

Методика представления омонимии в информационной технологии исследований в ассоциативно- вербальной сети

Оглавление

| | |
|--|---|
| 1. Метод свободных ассоциаций. | 2 |
| 2. Омонимия и многозначность естественного языка. | 2 |
| 3. Информационная технология исследований в АВС. Методика представления омонимии. | 3 |

1. Метод свободных ассоциаций.

Метод свободных ассоциаций первоначально возник как средство психодиагностики. Методика проведения свободного ассоциативного эксперимента составлена таким образом, чтобы минимизировать влияние предшествующих стимульно-реактивных пар на текущую, исключить сознательное искажение реакций респондентом, обеспечить равномерную проработку понятийного пространства и его высокую степень замкнутости.

По результатам массового ассоциативного эксперимента может быть построена ассоциативно-вербальная сеть (далее АВС), относящаяся к классу событийно-статистических моделей. Допущения ассоциативного эксперимента, позволяющие составить множества узлов и ребер ассоциативно-вербальной сети:

- Стимул предъявляется в текстовом виде;
- Реакция фиксируется в текстовом виде;
- Элементы с одинаковым написанием считаются эквивалентными.

Можно выделить ключевые направления практического использования ассоциативно-вербальных сетей:

- Обучение иностранному языку;
- Информационно-поисковые системы;
- Прогнозирование реакции на сообщение;

2. Омонимия и многозначность естественного языка.

Язык как средство передачи информации имеет ряд особенностей:

1. Ориентация на психофизические возможности человека, прежде всего формирования и восприятия речи.
2. Язык является результатом функционирования самоорганизующейся системы носителей и обслуживает их взаимодействие.

3. Способность включения носителем новых элементов языка в ряд воспринимаемых и используемых.

Согласно Мандельброту, баланс интересов передающего и слушающего соблюдается принципом наименьших усилий. Согласно ему, наиболее высокочастотны слова с наименьшим числом знаков, они же имеют наибольшее число значений. Это позволяет кратко формулировать сообщение, до той степени, чтобы оно было эффективно интерпретировано принимающей стороной.

Лексическая омонимия тесно связана с понятием многозначности. В рамках созданной информационной системы данные термины равнозначны и характеризуют неопределенность смысловой нагрузки лексической единицы. Проблемы неопределенности при автоматизации обработки естественного языка возникают на любом уровне рассмотрения и разрешаются, как правило, привлечением дополнительной информации, имеющей статистическую природу.

3. Информационная технология исследований в АВС. Методика представления омонимии.

Традиционными действиями в ассоциативно-вербальной сети являются статистическая обработка ассоциативных полей и анализ связности сети.

Рассматриваемая информационная технология (далее ИТ) исследований АВС в основе своей имеет программный комплекс двухуровневой архитектуры. Базовой операцией, является поиск путей между естественно-языковыми пропозициями. Система ориентирована на групповую обработку. В качестве примера подобного запроса можно привести поиск путей в АВС между понятием и его смысловой интерпретацией (на материалах кроссвордов). При этом необходимо перевести начальную и конечную пропозиции в пространство узлов ассоциативно-вербальной сети.

АВС содержит в себе полную грамматику языка, передаваемую словоформами узлов сети. Это вносит в работу дополнительную сложность, поскольку множество словоформ естественно-языкового запроса к АВС превосходит разнообразие узлов сети.

В ИТ производится лемматизация ассоциативно-вербальной сети с помощью орфографического словаря iSpell, включающего в себя 127 000 лемм и 1 300 000 словоформ. В результате лемматизации число узлов сети сокращается на 20%, число связей на 7%.

Для увеличения доли автоматически приводимых к АВС пропозиций могут использоваться дополнительные словари: метафор, аббревиатур, устойчивых словосочетаний и прочие.

Так как элементы языка при вводе в систему представлены своим текстовым обозначением, омонимия присутствует на различных этапах работы:

- 1) В лемматизаторе ассоциативно-вербальной сети;
- 2) В исходной ассоциативно-вербальной сети (не специфицирован омоним);
- 3) В дополнительных лингвистических данных;
- 4) При приведении естественно-языкового запроса во множество узлов АВС (неоднозначность исходного запроса).

С точки зрения информационной технологии выделяются 2 типа омонимов:

- 1) Полные омонимы – слова, все словоформы которых совпадают.
- 2) Частичные омонимы – совпадение нескольких словоформ (чаще всего одной) у разных по написанию и смыслу лемм.

Методика представления омонимии в информационной технологии включает:

- 1) Разрешение частичных омонимов лемматизатора.

- 2) Набор программных средств представления омонимов в ассоциативной сети и естественно-языковых данных (экранные формы, программные функции).
- 3) Хранение дополнительных спецификаторов лексем – реляторов.
- 4) Поддержку автоматизации разрешения лексической неопределенности событийно-статистической моделью ABC.

ABC содержит 1030 частичных омонимов, что делает неопределенными 25 880 связей. В результате диалога с экспертом сформирован набор инструкций обработки ABC.

Для различения полных омонимов лексические единицы в базе данных ИТ, алгоритмах комплекса и экранных формах сопровождаются релятором. В силу трудоемкости (десятки тысяч полных омонимов в русском языке), обработка лингвистических данных на предмет полной омонимии не была проведена.

В обработке запросов ИТ играет роль именно лексическая многозначность. Разрабатываемая в работе событийно-статистическая модель ассоциативно-вербального эксперимента в своей основе имеет установление соответствия между различными единицами текста и ABC.

Возможными могут оказаться следующие сопоставления:

1. Текст - ABC в целом или какая-либо ее часть;
2. ЯЕ текст - пропозиция ЯЕ как ближайшая, удаленная на K шагов стимульно-реактивная окрестность некоторого слова-стимула;
3. Пропозиция предложения - стимульно-реактивная цепочка длины L;
4. Конкретные синтагматические или парадигматические отношения - ассоциативные отношения.

Переход от рассмотрения отдельных ассоциаций и ассоциативных полей к построению ассоциативных пропозиций и текстов позволяет обратиться к существующим методам разрешения многозначности, применяемым в обработке естественного языка в целях разрешения неопределенности ABC.