

ПРИМЕНЕНИЕ СЕМАНТИЧЕСКОЙ СЕТИ ДЛЯ ПРЕДСТАВЛЕНИЯ И ОБРАБОТКИ ДАННЫХ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Потараев В.В.

Серебряная Л.В. – канд. техн. наук, доцент

В современных информационных системах содержится огромное количество текстовой информации. Это обуславливает актуальность автоматизированной обработки информации, в том числе текстовой. Рассмотрим представление текстовых данных в виде семантической сети и возможные способы применения этой сети.

В современных информационных системах обрабатываются большие объёмы информации. Поэтому актуальным является повышение эффективности автоматизированной обработки информации. Семантический анализ информации является довольно эффективным способом обработки данных, учитывающим их смысловую структуру.

Целью данной работы является разработка эффективных методов автоматизированной обработки информации.

Одним из инструментов для осуществления семантического анализа данных являются базы 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].

Существует четыре основных модели базы 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].

Модель, основанная на семантической сети, наиболее соответствует современным представлениям об организации долговременной памяти человека [3]. Если учесть, что человеческая память представляет информацию довольно эффективно, то и её искусственная реализация в виде семантической сети также, вероятно, будет эффективной.

Семантическая сеть может представлять собой граф, в котором узлами являются некоторые понятия, выражаемые подлежащими и дополнениями, а связями являются сказуемые. Такая модель может быть использована для классификации текстовой информации [4]. Рассмотрим модель, имеющую большее количество типов отношений. Очевидно, что увеличение их количества может позволить более точно решать поставленные перед сетью задачи, а также решать более широкий круг задач.

Количество типов отношений в семантической сети определяется её создателем, исходя из конкретных целей. В реальном мире их число 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].

Этот список возможных отношений можно расширить. Добавим в него следующие типы: “объект-свойство”, “принадлежность”, “синонимичность”, “признак предмета”, “действие”, “место действия”.

На основе перечисленных типов отношений можно создать семантическую сеть. Процесс её создания должен по возможности просто поддаваться автоматизации, поэтому имеет смысл отдельно выделить те типы отношений, определение которых легко автоматизируется.

Алгоритм создания 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 дополнен для поддержки большего числа отношений. Для определения членов предложения в простейшем случае могут использоваться словари (существительных, глаголов, прилагательных и т.п.).

Рассмотрим построение семантической сети на конкретном примере.

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

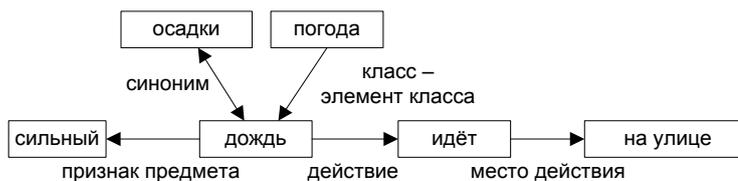


Рис. 1 – Пример семантической сети

В соответствии с алгоритмом, в сеть последовательно добавлены слова «дождь», «осадки», «погода», «идёт», «сильный», «на улице». Словарь «класс-элемент класса» может быть создан пользователем-экспертом либо вычеркнут из рассмотрения в зависимости от специфики решаемой задачи. Следует отметить, что предлог «на» перед существительным («на улице») означает указание места. Для повышения точности определения данного отношения можно использовать некоторый алгоритм вычисления падежа существительного.

Покажем, что данная сеть не только довольно подробно учитывает смысл предложения, но и содержит достаточно информации 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.

Таким образом, информации, отражённой в данной сети, достаточно, чтобы в ответ на вопрос «Где идут осадки?» можно было сформировать предложение «Дождь идёт на улице». Для ответа на вопрос понадобились отношения «Синоним», «Действие», «Место действия». Следовательно, данных типов отношений может быть достаточно для ответа на вопрос определения места.

Итак, построение рассмотренной сети является вполне автоматизируемой задачей. Предложенная семантическая сеть способна довольно подробно отразить смысл текста. Благодаря достаточно большому количеству типов отношений она позволяет сформировать ответ на вопрос. Тем не менее, её реализация представляет собой довольно сложную задачу.

Список использованных источников:

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.

ЭВОЛЮЦИОННЫЙ ПОДХОД В АВТОМАТИЗАЦИИ ПРОЦЕССА НАГРУЗОЧНОГО ТЕСТИРОВАНИЯ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Проведенцев Е. С.

Скобцов В. Ю. – канд. техн. наук

Невозможно представить успешное существование современного программного продукта и его конкурентоспособности без тщательного планирования развития. Одним из ключевых аспектов подобного планирования является инженерия производительности. Она включает в себя тестирование и прогнозирование загрузки, надежности работы продукта, его стабильности и производительности. Подобные процессы весьма трудоемки и требуют, как правило, участия квалифицированных и компетентных в этой области кадров, что влечет за собой вложение немалых денежных средств. Поэтому важной задачей является разработка алгоритмов и методов, позволяющих автоматизировать процессы тестирования и прогнозирования загрузки и производительности веб-сервиса, а также построение запросов, которые могут применяться для дальнейшего мониторинга производительности веб-сервиса.

В данной работе представлен способ автоматизации процессов тестирования и прогнозирования максимальной нагрузки и производительности веб-сервиса, а так же построение тестовых сценариев для мониторинга производительности веб-сервиса. На вход алгоритма подается JSON-schema сервиса, а также набор запросов к сервису – начальная популяция. Целью алгоритма является выведение популяции