



ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ ИДЕНТИФИКАЦИИ ФАКТОГРАФИЧЕСКОЙ ИНФОРМАЦИИ

Дорошенко А.Ю.

*Национальный технический университет НТУ «ХПИ»
г. Харьков, ул. Пушкинская, 79/2, тел. 707-63-60,
e-mail: doroshenkoanastasiia@gmail.com*

В работе обсуждается технология фактографического поиска, предлагается подход, основанный на представлении содержания текста в форме семантической сети, позволяющий искать факт в семантической сети определенного текста. Рассмотрена задача экстракции и идентификации знаний фактографической информации об ЭВМ.

Цель работы – построение семантической сети фактографической информации об ЭВМ на основе анализа и обработки текстов, а также решение задачи с помощью алгебры предикатов.

В общепринятом смысле под семантической сетью понимается модель представления знаний посредством сети узлов, связанных дугами, где узлы соответствуют понятиям или объектам, а дуги – отношениям между узлами [5].

Научной основой построения семантических сетей является теория графов. Семантические сети представляют знания в виде графовой структуры, которая является более наглядной и естественной по сравнению с другими структурами знаний. Решаемая задача использует структуру, моделирующую семантические связи, которые мы используем для получения одних фактов на основе других [5]. Построение графа помогает находить противоречия в знаниях, а также выявлять недостающие фрагменты знаний. Среди особенностей семантических сетей можем выделить:

- описание объектов ПрО (полной семантической сети) осуществляется средствами естественного языка;
- все факты, включая и вновь поступившие, накапливаются в относительно однородной структуре памяти;
- на сетях определяют ряд унифицированных семантических отношений между объектами и соответственно унифицированные методы вывода;
- структурное представление семантических знаний позволяет определить на них дополнительную семантику, определяющую относительную силу семантических связей, облегчающую процесс вывода в сетях.

К фактографической информации мы относим информацию о фактах [3]. Фактографическую информацию обычно сознательно трактуют просто как конкретные сведения или данные независимо от того, являются ли они фактическими или прогнозируемыми. Главное, что эти сведения сообщают о какой-то предметной области, а не о документах, посвященных этой области. Исходя из такого понимания, фактографическую информацию можно классифицировать следующим образом:

- 1) фактическая и прогнозная (гипотетическая) информация;



- 2) количественная и качественная фактографическая информация;
- 3) хорошо структурированная фактографическая информация и плохо (слабо) структурированная фактографическая информация.

К хорошо структурированным сведениям об ЭВМ относятся, прежде всего, сведения количественного характера, а также качественные (словесно выраженные) сведения, имеющие хорошо регламентированную форму: параметры оборудования и их значения и т. п. К плохо структурированным относятся сведения, представленные разнообразными нерегламентированными словесными инструкциями, т. е. различные описания отдельных фактов, изложение концепций и теорий, сделанных на естественном языке [2,3].

Заключение. Знание не такое определенное понятие, как факт. Оно лишь ограничивает множество возможных состояний мест предметного пространства. Поиск факта есть поиск в семантической сети текста такой подсети, которая изоморфна одному из шаблонов. Если подсеть найдена, факт считается установленным, после чего производится извлечение сущностей и их маркировка ролями, заданными в соответствующих узлах лингвистических описаний [1,4]. Таким образом, результатом поиска является имя факта и набор указателей на сущности семантической сети с указанием соответствующих им ролей в лингвистическом описании.

Список литературы

1. *Алисейко З. А.* Использование алгебры предикатов и предикатных операций для формализации декларативной и процедурной составляющих знаний / З. А. Алисейко, В. И. Булкин, О. В. Канищева, Н. В. Шаронова // Біоніка інтелекту. – Харків : ХНУРЕ, 2006. – № 1(64). – С. 59-63.
2. *Амамия М.* Архитектура ЭВМ и искусственный интеллект / М. Амамия, Ю. Танака. – М. : Мир, 1993. – 400 с.
3. *Бондаренко М. Ф.* О мозгоподобных ЭВМ / М. Ф. Бондаренко, З.В. Дударь, И.А. Ефимова, В.А. Лещинский, С.Ю. Шабанов-Кушнарченко // Радиоелектроника и информатика. – Харьков : ХНУРЭ, 2004. – № 2. – С. 89-105.
4. *Булкин В.И.* Математические модели знаний и их реализация с помощью алгебропредикатных структур / В. И. Булкин, Н.В. Шаронова: монография. – НТУ «ХПИ», МЭГИ. : Донецк, 2010. – 304 с.
5. *Sussna M.* Word sense disambiguation for free-text indexing using a massive semantic network // The Second International Conference on Information and Knowledge Management. – 1993. – P. 67-74.