## Задание 2

#### Поиск семантических связей между словами

WordNet - это электронный словарь-тезаурус.

WordNet состоит из слов, объединенных в синонимические ряды - синсеты. Синсет включает слова со схожим значением, слова, имеющие несколько значений, включены в несколько синсетов и могут быть причислены к различным синтаксическим и лексическим классам. Синсеты в WordNet связываются между собой семантическими отношениями:

* **гипероним**- от частного к общему (мышь -> грызун);
* **гипоним**- от общего к частному, отношение обратное гиперонимии;
* **антоним**- противоположность понятий (высокий - низкий);
* **мероним**- можно охарактеризовать данной отношение как "часть-целое" (например ножка - стол).
* и пр.

Для выполнения задания достаточно ограничиться следующими отношениями:

1. **Гипероним**.
2. **Member meronym**. Вид меронимии, характеризующий отношение между некоторой группой и членом группы. Например: вратарь - член футбольной команды.
3. **Part meronym**. Вид меронимии, характеризующий отношение между целым и его частью. Например колесо - часть автомобиля.

Используя словарь WordNet разработайте программу для поиска связей между словами. По заданной паре слов W1 и W2 необходимо найти последовательность семантических связей (если таковая существует), ведущих от одного слова к другому. Например, пусть требуется найти связь между словами "мышь" и "белка". Согласно информации из WordNet и мышь и белка являются членами синсетов, связанных отношением гиперонимии с синсетом, содержащим слово "грызун". В итоге получим связь следующего вида: (mouse ---hyp--- rodent, squirrel ---hyp--- rodent), т.е. мышь и белка связаны друг с другом через понятие "грызун".

##### Реализация и требования к коду

Вам даны реализация словаря WordNet на языке Prolog (**wordnet.zip**) и интерфейс для его использования (**wnload.zip**).

* Распакуйте wordnet.zip и wnload.zip в желаемое место на диске.
* Для использования WordNet в вашем коде необходимо подключить файл **wn.pl**, находящийся в директории **wnload/prolog.** Используем директиву :- use\_module, не забываем, что при подключении модуля **НЕ** надо указывать расширение подключаемого файла: :- use\_module(PATH/wnload/prolog/wn). Если в пути есть пробелы или нелатинские символы, то забираем путь в одинарные кавычки.
* Для того, чтобы wnload нашел wordnet, необходимо выполнить на выбор **ОДНО**из следующих действий:

1. Распаковать содержимое архива wordnet.zip в директорию по умолчанию. Для Windows это **C:\Program Files\WordNet\3.0**, для Linux и MacOS **/usr/local/WordNet-3.0**.
2. Установить переменную окружения **WNDB**, значением которой указать путь, куда вы распаковали wordnet.zip.
3. Добавить в начало своей программы следующую директиву:

:- prolog\_load\_context(directory,Dir), atom\_concat(Dir, 'WN\_PATH', WNDB\_), absolute\_file\_name(WNDB\_, WNDB), asserta(user:file\_search\_path(wndb, WNDB)).

Здесь WN\_PATH - относительный путь из директории, где лежит ваш код, в директорию, куда вы распаковали wordnet.zip.

После успешного подключения WordNet вам доступны следующие предикаты:

* wn\_s(Synset, \_, Word, PoSTag, Sense, \_). Synset - это код синсета, Word - слово из этого синсета, PoSTag и Sense - код части речи и значения, с которыми данное слово входит в данный синсет. Одно и то же слово может иметь разные значения и входить в разные синсеты.
* wn\_g(Synset, Description). Synset - код синсета, Description - описание синсета.
* wn\_hyp(Synset1, Synset2). Отношение гиперонимии между синсетами. Данное отношение означает, что синсет Synset2 содержит гиперонимы слов, содержащихся в синсете Synset1.
* wn\_mm(Synset1, Synset2). Отношение member meronym. То есть синсет Synset1 содержит меронимы слов из Synset2.
* wn\_mp(Synset1, Synset2). Отношение part meronym.

Таким образом, необходимо выразить семантические связи между словами, используя семантические связи между синсетами. Ход решения может быть следующим: по заданным синсетам научиться определять, находятся ли они в одном из требуемых отношений (hyp, mm или mp), научиться искать цепочки отношений между синсетами, а затем преобразовывать цепочки связей между синсетами в цепочки связей между словами. Другими словами, воспользоваться поиском в пространстве состояний.

**Для прохождения тестирования** НЕОБХОДИМО реализовать предикат

related\_words(Word1/PoS1/Sense1/Syn1, Word2/PoS2/Sense2/Syn2, MaxDist, Connection).

Здесь Word1, Word2 - слова, между которыми ищем связь.

PoS1, PoS2, Sense1, Sense2 - коды частей речи и значений слов, при которых данная связь существует.

Syn1, Syn2 - синсеты, в которых данные слова содержатся, когда между ними существует данная связь.

MaxDist - максимальная длина связи.

Connection - связь слов. Должна быть получена в виде последовательности семантических отношений, ведущих от слова Word1 к слову Word2. Семантические отношения должны быть представлены в виде r(W1/P1/Sen1, Rel, W2/P2/Sen2), где Rel - одно из отношений (hyp, mm или mp).

**Обратите внимание**, что предикаты, задающие данные отношения записываются как wn\_hyp, wn\_mm и wn\_mp, но в представлении отношения с помощью предиката r/3 отношения должны записываться без префикса wn\_. Более короткие связи должны выдаваться раньше, чем длинные.

**Обратите внимание**, что найденные связи между словами должны быть последовательными. Это означает, что каждая следующая пара связанных слов обязательно должна содержать слово из предыдущей пары, чтобы можно было визуально проследить последовательность связей в цепочке.

**Обратите внимание**, что цепочка должна начинаться и заканчиваться на словах, заданных в изначальном запросе. Например, в случае следующего запроса:

?- related\_words(mouse/PoS1/Sense1/Syn1, squirrel/PoS2/Sense2/Syn2, 5, WordConnection)

первый элемент списка WordConnection должен содержать слово mouse, а последний - слово squirrel.

#### Именование файлов с решениями

**НомерГруппы**-**ФамилияИнициалы**-**A2.pl**

**21215-IvanovII-A2.pl**

Примеры работы:

?- related\_words(mouse/PoS1/Sense1/Syn1, squirrel/PoS2/Sense2/Syn2, 5, Connection).

PoS1 = PoS2, PoS2 = n,

Sense1 = Sense2, Sense2 = 1,

Syn1 = 102330245,

Syn2 = 102355227,

Connection = [r(mouse/n/1, hyp, rodent/n/1), r(squirrel/n/1, hyp, rodent/n/1)] ;

PoS1 = PoS2, PoS2 = n,

Sense1 = Sense2, Sense2 = 1,

Syn1 = 102330245,

Syn2 = 102355227,

Connection = [r(mouse/n/1, hyp, gnawer/n/1), r(squirrel/n/1, hyp, gnawer/n/1)] ;

PoS1 = PoS2, PoS2 = n,

Sense1 = Sense2, Sense2 = 1,

Syn1 = 102330245,

Syn2 = 102355227,

Connection = [r(mouse/n/1, hyp, rodent/n/1), r('prairie dog'/n/1, hyp, rodent/n/1), r('prairie dog'/n/1, mm, 'Cynomys'/n/1), r('Cynomys'/n/1, mm, 'Sciuridae'/n/1), r(squirrel/n/1, mm, 'Sciuridae'/n/1)] ;

...