**Автоматизация сбора данных в системе Competition**

В работе рассматривается модуль автоматизированного сбора данных в составе программно-аппаратного комплекса (ПАК) Competition, разрабатываемого в виде SaaS-приложения (Software as a Service), которое позволит лицам, принимающим решения (ЛПР) от инновационных компаний проектировать конкурентную стратегию на основе анализа и прогнозирования состояния отраслевых рынков в соответствии с моделью глобальной конкуренции, которая описывает поведение системы интеллектуальных агентов.

**Обзор ПАК Competition**

В основе ПАК Competition [1] лежит шаблон проектирования MVC, разделяющий ПО на следующие компоненты: база данных системы, пользовательский интерфейс, вычислительный модуль системы поддержки принятия решений (СППР) и автоматизированного сбора данных. Диаграмма DFD (Data Flow Diagrams) иллюстрирует принцип работы ПАК (рис. 1).



Рис. 1. Диаграмма DFD, показывающая принцип работы ПАК

Отличительной чертой ПАК Competition от типичного BI-приложения[[1]](#footnote-1) является его предсказательная (прогнозирующая) компонента. Если решение предсказательных задач методами Data Mining, используемыми в BI, основано на регрессионном анализе данных (прогноз на основе статистики исторических данных), то в данном случае прогнозирование состояния рынка основывается на результатах поведения интеллектуальных агентов, моделирующих основные рыночные силы в соответствии с моделью глобальной конкуренции (рис. 2), где основные игроки F1, новые участники рынка F2, производители товаров-заменителей F3, поставщики F4, покупатели F5 и комплементоры F6 участвуют в процессе имитационного моделирования [2].

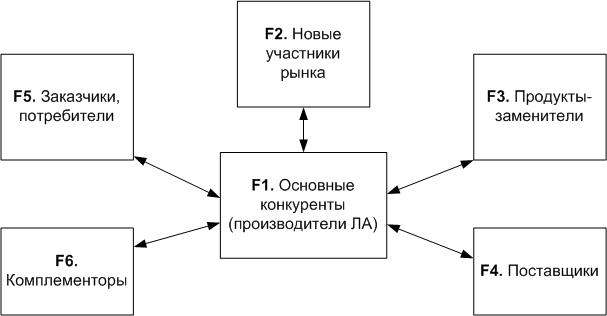


Рис. 2. Модель глобальной конкуренции

Для обеспечения эффективной работы СППР требуется располагать большими объемами релевантных данных, описывающих ситуацию на отраслевом рынке. Для накопления таких данных используется модуль автоматизированного сбора данных (МАСД). Его основные компоненты построены на основе ETL-процесса (**E**xtract – Извлечение, **T**ransform – Преобразование, **L**oad – Загрузка). Диаграмма DFD иллюстрирует принцип работы модуля (рис. 3). Аналитик передает начальные данные (полезную информацию, приоритетные сайты для поиска и пр.) в базу данных (БД) системы или же выбирает ранее собранные данные для проведения нового цикла поиска. На основе начальных данных с использованием внешнего ПО/сервисов (Avalanche [3], SiteSputnik [4], Y.Market [5]) собираются новые данные, которые помещаются во временные таблицы БД – процесс извлечения. Затем полученные данные анализируются и преобразуются в формат, пригодный для работы в среде СППР. Процесс продолжается до тех пор, пока не будет остановлен или пока не выполнится критерий избыточности информации, т.е. до момента, когда новые данные перестанут извлекаться, или же их объем станет незначительным.

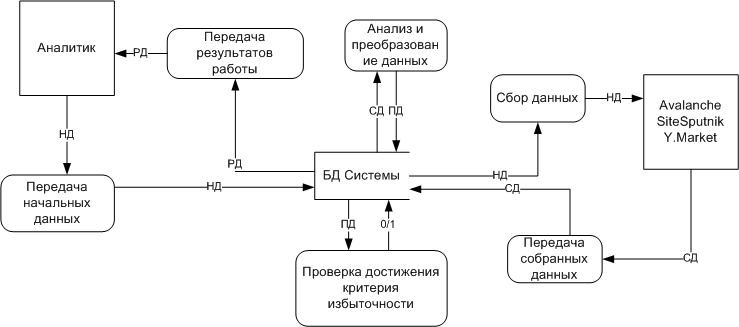


Рис. 3. Диаграмма DFD, показывающая принцип работы МАСД

В системе предусмотрена возможность сбора данных в фоновом режиме, что позволяет накапливать полезные и актуальные данные без участия аналитика.

**Заключение**

Модуль автоматизированного сбора данных является независимым вспомогательным инструментом сложного программно-аппаратного комплекса Competition. Этот модуль позволяет накапливать знания об участниках данного сектора отраслевого рынка, наращивая связанность исследуемых элементов системы рынка и определяя их свойства. Сформированная БД используется в составе СППР на основе модели глобальной конкуренции, позволяющей аналитикам проектировать конкурентную стратегию инновационной компании и передавать рекомендации ЛПР.

**Список литературы**

1. Бабенко Е.А., Клёнов Е.А., Ершов Д.М., Скородумов В.С. Свидетельство № 12-416 о регистрации объекта интеллектуальной собственности «Программно-аппаратный комплекс Competition конкурентного анализа сегмента рынка» / Зарегистрирован в Государственном реестре Госстандарта России 25 дек 2012. Москва 2012.
2. Бабенко Е.А. Разработка SaaS-приложения конкурентного анализа в секторе объектов авиационной техники. Журнал «Вестник МАИ». Том 20. Номер 1.
3. Avalanche – система интернет-мониторинга и конкурентной разведки. URL: <http://www.tora-centre.ru/avl3.htm> (дата обращения: 19.01.2014).
4. SiteSputnik – программа, предназначенная для поиска, сбора, мониторинга и анализа информации, размещенной в интернете. URL: http://sitesputnik.ru/ (дата обращения: 19.01.2014).
5. Y.Market (Яндекс.Маркет) – Система поиска и подбора различных товаров.URL: <http://market.yandex.ru/> (дата обращения: 19.01.2014).

1. BI (Business Intelligence) – бизнес-аналитика – это инструменты, используемые для преобразования, хранения, анализа, моделирования, доставки и трассировки информации в ходе работы над задачами, связанными с принятием решений на основе фактических данных. [↑](#footnote-ref-1)