



Заключение

Исходя из всего вышесказанного, можно отметить, что внедренные на АО «РКЦ «Прогресс» передовые информационные технологии способствуют решению поставленных перед предприятием сложных научно-технических и организационных задач, а именно - сокращению сроков проектирования и конструирования РКТ, а также повышению качества выпускаемых изделий. Поэтому дальнейшая цель – переход на полностью электронный документооборот и отказ от бумажных носителей информации – является актуальной на сегодняшний день для предприятия оборонно-промышленного комплекса.

Литература

1. А.А. Суханов, О.Н. Рязанцев, С.А. Артизов, А.Н. Бриндигов, Н.И. Незаленов, А.В. Карташев, П.М. Елизаров, Е.В. Судов. Концепция развития ИПИ-технологий для продукции военного назначения, поставляемой на экспорт. – М.: НИЦ CALS «Прикладная логистика», 2013.
2. Кременецкая М.Е. Модели управления единым информационным пространством научно-производственного предприятия // Мат-лы научно-технической конференции «Перспективные информационные технологии в научных исследованиях, проектировании и обучении-2012». Самара, СГАУ, 2012. – с.203-207.
3. Никашина И.В., Филатов А.Н., Микушкина С.М., Рогова Т.С. Электронный технический документооборот конструкторской документации как основа единого информационного пространства ОАО «РКЦ «Прогресс» // Мат-лы XX Научно-технической конференции молодых ученых и специалистов, г. Москва, ОАО РКК «Энергия» имени С.П. Королева, 2014г. – с.627-629.
4. Космодемьянский Е.В. Разработка технологии информационной поддержки проектирования и конструкторской подготовки производства космических аппаратов дистанционного зондирования земли // Автореферат диссертации на соискание ученой степени кандидата технических наук, г. Самара, 2014.
5. И.В. Никашина, Е.В. Головченко, А.Н. Филатов, С.М. Микушкина. Особенности информационной поддержки конструкторской подготовки производства в едином информационном пространстве предприятия // Мат-лы IV Всероссийской научно-технической конференции «Актуальные проблемы РКТ» (IV Козловские чтения), Самара, 2015 – т.2, с.156-158.



В.В. Ворошилов, С.А. Пиявский

ЭВРИСТИЧЕСКИЙ АЛГОРИТМ ОПТИМИЗАЦИИ ПОИСКОВОЙ ДЕЯТЕЛЬНОСТИ В ФАКТОГРАФИЧЕСКИХ ПОИСКОВЫХ СИСТЕМАХ

(Самарский государственный архитектурно-строительный университет)

В последнее время широкое распространение получило направление исследований «big data», связанное с использованием структурированных и неструктурированных данных огромных объемов циркулирующих в корпоративных сетях и сети интернет. Это совпало со все большим интересом, проявляемым к созданию поисковых систем, оперирующих не документами а фактами, так называемых фактограф информационных поисковых систем. Отдельные элементы таких систем уже реализованы в поисковых серверах: Яндекс, Google, WolframAlpha и других. При этом актуальным представляется разработка математических моделей и алгоритмов, позволяющих пользователю наиболее эффективно осуществить подбор фактографической информации используя для этого спец организацию баз данных и интеллектуальные алгоритмы поиска.

Развиваемый авторами в этом направлении так называемый атомарный подход [1] основан на организации информационного массива фактографической информации в виде так называемых «атомов» информации. Элементом, единицей фактографической информации является «атом», состоящий из совокупности некоторого числового значения, его размерности и описания – набора лексем, характеризующего смысл этого значения (рисунок 1). Лексема представляет собой словоформу (термин, понятие) из семантического тезауруса.

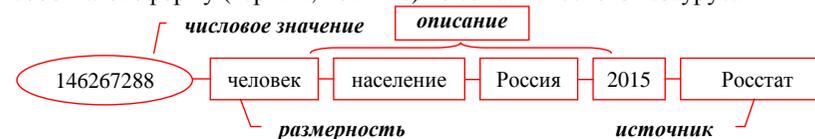


Рисунок 1 – Пример «атома» фактографической информации

Суть атомарного подхода заключается в том, чтобы положить в основу фактографического поиска базу данных, информационными единицами которой являются отдельные «атомы» фактографической информации. Будем называть такую базу данных атомарной (АтБД).

В [2-4] предложен ряд алгоритмов, позволяющих пользователю получать от АтБД блоки фактографической информации (кластеры), обладающие максимальной информативностью. Эти блоки можно характеризовать булевыми таблицами, в заголовках столбцов которых располагаются некоторые лексемы тезауруса, в заголовках строк числовые значения атомов, а в ячейке таблицы представляется единица, если лексема соответствующей ячейке столбца входит в описание атома, соответствующего строке ячейке; в противном случае представ-



ляется ноль. Информативность такого кластера определяется отношением суммы значений ячеек таблицы, отнесенным к общему числу ее ячеек. Кластеры с большей информативностью представляют для пользователя наибольший интерес. Разработаны математические модели, позволяющие формировать кластеры максимальной информативности путем численного решения задач булевого нелинейного и линейного программирования. Однако в условиях «big data» время решения таких задач достаточно велико, что не позволяет при существующих вычислительных возможностях использовать их для широкого круга пользователей.

Поэтому предлагается эвристический алгоритм, позволяющий алгоритмизировать деятельность пользователя в поисковой фактографической системе, использующей атомарную БД для получения наиболее релевантной и полной информации в интересующей его предметной области. Схема алгоритма представлена на рисунке 2.

Дадим его краткое описание.

Этап 1. Пользователь формулирует поисковый запрос (ПЗ) из лексем, входящих в тезаурус АТБД.

Этап 2. Для каждого атома, содержащего хотя бы одну лексему ПЗ рассчитываются две характеристики: релевантность и полнота. Релевантность равна отношению числа лексем ПЗ, входящих в описание атома, к общему числу лексем ПЗ. Полнота равна отношению числа лексем ПЗ, входящих в описание атома, к общему числу лексем в описании атома.

Этап 3. Пользователь задается коэффициентом «релевантность-полнота» (коэффициент Р-П) заключенным в пределах от нуля до единицы и отражающим приоритетность для него каждого из этих факторов.

Этап 4. С использование этого коэффициента рассчитывается индекс атома как среднее взвешенное значение его релевантности и полноты, и атомы упорядочиваются по убыванию их индекса.

Этап 5. Пользователю сообщается зависимость средней и гарантированной релевантности и полноты, выдаваемой ему упорядоченной по индексу атомов информации в зависимости от объема выдачи. Он выбирает интересующий его объем и получает соответствующую информацию.

Этап 6. Если пользователя не устраивают низкие значения полноты, выданной ему информации, это свидетельствует о том, что в описаниях ее атомов преобладают лексемы, не входящие в ПЗ. В этом случае возникает необходимость расширить круг активно привлекаемых (так называемых косвенных) лексем. Осуществляется переход к этапу 7.

Этап 7. Формируется набор косвенных лексем, в которые включаются не входящие в ПЗ из описания атомов, включенных в выбранную пользователем выдачу. Пользователь задает коэффициент важности учета косвенных лексем при расчете релевантности и полноты атомов. Затем алгоритм возвращается к этапу 2, расширенному с учетом наличия косвенных лексем.

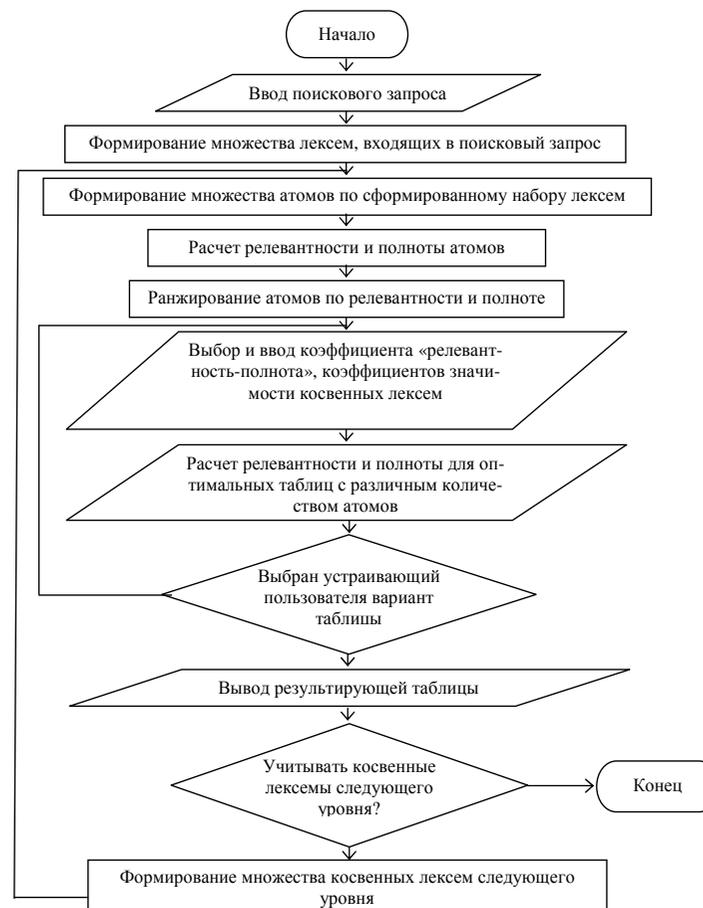


Рисунок 2 – Схема эвристического алгоритма оптимизации поисковой деятельности в фактографических поисковых системах

Литература

1. Пиявский С. А., Ворошилов В. В. Технология поиска и систематизации фактографической информации в информационном пространстве, Научное обозрение 2/2014, стр. 167-176, 2014.
2. Пиявский С.А., Ворошилов В.В. Математические модели оптимальной систематизации фактографической информации, Инфокоммуникационные технологии №1/2014, стр. 77-85.
3. В.В. Ворошилов Математические модели синтеза фактографической информации. Наука глазами молодежи: сборник научных статей студентов вузов, ссузов и учащихся общеобразовательных учреждений / – Тольятти; Самара: ООО «Издательство Ас Гард», 2012. – 561 с. 498-501



4. С.А. Пиявский, В.В. Ворошилов Математические модели оптимальной систематизации фактографической информации в электронных библиотеках, Интернет и современное общество: сборник научных статей. Труды XVI Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2013), Санкт-Петербург, 9 – 11 октября 2013 г. — СПб.: НИУ ИТМО, 2013. — 304 с.

Л.Р. Габдрахманова

ПОСТРОЕНИЕ СИСТЕМЫ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ СИМВОЛОВ

(Казанский национальный исследовательский
технический университет им А. Н. Туполева – КАИ)

Проблема распознавания символов остро стоит при компьютерном чтении, распознавании отсканированного текста, выделении и распознавании символов на фотографии (например, чтение автомобильного номера нарушителя дорожного движения).

Целью работы является разработка программы распознавания цифр и букв, обеспечивающих анализ и обработку информации на изображении с целью идентификации номерного знака отечественных автомобилей.

Задача распознавания номеров требует описания 10 цифр и 12 букв. Общая структура типовой системы распознавания автомобильных номеров состоит из устройства съема изображения, предварительной обработки изображения, обнаружения номерной пластины, извлечения символов и собственно распознавания символов.

Алгоритм распознавания номерного знака состоит из следующих этапов: предварительная обработка изображения, сегментация и распознавание.

Предварительная обработка изображения заключается в обработке изображения различными фильтрами с целью улучшения качества. Процедуры сегментации и распознавания работают с бинарным изображением, т.е. только с черными и белыми пикселями. Поэтому прежде чем передавать работу этим процедурам, нужно исходное цветное изображение привести к бинарному виду. Эта задача решается в два этапа.

На первом этапе цветное изображение превращается в черно-белое и представляется в градациях серого. Для каждого пикселя вычисляется его яркость в пределах от нуля до 255. Уровню яркости 0 соответствует черный цвет, уровню 255 – белый. Таким образом, для хранения изображения приходится один байт на пиксель. Яркость пикселя вычисляется по одной из следующих формул:

$$I = \frac{R+G+B}{3}, I = \frac{\max(R, G, B) + \min(R, G, B)}{2}$$

где R, G, B – нормированный на 256 (один байт) красный, зеленый и голубой



компонент цвета пикселя соответственно.

Вторым этапом является собственно бинаризация. Примеры бинаризации изображения с различными параметрами приведены на рис. 2.

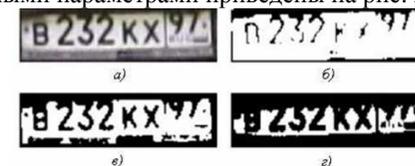


Рис. 2. Бинаризация, а) исходное изображение, б) 25% черных пикселей, в) 50% черных пикселей, г) 75% черных пикселей

Недостатком предобработки изображения с бинаризацией является то, что выбираемый порог бинаризации не обеспечивает необходимого качества для любого типа изображения. Такие факторы, как освещение или даже цвет транспортного средства влияют на качество бинаризации изображения.

На этапе сегментации выделяются символы, которые затем распознаются каким-либо методом. Для решения задачи сегментации изображения используется кластеризация К - средних. Метод К - средних считается наиболее быстрым [1], [2].

Процесс распознавания символов автомобильного номера базируется на обучении нейронной сети, а именно, обучение нейронной сети, используя генетические алгоритмы для поиска весов синаптических связей и/или структуры сети.

Использование генетического алгоритма для обучения нейронных сетей обладает следующими достоинствами [4,5]:

- изучение пространства поиска при помощи множества решений, а не одного. Это дает значительный эффект при поиске глобальных минимумов адаптивных рельефов;

- генетические алгоритмы малочувствительны к росту размерности множества оптимизации.

Управляемыми параметрами в генетическом алгоритме являются:

- длина хромосомы;
- наполнение хромосомы (локусы и аллели);
- параметры оператора кроссовера;
- параметры оператора мутации;
- параметры оператора инверсии;
- параметры выбора лучших особей;
- параметры генерации начальной и последующих популяций и т.д.

Были рассмотрены варианты обучения нейронной сети с помощью алгоритма обратного распространения ошибки и с помощью генетических алгоритмов. Обучение в обоих случаях заканчивается, когда на всем обучающем реальный выходной сигнал был равен требуемому выходному сигналу.

Результаты работы операторов генетических алгоритмов были сравнены с