

image not found or type unknown



Технологии интеллектуального анализа данных обеспечивают формирование аналитических данных посредством выполнения операции очищения данных локальных баз организации, применения статистических методов и других сложных алгоритмов. Появлению аналитических систем способствовало осознание руководящим звеном предприятий факта, что в базах данных содержится не только информация, но и знания (скрытые закономерности). Последние позволяют охарактеризовать процесс управления предприятием и дать интеллектуальную информацию для более обоснованного принятия решений.

Можно выделить следующие технологии интеллектуального анализа данных:

- Оперативный анализ данных посредством OLAP-систем;
- Поиск и интеллектуальный выбор данных Data Mining;
- Деловые интеллектуальные технологии BIS;
- Интеллектуальный анализ текстовой информации.

Интеллектуальный анализ данных Data Mining — это современная технология анализа информации. Цель данной технологии — нахождение в накопленных данных, необходимых для принятия оптимальных решений в различных областях человеческой деятельности. И в последнее время становится все более популярной у предприятий малого и среднего бизнеса, а не только в узком кругу специалистов, как это было несколько лет назад.

Компоненты интеллектуального анализа данных MS Office 2007 доступны для бесплатной загрузки и состоят из:

- средств анализа табличных данных для Excel, которые позволяют использовать Analysis Services для анализа и выявления закономерностей в данных из электронных таблиц Microsoft Excel;
- клиентских компонентов Data Mining для Excel, предоставляющих возможность для создания, настройки и сопровождения проекта Data Mining. В этой части надстройки предоставляются также средства для: — подготовки данных; —

построения, оценки и управления моделями; — получения результатов прогнозирования с использованием как табличных данных Excel в качестве источника, так и внешних источников, доступных базе данных Analysis Services;

- шаблонов Data Mining для Visio, которые предоставляют возможность графического отображения, форматирования, описания и распространения результатов моделирования Data Mining в виде диаграмм.

Интеллектуальный анализ извлекает шаблоны и тренды, существующие в данных. Такие шаблоны и тренды могут быть собраны воедино и определены как модель интеллектуального анализа данных. Модели интеллектуального анализа данных могут применяться к конкретным бизнес-сценариям, например: • прогнозирование продаж; • почтовая рассылка определенным клиентам; • определение продуктов, которые с высокой долей вероятности могут быть проданы вместе; • выявление последовательностей в том порядке, в котором клиенты добавляют продукты в корзину для покупок. Нужно понимать, что построение модели интеллектуального анализа данных является составной частью более масштабного процесса, начинающегося с определения базовой проблемы, которую модель будет решать, и заканчивающегося развертыванием этой модели в рабочей среде. Данный процесс может быть задан при помощи следующих шести базовых шагов:

1. Постановка задачи.
2. Подготовка данных.
3. Просмотр данных.
4. Построение моделей.
5. Исследование и проверка моделей.
6. Развертывание и обновление моделей.

Первым шагом процесса интеллектуального анализа данных является четкое определение бизнес-задачи. Этот шаг включает анализ бизнес-требований, определение масштаба проблемы, метрик, по которым будет выполняться оценка модели, а также определение конечной цели проекта интеллектуального анализа данных. Эти задачи можно сформулировать в виде следующих вопросов:

- Что необходимо найти?

- Какой атрибут набора данных необходимо предсказать?
- Какие типы связей необходимо найти?

И т.п.

Чтобы ответить на эти вопросы, возможно, потребуется исследовать уровень доступности данных, изучить потребности пользователей в отношении доступных данных. Если данные не поддерживают потребности пользователей, то может возникнуть необходимость в изменении определения проекта.

Вторым шагом процесса интеллектуального анализа данных является объединение и очистка данных, которые были определены во время первого шага. Службы Microsoft SQL Server 2005 Integration Services (SSIS) содержат все средства, необходимые для завершения данного шага, включая преобразования для очистки и объединения данных.

Третьим шагом процесса интеллектуального анализа данных является просмотр подготовленных данных.

Четвертым шагом процесса интеллектуального анализа данных является построение моделей интеллектуального анализа данных. Набор обучающих данных используется для построения модели, а контрольный набор данных — для проверки точности модели создания прогнозирующих запросов.

Пятым шагом процесса интеллектуального анализа данных является исследование построенных моделей и проверка их эффективности. Прежде чем развернуть модель в рабочей среде, необходимо проверить эффективность работы модели. Кроме того, возможно, было создано несколько моделей, и необходимо выяснить, какая из них наиболее эффективна. Если ни одна из моделей, созданных при выполнении шага 4 «Построение моделей», не обладает нужной эффективностью, может возникнуть необходимость вернуться к предыдущему шагу процесса и либо изменить постановку задачи, либо выполнить повторное изучение данных в исходном наборе данных.

Последним шагом процесса интеллектуального анализа данных является развертывание в рабочей среде наиболее эффективных моделей. После развертывания в рабочей среде моделей интеллектуального анализа данных можно выполнять множество различных задач, соответствующих потребностям пользователя. Ниже перечислены некоторые задачи, доступные для выполнения:

- Используйте модели для создания прогнозов, которые затем могут быть использованы для принятия бизнесрешений. SQL Server предоставляет язык расширений интеллектуального анализа данных, который можно использовать для создания прогнозирующих запросов, и построитель прогнозирующих запросов для помощи в построении запросов.
- Используйте службы Integration Services для создания пакета, в котором модель интеллектуального анализа данных используется для интеллектуального распределения входящих данных по разным таблицам.