

В.В. Александров, С.В. Кулешов

АНАЛИТИЧЕСКИЙ МОНИТОРИНГ INTERNET КОНТЕНТА. ИНФОЛОГИЧЕСКИЙ ПОДХОД

Рассматривается новое направление представления и обработки информации в Internet, основанное на инфологическом подходе. Производится сравнение истории создания энциклопедии в XVIII веке с современным состоянием развития информационных систем. Предлагается подход аналитического мониторинга Internet контента.

V.V. Alexandrov, S.V. Kuleshov

ANALYTICAL MONITORING OF INTERNET CONTENT. INFO LOGICAL APPROACH

A new direction of Internet information representation and processing based on info logical approach is considered. A comparison of encyclopedia creation history in XVIII century with modern state of information systems is shown. An approach of analytical monitoring of Internet content is proposed.

*Если текст состоит из метафор,
это загадка, а если из глосс – варваризм.
Аристотель «Поэтика»*

Общеизвестен исторический факт, что библиотеки возникали при храмах, монастырях, церквях – в местах коммуникативного общения. Библиотеки, в свою очередь, порождали образовательный процесс, отсюда берут свое начало Кембридж, Оксфорд.

Проведем исторические параллели между возникновением энциклопедии как формы представления знания и современными Internet-поисковыми системами.

В 1728 г. в Англии Эфраим Чамберс впервые создает «Энциклопедию, или Всеобщий словарь искусств и наук». По сути же, эта энциклопедия объединяла в себе лишь разрозненные фактографические сведения.

Но уже через некоторое время такая энциклопедия перестала удовлетворять практические потребности: *...Какой прогресс совершили с тех пор науки и искусства! Сколько ныне открыто истин, которые тогда и не снились! Истинная философия была в колыбели; геометрия бесконечного еще не существовала; экспериментальная физика едва появилась; диалектики не было вовсе; отсутствовал критический подход к источникам. Декарт, Бойль, Гюйгенс, Ньютон, Лейбниц, Бернулли, Локк, Бейль, Паскаль, Корнель, Расин и другие либо еще не существовали на свете, либо еще не писали. Дух исследования и соревнования не воодушевлял ученых; дух точности и метода еще не подчинил себе различные виды описаний...* [1].

Научные открытия и достижения требуют систематизации и отражения в литературе, чтобы познать с ними широкий круг читающей публики.

И вот, через 40 лет после Чамберса, энциклопедисты Даламбер, Дидро, Руссо на основании новых накопленных знаний решают создать новую энциклопедию «Энциклопедия, или Толковый словарь наук, искусств и ремесел». **Заметим существенную разницу в терминах, поясняющих предназначение энциклопедии.** В этом названии вместо понятия «всеобщий» используется понятие «толковый», а также виден инновационный характер технологий представления и распространения знаний.

Сегодня отчетливо ясно, что совместный прогресс телекоммуникационных каналов связи и компьютерная обработка, систематизация и представление информации дают возможность создать новый научный подход, а также эффективный инструмент систематизации знаний, который был сформулирован в [2, 3], а его прототип реализован в виде Интернет-проекта VisualWorld.ru.

Для своего времени «Энциклопедия» блестяще разрешила эту задачу – она дала синтез идей и знаний XVIII века.

Приведем дискуссию [1] Даламбера, Дидро, Руссо – создателей первой энциклопедии знаний:

– *Общеизвестно, что за короткое время словарь Чамбреса выдержал в Англии пять изданий. Как вы думаете, почему? Да потому, что в своем роде он уникален. Хотя в разное время и было сделано множество попыток создать нечто в подобном роде, но до хитроумного англичанина никто в этом не преуспел...*

– *Разве «Исторический и критический словарь» Бейля, опубликованный еще в конце прошлого века, уступает «Энциклопедии» Чамберса?*

— Вы бы вспомнили еще об Альтиштедте или Роджере Бэконе с его «Компендиумом»... Нет, Пьеру Бейлю и его предшественникам далеко до Чамберса. Бейль ведь не хотел, да и не мог создать действительно универсальный труд, основанный на единстве наук. Он преследовал гораздо более узкую цель — критический разбор религиозных и философских учений различных эпох. К тому же, выступая против религиозного ханжества, сражаясь с метафизиками и богословами, он не мог опереться на опыт, он не знал естественных наук...

— Дидро: «...Наша цель состоит в том, чтобы рассмотреть различные предметы не только в отдельности, но и во взаимной их связи, в том, чтобы рассмотреть каждый из них как нечто целое и как часть еще большего целого».

— Чамберс впервые попытался осуществить идею «Нового органа» великого Ф.Бэкона и дать стройный свод человеческих знаний, потому что Чамберс сделал больше, чем кто-либо из его предшественников, потому, наконец, что дальше речь пойдет о переводе двухтомника Чамберса на французский язык...

Их дискуссия **не отличается** от современных обсуждений в Сети альтернативных поисковых сервисов (s.liveinternet.ru, webalta.ru, visualworld.ru, povoteka.ru), которые могут не просто находить, а еще систематизировать и более вдумчиво отбирать результаты. Некоторые из них предлагают и различные варианты визуализации процесса поиска. Так поисковая система VisualWorld добавила к своему обычному поиску ассоциативный [4]. При выборе этой опции поисковик подбирает ассоциации к введенному запросу и осуществляет поиск с их учётом. Последовательно выбирая варианты, пользователь может сконструировать наиболее точный поисковый запрос.

Современные телекоммуникационные каналы и компьютер привели к новой информационной технологии представления и работы со знаниями. Возникает вопрос: где «современные» Даламбер, Дидро, Руссо, которые систематизируют огромный современный информационный поток?

Рост количества информационных ресурсов Internet-пространства уже достиг критического уровня эффективного функционирования. Увеличение количества информационных ресурсов происходит за счет дублирования, а также за счет примитивной компиляции. Это естественный необратимый процесс, развивающийся за счет новых публикаций, а также неотъемлемого элемента «журнализации событий» современного общества — новостей. Подобное экстенсивное увеличение количества доступных ресурсов создает большую нагрузку на поисковые

машины, которые должны адекватно отвечать на запросы пользователей и выдавать адекватные, актуальные и достоверные данные.

При этом возникают те же проблемы, что и во времена создания Энциклопедии — какие данные и сведения являются фактографической справкой, какие формируют, выявляют и систематизируют знания (т.е. участвуют в инновационном процессе), а какие являются просто мусором. При этом огромная скорость поступления, обработки и накопления данных ограничивается медленным процессом чтения текста с экрана.

Классические поисковые системы направлены на инструмент (компьютер), оптимизированы под его устройство, чтобы эффективно справляться с огромным потоком данных. Соответственно, они представляют данные для компьютера, а не для человека.

Инфологический механизм, реализуемый сервисом VisualWorld.ru, опирается на потребности человека, исходя из задач систематизации знаний в той форме, которая удобна человеку. Проект VisualWorld.ru представляет развитый интерактивный интерфейс для получения «выборки знаний» из актуального набора документов, доступных в Internet.

Принцип работы системы основан на инфологическом подходе к исследованию текущего состояния информационных ресурсов (Интернет, базы данных и др.) и его востребованности различными социальными группами. Предполагается, что собранная информация поможет при решении существующих проблем и позволит предотвратить информационный хаос.

Рассматривается визуально-динамическое представление данных как итерационный процесс $I = i(D, S, T)$ последовательного постепенного формирования знаний, где обработка данных D превращается в информацию I на заданном интервале времени T , формирующую онтологические знания S в виде антологий по проблемно-предметным областям, i — процесс их интерпретаций — топика. Каждая следующая итерация повышает уровень знаний, т.е. порождает новую порцию информации, организовывая инфологию как самовоспроизводимый ресурс развития знаний.

Можно утверждать, что образ или событие любой природы имеют некоторую собственную семантику. Эта собственная семантика универсальна и контекстно-интерпретируема в некотором конечном наборе смыслов или событий. Возможность осмысления практически любого образа или события внешнего мира не зависит от специфической конструкции языков: синтаксиса, лексики и прагматики. Вот почему

постоянно воспроизводится достаточно малое число повторяющихся мифов и сюжетов, отражающих социальные проблемы, будущее цивилизации и научно-технического прогресса вне зависимости от пространства сред обитаний, времени и языковых средств.

Когда что-либо обозначается – это означает, что смысл понят. Что мы не переходим от звуков к образам и от образов к смыслу: мы с самого начала помещены в смысл. Смысл предписывает возможные обозначения и условия и, в свою очередь, является объектом следующего предложения. Тогда, если принять предложение за некое имя, ясно, что каждое имя, обозначающее объект, само может стать объектом нового имени, обозначающего его смысл: n1 отсылает к n2, которое обозначает смысл n1; n2 отсылает к n3, и так далее. Для каждого из своих имён язык должен содержать некоторое имя для смысла этого имени. Такое бесконечное размножение вербальных сущностей известно как *парадокс Фреге*.

Следовательно, сами по себе, как *самостоятельные субстанции*, не существуют у человека смысл, память, знание, познание. Это просто имена (метки), синонимы, ассоциативно отражающие *продукт функционирования Мозга*, как инструмент (процессор) этерификаций.

Понятие этерификации лежит в основе технического и цивилизационного прогресса. Согласно Дж. Тойнби «...в истории письменности наблюдается не только соответствие между развитием техники письма и упрощением формы, но эти две тенденции фактически тождественны друг другу, поскольку вся техническая проблема, которую должно решить письмо как фиксатор, посредник человеческой речи, – это отчётливая репрезентация широчайшей сферы человеческого языка с максимальной экономией визуальных символов, т.е. **этерификация есть закон прогрессирующего упрощения...**».

«...*Думай о смысле, слова придут сами...*» – эта классическая фраза Льюиса Кэрролла по мере ее осознания становится все интереснее и важнее. Так, в исследованиях типа искусственного интеллекта сегодня важным становится не нахождение формального представления контекстно-независимого языка, а исследование логики смысла.

В отличие от большинства современных систем, основанных на использовании тематических и индексных каталогов, предлагаемая система основана на поисковой информационной системе, объединяющей систему семантического поиска в сетях (базах данных), поиск по нормативным документам и систему мониторинга информационных ресурсов.

Основная особенность системы состоит в организации поиска документов, удовлетворяющих условию наличия семантических связей в документе между всеми словами поискового запроса. Система находит страницы не с точным вхождением строки запроса, а отвечает на вопрос «что есть то, что написано в запросе?».

Отличия предлагаемой системы от существующих:

- Эта система использует более углубленные понятийные ассоциативные связи, что позволяет при аналитическом реферировании сужать область запроса и поиска документа.
- Наличие ассоциативного поиска, позволяющего осуществлять доступ к документам пользователям, не являющимся специалистами в данной предметной области (не владеющими полной терминологией предметной области).
- Формирование аналитических данных по имеющимся документам, а не просто представление и поиск документов.
- Возможность пополнения базы документов в автоматическом режиме сразу после их опубликования в электронном виде.
- Использование семантико-смысловых «ключей» при поиске документа.

ЛИТЕРАТУРА

1. Левандовский А.П. Наследники господина Чамберса. Повесть. – М.: «Дет.лит.», 1977. 239 с.
2. Александров В.В., Сарычев В.А. ИНФОТВАРЬ – НОМО INFOCREATURE // Системные проблемы надёжности, качества информационных и электронных технологий в инновационных проектах (Инноватика – 2006). / Материалы Международной конференции и Российской научной школы. Часть 3. Т. 1. – М.: Радио и связь, 2006. С. 3 – 11.
3. Александров В.В., Андреева Н.А., Кулешов С.В. Системное моделирование. Методы построения информационно-логистических систем: Учеб. пособие. – СПб.: Изд-во Политехн. ун-та, 2006. 95 с.
4. Александров В.В. Поиграем в ассоциации? – газета «Компьютер - Информ» 23.04.2007, №8.

Александров В.В.
Кулешов С.В.

СПИИРАН, alexandr@iias.spb.su