

УДК 004.896

РЕАЛИЗАЦИЯ АССОЦИАТИВНОГО ПОИСКА И ЛОГИЧЕСКОГО ВЫВОДА В БЗ С ПОМОЩЬЮ SQL-ЗАПРОСОВ

Кушнир А.Ю., Андрюхин А.И.

*Донецкий национальный технический университет
кафедра прикладной математики и информатики*

E-mail: artem.kushnir@hotmail.com

В статье рассмотрена задача автоматизированного поиска новых технических решений с использованием базы знаний физико-технических эффектов. Предложен многоэтапный метод решения, основанный на логическом выводе из базы знаний. Описана структура базы знаний и методика ее заполнения. Рассматривается приложение, реализующее разрабатываемый метод. Обсуждаются этапы решения и использование языка SQL на каждом из них.

Общая постановка проблемы

Изобретательская задача – это такая инженерная задача, которая содержит техническое противоречие, неразрешимое известными техническими средствами и знаниями, причем условия задачи исключают компромиссное решение. Если техническое противоречие преодолено – изобретательская задача решена, получено изобретение. Появление новых изобретений – основная форма развития и создания новой техники и технологии [1].

Деятельность исследователя – это нелегкий труд, получение результатов которого тяжело предсказать или запланировать. В XX веке возрос интерес к повышению интенсивности инженерного творчества. Были исследованы и улучшены уже известные эвристические методы. С развитием возможностей компьютеров стали появляться алгоритмы, основанные на переборе и комбинировании уже известной информации.

Известными являются методы Теории Решения Изобретательских Задач (ТРИЗ). Создатель ТРИЗ – Генрих Саулович Альтшуллер поставил задачу так: «Как без сплошного перебора вариантов выходить сразу на сильные решения проблемы?». Он и его коллеги проанализировали больше 200000 патентов, чтобы определить каким образом в основном происходит усовершенствование технических систем. В результате были открыты 40 изобретательских принципов, многие законы эволюции технических систем, концепции технического и физического противоречий и множество других теоретических и практических подходов [1]. ТРИЗ – это метод разрешения проблем, основанный на логике и данных, не интуиции, который усиливает способность команды исследовательского проекта решать проблемы креативно. Благодаря своей структуре и алгоритмическому подходу ТРИЗ также обеспечивает повторяемость, предсказуемость и надежность [2].

Часто случается, что при работе над определенной проблемой находятся решения других проблем, поэтому исследователю в своей деятельности также следует учитывать статистический закон Парето.

Закон Парето, или Принцип Парето, или принцип 20/80 – эмпирическое правило, введенное социологом В. Парето, в наиболее общем виде формулируется как «20 % усилий дают 80 % результата, а остальные 80 % усилий – лишь 20 % результата». Может использоваться как базовый принцип для оптимизации какой-либо

деятельности: правильно выбрав минимум самых важных действий, можно быстро получить значительную часть от планируемого полного результата, при этом дальнейшие улучшения неэффективны и могут быть не оправданы (согласно кривой Парето) [3].

Любой технический объект можно разложить на элементарные составные части, которые реализуют определенный физический закон. На современном уровне развития техники известно множество физических законов и возможностей их использования. Комбинируя их между собой, можно получить множество различных решений, в том числе представляющих практический интерес [4].

Чем больше знаний будет использовать человек в процессе изобретательской деятельности, тем более продуктивной она будет. Например, инженер обычно знает около 200 различных физико-технических эффектов, может применять не более 100, а в технической литературе их описано более 3000 [5]. Кругозор и возможности инженера в процессе изобретения возрастают вместе с числом физико-технических эффектов, которые он может использовать.

Таким образом, для повышения эффективности работы исследователя необходим инструмент, автоматизирующий поиск потенциальных технических решений, способный использовать большинство известных физико-технических эффектов, работающий в соответствии с принципами ТРИЗ и позволяющий видеть появляющиеся в процессе решения определенной проблемы потенциальные решения других проблем.

Описание решения задачи

Разрабатываемая в рамках исследования дипломного проекта магистра система WinRetriever, призвана помочь исследователю находить новые технические решения. В начале поисковой сессии в системе исследователь вводит описание того объекта, способ реализации которого он ищет, после чего он проходит последовательный ряд этапов, на каждом из которых определяет дальнейшее направление и приоритеты процесса поиска.

Процесс поиска в системе состоит из 5 этапов:

1. **Задание цели поиска.** На данном этапе вводится описание объекта поиска, его функций, возможностей, а также задаются входной и выходной потоки объекта, в качестве которых может выступать поток вещества, физическое или полевое воздействие и т.д.
2. **Настройка образа объекта поиска.** На основе данных, введенных исследователем на первом этапе, и текущей базы знаний физико-технических эффектов (ФТЭ), система создает образ объекта поиска, который будет использован при отборе потенциально полезных ФТЭ. Образ содержит информацию о словах, распознанных в описании объекта. На этом этапе пользователь настраивает силу влияния на выбор ФТЭ групп слов, отдельных слов, частей речи.
3. **Работа с отобранными ФТЭ.** После того, как настройка образа объекта поиска завершилась, производится составление рейтинга ФТЭ. Он показывает степень, в которой тот или иной ФТЭ имеет отношение к описанию объекта поиска. На данном этапе пользователь может вручную модифицировать рейтинг ФТЭ и удалить ненужные в дальнейшем эффекты.
4. **Работа с цепочками ФТЭ.** Каждый ФТЭ имеет входной и выходной потоки. Из отобранных ФТЭ строятся цепочки, так, чтобы типы входного и выходного потоков соседних ФТЭ совпадали. У ФТЭ может быть несколько

входных и\или выходных потоков, однако совпадения с одним из множества достаточно для создания связи в цепочке.

5. **Работа с цепочками устройств.** База знаний ФТЭ системы WinRetriever содержит также информацию об устройствах, реализующих физические эффекты и их характеристиках. Указывая необходимые значения характеристик (максимальная цена, температурные ограничения, вес и т.д.) исследователь получает цепочки устройств, построенные по цепочкам эффектов.

На каждом из этапов для взаимодействия с базой знаний используется язык SQL. На последних двух исследователь может пользоваться им вручную для задания необходимых параметров отбора. Вид приложения при работе исследователя на первом этапе приведен на рис. 1.

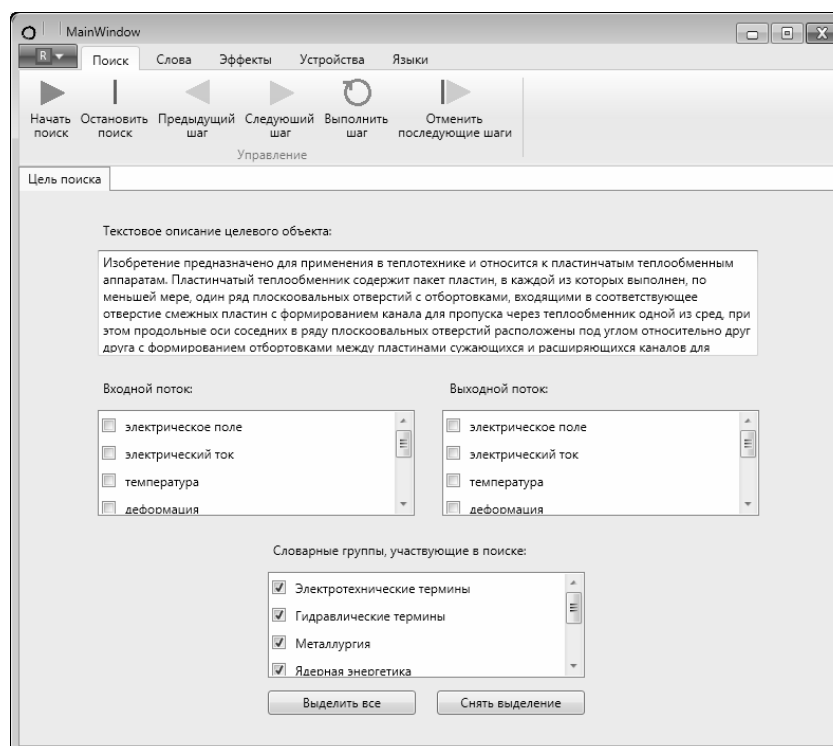


Рисунок 1 – Работа на первом этапе процесса поиска

Во время поисковых сессий система собирает статистическую информацию о тех решениях, которые принимает исследователь и результатах, к которым они приводят. На последних этапах исследователь отмечает полезные на его взгляд цепочки. Сбор и анализ этой информации позволяет отвечать на различные статистические вопросы и видеть слабые части системы и базы знаний.

Схема базы данных приведена на рис. 2.

Слова содержатся в таблице Words. В ней также хранится информация о части речи каждого слова. Эта информация используется при построении образов эффектов и объектов поиска. С помощью части речи определяются возможные формы, которые может принимать слово в тексте. Поскольку это довольно упрощенная система распознавания слов, то существует таблица WordForms, позволяющая указывать нестандартные формы слова.

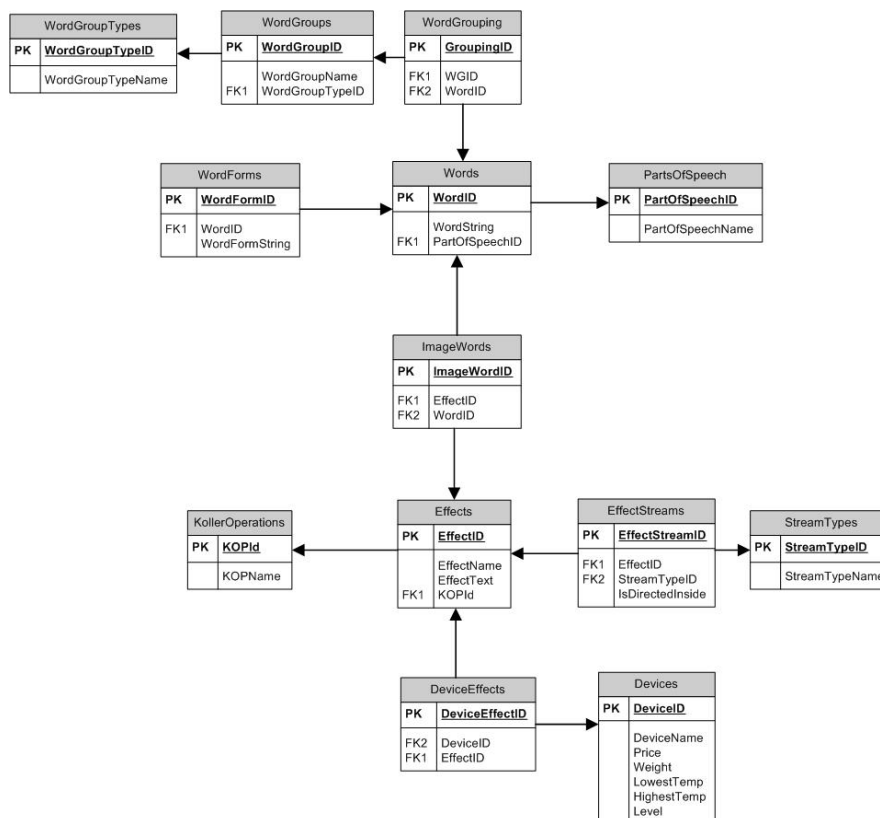


Рисунок 2 – Схема базы данных системы WinRetriever

Интерфейс пользователя реализован на 3 языках: украинском, английском и русском. Процесс поиска также осуществляется на одном из этих языков. Для поиска на языке необходима база знаний на этом языке. Разработка эффективной базы знаний – это трудоемкий и длительный процесс. Поэтому пока база знаний реализуется только на одном языке и соответственно осуществлять поиск можно будет только на одном языке. Для определения языка этой начальной базы знаний с помощью поисковых систем Интернет было проведено статистическое исследование, показывающее частоту встречаемости терминов ТРИЗ в Интернет. Например, поиск самого термина ТРИЗ дает результаты, показанные в таблице 1.

Таблица 1 – Результат поиска термина «ТРИЗ» на разных языках

	meta.ua	google.com	yandex.ru	rambler.ru	yahoo.com
Теорія рішення винахідницьких задач	472	3200	1243	1211	89
Теория решения изобретательских задач	4400	37300	133000	101000	92300
Theory of Inventive Problem Solving	111	34000	14100	298	573000

На основании результатов исследования было принято решение первой разрабатывать базу знаний на русском языке.

Система разрабатывается с помощью языка программирования Microsoft Visual C# 2010. Интерфейс пользователя реализуется посредством библиотеки Windows Presentation Foundation 4 (WPF 4).

Есть выражение: «Сначала горы – это горы, ручьи – это ручьи. Затем горы – это не горы, и ручьи – не ручьи. Но в конце горы – снова горы, и ручьи – снова ручьи». Одной из основных целей при разработке данной системы является помощь пользователю в видении изобретательской ситуаций с разных сторон, в отходе от стереотипного мышления и нахождении новых неочевидных связей между сущностями.

В соответствии с одной из известных гипотез человек, на первых этапах опознавания предмета мыслит по принципу наличия-отсутствия его признаков [6]. Также и WinRetriever ориентируется на наличие-отсутствие слов (признаков), которые содержатся в описаниях физико-технических эффектов.

Хорошо поставленный вопрос содержит в себе половину ответа, поэтому решение изобретательского противоречия может прийти к пользователю на любом из этапов работы с системой, в том числе и на первом. Система, можно сказать, помогает исследователю структурировать мысли и увидеть закономерности более четко и быстро.

Выводы

Одним из решающих факторов научно-технического прогресса является трансформация научных знаний и результатов творческой деятельности в производственные процессы.

Один человек или даже группа людей не может обладать или использовать весь объем знаний человечества. В то же время, чем больше знаний будет использовать человек в процессе изобретательской деятельности, тем более продуктивной она будет. Разрабатываемое решение направлено на повышение эффективности работы исследователя и использование как можно большего объема знаний.

Литература

- [1] Кушнир А. Ю., Андрюхин А.И. Реферат по теме выпускной работы «Реализация ассоциативного поиска и логического вывода в БЗ с помощью SQL-запросов» // Портал магистров ДонНТУ. [Электронный ресурс] / Кушнир А. Ю., Андрюхин А.И. – Режим доступа: <http://masters.donntu.edu.ua/2010/fknt/kushnir/diss/index.htm>
- [2] Katie Barry, Ellen Domb, Michael S. Slocum TRIZ - What Is TRIZ? // The TRIZ Journal [Электронный ресурс] / Katie Barry, Ellen Domb, Michael S. Slocum – Режим доступа: http://www.triz-journal.com/archives/what_is_triz/
- [3] Закон Парето // Материал из Википедии – свободной энциклопедии. [Электронный ресурс] / Режим доступа: http://ru.wikipedia.org/wiki/Закон_Парето
- [4] Кушнир А. Ю., Андрюхин А.И. Построение базы знаний физико-технических эффектов для поиска новых технических решений / «Інформаційні управляючі системи та комп'ютерний моніторинг» (ІУС та КМ-2010) / Матеріали І всеукраїнської науково-технічної конференції студентів, аспірантів та молодих вчених – ДонНТУ, Донецьк, 19-21 травня 2010. – 224-228 с.
- [5] Половинкин А.И. Основы инженерного творчества: Учеб. пособие для студентов вузов. – М.: Машиностроение, 1988. – 368 с.
- [6] Хофман И. Активная память: Экспериментальные исследования и теория человеческой памяти. – М.: Прогресс, 1986. – 312 с.