

А. МАЦИОЛ, Б. РЕМБЯШ, Е. ДУДА
(Горно-металлургическая академия, г. Краков, Польша)

УПРАВЛЕНИЕ ИНФОРМАЦИЕЙ В СИСТЕМАХ ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ

Наукові праці МАУП, 2002, вип. 3, с. 55–57

1. Введение

В эпоху всеобщего использования информатики и дальней связи в управлении понятие инноватики приобрело новое значение. Появились возможности прямой поддержки этой концепции компьютерными инструментами. Инновационное управление с поддержкой компьютерными инструментами сейчас называют “управление знаниями”.

Основной проблемой, связанной с приложениями данной концепции, является отсутствие достаточно эффективных информатических инструментов для приобретения знаний. Общеизвестно, что единственной формой хранения и предоставления информации, которая дает возможность их прямого применения, являются базы данных. К сожалению, в информационных ресурсах организации — как внутренних, так и внешних — встречаем чаще всего источники, которые носят характер свободного текста либо более или менее структурированных документов. Поиск информации в таких источниках, даже при использовании самых изысканных инструментов, не отвечает всем требованиям пользователей.

2. Управление знаниями

Все чаще и больше считают, что конкурентоспособные преимущества организации будут определяться (возможно, уже определяются) накопленными в ней знаниями. Однако не совсем ясно, что следует понимать под понятием “знания организации”. Т. Давенпорт и Л. Прусак толкуют знания как плавное сочетание опыта, оценок ценностей, информации о контексте, а также анали-

тического взгляда на вопросы, и это сочетание обеспечивает возможность оценивания и включения новых опытов и информации [2]. Эти исследователи также констатируют, что знания организации характерны для человеческих умов.

Такое толкование обходит самую существенную черту сведений (констатаций, информации, данных), составляющих знания, накопленные человеческим умом, которые мы бы хотели превратить в полезные для организации ресурсы [4]. Этой чертой является порядок, который искали философы, занимающиеся метафизикой, или, как мы предпочитаем говорить сейчас, — онтологией, в задачу которой входит простое описание действительности, т. е. всего существующего. Традиционной целью онтологии является разделение сведений о мире (бытии) на категории как на составляющие. Онтологический анализ необходим при попытках формирования образцов, применяемых для описания созданных человеком объектов, позволяющих толковать термины и их логическую взаимосвязь. Главное здесь состоит в точном определении выражений и сочетаний между ними, которые как единственные могут быть использованы для описания систем в виде однозначных определений и аксиом. Поэтому здесь речь идет об онтологиях как однозначных описаниях определенных отраслей знаний, содержащих однозначное описание феноменов, т. е. явлений, которые являются объектом восприятия. Условием для преобразования приведенных онтологией знаний в ресурсы, полезные для организации, является обладание тех-

ническими возможностями их записи, распространения и применения.

Накопленный к нашему времени опыт указал на области, в которых эффекты управления знаниями самые заметные: вспомоществование групповой работы, регистрирование самых хороших опытов (best practices), образование компетенционных центров, подписка на знания (крупные консалтинговые фирмы все чаще предлагают для продажи результаты своих анализов посредством Интернета), новые формы управления человеческими ресурсами.

3. Информационные источники знаний

Знания, необходимые для реализации инновационных концепций, можно собирать, опираясь на конкретные данные, однако главным их источником являются сведения, сообщаемые в разговорной форме либо представленные в форме стандартов (образцовых документов). В этой формулировке особое внимание следует обратить на четкое различие между информацией, составляющей сведения, полезные с точки зрения организации, но не обладающие четко определенной структурой, и данными, которые относятся к конкретным объектам и описывают его однозначно выделенную черту [5]. Отсюда вытекает необходимость конструирования диаметрально отличающихся друг от друга инструментов для управления информацией и управления данными.

Процессом, позволяющим преобразовывать информацию с целью приобретения знаний, является автоматический анализ текста (индексирование). Анализ текста должен привести к его индексации, т. е. к определению ценности указанных заранее (либо созданных во время анализа) атрибутов, представляющих собой отличительное свойство документа, позволяющее найти его в пакете всех индексированных источников. Индексирование можно вести вручную или автоматически. Уже с 50-х годов XX в. велись работы по созданию искусственных систем индексирования [1]. В основном можно выделить два способа решения вопроса автоматического индексирования: 1) независимый от возможных вопросов, 2) направленный на специфическую домену применений. В первом случае основой индексирования принимаются частотность и место применения слов в анализируемом тексте. Дополнительно принимается необходимость "очистки слов", т. е. выделения их основных форм или устранения префиксов. Принимается также, что слова, применяемые в документе очень часто или очень редко, следу-

ет отбросить как повсеместные или как мало-значачие. В литературе описаны многие алгоритмы, основанные на разных мерах оценки существенности слов в документе, которые в результате ведут к автоматическому определению ключевых слов и их возможного веса, т. е. представляющих собой отличительные свойства (индексы) документа [5]. Второй способ индексирования документов, или, скорее, различения их содержания, опирается на методы преобразования натурального языка (Natural Language Processing — NLP). Они представляют собой набор инструментов, искусственно созданных и позволяющих преобразовывать разговорную речь в однозначные образцы [1]. Отсюда видно, что данный метод оправдан в тех случаях, когда можно истолковать образцы формулировок, т. е. когда текст анализируется с точки зрения его семантического соответствия с определенной темой — доменой. Результатом автоматического лингвистического анализа является квалификация предложений к отдельным образцам, а затем описание всего документа с помощью этих образцов.

Основная аппликационная проблема связана с приданием документам однородной электронной формы, позволяющей применить автоматическую обработку компьютерными программами. Существуют, правда, программы обработки информации, действующие в любых форматах записи документов, однако самым популярным решением, позволяющим не только привести к единой форме разного рода источники, но также подготовить и облегчить процедуру их дальнейшего анализа, является XML (Extensible Markup Language). Этот язык служит для применения в более общей концепции SGML (Standard Generalized Markup Language) для публикации в Интернете [3]. Стандарт SGML как международный, принятый в 1986 г. (ISO 8879), касается структуризации электронных документов. Слово "markup" является техническим термином, связанным с издательской деятельностью и обозначающим процесс маркирования копии рукописи, которая предназначена для набора с указанием применения видов и размеров шрифта, интервалов, абзацев и т. п. Однако относительно SGML их следует понимать скорее как маркировку, ведущую к приданию документу определенной не только типографической, но в основном логической структуры. Стандарт SGML, вопреки тому, что следует из его названия, не представляет собой язык маркировки, а служит, скорее, метаязыком, применяемым для толкова-

ния языков маркировки, например: HTML или XML. Язык XML также не предлагает готовых наборов маркеров, но является языковой структурой, позволяющей свободно формулировать наборы таких маркеров. Благодаря этому он значительно лучше, чем ограниченный HTML, пригоден для записи документов, которые должны подвергнуться процессу автоматической индексации.

4. Заключение

Видимо, дальнейшее развитие приложений для переработки текстовых информации будет следовать в направлении специализации инструментов. Доступные сейчас на рынке системы поиска информации или даже системы управления знаниями с элементами поиска информации обладают существенным недостатком, вытекающим из их универсального характера. Для достижения удовлетворительного уровня их полезности требуется сложная подготовительная и имплементационная работа, без чего эффективность поиска будет невелика. Образование приложений, предназначенных для специфичного применения,

облегчает процесс “учения” индукционных процессов, однако требует индивидуального подхода к процессу построения знаний, который в большей степени использует обработку натурального языка. На факультете управления Горно-металлургической академии в Кракове сейчас ведутся работы по созданию языка маркировки (имплементации XML) для описания рабочих постов.



Литература

1. Adams K. C. Extracting Knowledge. Artificial intelligence tools offer new ways to explore Web content // <http://www.intelligentkm.com/feature>.
2. Davenport T. H., Prusak L. Working Knowledge. — Boston: Harvard Business School Press, 1998.
3. Extensible Markup Language (XML) Activity Statement // <http://www.w3.org/xml>.
4. Мациол А. Управление знаниями в фирме. Теория и практика управления. — Краков, 2000 (на польск. яз.).
5. Rijsbergen C. J. Information Retrieval, 1999 // <http://www.dcs.gla.ac.uk/~iain/keith>.