабатывать всего 1 или 2 одновременно поданных запроса, тогда как онлайн Cleverbot обычно общается с примерно 10 000–50 000 человек одновременно. Cleverbot постоянно увеличивает объем данных со скоростью от 400 до 7 миллионов взаимодействий в секунду. Обновления программного обеспечения в основном происходили негласно. В 2014 году Cleverbot был модернизирован для использования технологий обслуживания GPU. В отличие от ELIZA, программа не отвечает фиксированным образом, вместо этого она выбирает свои ответы эвристически с использованием нечеткой логики, при этом весь разговор сравнивается с миллионами, ранее проводившимися разговорами. Сейчас Cleverbot использует более 279 миллионов взаимодействий, что составляет около 3-4% данных, которые он уже накопил. Разработчики Cleverbot пытаются создать новую версию, используя методы машинного обучения.

Значительная часть движка Cleverbot и API для доступа к нему была предоставлена разработчикам в форме Cleverscript. Сервис для прямого доступа к Cleverbot стал доступен разработчикам в виде Cleverbot.io.

Приложение, которое использует движок Cleverscript для игры в 20 вопросов, было запущено под названием Clevernator. В отличие от других подобных игр, игрок задает вопросы, и задача ИИ - понять и дать фактический ответ. Было запущено приложение, которое позволяет владельцам создавать и общаться со своим собственным маленьким ИИ, похожим на Cleverbot, под названием Cleverme! для продуктов Apple.

В начале 2017 года на платформе Twitch была запущена трансляция двух устройств Google Home, модифицированных для общения друг с другом с помощью Cleverbot.io, трансляция собрала более 700 000 посетителей и более 30 000 одновременных зрителей.

В 1984 году компанией Mindscape был опубликован проект – Racter [4]. По сути, он является искусственным интеллектом – компьютерной программой, которая генерирует на английском языке предложения в случайном порядке.

Racter было написано Уильямом Чемберленом и Томасом Эттером. Существование программы было раскрыто в 1983 году в книге под названием «The Policeman's Beard Is Half Constructed» (ISBN 0-446-38051-2), которая описывалась как полностью созданная программой. Согласно введению Чемберлена к книге, программа явно работала на микроконтроллере CP/M; он был написан на BASIC на Z80 micro с 64Мб оперативной памяти. Эта версия программы, которая якобы написала книгу, не была выпущена для широкой публики. Достижения, заявленные для программы, были, вероятно, преувеличены, что можно было увидеть при исследовании системы шаблонов генерации текста.

Однако в 1984 году Mindscape выпустила интерактивную версию Racter, разработанную Inrac Corporation, для компьютеров, совместимых с IBM, Amiga и Apple II. Опубликованный Racter был похож на болтуна. Программа BASIC, выпущенная Mindscape, была гораздо менее сложной, чем всё, что могло бы написать такую же довольно изощренную статью «The Policeman's Beard». Коммерческую версию Racter можно сравнить с компьютеризированной версией Mad Libs, игры, в которой вы заранее заполняете пробелы, а затем вставляете их в текстовый шаблон, чтобы создать сюрреалистическую сказку. Коммерческая программа попыталась проанализировать вводимый текст, идентифицировать значимые существительные и глаголы, которые затем использовала, создавая «разговоры», вставляя вводимые пользователем данные в шаблоны фраз, которые затем объединяла вместе с модулями, которые спрягали английские глаголы.

Напротив, текст в «The Policeman's Beard», не считая того, что он был отредактирован из-за большого количества выходных материалов, был бы продуктом собственных специализированных шаблонов и модулей Чемберлена, которые не были включены в коммерческий выпуск программы. Но все же книга «The Policeman's Beard Is Half Constructed» [1] была первой в мире книгой, написанной компьютером. Обложка книги представлена на рисунке 5.

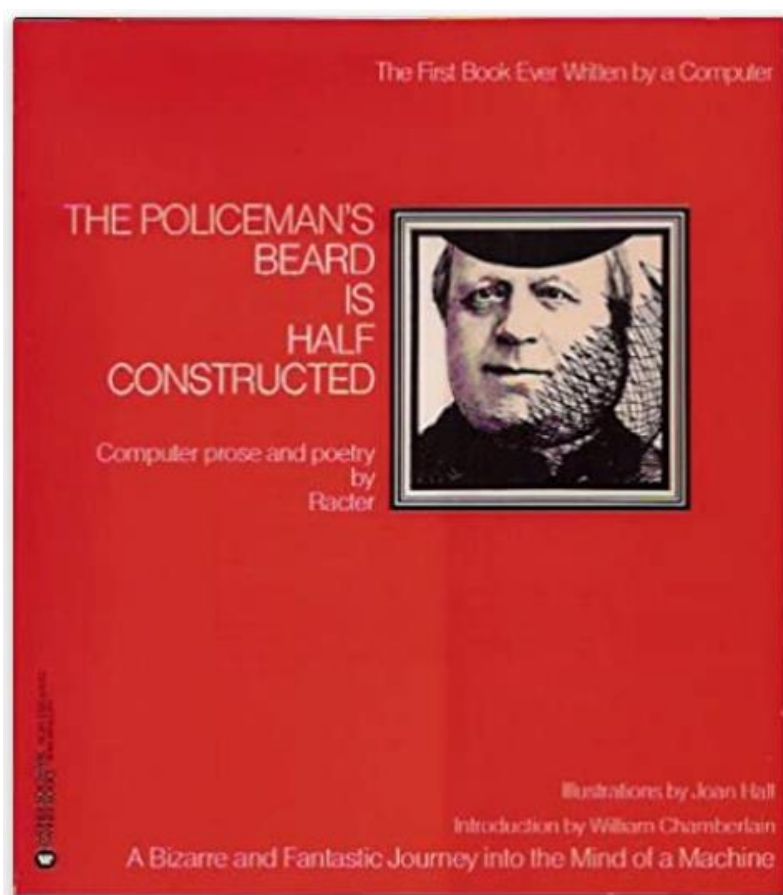


Рисунок 5. Обложка книги «The Policeman's Beard Is Half Constructed»

В 1991 году компания Creative Labs создала программу под названием Dr. Sbaitso. Программа представляет собой искусственный интеллект синтеза речи. Программа «разговаривала» с пользователем, как если бы она была психологом, хотя большинство ее ответов были примерно такими: «ПОЧЕМУ ВЫ ТАК ЧУВСТВУЕТЕ?», и она не представляла какое-либо сложное взаимодействие. Когда ей встречалась фраза, которую она не могла понять, она часто отвечала чем-то вроде «ЭТО НЕ МОЯ ПРОБЛЕМА». Dr. Sbaitso повторяла вслух текст, набранный после слова ключевого слова «СКАЗАТЬ». После запуска программы, она представлялась так: «привет мое имя доктор Dr. Sbaitso. я здесь, чтобы помочь вам. свободно рассказывайте, что в вашем уме, наш разговор будет

сохраняется и будьте уверены, что содержание памяти будет удалено после выхода, расскажите мне о ваших проблемах.»

Программа была разработана для демонстрации оцифрованных голосов хотя качество было далеко от реалистичного. Кроме того, была версия этой программы для Microsoft Windows, в которой использовалась программа под названием Prody Parrot – эта версия программного обеспечения отличалась более подробным графическим пользовательским интерфейсом.

В 2006 году был представлен суперкомпьютер от компании IBM – IBM Watson (Рисунок 6). Он был назван в честь основателя и первого генерального директора IBM, промышленника Томаса Дж. Ватсона [5].



Рисунок 6. Суперкомпьютер от компании IBM

«Jeopardy!» – американское телевизионное игровое шоу, созданное Мервом Гриффином. Шоу включает в себя викторину, в которой участникам предоставляются общие сведения в форме ответов, и они должны сформулировать свои ответы в форме вопросов. IBM Watson изначально была разработана для ответов на вопросы викторины Jeopardy! В 2011 году компьютерная система Watson соревновалась в «Jeopardy!» против чемпионов Брэд Раттера и Кен Дженнинга. В феврале 2013 года IBM объявила, что первое коммерческое приложение системы программного обеспечения Watson будет использоваться для принятия управленческих решений при лечении рака легких в онкологическом центре Memorial Sloan Kettering. Watson [5] был создан как вычислительная система с ответами на

вопросы (QA), которую IBM построила для применения передовых технологий обработки естественного языка, поиска информации, представления знаний, автоматизированного мышления и машинного обучения в области ответов на вопросы в открытой предметной области. При создании IBM заявила, что «более 100 различных методов используются для анализа естественного языка, идентификации источников, поиска и генерации гипотез, поиска и оценки доказательств, а также объединения и ранжирования гипотез» [11].

Философ Джон Сирл утверждает, что Ватсон, несмотря на впечатляющие способности, на самом деле не может думать. Опираясь на мысленный эксперимент в китайской комнате, Сирл утверждает, что Ватсон, как и другие вычислительные машины, способен только манипулировать символами, но не способен понимать значение этих символов, однако у эксперимента Сирла есть противники.

5. Семантический анализ текстовой информации в повседневной жизни

В настоящее время идет плотное развитие информационных технологий, цифровые помощники использующие передовые технологии распознавания текста, добавляются практически в каждое устройство.

Виртуальные помощники работают через:

1. Текст, в том числе: онлайн-чат (особенно в приложении для обмена мгновенными сообщениями или другом приложении), текст SMS, электронная почта или другой канал связи, например интеллектуальные виртуальные помощники Conversica для бизнеса.
2. Голосовая связь, например, с помощью Amazon Alexa на устройстве Amazon Echo, Siri на iPhone или Google Assistant на мобильных устройствах с поддержкой Android.
3. Снимая или загружая изображения, как в случае с Samsung Bixby на Samsung.

Виртуальные помощники используют обработку естественного языка для сопоставления пользовательского текста или голосового ввода с исполняемыми командами. Многие постоянно учатся, используя методы искусственного интеллекта, включая машинное обучение. Некоторые из этих помощников, такие как Google Assistant (который содержит Google Lens) и Samsung Bixby, также имеют дополнительную возможность обрабатывать изображения для распознавания объектов на изображении, чтобы помочь пользователям получить лучшие результаты от нажатых изображений.

Семантический анализ активно используется в различных отраслях. Модели семантической классификации, используемые в настоящее время в голосовых помощниках, а также в системах принятия решений:

1. Классификация тем: сортировка текста по заранее определенным категориям в зависимости от его содержания. Команды обслуживания клиентов могут захотеть классифицировать заявки в службу поддержки по мере их поступления. С помощью семантического анализа инструменты машинного обучения могут определить, следует ли классифицировать билет как «Проблема с оплатой» или «Проблема с доставкой».
2. Анализ настроений: обнаружение в тексте положительных, отрицательных или нейтральных эмоций для обозначения срочности. Например, отметьте упоминания в Twitter по настроениям, чтобы получить представление о том, что клиенты думают о вашем бренде, и иметь возможность идентифицировать недовольных клиентов в режиме реального времени.

3. Классификация намерений: классификация текста на основе того, что клиенты хотят делать дальше. Вы можете использовать это, чтобы пометить электронные письма с продажами как «Заинтересовано» и «Не интересно», чтобы активно обращаться к тем, кто может захотеть попробовать ваш продукт.

ЗАКЛЮЧЕНИЕ

Достижения в науке семантического анализа текстовой информации начались не так давно, всего век назад. Семантический анализ — это процесс извлечения значения из текста. Он позволяет компьютерам понимать и интерпретировать предложения, абзацы или целые документы, анализируя их грамматическую структуру и выявляя отношения между отдельными словами в определенном контексте.

Начиная с 1954 года начали появляться программы, в основе которых заложен семантический анализ текстовой информации. (Таблица 1)

Таблица 1. Программное обеспечение, использующее основы семантического анализа

Название программы	Год	Создатель	Описание
Джорджтаунский эксперимент	1954	Джорджтаунский университет и IBM	Полностью автоматический перевод более шестидесяти русских предложений на английский язык.
ELIZA	1964	Йозеф Вайценбаум	Симуляция психотерапевта из Роджера, перефразировавшего свою реакцию с помощью нескольких грамматических правил.
LIFER	1978	Хендрикс	Естественный язык интерфейса к базе данных о кораблях ВМС США.
Jabberwacky	1982	Ролло Карпентер	Чат-бот с целью «имитировать естественный человеческий чат в интересной, развлекательной и юмористической манере».
Racter	1983	Уильям Чемберлен и Томас Эттер	Произвольно генерировал англоязычную прозу.

Все вышеизложенные программы повлияли на изобретение современных программ, использующих семантический анализ как основной алгоритм для работы.

Семантический анализ является важной подзадачей обработки естественного языка и движущая сила таких инструментов машинного обучения, как чат-боты, поисковые системы и анализ текста.

Список используемой литературы

1. Chamberlain, Bill (1984). The Policeman's Beard Is Half Constructed. UbuWeb, Warner Books. ISBN 0-446-38051-2. Retrieved 2009-07-02
2. Jabberwacky [Электронный ресурс] URL: <http://www.jabberwacky.com/> (дата обращения: 15.12.2020)
3. Jason Eisner. Natural Language Processing, Course # 601.465/665, 2020.
4. Nasta, Terry (1984-12-25). "Thief of Arts". PC Magazine. p. 63. Retrieved 25 October 2013.
5. Watson, A System Designed for Answers: The Future of Workload Optimized Systems Design". IBM Systems and Technology. February 2011. p. 3. Retrieved September 9, 2015.
6. Батура Т.В. : Семантический анализ и способы представления смысла текста в компьютерной лингвистике // Программные продукты и системы - 2016. Т. 29. № 4. С. 45–57
7. Батура Т.В. Методы автоматической классификации текстов // Программные продукты и системы – 2017. № 1. – С. 3-6.
8. Батура Т.В. Методы и системы семантического анализа текстов. // Программные продукты, системы и алгоритмы № 4, 2016
9. История и философия науки : учеб. пособие / [М. Г. Федотова и др.] ; под общ. ред. М. Г. Федотовой ; Минобрнауки России, ОмГТУ. – Омск : Изд-во ОмГТУ, 2018.
10. Лахути Д.Г., Рубашкин В.Ш. Семантический (концептуальный) словарь для информационных технологий // Научно-техническая информация. 2000. № 7. С. 1–9.
11. Мельчук И.А. Опыт теории лингвистических моделей «Смысл-Текст». М.: Языки русской культуры, 1999. 346 с.
12. Наумов М.Ю., Чистяков А.С. Применение систем искусственного интеллекта в различных сферах деятельности // Постулат. 2017. №5.
13. Падучева Е.В. Динамические модели в семантике лексики. М.: Языки славянской культуры, 2004. 608 с.
14. Поначугин А.В., Пичужкина Д.Ю., Смекалова Е.С. Голосовой помощник как технология обработки данных. // НАУКА БЕЗ ГРАНИЦ. М.: НГПУ им. К. Минина, 2020 С. 96-100.
15. Тузов В.А. Компьютерная семантика русского языка. СПб: Изд-во СПбГУ, 2003. 391 с