

**МЕТОДИКА ПРЕДСТАВЛЕНИЯ
ОМОНИМИИ В ИНФОРМАЦИОННОЙ
ТЕХНОЛОГИИ ИССЛЕДОВАНИЙ
АССОЦИАТИВНО-ВЕРБАЛЬНОЙ СЕТИ**

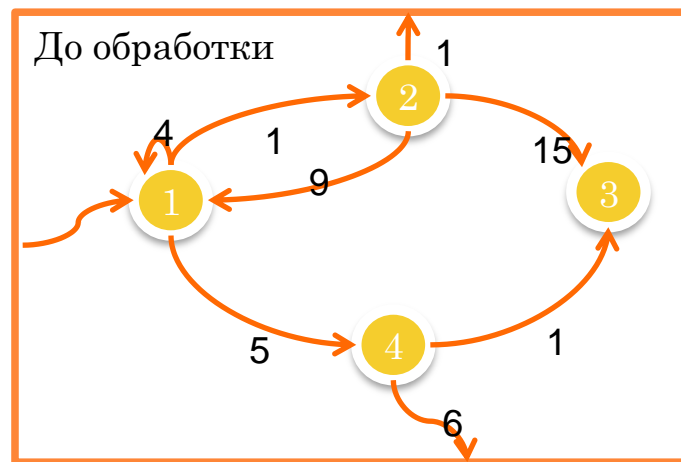
Сиренко Александр Викторович

19.05.2010

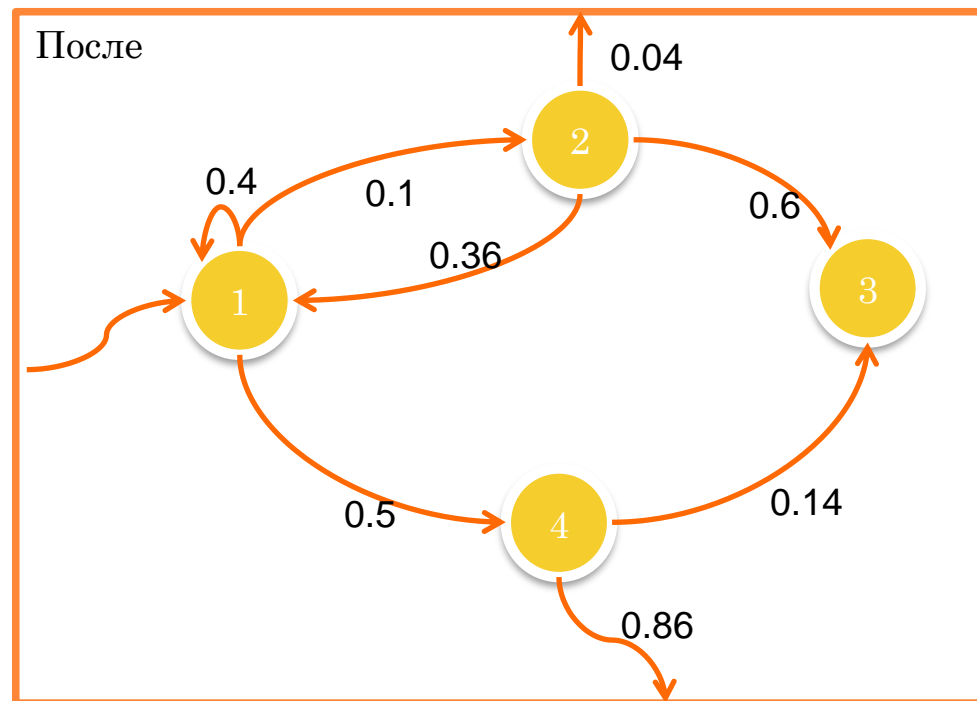
МГУП

Метод свободных ассоциаций

Стимул	Реакция	Число повторов ассоциации
ворота	открыты	6
ворота	в ад	4
ворота	дверь	4
ворота	открывать	4
ворота	рая	3
ворота	закреты	3
ворота	резные	2
ворота	к знанию	1
ворота	квадратные	1



Переход от числа ассоциаций к вероятности перехода



Особенности ассоциативно-вербальной сети

- Реально существующее и имеющее, вероятно, нейронную основу образование;
- Связи сети содержат в себе многообразные отношения: семантические, деривационные и прочие;
- Широкая вариативность элементов: словоизменительная, словообразовательная, семантическая;
- Полноценное присутствие грамматики языка в сети;
- Высокая степень семантической связности;
- Способность к самоорганизации и адаптации, освоению носителем новых элементов за счет имеющихся.



Омонимия и многозначность

Мандельброт:

$$fr^\gamma = c,$$

где f – частота слова,

r – ранг,

c – постоянная величина,

γ – величина, зависящая от текста

Ципф:

Связь между числом значений слова и его частотой $m \sim \sqrt{f}$

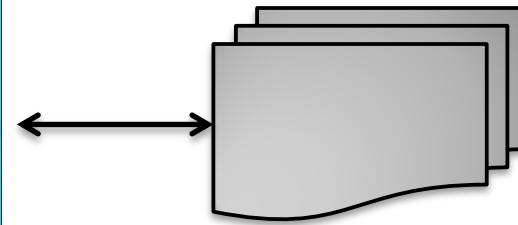
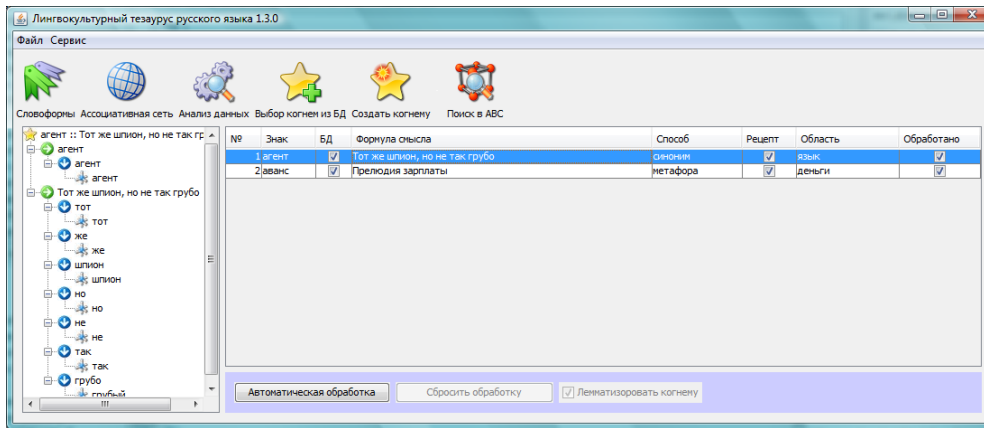
Основные методы разрешения неопределенности:

1. Вероятностные модели на основе марковских моделей.
2. Контекстно-свободные грамматики.
3. Применение тезаурусов, конкордансов.
4. Билингвальные тексты.
5. Прочие методы.

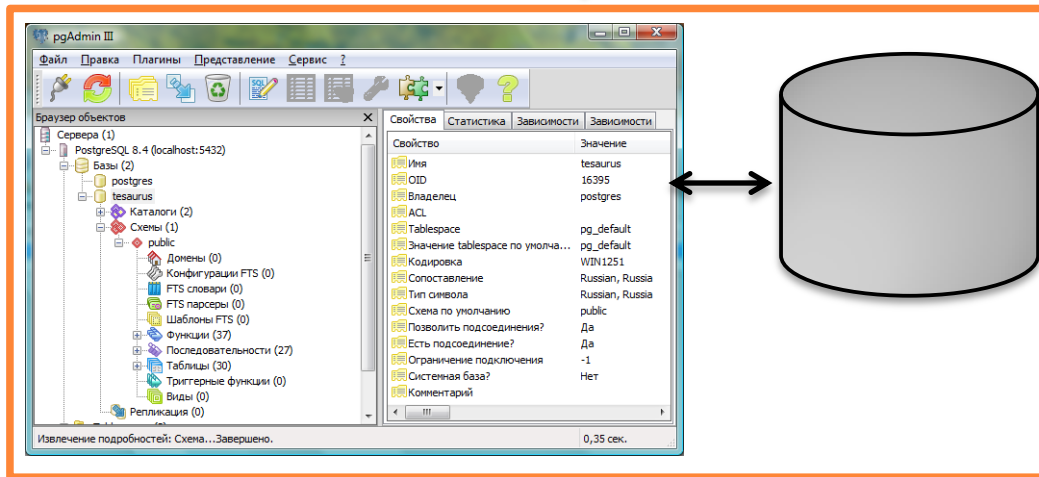


ИТ исследований в ассоциативно-вербальной сети

Программный комплекс «Лингвокультурный тезаурус русского языка»



Файлы импортируемых лингвистических данных, словари.



Реляционная база данных (PostgreSQL)



Обработка запросов

Поиск путей

Поиск кратчайших путей

Поиск путей фиксированной длины

Поиск путей диапазона длин

Лемматизированная сеть

Поиск в когнемах

Произвольные пропозиции

Стимул

Реакция

Обработка запроса: [марка :: Торговый знак, клеймо на изделиях, товарах]
Генерация волн завершена. Формирование путей
Запрос обработан

Результаты поиска в ABC

Обработка
Генерация

★ Мед :: Продукт пчел

- Мед
 - мед
 - мед
- Продукт пчел
 - продукт
 - продукт
 - пчел
 - пчел

№	Знак	БД	Формула смысла	Способ	Рецепт	Область	Обработ.
1	марка	<input checked="" type="checkbox"/>	Торговый знак, клеймо на...	описание	<input checked="" type="checkbox"/>	производство	<input checked="" type="checkbox"/>
2	марал	<input checked="" type="checkbox"/>	Крупный сибирский олень	описание	<input type="checkbox"/>	зоология	<input checked="" type="checkbox"/>
3	Мед	<input checked="" type="checkbox"/>	Продукт пчел	описание	<input checked="" type="checkbox"/>	кулинария	<input checked="" type="checkbox"/>
4	агент	<input checked="" type="checkbox"/>	Тот же шпион, но не так г...	синоним	<input checked="" type="checkbox"/>	язык	<input checked="" type="checkbox"/>
5	Марш	<input checked="" type="checkbox"/>	Свадебный Мендельсона	гипероним - гип...	<input checked="" type="checkbox"/>	музыка	<input checked="" type="checkbox"/>
6	аванс	<input checked="" type="checkbox"/>	Прелюдия зарплаты	метафора	<input checked="" type="checkbox"/>	деньги	<input checked="" type="checkbox"/>
7	микрон	<input checked="" type="checkbox"/>	мера длины	описание	<input checked="" type="checkbox"/>	мера	<input checked="" type="checkbox"/>
8	Мел	<input checked="" type="checkbox"/>	Известняк	синоним	<input checked="" type="checkbox"/>	геология	<input checked="" type="checkbox"/>

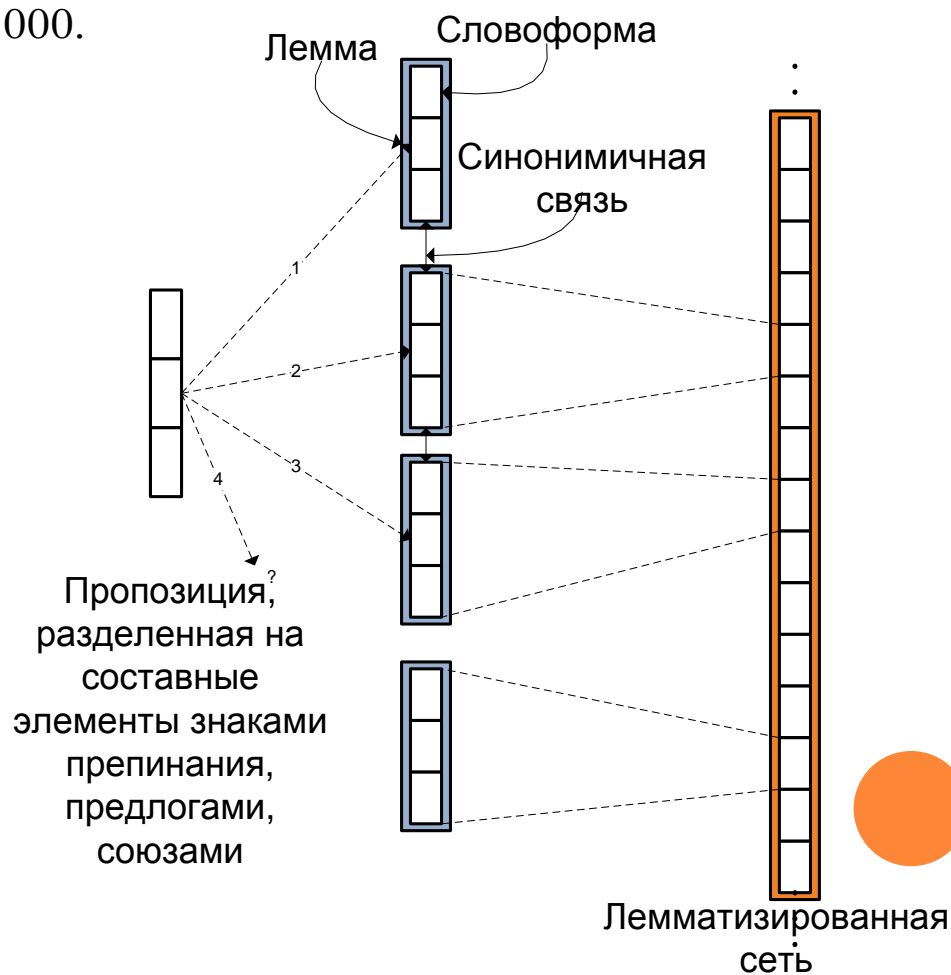
№	Путь
1	пчел = [0.011] > пчелка = [0.165] > мед
2	продукт = [0.0090] > вкусный = [0.0050] > мед
3	продукт = [0.0090] > плохой = [0.0020] > мед
4	продукт = [0.0090] > масло = [0.0040] > мед
5	пчел = [0.011] > пчелиный = [0.456] > мед
6	пчел = [0.183] > улей = [0.05] > мед
7	продукт = [0.0050] > чистый = [0.0010] > мед
8	пчел = [0.022] > пасека = [0.121] > мед
9	пчел = [0.011] > улы = [0.03] > мед

Лемматизация

Орфографический словарь iSpell

- Электронный орфографический словарь «Корректор» (120 тыс. словоформ);
- Грамматический словарь русского языка: словоизменение. (110-тыс. словоформ);
- Сводный словарь современной русской лексики 1991г. (170 тыс. словоформ);
- Русский орфографический словарь. Под ред. В. Лопатина (180 тыс. словоформ);
- Прочие источники.

Всего лемм: 127 000, словоформ: 1 300 000.



Методика представления омонимии

1. Обработка частичных омонимов лемматизатора;
2. Использование программных средств представления омонимов в сети и лингвистических данных;
3. Хранение дополнительных спецификаторов лексем – реляторов. Использование реляторов алгоритмами поиска.
4. Поддержка автоматизации разрешения лексической неопределенности событийно-статистической моделью АЭ.

```
СЛОВОФОРМА
# ЛЕММА1->РЕЛЯТОР_ЛЕММЫ1
СЛОВОФОРМА_СТИМУЛ-
>СЛОВОФОРМА_РЕАКЦИЯ-
>ЧИСЛО_СВЯЗЕЙ
...
# ЛЕММА2->РЕЛЯТОР_ЛЕММЫ2
СЛОВОФОРМА_СТИМУЛ-
>СЛОВОФОРМА_РЕАКЦИЯ-
>ЧИСЛО_СВЯЗЕЙ
...
```

Пример:

```
шерсти
# шерстить->
# шерсть->
комок->шерсти->1
клубок->шерсти->7
```



Событийно-статистическая модель ассоциативного эксперимента

Текст	Ассоциативно-вербальная сеть
Естественно-языковой текст	АВС полностью, либо часть
Естественно-языковой контекст	Пропозиция ЯЕ как ближайшая, удаленная на K шагов стимульно-реактивная окрестность некоторого слова-стимула
Пропозиция	Стимульно-реактивная цепочка длины L
Конкретные синтагматические или парадигматические отношения	Ассоциативные отношения