Применение ассоциативно-семантического препроцессора в интерактивных диалоговых системах на естественном языке

В.Е. Сачков <megawatto@mail.ru>
ФГБОУ ВО МИРЭА – Российский технологический университет,
119454, г. Москва, Проспект Вернадского, д. 78

Аннотация. В статье исследуется возможность применения ассоциативно-семантического препроцессора специальной обработки текста на естественном языке в диалоговых системах. Применение в препроцессоре ассоциаций позволяет абстрагироваться от прямого значения слова и заменить его на набор других слов. Этот эффект имеет и обратное действие: по набору слов (ассоциаций) можно восстановить искомое слово, что позволяет человеку формировать запрос на естественном языке, не зная ключевых слов или терминов той или иной предметной области, но при этом получать нужный ему результат. При семантической обработке текста с использованием ассоциаций совершенно не важен порядок слов и их количество, что позволяет человеку общаться с машиной, не формируя фразы специальным образом, так как интерактивная диалоговая система сама обработает запрос, очистив его от всего лишнего. Применение специального текстового препроцессора на основе ассоциативно-семантической обработки текста позволяет наделить интерактивные диалоговые системы способностью к пониманию темы диалога машины с пользователем, улучшить взаимодействие путем общения на естественном языке, а также упростить процесс создания и разработки диалоговых систем.

Ключевые слова: семантика; диалоговая система; метрика EMD; поисковая система; ассоциации

DOI: 10.15514/ISPRAS-2018-30(4)-13

Для цитирования: Сачков В.Е. Применение ассоциативно-семантического препроцессора в интерактивных диалоговых системах на естественном языке. Труды ИСП РАН, том 30, вып. 4, 2018 г., стр. 195-208. DOI: 10.15514/ISPRAS-2018-30(4)-13

1. Введение

Текст на естественном языке понятен пользователю, но для машины он представляется не более чем набором закодированных символов. Для

извлечения ценных данных машине необходимо решить множество задач по обработке текста на естественном языке. Такими задачами занимается специальный раздел прикладной лингвистики «Автоматическая обработка текстов на естественном языке» (Natural Language Processing, NLP). На сегодняшний день существует большое разнообразие задач обработки естественного языка, наиболее распространенными среди которых являются следующие [1]:

- поиск фрагментов текста – разделение текста на различные элементы разных типов: слова, предложения, абзацы и т. д.;
- поиск предложений (Sentence Boundary Disambiguation, SBD) – определение границ предложения;
- поиск именованных объектов (Named entity recognition, NER) – механизм поиска адресов, названий, имен, дат, или любых других именованных сущностей;
- определение частей речи (Parts of speech, POS) – классификация элементов текста на уровне предложения; предложение может быть разделено на отдельные слова и словосочетания по таким категориям, как существительные, глаголы, наречия, предлоги и т. д.;
- классификация текстов и документов – цель данной классификации в присвоении меток фрагментам, найденным в текстах и документах;
- выделение взаимоотношений – выявление связей между словами или словосочетаниями для построения семантического дерева.

Одной из приоритетных задач автоматической обработки текста, является семантический анализ – понимание содержательной и смысловой части текста. При создании диалоговых систем на естественном языке широко применяется семантика, поскольку от этих систем в первую очередь требуется понимание запроса пользователя. Эталонным примером может служить первая диалоговая система «Eliza», созданная в 1966 году и имитирующая диалог с психоаналитиком [2]. Для своего времени программа «Eliza» была прорывом, но ее алгоритм основывался на простом перефразировании вопросов и ей было далеко до понимания «смысла».

Современный пример развития диалоговых систем можно представить программой «A.L.I.C.E» [3] – это виртуальный собеседник, способный вести диалог на естественном языке. Основой «A.L.I.C.E» является язык разметки искусственного интеллекта (Artificial Intelligence Markup Language, AIML), более детально рассматриваемый в статье далее.

При создании «Eliza», «A.L.I.C.E» и тому подобных систем ключевым моментом всегда оставалась потребность в понимании «смысла» запроса пользователя на естественном языке, т.е. потребность в инструменте, способном выявить из текста информацию, о чем пользователь спрашивает систему, и привести эту информацию в формальный и понятный

вычислительной машине вид. Разработка такого инструмента до сих остается актуальной проблемой.

2. Анализ задачи

Самым сложным этапом автоматической обработки текста на естественном языке считается содержательный анализ. Для его успешного выполнения необходимы знания о том, что такое значения слов и предложений, как эти значения описать формально, как представлять и хранить смысловое содержание текста в памяти компьютера, как производить операции со значениями, переводить значения с естественного языка на формальный язык и обратно. Ответы на эти и многие другие вопросы дает компьютерная семантика, в ведении которой находится разработка моделей семантического уровня естественного языка [4].

Применение интерактивных диалоговых систем (ИДС) значительно упрощает взаимодействие вычислительной машины с пользователем за счет того, что общение между ними происходит на естественном языке. Пользователь такой системы не обязан обладать специальными навыками и знаниями, может вести с ИДС осмысленный диалог, что обеспечивает удобство использования системы и повышает уровень доверия пользователя к этой системе. Для реализация таких возможностей в ИДС часто требуется разработка сложных интеллектуальных систем, основанных на базах знаний, правилах, словарях и т.д.

Современная тенденция развития ИДС по-прежнему предполагает использование значительных инженерных и экспертных знаний. В статье [5] представлен обзор такого подхода с детальным разбором наборов данных, корпусов документов на естественном языке, с помощью которых можно создавать диалоговые системы на базе моделей машинного обучения для разных тематик.

Одним из важных свойств диалоговых систем является корректность сформированного ответа на вопрос пользователя, и очень важно уметь оценивать ответ, выдаваемый пользователю на естественном языке. Конечном результатом должна быть оценка удовлетворённости пользователя [6]. В работе [7] проведены сравнения нескольких машинных метрик оценки результатов диалоговой системы с оценками человека.

Генерация естественного языка является критическим компонентом разговорного диалога и оказывает значительное влияние как на удобство использования, так и на воспринимаемое качество. В большинстве случаев используются правила и эвристики, генерируются жесткие и стилизованные ответы без естественного изменения человеческого языка. В работе [8] описывается генератор ответов на основе семантически контролируемой структуры нейронной сети «Long Short-term Memory» (LSTM) [9], позволяющий приблизить машинные ответы к естественному языку. Другим примером построения диалоговой системы может служить работа [10], где

используются большие наборы данных для тренировки рекуррентных нейронных сетей.

Чтобы научить компьютер понимать естественный язык, его требуется оснастить механизмом, позволяющим получать, извлекать и обрабатывать содержание документа, понимать смысл слов, словосочетаний и отдельных предложений. Для этого требуется применять методы компьютерной семантики, и компьютеру нужна некая структура, чтобы хранить соответствующие знания.

Один из подходов основан на использовании концепции Semantic Web [11], на основе которой была разработана модель документов RDF [12], позволяющая хранить семантические структуры. Существенной проблемой модели RDF является сложная структура онтологий, а у языка SPARQL, предложенного в качестве стандарта для работы с RDF, имеется ряд существенных синтаксических и семантических недостатков. Кроме того, концепция Semantic Web получила ограниченное распространение среди разработчиков, поскольку приходилось создавать два одинаковых по содержанию документа, но один для «людей», а второй для «машины».

Другой современный подход к пониманию семантики текста, поиску близких по смыслу слов и определению тематики текста основан на обучении моделей нейронных сетей. Существует два конкурирующих подхода, один из которых основан на частотном вхождение в корпус документов, а другой – на модели прогнозирования. Популярным представителем первого подхода является латентно-семантический анализ (Latent semantic analysis, LSA) [13], представитель второго – набор алгоритмов word2vec [14].

Модель LSA основана на частотных подсчетах, где аналогичные слова имеют одинаковые значения в разных документах. LSA широко используется в поисковых системах для индексирования и поиска близких по смыслу слов и документов. Но у этой модели имеется ряд недостатков, в частности, разрежённость данных, игнорирование семантических ассоциаций между словами. Кроме того, модель плохо работает с большими массивами данных из-за большого потребления памяти при проведении расчетов и проигрывает word2vec по качеству результатов [15].

В модели word2Vec слово представляется в виде вектора, а сама модель предсказывает набор векторов, ближайших к исходному, по дистрибутивным признакам. Преимуществом и одновременно недостатком модели является необходимость в обучении на большом корпусе данных, при этом она более экономна по вычислительным ресурсам. Хотя модель word2Vec показывает результаты, лучшие, чем LSA, ей свойственна проблема частотного перекрытия слов, что порождает семантическую неоднозначность.

Рассмотренные модели обладают рядом других преимуществ и недостатков. Более детальное сравнение приводится, например, в работе [16].

Подводя итоги, можно заключить, что для разработки современной ИДС необходимо решение, которое учитывает семантические ассоциации, основано

на прогнозируемой модели сходства слов, решает задачу частотного перекрытия, а также обладает алгоритмом сравнения двух документов на естественном языке. С учетом этих требований было разработано специальное программное обеспечение «Ассоциативно-семантический текстовый препроцессор». В статье описываются результаты его применения при создании современных ИДС.

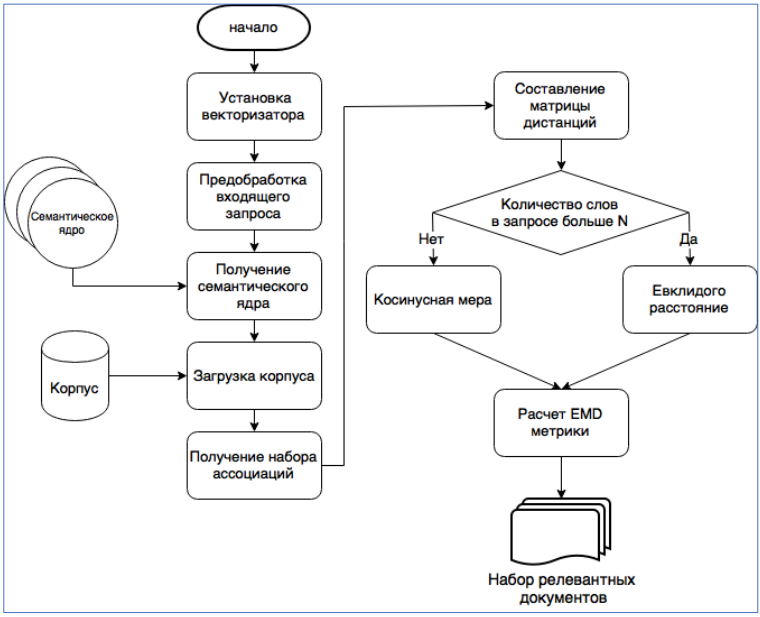


Рис. 1. Алгоритм ассоциативно-семантического поиска
Fig. 1. Algorithm of associative-semantic search

3. Ассоциативно-семантический текстовый препроцессор

Ассоциативно-семантический текстовый препроцессор (АСТП) – предназначен для предварительной нормализации текстов с целью преобразования в наборы ассоциативных семантических векторов с заданной смысловой аннотацией. Дополнительно поддерживаются функции обработки естественного языка в процессах взаимодействия естественного языка и языков компьютерных систем и роботизированных устройств (токенизация и стеммизация) для последующей обработки средствами NLP корпуса документов для проверки гипотез, обучения и статистического лингвистического анализа.

Препроцессор АСТП может использоваться как для встраивания в пакеты прикладных программ, так и в качестве самостоятельного приложения. На рис. 1 показан один из алгоритмов АСТП, отвечающий за семантический поиск близких по смыслу текстов.

Рассмотрим наиболее важные элементы алгоритма.

- Семантическое ядро – это векторное пространство семантического поля, в котором производится поиск близких по смыслу слов и ассоциаций. Оно представляет упорядоченный набор слов или словосочетаний, наиболее точно характеризующий предметную область, вид деятельности или предмет, и позволяет создавать наиболее релевантные поисковые запросы. Семантическое ядро имеет центральное ключевое слово, как правило, высокочастотное, и все остальные ключевые слова в нем ранжируются по мере убывания частоты совместного использования с центральным запросом в общей коллекции документов [17].
- Корпус – это набор специально подготовленных документов, среди которых производится семантический поиск. В наших экспериментах корпус собирался с новостного портала «РИА НОВОСТИ» (https://ria.ru), с помощью веб-скраппера (web-scraping). Корпус состоит из 250 тысяч документов на естественном языке по 13 основным разделам сайта за 3 года новостных публикаций портала.
- Модуль оценки ассоциаций – это модуль, который производит оценку и фильтрацию найденных ассоциаций в семантическом ядре.
- Метрика EMD (Earth Mover’s Distance) – это метод оценки несходства между двумя многомерными распределениями в некотором пространстве признаков, для которого задана дистанционная мера между одиночными признаками [18]. Метрика EMD вычисляет минимальную стоимость изменений или работы, требуемой для преобразования одного документа в другой. Вычисление значения EMD базируется на решении транспортной задачи.
- Матрица расстояний –матрица весов каждого слова в документе для расчёта метрики EMD

Упрощенный алгоритм ассоциативно-семантического поиска выполняет следующие шаги:

1. устанавливается тип вектора, в котором будут храниться данные;
2. производиться предобработка текста: токенизация, стеммизация, лемматизация, удаление стоп слов и т.д.;
3. загружается необходимое семантическое ядро в зависимости от темы входящего сообщения; выбор семантического ядра осуществляет обученная модель по алгоритму мультиномиального байесовского классификатора [19];
4. загружается необходимый корпус документов, в котором производится поиск;

5. для поискового запроса рассчитывается матрица расстояний в векторном пространстве семантического ядра;
6. для каждого документа корпуса создается вектор ассоциаций;
7. рассчитывается дистанция EMD;
8. возвращается отсортированный ревалентный список документов, удовлетворяющих поисковому запросу.

```
<aiml version="1.0.1" encoding="UTF-8">

    <!--HELLO-->

    <category>
        <pattern>ПРИВЕТ</pattern>
        <template>
            <random>
                <li>Привет!</li>
                <li>Добрый день!</li>
                <li>Рада Вас видеть!</li>
            </random>
            <random>
                <li>Как вас зовут?</li>
                <li>Ваше имя?</li>
                <li>Как мне вас называть?</li>
            </random>
        </template>
    </category>

    <category>
        <pattern>*</pattern>
        <that>_ КАК ВАС ЗОВУТ</that>
        <template>
            Я запомню вас как
            <set name="user">
                <star/>
            </set>
        </template>
    </category>
```

Рис. 2. Шаблон AIML
Fig. 2. AIML Template

В качестве основы для создания ИДС был выбран специальный стандарт языка шаблонной разметки AIML AIML – это язык разметки искусственного интеллекта, основанный с XML и позволяющий создавать виртуальных собеседников [20]. Использование этого языка разметки ускоряет создание диалоговой системы с применением препроцессора АСТП для ассоциативно-семантического поиска. Пример документа AIML представлен на рис. 2.

Для создания AIML документа, который позволил бы ввести диалог ассоциативно-поисковый системы с пользователем на естественном языке, были разработаны следующие смысловые блоки:

- приветствие – этот блок отвечает за начало работы с пользователем, здесь диалоговая система запоминает имя пользователя для дальнейшего использования в диалоге;
- о себе – в этом блоке диалоговая система рассказывает о себе и своих возможностях;
- ассоциации – этот блок отвечает за обработку запросов, связанных с поиском ассоциаций, с использованием препроцессора АСТП;
- поиск – этот блок отвечает за ассоциативный поиск;
- ответ по умолчанию – этот блок срабатывает только в том случае, когда ни один из предыдущих блоков не смог обработать запрос пользователя.

Упрощенная схема ведения диалога на естественном языке представлена на рис. 3.

ИДС значительно упрощает поиск близких по смыслу текстов и ассоциаций, так как позволяет пользователям общаться с поисковой системой на естественном языке, увеличивая количество формулировок запроса, которые можно использовать. Применение семантических ядер позволяет ИДС довольно точно определять тематику диалога с пользователем, и если тематика не определяется однозначно, то ИДС задает уточняющий вопрос, корректирующий подбор семантического ядра. После определения тематики диалога ИДС способна выполнить два действия.

1. Поиск ассоциаций – поиск производится путем вычисления семантического расстояния в семантическом ядре. Вычисление базируется на дистрибутивной семантике – области научных исследований, занимающейся вычислением степени семантической близости между лингвистическими единицами на основе их дистрибуционных признаков в больших массивах лингвистических данных [21]

2. Поиск близких по смыслу текстов: происходит по алгоритму, рассмотренному ранее, с вычислением значений метрики EMD.

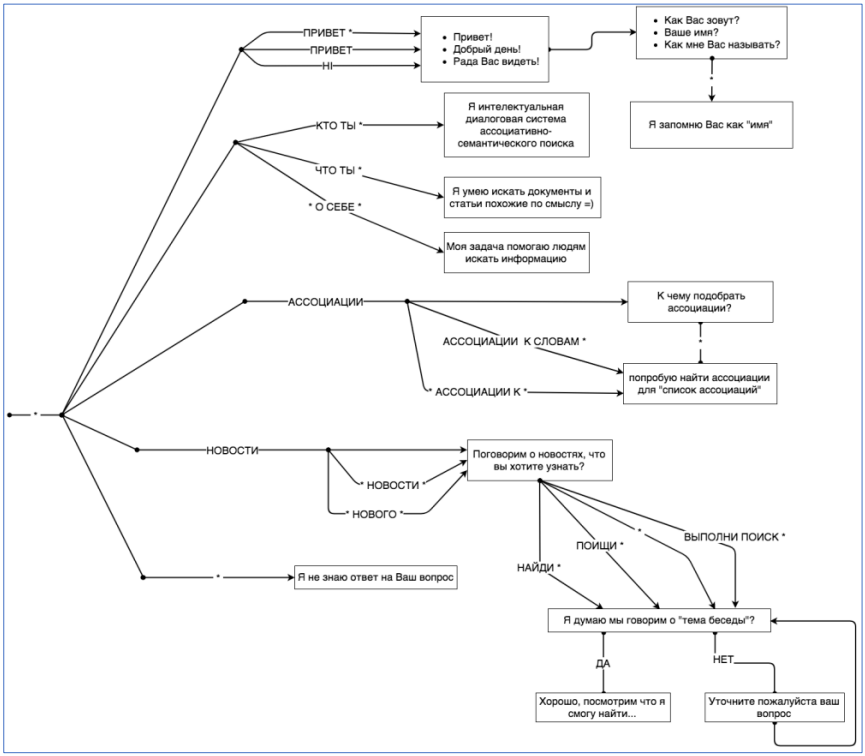


Рис. 3. Схема ведения диалога на естественном языке с пользователем
Fig. 3. Diagram of dialogue in natural language with the user

Примером применения алгоритмов АСТП с языком разметки AIML может служить работа небольшого приложения для поиска ассоциаций в семантических ядрах новостного корпуса документов. Соответствующий диалог представлен на рис. 4.

```
Loading bot/astp_bot.aiml...done (0.00 seconds)
User >>> привет
Bot  >>> Рада Вас видеть! Как вас зовут?
User >>> Watto
Bot  >>> Я запомню вас как  Watto
User >>> помоги мне найти ассоциации
Bot  >>> Скажите  Watto  к чему подобрать ассоциации?
User >>> истребитель
Bot  >>> Хорошо  Watto  попробую найти ассоциации для:  истребитель
Bot  >>>  Я думаю следующие ассоциации подходят вам: {'истребител', 'штурмовик',
 'миг', '30мки', '25см3'}
```

Рис. 4. Разговор с диалоговой системой
Fig. 4. General conversation with the dialogue system

В завершение продемонстрируем работу ИДС по поиску близких по смыслу новостей из корпуса документов на тему «Крушения самолета». Диалог представлен на рис. 5.

```
Bot  >>> Хорошо  Watto  поговим о новостях, что вы хотите узнать?
User >>> расскажи мне о крушении самолета в донбассе на украине
Bot  >>> Хорошо  Watto  ищу:  расскажи мне о крушении самолета в донбассе на укр
аине
Bot  >>> Я думаю мы говорим о политике?
User >>> да
Bot  >>> Хорошо, посмотрим что я смогу найти...
Bot  >>> Смотрю документ 19996 из 19997       doc_num         EMD
                                  text
18348    18348   5.114148  москв 18 – причин крушен малайзийск лайнер нес...
18331    18331   5.114148  москв 18 – причин крушен малайзийск лайнер нес...
18365    18365   5.114148  москв 18 – причин крушен малайзийск лайнер нес...
18343    18343   5.360942  москв 18 – заявлен киевск власт крушен малайзи...
18315    18315   5.360942  москв 18 – заявлен киевск власт крушен малайзи...
18326    18326   5.360942  москв 18 – заявлен киевск власт крушен малайзи...
18360    18360   5.360942  москв 18 – заявлен киевск власт крушен малайзи...
18333    18333   5.687893  москв 18 – лидер справедлив росс серг мирон уб...
18350    18350   5.687893  москв 18 – лидер справедлив росс серг мирон уб...
18367    18367   5.687893  москв 18 – лидер справедлив росс серг мирон уб...
18164    18164   6.503808  москв 28 – основан полага ход расшифровк черн ...
18366    18366   6.677729  москв 18 – росс готов оказа логистическ содейс...
18332    18332   6.677729  москв 18 – росс готов оказа логистическ содейс...
18349    18349   6.677729  москв 18 – росс готов оказа логистическ содейс...
18215    18215   6.754733  москв 24 – сша предостав документирова дан под...
```

Рис. 5. Поиск документов в ИДС
Fig. 5. Searching documents in an interactive dialog system

4. Заключение

Как показывают исследования возможности применения специализированного программного обеспечения АСТП в ИДС, АСТП повышает уровень взаимодействия пользователя и машины, так как позволяют вести диалог на естественном языке. В отличие, например, от модели RDF, применение данного программного обеспечения не требует большой сложности в составлении онтологий и сложных поисковых запросов по корпусам документов, не обладают такой же большой сложностью в реализации и обучении, что позволяет наделить ИДС минимально необходимым пониманием темы диалога для решения типовых задач, не привлекая большие вычислительные и интеллектуальные ресурсы.

Использование семантики, ассоциаций и современных методов компьютерной обработки текстов на естественном языке позволяет улучшить качество поиска, а также увеличить степень понимания диалога машины с человеком и наделить компьютер дополнительными инструментами для общения на естественном языке.

Благодарности

Работа выполнена за счет финансирования Министерством образования и науки РФ конкурсной части государственных заданий высшим учебным заведениям и научным организациям по выполнению инициативных научных проектов. Номер проекта 28.2635.2017/ПЧ, наименование «Разработка моделей стохастической самоорганизации слабоструктурированной информации и реализации памяти при прогнозировании новостных событий на основе массивов естественно-языковых текстов».

Список литературы

[1]. Сачков В.Е., Гильмутдинова Е.Ф., Матяш Е.Д., Акимов Д.А. Обработка и компьютерный анализ текста на естественных языках. Современная наука: актуальные проблемы теории и практики, серия Естественные и технические науки, № 12, 2016 г., стр. 57-64

[2]. Weizenbaum J. Computer Power and Human Reason: From Judgment to Calculation. Freeman and Company, New York, 1976, 281 p.

[3]. Wallace R. S. The Anatomy of A.L.I.C.E. In Parsing the Turing Test. Springer, 2009, pp. 181-210

[4]. Прикладная и компьютерная лингвистика. Под ред. И. С. Николаева, О. В. Митриениной, Т. М. Ландо. М., ЛЕНАНД, 2016, 320 стр.

[5]. Serban I. V., Lowe R., Henderson P., Charlin L., Pineau J. A Survey of Available Corpora for Building Data-Driven Dialogue Systems [электронный ресурс] URL: https://arxiv.org/abs/1512.05742 (дата обращения: 01.06.18)

[6]. Walker M., Litman D., Kamm C., Abella A. Paradise: A framework for evaluating spoken dialogue agents, In Proc. of the Eighth Conference on European chapter of the Association for Computational Linguistics, Spain, 1997, pp. 271–280

[7]. Chia-Wei Liu, Lowe R., Serban I. V., Noseworthy M., Charlin L., Pineau J. How NOT to Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation [электронный ресурс] URL: https://arxiv.org/abs/1603.08023 (дата обращения: 01.06.18)

[8]. Henderson M., Thomson B., Williams J. The Second Dialog State Tracking Challenge [электронный ресурс] URL: http://camdial.org/~mh521/dstc/downloads/handbook.pdf (дата обращения: 01.06.18)

[9]. Hochreiter S., Schmidhuber J. Long Short-Term Memory. Neural Computation, vol. 9, issue 8, 1997, pp.1735-1780

[10]. Serban I. V., Sordoni A., Bengio Y., Courville A., Pineau J. Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models [электронный ресурс] URL: https://arxiv.org/abs/1507.04808 (дата обращения: 01.06.18)

[11]. W3C Semantic web activity [электронный ресурс]. URL: https://www.w3.org/2001/sw/ (дата обращения: 25.06.2018)

[12]. Среда Описания Ресурса (RDF): Понятия и Абстрактный Синтаксис [электронный ресурс]. URL: https://www.w3.org/2007/03/rdf_concepts_ru (дата обращения: 20.02.2017)

[13]. Landauer T., Foltz P. W., Laham D. Introduction to Latent Semantic Analysis. Discourse Processes, vol. 25, issue 2-3, 1998, pp. 259–284

[14]. Tomas Mikolov, Quoc V. Le, Ilya Sutskever Exploiting Similarities among Languages for Machine Translation [электронный ресурс] URL: https://arxiv.org/pdf/1309.4168.pdf (дата обращения: 20.04.2017)

[15]. Levy O., Golberg Y., Dagan I. Improving Distributional Similarity with Lessons Learned from Word Embeddings [электронный ресурс] URL: http://www.aclweb.org/anthology/Q15-1016 (дата обращения: 01.06.2018)

[16]. Altszyler E., Sigman M., Ribeiro S., Slezak D.F. Comparative study of LSA vs Word2vec embeddings in small corpora: a case study in dreams database [электронный ресурс] URL: https://arxiv.org/pdf/1610.01520.pdf (дата обращения: 01.06.2018)

[17]. Семантическое ядро [электронный ресурс] URL: https://ru.wikipedia.org/wiki/Семантическое_ядро (дата обращения: 25.05.2017)

[18]. Rubner Y., Tomasi C., Guibas L. J. A metric for distributions with applications to image databases. In Proc. of the Sixth International Conference on Computer Vision, 1998, pp. 59-66

[19]. AIML: Artificial Intelligence Markup Language [электронный ресурс] URL: http://www.alicebot.org/aiml.html (дата обращения: 13.06.2017)

[20]. Jason D. M., Rennie, Shih L., Teevan J., Karger D. R. Tackling the Poor Assumptions of Naive Bayes Text Classifiers. In Proc. of the Twentieth International Conference on Machine Learning (ICML), 2003, pp. 616-623

[21]. Морозова Ю.И. Построение семантических векторных пространств различных предметных областей. Третья школа молодых ученых ИПИ РАН. Сборник докладов, 2012, стр. 4-11

The use of associative semantic preprocessor in the interactive dialogue systems in natural language

V.E. Sachkov <megawatto@mail.ru>
MIREA - Russian Technological University,
78 Vernadsky Avenue, Moscow 119454

Abstract. The article explores the possibility of using an associative-semantic preprocessor for special text processing in natural language. The use of associations allow to abstract from the direct meaning of a word and to replace it with a set of other words. This has also the opposite effect: by typing words (associations) a person is able to restore the search word, which allows to form a query in a natural language without knowing the keywords or terms of a particular domain but at the same time to receive the required result, in contrast to systems oriented to frequency occurrences of words. In the semantic processing of text using associations, the order of words and their number are not important, which allows a person to communicate with the machine without formulating phrases in a special way, since the interactive dialog system itself will process the request clearing everything else. The use of a special text preprocessor based on the associative-semantic processing of text allows interactive systems to be able to understand the topic of the machine's dialogue with the user, improve interaction by communicating in a natural language, and also to simplify the process of system creation and development.

Keywords: semantics; dialogue System; EMD; search engine; associations

DOI: 10.15514/ISPRAS-2018-30(4)-13

For citation: Sachkov V.E. The use of associative semantic preprocessor in the interactive dialogue systems in natural language. Trudy ISP RAN/Proc. ISP RAS, vol. 30, issue 4, 2018. pp. 195-208 (in Russian). DOI: 10.15514/ISPRAS-2018-30(4)-13

References

[1]. Sachkov V.E, Gilmutdinova E.F, Matyash E.D, Akimov D.A. Processing and computer analysis of the text in natural languages. Sovremennaja nauka: aktual'nye problemy teorii i praktiki, serija Estestvennye i tehnicheskie nauki [Journal of Contemporary Science: Actual Problems of Theory and Practice, Series of Natural and Technical Sciences], № 12. 2016. pp. 57-64 (in Russian)

[2]. Weizenbaum J. Computer Power and Human Reason: From Judgment to Calculation. Freeman and Company, New York, 1976, 281 p.

[3]. Wallace R. S. The Anatomy of A.L.I.C.E. In Parsing the Turing Test. Springer, 2009, pp. 181-210

[4]. Applied and Computational Linguistics, I.S Nikolaeva, O.V Mitrenina, T.M Lando, eds. Moscow, LENAND, 2016, 320 p. (in Russian)

[5]. Serban I. V., Lowe R., Henderson P., Charlin L., Pineau J. A Survey of Available Corpora for Building Data-Driven Dialogue Systems [online] URL: https://arxiv.org/abs/1512.05742 (accessed at: 01.06.18)

[6]. Walker M., Litman D., Kamm C., Abella A. Paradise: A framework for evaluating spoken dialogue agents, In Proc. of the Eighth Conference on European chapter of the Association for Computational Linguistics, Spain, 1997, pp. 271–280
[7]. Chia-Wei Liu, Lowe R., Serban I. V., Noseworthy M., Charlin L., Pineau J. How NOT to Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation [online] URL: https://arxiv.org/abs/1603.08023 (accessed at: 01.06.18)
[8]. Henderson M., Thomson B., Williams J. The Second Dialog State Tracking Challenge [online] URL: http://camdial.org/~mh521/dstc/downloads/handbook.pdf (accessed at: 01.06.18)
[9]. Hochreiter S., Schmidhuber J. Long Short-Term Memory. Neural Computation, vol. 9, issue 8, 1997, pp.1735-1780
[10]. Serban I. V., Sordoni A., Bengio Y., Courville A., Pineau J. Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models [online] URL: https://arxiv.org/abs/1507.04808 (accessed at: 01.06.18)
[11]. W3C Semantic web activity [online]. URL: https://www.w3.org/2001/sw/ (accessed at: 25.06.2018)
[12]. Resource Description Framework (RDF): Concepts and Abstract Syntax [online]. URL: https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/ (accessed at 20.02.2017)
[13]. Landauer T., Foltz P. W., Laham D. Introduction to Latent Semantic Analysis. Discourse Processes, vol. 25, issue 2-3, 1998, pp. 259–284
[14]. Tomas Mikolov, Quoc V. Le, Ilya Sutskever. Exploiting Similarities among Languages for Machine Translation [online] URL: https://arxiv.org/pdf/1309.4168.pdf (accessed at: 20.04.2017)
[15]. Levy O., Golberg Y., Dagan I. Improving Distributional Similarity with Lessons Learned from Word Embeddings [online] URL: http://www.aclweb.org/anthology/Q15-1016 (accessed at: 01.06.2018)
[16]. Altszyler E., Sigman M., Ribeiro S., Slezak D.F. Comparative study of LSA vs Word2vec embeddings in small corpora: a case study in dreams database [onine] URL: https://arxiv.org/pdf/1610.01520.pdf (accessed at: 01.06.2018)
[17]. Semanticheskoe yadro [Semantic kernel] [online] URL: https://ru.wikipedia.org/wiki/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B5_%D1%8F%D0%B4%D1%80%D0%BE, accessed at: 25.05.2017 (in Russian)
[18]. Rubner Y., Tomasi C., Guibas L. J. A metric for distributions with applications to image databases. In Proc. of the Sixth International Conference on Computer Vision, 1998, pp. 59-66
[19]. AIML: Artificial Intelligence Markup Language [online] URL: http://www.alicebot.org/aiml.html (accessed at: 13.06.2017)
[20]. Jason D. M., Rennie, Shih L., Teevan J., Karger D. R. Tackling the Poor Assumptions of Naive Bayes Text Classifiers. In Proc. of the Twentieth International Conference on Machine Learning (ICML), 2003, pp. 616-623
[21]. Morozova Yu.I. The construction of semantic vector spaces of different subject domains. In Proc. of the Third School of Young Scientists of the IPI RAS, 2012, pp. 4 – 11 (in Russian)