

П.А. Мальцев, Т.В. Воронина¹

Национальный исследовательский университет
«Высшая школа экономики» (Пермский филиал)

Pavel_Maltsev@mail.ru, TVVoronina@gmail.ru

ОНТОЛОГИЯ BUSINESS INTELLIGENCE

Первое из общепризнанных определений Business Intelligence (BI) было дано аналитиками Garther в конце 80-х годов прошлого века. Тогда термин Business Intelligence был определён как «пользователецентрический процесс, включающий доступ и исследование информации, ее анализ, выработку интуиции и понимания, которые способствуют принятию правильных и обоснованных решений», [1]. Другими словами, в конце 80-х годов прошлого века понятие Business Intelligence представлялось неким концептуальным процессом, который должен был лежать в основе систем поддержки принятия решений того времени.

В 1996 г. аналитики Gartner уточнили определение: теперь Business Intelligence понимается как комплекс инструментов для анализа данных, построения отчетов и запросов, призванных помочь бизнес-пользователям извлечь значимую информацию из большого массива данных.

Таким образом, под термином Business Intelligence будем понимать и комплекс инструментальных средств анализа больших массивов данных с целью поддержки принятия эффективных решений, и область знаний, изучающую технологии и методы анализа данных с целью поддержки принятия эффективных решений (рис. 1).

На сегодняшний день следует выделить следующие классы средств Business Intelligence:

- средства построения аналитических отчетов;
- средства оперативно-аналитической обработки данных;

¹ Работа выполнена при поддержке Программы «Научный фонд НИУ ВШЭ» (проект № 12-09-0102)

- средства глубинного анализа данных;
- предметно-ориентированные информационно-аналитические системы и системы поддержки принятия решений;
- платформы Business Intelligence.

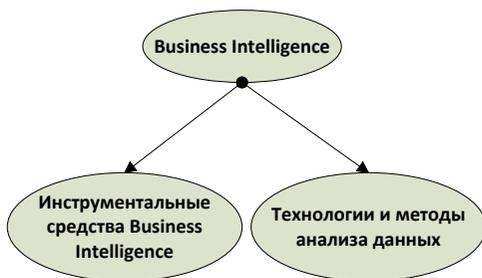


Рис. 1. Фрагмент верхнего уровня онтологии

Особое внимание в данной классификации заслуживают системы поддержки принятия решений и информационно-аналитические системы.

Аналитики аналитического агентства TAdviser дают следующее определение системам поддержки принятия решений: *система поддержки принятия решений*, или *СППР (Decision Support Systems, DSS)* – это компьютерная система, которая путем сбора и анализа большого количества информации может влиять на процесс принятия решений организационного плана в бизнесе и предпринимательстве. Интерактивные системы позволяют руководителям получить полезную информацию из первоисточников, проанализировать ее, а также выявить существующие бизнес-модели для решения определенных задач. С помощью СППР можно проследить за всеми доступными информационными активами, получить сравнительные значения объемов продаж, спрогнозировать доход организации при гипотетическом внедрении новой технологии, а также рассмотреть все возможные альтернативные решения, [2].

По *способу взаимодействия с пользователем* выделяют следующие типы СППР:

- *пассивные* системы не позволяют выдвинуть конкретное предложение, хотя реализуют средства, в различной степени поддерживающие пользователя при поиске наиболее эффективного решения;

- *активные* системы непосредственно участвуют в поиске и подготовке наиболее оптимального решения;
- *кооперативные* системы предоставляют пользователю возможность доработать найденные ими решение, а затем проверить внесённые пользователем коррективы, [2].

По *способу поддержки* различают:

- *модельно-ориентированные СППР*, выполняют поиск оптимальных решений основываясь на специально разработанные модели (статистические, финансовые и т.п.);
- *СППР, ориентированные на данные*, организуют поддержку пользователя при поиске эффективных решений, агрегируя большие объёмы данных из гетерогенных источников;
- *СППР, ориентированные на знания*, выполняют поиск оптимальных решений основываясь на специально разработанной базе знаний, [2].

Чаще всего встречаются СППР смешанного типа.

По *сфере использования* выделяют:

- *настольные СППР* – небольшие системы, ориентированные на использование одним пользователем, работающим на персональном компьютере;
- *общесистемные СППР*, используют в своей работе большие хранилища данных и ориентированы на использование многими пользователями, [2].

Информационно-аналитическими системами, или *ИАС*, чаще всего называют системы, предоставляющие доступ к предметно-ориентированному набору данных, полученных из гетерогенных источников. Другими словами, информационно-аналитическими системами будем называть пассивные, ориентированные на данные СППР.

На рис. 2 представлен фрагмент онтологии, соответствующий системам поддержки принятия решений и информационно-аналитическим системам.

Вопросы, посвященные различным аспектам разработки и применения СППР, подробно рассматриваются в публикациях, представленных в следующих журналах:

- Expert Systems with Applications;
- Decision Support Systems;
- Information Systems Management;
- Data Mining and Knowledge Discovery;
- Knowledge and Information Systems;

- Intelligent Data Analysis;
- Journal of Intelligent Information Systems.

Такое внимание, оказанное системам поддержки принятия решений и информационно аналитическим системам, вызвано тем, что данный класс средств Business Intelligence занимает особое место среди остального BI-инструментария. Дело в том, что СППР и ИАС ориентированы на конечных пользователей, принимающих решения, а значит, являются воплощением потребностей рынка в продуктах Business Intelligence. Остальные BI-инструменты играют исключительно вспомогательную роль.

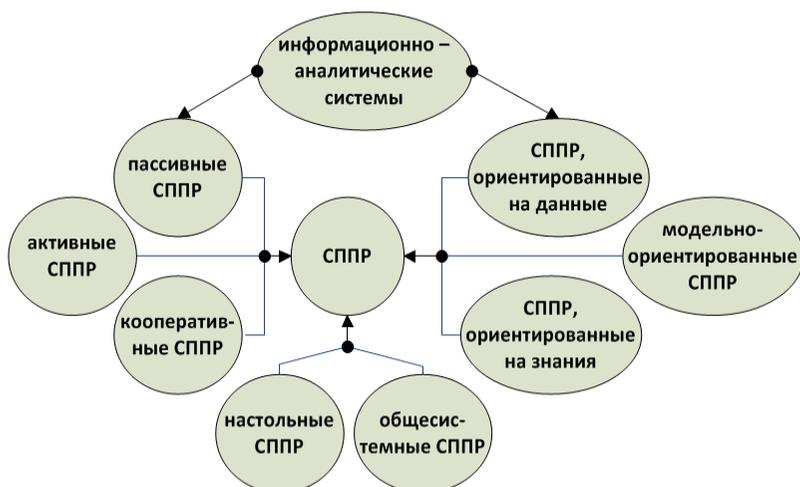


Рис. 2. Фрагмент онтологии «Классификация СППР»

Другим важным инструментом Business Intelligence являются BI-платформы. Аналитики компании Gartner определили *BI-платформу* как инструмент, который даёт организациям возможность строить приложения, позволяющие изучать и понимать бизнес. Аналитики Gartner в своём ежегодном отчёте [4] выделяют 12 основных функций современных платформ Business Intelligence, которые объединены в три группы:

1. *Средства интеграции.* В данную группу включены следующие функции:
 - *Инфраструктурное единство.* Все компоненты платформы должны использовать единые средства обеспечения

безопасности, метаданные, средства администрирования, средства генерации запросов, а также иметь однотипные интерфейсы.

- *Управление метаданными.* Все компоненты платформы должны не только опираться на одни и те же метаданные, но также должны обеспечивать быстрый поиск, хранение, использование и публикацию объектов метаданных.
 - *Средства разработки.* ВІ-платформа должна предоставлять средства программной разработки и интеграции ВІ-приложений. ВІ-платформа должна предоставлять разработчикам возможность создания приложений без кодирования, на основе применения мастеров.
 - *Совместная работа и управление рабочими процессами.* Платформа должна предоставлять пользователям возможность разделять информацию и обсуждать её с помощью общих папок и средств ведения дискуссионных «тредов» (discussion threads). В дополнение ВІ-приложения могут давать возможность назначать и отслеживать события или задачи, возложенные на отдельных пользователей, на основе неких заранее определенных бизнес-правил. Обычно данная функциональность предоставляется на базе интеграции с отдельным Workflow-инструментом.
2. *Средства визуализации данных.* К данной группе относят следующие функции:
- *Средства создания отчётов.* Современные ВІ-платформы предоставляют возможность создавать форматированные интерактивные отчеты. В дополнение к этому поставщики ВІ-платформ должны предоставлять широкий набор типов отчетов (финансовых, операционных и т.п.).
 - *Дэшборды, или приборные панели.* Приборные панели – это специальное средство визуализации данных, представляющее информацию в виде интуитивно понятного графического изображения (диаграммы, круговые шкалы, светофоры и т.п.). Данные индикаторы показывают состояние анализируемых параметров на фоне их целевого назначения. Руководитель или аналитик, подобно пилоту самолета, видит перед собой «доску приборов» и управляет системой, ориентируясь на значения индикаторов. При этом ключевые факторы, необходимые для управле-

ния предприятием, должны быть так или иначе измерены и представлены в виде показателей.

- *Генератор нерегламентированных запросов.* Платформа должна предоставлять пользователям возможность получать ответы на возникающие вопросы. Система предоставляет средства навигации по доступным ресурсам данных.
 - *Интеграция с офисными приложениями.* Часто BI-платформы используются как промежуточное звено в цепочке анализа информации, а Microsoft Office (в частности, Excel) выступает как BI-клиент. В этих случаях очень важно, чтобы BI-платформа обеспечивала интеграцию с Microsoft Office, включая поддержку форматов документов, формул и сводных таблиц.
3. *Средства анализа данных.* В данную группу объединены следующие функции:
- *Оперативно-аналитическая обработка данных (OLAP).* На сегодняшний день технология OLAP стала концептуальной основой при реализации средств обработки и манипуляции данными в информационно-аналитических системах и системах поддержки принятия решений.
 - *Продвинутой визуализации или средства визуального анализа.* Инструменты продвинутой визуализации позволяют представлять данные более эффективного их восприятия посредством использования интерактивных картинок и диаграмм вместо таблиц.
 - *Средства глубинного анализа данных.* В последнее время пользователи проявляют всё больший интерес к средствам статистического и интеллектуального анализа данных. Это требует от BI-платформ включения в них данного инструментария.
 - *Средства скоринг-анализа.* Используются контрольные показатели, отображаемые на информационной панели, для более глубокого анализа путем наложения их на некоторую стратегическую карту, которая увязывает ключевые параметры производительности со стратегическими задачами, [3, 4].

Таким образом, современная BI-платформа представляет собой наиболее полный набор инструментов Business Intelligence. Функциональные требования, предъявляемые к BI-платформам, диктуются не столько самими вендорами BI-платформ, сколько разработчиками си-

стем поддержки принятия решений и информационно-аналитических систем. На рис. 3 представлен фрагмент онтологии для соответствующей классификации инструментов Business Intelligence.

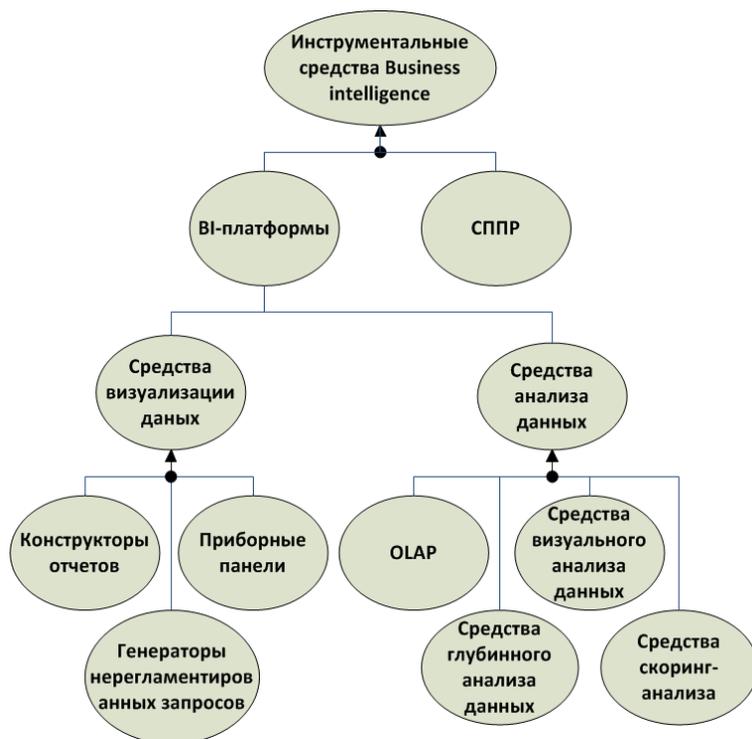


Рис. 3. Фрагмент онтологии «Классификация инструментальных средств Business Intelligence»

Вопросы, относящиеся к тематике современных BI-платформ, достаточно подробно освещаются на сайтах BI-вендоров:

- Microsoft (<http://www.microsoft.com/en-us/bi/default.aspx>);
- Oracle (<http://www.oracle.com/us/solutions/business-analytics/business-intelligence/overview/index.html>);
- MicroStrategy (<http://www.microstrategy.com/software/business-intelligence/>);
- IBM (<http://www-142.ibm.com/software/products/us/en/category/SWQ20>);

- SAP (<http://www.sap.com/cis/platform/businessintelligence/index.epx>);
- SAS (<http://www.sas.com/technologies/bi/>);
- PROGNOZ (<http://www.prognoz.ru>);
- BaseGroup Labs (<http://basegroup.ru>).

Особого внимания заслуживают *технологии и методы анализа*. Именно они лежат в основе любого BI-решения. Большинство современных BI-решений строится на основе концепции Knowledge Discovery in Databases (KDD). KDD – это концепция, которая описывает, как следует осуществлять анализ «сырых» данных.

Принято выделять 5 основных этапов процесса KDD:

- Подготовка исходного набора данных.
- Предварительная обработка данных, которая включает в себя решение задач выявления и устранения шумов и аномалий.
- Нормализация и трансформация данных. На данном этапе решаются задачи приведения данных к общим единицам и масштабу, а также подготовка данных к применению методов интеллектуального и статистического анализа.
- Применение методов интеллектуального, статистического анализа и технического анализа.
- Постобработка данных, которая включает решение задач интерпретации результатов анализа и визуализации данных.

Таким образом, все *методы аналитической обработки данных* следует подразделять на

- методы предварительной обработки (устранение шумов и аномалий, нормализация и трансформация данных);
- методы анализа (статистического и интеллектуального);
- методы постобработки данных.

Среди *методов анализа данных* следует выделять две основные группы:

1. *Методы статистического анализа данных*. Методы данной группы широко применяются при решении задач, связанных с анализом данных при наличии случайных и непредсказуемых воздействий. Неоспоримым достоинством данной группы методов является их глубокая теоретическая проработка, сделанная математиками и другими исследователями за последние двести лет. За это время накоплен большой опыт успешного применения этих методов в разных сферах человеческой деятельности, от экономики до космических исследований. При определенных условиях эти методы позволяют получать оп-

тимальные решения.

2. *Методы интеллектуального анализа данных.* Методы данной группы решают различные задачи анализа данных, опираясь на подходы из области искусственного интеллекта. Как и остальные методы искусственного интеллекта, методы интеллектуального анализа следует делить на
 - *когнитивные методы* – методы, в основе которых лежит идея моделирования мыслительного процесса человека;
 - *бионические методы* – методы, в основе которых лежит идея моделирования физиологической работы человеческого мозга.

Особой популярностью сегодня пользуются *методы глубинного анализа*, или *Data Mining*. Часто методы Data Mining отождествляются с методами интеллектуального анализа, но, по мнению авторов, это неверно. Data Mining – это процесс обнаружения в «сырых» данных ранее неизвестных нетривиальных практически полезных и доступных для интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [6]. Место методов глубинного анализа среди остальных методов анализа данных представлено на рис. 4.



Рис. 4. Место Data Mining среди методов анализа данных

На рис. 5 представлен фрагмент онтологии соответствующий классификации технологий и методов анализа данных.

Кроме приведённой выше классификации инструменты и методы Business Intelligence можно классифицировать также по областям применения:

- государственное и муниципальное управление;
- банковский сектор;
- оптовая и розничная торговля;
- корпоративное управление и др.

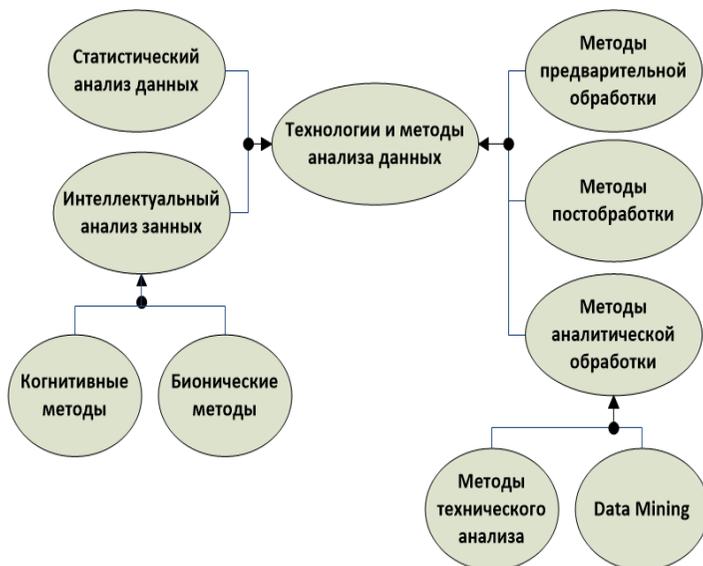


Рис. 5. Фрагмент онтологии «Классификация технологий и методов анализа данных»

В данной статье было дано определение понятия «Business Intelligence». По мнению авторов, следует обратить внимание читателя на двойственность данного понятия: в современном состоянии термин Business Intelligence понимается и как комплекс аналитического инструментария, и как область знаний, изучающая вопросы применения различных методов анализа данных.

В данной работе авторы предприняли попытку дать классификацию различных типов инструментов Business Intelligence и методов анализа данных. Приведённая классификация может быть полезна для студентов, начинающих изучать методы и средства BI, пользователей и разработчиков.

Разработанные онтологии предполагается использовать для реализации средств интеллектуального поиска и обработки данных учебно-исследовательского портала, посвящённого вопросам исследования и разработки средств создания информационных систем, основанных на моделировании. Материалы исследования будут также опубликованы на портале.

Библиографический список

1. *Артемьев В.* Что такое Business Intelligence? // Открытые системы. М.: 2003, № 04.
2. Системы поддержки принятия решений. [Электронный ресурс] М.: 2010. [Режим доступа: <http://www.tadviser.ru/index.php/DSS>; свободный].
3. *Харатишвили Д.* Рынок BI-платформ: претенденты и победители // Компьютер пресс. М.: 2008. № 7.
4. Magic Quadrant for Business Intelligence Platforms. [Электронный ресурс]. [Режим доступа: <http://businessintelligence.info/docs/estudios/Magic-Quadrant-for-Business-Intelligence-Platforms-2012.pdf>].
5. *Некипелов Н., Шахиди А.* Онтология анализа данных [Электронный ресурс]. [Режим доступа: <http://www.basegroup.ru/library/methodology/ontology/>].
6. Материалы сайта BaseGroup Labs / [Электронный ресурс] [URL: www.basegroup.ru].