

– оценка построенной модели. Имеющиеся данные разбиваются на два множества: обучающее и тестовое. Построение модели выполняется на обучающем множестве, а затем с ее помощью строят прогноз на тестовом множестве. Спрогнозированные результаты сравнивают с реальными данными и по степени ошибки оценивают модель;

– если построенная на первом этапе модель получила удовлетворительную оценку, то ее можно использовать для прогноза будущих событий.

Данная задача может решаться как методами математической статистики (экстраполяция, экспоненциальное сглаживание и др.), так и методами Data Mining, например методом скользящего окна.

Основная идея метода скользящего окна представлена гипотезой существования некоего закона, по которому можно определить значение очередного члена ряда как функцию от нескольких предыдущих членов. Обычно из каких-то соображений фиксируют число k и предполагают, что только k предшествующих членов влияют на дальнейшее поведение ряда, а зависимостью от остальных пренебрегают, т. е.

$$e_{n+1} = f(e_n, e_{n-1}, \dots, e_{n-k}).$$

При этом говорят об «окне» размером k , в пределах которого рассматривается ряд. Для нахождения функции f временной ряд «нарезается» на множество окон (каждое из которых сдвигает-

ся на один элемент). На полученном множестве выполняется поиск искомой функции.

Необходимо заметить, что если функция f используется для предсказания численных значений, то говорят о задаче регрессии. В случае категориальных значений ряда речь идет о классической задаче классификации. Решаются эти задачи методами Data Mining. Например, задача регрессии может быть решена методом SVM, а задача классификации – методами построения деревьев решений. При использовании метода деревьев решений полученные результаты легко представить в виде правил < если-то >. В этом случае в условной части (если) указываются уже прошедшие события, а в заключительной (то) – предсказываемые.

Описанные задачи оценки рисков невыполнения программных мероприятий не ограничиваются рассмотренным списком. Другие задачи в этой области также могут решаться методами Data Mining. Основным требованием для этого является доступность всей необходимой для анализа информации.

СПИСОК ЛИТЕРАТУРЫ

Анализ данных и процессов / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. СПб.: БХВ-Петербург, 2009.

I. I. Kholod

APPLICATION OF DATA MINING FOR PERFORMANCE EVALUATION PROGRAM ACTIVITIES DEFENSE COMPANIES

The article describes the application of the approaches of Data Mining methods for solving various problems associated with the assessment of the feasibility of the program activities of defense-industrial complex.

Data Mining, risks, risk assessment

УДК: 20.53.19, 28.23.13

И. И. Холод, С. В. Родионов

Построение централизованных хранилищ данных для систем управления предприятиями ОПК

Описывается построение централизованных хранилищ данных для систем поддержки принятия решений, используемых в управлении предприятий оборонно-промышленного комплекса.

Системы поддержки принятия решений, хранилища данных

В настоящее время существует множество информационных систем содержащих различную информацию о деятельности предприятий обо-

ронно-промышленного комплекса (ОПК). Такие системы разрабатываются по заказам различных ведомств, на разных уровнях (отраслевом и феде-

ральном), в разное время, построены на разных технологиях и работают в различных средах. Такое разнообразие обусловлено независимостью формирования требований к этим системам. Несмотря на это целевые задачи у большинства из них схожие: сбор и анализ информации о предприятиях ОПК для эффективного принятия решений по управлению ими как на уровне министерств, так и на отраслевом уровне. В результате во всех них хранится информация об одних и тех же предприятиях. Информация, во многом дополняющая друг друга, а иногда и противоречивая.

В целом вся собираемая информация имеет большой интерес с точки зрения ее анализа для выработки эффективных управленческих решений в развитии как отдельного предприятия ОПК, так и всей отрасли в целом. Для ее использования необходимо обеспечить сбор и хранение актуальной информации о предприятиях ОПК на уровне Министерства промышленности и торговли РФ. Решение данной задачи имеет два уровня:

1) организационный – определяющий нормативную базу, обязывающую предприятия и ведомства предоставлять актуальную информацию;

2) технический – обеспечивающий возможность передачи информации о предприятиях в единое информационное хранилище.

На организационном уровне процесс сбора информации можно разделить на три логических этапа:

– формирование нормативной базы для сбора информации;

– согласование регламента предоставления данных;

– подключение организации к информационному взаимодействию.

На техническом уровне определяются и реализуются технические средства, обеспечивающие сбор актуальной информации. При этом должны решаться следующие вопросы:

– каналы передачи информации;

– форма предоставления информации;

– сценарии передачи информации.

Каналами передачи информации могут быть:

– сети общего пользования – Интернет;

– ведомственные/закрытые сети передачи данных;

– системы передачи данных, предназначенные для передачи секретной информации (например, СПД «Атлас»).

При этом все перечисленные каналы связи могут использоваться совместно в зависимости от

их технической реализации на момент передачи информации, а также от уровня секретности передаваемой информации.

В зависимости от используемых каналов связи будут различаться режимы передачи информации:

• on-line – информация будет запрашиваться модулем загрузки по необходимости посредством вызова web-сервиса (протокола SOAP), в результате на стороне источника информации будет формироваться «обменный файл», который и будет передаваться;

• off-line – передача информации будет осуществляться периодически (регламентировано) посредством выгрузки данных в «обменный файл» на стороне источника информации и его загрузки в единое хранилище.

Реализация обоих вариантов не будет различаться практически ничем, за исключением непосредственно способа передачи «обменного файла», что позволяет применять для этих вариантов одни и те же программные модули.

Использование для передачи информации «обменного файла» позволит передавать информацию непосредственно из информационной системы источника информации. Для этого необходима доработка информационных систем в части формирования «обменного файла» и его передачи в единое хранилище.

Учитывая разнородность информационных систем как с точки зрения платформ, на которых они функционируют, так и с точки зрения технологий на которых они построены; наиболее эффективным представляется использование в качестве формата «обменного файла» формат XML (eXtensible Markup Language).

Уровень организации хранилища данных должен обеспечивать организацию средствами СУБД хранения следующих типов данных (см. рисунок) [см. лит.]:

– детальные данные;

– агрегированные данные;

– метаданные.

Детальные данные должны соответствовать данным, поступающим непосредственно из оперативных источников данных. Они должны извлекаться из отчетов, предоставляемых предприятиями отрасли.

Агрегированные данные должны вычисляться на основании детальных данных и соответствовать наиболее часто запрашиваемой при анализе информации. Архитектура хранилища данных должна предоставлять быстрый и удобный способ получать

информацию, интересующую пользователя. Для этого необходимо часть агрегированных данных хранить в хранилище данных, а не вычислять их при выполнении аналитических запросов.

Метаданные должны обеспечивать пользователя информацией о содержащихся в хранилище данных. Метаданные должны включать в себя описания:

- объектов: метаданные должны описывать объекты предметной области, сведения о которых хранятся в хранилище данных. Такое описание должно включать: атрибуты объектов, их возможные значения, соответствующие поля в информационных структурах хранилищ данных, источники информации об объектах и т. п.;

- пользователей: метаданные должны описывать категории пользователей, использующих данные. Они должны описывать права доступа к данным, а также включать в себя сведения о пользователях, выполнявших над данными различные операции (ввод, редактирование, загрузку, извлечение и т. п.);

правление ошибок, расщепление полей и т. п.), так и в процессе их эксплуатации в хранилище данных;

- времени: метаданные должны описывать время выполнения разных операций над данными (например, загрузка, агрегирование, архивирование, извлечение и т. п.);

- причин: метаданные должны описывать причины, повлекшие выполнение над данными тех или иных операций. Такими причинами могут быть требования пользователей, статистика обращений к данным и т. п.

Средства центрального хранилища данных (ЦХД) должны обеспечивать циркуляцию следующих потоков информации:

- входной поток: образуется данными, копируемыми из оперативного источника данных (ОИД) в хранилище данных;

- поток обобщения: образуется агрегированием детальных данных и их сохранением в хранилище данных;

- архивный поток: образуется перемещением детальных данных, количество обращений к которым снизилось;



- места хранения: метаданные должны описывать местоположение серверов, рабочих станций, оперативных источников данных, размещенных на них программных средств и распределение между ними данных;

- действий: метаданные должны описывать действия, выполняемые над данными, действия, которые выполняются как в процессе переноса из оперативных источников данных (например, ис-

- поток метаданных: образуется потоком информации о данных в репозиторий данных;

- выходной поток: образуется данными, извлекаемыми пользователями.

Средства центрального хранилища данных должны обеспечивать очистку данных, поступающих из различных источников – федеральных органов исполнительной власти, предприятий

ОПК и других источников. Очистка данных должна обеспечивать снижение вероятности попадания в хранилище ошибочной и/или противоречивой информации, связанной с неправильными преобразованиями форматов, ошибками ввода и т. п.

Реализация центрального хранилища данных, обеспечивающего хранение всей доступной информации по предприятиям ОПК, позволит решать различные задачи как по управлению отдельными предприятиями и отраслями в целом, так и по эффективному размещению государственного обо-

ронного заказа с минимальными рисками невыполнения соответствующих контрактов. Для решения этих задач к созданному хранилищу данных должен быть предоставлен доступ различных аналитических систем, использующих хранимую в нем информацию.

СПИСОК ЛИТЕРАТУРЫ

Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. СПб.: БХВ-Петербург, 2004. 336 с.

I. I. Kholod, S. V. Rodionov

CREATION OF CENTRALIZED STORAGE FOR DECISION SUPPORT SYSTEMS FOR DEFENSE COMPANIES

This paper describes the construction of a centralized data warehouse for decision support systems used in the management of the defense-industrial complex.

Decision support systems, data warehouses

УДК 681.5.01

С. А. Беляев

Реализация положений ITIL/ITSM в подсистеме управления и контроля функционирования ГАС «Выборы»

Описываются особенности реализации библиотеки инфраструктуры информационных технологий (ITIL/ITSM) в рамках государственной автоматизированной системы «Выборы».

Процессы предоставления услуг, поддержка услуг, информационная автоматизированная система

Успешность применения информационных автоматизированных систем в соответствии с современными подходами определяется тем, насколько они эффективно обеспечивают выполнение бизнес-процессов организации. Применительно к государственной автоматизированной системе (ГАС) «Выборы» это означает, что ее методы и средства должны быть направлены на эффективное и непрерывное обеспечение задач, связанных с подготовкой и проведением избирательных кампаний и референдумов различных уровней [см. лит]. Подсистема управления и контроля функционирования как раз и обеспечивает работоспособность ГАС «Выборы» в целом, а следовательно, решение основной задачи.

Современные тенденции по формированию стандартов, таких как ISO-9000, СММ, СММІ, а также подходов РМВОК и SWEВОК описывают выполнение требований к организации, проекту или процессу разработки программного обеспечения с точки зрения составляющих процессов. Не является исключением и ITIL (the IT Infrastructure Library – библиотека инфраструктуры информационных технологий), описывающая лучшие из применяемых на практике способы организации работы подразделений или компаний, предоставляющих услуги в области информационных технологий. В настоящее время ITIL становится стандартом для Российского ИТ-рынка, фактически став международным стандартом в