



Таблица 1. Полученные показатели на фрезированном фоне

Наименование сортов	5 см			10 см			15 см			5-15 см		
	период вегетации, день	урожайность, ц/га	расход воды, м ³	период вегетации, день	урожайность, ц/га	расход воды, м ³	период вегетации, день	урожайность, ц/га	расход воды, м ³	период вегетации, день	урожайность, ц/га	расход воды, м ³
Гулжахон	105,0	54,8	7900	108,0	61,3	16250	110,0	62,1	25000	108,0	62,0	14000
Илгор	120,0	61,2	8950	122,5	67,0	18200	125,0	67,9	27900	122,5	67,8	16550
Искандар	119,0	61,2	8900	122,0	69,4	18000	122,5	69,8	27250	120,0	69,6	16000
Лазурный	123,5	61,2	9300	125,0	66,4	18400	122,5	67,1	28000	124,5	67,9	17000
Мустакиллик	132,0	68,5	9800	134,5	77,4	20000	135,5	77,7	30300	134,0	78,0	18000
УзРос-7-13	132,0	67,9	9800	134,5	76,7	20000	135,5	74,6	30300	134,0	77,2	18000
	НСР _{0,05} =2,2 ц/га			НСР _{0,05} =2,1 ц/га			НСР _{0,05} =2,2 ц/га			НСР _{0,05} =2,2 ц/га		

Литература

1. Выращивание риса в Узбекистане. Рекомендации по выращиванию риса в Узбекистане, УзНИИРиса, Ташкент, 2009
2. Katayama T. Analytical studies of tillering in paddy rice. J. Imp Agri. Exp. Exp. Stat. Tokio, 1931, vot. No,7.
3. Хожамкулова Ю.Ж Годовые научные отчеты за 2014-2015 гг.

И.В. Чеховских, Е.В. Симонова

РАЗРАБОТКА АЛГОРИТМА ПОСТРОЕНИЯ ОНТОЛОГИИ ПО ТЕКСТОВОМУ ОПИСАНИЮ

(Самарский национальный исследовательский университет имени академика С.П. Королева)

Введение

Из-за постоянного роста объема информации, представленной на естественном языке, задача эффективного поиска в огромных коллекциях текстовых документов продолжает оставаться актуальной. Человеку приходится тратить большое количество времени и усилий на поиск требуемой информации в тексте.

Обработка естественного языка, или NLP (Natural Language Processing), позволяет машинам читать и понимать. Сложность задачи состоит в том, что, как правило, компьютеры «заставляют» людей говорить с ними на специализированном языке. Он должен быть однозначен и хорошо структурирован, а все



правила языка — неукоснительно выполняться. Лингвистическая структура естественного языка во много раз сложнее, и в ней постоянно присутствует целый комплекс меняющихся параметров. Контекст разговора определяет значение той или иной фразы, что означает невозможность однозначно трактовать часть словесной конструкции. Все, что может помочь в данном случае, — это «большие данные» [1].

Поэтому требуются средства для структурирования для быстрого поиска и наглядного отображения, извлечённых фактов и их отношений из текста, с помощью обработки естественного языка.

Существуют множество подходов, моделей и языков описания данных и знаний. Но наибольшую популярность последнее время набирают онтологии. До сих пор не существует единого определения для понятия онтология. В качестве классического определения, принято считать определение Томаса Грубера [2]: онтология – спецификация концептуализации, формализованное представление основных понятий и связей между ними. Построение онтологии предполагает определение классов объектов и описание их отношений с помощью одного из формальных языков.

Постановка задачи

Необходимо разработать методы и средства автоматизированного построения онтологии на основе корпуса текстов. Для этого в первую очередь требуется получить факты из текста, которые и станут основой онтологии. Существуют различные методы извлечения фактов из текста, большинство из которых основаны на машинном обучении. Но достигнуть полностью верной обработки текста невозможно, так как для каждого правила есть сотни исключений, а для каждого исключения тоже есть сотни исключений []. Однако для большинства прикладных задач достаточно и частичного разбора.

Новизна данной работы состоит в том, что построение такой системы выполняется с учетом особенностей русского языка.

Предлагаемое решение

Извлечение объектов и фактов из текстов – это часть NLP. Извлечение структурированной информации из неструктурированного текста называется Text Mining. Основная часть этого процесса посвящена определению объектов, их отношениям и свойствам в текстах [3].

На вход системы подается текст на естественном языке. Анализировать его необходимо на нескольких лингвистических уровнях, таких как лексический, морфологический, синтаксический, семантический. Перед обработкой текста, требуется точно определить границы предложения. Затем слова нормализуются (приводят их в начальную форму) и проводится полный или частичный синтаксический разбор, определяются зависимости и связи между словами в каждом их предложений предложениях.

Изначально кажется, что выделить предложения из текста не составляет никакого труда. Достаточно просто разбить текст согласно знакам препинания, маркирующие конец предложения. Но на деле все оказывается немного иначе, ведь, например, точка может обозначать и сокращение, и использоваться в



дробных числах или URL. Поэтому следует использовать более сложные алгоритмы и, для получения большей точности, завести словарь общеизвестных сокращений.

Следующий этап — полный синтаксический разбор. Это необходимо для выстраивания графа зависимостей и отношений между словами внутри предложения.

Последним этапом обработки текста является непосредственно к извлечению фактов. При помощи алгоритмов получать из неструктурированного отрывка текста, в котором все нужные нам объекты и факты будут размечены и категорированы в соответствующей онтологии. Для этого можно использовать различные обученные классификаторы.

Большинство методов основано на машинном обучении. Оно требует большого объема вводных данных. Нужно максимально покрыть лингвистической информацией обучающую выборку текстов: разметить всю морфологию, синтаксис, семантику, онтологические связи. Плюсы этого подхода в том, что он не требует ручного труда помимо создания размеченного корпуса. При необходимости такая система легко перенастраивается и переобучается. Правила получаются более абстрактными. Однако есть и минусы. Инструменты для автоматической разметки русскоязычных текстов пока не очень развиты, а существующие не всегда легко доступны. Корпуса должны быть достаточно объемными, размечены верно, единообразно и полностью. А это достаточно трудоемкий процесс. Кроме того, если что-то пошло не так, сложно отследить, где именно возникла ошибка, и точно ее исправить [3].

Заключение

Для качественной обработки текстов необходимо иметь детальное описание проблемной области с множеством логических связей, которые показывают соотношения между терминами области. Использование онтологий позволяет представить естественный языковой текст в таком виде, что он становится пригодным для автоматической обработки [4].

В дальнейшем планируется практическая реализация представленного теоретического подхода к автоматическому построению онтологии по текстовому описанию, его оценка и сравнение с существующими подходами из других групп.

Литература

1 Gruber T.R. The role of common ontology in achieving sharable, reusable knowledge bases // Principles of Knowledge Representation and Reasoning. // Proceedings of the Second International Conference. J.A. Allen, R. Fikes, E. Sandewell – eds. Morgan Kaufmann – 1991 - pp. 601-602.

2 Обработка естественного языка [Электронный ресурс]: ООО «Компьютерра-Онлайн», 1997-2018. — Режим доступа: <http://www.computer-ra.ru/92655/obrabotka-estestvennogo-yazyika-trudnosti-ponimaniya-i-sotsialnyie-seti>



3 Ландо Т. Извлечение из текстов объектов и фактов (Text mining) [Текст] – Яндекс, Отдел лингвистических технологий, 2015. – 49 с.

4 Никоненко, А.А. Обзор баз знаний онтологического типа [Текст] – Киев: Киевский национальный университет имени Т. Шевченко, 2009. – 12 с.

А.А. Шарипов, А.Р. Мавлютов, А.Ф. Атнабаев

АНАЛИЗ И ИЗВЛЕЧЕНИЕ СОДЕРЖИМОГО ИНФОРМАЦИОННЫХ РЕСУРСОВ СЕТИ INTERNET СРЕДСТВАМИ ЯЗЫКА PHP И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

(Уфимский государственный авиационный технический университет, г.Уфа)

К концу 2016 года по всему миру будет передано 1,1 зеттабайта данных. Это настолько много информации, что ни один человек даже за тысячу жизней не смог бы проанализировать. Не сложно заметить, что сайты, принадлежащие либо магазинам, либо разного рода организациям не содержат в себе столько данных, сколько содержат гиганты как Википедия или Ютуб, но даже для их анализа у человека потребуется большое количество времени.

Чаще всего найти и собрать полезную информацию вручную просто нереально, проблема не только в объеме информации, но и в допущении человеком ошибок, и это нормально, ведь человек не машина и ему свойственно совершать ошибки. Поэтому авторами была поставлена цель разработать программу для сбора, анализа и сохранения проанализированной полезной информации, которая бы решала все возможные проблемы. Изучив информацию по данной тематике стало известно, что такого рода программы в англоговорящих странах называют “Parser”. Они могут быстро обрабатывать информацию, но чаще всего платные и стоят не малых денег.

В информационных технологиях программы синтаксического анализа называются “Parser” (далее – парсер), а сам процесс получения информации с помощью программы — парсинг (от англ. Parsing, далее – парсинг), принятое в информатике определение синтаксического анализа, то есть сопоставление лексем с формальной грамматикой. Работу парсера можно сравнить с человеком, который ищет необходимую ему информацию и записывает ее, или сохраняет, чтобы использовать ее в дальнейшем. Алгоритм сопоставления описывается в математической модели на одном из языков программирования. Например, PHP, Perl, Ruby.

Независимо от того на каком формальном языке программирования написан парсер, алгоритм его действия остается одинаковым:

- выход в интернет, получение доступа к коду веб-ресурса и его скачивание;
- чтение, извлечение и обработка данных;
- представление извлеченных данных в удобоваримом виде – файлы .txt, .sql, .xml, .html и других форматах.