

# ОРГАНИЗАЦИЯ РАБОТЫ СЕМАНТИЧЕСКОЙ СЕТИ В РЕЖИМЕ ДИАЛОГА

Огородник Р. В., Серебряная Л. В.

Кафедра программного обеспечения информационных технологий, Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: aharodnik@gmail.com, l\_silver@mail.ru

*Работа посвящена построению семантической сети, отвечающей на вопросы с помощью накопленных знаний. Все вопросы классифицируются в соответствии с точностью ожидаемого ответа. В докладе предложен способ структурирования информации, автоматизации процесса получения новых знаний, а также определены основные нерешенные задачи в рамках исследования и обозначены пути их решения.*

## ВВЕДЕНИЕ

Ситуация, сложившаяся в современном информационном пространстве, требует увеличения скорости и качества поиска информации, которую предоставляет информационное общество. В связи с этим одной из главных задач, связанных с информацией, является процесс выделения знаний из информационного многообразия. Отсутствие структурированных знаний, ввиду лавинообразно растущего количества новых данных и знаний влечет за собой проблемы для решения задач такого рода и налагает дополнительные трудности.

Цель работы заключается в изучении реакции семантической сети на различные вопросы, а также их классификация с последующим анализом алгоритмов работы сети с каждым типом вопросов.

Для более точного анализа вопросов была предложена их классификация по ожидаемой точности выдаваемого ответа. В результате все ответы делятся на следующие: точная цитата, примерное воспроизведение, ответ минимально допустимой точности.

## I. ЭТАПЫ РАБОТЫ

На первом этапе происходит сбор информации и наполнение модели данными. Для этого выполняется автоматический или ручной сбор информации, подтверждаемой экспертом и осуществляющим обучение сети. Человек задает достоверность получаемых системой знаний и прорабатывает ситуации возможных ошибок системы. Система собирает данные по строго обозначенному шаблону и интегрирует его в сеть на основе заданных вопросов и ранее полученной информации от эксперта, формирует связи семантической сети. Информация, получаемая от эксперта на этапе обучения, не подвергается сомнению, но в случае противоречий, возникающих в дальнейшей работе, система способна указывать на несоответствия и исправлять ошибки. Требования к предоставляемой информации для обработки анализатором – это однозначность, струк-

турированность, отсутствие ложных и противоречивых данных. При нарушении этих требований система может тратить довольно много ресурсов на обнаружение ошибок, вполне может принять ложные знания за истинные, впоследствии неправильно интерпретировав новые знания на основе старых.

Семантическая сеть представляет собой структурированное хранилище данных, которые связаны отношениями, и ответ на вопрос определяется как наиболее вероятный для данного вопроса с данной степенью точности. Связи сети можно разделить на категории: определяющие тип объектов, количественные, временные, атрибутивные и логические. Функциональные связи ввиду их большого количества вынесены в структуру данных, что ограничивает возможности поиска. Однако наложение дополнительных связей, определяющих тип объекта для этих данных, позволяет пользоваться ими при поиске правильного ответа с высокой степенью свободы. Пространственные связи используются только в качестве возможности их объективного определения (например, «граничит» вместо «недалеко»), чтобы избежать субъективной оценки как на этапе обучения системы, так и в ходе работы с ней. Таким образом, предложенную семантическую сеть можно классифицировать как неоднородную, бинарную сеть с вышеперечисленными типами отношений.

Во время тестовой работы сеть допускала ошибки в ходе получения новых знаний. Это связано с неоднозначностью лексических значений некоторых слов и недостаточностью определения некоторых отношений. Данные проблемы решались дополнительными уточнениями у эксперта на этапе обучения и введением новых отношений для более сложных утверждений, чтобы в дальнейшем интегрировать новые знания в семантическую сеть.

## II. РАБОТА С НЕДОСТОВЕРНЫМИ ДАННЫМИ

В ходе построения семантической сети было уделено внимание работе с недостоверными дан-

ными и представлению нечетких знаний. Информация, предлагаемая семантической сети экспертом в режиме обучения, заносится в неё с достоверностью 1. Информация, получаемая сетью в режиме работы, заносится в сеть с достаточно малой достоверностью  $\epsilon$  (было выбрано значение  $\epsilon = 0.01$ ). В случае обнаружения прямого несоответствия новых данных текущим, производится декрементирование их значений на величину  $\epsilon$ . В случае, если полученные знания базировались на нечетких данных, и это могло привести к получению нескольких вариантов ответов с разной степенью вероятности, выводятся все варианты ответов, вероятность которых не превышает  $\epsilon$ , с выводом соответствующих вероятностных значений. Если некоторые данные в процессе работы получили высокий процент недостоверности (в работе эта величина была взята равной  $10 \cdot \epsilon$ ), система сообщает об этом, и эксперт может пересмотреть достоверность этих данных, восстановив их в 1. В случае, если эксперт пересмотрел достоверность данных для какой-либо связи, то все связанные вершины также проверяются на достоверность.

### III. Недостатки и улучшения

Предложенная семантическая сеть может работать только с данными, представленными на естественном языке и соответствующими заранее определенному шаблону. Кроме того, для корректной работы сети рекомендовался шаблон вопросов, не допускающих введения в систему ложных знаний из-за возможных ошибок при разборе текстовой информации. Поскольку в ходе работы не предлагалось исправлять ошибки, дополнительным требованием к эксперту было построение грамматически и орфографически корректных предложений.

Лингвистический аспект получения знаний был разделен на две категории: морфологическую и синтаксическую. Первая из которых включает в себя распознавание слов, словоформ, определение на основе грамматического разбора слова его формы. Определение возможной части речи для последующего связывания слов при синтаксическом разборе возможно уже на этапе морфологического анализа. Синтаксический анализ предложения с определением роли слова в предложении помогает найти связи, понять структуру предложения и отношения между словами для последующего построения соотношений. В качестве хранилища слов и словоформ предлагается использовать префиксные деревья, где будут храниться основы слов. В смешанных

структурах с использованием списков и деревьев планируется хранить возможные грамматические формы слова. Преимущество такого подхода в том, что не хранятся длинные цепочки повторяющихся символов для одного и того же слова, стоящего в разных формах. Таким образом, уменьшается размер файла со словарем. К примеру, для русского языка это дерево вместе с дополнительными данными занимает порядка 2 мегабайт. Скорость поиска по такому дереву значительно выше, чем по отсортированному файлу. Теперь она зависит не от размера всего словаря, а только от длины искомого слова. Другая решаемая задача – это исправление ошибок. Как только слово не находится в словаре, зато обнаруживается там другое слово на коротком расстоянии Левенштейна до искомого меньше заданного значения точности, то принимается решение об исправлении.

В качестве развития данной семантической сети предполагается внедрение в систему скрытых структур знаний на основе обобщений ранее полученных данных. Имплиционные знания будут определяться экспертом на основе обобщений имеющихся структур. Затем во время работы система сама определит эти структуры и построит на их основе различные аналитические методы получения знаний.

### ЗАКЛЮЧЕНИЕ

В ходе исследования были определены различные типы вопросов, с которыми придется работать семантической сети. Предложенная модель организации знаний позволяет наиболее выгодно использовать знания экспертов, минимизировав временные затраты человека.

Спроектированная и реализованная семантическая сеть отражает принципы работы с каждым типом вопросов.

Были предложены пути улучшения работы семантической сети, а также вскрыты недостатки метода обработки данных, мешающие более широкому использованию семантической сети и распознаванию ею данных и знаний.

### СПИСОК ЛИТЕРАТУРЫ

1. Бернерс-Ли Т., Хендлер Дж., Лассила О. [Электронный ресурс] /Семантический Веб. – Режим доступа: [http://ezolin.pisem.net/logic/semantic\\_web\\_rus.htm](http://ezolin.pisem.net/logic/semantic_web_rus.htm) – Дата доступа: 22.03.2013.
2. Тельнов Ю. В. Интеллектуальные информационные системы. М., 2004. С.67–69
3. Амосов Н. М., Касаткин А. М., Касаткина Л. М., Талаев С. А. - Автоматы и разумное поведение. Опыт моделирования. - Киев: "Наукова думка 1973.– 261с.