

Разработка нечетких баз данных связанных с риэлтерской деятельностью

Ананьев Егор Андреевич

*Волжский политехнический институт (филиал) Волгоградского государственного технического университета
Студент*

Рыбанов Александр Александрович

*Волжский политехнический институт (филиал) Волгоградского государственного технического университета
к.т.н., доцент, заведующий кафедрой информатики и технология программирования*

Аннотация

В статье рассматриваются проблемы поиска объектов недвижимости, связанных с четкостью атрибутов поиска. Были приведены анализы существующих баз данных, связанных с риэлтерской деятельностью, и систем, работающих с нечеткими множествами. По данным результатам, были построены и сравнены две базы данных, построенных по принципам синтезирования и резюмирования.

Ключевые слова: нечеткие базы данных, принцип синтезирования, принцип резюмирования, риэлтерское агентство, лингвистическая переменная, лингвистическая терма.

Development of fuzzy databases related to real estate activities

Ananiev Egor Andreevich

*Volzhskiy Polytechnic Institute (Branch) Volgograd State Technical University
Student*

Rybanov Aleksander Aleksandrovich

*Volzhskiy Polytechnic Institute (Branch) Volgograd State Technical University
Ph. D., associate Professor, head of Department of computer science and software engineering*

Abstract

The article deals with the problems of searching for real estate objects related to the clarity of search attributes. Analyzes of existing databases related to real estate activity and systems working with fuzzy sets were given. According to the results, two databases were constructed and compared, based on the principles of synthesis and summarization.

Keywords: fuzzy databases, the principle of synthesis, the principle of summary, real estate agency, linguistic variable, linguistic term.

Нечеткая логика сегодня используется во всех автоматизированных системах, на примере нечетких контроллеров - кулеры в компьютерах, чем больше температура, тем больше оборотов в секунду совершает вентилятор и тем больше энергопотребления он требует. Также нечеткая логика применима в специфичной сфере как нейронные системы.

Однако теорию нечетких множеств, возможно, применить и в поисковых системах, переводя нечеткий, неформализованный запрос в четкий. Особенно такое применение будет полезно для риэлторов, работающих с нечеткостью, то есть запросами клиентов. При обращении клиента в риэлтерскую фирму, риэлтор фирмы переводит человеческий запрос клиента в числовые данные, которые переводятся в запрос, который поймет база данных. Благодаря нечетким множествам Заде, критерии поиска клиента, возможно, сразу передать в нечеткую базу данных и извлечь данные, которые будут удовлетворять нечеткому запросу[1].

Постановка проблемы

Проблемы поиска информации связаны с упущением объектов, которые находят риэлторы фирм, используя запросы к реляционной базе данных. Профессиональный риэлтор не упустит подходящие объекты жилья, однако займет это значительное количество времени. Система поиска объектов недвижимости нечетким запросом к базе данных, снижают затраты на поиск объектов, расширяя границы поиска в нужных пропорциях.

Анализ существующих риэлтерских программ, и программ, работающих с нечеткой логикой

Существующие риэлтерские фирмы используют несколько типов программных продуктов, различных по способу использования. Такими являются CRM, системы, предназначенные для улучшенного взаимодействия с клиентами, которые разрабатываются для одной конкретной риэлтерской фирмы и в последующем использовании клиентами и сотрудниками данного бизнес процесса[2]. Доски объявлений в сети интернет, очень просты в показе информации и использовании, являются наилучшим средством для продажи собственниками, однако и риэлторами не имеющих CRM, но имеющих базу данных клиентов, способна успешно производить трудовую деятельность на данных интернет порталах. Однако все данные системы имеют примитивный поиск с указанием дискретных значений в поисковых полях. Например, указываются количество квартир, цена в виде «от и до» и т.п.. Данные системы для работы с объектами недвижимости, дали представление об атрибутах, которые возможно хранить в базе данных риэлтерской фирмы.

Систем использующих нечеткую логику при поиске объектов недвижимости, обнаружено не было, однако были найдены системы,

использующие системы нечеткого вывода. Такими оказались FuzzyTech и инструментарий Fuzzy в программе Matlab. Благодаря данным системам был сформирован модуль настройки лингвистических переменных.

База данных основанная на четкой логике

При экспериментальной оценке баз данных, классифицирующийся по различным принципам хранения данных, которыми являются принципы резюмирования и синтезирования, было выявлено, что отличием их, было в том, что первая позволяла хранить вычисляемые данные (рис 1), а вторая нет (рис 2).

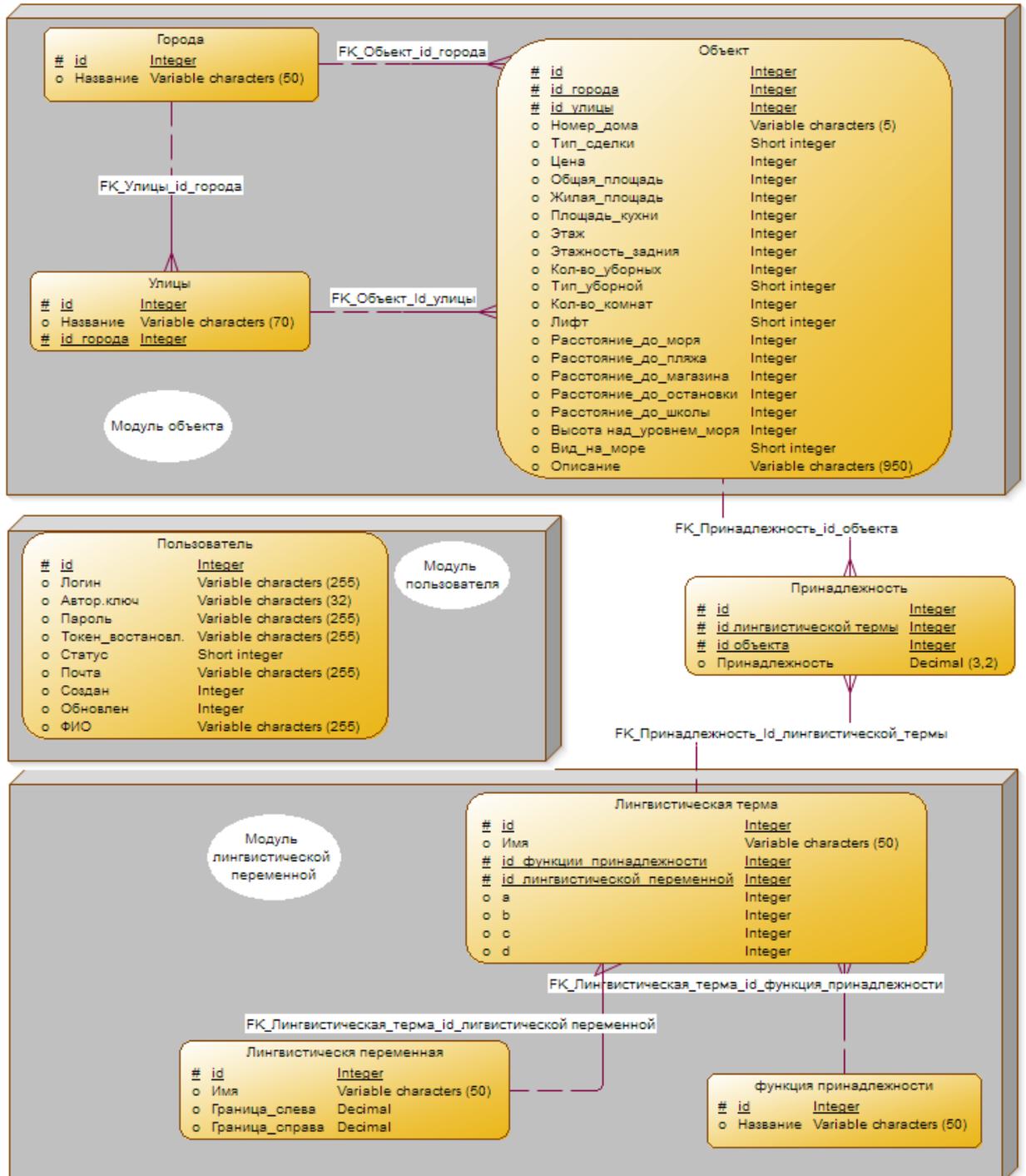


Рис.1. Схема базы данных, построенная по принципу резюмирования

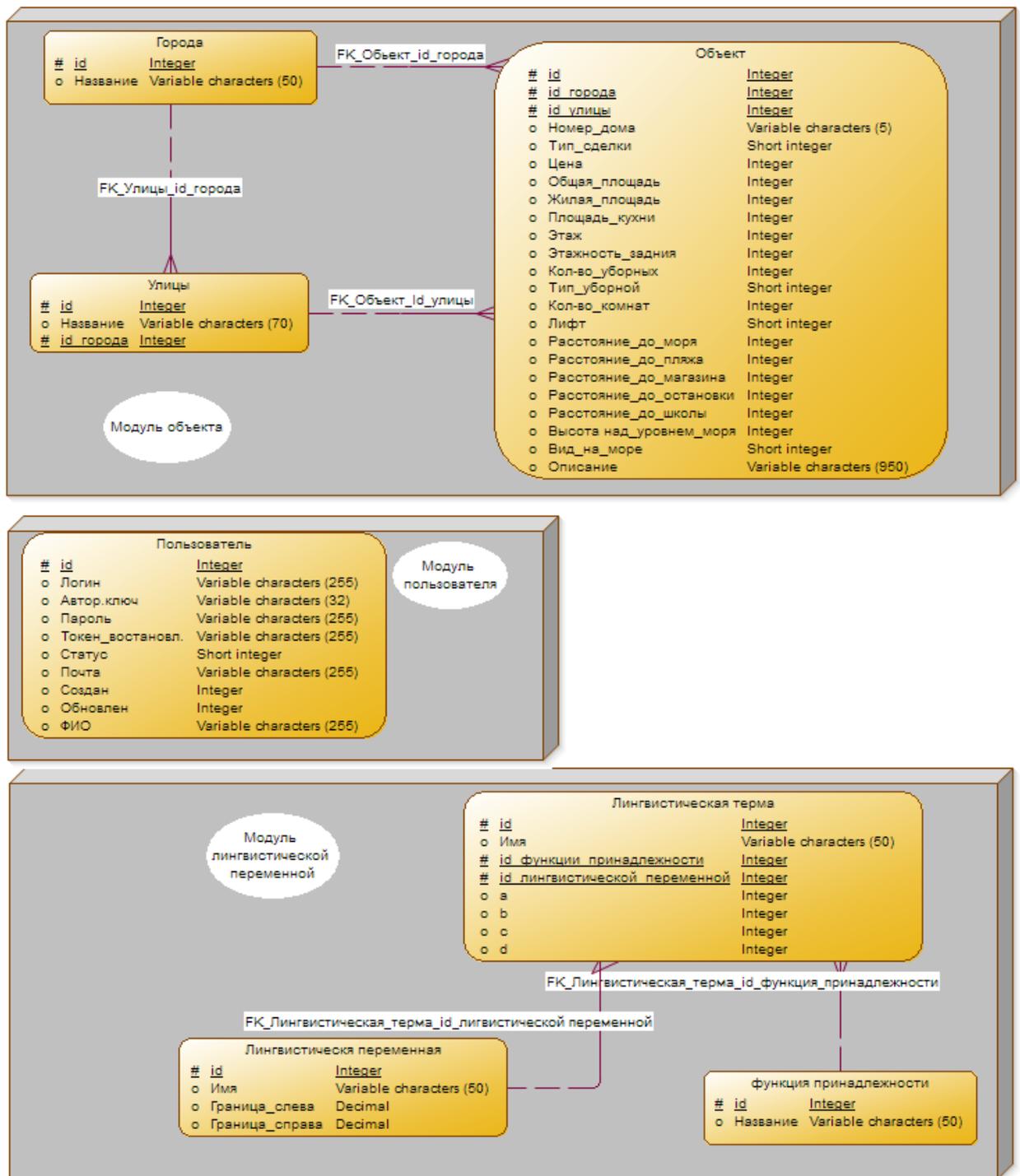


Рис.2. Схема базы данных, построенной по принципу синтезирования

Конечные вычисления в базе, построенной по принципу синтезирования, не хранятся, а вычисляются по необходимости. При осуществлении нечеткого запроса к базе данных, выполняется выборка из таблицы объектов и лингвистических терм, по найденным данным создается двумерный массив, значениями столбцов которых будет ссылка на объект, ссылка на терм, и принадлежность. Вычисляется принадлежность с помощью функций принадлежности для соответствующего атрибута выбранного объекта. Таким образом, количество строк в массиве принадлежностей,

является декартово произведение таблиц объектов и лингвистических терм. Далее поиск идет по массиву с помощью написанного дополнительного кода, который обращается к массиву, фактически, сколько лингвистических терм участвует в нечетком запросе, столько обращений и будет к массиву принадлежностей.

По принципу резюмирования, база данных уже содержит таблицу принадлежностей, заполняется она также как и выше массив принадлежностей. При добавлении или изменения лингвистических терм и объектов недвижимости, соответствующие вычисления заносятся в таблицу принадлежностей. При выполнении нечеткого запроса к базе данных, количество выборок в таблице принадлежностей, зависит от количества лингвистических терм, участвующих в запросе, как, впрочем, и в предыдущем варианте, разница лишь в обращении к массиву.

Разница между базами данных представлены в таблице 1.

Таблица 1. Разница между базами данных построенных по принципу резюмирования и синтезирования.

Синтезирования	Резюмирования
База хранит лишь минимальные данные для работы системы. Что означает меньшее количество таблиц.	База хранит все данные, необходимые для работы, а также вычисляемые данные. Число таблиц возрастает.
Вычисление принадлежности происходит лишь при соответствующем запросе и хранит их в оперативной памяти.	Вычисление принадлежностей происходит при добавлении, изменении различных данных, связанных с лингвистическими терминами или объектами. Принадлежность заносится в отдельную таблицу.
Данные извлекаются из нужных таблиц, вычисляются производные данные и далее идет работа с массивом.	Данные берутся из соответствующих таблиц, и затем используются.
При изменение ячеек, не производится дополнительных действий.	При изменении ячеек, происходит перерасчет производных данных.

Пример. За месяц в риэлтерскую фирму поступает 15 новых объектов с имеющимися 200 объектами, обращается 20 клиентов (примерно по 2,5 поиска на каждого клиента с корректировкой, получим 50 нечетких запросов). Имеется 39 терм, из которых изменяют в конце месяца 5 терм (связанных с ценой), Условно – вычисление одной принадлежности, обращение или заполнение одной строки 1 строки происходит в 1 «вычисление». Теоретическое сравнение баз данных построенных по

принципу синтезирования и резюмирования продемонстрировано в таблице №2.

Таблица 2. Теоретическое сравнение баз данных построенных по принципу синтезирования и резюмирования.

Операция	Синтезирование	Резюмирование
Добавление 15 новых объектов	Заполнение 15 строк в таблице объектов = 15 «вычислений».	Кол-во вычислений принадлежностей и добавлений строк в таблицу принадлежностей является декартовым произведение терм и новых объектов = $15 \cdot 39 \cdot 2 = 1170$ «вычислений» При этом обычный 1 запрос = 78 «вычислений»
Изменение 5 терм	Изменение 5 строк в таблице терм = 5 «вычислений».	Кол-во вычислений принадлежностей и добавлений строк в таблицу принадлежностей является декартовым произведение измененных терм и всех объектов = $215 \cdot 5 \cdot 2 = 2150$ «вычислений». При этом обычный 1 запрос = 430 «вычислений».
Поиск 50 раз	Количество строк в двумерном массиве = декартовому произведению терм и объектов: $215 \cdot 39 = 8385 \cdot 2 = 16770$ «вычислений» - 1 запрос. $16770 \cdot 50 = 838500 - 50$ запросов.	Выборка из таблицы принадлежностей = $215 \cdot 39 = 8385$ «вычислений» – 1 запрос. $8385 \cdot 50 = 419250 - 50$ запросов.
	Однотипные операции	
Итого	$838500 + 15 + 5 = 838520$ «вычислений»	$419250 + 1170 + 2150 = 422570$ «вычислений»

По данным приведенным в таблице, как минимум по количеству вычислений, можно сформулировать следующие выводы:

- Время поиска растет с увеличением количества объектов и лингвистических терм.

- При одновременном поиске несколькими риэлторами, время поиска может увеличиться из-за хранения большого массива в оперативной памяти.

- Из-за специфичной сферы деятельности, количество операций поиска значительно больше, чем запросов на изменение или добавление данных, что влияет непосредственно на нагрузку веб-сервера, при базе данных построенной по принципу синтезирования.

Выводы

Была выбрана база данных по типу резюмирования, по следующим причинам:

- Количество таблиц незначительно влияет на производительность системы.

- Данные вычисляются намного реже и равномернее, при построении базы по принципу резюмирования, так как изменение или добавление лингвистических терм или объектов приводит к перерасчету, которые происходят в разное время и зависят лишь от количества объектов или лингвистических терм.. В тоже время, в базе построенной по принципу синтезирования, при нечетком поиске увеличилось бы время запроса из-за вычислений принадлежности, а также к самому поиску будет обращений больше, чем при редактировании объектов или терм.

- время и нагрузка на сервер будут увеличиваться при увеличении объектов недвижимости в базе данных, причем разница между разными базами, построенных по принципу резюмированию и синтезирования будет увеличиваться, оставляя преимущество за первым типом базы данных.

Библиографический список

1. Паклин, Н. Нечеткие запросы к реляционным базам данных. URL: <https://basegroup.ru/community/articles/fuzzylogic-queries> (дата обращения: 04.10.2017).
2. Ананьев Е.А., Рыбанов А.А. Исследование методов работы с нечеткими базами данных (на примере базы данных риэлтерского агентства) // Материалы X Международной студенческой электронной научной конференции «Студенческий научный форум», 2018