

Агрегация данных для формирования производственных решений на промышленных предприятиях с использованием онтологических систем

Шостак Игорь Владимирович

Национальный аэрокосмический университет им. Н. Е. Жуковского «Харьковский авиационный институт», профессор кафедры инженерии программного обеспечения, доктор технических наук, профессор, Украина

Собчак Андрей Павлович

Национальный аэрокосмический университет им. Н. Е. Жуковского «Харьковский авиационный институт», доцент кафедры менеджмента, кандидат технических наук, доцент, Украина

Фирсова Анна Викторовна

Национальный аэрокосмический университет им. Н. Е. Жуковского «Харьковский авиационный институт», аспирант кафедры менеджмента, Украина

Кушнарченко Александр Сергеевич

Национальный аэрокосмический университет им. Н. Е. Жуковского «Харьковский авиационный институт», ассистент кафедры документоведения, Украина

Аннотация. Рассмотрены вопросы использования средств онтологического инжиниринга при создании информационных структур для агрегированного хранения больших объемов разнородных данных о состоянии производства, а также возможности применения таких структур в системах поддержки принятия решений на производственных предприятиях.

Ключевые слова: онтологический инжиниринг, агрегация данных, OLAP-система.

Введение

Реализация сложных прикладных задач, имеющих большую практическую значимость, как правило, связана с преодолением двух фундаментальных проблем в практике человека – проблемы размерности и проблемы разнообразия. Влияние указанных проблем в ряде случаев усугубляется дефицитом времени на принятие решений и рисками значительного ущерба от неверных решений. Типичной задачей, при реализации которой в полной мере проявляются проблемы размерности и разнообразия, является мониторинг текущего состояния производства для принятия решений топ-менеджментом производственных предприятий. Эффективность решения этой задачи напрямую зависит от своевременности поступления и полноты учета информации о состоянии рынка, анализа эффективности работы собственных организационно-технических структур, оценки взаимоотношений с бизнес-партнерами и конкурентами, прогнозирования перспектив развития производства.

Современный уровень развития аппаратных и программных средств с некоторых пор сделал возможным повсеместное ведение баз данных оперативной информации на всех уровнях управления производством. Эффективность реализации процессов непосредственным образом связана с преодолением проблемы размерности (необходим учет большого числа релевантных, частично взаимосвязанных, разнородных (экономических, технических, социальных) показателей, источники получения которых топологически распределены в информационном пространстве), а также проблемы разнообразия, которая проявляется в данном случае, в необходимости рассмотрения большого числа прецедентов, имевших место в практике предприятия. Решение указанных проблем на сегодня предполагает использование методов и средств искусственного интеллекта в частности, онтологического инжиниринга. Применение онтологического подхода к представлению и обработке информации обусловлено тем обстоятельством, что практически всегда у специалиста, который решает подобные задачи, возникает необходимость в интеграции используемой информации, и данных, которые ее характеризуют, на основе тематических свойств информационных объектов, определяющих выбранную стратегию решения.

Вопросам создания и применения возможностей сложных и знание-ориентированных производственных информационных систем посвящен ряд работ таких ученых, как В. М. Глушков, В. П. Гладун, А. В. Палагин, В. А. Широков, В. Ф. Хорошевский, Д. А. Поспелов, А. Н. Авдий, В. Д. Соловьев, Т. А. Гаврилова и др. Однако вопрос интеграции всех составляющих информационного пространства предприятия при формировании решений в задачах, связанных с учетом большого числа разнородных факторов, оказывающих влияние на эффективность реализации производственных процессов, на сегодняшний день еще не имеет окончательного решения.

Таким образом, на сегодня, обобщенного решения по обеспечению процессов повышения эффективности интегрированного использования разнородных, топологически распределенных информационных ресурсов при формировании решений сложных производственных задач, нет. В связи с этим имеет место сложная научно-техническая проблема, которая связана с обеспечением процессов интегрированного взаимодействия сложных информационных объектов в процессе решения специалистами комплекса задач по идентификации текущего состояния производства и прогнозированию его дальнейшего развития. Наиболее прогрессивным инструментом создания информационных структур большого объема для хранения разнородных данных с учетом их семантики, является онтологический инжиниринг.

Цель статьи состоит в описании подхода к представлению разнородных данных о ходе реализации бизнес процессов на производственном предприятии, в виде однородной информационной структуры большого объема, обладающей возможностями эффективного доступа к релевантным данным и предварительной их обработке с целью обобщения и уточнения.

Решение рассматриваемой проблемы путем разработки соответствующей онтологической системы, предполагает реализацию последовательности взаимосвязанных прикладных задач:

1. Разработать методические средства онтологического представления управляющих параметров в задачах идентификации текущего состояния производства.

2. Адаптировать существующие модели, методы и средства формирования состояний онтологической системы для поддержки принятия решений по оцениванию текущего состояния производства.

3. Разработать метод формирования, в среде онтологической системы, решений по оценке экономической целесообразности изменения управляющих параметров нижнего уровня для поддержания уровня конкурентоспособности изделий.

4. Разработать и реализовать в онтологической среде прогностическую модель, отражающую влияние временных интервалов разработки и внедрения в производство новых изделий на жизненный цикл производственного предприятия.

Результаты исследования

Применение средств онтологического инжиниринга для агрегации данных на производственных предприятиях. Идея использования онтологий для интеграции разнородных информационных объектов заключается в том, что при принятии решения об отождествлении некоторой пары информационных атрибутов из разных источников, осуществляется анализ их смыслового (семантического) родства путем выявления связей между соответствующим этим атрибутам понятием (концептам) в онтологии предметной области [1]. Подход к агрегации данных с использованием онтологических технологий, в целом дает принципиальную возможность решить эту проблему, поскольку он лишен многих недостатков, присущих традиционным, чисто техническим методам, и предоставляет возможность разработки приложений, работающих с информацией на семантическом уровне.

На практике, проблема агрегации данных имеет много аспектов [2]. Сложность и характер используемых способов ее решения существенным образом зависят от уровня интеграции, который необходимо обеспечить, от свойств отдельных источников данных и всего множества источников в целом, а также от необходимых способов интеграции.

Системы агрегации данных могут обеспечивать интеграцию данных на следующих трех уровнях [3, 4].

1. Интеграция физически сводится к приведению данных из различных источников к единому, в данном случае рациональному формату их физического представления.

Проблема интеграции на рассматриваемом уровне состоит в том, что в источниках данных могут использоваться различные форматы представления данных.

2. Интеграция на логическом уровне предполагает возможность доступа к данным, содержащимся в различных источниках, в терминах единой схемы, описывающей их совместное представление с учетом структурных и, возможно, поведенческих (при использовании объектных моделей) свойств данных. Семантические свойства данных при этом не учитываются.

Проблема интеграции на этом уровне состоит в том, что возникают для различных источников неоднородность используемых моделей данных, или будут различаться схемы этих данных, при использовании одной и той же модели данных.

3. Интеграция на семантическом уровне обеспечивает поддержку единого представления данных с учетом их семантических свойств в контексте единой онтологии модели предметной области.

Проблема интеграции на этом уровне состоит в том, что различным источникам данных могут соответствовать разные онтологии.

Анализ ситуации в решении проблемы интеграции данных на семантическом уровне позволяет на основе работ [5, 6] сделать вывод о недостаточной изученности проблемы интеграции онтологий в рамках одной онтологической системы.

Технологические трудности агрегации данных: чаще всего при этом имеются в виду способы хранения и представления знаний, ориентированных на компьютерную обработку информации и, в частности, представление, которые состоят из явных объектов и из соображений или утверждений о них. Представление в подобной явной форме позволяют генерировать дедуктивные выводы из знаний, которые были сохранены в системе ранее.

Общая проблема представления знаний включает ряд частных проблем:

- представление декларативных знаний (фактов), как данных, наделенных семантикой;
- представление процедурных знаний, как отношений между элементами модели, в том числе в виде процедур и функций,
- представление метазнаний – правил обработки фактов, способов организации вывода на знаниях, методов решения задач пользователем, порождение новых знаний и т. д.

В настоящее время существуют новые подходы к представлению информации – визуальные способы, которые дают возможность максимального использования всего лингвистического аппарата конкретного языка, а именно: онтологии, таксономии, классификаторы, семантические сети, тезаурусы, словари для визуализированного представления знаний в виде онтологических графов, семантических сетей и тому подобное.

Правильный выбор метода представления знаний в определенной степени может упростить решение сложных задач, часто осложняется еще и трудностями для эксперта по формулированию в явном виде имеющихся у него знаний. В этом случае необходимо использовать инструментальные средства автоматизированного анализа документа и построения базовой («первоначальной версии») онтологии, которую, в дальнейшем, эксперт должен анализировать и редактировать согласно поставленной цели.

По своей сути, задача агрегации данных состоит в преодолении многочисленных проявлений неоднородности, что присуще производственным информационным системам, которые создавались и создаются, не на унифицированном отношении к данным, поскольку существующие корпоративные системы имеют разную функциональность, в них используются различные типы данных (алфавитно-цифровые, медийные, структурированные и неструктурированные), их компоненты различаются по автономности, имеют различную производительность, строятся на различных аппаратных платформах, располагают различными средствами управления данными, используют различные модели данных, интерфейсы и многое другое.

Агрегация данных предоставляет лицу, принимающему решения единый взгляд на разнородные источники данных, предусматривает общую модель и общее отношение к семантике, с тем, чтобы обеспечить возможность для доступа к информации, а в случае необходимости, предоставляет возможность преодолевать конфликтные ситуации и вырабатывать правильные решения поставленных задач.

Достаточно подробно стандартная методика разработки онтологий представлена в работе [7]. основополагающие правила разработки онтологии авторы формулируют следующим образом:

1. Не существует единственно правильного способа моделирования предметной области – всегда существуют жизнеспособные альтернативы.

2. Разработка онтологии – это обязательно итеративный процесс. Под итеративным процессом понимается неоднократный проход по онтологии с целью ее уточнения: на начальном этапе строится первичный, «черновой» вариант; затем составленная онтология подвергается проверке и уточнению, при этом возможна частичная или полная ревизия первоначального варианта онтологии.

3. Элементы онтологии должны быть близки к объектам (физическим или логическим) и отношениям в заданной предметной области.

Описанную выше форму представления принято называть онтологией задач. Необходимо отметить, что онтология сама по себе не предназначена для решения конкретных инженерных задач. Для этого предназначены экспертные, поисковые, вопросно-ответные, Text Mining системы концептуального доступа к базам данных и другие проблемно-ориентированные системы. Онтология должна лишь обеспечивать им семантическую поддержку при формировании ответов на вопросы о предметной области [8, 9].

Агрегация разреженного гиперкуба данных о текущем состоянии производства. Многомерная база данных называется разреженной, если относительно большой процент ячеек содержит пустые (утраченные) данные. Вполне обычны такие наборы данных, которые содержат 1 %, 0,01% и даже меньшую долю возможных данных.

Агрегация данных – суть получение значений, соответствующих меткам некоторого уровня $l, l \geq 1$ иерархического измерения D на основе значений уровня $l - 1$.

Рассмотрим иерархическое измерение D с L уровнями (рис. 1).

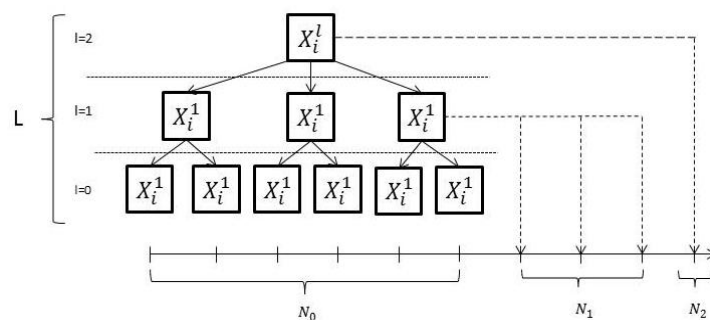


Рисунок 1 – Агрегация гиперкуба данных. Одномерное представление

Первичные данные (факты) соответствуют нижнему уровню иерархии ($l = 0$) [1]. Вычисление агрегатов производится в соответствии с применяемым методом агрегирования. Например, в случае суммирования значение агрегата X_j^l на уровне иерархии $l = 1$ может быть вычислено по формуле (1):

$$X_j^1 = \sum_{i=1}^{M_j} x_i^0, \quad (1)$$

где M_j – количество фактов, соответствующих меткам, являющимся дочерними по отношению к метке j .

Обобщая, получим формулу вычисления агрегатов по методу суммирования на остальных уровнях иерархии (2):

$$X_j^l = \sum_{i=1}^{M_j} x_i^{l-1}, \quad l = 1, \dots, L; j = 1, \dots, N_l. \quad (2)$$

Ось измерения D , первоначально содержащая метки, соответствующие нижнему уровню иерархии ($l = 0$), может быть дополнена метками, соответствующими уровням иерархии, начиная с $l = 0$. Таким образом, различие между метками, соответствующими первичным данным и метками, соответствующим агрегатам является условным.

Операция свертки данных в этом случае представляет собой построение среза гиперкуба данных, соответствующего смене метки уровня агрегации $l_1 | 0 \leq l_1 < L$ на уровень $l_2 | l_1 < l_2 \leq L$. Операция детализации соответствует смене метки уровня $l_1 | 1 \leq l_1 \leq L$ на уровень. Количество агрегатов для одного измерения $N_A = \sum_{i=1}^L N_i$.

Покажем, каким образом повлияет отсутствие показателя нижнего уровня на формирование агрегата [2].

Выбор того или иного метода агрегации данных зависит от конкретной решаемой задачи. Технологически процедура подсчета агрегатов выполняется с использованием т.н. карт агрегации, которые включают стандартные методы агрегации. Во многих популярных OLAP-системах как метод

агрегации «по умолчанию» используется метод добавления, предусматривающий наличие первичных данных на нижнем уровне иерархии.

Однако возникает вопрос о применимости данных методов при агрегации данных в разреженных гиперкубах [3].

При решении аналитических задач важно знать не только значение показателя, но и то, насколько он может доверять полученному значению. Вычисление агрегата по методу среднего значения при наличии первичных данных по всем значениям нижнего уровня в иерархии дает 100 % достоверность, поскольку нет причин полагать, что это среднее значение могло быть чем-либо искажено. Очевидно, что в стандартных методах агрегации не учитывается ситуация неопределенности первичных данных, которые отвечают некоторым меткам нижнего уровня иерархического измерения.

Именно такую ситуацию являет собой агрегация данных в разреженном гиперкубе. Во время выполнения агрегации в разреженном гиперкубе по методу вычисления среднего необходимо введение дополнительного параметра, который характеризует уровень достоверности полученного результата. Технологически эта операция может осуществляться путем создания дополнительной карты агрегации, которая включает расчет уровня достоверности для каждого, полученного в ходе агрегации значения.

Вычисление агрегата на первом уровне иерархии осуществляется по формуле (3):

$$Ag_j^l = \frac{\sum_{i=1}^{V_{jj}} ag_i^0}{V_j}, \quad (3)$$

где V_j – число фактов, соответствующих атрибутам, являющимся дочерними по отношению к атрибуту j .

Обобщая, получим формулу (4) вычисления агрегатов на остальных уровнях иерархии:

$$Ag_j^l = \frac{\sum_{i=1}^{V_{jj}} ag_i^{l-1}}{V_j}, \quad l = 1, \dots, N. \quad (4)$$

Рассмотренный метод может быть применен для построения карт агрегации в разреженных гиперкубах данных и дает возможность оценить уровень достоверности полученных результатов на этапе анализа. При этом может применяться для построения карт агрегации в разреженных гиперкубах данных и предоставляет возможности оценки степени достоверности полученных результатов на этапе анализа.

Применение OLAP-технологии при разработке онтологии задач выбора в процессе создания онтологической системы поддержки принятия производственных решений. Рассмотрение процесса агрегации разнородных данных может быть сведено к выделению линейно-упорядоченного множества про-

странства решений задачи выбора, заданной над всеми общими концептами взаимодействующих систем. Тогда онтология задачи выбора определяет для каждого SN пространства отображения ее информационных состояний. Это пространство можно представить в виде определенного OLAP-куба. Для таких информационных объектов справедливо следующее:

Утверждение 1. Онтология задачи выбора вида (5) унивалентна любому объекту OLAP-куба.

$$O_{TPCh}; W : (X, Y, Z) \rightarrow W. \quad (5)$$

R^+ является минимальной множеством свойств множества концептов X онтологической системы O_n , между которыми существует отношение бинарной частичной упорядоченности \check{p} . Описание OLAP-куба вида $W : (X, Y, Z) \rightarrow W$ может быть представлено в виде проекции отношений, согласно положениям реляционной алгебры [10]. Между этими проекциями существует бинарное отношение линейной упорядоченности p . Если применить метод конверсии η для элементов, представленных за счет проекций (X, Y, Z) , и сформировать на их основе выражение абстракции Λ , то мы переходим к определению бинарных структур и можем формировать онтологию задачи выбора вида (5).

Утверждение 2. Онтология задачи выбора вида (6) унивалентна любому объекту векторного пространства любой размерности.

$$O_{TPCh}; V(F) = (V, F, +, \times). \quad (6)$$

Любое векторное пространство включает в себя отношение линейной упорядоченности p , на основе которого можно выделить бинарное отношение частичной упорядоченности. Исходя из этого обстоятельства, из элементов, которые непосредственно составляют векторное пространство, можно сформировать онтологию задачи выбора вида (6).

Выводы

Главными особенностями информационных структур, применяемых в производственных информационных системах, является разнородность данных и значительный их объем. При создании систем поддержки принятия решений на производственных предприятиях для хранения разнородных данных о текущем состоянии бизнес процессов целесообразно использовать средства онтологического инжиниринга. Разработку онтологии задач выбора, которая является центральной в онтологической системе поддержки принятия производственных решений, и предполагает решение задачи агрегации данных, можно провести с использованием OLAP-технологии.

Перечень использованных источников

1. Заботнев М. С. Методы представления информации в разреженных гиперкубах данных [Электронный ресурс] / М. С. Заботнев // OLAP.ru : Business intelligence – effective data mining & analysis. – Режим доступа: <http://www.olap.ru/home.asp?artId=77>. – По состоянию на 01.03.2016. – Загл. с экрана.
2. Шаховська Н. Б. Опрацювання невизначеності у великих даних / Н. Б. Шаховська, Ю. Я. Болюбаш // Радіоелектроніка, інформатика, управління. – 2014. – № 1. – С. 96-105.
3. Шаховська Н. Б. Робота з великими даними – показниками соціо-еколого-економічного розвитку регіону / Н. Б. Шаховська, Ю. Я. Болюбаш // Восточно-Европейский журнал передовых технологий. – 2013. – Том 5. – № 2(65). – С. 4-8.
4. Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсегян [и др.]. – Санкт-Петербург : БХВ-Петербург, 2004. – 336 с.
5. Шостак И. В. Применение логических моделей представления знаний в интеллектуальных компонентах систем управления машиностроительными предприятиями / И. В. Шостак, О. В. Палун, Д. А. Бастеев / Авиационно-космическая техника и технология. – 2006. – № 7(33). – С. 145-151.
6. Соколов А. Ю. Сравнительная оценка эффективности онтологической и реляционной моделей данных и знаний / А. Ю. Соколов, И. В. Шостак, Д. А. Бастеев // Вісті Академії інженерних наук України «Машинобудування та прогресивні технології». – 2006. – № 3(30). – С. 173-177.
7. Бастеев Д. А. Синтез онтологии управления проектом технологической подготовки производства на машиностроительном предприятии / Д. А. Бастеев // Открытые информационные и компьютерные интегрированные технологии. – 2006. – Вып. 30. – С. 97-106.
8. Шостак И. В. Применение онтологического представления знаний в многоуровневой системе управления проектами технологической подготовки производства / И. В. Шостак, Д. А. Бастеев // Открытые информационные и компьютерные интегрированные технологии. – 2004. – Вып. 24. – С. 136-147.
9. Бастеев Д. А. Интеллектуальная поддержка принятия решений для управления проектом технологической подготовки авиационного производства с использованием онтологического представления знаний / Д. А. Бастеев, И. В. Шостак // Автоматизированные системы управления и приборы автоматики. – 2004. – Вып. 129. – С. 94-101.
10. Вебер А. В. Knowledge-технологии в консалтинге и управлении предприятием / А. В. Вебер, А. Д. Данилов, С. И. Шифрин. – Санкт-Петербург : Наука и Техника, 2003. – 176 с.

© И. В. Шостак, А. П. Собчак, А. В. Фирсова, А. С. Кушнаренко

Data aggregation for the formation of manufacturing solutions for industrial enterprises using ontology systems

Shostak Igor

*N. E. Zhukovsky National Aerospace University "Kharkiv Aviation Institute", Professor of Department
of Computer Engineering, Doctor of Science (Technique), Professor, Ukraine*

Sobchak Andrii

*N. E. Zhukovsky National Aerospace University "Kharkiv Aviation Institute", Associate Professor
of Department of Management, PhD in Technical Sciences, Associate Professor, Ukraine*

Firsova Anna

*N. E. Zhukovsky National Aerospace University "Kharkiv Aviation Institute",
graduate student of Department of Management, Ukraine*

Kushnarenko Alexander

*N. E. Zhukovsky National Aerospace University "Kharkiv Aviation Institute",
Assistant Professor of Department of Documentation, Ukraine*

Abstract. The article deals with the problems of using ontological engineering resources to create aggregated information structures for the storage of large volumes of heterogeneous data on the status of production, as well as the use of such structures in the decision support systems in manufacturing plants.

Keywords: ontological engineering, data aggregation, OLAP-system.

© I. Shostak, A. Sobchak, A. Firsova, A. Kushnarenko