

УДК 004.658.6

И.В.Бутенко (асп., каф. ИУС), Д.Ф.Дробинцев, ст. преп.

## ВЫБОР СРЕДСТВ ОРГАНИЗАЦИИ МЕТАДААННЫХ ДЛЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ

В настоящее время разработка информационных систем успешно решает вопросы оперативной деятельности предприятий и активно внедряется в область аналитических задач. Проблемы, с которыми они при этом сталкиваются, обусловлены как значительными объемами информации, так и многообразием структур исходных данных, которые необходимо получать из различных источников.

Хранилища данных являются фундаментом для систем интеллектуального управления бизнесом и управления связями с потребителем. Поэтому так важно правильно спланировать и реализовать систему хранения.

Это означает, что необходимо:

- точно преобразовать потребности бизнес-пользователей в рабочие модели данных;
- создать легкодоступное хранилище данных, которое будет непрерывно черпать информацию из различных приложений и источников данных;
- эксплуатировать хранилище таким образом, чтобы оно было в любой момент готово к изменению потребностей пользователя, т.е. к изменению характера его запросов.

Именно из-за этих требований в последнее время так возросла значимость данных о данных – метаданных. Метаданными явно или неявно пользуются все группы пользователей хранилища, начиная с наименее подготовленных конечных пользователей, приложения для которых управляются метаданными, и кончая разработчиками и администратором хранилища.

Спецификация Common Warehouse Metamodel (Общая метамодель Хранилища данных, далее CWM) – это стандарт, который описывает метаданные при использовании технологий Хранилищ данных (ХД), Business Intelligence, Knowledge Management (Управление знаниями). Опираясь на базовую метамодель, стандарт добавляет метамодели для реляционных, многомерных данных и данных таблиц, а также для преобразования, функций OLAP, data mining и ХД, включая процессы и операции.

Главная цель CWM – определить метамодель и обеспечить лёгкий обмен метаданными между инструментами хранилища, платформой хранилища (warehouse platforms) и репозиторием в распределённых гетерогенных средах. CWM основан на трех ключевых промышленных стандартах:

- UML – Unified Modeling Language, an OMG modeling standard;
- MOF – Meta Object Facility, an OMG metamodeling and metadata repository standard;
- XMI – XML Metadata Interchange, an OMG metadata interchange standard.

Поскольку данные в ХД попадают из внешних источников (ВИ), то вполне логично и метаданные ХД определять, как функцию от метаданных ВИ. В таком случае мы сможем упростить описание преобразований и очистки данных, т.к. и в исходной, и в конечной системе мы будем оперировать одними и теми же понятиями. Трудности такого подхода заключаются в том, что необходимо:

- Привести метаданные ВИ к стандарту CWM. Модель ВИ отображается UML-диаграммами.
- Загрузить метаданные ВИ в репозиторий метаданных ХД, при этом обязательно нужно отслеживать повторяющиеся метаданные. В данном случае этот этап аналогичен этапу

загрузки данных в ХД. Т.е. метаданные должны пройти этап очистки и согласования. Загрузка осуществляется с помощью стандарта ХМІ.

- На основе уже имеющихся метаданных построить метамодель ХД. Для этого этапа будет использоваться стандарт MOF.

Предложенный в данной работе алгоритм в настоящее время используется для построения аналитической системы для коммерческих банков.