

## АЛГОРИТМИЗАЦИЯ ИНТЕГРАЦИИ ОНТОГРАФОВ

Гусева О. В., Волосюк О. В., Тихонов Ю.Л.

Луганский национальный университет имени Тараса Шевченка

кафедра информационных технологий и систем

E-mail:its@luguniv.edu.ua

### *Аннотация*

*Гусева О. В., Волосюк О. В., Тихонов Ю.Л. Могильный Г.А. Алгоритмизация интеграции онтографов. Рассматривается проблема интеграции двух и более онтологий. В работе введено определение онтографа, а также представлен план алгоритма объединения онтографов и разработан первый этап алгоритмизации интеграции онтографов.*

**Общая постановка проблемы.** Новым шагом развития междисциплинарных научных исследований, и теории баз знаний в частности, должно стать теоретически обоснованное объединение (интеграция) уже разработанных как общедоступных онтологий, так и коммерческих баз знаний для разнообразных прикладных задач, проблем, целых предметных областей и трансдисциплинарных знаний общего характера. Сущность интеграции сформулирована следующим образом: «Устойчивые знания совокупности научных дисциплин можно представить в форме интегрированной иерархической сети научных теорий (разного уровня развития, содержательности и охвата действительности), составляющие которых возможно связаны посредством общих объектов действительности».

**Цель работы** состоит в разработке операций и построения алгоритма *объединения онтографов*, в частности первого этапа - сравнения, который заключается в описании предметных дисциплин (ПдД).

В работе впервые вводится определение онтографа ПдД и рассмотрены операции необходимые для объединения онтографа, определены основные этапы алгоритма объединения.

Понятие онтология происходит от др. греч. «онтос» - сущее, «логос» - учение, понятие, т.е. это раздел философии, изучающий бытие. Онтологию определяют как попытку всеобъемлющей и детальной формализации некоторой области знаний с помощью концептуальной схемы. Под концептуальной схемой подразумевается *набор понятий (концептов) + информация о понятиях* (свойства, отношения ограничения, аксиомы и утверждения о понятиях необходимых для описания процессов решения задач в избранной предметной области).[1]

Формально онтология определяется как  $O = \langle X, R, F \rangle$ , где  $X$  - конечное множество понятий предметной области,  $R$  - конечное множество отношений между понятиями,  $F$  - конечное множество функций интерпретации (совокупность значений, придаваемых элементам). [3]

Онтологические системы строят на основе таких принципов [2]:

- формализации, что подразумевает описание объективных элементов реальности с применением единственных, строго определенных, образцов (сроков, моделей и тому подобное);
- использования ограниченного количества базовых сроков (сущностей), на основе которых конструируют все другие понятия;
- внутренней полноты;
- логической несогласованности.

В отличие от обычного словаря, для онтологической системы характерны внутреннее единство, логическая взаимосвязь и несогласованность используемых понятий.

Поскольку онтологии были ответом науки на нужды своего времени, то и появление их произошло сразу в нескольких областях знания. Соответственно, в каждой из них ресурсы онтологического типа были сформированы по своим, специфичным для области знаний, правилам. Эту же точку зрения разделяют авторы работы [2], которые считают, что в проектировании и исследовании онтологий условно можно выделить два направления, до некоторого времени развивавшихся отдельно.

Первое связано с представлением онтологии как формальной системы, основанной на математически точных аксиомах (т.е. это ресурсы онтологического типа, созданные в различных областях математики).

Второе направление развивалось в рамках компьютерной лингвистики и когнитивной науки. Здесь онтология понималась как система абстрактных понятий, существующих только в сознании человека, которая может быть выражена на естественном языке (или средствами какой-то другой системы символов). При этом обычно не делается предположений о точности или непротиворечивости такой системы.

Таким образом, существует два альтернативных подхода к созданию и исследованию онтологий. Первый (формальный) – основан на логике (предикатов первого порядка, дескриптивной, модальной и т.п.). Второй (лингвистический) – основан на изучении естественного языка (в частности, семантики) и построении онтологий на больших текстовых массивах, так называемых корпусах.[1]

Выделяют различные уровни интеграции онтологий, в зависимости от числа изменений, которые необходимо сделать, чтобы получить некую общую онтологию из частных [8]:

- соответствие (alignment). Соответствие есть отображение понятий и отношений одной онтологии на другую. Соответствие может быть определено не полностью, так может существовать несколько понятий в одной онтологии, не имеющих своих эквивалентов в другой. Иногда для приведения онтологий в соответствие в них добавляют новые подклассы и надклассы понятий. Никаких других изменений аксиом, определений, доказательств или вычислений не производится;

- частичная совместимость (partial compatibility). Частичная совместимость есть соответствие онтологий, которое поддерживает также эквивалентные выводы и вычисления для всех эквивалентных понятий и отношений. Если две онтологии являются частично совместимыми, то любой вывод или вычисление, которые могут быть выражены в одной онтологии с использованием только соответствующих понятий и отношений, могут быть транслированы в эквивалентный вывод или вычисление в другой онтологии;

- унификация (unification). Унификация есть взаимно-однозначное соответствие всех понятий и отношений в двух онтологиях, которое позволяет любой процесс вывода или вычислений, выраженных в одной онтологии, отображать в эквивалентный процесс вывода или вычислений в другой. Обычным способом унификации двух онтологий является усовершенствование каждой из них в более детальные онтологии, чьи категории взаимно-однозначно эквивалентны.

Мы будем рассматривать онтологию данных как средство разносторонней и детальной формализации знаний о ПдД с помощью онтографа ПдД. [4] Введем определение: онтограф ПдД - это представление онтологии, имеющее графическую интерпретацию и понимающие под собой тройку  $S = \langle A, B, F \rangle$ , где  $A$  - конечное множество вершин, соответствующих понятиям предметной области,  $B$  - конечное множество ребер между вершинами, соответствующих отношениям между понятиями,  $F$  - конечное множество функций интерпретации (совокупность значений, придаваемых вершинам и ребрам).

При интеграции онтографов могут возникнуть несогласованности:, связанные с совпадениями и различиями содержания, видов, отношений и классов.

**Постановка задачи** объединения онтографов. Даны два онтографа  $O^1, O^2$ , требуется произвести онтограф  $O^3$ , который представляет понятия из двух исходных онтографов, а также дополнительные ограничения и взаимосвязи, если они требуются.

Решение этой задачи можно осуществить просмотром двух онтографов  $O^1, O^2$ , обнаружением синонимов, разрешением противоречий и созданием онтографа  $O^3$ . Этот процесс можно, автоматизировать. Для любых двух онтографов, существует много способов их объединения без создания противоречий. В общем случае, почти любое понятие в одном словаре «Определения понятий» *может быть* связано с любым понятием в другом словаре «Определения понятий» для второго онтографа. По этой причине онтографы обычно не могут объединяться без некоторых ограничений со стороны пользователя.

В ходе построения искомого онтографа предлагается использовать следующие логические операции: объединение, слияние, разность, Пересечение, ограничение, переименования, ограничение, переименования, присваивание, соединение

Алгоритм объединения онтографов можно разделить на три этапа,

1. Этап сравнения онтографов
2. Этап объединения онтографов (составление формул математической логики, применение аксиом и теорем, объединение с помощью логических операций )
3. Проверка результата на корректность (применение функций интерпретации, т.к. они показывают истинность или ложность объединения).

Сравнение онтографов разобьем на три пункта:

1. Сравнение по именам
2. Сравнение по классам
3. Сравнение по видам

**Сравнение по имени.** Наиболее простым способом сравнения классов является *сравнение по имени*. Данный подход является основным критерием большинства существующих модулей объединения онтологии. В наиболее развитых из них существует возможность нечеткого сравнения имен классов и сравнения с учетом синонимов, однако в случае сложных синонимичных и омонимичных конструкций сравнение классов по данному критерию дает неверные результаты. Причиной таких ошибок является предположение о роли класса в онтологии по его имени, тогда как имя не определяет класс (как термин не определяет понятия), а служит удобной для человека меткой класса.[5] При сравнении имен необходимо обратить внимание на содержание, которое в свою очередь включает отношения, функции интерпретации, аксиомы и определения.

**Сравнение по классам.** Классы используются в широком смысле. Классы — это абстрактные группы, коллекции или наборы объектов. Они могут включать в себя экземпляры, другие классы, либо же сочетания и того, и другого. [1]. Например, классы «Технология БД» и Технология файловой системы являются подклассами класса «Информационные технологии», который в свою очередь включен в класс «Информатика».

При сравнении отношений эквивалентности *классов* можно наблюдать отношение родовидовой зависимости: вместо взаимного позиционирования двух классов как эквивалентных или различны рассматриваются классы как пару эквивалентных, пару частное-общее, общее-частное, пару классов с общей частью или же пару полностью различных классов (без общей части).

**Сравнение по видам.** В свою очередь сравнение по видам является важным фактором, т.к. некоторые понятия можно разделить по характеру признаков, играющих роль видового отличия обобщаемых предметов. Виды понятий - это различные формы или способы, какими пользуются для того, чтобы выделить и выразить в мысли те или иные предметы. В многообразии видов понятий выражается активный и сложный характер отражения мира в мышлении, соответствующий сложности и многосторонности

познаваемой нами действительности. Можно обобщать в одном понятии многие предметы по отдельным их сторонам или эти отдельные стороны (качества, свойства, отношения предметов). Можно обобщать предметы по наличию тех или иных качеств, свойств, отношений и по отсутствию их и т.д.[7]

Например, понятие «компьютер» можно рассматривать:

- по назначению (игровые, пользовательские, сервера);
- по архитектуре;
- по форме (карманный персональный компьютер, ноутбук, нетбук, смартфон, сервер) и др.

**Сравнение по отношениям.** Отношения представляют тип взаимодействия между понятиями предметной области, поэтому в объединении онтографов необходимо учитывать типы отношений, т.к. типы отношений между понятиями могут быть любыми. Среди них можно выделить следующие группы:

1) *категориальные отношения;*

2) *логические отношения* (конъюнкция, дизъюнкция, отрицание, импликация);

3) *теоретико-множественные отношения* включают в себя отношения типа "множество — подмножество" ("род — вид", "класс — подкласс"), "целое — часть", "элемент - множество" и др.

4) *квантифицированные отношения*

Чаще всего применяются следующие основные связи (отношения): "*род — вид*", "*является представителем*", "*является частью*". Наличие связи типа "род — вид" между обобщенными объектами *A* и *B* означает, что понятие *A* более общее, чем понятие *B*. Любой объект, отображаемый понятием *B*, отображается и понятием *A*, но не наоборот. Связь "является представителем" существует обычно между обобщенным и индивидуальными объектами, когда последний выступает в роли представителя некоторого класса. Экземпляр может быть представителем нескольких обобщенных объектов. В этом случае ему присущи свойства нескольких обобщенных объектов, что соответствует множественному наследованию.

Рассмотренные выше типы отношений ("род — вид", "быть представителем", "быть частью", падежные отношения) образуют основу для решения объединения отношений онтографов. [6].

**Выводы.** В данной статье рассматриваются некоторые аспекты проблемы объединения онтографов ПдД. Обозначены основные факторы, которые необходимо учитывать при сопоставлении онтографов различных онтологий ПдД, такие как сходство символических имен, положение в иерархии терминов и набор необходимых и достаточных свойств.

Введено понятие онтографа ПдД. Первым этапом объединения онтографов является их сравнение, это достаточно емкий процесс, т.к. требуется рассмотреть сравнение по различным фактам. Рассмотренный первый этап алгоритма объединения двух онтологий позволяет теоретически автоматизировать формирование общей онтологической модели и поддерживать логическую согласованность общей объединенной онтологии.

Дальнейшая работа предусматривает продолжение разработки второго и третьего этапов: объединения онтографов и проверки результатов на корректность.

#### Литература

1. <http://www.intuit.ru/department/expert/ontoth/1/2.html>
2. Никоненко А.А., Обзор баз знаний онтологического типа. [Электронный ресурс]
3. <http://www.nbuu.gov.ua/portal/natural/ii/2009/Nikonenko AA.pdf>
4. Тихонов Ю.Л., Петренко Н.Г. Знание ориентированные системы разработки электронных курсов. // Information model of knowledge. Kiev – 2010. С.304