

**Амрина Асемгуль Козубаевна**

Магистрант

**Направление:** Информатика и вычислительная техника

**Магистерская программа:** Распределенные автоматизированные системы

### **Архитектура информационно-поисковой системы на основе семантической сети**

**Аннотация.** Эффективность любой информационно-поисковой системы определяется ее способностью служить тем целям, для которых она была разработана.

**Ключевые слова:** информационно-поисковая система, пертинентность.

Основная цель информационно-поисковых систем (ИПС) – отображать соответствующую информацию по введенному запросу пользователя. Пользователь хотел бы видеть точную и актуальную информацию на одной странице, а не осуществлять ее поиск по списку предоставленных ссылок.

В настоящее время для достижения указанной цели разработчики ИПС совершенствуют алгоритмы поиска, пользовательский интерфейс, который позволяет формировать сложные запросы, создают дополнительные функции и сервисы [1].

В данной статье автор предлагает архитектуру ИПС на основе семантических сетей.

Семантическая сеть или смысловая сеть – это модель предметной области представления знаний, представленная в виде графа, позволяет описывать объекты, явления и понятия предметной области с помощью сетевых структур, основанных на теории графов.

На рис. 1 представлена архитектура ИПС на основе семантической сети.

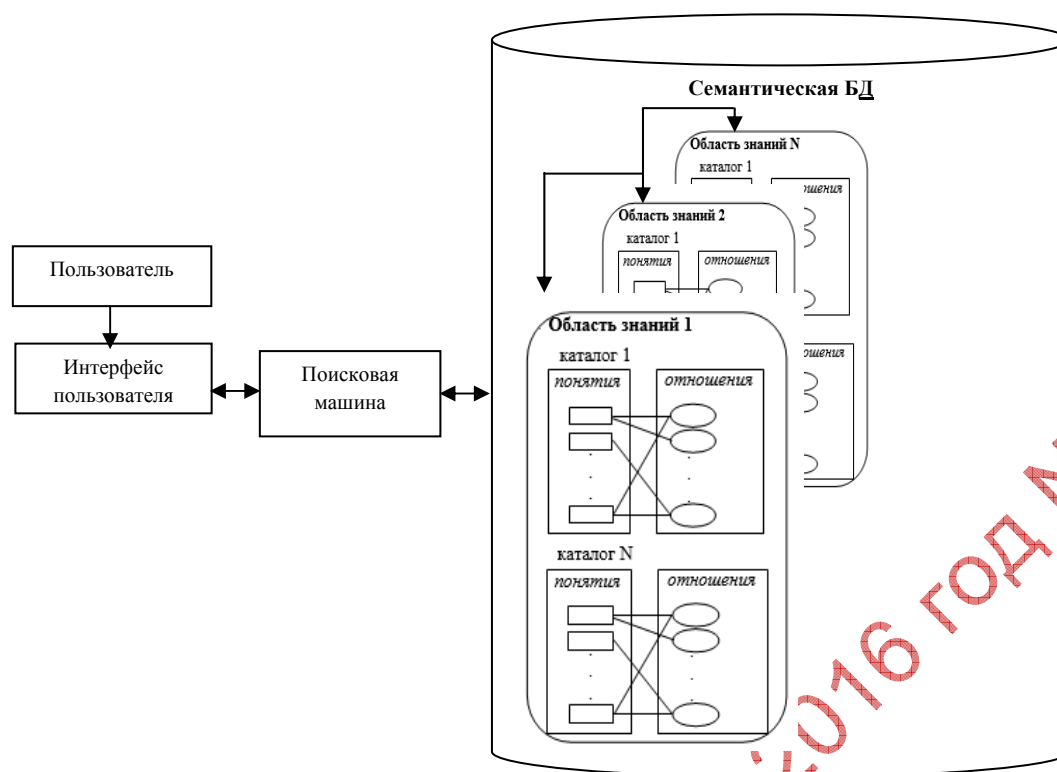


Рис. 1. Архитектура ИПС на основе семантических сетей

Интерфейс ИПС должен иметь не просто обычное поле поиска, а включать функцию кластеризации и расширенного поиска, таким образом, учитывать не только ввод простых запросов, но и сложных, ведь поисковой системой должны пользоваться не только профессионалы, но и обычные пользователи.

Поисковая машина осуществляет поиск информации в информационной среде. Критериями качества поиска являются релевантность, учет морфологии и семантики языка, полноты индекса.

Поиск осуществляется в базе данных, в основе которой лежит семантическая модель. Семантическая база данных (БД) состоит из совокупностей всех областей знаний. Каждая область знаний представлена в виде дерева. В поисковой системе знания удобнее всего, по мнению автора работы, представлять в виде наборов концептуальных графов, которые состоят из понятий и отношений. Причем, не существуют ребра, соединяющего 2 вершины из одной и той же части.

Существуют различные виды отношений между понятиями [2]:

1. Таксономические (класс-подкласс-экземпляр, множество-подмножество-элемент);

2. Структурные (часть-целое), агрегация, композиция, мероним;

3. Родовые (потомок-предок);

4. Производственные;

5. Функциональные (например, производит, влияет);

6. Количественные;

7. Временные;

8. Пространственные;

9. Атрибутивные;

10. Логические;

11. Казуальные (причинно-следственные).

Каждое понятие имеет определенный список документов (упорядоченный информационный массив), внутри которого по запросу пользователя осуществляется поиск.

При поиске может оказаться, что искомое понятие относится к другой области знаний, поэтому все существующие области знаний связаны между собой. Каждая область знаний может состоять из множества подобластей, а те из разделов областей знаний (на схеме они же называются каталогами).

Если при поиске информации оказывается, что искомого понятия с соответствующим отношением отсутствует, то на основе существующих графов можно создавать новые концептуальные графы [3] путем таких операций как копирование, ограничение, объединение, упрощения. Таким образом, можно расширять семантическую базу данных.

© На рис. 2 представлен алгоритм поиска информации в области знаний семантической базы данных.



**Рис. 2.** Алгоритм поиска информации

Конечное число уровней может быть разным – это зависит от конкретной области знаний.

Элементы уровней выше пятого являются обычными множествами. Компоненты пятого и последующих уровней являются нечеткими множествами.

Поскольку семантическая сеть отображает 1 предметную область, то ответ на запрос пользователя должен отображать на странице поисковой системы, а не выводить список ссылок на другие сайты. Но все же необходимо заметить, что качество ответа напрямую зависит от качества введенного запроса.

Организация поиска на основе данной архитектуры ИПС может позволить существенно повысить пертинентность выдаваемых документов.

Также можно отметить, что архитектура поисковых систем должна учитывать социальный аспект современности – это переход на мобильные носители (в настоящее время такую тенденцию учитывает поисковая система Google). Все меньше пользователей хотят быть привязаны к конкретному рабочему месту или дому для того, чтобы выйти в сеть и воспользоваться различными сервисами. Пользователи желают быть в курсе мировых новостей,

находясь в поездках, по дороге на работу, в любом месте, где есть мобильная связь и зона Wi-Fi.

Для привлечения новых и удержания старых пользователей разработчики должны совершенствовать алгоритм ранжирования, добавлять новые сервисы и функции, при этом не забывать про интуитивно понятную навигацию системы. При этом контент поискового сайта должен быть уникален и не иметь повторов.

### Литература

1. Андрейчиков А.В., Андрейчикова О.Н. Интеллектуальные информационные системы: учебник. М.: Финансы и статистика, 2004.
2. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем СПб.: Питер, 2000.
3. Лорьер Ж.-Л. Системы искусственного интеллекта. М.: Мир, 1991.

© Бюллетень магистранта № 3