

АРХИТЕКТУРА ИНТЕГРИРОВАННОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРОГНОЗНО-АНАЛИТИЧЕСКИХ ИССЛЕДОВАНИЙ

д.т.н. Д.В. Ландэ, к.ф.-м.н. С.М. Брайчевский, А.Н. Григорьев

Одна из проблем современных прогнозно-аналитических исследований – наличие больших массивов слабо структурированной информации, которую необходимо обрабатывать, анализировать и обобщать в оперативном режиме.

Решение данной проблемы может основываться на использовании интегрированных систем, позволяющих компоновать и обобщать разнородные данные, относящиеся к объекту исследования [1]. Один из подходов к созданию таких систем состоит в объединении фактографических баз данных с мониторингом информационных потоков [2], поступающих из различных источников, в том числе и с веб-сайтов сети Интернет. При этом системы управления базами данных задают структуру сохраняемой фактографической информации и обеспечивают прямой доступ к требуемым данным, а средства мониторинга составляют динамическую компоненту системы, используемую для накопления данных и выполнения различных статистико-аналитических исследований.

Интегрированная система, построенная подобным образом, может включать в себя три основные компоненты: базу данных реквизитов объектов исследования, систему автоматического мониторинга сетевой информации и ретроспективную базу данных, содержащую определенным образом отобранные документы, имеющие отношение к предмету исследования. К инструментальным средствам системы должна включать развитый поисковый механизм, средства агрегирования данных, визуализации результатов и генератором отчетов. Остановимся подробнее на отдельных компонентах интегрированной системы.

Фактографическая база данных, ядром которой является база данных реквизитов, содержит полностью формализованные и верифицированные данные, относящиеся к объектам исследования. Ее сопровождение осуществляется в автоматизированном режиме, поскольку предполагается, что только оператор, обладающий соответствующей квалификацией, обеспечивает необходимый уровень корректности вводимых данных. Этот режим может включать в себя две модификации: интерактивную и пакетную. В первом случае данные вводятся и редактируются с помощью специально предназначенных для этого экранных форм, а во втором – данные загружаются из специального файла, который предварительно готовится оператором (операторами) с помощью штатных средств (текстовых редакторов, электронных таблиц и т. д.) или экспортируется из внешних систем.

Мониторинговая компонента фиксирует место объектов исследования и роль в конкретном, динамически меняющемся контексте. Имеется ввиду, что устойчивое упоминание объекта исследования в средствах массовой информации, публикация информации о его деятельности на протяжении заметного промежутка времени отражает реальные общественные процессы, в которых он так или иначе принимает участие (эпизодические упоминания исключаются за счет корректного применения статистических методов). При этом сопровождение мониторинговой компоненты осуществляется путем выполнения ряда требуемых настроек. Ее функционирование в рамках интегрированной системы полностью автоматическое, однако при необходимости может быть предусмотрен интерактивный режим работы с соответствующими программными средствами (например, поисковыми).

Ретроспективная компонента является коммуникативной составляющей, обеспечивающей взаимосвязь фактографической базы данных и средств мониторинга, а также обеспечивает выдачу по запросам необходимых документов из ретроспективного фонда интегрированной системы.

Такой подход к построению интегрированных систем применялся авторами на практике при выполнении двух проектов, реализуемых в последнее время ИЦ «ЭЛВИСТИ». Первый проект по созданию инструментальных средств автоматического группирования информации о деятельности больших структурированных субъектов хозяйствования (заказчик - Центр комплексных исследований по вопросам антимонопольной политики) предполагал изучение хозяйственной деятельности структурированных субъектов хозяйствования в конкурентной среде [3, 4]. Второй проект, относящийся к созданию системы мониторинга веб-сайтов органов государственной власти, заказчиком которого был Государственный комитет телевидения и радиовещания Украины, был предназначен для отслеживания динамики информационного наполнения Интернет-ресурсов органов государственной власти.

Проиллюстрируем общие характеристики интегрированных систем группирования информации на примере первого проекта. Фактографическая база данных содержит информацию о компаниях и рынках, на которых эти компании представлены. Эти данные позволяют сгруппировать компании в соответствии с их иерархической подчиненностью и, таким образом, построить систему структурированных субъектов хозяйствования. Кроме того, система позволяет вести данные о ценах на различные товары, назначаемые субъектами хозяйствования, и об их долях на соответствующих товарных рынках.

Компонента мониторинга обеспечивает:

- постоянное пополнение хранилища оперативными сообщениями;
- эффективный одновременный многопользовательский доступ к базам данных
- удобные инструменты и средства и аналитического обобщения информации.

Компонента мониторинга соответствует потребностям надежного приема информации, поступающей в электронном виде, ее структуризации, классификации и предоставления к ней доступа режиме поиска. Компонента базируется на технологии InfoStream® [5], элементы которой, программные компоненты, подсистемы и информационное наполнение адаптируется к проблематике заказчика. При этом реализуется полномасштабное информационное хранилище, в котором накапливается и надежно хранится информация для использования в аналитической работе.

В результате функционирования компоненты мониторинга из общего потока информации извлекаются документы, содержащие сведения о компаниях, включенных в фактографическую базу данных. Эти документы помещаются в специальную ретроспективную полнотекстовую базу данных, позволяющую при необходимости получать доступ к тексту требуемого документа. Комплекс мониторинга позволяет также выполнять ряд дополнительных операций по аналитической обработке данных. К их числу относятся возможности построения информационных портретов, выделение сюжетов событий, определение динамики упоминаний компаний в документах. Примером содержательной обработки данных мониторинга новостной информации может служить построение таблицы взаимосвязей между фирмами на основе учета данных об освещении их деятельности в средствах массовой информации. Для этого аналитику требуется выбрать из фактографической базы данных набор фирм, а средства системы определяют для каждой пары фирм число их совместных упоминаний в документах

ретроспективной базы данных. Результаты отображаются на экране в виде таблицы, ячейки которой закрашиваются с различной интенсивностью в зависимости от частоты совместных упоминаний. Кроме того, формализованные данные о документах помещаются в специальную базу данных, осуществляющую связь между системой мониторинга, ретроспективной и фактографической базами данных.

В результате накопленного опыта разработки и внедрения подобных интегрированных систем поддержки принятия решений определены дополнительные требования [6], реализация которых возможна путем:

- использование материалов из средств массовой информации с ретроспективой не менее 10 лет;
- интеграции информации, которая получена из сети Интернет с другими электронными и печатными источниками в единой базе данных;
- интеграции существующих баз данных с внешними базами данных фактографической информации, которая позволит получать более полные количественные и качественные срезы, реально отражающие ситуацию об объектах исследования.

Для принятия серьезных решений в области экономики, информатизации, государственного управления, использование интегрированных систем, позволяющих находить, группировать и обобщать полученную из разных источников информацию об объектах исследований, в настоящее время является необходимым условием. Предложенный подход может быть одним из звеньев решения задачи информационной поддержки принятия таких решений.

Литература

1. Ланде Д.В., Фурашев В.М, Григор'ев О.М. Програмно-апаратний комплекс інформаційної підтримки прийняття рішень. - К. : Інжиніринг, 2006. - 48 с.
2. Ланде Д.В. Основы интеграции информационных потоков - К.: Инжиниринг, - 2006. - 240 с.
3. Ланде Д.В., Безулик В.О. Принципи функціонування комплексу інформаційної підтримки антимонопольної діяльності // Антимонопольно-конкурентна політика: теорія та практика: збірник наукових праць. Вип. 1. - К.: Фенікс, 2006. - С. 229-252.
4. Ланде Д.В., Безулик В.О. Концепція створення комплексу інструментальних засобів автоматичного групування інформації щодо діяльності великих структурованих суб'єктів господарювання // Антимонопольно-конкурентна політика: теорія та практика: збірник наукових праць. Вип.2. -К.: Фенікс, 2007. - С. 129-138.
5. Григорьев А.Н., Бородин С.А. и др. InfoStream. Мониторинг новостей из Интернет: технология, система, сервис: Научно-методическое пособие - Киев: Старт-98, 2007. - 40 с.
6. Бараш М.Я, Ланде Д.В., Безулик В.О. Програмно-апаратний комплекс інформаційної підтримки антимонопольної діяльності // Конкуренція. Вісник антимонопольного комітету України. - К., 2006. - № 1(20). - С. 18-26.