

Маркелов Константин Сергеевич, Нейман Андрей Борисович

ТЕЗАУРУС ПРЕДМЕТНЫХ ОБЛАСТЕЙ

Адрес статьи: www.gramota.net/materials/1/2012/4/47.html

Статья опубликована в авторской редакции и отражает точку зрения автора(ов) по данному вопросу.

Источник

Альманах современной науки и образования

Тамбов: Грамота, 2012. № 4 (59). С. 154-155. ISSN 1993-5552.

Адрес журнала: www.gramota.net/editions/1.html

Содержание данного номера журнала: www.gramota.net/materials/1/2012/4/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: almanac@gramota.net

УДК 681.3

Технические науки

*Константин Сергеевич Маркелов, Андрей Борисович Нейман**Московский государственный технический университет радиотехники, электроники и автоматики (МГТУ МИРЭА)*ТЕЗАУРУС ПРЕДМЕТНЫХ ОБЛАСТЕЙ[©]

Накопленные к настоящему времени колоссальные объёмы информации, в совокупности с непрерывно увеличивающимися темпами её роста определяют актуальность и значимость исследований в области построения некоторого базисного описания предметных областей. Таким описанием может выступать созданный тезаурус той или иной предметной области. Бурное развитие сетевых технологий, в том числе и Интернета, способствуют значительному увеличению доступных информационных ресурсов и объёмов передаваемой информации, объёмов информации по предметным областям. Зачастую много разнородной, слабо структурированной, избыточной информации и информации, которая может быть сложна в понимании. И для того, чтобы любому человеку хорошо изучить предметную область, нужно иметь тезаурус. В любой момент времени можно обратиться к нему, изучить необходимый материал и в дальнейшем получить полное представление о предметной области.

Что такое тезаурус? Определить, что же такое тезаурус, достаточно сложно, но все же можно выделить несколько определений, которые представлены в различных Интернет словарях, а также предоставить различные подходы к этому понятию.

Тезаурус в современной лингвистике - особая разновидность словарей общей или специальной лексики, в которых указаны семантические отношения (синонимы, антонимы, паронимы, гипонимы, гиперонимы и т.п.) между лексическими единицами. Таким образом, тезаурусы, особенно в электронном формате, являются одним из действенных инструментов для описания отдельных предметных областей.

В отличие от толкового словаря, тезаурус позволяет выявить смысл не только с помощью определения, но и посредством соотнесения слова с другими понятиями и их группами, благодаря чему может использоваться в системах искусственного интеллекта.

В теории информации термин тезаурус употребляется для обозначения совокупности всех сведений, которыми обладает субъект [2].

Другое определение и трактовка тезауруса. **Тезаурус** - множество смысловыражающих единиц некоторого языка с заданной на нём системой семантических отношений. Тезаурус фактически определяет семантику языка (национального языка, языка конкретной науки или формализованного языка для автоматизированной системы управления) [1]. Здесь уже четко прослеживается назначение тезауруса, как некоторого базисного словаря для какой-либо предметной области в применении в программных средах, интеллектуальных системах, автоматизированных системах и в системах обучения.

В широком смысле тезаурус интерпретируют как описание системы знаний о действительности, которыми располагает индивидуальный носитель информации или группа носителей. Этот носитель может выполнять функции приёмника дополнительной информации, вследствие чего изменяется и его тезаурус. Исходный тезаурус определяет при этом возможности приёмника при получении им семантической информации. В психологии и при изучении систем с искусственным интеллектом рассматривают свойства тезауруса индивидуумов, проявляющиеся при восприятии и понимании информации. В социологии и теории коммуникаций изучают свойства тезауруса индивидуумов и коллективов, обеспечивающие возможность взаимопонимания на основе общности тезаурусов. В этих ситуациях в тезаурус приходится включать сложные высказывания и их семантические связи, определяющие запас сведений, которыми располагает сложная система. Тезаурус фактически содержит не только информацию о действительности, но и метаинформацию (сведения об информации), обеспечивающую возможность приёма новых сообщений [Там же].

Представим ещё некоторые из определений понятия тезауруса из различных словарей.

Под **тезаурусом** понимается сложный компонент словарного типа, в котором все значения словаря связаны между собой семантическими отношениями, отражающими основные соотношения понятий в описываемой предметной области знаний.

В разных системах тезаурус может выполнять разные функции:

- источник специальных знаний в узкой или широкой предметной области, способ описания и упорядочения терминологии предметной области;
- инструмент поиска в информационно-поисковых системах;
- инструмент ручного индексирования документов в информационно-поисковых системах (так называемый контролирующий словарь);
- инструмент автоматического индексирования текстов [6].

Классическое толкование определения тезаурус в научной литературе состоит в определении его как словаря особого типа, отражающего словарный состав языка в полном объёме. Структура хорошо разработанного тезауруса подробно описывается в работах Ю. Н. Караулова [3; 4].

Понятие тезауруса может толковаться расширительно, если в качестве языковых единиц, представляемых в нём рассматривать предложения, сверхфразовые единства и тексты. В этом случае тезаурус будет являться своеобразным гипертекстом, т.е. естественно-языковым описанием.

В качестве элементов тезауруса можно рассматривать отдельные суждения, а в качестве семантических отношений выбрать логические операции, тогда тезаурус будет представлять собой некий свод возможных логических рассуждений, некоторую логическую модель определенной мыслительной деятельности.

Понимание тезауруса как максимально полного собрания слов языка расширяет его лексический состав путем включения в него не только слов литературного языка данной эпохи, но и архаизмов, неологизмов, слов высокого слога, просторечные слова, иноязычную лексику, а также областные слова. Однако такое определение стирает различия между тезаурусом и другими словарями. Более того, в этом смысле к тезаурусам относятся все типы идеографических словарей (тематические, аналогические, собственно идеографические).

Тезаурус является лексическим инструментом информационно-поисковых систем. Он состоит из контролируемого, но изменяемого словаря терминов, между которыми указаны смысловые связи. Такой словарь исчерпывающим образом покрывает некоторую специфическую область знаний, представляет собой перечень дескрипторов и недескрипторов (вспомогательных слов), который упорядочен по систематическому и алфавитному принципу и отражает смысловые отношения как иерархического типа (например, «родовидовые»), так и неиерархического типа. Если в общепринятом определении специальных тезаурусов как общетехнических и информационно-поисковых классификаторов и рубрикаторов снять терминологические и отраслевые ограничения, то принципиальной разницы между общеязыковым и информационно-поисковым тезаурусом нет [3]. Зафиксированный набор языковых единиц (слов и словосочетаний) информационно-поискового тезауруса и правила их соотношения друг с другом (исчисления отношений), в особенности, если это касается терминологической лексики, может быть приравнен к некоторому информационно-поисковому языку, формально-языковой системе [5].

В итоге тезаурус является некоторым основным компонентом для обработки и анализа текстов, легко изучить любую предметную область, если под рукой есть вся необходимая информация и все определения и понятия не только присутствуют в словаре тезауруса, но и существует взаимосвязь между ними как по иерархии, так и по тексту. По иерархии означает то, что определения включаются в себя уже ранее раскрытые понятия, то есть не может быть такого, что в одном из определений присутствует что-то непонятное тому или иному человеку, изучающему предметную область. Человек всегда может обратиться ко всем понятиям раскрывающие предыдущие понятия. По тексту означает то, что определения в тезаурусе следуют в том порядке, в котором впервые употребляются в тексте.

Использование тезауруса позволяет упростить изучение любой предметной области. Существуют различные методы построения тезаурусов предметных областей. К таким методам обычно относят статистические, в которых анализируется весь текст и строится частота встречаемости всех определений и слов в тексте. Таким образом, тезаурус становится возможным построить и автоматизировано, а не полностью в ручную, как, например, если составлять обычный словарь терминов.

Список литературы

1. **Большая советская энциклопедия** [Электронный ресурс]. URL: <http://slovari.yandex.ru/~книги/БСЭ/Тезаурус/> (дата обращения: 29.11.2010).
2. **Википедия** - **свободная энциклопедия** [Электронный ресурс]. URL: <http://ru.wikipedia.org/wiki/%D2%E5%E7%E0%F3%F0%F3%F1> (дата обращения: 29.11.2010).
3. **Караулов Ю. Н.** Лингвистическое конструирование и тезаурус литературного языка. М.: Наука, 1981. 366 с.
4. **Караулов Ю. Н.** Частотный словарь семантических множителей русского языка. М.: Наука, 1980. 207 с.
5. **Филиппович Ю. Н., Прохоров А. В.** Семантика информационных технологий: опыты словарно-тезаурусного описания. М.: Изд-во МГУП, 2002. 368 с.
6. **Шрейдер Ю. А.** Информация в структурах с отношениями // Исследования по математической лингвистике, математической логике и информационным языкам. М.: Наука, 1972. С. 147-159.

УДК 338.482

Экономические науки

Наталья Степановна Мартышенко

Владивостокский государственный университет экономики и сервиса

МЕТОДИЧЕСКИЕ ПОДХОДЫ К ОБЕСПЕЧЕНИЮ КАЧЕСТВА ТУРИСТСКИХ УСЛУГ, ПРОИЗВОДИМЫХ В РЕГИОНАЛЬНОМ ТУРИСТСКОМ КОМПЛЕКСЕ[©]

Рекреация и туризм являются одними из базовых потребностей человека. Технический прогресс только повышает значимость этих потребностей. Это подтверждается неуклонным ростом объемов реализуемых туристских и рекреационных услуг в мире. Темпы роста туризма в России не отвечают потребностям