

АЛГОРИТМ ПОСТРОЕНИЯ СЕМАНТИЧЕСКОЙ СЕТИ И ЕЁ ПРИМЕНЕНИЕ

Потараев В. В.

Кафедра программного обеспечения информационных технологий, Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: vic229@rambler.ru

На протяжении последних десятилетий объём текстовой информации, хранимой в информационных системах, резко возрос. Это обуславливает актуальность автоматизированной обработки информации, в том числе текстовой. Рассмотрим алгоритм построения семантической сети для произвольного текста и способы применения этой сети.

ВВЕДЕНИЕ

В современных автоматизированных системах хранятся и обрабатываются значительные объёмы информации. Поэтому актуальным является повышение эффективности автоматизированной обработки данных, как по скорости, так и по точности обработки. Одним из инструментов улучшения эффективности обработки данных является учёт их смысловой структуры, то есть семантический анализ.

Целью данной работы является разработка методов автоматизированной обработки информации, основанных на семантическом анализе.

I. СЕМАНТИЧЕСКАЯ СЕТЬ КАК ИНСТРУМЕНТ АНАЛИЗА ДАННЫХ

Одним из способов осуществления семантического анализа данных является применение баз знаний.

База знаний – это компонент экспертной системы, предназначенный для хранения долгосрочных данных, описывающих определенную предметную область, и правил, описывающих целесообразные преобразования данных этой области [1].

Можно выделить четыре распространённые модели баз знаний:

1. Логическая модель (основанная на формулах).
2. Продукционная модель (основанная на правилах).
3. Фреймы (фрейм - это минимально возможное описание сущности объекта).
4. Семантическая сеть (ориентированный граф, отражающий понятия и их отношения) [2].

Семантическая сеть является моделью, которая наиболее соответствует современным представлениям об организации долговременной памяти человека [3]. Учитывая, что человеческая память представляет информацию относительно эффективно, то и её искусственная реализация, основанная на семантической сети, вероятно, также будет эффективной.

II. ТИПЫ ОТНОШЕНИЙ В СЕМАНТИЧЕСКОЙ СЕТИ

Одной из основных характеристик модели семантической сети является набор используемых отношений.

Например, для классификации текстовой информации может быть использована сеть, в которой узлами являются некоторые понятия, выражаемые подлежащими и дополнениями, а связями являются сказуемые [4].

Рассмотрим модель с большим количеством типов отношений. Очевидно, что увеличение их количества приведёт к усложнению построения и обработки сети, однако может позволить точнее решать поставленные перед сетью задачи, а также решать более широкий круг задач. Поэтому важным является отбор отношений, которые существенны при решении конкретной задачи.

Число связей в семантической сети и их типы выбираются её создателем исходя из конкретных целей. Необходимо отметить, что в реальном мире число связей между объектами стремится к бесконечности. Каждое отношение является, по сути, предикатом (утверждением), простым или составным [5].

К основным типам отношений, которые наиболее часто применяются в семантических сетях, можно отнести следующие:

1. Класс – элемент класса (например, «автомобиль» – «Волга»).
2. Свойство – значение («цвет» – «красный») [3].

III. АЛГОРИТМ ПОСТРОЕНИЯ СЕТИ

Покажем, что на основе приведённых типов отношений возможно построить семантическую сеть. Процесс её создания должен по возможности просто поддаваться автоматизации, поэтому стоит обратить внимание на те типы отношений, выявление которых легко автоматизируется.

Алгоритм построения сети в простейшем случае может быть описан следующим образом:

1. Загрузить словари для каждого известного типа отношений.

2. Перейти к очередному предложению текста (начиная с первого).
3. Найти в предложении подлежащее. Добавить его в сеть, если ещё не добавлено.
4. В соответствии со словарями, добавить связанные с подлежащим слова (например, добавить синонимы из словаря).
5. Найти сказуемое. Добавить его в сеть, если ещё не добавлено.
6. Найти определения. Добавить в сеть, если ещё не добавлены.
7. Найти обстоятельства.
8. Определить типы обстоятельств (например, обстоятельство места). Добавить в сеть найденные обстоятельства, если они ещё не добавлены.
9. Повторять шаги 2-8 для каждого предложения в тексте.

Данный алгоритм может быть модифицирован для выявления большего числа отношений. При определении членов предложения могут использоваться словари (существительных, глаголов, предлогов и т.п.).

IV. ПРИМЕР

Для предложения «На улице идёт сильный дождь» можно выделить отношения, представленные на рисунке 1.

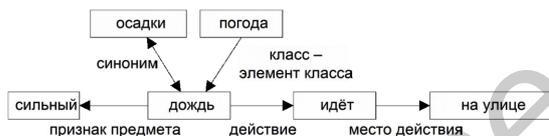


Рис. 1 – Пример семантической сети

Согласно алгоритму, в сеть по очереди добавлены слова «дождь», «осадки», «погода», «идёт», «сильный», «на улице». Словарь «класс-элемент класса» может быть задан пользователем-экспертом либо убран из рассмотрения в зависимости от особенностей решаемой задачи. Следует заметить, что предлог «на» перед существительным («на улице») обозначает указание места. Для повышения точности определения данного отношения можно использовать некоторый алгоритм вычисления падежа существительного.

Покажем, что предложенная сеть не только достаточно подробно учитывает смысл предложения, но и содержит всю информацию, необходимую для ответа на вопрос «Где идут осадки?».

«Где» - вопрос определения места. «Осадки» - это синоним слова «дождь». «Идёт» - действие, выполняемое дождём. Место, связанное с «идёт» - на улице.

Итак, информации, представленной в данной сети, достаточно, чтобы в ответ на вопрос «Где идут осадки?» сформировать предложение «Дождь идёт на улице». Для этого понадобились отношения «Синоним», «Действие», «Место действия». Таким образом, данных типов отноше-

ний может быть достаточно для ответа на вопрос определения места.

V. СПОСОБЫ РЕАЛИЗАЦИИ

Программное средство, реализующее обработку информации описанным способом, должно выполнять две основные задачи:

- Представление взаимосвязанных понятий некоторого текста (или некоторой предметной области) в виде семантической сети.
- Извлечение информации из построенной сети.

Для решения данных задач можно написать отдельные программные модули, решающие данные задачи. Кроме того, есть альтернативные варианты - использование языков описания онтологий (таких, как OWL - Web Ontology Language) и соответствующих механизмов получения новых фактов [6].

К преимуществам первого подхода можно отнести возможность максимально адаптировать построенную модель под решаемую задачу, в то время как использование уже существующих средств может значительно упростить решение поставленной задачи.

VI. ВЫВОД

Итак, в зависимости от поставленной задачи могут быть выбраны различные виды отношений. Семантическая сеть может быть использована для решения различных задач - классификации текстов, ответ на вопрос и др.

Предложенная семантическая сеть способна довольно подробно отразить смысл текста. Её построение может быть автоматизировано. Благодаря относительно большому количеству видов отношений она позволяет сформировать ответ на вопрос. Данная сеть может быть реализована различными способами, хотя её реализация является достаточно сложной задачей.

1. Базы знаний экспертных систем. [Электронный ресурс]. - Электронные данные. - Режим доступа: <http://daxnow.narod.ru/index/0-18>. Дата доступа : 25.03.2017.
2. Гаврилова, Т. А. Базы знаний интеллектуальных систем / Т. А. Гаврилова, В. Ф. Хорошевский - СПб. : Питер, 2000. - 384 с.
3. Масленникова, О. Е. Основы искусственного интеллекта [Электронный ресурс] : учеб. пособие / О. Е. Масленникова, И. В. Гаврилова. - М.: ФЛИНТА, 2013. - Режим доступа: <http://search.rsl.ru/ru/record/01007574162>.
4. Потараев, В. В. Метод классификации текстовой информации на основе семантической сети / В. В. Потараев // Апробация.- 2016. - №1. - С. 56-58.
5. Рахимова Д. Р. Построение семантических отношений в машинном переводе // Вестник КазНУ им. аль-Фараби. Серия «Математика, механика и информатика». - Алматы, 2014. - №1. - С. 90-101.
6. Мошкин, В. С., Ярушкина, Н. Г. Методики построения нечётких онтологий сложных предметных областей // Материалы V международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем». - Минск : БГУИР, 2015. - С. 401 - 406.