

ность определения параметров каждого бревна в отдельности. В настоящее время измерение экспортного леса производится только контактными методами, погрешность которых не удовлетворяет нормам стандарта. Рассмотрев описанные выше бесконтактные методы, можно сделать вывод, что наиболее перспективным для измерения объема бревна без учета коры является развитие фотометрических методов, ввиду простоты процесса измерения, точности определения границ коры дерева и несложных в применении устройств.

Основываясь на предложенной классификации можно сделать вывод, что в настоящее время появились два новых направления измерения круглого леса с использованием оптических методов измерения и фотометрических методов измерения. Первое направление применяется для измерения объема круглого лесоматериала с учетом коры и в настоящее время глубоко теоретически и практически проработано [4]. Область применения второго направления – измерения объема делового лесоматериала. Данное направление находится на стадии развития и в этом направлении необходимо провести теоретические и практические исследования.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Шегельман И.Р., Быков Е.Н. Поштучный учет и приемка лесоматериалов. Пороки и дефекты древесины. – С-Пб: «Профикс», 2006. – 25 с.
2. Справочник по круглым лесоматериалам. – М: Лесэксперт, 2003. – 112 с.
3. Самойлов А.Н. Анализ существующих методов измерения лесоматериалов // МНТК «Интеллектуальные системы» (AIS'06) и «Интеллектуальные САПР» (CAD-2006): Научное издание в 3- томах. – М.: Физматлит, 2006, т. 2. – 334 с.
4. «Объем имеет значение». Российская лесная газета, выпуск № 43 (121) от 12.11.2005.

Р.Ю. Вишняков

ОБ ОДНОМ ПОДХОДЕ К ИНТЕЛЛЕКТУАЛИЗАЦИИ ИНФОРМАЦИОННО-ПОИСКОВЫХ СИСТЕМ

Введение. Задачи информационного поиска. В настоящее время создание интеллектуальной информационно-поисковой системы является одной из актуальных проблем в информационных технологиях. Информационно-поисковые системы (ИПС) появились задолго до компьютеров и более известны как методы по учету, контролю и поиску информации. Они вышли на качественно новый уровень с появлением компьютерных систем, которые стали применяться в различных областях человеческой деятельности. Новые возможности и перспективы, вносимые увеличением объема поисковой информации и увеличивающимися возможностями компьютерной техники, породили спрос на интеллектуализацию ИПС.

Поскольку главная задача поиска – быстрое и точное нахождение необходимой информации, то одной из основных проблем, возникающих при работе с большими объемами документов (при анализе содержания), является проблема минимизации времени. В принципе очевидна прямая зависимость времени поиска от объема информации, содержащегося в документе, и количества документов. Актуальной становится задача автоматического анализа документов с целью выявления смысловых единиц и структур, в сжатом виде описывающих текстовую информацию (материал). Другими словами, задача построения универсального индекс-дескриптора (или репозитория образов), в котором образ выражает, так называемую, «основную мысль» обработанного документа, актуальна.

Понимая то, что ИПС с систематизированной структурой данных требует значительно меньшего времени на поиск информации, тем не менее, система не в состоянии до конца определить соответствие найденной информации запросу и требует дополнительных мер по увеличению эффективности поиска.

Процесс поиска текстовой информации. Он может быть представлен следующими этапами:

1. Формализация запроса.
2. Формирование предварительного перечня найденной информации;
3. Анализ отобранного спектра документов.
4. Оценка релевантности.

Этап формализации запроса является важным этапом в работе с ИПС и является результатом интеллектуальной деятельности человека. Важно понимать, что полностью могут быть формализованы лишь элементарные области с простой логической структурой и небольшим запасом понятий (например, исчисление высказываний и узкое исчисление предикатов – в логике, элементарная геометрия – в математике). Если же область запроса сложна, ее формализация усложняется и принципиально не может быть полной. Таким образом, одна из проблем увеличения эффективности работы с информационно-поисковой системой (оцениваемой по релевантности возвращаемых результатов), напрямую зависит от интеллектуального уровня человека, формирующего запросы к ИПС.

Этап формирования предварительного перечня найденных документов, содержащих формальные признаки присутствия искомой информации, отвечает за сведение понятий запроса (в области значений). Важно понимать, что этап формирования предварительного перечня зависит от вычислительных мощностей, объема информации, количества и качества агентов-индексирования, а также используемых при реализации ИПС алгоритмов. В таком случае, при обработке поисковых запросов в глобальной сети целесообразнее формировать предварительный перечень документов, пользуясь статистико-вероятностными методами. Такие базовые математические понятия, как статистика и вероятность не требуют дополнительных комментариев о применимости данного подхода к глобальному поиску. Важно помнить, что важнейшими условиями получения достоверного результата являются: (1) независимость конкретных состояний объекта от его предыдущих состояний; (2) однородность сопоставляемых (включенных в выборку) элементов наблюдения.

Лингвистический анализ отобранного спектра документов разделим на следующие уровни (в порядке возрастания сложности анализа):

- ◆ Лексический анализ - заключается в разборе текстовой информации на отдельные абзацы, предложения, слова, определении языка изложения (например, позволяет определить тематическую направленность), типа предложения, выявлении типа лексических выражений и т.д. Но сам по себе лексический анализ неоднороден.
- ◆ Морфологический анализ - сводится к распознаванию морфем, словоформ и частей речи.
- ◆ Синтаксический анализ - заключается в выделении семантических элементов предложения (именных групп, терминологических целых, предикативных основ).
- ◆ Семантический анализ - заключается в определении информативности текстовой информации и выделении информационно-логической основы текста. Семантический анализ подразумевает выявление и оценку смыслового содержания текста.

Этап оценки релевантности полученных результатов заключается в осмыслении (пользователем) найденной информации и оценки этой информации на соответствие предлагаемой информации запросу. Оценка релевантности, как и этап формализации запроса, является продуктом интеллектуальной деятельности человека.

Глобальный и специализированный поиск. Google, Yahoo!, Yandex, Rambler и пр. для более качественного отбора текстовых ресурсов, соответствующих запросу пользователя, используют приемы до-аналитического назначения релевантности ресурсу (индексы цитирования, частота встречаемости ключевых слов, обратные связи и т.п.). Лингвистические методы обработки информации в глобальных поисковых системах представлены в виде морфологических характеристик лексических единиц. Также применяется поверхностный синтаксический анализ запроса пользователя для выделения ключевых слов и поиска ответа по ним. Достаточно неглубокий анализ текста позволяет находить верный ответ с ощутимо большей вероятностью. Однако данный уровень синтаксического анализа определяет недостаточную точность поиска в специализированных системах ИПС. Для достижения большей точности необходим углубленный синтаксический анализ предложений в сочетании с семантическим анализом обрабатываемых текстов. Специализированные системы ИПС оперируют с меньшим объемом индексируемой информации, по большей части текстовая информация в таких областях является семантически однородной. В таком случае, очевидно, что для поиска точного ответа в подобных системах вероятностно-статистические методы не подходят. Учитывая прикладную направленность данных таких поисковых систем, наиболее актуальным является применение алгоритмов поиска смысловой информации. Качественный поиск в подобных специализированных областях может быть реализован на основе многоуровневого лексико-грамматического анализа содержания области данных и запросов пользователей к ИПС, а также широкого привлечения различных методов оценки их смыслового соответствия.

Поисковые системы. Практически во всех существующих системах ИПС оценка смыслового содержания найденной информации возлагается на пользователя. Основываясь на этом, разделим роли процесса поиска на два этапа.

1. ИПС осуществляет поиск ресурсов в соответствии с запросом.
2. Пользователь осуществляет процесс поиска смысловой информации в перечне выдаваемых системой результатов.

Данные процессы можно проклассифицировать как отдельные задачи поиска, которые определяют то, что ищет пользователь. Поиск текстовых ресурсов характерен для процесса поиска пользователем информации вслепую, касающейся искомой им тематики. Этот частный случай поиска возникает в случае, если пользователь не знаком с предметом или предметной областью поиска. Как следствие, пользователь не в состоянии формализовать запрос или не находит искомой информации среди предлагаемой ИПС и пытается найти зацепки по интересующей тематике. Поиск смысловой информации, как правило, проводится пользователем, знакомым с прикладной областью и уверенным, что искомая информация может содержаться в базе индекс-дескриптора.

Заключение. Эффективность работы систем ИПС напрямую зависит от объема, содержания и семантической однородности информации, а также от интеллектуального уровня пользователя и его умения и/или возможности формализовать запросы на интересующие его темы. Улучшение результативности поиска можно добиваться с введением некоторых новых интеллектуальных свойств систем ИПС, позволяющих накапливать и впоследствии анализировать информацию о конкрет-

ном пользователе, обучая и подстраивая систему под конкретного человека. Стоит также обратить внимание на обучение пользователей работе с информационно-поисковыми системами, что позволит говорить о переходе поиска на качественно новый уровень. Семантический анализ является трудно формализуемым. Реализация семантического анализа текстовой информации предполагает обязательное использование экспертных систем, систем искусственного интеллекта для выявления смыслового содержания информации. Хотя суть и принципы построения ИПС достаточно прозрачны и открыты широкому кругу специалистов, но до сих пор не создано универсальных подходов к построению интеллектуальных информационно-поисковых систем, не найден консенсус по адекватному и эффективному инструментарию.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Черный А.И.* Введение в теорию информационного поиска. – М.: Наука, 1975.
2. *Налимов В.В.* Вероятностная модель языка // О соотношении естественных и искусственных языков. – М.: Наука, 1979.
3. *Рубашкин В.Ш.* Представление и анализ смысла в интеллектуальных информационных системах. – М.: Наука, 1989.
4. *Совпель И.В.* Инженерно-лингвистические принципы, методы и алгоритмы автоматической переработки текста. – Минск, 1991.
5. *Вишняков Ю.М., Вишняков Р.Ю.* Проблемы семантического информационного поиска // Труды международных научно-технических конференций «Интеллектуальные системы» (AIS'06) и «Интеллектуальные САПР» (CAD-2006). Научное издание в 3-х томах. – М.: ФИЗМАТЛИТ, 2006, Т.2. – С. 308-314.
6. *Андреев А.М., Березкин Д.В., Брик А.В.* Лингвистический процессор для информационно-поисковой системы // Компьютерная хроника, 1998. № 11. – С. 79-100.
7. *Тузов В.А.* Семантический анализатор текстов на русском языке. // Информационные технологии в гуманитарных и общественных науках. Вып. 9: Семантико-синтаксический анализ текстов. – СПб.: СПб Экономико-математический институт РАН, 2000.
8. *Bobrow D.G.* Natural Language Input for a Computer Problem Solving System, Semantic Information Processing, Cambridge, M.I.T., 1968.
9. *Попов Э.В.* Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ. – М., 1987.