Информационное агентство INNOV.RU | Понедельник, 10 июля 2017 г. 13:51

INNOV
РОССИЙСКИЙ БИЗНЕС ON-LINE
Иннов: электронный научный журнал

Разработка вопросно-ответной системы с нейросетевым обучен современных свободных технологий

Design of Question Answering System Based on Neural Networks and Modern F Technologies

15.06.17 10:45
233

Выходные сведения: *Науменко А.М., Шелудько С.Д., Юлдашев Р.Ю., Хлебников Н.О., Радыгин В.Ю. Разработка вопросно-ответной системы с нейросетевым обучением на базе современных свободных технологий // Иннов: электронный научный журнал, 2017. №2 (31). URL: http://www.innov.ru/science/tech/razrabotka-voprosno-otvetnoy-sistem/*

Авторы:
Науменко А.М.1, Шелудько С.Д.2, Юлдашев Р.Ю. 3, Хлебников Н.О. 4, Радыгин В.Ю.5

1 студент 4-го курса бакалавриата по направлению «Информационные системы и технологии», ФГАОУ ВО Национальный исследовательский ядерный университет «МИФИ», Москва, Российская Федерация (115409, г. Москва, Каширское ш., 31), e-mail: naumenko.mephi@gmail.com.

2 студент 4-го курса бакалавриата по направлению «Информационные системы и технологии», ФГАОУ ВО Национальный исследовательский ядерный университет «МИФИ», Москва, Российская Федерация (115409, г. Москва, Каширское ш., 31), e-mail: sheludko.serg@gmail.com.

3 студент 4-го курса бакалавриата по направлению «Информационные системы и технологии», ФГАОУ ВО Национальный исследовательский ядерный университет «МИФИ», Москва, Российская Федерация (115409, г. Москва, Каширское ш., 31), e-mail: romanyuldashev@gmail.com.

4 студент 4-го курса бакалавриата по направлению «Информационные системы и технологии», ФГАОУ ВО Национальный исследовательский ядерный университет «МИФИ», Москва, Российская Федерация (115409, г. Москва, Каширское ш., 31), e-mail: nikolay.khlebnikoff@gmail.com.

5 к.т.н., доцент кафедры финансового мониторинга, ФГАОУ ВО Национальный исследовательский ядерный университет «МИФИ», Москва, Российская Федерация (115409, г. Москва, Каширское ш., 31), e-mail: vyradygin@mephi.ru.

Authors:
Naumenko A.M.1, Sheludko S.D.2, Uldashev R.Yu. 3, Hlebnikov N.O. 4, Radygin V.Yu.5

1 fourth year BA-student, specialty “Information systems and

technologies", National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russian Federation (115409, Moscow, Kashirskoe shosse, 31), e-mail: naumenko.mephi@gmail.com.

2 fourth year BA-student, specialty "Information systems and technologies", National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russian Federation (115409, Moscow, Kashirskoe shosse, 31), e-mail: sheludko.serg@gmail.com.

3 fourth year BA-student, specialty "Information systems and technologies", National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russian Federation (115409, Moscow, Kashirskoe shosse, 31), e-mail: romanyuldashev@gmail.com.

4 fourth year BA-student, specialty "Information systems and technologies", National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russian Federation (115409, Moscow, Kashirskoe shosse, 31), e-mail: nikolay.khlebnikoff@gmail.com.

5 Ph.D., assistant professor of dept. Financial Monitoring, National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russian Federation (115409, Moscow, Kashirskoe shosse, 31), e-mail: vyradygin@mephi.ru.

Ключевые слова: *машинное обучение, вопросно-ответная система, синтаксическое дерево, word2vec, nltk, pymorphy2, python, нейросетевой анализ данных*

Keyword: *machine learning, question answering system, syntax tree, word2vec, nltk, pymorphy2, python, neural network analysis*

Аннотация: *Данная статья посвящена исследованию вопроса разработки автоматизированной вопросно-о на базе современных открытых технологий семантического сжатия текста. Выполнен подробный анализ су подходов к решению задачи извлечения информации из больших объёмов текста. Показано отсутствие гото вида. Исследован вопрос построения автоматизированного решения задачи часто задаваемых вопросов на ог массивах данных. Выполнен подробный анализ современных научных работ и технологий в данной области. Р основные модели частично автоматизированного и полностью автоматизированного поиска ответов на во недостатки применяемых на сегодняшний день подходов и подчеркнута ограниченность всех решений, имеющ доступе. Рассмотрены основные семантические модели представления текста. Подробно раскрыты идеи пои основе синтаксических деревьев и векторного описания вопросов. Исследованы технологии word2vec, NLTK, ру их преимущества при реализации вопросно-ответной системы. Раскрыты основные идеи нейросетевого анал алгоритмов Continuous Bag of Words (CBOW) и Skip-gram. Приведено подробное описание разработанного про стадий его проектирования. Показаны примеры найденных решений для семантически сложных вопросов. Оп полученного полноценного программного продукта. Подчёркнута возможность применения отработанных т решения аналогичных задач. Отмечено экономическое преимущество исследованных технологий, достигаемо использования бесплатного программного обеспечения.*

Annotation: *This article is devoted to design of automatic question-answering system based on free technologies of se compression. A detailed analysis of modern solutions of knowledge extraction from big text data sets is performed. Lac using program products is shown. The ways to design of automatic frequently answering questions system are investig solutions for limited questions sets is discussed. A detailed analysis of scientific works in this field is performed. The ma semiautomatic and fully automatic semantic questions analysis are described. The disadvantages of all popular metho ways to semantic representation of text phrases are revealed. The Continuous Bag of Words (CBOW) and Skip-gram al discussed. The detailed description of designed question answering software is shown. Examples of automatically matc and semantically difficult real questions are given. The advantages of developed product are emphasized. The possibili technologies for solving similar tasks are discussed. In conclusion, the economic benefits of using free soft are emphasiz*

Введение

Анализ информации – это одна из наиболее востребованных задач, возникающих сегодня во всех областях де Современные объёмы данных и требуемая скорость их обработки побуждают людей всё чаще использовать сре анализа, базирующиеся на компьютерных технологиях. Некоторые аспекты данной области на сегодняшний де обеспечены соответствующим программным обеспечением. Это, например, статистический анализ чис распознавание печатного текста и т.д. К сожалению, до сих пор остаётся ряд широко востребованных в повсе анализа данных, не имеющих качественного и общедоступного автоматизирующего программного обеспечения. актуальна в настоящее время задача семантического сжатия больших объемов текстовой информации по опред

одним из наиболее значимых подмножеств которой является задача поиска в тексте ответа на заданный вопрос (задача полнотекстового поиска).

Одним из решений задачи полнотекстового поиска можно считать современные поисковые системы. Например, исследованию особенностей семантического поиска с использованием технологий Google посвящены работы В. Мала (V. Mala) [2], В.Н. Пху (V.N. Phu) [3]. Построение вопросно-ответной системы (question-answering system – QA system) для Википедии рассмотрено в работе Ф. Аббас (F. Abbas) [4].

К сожалению, результаты полнотекстового поиска, полученные с помощью механизмов современных поисковых систем, неудовлетворительны для тематических задач поиска. Причинами этого являются использов речевых конструкций естественного языка (например, вопросительных слов), высокая частота встречаемости ис в просматриваемом тексте и другие проблемы, связанные с отсутствием семантического анализа просматриваемых данных.

В общем случае говорить об автоматизированном поиске ответов на вопросы нельзя, так как всесторонний сем использованием только компьютерных средств является AI-полной задачей [5], предполагающей разраб интеллекта, сопоставимого по возможностям с человеческим. Тем не менее, данная задача может иметь определённых ограничений.

Альтернативой полноценного семантического поиска ответов на вопросы является подход, базирующийся на данных и наличии достаточной для выявления закономерностей подборке запросов к ним. Простейшим приме является решение задачи семантического поиска в виде набора часто задаваемых вопросов (frequently asked исследуемой тематике. В данном случае есть два возможных направления автоматизации: решения с частичной исключения человеческого фактора и решения с полноценной компьютерной автоматизацией.

Частично автоматизированные подходы

Идея подхода с использованием часто задаваемых вопросов очень проста. При таком подходе пользо осуществлять просмотр всего массива данных по искомой тематике, предлагается обратиться к краткой выдерж которой часто обращались другие посетители. Данная технология требует от пользователя времени на прочтени выявление среди них аналогичного своему, не гарантируя нахождения такового.

Решением задачи с частичной автоматизацией может быть подход на основе иерархических инструкци специалистом. Причем реализация данных инструкций может быть, как компьютеризованной, так и «бумаж например, используется в системах телефонной поддержки клиентов банков или операторов сотовой предлагается выбрать тему своего вопроса из списка предложенных тем. Дальнейший поиск ответа осуществляет с оператором поддержки, имеющим инструкцию для решения часто возникающих ситуаций, заданную в виде дре Оператор, задавая пользователю вопросы и анализируя его ответы, продвигается по дереву сверху вниз и сведения, полученные в конечном листе инструкции. Такая система эффективно отвечает на вопросы п организация и поддержка предполагают большие финансовые затраты на персонал.

Автоматизированные подходы

Автоматизированный подходы к задаче построения вопросно-ответной систему обычно также оперируют некото массивами данных и подборкой заранее известных вопросов и ответов, позволяющей осуществить обучение с работе Д. Бхардважа (D. Bhardwaj) [6] рассматривается модель построения автоматизированной вопросно формате FAQ, основанной на основе простейшего OR/AND-поиска и методов комбинаторики. Х. Баотьян (H. Baoti используют для построения вопросно-ответной системы технологию нейронных сетей.

Особую сложность вопросу построения вопросно-ответной системы могут добавлять национальные особенност на сегодняшний день есть большой ряд узкоспециализированных работ, посвящённых разработке вопросно- конкретных языковых групп. Например, в работе Силана А. (Saelan, A) [8] рассматривается построение по индонезийского языка. В работе Мегьюхота Х. (Meguehout, H) [9] показано построение вопросно-ответной си языка. Работы Медведя М. (Medved', M.) [10] и Фама С.Т. (Pham, S.T.) [11] посвящены исследованиям в данной об вьетнамского языков, соответственно.

Тем не менее, у большинства данных работ есть существенные недостатки. В основном это узко специали разработки, применение которых для задач других областей или других языковых групп невозможно. существующие на сегодняшний день промышленные разработки являются преимущественно закрытыми програ недоступными для широкого использования. В какой-то мере, к готовым технологиям, использующимися для се текста, можно отнести проект IBM Watson [12]. Данная разработка является дорогостоящим уникальным проектом, использование которого в построение вопросно-ответных систем современного интернет сообществ экономически обоснованным.

Таким образом, задача разработки полноценной вопросно-ответной системы является актуальной и востребова день. Исходя из востребованности данной тематики, в НИЯУ МИФИ была разработана вопросно-ответная систе

себе свободные технологии семантической обработки текста, предоставляемые современными вс
разработчиками, с актуальными алгоритмическими подходами обработки больших объёмов текстовой информац

Разработка вопросно-ответной системы на основе расширенной модели часто задаваемых вопросов

В основе разработанной системы лежит практика часто задаваемых вопросов в расширенном виде. В структуру
вопросов по заданной тематике, которые могут задать пользователи. Каждому из них ставится в соответствие
ответу могут соответствовать несколько вопросов. Таким образом, описанный подход сводит задачу нахожде
вопроса из базы, семантически близкого к заданному. Для того, чтобы решить эту задачу, формируются модели
каждого вопроса из базы. В качестве моделей используются синтаксические деревья и геометрические
рассматриваемой задачи сравнение подобных моделей является объективным показателем семантической близо

Синтаксическое дерево — это построенный по определенному алгоритму граф, узлами которого являют
предложения. Ребрам, соединяющим узлы, соответствует их синтаксическая связь. В узле дерева
таксономические единицы, отдельные слова предложения, или функциональные единицы, сочетания слов, котор
перестают выполнять синтаксическую функцию. Существуют четыре основных алгоритма расположения узлов в
предложения, грамматика Теньера, грамматика зависимостей, грамматика непосредственно составляющих
предложения — это алгоритм, в котором в качестве вершины дерева выступает член предложения, не являющ
отношению ни к одной другой синтаксической единице. В соответствии с грамматикой Теньера вершиной си
является глагол-сказуемое. Кроме того, вводятся понятия актанты — функциональной единицы, обязательн
сказуемому и сирконстанты — необязательной (факультативной) функциональной единицы. Грамматика
алгоритм, при котором в узлах дерева располагаются таксономические единицы. Вершиной дерева является гла
аналитическая часть. В случае составного глагола, все связи в дереве подчинительные. Грамматика непосредст
— это алгоритм, в ходе выполнения которого каждая грамматическая единица делится на две более просты
деление происходит вплоть до выделения в качестве узла отдельного слова, каждому узлу соответствует гр
среди которых все части речи, а также именная и глагольная составляющие.

Представление слова в векторном виде — сопоставление слова из словаря геометрическому вектору в прос
словарём понимается пространство конечной размерности N, равной количеству всех представляемых в в
Задачей определения семантической близости между словами занимается дистрибутивная семантика. Увел
словесного векторного пространства способствует повышению точности определения смысловой близости
некоторая критическая размерность, превышая которую, модель не приносит заметного увеличения точности.
вектора устанавливается в диапазоне от 100 до 1000. Любой алгоритм построения векторного простр
максимизации косинусного сходства между векторами семантически близких слов. Косинусное сходство определя

$$similarity = cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}},$$

где A и B — вектора, расстояние между которыми вычисляется, θ — угол между ними. Одним из наи
семантических анализаторов на сегодняшний день является word2vec — программное средство для построения
пространств, разработанное компанией Google в 2013 году [14].

Word2vec основан на двухслойной нейронной сети прямого распространения, поэтому у пользователей суш
обучить сеть на собственных текстовых корпусах и, таким образом, получить наиболее подходящую для реш
векторную модель. Результаты обучения модели зависят от выбранной пользователем модельной архитектур
реализованы два алгоритма обучения: Continuous Bag of Words (CBOW) и Skip-gram.

При использовании архитектуры CBOW алгоритм предсказывает слово, исходя из его контекста, т.е. анал
находящиеся по правую и левую стороны от данного. При этом результат работы алгоритма не зависит от поряд
Входным элементом в нейронную сеть выступает набор контекстных векторов w(t-k),.., w(t-1), w(t+1), ..., w(t+k) , а
— w(t), где w(t) — вектор предсказанного на основе контекста слова. Архитектура Skip-gram отличается
предсказывает набор слов вокруг, основываясь на данном слове. Входным вектором выступает w(t), а выхо
множество M = {w(t-k),.., w(t-1), w(t+1), ..., w(t+k)}, где M — множество векторов. Каждое слово, соответст
множества M, характеризует слово, соответствующее входному вектору. Схема работы алгоритмов CBOW и S
рисунке 1.

Работу word2vec можно разделить на пять этапов. На первом этапе происходит статистическая обработка
корпуса, то есть для каждого слова рассчитывается количество вхождений его в исходный корпус.

На втором этапе происходит сортировка слов по частоте вхождения, а также, в целях оптимизации работы с па
называемые гапаксы — слова, встречающиеся редко в сравнении с другими словами текста. Результаты р
сохраняются в хеш-таблице.

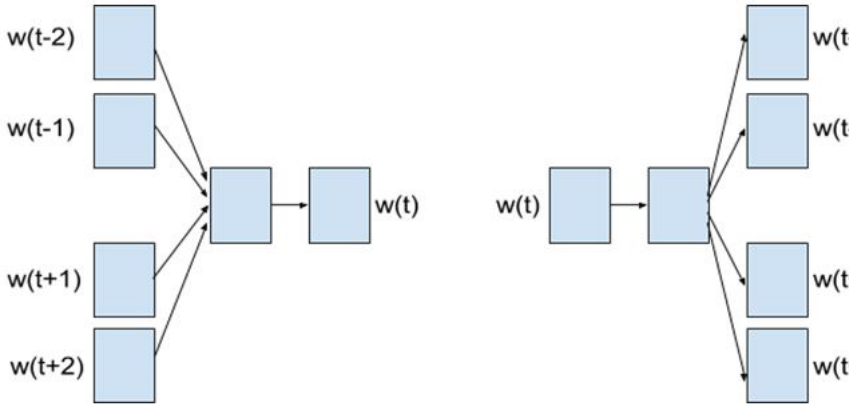

Рис. 1. Схема работы алгоритмов CBOW и Skip-gram

На третьем этапе для сжатия данных к полученной хеш-таблице применяется код Хаффмана — алгоритм оптим кодирования. В результате применения данного алгоритма чаще встречающиеся слова кодируются меньшим кол встречающиеся — большим.

Четвертый этап заключается в суб-сэмплировании самодостаточной выборки из текстового корпуса (наприме абзаца). В ходе данного процесса из выборки удаляются наиболее часто встречающиеся слова, так как с значимого смысла. Операция суб-сэмплирования применяется для уменьшения времени обучения модели.

На пятом этапе к получившейся выборке применяется один из алгоритмов обучения, рассмотренных выше: CBOV

Разработанная в НИЯУ МИФИ вопросно-ответная система включает в себя следующие элементы: управляющи на Национальном корпусе русского языка векторная Skip-gram-модель в 300-мерном пространстве [15, 16], содер лемм; база ответов на вопросы, сами вопросы, а также их векторное представление.

Так как технология Word2vec применима как к отдельным словам, так и к текстам, в данном случае являющим предложениями, то с её помощью на основе любой текстовой выборки можно построить соответствующее векто путём суммирования векторов слов, входящих в выборку. Каждое слово из предложения должно быть леммат векторную модель входят только леммы, то есть слова в словарной форме, а сама форма не имеет значения в слова. Некоторые части речи, такие как существительные, глаголы и прилагательные, в большинстве случаев яв значимыми, тогда как другие части речи, например, предлоги, союзы и местоимения не несут смысловой на анализе предложения к ним применяется фильтрация, в ходе которой из него исключаются так называемые ст препинания. Пример подготовки предложения к построению векторной модели изображен на рисунке 2.

Рис 2. Пример анализа предложения

Управляющие скрипты написаны на языке Python. Лемматизация производится средствами библиотеки pymo настоящий момент способна обрабатывать до 100000 слов в секунду, при этом потребление оперативной памяти 20 Мб [19]. Набор используемых стоп-слов взят из библиотеки Natural Language Toolkit [20], которая испо связанных с компьютерной лингвистикой и машинным обучением, и предназначена для обработки естественного

Процесс реализации вопросно-ответной системы был разделён на несколько этапов. Так как система базирует задаваемых вопросов, то на первом этапе было осуществлено проектирование базы вопросов и ответов.

На втором этапе в базу были внесены векторные представления каждого вопроса, построенные по одному алгоритмов. Затем этот же алгоритм применяется к вопросу, заданному пользователем системы. Таким образо необходимую для определения семантической близости информацию, на основе которой может делать предпол сходстве или различии заданного вопроса и вопросов из базы.

На заключительном этапе для каждого вопроса из базы и заданного пользователем вопроса были вычислены сходства. Результатом работы системы является ответ на тот вопрос из базы, косинусное сходство с которым принципа работы системы изображена на рисунке 3.

Разработанная система оформлена в виде удалённого робота (бота), отвечающего на вопросы посредством тестирования разработанной системы использовалась технология Microsoft Bot Framework Channel En тестирования вопросно-ответной системы показан на рисунке 4.

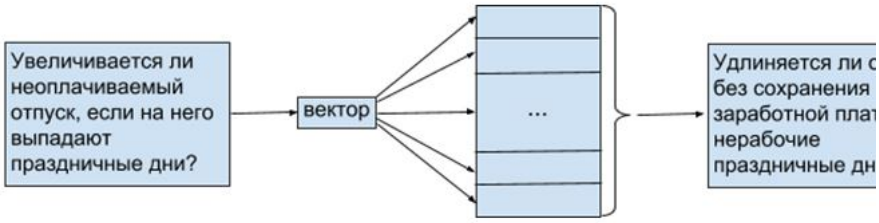

Рис 3. Схема принципа работы системы

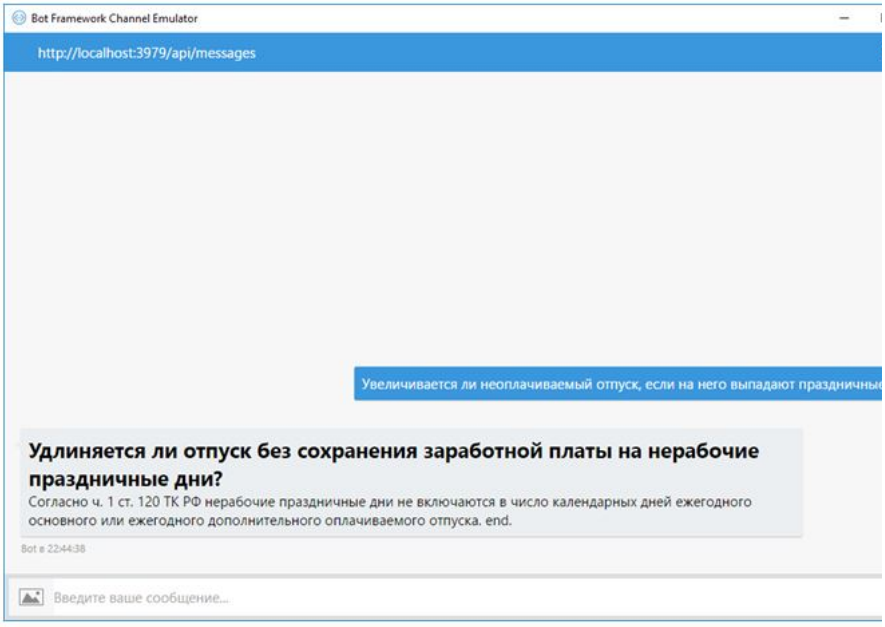

Рис 4. Пример работы системы

Тестирование системы выявило высокое качество поиска ответов, даже для вопросов семантически слож сопоставления с вопросами, для которых готовы ответы. Примеры заданных вопросов и поставленных им в соот базы FAQ показаны в таблице 1.

Примеры заданных вопросов и поставленных им в соответствие вопросов из базы FAQ

Вопрос пользователя	Вопрос в базе
Должен ли сотрудник находиться на рабочем месте, если сейчас работы нет?	Обязан ли работник находиться во врем простоя на рабочем месте?
Может ли сотрудник сменить банк, который выплачивает заработную плату?	Имею ли я право поменять банк для выплат заработной платы?
Включены ли в оклад надбавки за учёную степень?	Надбавки за ученую степень должн выплачиваться в виде премий или они уж включены в оклад?
Может ли руководитель не проходить обучение по охране труда?	Обязательно ли руководителю организаци проходить обучение по охране труда?
Является ли договор без паспортных данных работника легитимным?	Если в договоре отсутствуют паспортны данные работника, будет ли договор считатьс заключенным?
Ограничен ли период действия ученического договора?	Установлен ли максимальный срок действи ученического договора?
Можно ли уволить сотрудника, если тот подал	Может ли работник быть уволен з

поддельный паспорт?	предоставление поддельных документов пр приеме на работу?
Как индексируется зарплата сотрудника?	Необходимо ли индексировать заработну плату работника?
Предоставляется ли выходной в другой день, если в праздник был рабочий день?	Может ли быть предоставлен другой ден отдыха работнику, работавшему в выходно или нерабочий праздничный день?
Должен ли работодатель давать сотрудникам отдыхать?	Обязан ли работодатель предоставлят работникам отпуска?

Заключение

Разработанная в НИЯУ МИФИ система семантического анализа является полноценным продуктом и представл эффективное решение задачи построения вопросно-ответных систем. Она способна найти применение во научных и бизнес задачах. Представленный подход к её реализации, основанный на технологиях word2vec, nlt значительные преимущества в сравнении с большинством использующихся на настоящий момент разрабо распространяются под свободной лицензией Apache 2.0. Pymorphy2 распространяется под свободной лицензией данные технологии могут быть беспрепятственно использованы при разработке коммерческих продуктов, ч обуславливает высокую экономическую эффективность созданного подхода. В итоге, можно отметить, что п ответных систем на основе нейросетей является перспективным направлением в практическом применении ме обучения.

Библиографический список

1. Ceglarek, D.: Semantic Compression for Text Document Processing.// Proceedings of Transactions on Computational XIV, Springer, Heidelberg, 2014. – C20–48.
2. Mala V., Lobiyal D.K. Semantic and keyword based web techniques in information retrieval // Proceedings of Comp and Automation (ICCCA), International Conference, 2016 – C23–26.
3. Phu, V.N., Chau, V.T.N., Dat, N.D., Tran, V.T.N., Nguyen, T.A. A valences-totaling model for English sentiment clas and Information Systems, 2017 – C1–58.
4. Abbas, F., Malik, M.K., Rashid, M.U., Zafar, R. WikiQA - A question answering system on Wikipedia using freebase, I Proceedings of Sixth International Conference on Innovative Computing Technology (INTECH), 2016 – C185–193.
5. Raymond E.S., The New Hacker's Dictionary — MIT Press, 1996. —547 c.
6. Bhardwaj, D., Pakray, P., Bentham, J., Saha, S., Gelbukh, A. Question answering system for frequently asked questio Proceedings, Vol. 1749, 2016 – C1–5.
7. Hu, B., Lu, Z., Li, H., Chen, Q. Convolutional neural network architectures for matching natural language sentences Information Processing Systems, № 3, 2014 – C2042-2050.
8. Saelan, A., Purwarianti, A., Widyantoro, D.H. Question analysis for Indonesian comparative question // Journal o Series, Vol. 801, Iss. 1, 2017 – C1–6.
9. Meguehout, H., Bouhadada, T., Laskri, M.T. Semantic role labeling for Arabic language using case-based reasoning a Journal of Speech Technology № 2, 2017, – C1-10
10. Medved', M., Horák, A. AQA: Automatic question answering system for Czech // Lecture Notes in Computer Scien Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9924, 2016, – C270-278.
11. Pham, S.T., Nguyen, D.T. A Computational and Inferential Method for Analyzing the Semantics of Phrase and Se Question Answering System Model (VietQASM) // Proceedings of Asia Modelling Symposium 2015 – Asia 9th Intern Mathematical Modelling and Computer Simulation, 2016, – C107-112.
12. Fan, J., Kalyanpur, A., Gondek, D.C., Ferrucci, D.A., Automatic knowledge extraction from documents // IBM Jo Development, Vol. 56, Iss. 3-4, 2012 – C5:1–5:10.
13. Касевич В.Б. Структура предложения. Элементы общей лингвистики. — М.: Наука, 1977. —183 с.
14. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed Representations of Words and Phrases and th Proceedings of the 26th International Conference on Neural Information Processing Systems, USA, 2013 – C3111–3119.
15. Kutuzov A., Andreev I. Texts in, Meaning Out: Neural Language Models in Semantic Similarity Task for Russian // Pro 2015 Conference. Moscow, Russia, 2015 – C143–154.
16. http://ling.go.mail.ru – официальный сайт проекта RusVectōrēs (дата последнего обращения 22.05.2017).
17. Manning C.D., Raghavan P., Schütze H. An Introduction to Information Retrieval, Cambridge University Press, Cam —547 c.
18. Korobov M.: Morphological Analyzer and Generator for Russian and
Ukrainian Languages // Proceedings of International Conference on Analysis of Images, Social Networks and Texts, 201
19. http://pymorphy2.readthedocs.io – официальный сайт проекта pymorphy2 (дата последнего обращения 22.05.20
20. http://www.nltk.org/ – официальный сайт проекта Natural Language Toolkit (дата последнего обращения 22.05.20
21. https://docs.microsoft.com/en-us/bot-framework/cognitive-services-bot-intelligence-overview – обзор технологии Mi Channel Emulator (дата последнего обращения 22.05.2017).

References

1. Ceglarek, D.: Semantic Compression for Text Document Processing in Proceedings of Transactions on Computationa XIV, Springer, Heidelberg, 2014. – pp. 20–48.
2. Mala V., Lobiyal D.K. Semantic and keyword based web techniques in information retrieval in Proceedings of Comp and Automation (ICCCA), International Conference, 2016 – pp. 23–26.
3. Phu, V.N., Chau, V.T.N., Dat, N.D., Tran, V.T.N., Nguyen, T.A. A valences-totaling model for English sentiment class and Information Systems, 2017 – pp. 1–58.
4. Abbas, F., Malik, M.K., Rashid, M.U., Zafar, R. WikiQA - A question answering system on Wikipedia using freebase, D Proceedings of Sixth International Conference on Innovative Computing Technology (INTECH), 2016 – pp. 185–193.
5. Raymond E.S., The New Hacker's Dictionary — MIT Press, 1996. —547 p.
6. Bhardwaj, D., Pakray, P., Bentham, J., Saha, S., Gelbukh, A. Question answering system for frequently asked questio Proceedings, Vol. 1749, 2016 – pp. 1–5.
7. Hu, B., Lu, Z., Li, H., Chen, Q. Convolutional neural network architectures for matching natural language sentences Information Processing Systems, # 3, 2014 – pp. 2042-2050.
8. Saelan, A., Purwarianti, A., Widyantoro, D.H. Question analysis for Indonesian comparative question in Journal o Series, Vol. 801, Iss. 1, 2017 – pp. 1–6.
9. Meguehout, H., Bouhadada, T., Laskri, M.T. Semantic role labeling for Arabic language using case-based r International Journal of Speech Technology # 2, 2017, – pp. 1-10
10. Medved', M., Horák, A. AQA: Automatic question answering system for Czech in Lecture Notes in Computer Scienc Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9924, 2016, – pp. 270-278.
11. Pham, S.T., Nguyen, D.T. A Computational and Inferential Method for Analyzing the Semantics of Phrase and Se Question Answering System Model (VietQASM) in Proceedings of Asia Modelling Symposium 2015 – Asia 9th Intern Mathematical Modelling and Computer Simulation, 2016, – pp. 107-112.
12. Fan, J., Kalyanpur, A., Gondek, D.C., Ferrucci, D.A., Automatic knowledge extraction from documents in IBM Jo Development, Vol. 56, Iss. 3-4, 2012 – pp. 5:1–5:10.
13. Kasevich V.B. Structure of Sentence. The Elements of Common Linguistic. — Moscow, Science, 1977. —183 p.
14. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed Representations of Words and Phrases and th Proceedings of the 26th International Conference on Neural Information Processing Systems, USA, 2013 – pp. 3111–311
15. Kutuzov A., Andreev I. Texts in, Meaning Out: Neural Language Models in Semantic Similarity Task for Russian Dialog 2015 Conference. Moscow, Russia, 2015 – pp. 143–154.
16. http://ling.go.mail.ru – official site of RusVectōrēs project (last access date 22.05.2017).
17. Manning C.D., Raghavan P., Schütze H. An Introduction to Information Retrieval, Cambridge University Press, Cam —547 p.
18. Korobov M.: Morphological Analyzer and Generator for Russian and
Ukrainian Languages in Proceedings of International Conference on Analysis of Images, Social Networks and Texts, 201
19. http://pymorphy2.readthedocs.io – official site of pymorphy2 project (last access date 22.05.2017).
20. http://www.nltk.org/ – official site of Natural Language Toolkit project (last access date 22.05.2017).
21. https://docs.microsoft.com/en-us/bot-framework/cognitive-services-bot-intelligence-overview – review of Microsoft B Emulator technology (last access date 22.05.2017).

архив: 2013 2012 2011 1999-2011 новости ИТ гость портала 2013 тема недели 2013 поздравления

Свидетельство Управления Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций по Нижегородской области ИА № ТУ 52-0604 от 29 февраля 2012 г.
ISSN: 2414-5122