

УДК 681.3.016

И.Е. Мазурок

ФОРМАЛЬНЫЕ ОНТОЛОГИИ В МОДЕЛИРОВАНИИ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННО-УПРАВЛЯЮЩИХ СИСТЕМ

Введение

Современное состояние развития информационно-управляющих систем в составе автоматизированных систем проектирования и производства имеет ряд принципиальных ограничений связанных с используемыми платформами концептуально-логического проектирования [1]. Для автоматизации решения сложных современных научно-практических задач, как правило, недостаточно возможностей обычных баз данных или систем Datawarehouse. Базы данных, обеспечивая хранение четко структурированных данных, создают удобную среду функционирования для детерминированных алгоритмически определяемых процессов, но не эффективны для представления слабоструктурированной информации. Напротив, системы Datawarehouse, предназначенные для хранения нетипизированной или слабо структурированной информации требуют существенной интеллектуализации систем обработки [1].

Рассмотрим проблемы связанные с использованием интеллектуальных информационных технологий в данной области [2].

Цель

Целью работы является разработка системы онтологий для моделирования содержательных когнитивных структур в интеллектуальных информационных системах и формальных методов ее построения. Такие модели должны стать базой для решения задач анализа и синтеза информационных структур.

Материал и результаты исследований

Построение некоторой онтологии $O = \langle T, D \rangle$ обычно подразумевает создание таксономии понятий T и системы определений узловых понятий D . Большой объем работы и высокие требования к квалификации разработчиков не позволяет широко

использовать онтологии в интеллектуальных информационно-управляющих системах общего назначения и сверхбольших репозиториях баз знаний. Автоматизация онтологических исследований происходит в основном за счет стандартизации документов и процессов с последующей разработкой документирующего инструментария. Хорошим примером может служить методология IDEF (Integrated DEFinition LMethods) из семейства Integrated Computer-Aided Manufacturing (ICAM) для решения подобных задач моделирования сложных систем [3]. Данная методология (IDEF 0- IDEF 14) позволяет решать задачи отображения и анализа моделей деятельности различного рода сложных систем. В стандарте онтологического исследования (Ontology Description Capture) сложных систем по методологии IDEF5 [4] онтология системы может быть описана при помощи словаря терминов и правил формирования достоверных утверждений о системе.

Основные проблемы алгоритмизации онтологических исследований находятся в плоскости обработки отношения «означаемое - означающее», поскольку семантика понятий обычно лежит непосредственно в предметной области, т.е. за пределами формальной системы.

Для решения задач формально алгоритмического построения онтологий в системах с обобщенным дискурсом (текстологические, музыкально нотационные, чертежно-графические) предлагается использовать в качестве исходного материала большой объем уже накопленных в машинной форме данных из соответствующей предметной области – стандарты, отчеты, схемы, и пр. Поскольку накопленный материал составлен специалистами соответствующей предметной области, полученный обобщенный дискурс является отражением исторически сложившихся онтологий. Сложившиеся когнитивные структуры можно извлечь методами кластерного анализа. Выбор дискурса для включения в базу анализа должен осуществляться разработчиком онтологии. Т.е. кластерный анализ и методы семантической классификации употребления понятий при большом объеме информации дают возможность построить многоуровневую формальную таксономию без непосредственного привлечения экспертов.

Покажем возможности и ограничения метода формальных онтологий. Обозначим через ${}^n\Lambda = \{\lambda_i | i = \overline{1, n_\lambda}\}$ базовое множество онтологии – множество базовых объектов для ее построения (слова, муземы). Тогда естественной онтологией можно назвать любую пару вида ${}^nO = \langle {}^nB({}^n\Lambda), \Sigma \rangle$, где база онтологии ${}^nB({}^n\Lambda) \subset \tilde{\beta}({}^n\Lambda)$ является несобственным подмножеством гипербулеана над базовым множеством, а семантическая функция $\Sigma : {}^nB({}^n\Lambda) \rightarrow {}^n\Lambda$ именуется семантическими категориями. Формальная онтология первого порядка ${}^1O = \langle {}^1B({}^n\Lambda), \Sigma \rangle$ не оперирует структурами семантических категорий и не поддерживает именования. Т.е. семантические категории являются простыми подмножествами базового множества ${}^1B({}^n\Lambda) \subset \beta({}^n\Lambda)$ и для них задана некоторая нумерация $\mu : {}^1B({}^n\Lambda) \rightarrow \mathbb{N}$. Очевидно, что построение формальных онтологий более высоких порядков потребует последовательной детализации структуры онтологий.

Для отображения структуры псевдокатегоризации, составляющей основу формальной онтологии можно использовать два подхода. Первый основан на методах поиска в ширину. На каждом шаге категоризации ограничивается единичный уровень кластеризации понятий. При этом понятия могут агрегировать понятия всех уже сформированных уровней

$$\begin{aligned} {}^nO &= \langle {}^nB(\Lambda), \Sigma \rangle \\ {}^nB(\Lambda) &\in 2^{\bigcup_{j=0}^{n-1} ({}^jB(\Lambda))} \end{aligned} \quad (1)$$

в которых реализуется необъединяющее агрегирование множеств.

Отсутствие именований и структур категорий допускает формальное построение системы онтологий первого уровня. Наличие онтологий, ориентированных на конкретные предметные области создает базу для построения семиотических моделей текстов с последующим выделением актантов. Второй подход использует идеи поиска в глубину. Т.е. категоризация происходит без ограничения степени вложенности агрегирования. Создаваемые при этом формальные понятия

$$O = \{k = \langle B(\Lambda), \Sigma \rangle\} \quad (2)$$

являются классами множеств, строящимися, в отличие от (1), на основании объединяющего агрегирования

$$B(\Lambda) \in 2^{\bigcup_{j=0}^{\infty} ({}^j B(\Lambda))} \quad (3)$$

В обоих подходах используется идентифицирующая именуемая функция

$$\begin{aligned} \Sigma : {}^n O \rightarrow N, \exists \Sigma^{-1} : N \rightarrow {}^n O \cup \{\emptyset\} \\ \Sigma : O \rightarrow N, \exists \Sigma^{-1} : N \rightarrow O \cup \{\emptyset\} \end{aligned} \quad (4)$$

Для которой должна существовать обратная функция Σ^{-1} – функция выборки. Обратная функция доопределена таким образом, чтобы для идентификаторов, не соответствующих формальным категориям, построенным в результате анализа, она возвращала фиктивную категорию $\{\emptyset\}$.

При необходимости функции именования для открытых систем общего назначения строятся методом естественных фолксономий [4].

Выводы

Построенные формальные категории позволяют создать формальную таксономию на основании полностью алгоритмизированной обработки больших и сверхбольших информационных массивов. Формальные таксономии обладают принципиальными отличиями от используемых в существующих понятиях онтологий. Первое принципиальное отличие состоит в том, что графовой моделью формальной онтологии является не корневое дерево, а ориентированный ациклический граф с заданным истоком и множеством стоков, соответствующих единичным и конкретным понятиям. Второе отличие состоит в отсутствии именования категорий, что снимает проблему релевантности естественного языка предметной области. Построенные таким образом системы понятий не удобны для использования людьми, но полностью соответствуют своему назначению – служить основой для анализа и синтеза когнитивных структур интеллектуальных информационных и управляющих систем.

ЛІТЕРАТУРА

1. Мазурок И.Е. Формальное описание класса информационных систем основанных на знаниях// 36.статей: Нові інформаційні технології навчання в учбових закладах України, 1998. - вип. VI. - С.217-226.

2. Мазурок И.Е. Метамоделирование содержательной части систем, основанных на знаниях// Материалы V семинара “Моделирование в прикладных научных исследованиях”. – Одесса: Од.гос.политехн.ун.-т.– 1998.– С.51-55.
3. Ziyu Liu, Lei Huang. Research on Process-Oriented Enterprise Knowledge Modeling and Integration Management Based on Ontology// Research and Practical Issues of Enterprise Information Systems II Volume 1. 2007, pp. 349-357.
4. Robert Jdschke, Andreas Hotho, Christoph Schmitz, Gerd Stumme. Analysis of the Publication Sharing Behavior in BibSonomy// Conceptual Structures: Knowledge Architectures for Smart Applications. – Volume 4604/2007, pp. 283-295.

Получено 15.03.2008 г.