



АВТОМАТИЧЕСКОЕ ПОСТРОЕНИЕ ЭЛЕМЕНТОВ СЕМАНТИЧЕСКОЙ СЕТИ НА БАЗЕ ТОЛКОВОГО СЛОВАРЯ ПРЕДМЕТНОЙ ОБЛАСТИ

Петрасова С. В.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707-63-60,
e-mail:svetapetrasova@gmail.com*

В настоящее время все более актуальным становится создание и использование интеллектуальных информационных систем, основой которых являются знания. Знания – это связи и закономерности предметной области, выявленные в результате практической деятельности и профессионального опыта, позволяющего специалистам ставить и решать задачи в данной области.

Центральным вопросом построения систем, основанных на знаниях, является выбор модели представления знаний. Представление знаний подразумевает способ формального выражения знаний о предметной области в компьютерно-интерпретируемой форме [1].

Каждая из моделей представления знаний позволяет получить интеллектуальную систему с некоторыми преимуществами, делая ее более эффективной в конкретных условиях, облегчая ее понимание и требуемые модификации.

Для оценки моделей представления знаний использовались следующие критерии: уровень сложности элемента знаний; универсальность представления знаний; естественность и наглядность представления знаний при использовании; способность модели к обучению и формированию новых, непротиворечивых знаний; удобство разработки системы на основе модели [2].

Сетевые модели формально можно задать в виде кортежа $H = (I, C_1, C_2, \dots, C_n, R)$, где I – множество информационных единиц; C_1, C_2, \dots, C_n – множество типов связей между информационными единицами; R – отображение, которое задает связи из заданного набора типов связей между информационными единицами, входящими в I . Сетевая модель, допускающая связи различного типа, называется семантической сетью.

Таким образом, семантическая сеть представляет собой информационную модель предметной области, имеющую вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги задают отношения между ними. Объектами могут быть понятия, события, свойства, процессы.

Семантические сети отражают семантику предметной области в виде понятий и отношений. Их ценность заключается в возможности учитывать в базе знаний, кроме форм утверждений, их семантику. Такие знания позволяют прийти до приемлемого логического вывода при работе с неоднозначными фактами [4].



В исследовании для построения семантической сети данной предметной области взят за основу толковый словарь, который, являясь лингвистическим ресурсом, в явном виде представляет смысловую нагрузку компьютерной лексики. В качестве исследуемого толкового словаря выбран толковый словарь компьютерных терминов Microsoft (Microsoft Computer Dictionary) [5].

Для реализации метода построения элементов семантической сети предлагается использовать следующий алгоритм.

Каждая структурная единица толкового словаря записывается в базу данных, представленную полем термина и полем его определения. На следующем этапе выполняется лингвистическая обработка исходного текста определений из базы данных, которая заключается в фильтрации и нормализации. Вручную анализируются типы отношений в толковом словаре и классифицируются по определенному признаку, затем формализуются типы отношений и для каждого типа формируется база данных.

Для определения типа отношений синонимии применяется метод расстояний. Вычисляется степень семантической близости в процентах, величина которой задается отношением количества общих слов к количеству всех слов в определениях[3]. Формируется база данных с типом отношения синонимии.

Затем проводится редактирование результатов с целью удаления некорректных записей из базы данных. После ввода пользователем запроса в виде термина словаря выполняется его поиск в базу данных, термины идентифицируются на наличие в базе данных типов отношений.

Реализация всех шагов позволяет создать элемент семантической сети, который будет представлен графом, состоящим из концептов и связей между ними.

В работе рассмотрены перспективные направления использования семантических сетей, в частности, в качестве отдельного этапа проектирования онтологий различного назначения и направления дальнейшего научного исследования.

Список литературы

1. Маннинг Кристофер Д., Рагхаван Прабхакар, Шютце Хайнрих Введение в информационный поиск: Пер. с англ. – М.: ООО "И.Д. Вильямс", 2011. – 528с.
2. Потапова Р.К. Новые информационные технологии и лингвистика: Учебное пособие. Изд. 2-е. – М.:Едиториал, 2004. – 320 с.
3. Федорченко Л.А., Хайрова Н.Ф., Довнар А.И., Булгаков С.О. Метод автоматизированного построения семантической сети терминов учебной дисциплины / Радиоэлектрон. и компьют. системы. – 2011. – N4. – С. 115-120.
4. Филиппович Ю.Н., Прохоров А.В. Семантика информационных технологий: опыты словарно-тезаурусного описания. / Серия «Компьютерная лингвистика». М.: МГУП, 2002. – 368 с.
5. Microsoft Computer Dictionary – Microsoft Press, 5-th Ed., 2011, p. 656.