

УДК 025.4

Мэн Цинань

АНАЛИЗ МЕТОДОВ КЛАССИФИКАЦИИ ИНФОРМАЦИИ В ИНТЕРНЕТЕ ПРИ РЕШЕНИИ ЗАДАЧ ИНФОРМАЦИОННОГО ПОИСКА

Воронежский институт высоких технологий

В связи с тем, что в Интернете хранится большое количество информации, должны быть использованы эффективные методы ее поиска. Чертами, присущими профессиональному поиску, являются полнота, достоверность и высокая скорость. Эти характеристики могут быть достигнуты при использовании соответствующих способов классификации. В статье рассмотрено несколько подходов. Кластерный метод основывается на разбиении элементов множества на группы. Расстояние между элементами задается метрикой. Лингвистический анализ основывается на возможностях извлечения информации из текста. Представлена схема лингвистической обработки. Статистические подходы исходят из определенных закономерностей частоты встречи слов. Анализ признаков заключается в изучении морфемных, морфологических и синтаксических признаков слов и предложений в тексте. Семантический анализ занимается разбором текста относительно значения слов внутри него. Комбинированный подход подразумевает использование нескольких из вышеописанных подходов в связке, последовательной или параллельной, для повышения точности анализа. В зависимости от возникающих задач, будет полезен соответствующий метод классификации.

Ключевые слова: поиск информации, интернет, статистика, классификация, семантика, признак, метрика, слово.

Интернет в 21-ом веке является неотъемлемой частью повседневной жизни. Экономическая, социальная и научная деятельность человечества в той или иной степени связана с интернет-технологиями. В наши дни можно проводить переговоры с партнёрами по бизнесу, денежные переводы, онлайн консультации, обучение и многое другое не выходя из дома. Мобильный интернет привязал человека к виртуальному миру - в любой момент времени и в любом месте на земном шаре, имея доступ к Интернету, можно быть в курсе всего, что происходит в реальном мире [1-4].

В связи с тем, что мы имеем дело с большими объемами данных, требуется использовать подходы, позволяющие ее классифицировать.

Целью работы является рассмотрение методов классификации информационного поиска (ИП) и информационных ресурсов (ИР), для персонализации информационного поиска в Интернете.

1.Кластерный метод

Кластеризация - это автоматическое разбиение элементов некоего множества на группы (кластеры) в зависимости от показателей их схожести.

Элементами множества может быть, что угодно: объекты с определённым набором данных или вектора характеристик.

Большинство исследователей считают, что прародителем кластерного анализа является Роберт Коати Трион (Robert Choat Tryon) - американский исследователь поведения животных, который предложил систематизировать методы анализа влияния окружающей среды (экология, социальный уровень и т.д.) на поведение субъектов исследования (животных, людей) и предложил группировать субъекты исследования в кластеры.

Предложенный им метод позволил с большой точностью определять причины и возможные последствия поведения человека в стрессовых ситуациях, исходя из его социального окружения.

У кластеризации существует большое количество практических применений. Кластеризация позволяет, например, провести анализ данных, поиск информации или группировку объектов по признакам и свойствам [5-8].

Так же кластеризация сама по себе является важной формой абстракции данных, и в этой области был получен ряд интересных научных результатов.

Говоря о кластеризации интернет - объектов, необходимо определить следующие базовые понятия.

Объект - элементарная единица, которая может быть представлена с помощью набора числовых характеристик, и с которой оперируют алгоритмы кластеризации.

Каждому объекту x_i , $i \in I$, сопоставляется вектор числовых характеристик $z_i = (z_{i1}, \dots, z_{ij}, \dots, z_{in})$.

Кардинальность вектора n определяет размерность пространства характеристик. Расстояние $p(x_i, x_k)$ между объектами x_i и x_k - результат применения выбранной метрики в пространстве характеристик.

В настоящее время, существует большое количество метрик для оценки расстояния между векторами одного и того же векторного пространства.

К простым метрикам можно отнести евклидово расстояние или его квадрат, манхэттенское расстояние, степенное расстояние и другие [9, 10].

К сложным метрикам можно отнести расстояние между центрами тяжести, групповое среднее расстояние, расстояние Чебышева и другие.

2.Метод анализа содержания текста

Лингвистический анализ - метод исследования текста, который может быть охарактеризован как лингвосмысловой анализ. Он представляет собой изучение методов, позволяющих автоматизировано «понимать» текст, т.е. уметь извлекать из него нужную информацию и отвечать на вопросы, заданные относительно текста. Лингвистический

анализ применим, в частности, к извлечению информации, машинному переводу, а также ко многим областям искусственного интеллекта, относящимся к общению с пользователем.

Существует множество подходов к лингвистическому анализу. Среди них можно выделить статистический анализ, анализ признаков, семантический анализ и комбинированный подход.

3. Статистический метод

Данный метод подразумевает изучение последовательности слов в предложениях текста, а также вывод определенных закономерностей на основании проведенного изучения. Для этого производится подсчет частоты встречи слов в тексте, а также вероятность появления слов друг за другом. С помощью статистических методов анализа текста решается проблема классификации текстов [11].

Соответственно, анализируя содержание слов, можно получать вероятности встречи \wedge - граммы или частоту терминов (TF или *TF-IDF*) в художественном или в научном тексте. После этого найденные вероятности переводятся в веса и складываются. Текст будет относиться к тому классу, вес которого окажется больше.

4. Лингвистическая обработка запросов интернет-пользователей и текстов интернет-ресурсов

Для применения методов лингвистического анализа необходимы леммы всех слов текста (анализ признаков) и частота встречи этих лемм в тексте (статистический анализ). Следует отметить, что анализ моделей связей между словами внутри предложений и предложениями внутри текста в данной работе не рассматривается.

Первым шагом к решению поставленной проблемы анализа содержания, как ИР, так и запросов ИП является разбиение текста интернет - страниц и интернет - запросов на отдельные слова (термины), т.к. текст в формальном определении является просто набором слов. Здесь уже возможны проблемы и неоднозначные трактовки: что считать словом, как относиться к сложным знакам пунктуации и т.д. Введём набор простых правил, описывающих большинство случаев, которые могут встретиться в априорно правильном тексте.

Словом или **термином** назовем последовательность символов букв, ограниченная с обеих сторон пробелами либо знаками препинания, в которой могут присутствовать цифры, в том числе и на первой позиции. Все знаки препинания и специальные символы («+», «-», «/», «=» и т.д.) заменяются пробелами, тем самым, непрерывная последовательность символов превращается в слова, отдельные друг от друга пробелами.

Леммой или **корнем слова** будем считать урезанную последовательность символов термина, получаемую помощью специально разработанного алгоритма отсечения окончаний с учетом двухуровневого

словаря, включающего глобальный словарь терминов и словарь лемм, который может заполняться с помощью существующих открытых словарей и динамически пополняться при появлении новых терминов. Очевидно, что нескольким терминам может соответствовать одна лемма.

Из сформулированных правил следует, любые тексты действительно подойдут под описания, приведённые выше. «Правильными» текстами в данном случае будут считаться тексты, как на русском, так и на английском языках, находящиеся в открытом доступе в сети Интернета и допускающие компьютерную обработку.

Априори будем считать, что ИР не содержат недопустимые для заданного языка символы либо тексты с синтаксическими опечатками. Все специальные символы фильтруются. Выявление синтаксических опечаток - отдельная серьёзная задача, которая может быть отведена лингвистическому эксперту, который, в свою очередь, может внести исправления в словарь лемм.

Пользуясь справочниками, удалось написать алгоритм заполнения словаря терминов и лемм, соответствующий приведенным выше правилам.

В диссертации, как уже было сказано, применяется комбинированный подход, основанный на статистическом методе, методе анализа признаков и особенностях *DOM-моделей* ИР.

Семантический анализ требует дополнительного исследования, и не будет рассматриваться в рамках текущей работы.

На Рисунке 1 представлены этапы процесса обработки содержания запросов ИП и текстов ИР от первичного «грязного» текста до лемм.

Результатом выполнения этого процесса является формирование статистики лемм и наращивания словаря с помощью лингвистического эксперта.

Схема алгоритма лингвистической обработки терминов (слов) приведена на Рисунке 2.

Так как на этапе работы рассматриваемого алгоритма происходит полный перебор слов из текста ИР или поискового запроса ИП, целесообразно проводить здесь же подсчёт встречаемости слов в тексте, применяя известные статистические методы. Затем, на основании полученной информации будут сформированы характеристические вектора классифицируемых и

интернет-объектов и глобальный словарь терминов, с которыми будут дальше работать алгоритмы кластерного анализа.

В алгоритме Рисунка 2, можно воспользоваться особенностями *DOM- модели* ИР, позволяющей локализовать позиции терминов и тем самым выделять особо важные термины (например, заголовки и наименование интернет - страниц), с целью повышения значений числовых

характеристик у ключевых терминов или даже исключения динамических элементов, «просочившихся» при чтении содержания текстов ИР.

С помощью описанного выше алгоритма, текст ИР разбивается на термины, которые уже можно обрабатывать. Необходимо обратить внимание на процесс обработки терминов, а точнее на стемминг (*stemming*) и лемматизацию терминов для приведения слова к начальной форме и тем самым сгруппировать их, что, в свою очередь, сделает статистику терминов более достоверной.

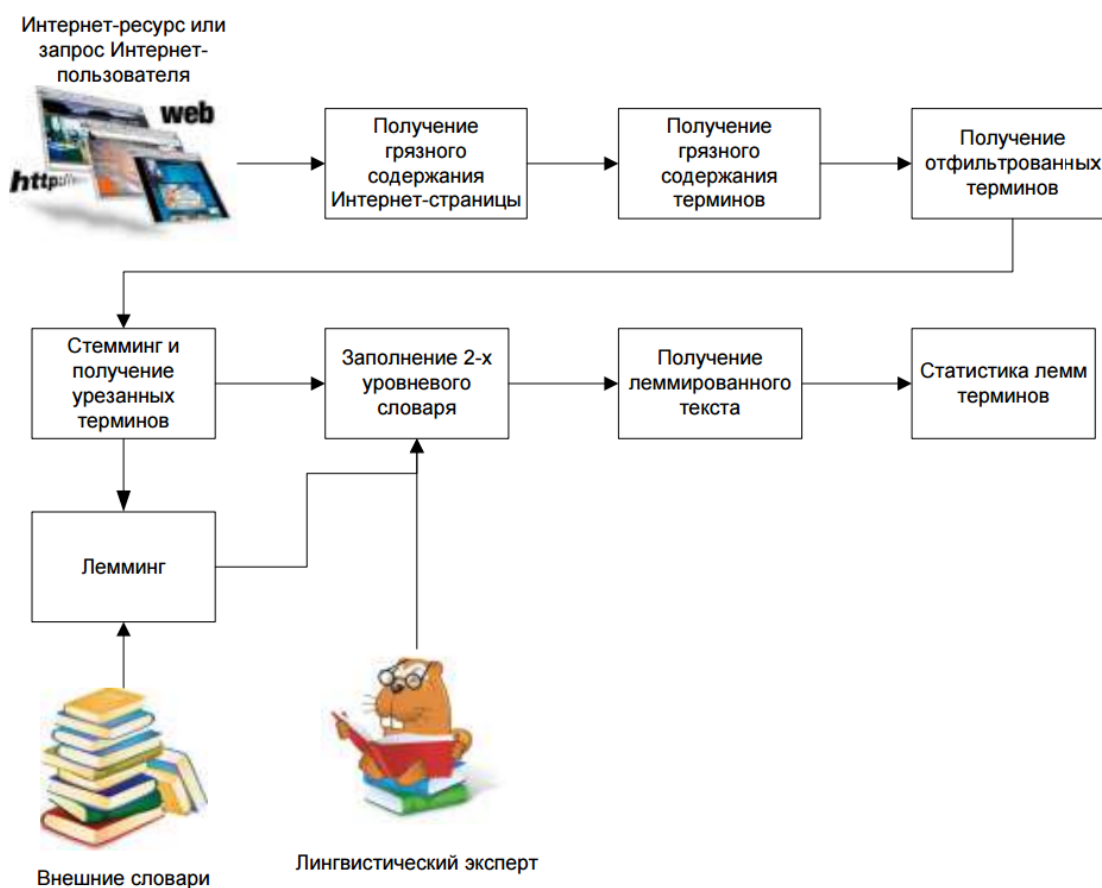


Рисунок 1 - Этапы процесса обработки содержания запросов

Начальной стадией работы почти всех лемматизаторов является стемминг.

Этот процесс подразумевает выделение корня слова, тогда как лемматизатор на основании корня подбирает базовую форму, подходящую для него. Одними из признанных фаворитов среди стеммеров является программы, реализующие алгоритмы усечения окончаний.

Алгоритмы усечения окончаний могут быть использованы как для русских, так и для английских терминов, а полученные впоследствии

усеченные термины могут быть применены в качестве лемм для формирования статистики первоначальных слов из считанного текста.

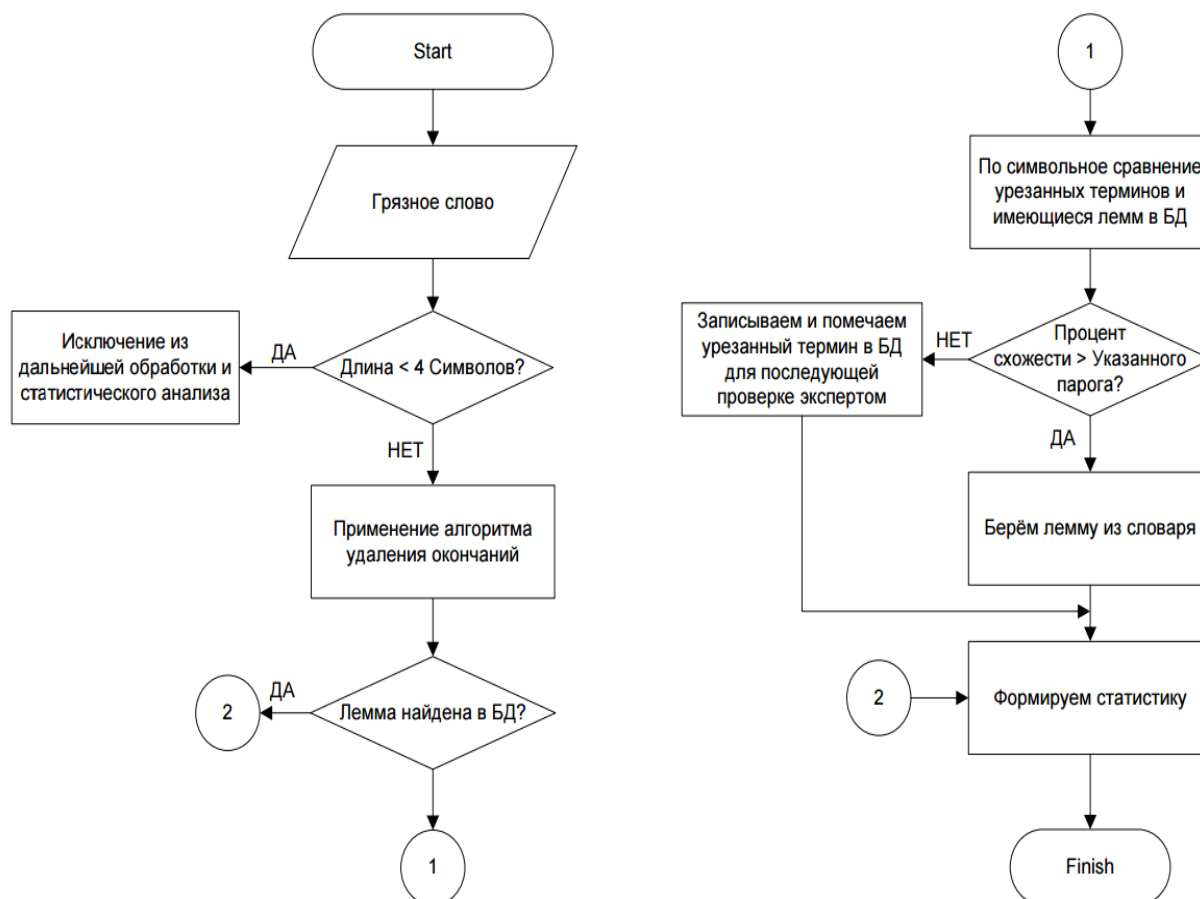


Рисунок 2 - Схема лингвистической обработки

В добавок алгоритм работает с множеством окончаний, разбитым на несколько подмножеств: окончания герундия («в», «вши», «вшись» - некоторые окончания деепричастий для русского языка или «ing» для английского языка), окончания прилагательных («ыми», «ми», «ие»), причастные суффиксы («ющ», «вш», «ивш»), возвратные постфиксы («сь», «ся»), глагольные окончания («ла», «ли», «ем»), префиксы («до», «за», «над») и окончания существительных.

Процесс стемминга легко реализуется с использованием БД, для чего достаточно воспользоваться специальным словарём масок (маски на *T-SQL* представляются набором символов, в которых должны присутствовать символы «%» и/или «_»), который может быть совмещён со списком специальных символов для фильтрации. Получившаяся в результате обработки усечённая форма терминов, проверяется по БД с помощью словаря лемм. Если лемма найдена, то оригинальный термин

привязывается к найденной лемме (Рисунок 2). Если урезанному термину невозможно было сопоставить лемму из словаря, то он посимвольно сравнивается с леммами из словаря. На последних шагах алгоритма в словарь лемм включаются урезанные термины, которым невозможно было сопоставить уже имеющиеся леммы. Они помечаются, чтобы лингвистический эксперт смог провести проверку и, при необходимости, поправить или назначить новые леммы. Такой алгоритм сам по себе универсален и весьма точен, однако и ошибки, как показала практика, случаются достаточно часто, например, при наличии у слов приставок, что может привести к получению несловарных корней либо к чрезмерному усечению слов. Однако эту проблему всегда может решить квалифицированный лингвистический эксперт.

В предложенном алгоритме достаточно воспользоваться двухуровневым словарём и для статистической обработки текстов и для формирования характеристических векторов. Если для обработки терминов потребуются другие промежуточные результаты стемминга, то алгоритм нужно будет перенастроить на систему словарей более высокого уровня (например, 3-х или 4-х уровневых словарей), добавляя в него блоки обработки дополнительного словаря.

Резюмируя всё сказанное выше, получим следующую последовательность действий, по преобразованию исходного запроса ИП/текста ИР в представление, пригодное для дальнейшей обработки алгоритмами кластерного анализа:

- а) выделение всех терминов из запроса ИП/текста ИР;
- б) стемминг и получение урезанных терминов после удаления окончаний;
- в) проверка урезанных терминов по словарю лемм БД. Если лемма найдена, переход к пункту д).
- г) сохранение помеченных урезанных терминов, для которых не определена лемма из БД. При необходимости лингвистический эксперт может подтвердить или изменить помеченные урезанные термины, превращая их в новые леммы;
- д) формирование статистики терминов и характеристических векторов;

После выполнения перечисленных пунктов, осуществляется переход к кластерному анализу лингвистически подготовленных запросов ИП и текстов ИР.

Анализ признаков заключается в изучении морфемных, морфологических и синтаксических признаков слов и предложений в тексте. Это изучение необходимо для того, чтобы дальше можно было строить модель (структуру) предложений и на её основании извлекать требуемую информацию.

Для построения такой структуры необходимо произвести несколько операций. Первая заключается в определении грамматических признаков всех слов. Для этого в языках, в которых существуют падежи, спряжения, склонения и времена, устанавливаются все эти признаки. Также определяется лемма каждого слова - его словарная форма. Для определения леммы слова используется алгоритм-лемматизатор, который либо с помощью сравнения со словарём, либо посредством последовательного отсечения окончаний и аффиксов (префиксов, суффиксов и постфиксов) и добавления нормализованного окончания выделяет основу слова. Вторая операция сводится к построению модели предложений в тексте. Посредством анализа грамматических признаков, найденных ранее, составляется зависимость слов друг от друга - сначала в пределах предложения, а затем и в пределах текста, если необходим более широкий анализ. Третья операция сводится к поиску искомой информации. Пользовательский запрос проходит лемматизацию, после чего происходит поиск необходимых лемм по всем предложениям. Однако то, что не представляет большого труда для лингвиста и просто человека, является большой проблемой для вычислительной системы. Описываемый выше анализ сложно чётко сформулировать, т.к., например, слово «истории» можно рассматривать не только как дательный падеж, но и как именительный/винительный падеж множественного числа слова «история». Отдельной препоной являются омонимы («ягуар» - это автомобильный бренд, энергетический напиток или хищник?), т.е. созвучные слова; если в тексте встречается несколько слов с одинаковыми леммами, нет гарантии того, что это - одно и то же слово. Для выяснения этого необходимо провести более сложный кластерный анализ для определения содержания текстов ИР в целом после получения характеристических векторов. Без кластерного анализа (или без применения методов классификации текстов) значений слов такие неоднозначности довольно сложно разрешить. Ещё одной значительной проблемой является разбор отсутствующих в словаре слов, когда нужно обращаться к лингвистическому эксперту, который в свою очередь вносит изменения в словарь лемм.

Чётко заданных правил для такого анализа не существует; их можно вывести экспериментальным путём, однако всё равно будут существовать исключения, отработать которые по правилам будет невозможно.

Семантический анализ занимается разбором текста относительно значения слов внутри него. Обычно этот вид анализа применяется после проведения грамматического анализа и дополняет его своими выводами. В частности, семантический анализ позволяет выявлять несвязность слов и предложений внутри текста, хотя они и могут быть согласованы грамматически. Также семантический анализ позволяет определять

метафоры, переносные значения, истинный смысл созвучных слов в зависимости от контекста, и т.д. К сожалению, серьёзных успехов в этой области пока достичь не удалось, т.к. этот вид анализа является наиболее сложным и наименее формализованным, хотя и самым востребованным. Простые способы семантического анализа позволяют классифицировать текст, выделять эмоциональную окраску текста (с помощью выявления определённых слов и анализа словосочетаний на предмет метафор и иносказаний) и его тему (по синтаксическим признакам и количеству повторяющихся слов в предложениях). В частности, с помощью семантического анализа происходит выдача контекстной рекламы на многих сайтах и в поисковых системах [12]. Страница, выдаваемая пользователю, исследуется на предмет наличия повторяющихся ключевых слов, после чего автоматизированный генератор рекламы выдаёт связанную с найденными ключевыми словами выборку.

Комбинированный подход подразумевает использование нескольких из вышеописанных подходов в связке, последовательной или параллельной, для повышения точности анализа. Чаще всего для сложного анализа текста применяют анализ признаков, совмещённый со статистическим анализом для ранжирования результата поиска и разрешения неоднозначностей [13-16]; реже используются вкрапления семантического анализа в любой из вышеописанных методов. В частности, такой подход используется в текстовых редакторах для выявления сложных ошибок (несогласованность текста, рекомендации по разбиению текста на абзацы, и т.д.).

Вывод. В данной работе проводится анализ основных методов классификации информации в Интернете. В зависимости от возникающих задач, требуется привлечение соответствующего метода. Предлагается использовать для повышения эффективности комбинированный подход, который использует для классификации несколько методов в связке.

ЛИТЕРАТУРА

1. Землянухина Н.С. О применении информационных технологий в менеджменте / Н.С.Землянухина // Успехи современного естествознания. 2012. № 6. С. 106-107.
2. Завьялов Д.В. О применении информационных технологий / Д.В.Завьялов // Современные наукоемкие технологии. 2013. № 8-1. С. 71-72.
3. Львович И.Я. Применение методологического анализа в исследовании безопасности / И.Я.Львович, А.А.Воронов // Информация и безопасность. 2011. Т. 14. № 3. С. 469-470.
4. Ермолова В.В. Архитектура системы обмена сообщений в немаршрутизируемой сети / В.В.Ермолова, Ю.П.Преображенский //

- Вестник Воронежского института высоких технологий. 2010. № 7. С. 79-81.
5. Свиридов В.И. Технологии, применяемые при подготовке современных инженеров / В.И.Свиридов // Вестник Воронежского института высоких технологий. 2012. № 9. С. 151-152.
 6. Зяблов Е.Л. Построение объектно-семантической модели системы управления / Е.Л.Зяблов, Ю.П.Преображенский // Вестник Воронежского института высоких технологий. 2008. № 3. С. 029-030.
 7. Исакова М.В. Об особенностях систем управления персоналом / М.В.Исакова, О.Н.Горбенко // Вестник Воронежского института высоких технологий. 2014. № 12. С. 168-171.
 8. Фомина Ю.А. Принципы индексации информации в поисковых системах / Ю.А.Фомина, Ю.П.Преображенский // Вестник Воронежского института высоких технологий. 2010. № 7. С. 98-100.
 9. Аттетков, А.В. Методы оптимизации: Учебное пособие / А.В. Аттетков, В.С. Зарубин, А.Н. Канатников. - М.: ИЦ РИОР, НИЦ ИНФРА-М, 2013. - 270 с.
 10. Чандра, А.М. Дистанционное зондирование и географические информационные системы / А.М. Чандра, С.К. Гош; Пер. с англ. А.В. Кирюшин. - М.: Техносфера, 2008. - 312 с.
 11. Шушкевич, Г.Ч. Компьютерные технологии в математике. Система Mathcad 14. В 2-х т.Т. 1. Компьютерные технологии в математике. Система Mathcad 14 / Г.Ч. Шушкевич. - Мн.: Изд.-во Гревцова, 2010. - 288 с.
 12. Уськов, В.В. Компьютерные технологии в подготовке и управлении строительством объектов: Учебно-практическое пособие / В.В. Уськов. - М.: Инфра-Инженерия, 2011. - 320 с.
 13. Паневин Р.Ю. Реализация транслятора имитационно-семантического моделирования / Р.Ю.Паневин, Ю.П.Преображенский // Вестник Воронежского института высоких технологий. 2009. № 5. С. 057-060.
 14. Преображенский Ю.П. Алгоритм нахождения оптимальной стационарной стратегии для марковских процессов принятия решений / Ю.П.Преображенский // Вестник Воронежского института высоких технологий. 2010. № 6. С. 81-82.
 15. Львович Я.Е. Адаптивное управление марковскими процессами в конфликтной ситуации / Я.Е.Львович, Ю.П.Преображенский, Р.Ю.Паневин // Вестник Воронежского государственного технического университета. 2008. Т. 4. № 11. С. 170-171.
 16. Преображенский Ю.П. Адаптивные алгоритмы для бесконечных стохастических игр / Ю.П.Преображенский // Вестник Воронежского института высоких технологий. 2010. № 7. С. 46-47.

Man Jinan

THE ANALYSIS OF METHODS OF INFORMATION QUALIFICATION IN INTERNET DURING SOLVING PROBLEMS OF INFORMATION SEARCH

Voronezh Institute of High Technologies

Due to the fact that the Internet stores a large amount of information must be used effective methods of search. The characteristics inherent in job search, are the completeness, reliability and high speed. These characteristics can be achieved with the use of appropriate methods of classification. The article considers several approaches. The cluster method is based on partitioning elements of the set into groups. The distance between the elements is set to metric. Linguistic analysis based on the ability to extract information from the text. The scheme of linguistic processing. Statistical approaches proceed from the specific patterns frequency of the meeting words. Symptom analysis is the study of the morphemic, morphological and syntactic features of words and sentences in the text. Semantic analysis is the analysis of the text concerning the meaning of the words inside it. The combined approach involves the use of several of the above approaches in tandem, series, or parallel, to increase the accuracy of the analysis. Depending on the emerging challenges, will be useful for the appropriate method of classification.

Keywords: information search, Internet, statistics, classification, semantics, sign, metric, the word.

REFERENCES

1. Zemlyanukhina N.S. O primeneniі informatsionnykh tekhnologiy v menedzhmente / N.S.Zemlyanukhina // Uspekhi sovremennogo estestvoznaniya. 2012. № 6. S. 106-107.
2. Zav'yalov D.V. O primeneniі informatsionnykh tekhnologiy / D.V.Zav'yalov // Sovremennye naukoemkie tekhnologii. 2013. № 8-1. S. 71-72.
3. L'vovich I.Ya. Primenenie metodologicheskogo analiza v issledovanii bezopasnosti / I.Ya.L'vovich, A.A.Voronov // Informatsiya i bezopasnost'. 2011. T. 14. № 3. S. 469-470.
4. Ermolova V.V. Arkhitektura sistemy obmena soobshcheniy v nemarshrutiziruемой seti / V.V.Ermolova, Yu.P.Preobrazhenskiy // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2010. № 7. S. 79-81.
5. Sviridov V.I. Tekhnologii, primenyaemye pri podgotovke sovremennykh inzhenerov / V.I.Sviridov // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2012. № 9. S. 151-152.
6. Zyablov E.L. Postroenie ob"ektно-semanticheskoy modeli sistemy upravleniya / E.L.Zyablov, Yu.P.Preobrazhenskiy // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2008. № 3. S. 029-030.

7. Isakova M.V. Ob osobennostyakh sistem upravleniya personalom / M.V.Isakova, O.N.Gorbenko // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2014. № 12. S. 168-171.
8. Fomina Yu.A. Printsipy indeksatsii informatsii v poiskovykh sistemakh / Yu.A.Fomina, Yu.P.Preobrazhenskiy // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2010. № 7. S. 98-100.
9. Attetkov, A.V. Metody optimizatsii: Uchebnoe posobie / A.V. Attetkov, V.S. Zarubin, A.N. Kanatnikov. - M.: ITs RIOR, NITs INFRA-M, 2013. - 270 c.
10. Chandra, A.M. Distantionnoe zondirovanie i geograficheskie informatsionnye sistemy / A.M. Chandra, S.K. Gosh; Per. s angl. A.V. Kiryushin. - M.: Tekhnosfera, 2008. - 312 c.
11. Shushkevich, G.Ch. Komp'yuternye tekhnologii v matematike. Sistema Mathcad 14. V 2-kh t.T. 1. Komp'yuternye tekhnologii v matematike. Sistema Mathcad 14 / G.Ch. Shushkevich. - Mn.: Izd.-vo Grevtsova, 2010. - 288 c.
12. Us'kov, V.V. Komp'yuternye tekhnologii v podgotovke i upravlenii stroitel'stvom ob"ektov: Uchebno-prakticheskoe posobie / V.V. Us'kov. - M.: Infra-Inzheneriya, 2011. - 320 c.
13. Panevin R.Yu. Realizatsiya translyatora imitatsionno-semanticheskogo modelirovaniya / R.Yu.Panevin, Yu.P.Preobrazhenskiy // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2009. № 5. S. 057-060.
14. Preobrazhenskiy Yu.P. Algoritm nakhozheniya optimal'noy statsionarnoy strategii dlya markovskikh protsessov prinyatiya resheniy / Yu.P.Preobrazhenskiy // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2010. № 6. S. 81-82.
15. L'vovich Ya.E. Adaptivnoe upravlenie markovskimi protsessami v konfliktnoy situatsii / Ya.E.L'vovich, Yu.P.Preobrazhenskiy, R.Yu.Panevin // Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta. 2008. T. 4. № 11. S. 170-171.
16. Preobrazhenskiy Yu.P. Adaptivnye algoritmy dlya beskonechnykh stokhasticheskikh igr / Yu.P.Preobrazhenskiy // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2010. № 7. S. 46-47.