Федеральное государственное бюджетное образовательное учреждение   
высшего образования

|  |  |
| --- | --- |
| Gerb-BMSTU_01 | «Московский государственный технический университет  им. Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана) |

ФАКУЛЬТЕТ – Информатика и управления

КАФЕДРА – Информационные системы и телекоммуникации

РАСЧЁТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

к курсовой работе на тему

«Разработка модели данных POSTagger русского языка для HypergraphDB с примерами»

Студент группы ИУ3-73 (подпись) 16.02.2017 C.C. Агальцев

Руководитель курсовой работы (подпись) 16.02.2017 А.М. Иванов

**Синонимы**

Для вывода синонимов слов предложения используется библиотека extjwnl (Extended Java WordNet Library, версия 3.1). Библиотека использует словари формата WordNet для работы со словами. Базовая единица словаря это synset описания слова, его часть речи, его индекс и связанных с ним слов.

Для загрузки словаря описан метод setUpDictionary(). Для загрузки используется метод библиотеки extjwnl Dictionary.getDefaultResourceInstance(). Затем данные о словаре сохраняются в переменную dictionary, для того чтобы к ней был доступ из других методов.

**public** **void** setUpDictionary() **throws** FileNotFoundException, JWNLException, CloneNotSupportedException {

Dictionary dictionary = **null**;

dictionary = Dictionary.*getDefaultResourceInstance*();

**if** (**null** != dictionary)

{

**this**.dictionary = dictionary;

}

}

Для работы с методами библиотеки extjwnl, которые анализируют слова требуется иметь объект класс IndexWord. Опишем метод getWord, который будет получать IndexWord по передаваемому ему слову, представленному в виде строки.

Если известна часть речь слова можно использовать метод dictionary.getIndexWord(pos, lemma), но поскольку часть речи неизвестна будем использовать метод dictionary.lookupAllIndexWords(lemma), которая находит схожие слова всех частей речи, и будем брать первое слово из получаемого списка.

**public** IndexWord getWord(String word) **throws** JWNLException {

Collection<IndexWord> wordSet = dictionary.lookupAllIndexWords(word).getIndexWordCollection();

**if** (wordSet.iterator().hasNext())

{

**return** wordSet.iterator().next();

}

**return** **null**;

}

Получив объект IndexWord можем узнать и вывести синонимы слова. Для этого используются методы PointerUtils.getSynonyms(word) или PointerUtils. getDirectHypernyms(word), в зависимости от части речи. Часть речи определяет слова определяет метод word.getPos().

**private** **void** printSynonyms(IndexWord word) **throws** JWNLException {

System.***out***.println("Синонимы " + word.getLemma() + ":");

// Для глаголов и существительных используются гиперонимы, для прилагательных синонимы. Так устроен extjwnl,

// см. PointerUtils.getSynonyms

**if** (word.getPOS() == POS.***NOUN*** || word.getPOS() == POS.***VERB***)

{

PointerTargetNodeList hypernyms = PointerUtils.*getDirectHypernyms*(word.getSenses().get(0));

hypernyms.print();

}

**else**

{

PointerTargetNodeList synonyms = PointerUtils.*getSynonyms*(word.getSenses().get(0));

synonyms.print();

}

}

Пользователь задает предложение в виде строки. Затем строка разбивается на отдельные слова по пробелам. После подключения словаря циклом проходим по всем словам и вызываем на них метод getWord(). Если слово найдено вызывается метод вывода синонимов printSynonyms().

**public** **synchronized** **void** bindDialog(IDialogueStemmerService service) {

String example = "Wizard take your magic wand and prepare!";

String[] exampleWords = example.split(" ");

// Синонимы

**try**

{

// Подключаем словарь

setUpDictionary();

**for** (**int** i = 0; i < exampleWords.length; i++)

{

IndexWord currentWord = getWord(exampleWords[i]);

**if** (currentWord != **null**)

{

printSynonyms(currentWord);

}

}

}

**catch** (Exception e)

{

System.***out***.println(e.toString());

}

}

**Пример вывода:**

**"Wizard take your magic wand and prepare!"**

Синонимы wizard:

[PointerTargetNode: [Synset: [Offset: 9640897] [POS: noun] Words: expert -- (a person with special knowledge or ability who performs skillfully)] [PointerType: [Label: hypernym] [Key: @] Applies To: noun, verb]]

Синонимы take:

[PointerTargetNode: [Synset: [Offset: 13276044] [POS: noun] Words: income -- (the financial gain (earned or unearned) accruing over a given period of time)] [PointerType: [Label: hypernym] [Key: @] Applies To: noun, verb]]

Синонимы magic:

[PointerTargetNode: [Synset: [Offset: 5961357] [POS: noun] Words: supernaturalism -- (a belief in forces beyond ordinary human understanding)] [PointerType: [Label: hypernym] [Key: @] Applies To: noun, verb]]

Синонимы wand:

[PointerTargetNode: [Synset: [Offset: 4107268] [POS: noun] Words: rod -- (a long thin implement made of metal or wood)] [PointerType: [Label: hypernym] [Key: @] Applies To: noun, verb]]

Синонимы prepare:

[PointerTargetNode: [Synset: [Offset: 126072] [POS: verb] Words: change, alter, modify -- (cause to change; make different; cause a transformation; "The advent of the automobile may have altered the growth pattern of the city"; "The discussion has changed my thinking about the issue")] [PointerType: [Label: hypernym] [Key: @] Applies To: noun, verb]]