

ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ SEMANTIC WEB В МАШИННОМ ПЕРЕВОДЕDOI: [10.31618/ESU.2413-9335.2019.2.65.265](https://doi.org/10.31618/ESU.2413-9335.2019.2.65.265)*Морозов Юрий Алексеевич**аспирант,**Санкт-Петербургский политехнический университет**Петра Великого (СПбПУ)***АННОТАЦИЯ**

В статье рассмотрен подход применения технологий Semantic Web для улучшения возможностей машинного перевода. Приведен и описан пример использования данных Linked Open Data для реализации этого подхода.

ABSTRACT

The article discusses the approach of applying Semantic Web technologies to improve the capabilities of machine translation. An example of using Linked Open Data to implement this approach is presented and described.

Ключевые слова: semantic web, семантический веб, машинный перевод, linked open data

Keywords: semantic web, machine translation, linked open data

Введение

В настоящее время мы наблюдаем дефицит систем обработки естественного языка и машинного перевода, использующих знания из ресурсов семантического веба, таких как Linked Open Data (LOD). [1]

Такая ситуация возникает, так как все основное внимание исследователей в области машинных переводов сосредоточено на усовершенствовании существующих правил и алгоритмов. А разработчики в области Linked Open Data занимаются проблемами связывания данных и усовершенствованием соответствующих ресурсов.

Мы рассмотрим возможную стратегию для использования LOD в системах обработки естественного языка.

1.1 Semantic Web

Понятие Семантической паутины (Semantic Web) было введено Тимом Бернсом-Ли в сентябре 1998 года [1]. Семантический веб – это концепция развития интернета, ключевую роль в которой играют метаданные ресурсов, характеризующие их свойства и содержание, для реализации машинной обработки информации.

При использовании в системах для обработки языка, используются следующие ключевые технологии семантического веба:

- RDF (Resource Description Framework) - модель для представления данных в Интернете как помеченный граф «триплетов» (отношения вида «субъект - предикат – объект»).

- Связанные открытые данные (Linked Open Data, LOD) — это опубликованные структурированные данные, каждый элемент которых имеет свой URI, представлен в виде Resource Description Framework и имеет связь с другими данными. Главное отличие LOD от обычных веб-страниц в том, что они предназначены не столько для прочтения человеком, сколько для обработки компьютерными программами.

- SPARQL – язык запросов, используемый в семантическом вебе, который позволяет работать с данными из RDF.

- NIF (Natural Language Processing Interchange Format) – можно перевести как – формат для обмена естественным языком или формат для обмена между системами обработки естественного языка. Это формат, основанный на RDF, предназначенный для обмена между инструментами NLP-систем, движков систем машинных переводов и ресурсов Linked Open Data.

1.2 Машинный перевод

Термин машинный перевод понимается по крайней мере в двух смыслах. Машинный перевод в узком смысле – это процесс перевода некоторого текста с одного естественного языка на другой, реализуемый компьютером полностью или почти полностью.

Машинный перевод в широком смысле – это область научных исследований, находящаяся на стыке лингвистики, математики, кибернетики, и имеющая целью построение систем, реализующих машинный перевод в узком смысле [2]

К предпосылкам возникновения машинного перевода можно отнести предложения Чарльза Бэббиджа в 19 веке, о создании машины, использующей память для хранения словарей.

Однако, первая документально зафиксированная концепция машинного перевода принадлежала советскому изобретателю Петру Смирнову-Троянскому, получившему патент на «машину для подбора и печатания слов при переводе с одного языка на другой» в 1933 году [3], но из-за прохладного отношения к его изобретению в научном сообществе, его труды остались в тени.

Первая публичная демонстрация машинного перевода состоялась в 1954 году в Нью-Йорке и получила название «Джорджтаунский эксперимент» в ходе которого «Девушка, которая не понимает ни слова на языке Советов, набрала русские сообщения на перфокартах. Машинный мозг сделал их английский перевод и выдал его на автоматический принтер с бешеной скоростью — две с половиной строки в секунду», — сообщалось в пресс-релизе компании IBM. [4]

1.2.1 Статистический машинный перевод

Существует разные виды машинного перевода: основанные на грамматических

правилах, на примерах, гибридный машинный перевод и статистический машинный перевод.

Статистический машинный перевод является наиболее распространенным, его особенность заключается в том, что он основывается на анализе большого объема параллельных текстов. Данная система не запоминает правила, но выдает наиболее вероятный вариант соответствия. Google популяризировал эту модель, собрав десятки миллионов символов параллельных текстов за счет индексации многоязычных сайтов. [5]

2 Инструментальные средства

Из ресурсов Linked Open Data, мы остановим свой выбор на самом крупном - DBpedia. DBpedia извлекает метаданные из самой популярной мировой интернет-энциклопедии – Wikipedia. Она насчитывает огромное количество онтологий.

В качестве статистической системы машинного перевода мы будем использовать программное обеспечение с открытым исходным кодом – Moses. Основной момент заключается в том, что при получении данных (перевода) из LOD, они имеют больший приоритет, чем данные полученные переводом с помощью декодера программы. [6]

3 Идея интеграции LOD и систем машинного перевода.

Для взаимодействия ресурсов LOD, а в частности DBpedia с системой машинного перевода

– Moses необходима следующая последовательность действий:

1. Необходимо конвертировать текст, который будет переводиться в формат NIF

2. Для каждого предложения, производить вызов NER, который предназначен специально для распознавания предложений и дальнейшего перевода их на другой язык.

3. Для каждой отдельной именованной сущности API вызывает службу предназначенную для их пометки в документ.

4. Для каждого элемента, на который есть ссылка в ресурсе DBpedia, получить напрямую перевод, при помощи языка SPARQL, указывая идентификатор языка.

5. Далее, производим интегрирование полученных переводов в декодер Moses. Полученные переводы из LOD, конвертируются в совместимый NIF формат для Moses и включаются в результат декодера.

6. На экран выводятся результаты перевода.

Продемонстрируем прототип данного алгоритма для высказывания

«Санкт-Петербург – хороший город» при переводе на английский язык.

Воспользуемся фреймворком FREGRE, который поддерживает формат NIF. При конвертации данного высказывания в NIF мы получим следующий синтаксис, показанный на рисунке:

```
<http://freme-project.eu/#char=0,29>
a nif:Context, nif:RFC5147String, nif: Sentence;
nif: anchorOf "Санкт-Петербург хороший город";
nif : beginIndex "0" ;
nif: endIndex "29";
nif: firstWord <http://freme-project.eu/#char=0,0>
nif:refContext <http://freme-project.eu/#char=0,29>
nif:word <http://freme-project.eu/#char=0,15>
<http://freme-project.eu/#char=16,23>
<http://freme-project.eu/#char=24,29>
```

Рисунок 1. Синтаксис NIF для «Санкт-Петербург - хороший город»

Из листинга на рисунке выше мы видим, как данное предложение разбивается на атрибуты, характеризующее слова, индексы и их позиции. BeginIndex и endIndex — начальный и конечный

индексы строки. В теге word три атрибута с индексами начала и конца каждого слова в строке.

При вызове NER (Named Entity Recognition) происходит извлечение из DBpedia, существующего в предложении объекта. Таким

образом, после вызова NER у нас выделится наименование «Санкт-Петербург» и для него NIF примет следующий формат:

```
<http://freme-project.eu/#char=0,29>
a nif:Context, nif:RFC5147String, nif: Sentence;
nif: anchorOf "Санкт-Петербург хороший город";
nif : beginIndex "0" ;
nif: endIndex "15";
nif:nextWord:" <http://freme-project.eu/#char=16,23>";
nif:referenceContext <http://freme-project.eu/#char=0,29>
nif:sentence <http://freme-project.eu/#char=0,29>
itsrdf:taIdentRef <http://dbpedia.org/resource/Санкт-Петербург>
```

Рисунок 2. Синтаксис NIF после применения NER

Следующим шагом, является написание запроса на языке SPARQL для извлечения информации о наименовании на другом языке.

Используя синтаксис SPARQL, это будет выглядеть следующим образом:

```
SELECT distinct * WHERE {
  <http://dbpedia.org/resource/Санкт-Петербург>
  rdfs:label ?label
  filter langMatches(lang(?label), "en")
}
```

Рисунок 3. Запрос SPARQL

После того, как мы извлекли метаданные из DBpedia о интересующей нас сущности, далее необходимо произвести интеграцию с декодером системы машинного перевода. Декодер Moses, при включении функции xml input может принудительно использовать имеющиеся данные о переводах.

Запустив команду «echo '<np translation="Saint-Petersburg">Санкт-Петербург</np> хороший город' | moses -xml-input exclusive -f moses.ini» в программе Moses, мы добьемся того, что Санкт-Петербург будет переводиться как Saint-Petersburg.

Таким образом, запустив с данной командой Moses мы получим перевод «Saint-Petersburg is a good city».

Данный подход очень удобен при переводе технических терминов, имен собственных и других наименований дословный перевод которых не нужен. Например, при использовании этого подхода словосочетание Green Street is a good будет переводиться как «Гринстрит является хорошей», не переводя дословно Green Street как Зелёная улица. Также такой подход отлично подойдет при переводе различных технических терминов и названий программ таких как Adobe Audition, Microsoft Paint, After Effects и др., дословный перевод которых не нужен.

Как мы видим, по данным примерам, существует возможность интегрировать Linked Open Data с системами машинного перевода, для этого требуется реализовать следующие компоненты:

- Конвертер в формат NIF;
- Клиент взаимодействия с API NIF;
- Взаимодействие с запросами SPARQL

Мы описали процедуру перевода с использованием Linked Open Data, доступных в семантическом вебе. В описанном примере мы показали подход, позволяющий улучшить и сделать более понятным машинный перевод, в частности проблему перевода имен собственных.

Список литературы

1. Hokamp, Chris. Leveraging NLP Technologies and Linked Open Data to Create Better CAT Tools. *International Journal of Localisation*, Vol 14, pages 14–18, 2014.

2. Зубов А.В., Зубова И.И. Основы искусственного интеллекта для лингвистов. М., 2007

3. Свидетельство на изобретение. Описание машины для подбора и печатания слов при переводе с одного языка на другой или на несколько других одновременно, 1933. Режим доступа:

http://www.findpatent.ru/img_show/3401576.htm

4. Назад, в 47-й: к 70-летию машинного перевода как научного направления / О. В. Митренина // Вестник Новосибирского государственного университета. - 2017. - Т. 15, № 3. - С. 5-12. - Библиогр.: с. 10-12.

5. Дроздова К.А. Машинный перевод: история, классификация, методы. // Вестник ОмГУ – 2015 – №3 (7) – С. 156-158

6. Philipp Koehn, Hieu Hoang - Moses: Open Source Toolkit for Statistical Machine Translation, 2007

ВЛИЯНИЕ КАЧЕСТВА РАБОЧЕЙ СИЛЫ НА ЭНЕРГОЭФФЕКТИВНОСТЬ ПРОИЗВОДСТВА

DOI: [10.31618/ESU.2413-9335.2019.2.65.269](https://doi.org/10.31618/ESU.2413-9335.2019.2.65.269)

Орел Татьяна Витальевна

кандидат техн. наук, доцент,

ООО «Кимкано-Сутарский горно-обогатительный комбинат, начальник отдела кадров, г. Биробиджан

АННОТАЦИЯ

Современное высокотехнологическое оборудование горно-обогатительных комбинатов требует постоянного и систематического внимания специалистов предприятия к подготовке качественного производственного персонала основных технологических рабочих профессий. Качество подготовки персонала оценивается различными показателями: уровень специального образования, тарифный разряд по профессии рабочего и др. Уровень подготовки производственного персонала оказывает существенное влияние на энергоэффективность технологического процесса, величину управляемого ими энергоматериального потока.

ABSTRACT

Modern high-tech equipment of mining and processing enterprises requires constant and systematic attention of the enterprise specialists to the training of high-quality production personnel of the main workers of technological professions. The quality of personnel training is assessed by various indicators: the level of special education, the tariff category by profession of a worker, etc. The level of training of production personnel has a significant impact on the energy efficiency of the process, the amount of energy flow controlled by them.

Ключевые слова: энергоэффективность, эксергия, энергоматериальный поток, технологический персонал, средний тарифный разряд.

Keywords: energy efficiency, exergy, energy material flow, process personnel, average tariff category.

Современное государство не способно эффективно развиваться без существенного и постоянного повышения энергоэффективности промышленного производства, особенно базовых отраслей промышленности - горно-металлургической, тяжелого машиностроения и др. Потребность в минеральном сырье предприятий горно-металлургического комплекса (ГМК) требует вовлечение в переработку труднообогатимых руд и шихтовых материалов. Для переработки труднообогатимых руд требуется большее количество энергоматериальных ресурсов, применение комплексных многоуровневых технологий, высокопрофессиональных специалистов и рабочих технологических профессий. Обучение

производственного персонала методам сбережения и эффективного использования энергоматериальных ресурсов оказывает существенное влияние на повышение энергоэффективности технологического процесса.

Повышение эффективности использования энергоматериальных ресурсов при анализе существующих технологических линий, при разработке новых технологий, совершенствовании и оптимизации структурно-управленческих звеньев предприятия невозможно без теоретической базы и инженерных методов комплексного анализа производства с учетом уровня знаний и умений производственного технологического персонала. Для объективной оценки энергоэффективности горно-обогатительного производства,