

Александров Дмитрий Сергеевич

Магистрант НАЧОУ ВПО СГА

Направление: Информатика и вычислительная техника

Магистерская программа: Распределённые автоматизированные
системы

**Создание базы данных для интеллектуальной информационной
системы**

Аннотация. Для обеспечения информатизации учебного процесса возникает задача разработки структуры представления знаний, алгоритмы для решения возникающих противоречий, алгоритмы поиска в семантической сети. Применение поисковых запросов обеспечивает возможность динамической актуализации семантической сети по мере увеличения количества публикаций, депонируемых в базе данных. Показано, что вычисляемая системой оценка смыслового сходства документов может быть использована для автоматизированного выявления взаимосвязей и конструировать семантические сети.

Ключевые слова: семантическая сеть, инфологическая модель, сущность, алгоритм.

Вступление человечества в XXI век требует перехода к новой стратегии развития общества на основе знаний и высокоэффективных технологий. Соответственно обеспечение эффективности системы образования – одна из важнейших задач. По признанию специалистов в XXI веке в образовании станут решающими именно информационные технологии. Важнейшей составляющей процесса обучения является методически правильное и обоснованное применение современных информационных технологий в учебном процессе. Главная способность компьютера заключается, несомненно,

в его пригодности для автоматизированного обучения, интеллектуального управления процессом обучения [2].

С позиций теории управления учебный процесс рассматривается как сложная система, где в роли объекта управления выступает учащийся со своими индивидуальными особенностями. В настоящее время стратегической линией исследования, разработки и внедрения образовательных информационных технологий является их интеллектуализация на основе моделей знаний учебного материала и моделей учебного процесса. Однако общепризнанных моделей и конструктивной теории решения этой проблемы в настоящее время нет, что и определяет актуальность проводимой работы [1].

Модель БД “ISCMK”

Анализ определенных объектов и атрибутов [1-3, 5-9] позволяет выделить сущности проектируемой БД и построить ее инфологическую модель (рис.1).

К стержневым сущностям можно отнести:

- Людей.

Эта сущность отводится для хранения данных о пользователях системы («Автор», «Эксперт», «Модератор», «Администратор»). Такое объединение допустимо, так как данные о разных людях выбираются из одного домена (фамилия, имя и пр.), оно исключает дублирование данных (один и тот же человек может быть и администратором системы и экспертом).

Так как фамилия и имена пользователей могут быть довольно громоздкими и будут многократно встречаться в разных таблицах БД, то их целесообразно нумеровать и ссылаться на эти номера. Для этого вводится суррогатный ключ – целочисленный атрибут «ИД», который будет автоматически наращиваться на единицу при вводе в БД нового человека.

Аналогично создаются суррогатные ключи для большинства сущностей модели.

- Дисциплины.

Данная сущность предназначена для хранения данных о дисциплинах. На данном этапе сущность «Дисциплины» заблокирована для редактирования, но при дальнейшем развитии системы данные ограничения сниматься.

- Разделы.

Сущность предназначена для хранения разделов и подразделов.

- Объекты содержания.

Данная сущность хранит информацию об объектах содержания, их описание, тип объектов и прочие.

- Объекты данных.

Сущность предназначена для хранения данных об объектах данных. Данная сущность связана с объектом содержания и без этой связи она представляла бы собой хаотичный набор данных.

- Словарь шаблонов.

Сущность предназначена для хранения информации о словах шаблонов. Использование данной сущности в системе значительно повышает качество поиска.

- Синонимы.

Данная сущность предназначена для хранения синонимов, что в разы повышает скорость нахождения нужных данных.

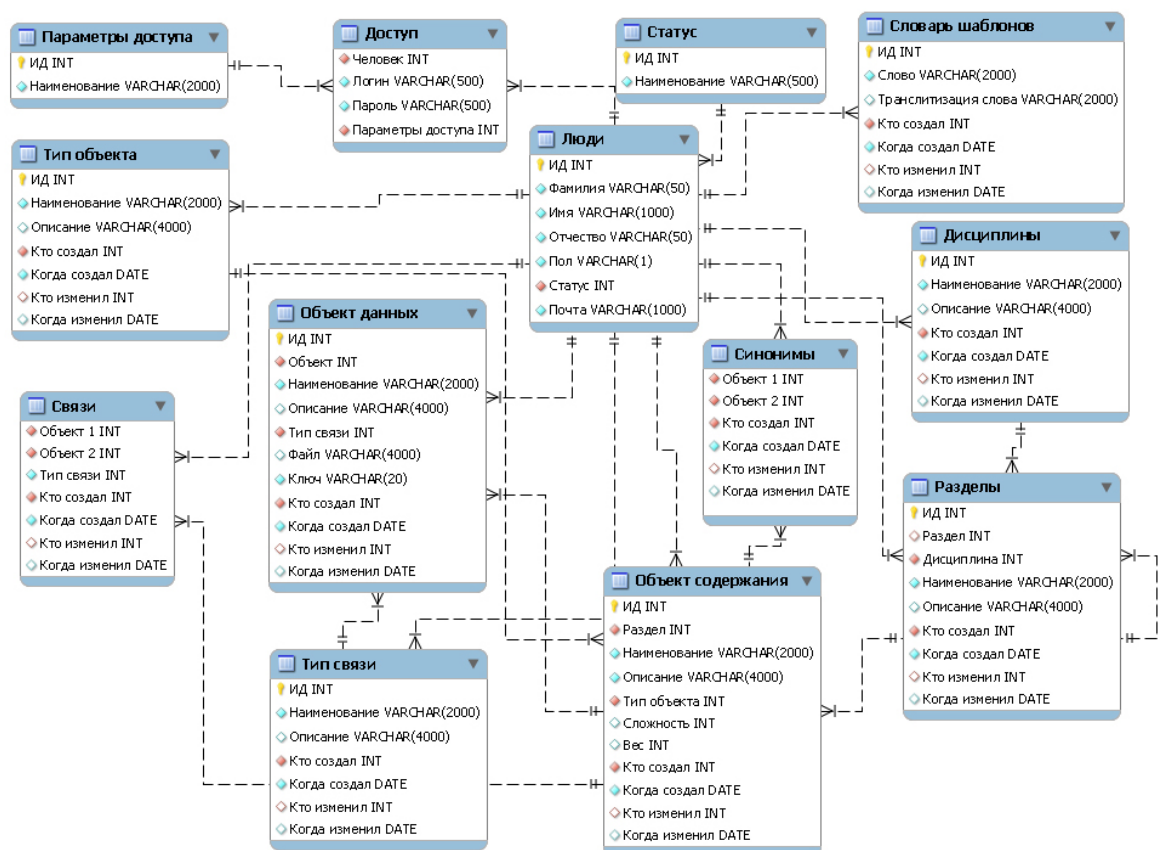


Рис. 1. Инфологическая модель БД «ISCMK»

Организация поиска в сети

За основу поиска был взят метод Левенштейна (расстояние Левенштейна) [4] – минимальное количество операций вставки одного символа, удаление одного символа и замены одного символа на другой, необходимы для превращения одной строки в другую.

На рис. 2 представлен пример матрицы Левенштейна. Взяты два слова «Объект» и «Объектов». Если сравнивать эти два слова напрямую, то система выдаст отрицательный результат. Используя метод Левенштейна, получим расстояние между этими объектами равно 2. Чем меньше расстояние, тем ближе искомый объект к шаблону.

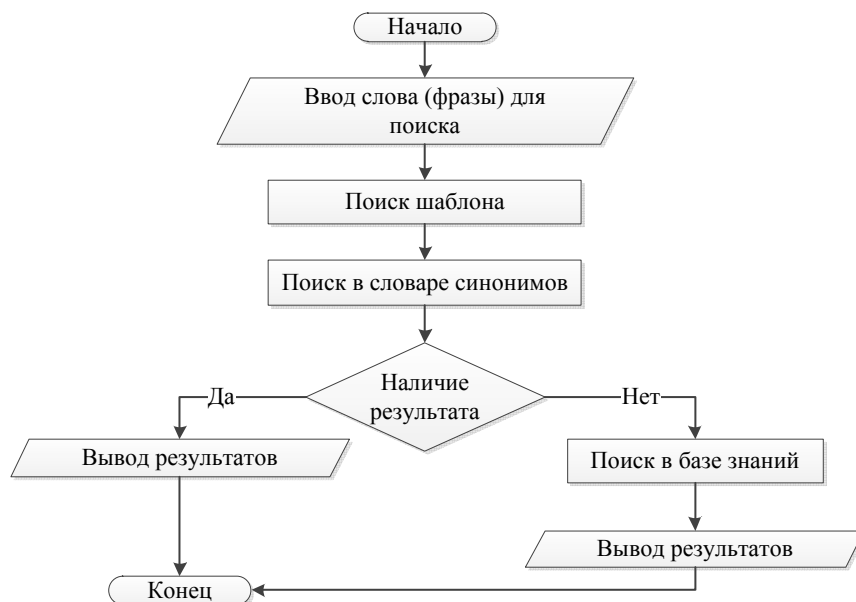


Рис. 2. Блок-схема поиска объектов в сети

На рис. 2 представлен принцип организации объектов в сети. Для быстрого и качественного поиска в системе предусмотрен словарь шаблонов (так как слова могут быть в разных падежах) и словарь синонимов для более быстрого поиска необходимой информации. Принцип поиска заключается в следующем: пользователь вводит слово или предложение, которое он планирует найти, и система проводит поиск шаблона. В случае если введено предложение, система разбивает введенное предложение на слова и находит для каждого слова шаблон. После поиска шаблона система проводит поиск в словаре синонимов, если поиск дает результат то система выводит его пользователю, если результата нет, система проводит поиск в базе знаний и выводит результат, как положительный, так и информацию о том, что на данный запрос данных нет в случае отрицательного результата.

Алгоритм исключения дублирования информации

Разработка алгоритма исключения дублирования информации является одной из главной задачи. При создании нового объекта, пользователь (человек, зарегистрированный в системе) может ввести уже существующую информацию в систему, что повлечет крайне нежелательное дублирование информации,

появятся два объекта схожих по содержанию. Для разрешения этой проблемы был разработан алгоритм исключения дублирования. Работу алгоритма можно разбить на следующие шаги.

На первом шаге происходит поиск по названию объекта, и выявляются наиболее вероятные объекты. Так как объект может быть назван произвольным образом, то не исключена ситуация, когда наименования объектов различны, а содержания схожи.

На втором шаге происходит поиск по ключевым словам.

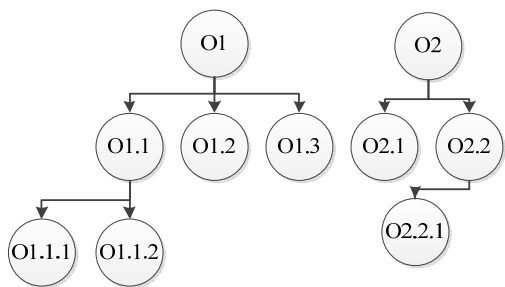


Рис. 3. Иерархия объектов в сети

На рис. 3 представлен наглядный пример иерархии объектов в сети. Характеристикой каждого объекта являются объекты, которые в него входят, то есть характеристикой объекта O1 являются объекты O1.1, O1.2 и O1.3. Аналогично характеристикой объекта O2 являются O2.1 и O2.2. Для объекта O1.1 характеристикой являются объекты O1.1.1 и O1.1.2. Объекты, характеризующие данный объект и являются его ключевыми словами.

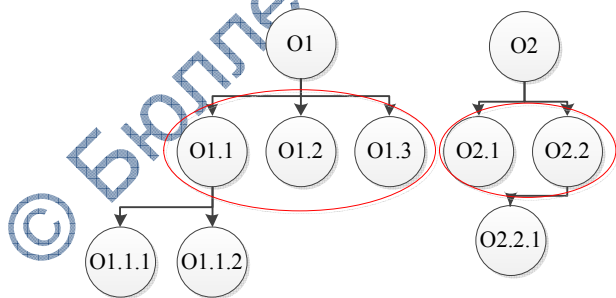


Рис. 4. Ключевые слова объектов O1 и O2

На рис. 4 выделены красным ключевые слова объектов O1 и O2. Эти ключевые слова являются первым шагом в поиски схожего объекта в базе знаний.

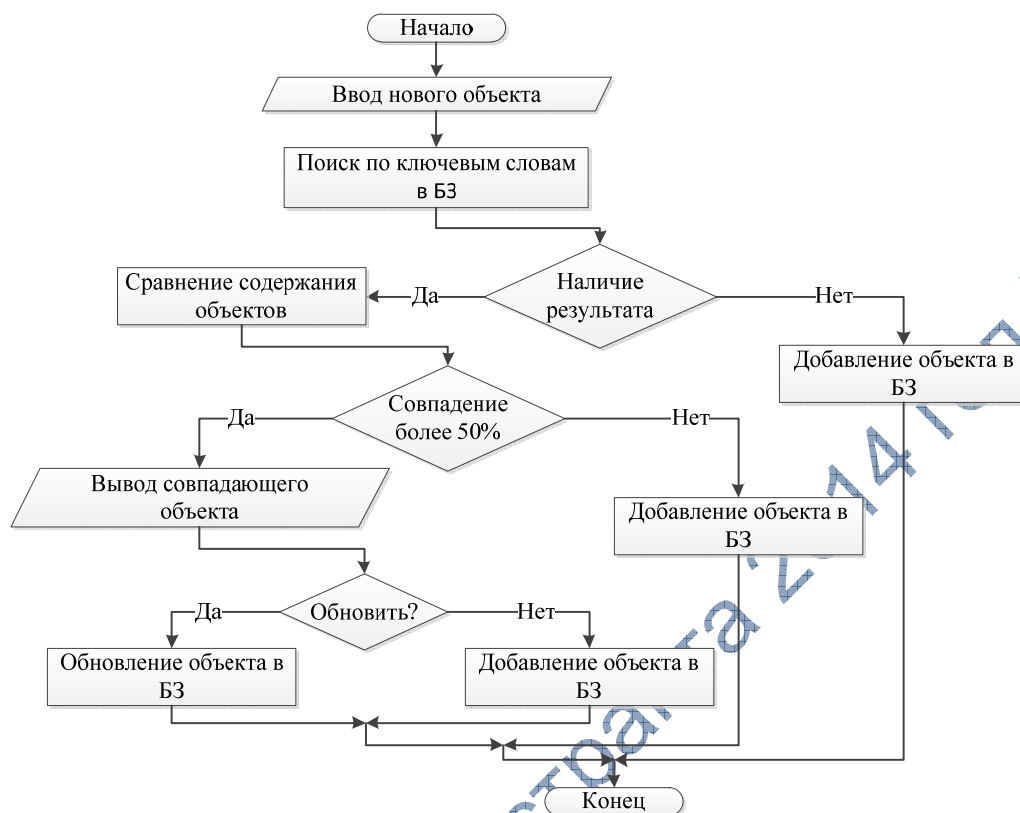


Рис. 5. Блок-схема алгоритма исключения дублирования информации

На рис. 5 приведена блок-схема алгоритма исключения дублирования информации. Вначале система осуществляет поиск по ключевым словам нового объекта в базе знаний. Если поиск по ключевым словам не дал результата, то объект считается новым, и система его добавляет в базу знаний. Если система нашла совпадения по ключевым словам, то на следующем шаге проверяется на совпадение содержимое объектов. В случае, когда содержимое объектов совпадает менее чем на 50%, объект считается новым и заносится в базу знаний, если же совпадение более 50% объект выводится для сравнения пользователю, и пользователь сам решает что необходимо сделать, обновить старый объект или проигнорировать и создать новый.

Заключение

В ходе исследования разработана информационная система для построения модели знаний, алгоритмы исключения дублирования информации и поиска в семантической сети, которые позволяют усовершенствовать систему для дальнейшего использования пользователями.

Литература

1. Башмаков А.М., Башмаков И.А. Механизмы наследования, выявления и разрешения противоречий в обобщенной модели представления предметной области. Часть I // Техническая кибернетика. 1994. №5.
2. Башмаков А.И. Обобщенная модель представления предметной области. М.: МЭИ, 1997.
3. Башмаков А.И., Башмаков И.А. Механизмы наследования, выявления и разрешения противоречий в обобщенной модели представления предметной области. Часть II // Теория и системы управления. 1995. №3.
4. Башмаков А.И., Башмаков И.А. Интеллектуальные информационные технологии. М.: МГТУ им. Н.Э. Баумана, 2005.
5. Гасфилд Д. Строки, деревья и последовательности в алгоритмах. Информатика и вычислительная биология. СПб.: Невский Диалект, БВХ-Петербург, 2003.
6. Кузнецов И.П. Расширенные семантические сети для представления и обработки знаний / Системы и средства информатики: Ежегодник. Вып. 4. М., 1993.
7. Осипов Г.С. Построение моделей предметных областей. Ч. I. Неоднородные семантические сети // Известия РАН. Техническая кибернетика. 1990. №5.
8. Перминов И.А. Нечеткая объектно-ориентированная семантическая сеть // Доклады Международной конференции «Информационные средства и технологии» Международного форума информатизации МФИ-99. Т. 3. М.: Станкин, 1999.

9. Перминов И.А. Объектно-ориентированный язык для оперирования семантическими сетями // Тезисы докладов Международной конференции «Информационные средства и технологии» Международного форума информатизации МФИ-2000. Т. 2. М.: Станкин, 2000.

© Бюллетень магистранта 2014 год № 2