

При наличии источника излучений производится определение координат источника излучений необходимым алгоритмом обработки данных и идентификация его изотопного состава. Результат обработки визуализируется на ЖКИ дисплее.

Таким образом, описываемая система позволяет обнаружить присутствие источника гамма-излучения в поле зрения системы без сканирования исследуемого участка местности и идентифицировать его изотопный состав.

Использование таких чувствительных и высоко автоматизированных систем позволяет осуществлять скрытый контроль непрерывного потока людей, транспорта, багажа или грузов на таможенных постах, автомагистралях, аэропортах, железнодорожных, морских, речных и автомобильных вокзалах.

1. *Акимов Ю.К.* Сцинтилляционные методы регистрации частиц больших энергий. Изд. МГУ, 1963.
2. *Батій В.Г., Єгоров В.В., Закревський Ю.А. та ін.* Пристрій для вимірювання кутового розподілу інтенсивності гамма-випромінювання МПК7 G01T 1/28. Патент на винахід № 51987 від 15.07.2004 г. - Бюл. "Промислова власність", № 7, 15.07.04.
3. Патент РФ на изобретение 2068184 от 20.10.1996 . G01T1/29.
4. *Забулонов Ю.Л., Лисиченко Г.В., Макарець М.В.* Спосіб динамічного аналізу нестационарних радіаційних полів - Заявл. № U 2006 12043 16.11.06

Поступила 2.08.2010р.

УДК 004.832.3

В.В. Зосимов, аспирант, А.С. Булгакова, аспирант,
Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины

ПРОЕКТИРОВАНИЕ ПРОГРАММНОГО МОДУЛЯ ДЛЯ ИССЛЕДОВАНИЯ РАБОТЫ АЛГОРИТМОВ ПОИСКОВЫХ СИСТЕМ

Аннотация. В данной статье рассмотрен принцип работы программного комплекса, предназначенного для сбора значений внешних и внутренних факторов ранжирования сайтов в выдаче поисковых систем и дальнейшего анализа этих данных с помощью многорядного алгоритма МГУА и его модификаций. На основе анализа будут определены основные факторы, учитываемые поисковыми системами при ранжировании сайтов, что позволит оценить эффективность работы поисковых систем и в дальнейшем оптимизировать существующие алгоритмы ранжирования.

Abstract. This article discusses the principle of the software complex intended for the collection of values of external and internal factors of ranking websites to obtain the search and subsequent analysis of these data using the multilayered GMDH algorithm and its modifications. The analysis will identify the main factors taken into account by search engines when ranking websites, which will help us to estimate the effectiveness of search engines and further optimize the existing ranking algorithms.

Ключевые слова. Алгоритм, ранжирование сайтов, МГУА, многоярный алгоритм, моделирование.

Введение. Мы уже давно привыкли находить почти любую, интересующую нас информацию в интернете. Это стало возможным во многом благодаря появлению и развитию поисковых систем, основная задача которых – сделать поиск информации в Интернете более эффективным. Для этого были разработаны алгоритмы ранжирования сайтов в поисковой выдаче, результатом работы которых является отсортированный по релевантности (наиболее точно удовлетворяющий условия поиска) список сайтов.

Современный Интернет все больше становится похож на огромную рекламную площадку, что значительно затрудняет поиск информации в сети. Алгоритмы работы поисковых систем развиваются вместе с Интернетом, они постоянно совершенствуются, но, не смотря на это, поиск нужной информации становится все более сложной задачей.

В данной работе описаны принципы работы программного комплекса, который позволит нам провести анализ существующих алгоритмов, используемых современными поисковыми системами и выявить основные факторы, которые влияют на ранжирование сайтов в результатах поисковой выдачи.

Объектом исследования являются процессы поиска информации в сети Интернет.

Предмет исследования – методы работы поисковых систем и алгоритмы ранжирования сайтов в поисковой выдаче.

Цель исследования – разработка и оптимизация алгоритмов ранжирования сайтов в поисковой выдаче.

Ожидаемые результаты – система оценки эффективности работы поисковых систем.

Основное направление нашей деятельности – анализ существующих алгоритмов ранжирования сайтов в поисковой выдаче с целью выявления их недостатков и дальнейшей их оптимизации.

1. Основные известные факторы, влияющие на ранжирование сайтов

Поисковая оптимизация — комплекс мер для поднятия позиций сайта в результатах выдачи поисковых систем по определенным запросам пользователей. Обычно, чем выше позиция сайта в результатах поиска, тем

больше заинтересованных посетителей переходит на него с поисковых систем. При анализе эффективности поисковой оптимизации оценивается стоимость целевого посетителя с учетом времени вывода сайта на указанные позиции, на который привлекаются целевые посетители.

Работа по оптимизации включает в себя работу с *внутренними факторами* (находятся под контролем владельца веб-сайта) — приведение текста и разметки страниц в соответствие с выбранными запросами, улучшение качества и количества текста на сайте, стилистическое оформление текста (заголовки, жирный шрифт), улучшение структуры и навигации, использование внутренних ссылок, а также *внешними факторами* — обмен ссылками, регистрация в каталогах и прочие мероприятия для повышения и стимулирования «ссылаемости» на ресурс.

Механизм ранжирования - это программа, которая определяет релевантность страницы (степень соответствия) поисковому запросу на основе семантического анализа документа, плотности и соответствия ключевых слов, ссылок с других Интернет-ресурсов и других параметров. От релевантности страницы зависит ее место при выводе результатов поиска.

Факторы, которые используют современные поисковые системы при ранжировании поисковой выдачи, делятся на внешние и внутренние.

Внешние факторы направлены на увеличение авторитетности сайта, его «общего веса». Их совокупность не влияет на релевантность сайта напрямую, но является важным коэффициентом при формировании поисковой выдачи.

К внутренним факторам относятся все внутреннее наполнение сайта. Поисковые системы в ходе индексации страниц сайта проводят семантический анализ его содержимого и на основе полученных данных определяют релевантность сайта тем или иным поисковым запросам.

2. Сбор данных о сайтах и приведение их к единому виду

Для проведение анализа эффективности работы алгоритмов ранжирования поисковой выдачи мы выбираем сайты с первых нескольких страниц поисковой выдачи и собираем значения всех внешних и внутренних факторов, которые могут повлиять на их ранжирование в поисковой выдаче.

Для контроля результата работы нашей программы мы вручную просматриваем анализируемые сайты и оцениваем, насколько они релевантны введенному нами поисковому запросу.

После того, как все значения внешних и внутренних факторов ранжирования выбранных нами сайтов были собраны, мы приступаем к приведению их к единому формату. Все данные формируются в таблицу, которая в последствии будет передана для анализа в следующую часть программного комплекса.

3. Проектирование системы анализа и моделирования исследуемых данных

Для проведения исследования методов поискового ранжирования, используемых в любой поисковой системе, используется программный

комплекс, основанный на МГУА алгоритмах.

На рис. 1 представлена структурная схема основной части программного комплекса, с помощью которого будут проходить исследования факторов, влияющих на поисковую выдачу, а именно, будет построена модель ранжирования, на основе которой можно определить значимость факторов, а также спрогнозировать результаты поисковой выдачи (для нахождения модели используются оптимизации многорядного алгоритма МГУА [1]). Существует два основных типа алгоритмов МГУА – комбинаторный и многорядный. Комбинаторный алгоритм МГУА реализует полный перебор всех возможных структур моделей, что занимает много времени, если входных переменных больше, чем 20. Для решения задач моделирования большой размерности необходимы алгоритмы укороченного перебора моделей. Одним из таких алгоритмов является классический многорядный алгоритм МГУА, который требует значительно меньшего количества времени для нахождения оптимальной модели.

Блок хранения данных



Рис.1 Построение прогнозной модели с использованием оптимизаций многорядного алгоритма МГУА

В данном исследовании присутствует большое количество входных переменных (больше чем 20), поэтому для решения поставленной задачи был выбран комбинаторный алгоритм и его модификации.

С помощью *блока хранения данных*, имея начальную выборку (сгенерированную или полученную из существующего файла), есть возможность разбивать данные на проекты, сохранять промежуточные расчеты, для дальнейшего продолжения процесса моделирования, сохранять результаты расчетов, а также использовать полученные модели на новых данных. Более детально блок хранения данных рассмотрен на рис.2.

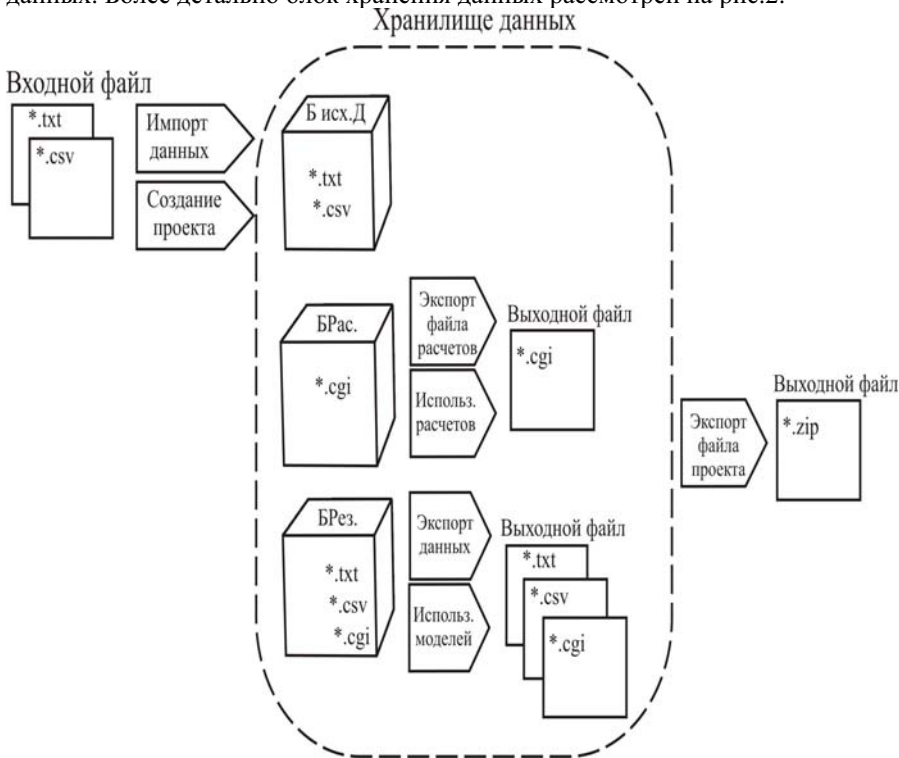


Рис.2. Блок хранения данных

После того как начальная выборка сгенерирована или получена, мы переходим к *блоку формирования задачи* (рис.3.). На этом этапе выборка делится на две части: на учебную и проверочную. Учебная подвыборка нужна для оценивания параметров модели, а проверочная – для определения прогнозирующей способности модели.

При генерировании выборки данных есть возможность задавать: тип разбиения выборки; уровень шума, выбор класса модели.

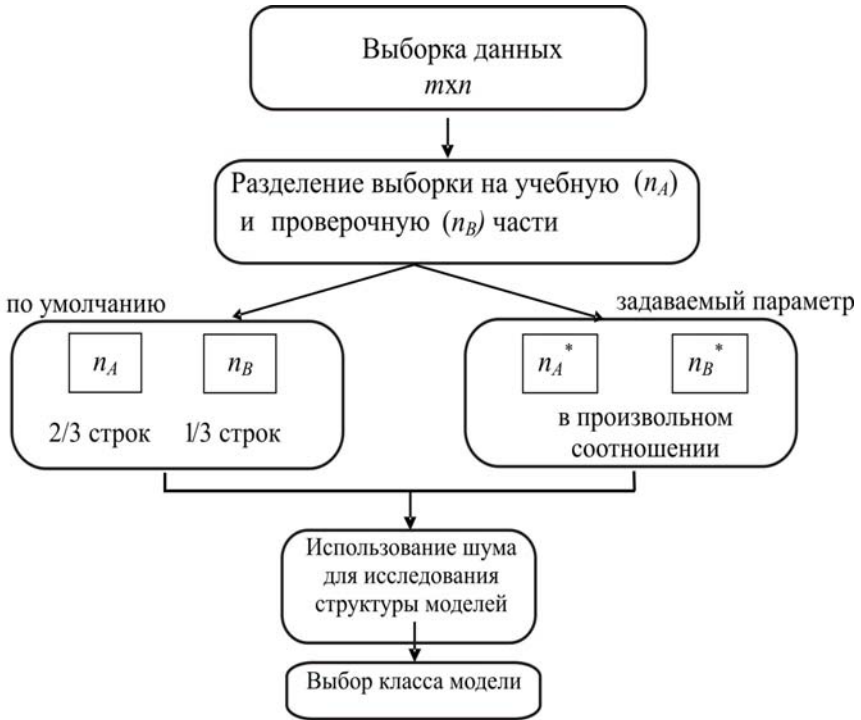


Рис.3. Блок формирования задачи

n_A – длина учебной части выборки данных;

n_B – длина проверочной части выборки данных;

Дальше, в зависимости от использования различных оптимизаций многоядного алгоритма МГУА, генерируются модели, различной сложности, для каждой из которой вычисляется значение критерия, по которому они отбираются для следующего слоя. В программном комплексе предусмотрено автоматическое управление и ручное управление. Ручное управление отличается тем, что на каждом слое можно изменять структуру сети (все возможные структуры описаны в [1]).

После того как оптимальная модель будет найдена можно проверить, как ведет себя модель на новых данных, выполнить имитационное моделирование и использовать построенные модели для принятия решений.

Данный программный комплекс дает возможность:

- самостоятельно конструировать методы моделирования по данным наблюдений;
- сравнивать полученные модели по заданным критериям;

- разрабатывать методики и планировать статистические испытания;
- решать задачи моделирования;
- проводить имитационные эксперименты с моделями, построенными по разным методам моделирования.



Рис.3. Блок решения задачи моделирования.

AR_1, AR_2, \dots, AR_V – значение критерия регулярности для V лучших моделей.

Заключение. На основе полученных в результате анализа данных планируется разработать методы, которые позволят оптимизировать существующие алгоритмы ранжирования сайтов в поисковой выдаче. Это позволит сделать поиск информации в сети Интернет более простым и удобным, что значительно сэкономит время пользователей сети.

1. *Volodymyr Stepashko, Oleksandra Bulgakova, Viacheslav Zosimov.* Modified multilayered GMDH algorithm with combinatorial optimization of partial descriptions complexity. – Proceedings of the International Workshop on Inductive Modelling IWIM-2010, Ukraine. – Yevpatoria, 2010.
2. *Байков Владимир Дмитриевич* Интернет. Поиск информации. Продвижение сайтов. — СПб.: БХВ-Петербург, 2000. — С. 288.
3. *Колисниченко Денис Николаевич* Поисковые системы и продвижение сайтов в Интернете. — М.: «Диалектика», 2007. — С. 272.
4. *Ашманов И. С., Иванов А. А.* Продвижение сайта в поисковых системах. — М.: «Вильямс», 2007. — С. 304.

Поступила 9.08.2010г.