

Работа выполняется при поддержке гранта РФФИ (проект офи-м 11-07-12058).

Вычислительный метод определения семантической близости концептов неоднородных онтологий

О.А. Бубарева, Ф.А. Попов
БТИ (филиал) АлтГТУ, г. Бийск

Для обеспечения интеграции разнородных информационных систем (ИС) актуально использование онтологического подхода [1]. В процессе интеграции онтологий, построенных на объектных схемах гетерогенных ИС, порождается ряд конфликтов, в частности: использование различных терминов для обозначения одних и тех же понятий; различного рода семантические конфликты; одни и те же сущности реального мира представляются в разных источниках разными структурами данных. Поэтому, для обеспечения семантически корректной интероперабельности неоднородных ИС в контексте предметной области задачи, необходимо выявить общность и различия онтологий, лежащих в их основе, согласовать неоднородные онтологические спецификации и на базе соответствий онтологических контекстов, осуществлять преобразование информации.

Для численной оценки семантической близости концептов онтологий авторами выбран подход, основанный на результатах исследований доктора наук, профессора университета Мангейма А. Maedche. В соответствии с этим рассматриваются атрибутивная, таксономическая и реляционная меры, результаты измерений с использованием каждой из них с учетом весовых коэффициентов и используются для комплексной оценки семантической близости.

Модификация данного метода заключается в способе нахождения атрибутивной меры и применении генетического алгоритма для определения весовых коэффициентов. При этом предлагается определять атрибутивную меру не как пересечение диапазонов числовых значений атрибутов концептов, а как отношение пересечения множеств атрибутов к объединению множеств атрибутов концептов. Основные преимущества предлагаемого подхода заключается в выявлении ключевых концептов для построения результирующей онтологии, устранении субъективности описаний понятий онтологии и зависимости от точек зрения разработчиков онтологий.

Определим $S^T(c_i, c_j)$ как мера близости двух концептов на основе их положения, $S^R(c_i, c_j)$ – мера близости двух концептов на основе сопоставления их отношений, $S^A(c_i, c_j)$ – мера близости двух концептов на основе сопоставления атрибутов и значений атрибутов концептов.

Мера близости $S(c_i, c_j)$ двух концептов c_i онтологии O и c_j онтологии O' определяется как:

$$S(c_i, c_j) = t \cdot S^T(c_i, c_j) + r \cdot S^R(c_i, c_j) + a \cdot S^A(c_i, c_j), \quad (1)$$

где t, r, a – весовые коэффициенты.

С учетом того, что $t, r, a \in [0; 1]$, $t + r + a = 1$, $S(c_i, c_j) \in [0; 1]$, причем если концепты идентичны $c_i = c_j$, тогда $S(c_i, c_j) = 1$, если концепты различны и не имеют общих характеристик, тогда $S(c_i, c_j) = 0$.

Для автоматического определения параметров t, r, a используется генетический алгоритм, где индивид представляется в виде тройки генов (t, r, a) . В роли функции приспособленности выступает целевая функция 1.

К сформированной популяции потенциальных решений со следующими ограничениями $t, r, a \in [0; 1]$, $t + r + a = 1$ применяются стандартные операторы отбора, кроссовера и мутации.

Критерий выбора: максимизация суммы мер семантической близости между концептами двух онтологий.

$$f_{t,r,a} = \sum_{c_i, c_j \in C} S(c_i, c_j)$$

Для выделения меры семантической близости, при которой концепты эквивалентны, необходимо выбрать пороговое значение меры близости. Разработан метод определения критерий подобия концептов для классификации отображений в пять групп: эквивалентность, частичная эквивалентность, обобщение, уточнение, неопределенность.

Проведен вычислительный эксперимент, в результате которого была создана интегрированная онтология, позволившая в короткие сроки объединить локальные базы данных упомянутых систем, исключить дублирование, а также обеспечить целостность и непротиворечивость представленных в них сведений.

Библиографический список

1. Бубарева О.А., Попов Ф.А., Ануфриева Н.Ю. Решение проблемы интеграции данных при построении интегрированной автоматизированной информационной системы вуза // Международный журнал экспериментального образования. – 2011. – №5. – С. 92-95.

Облачный сервис интеллектуального анализа данных

*А.Н. Вязьмина**, *С.С. Киргизов***, *С.И. Жилин**,
*П.А. Ледомский**, *Н.П. Мусиенко**, *П.В. Нуждин**,
*В.Д. Пятков**

** АлтаГУ, г. Барнаул*

*** Университет Пьера и Марии Кюри, г. Париж*

Системы интеллектуального анализа данных находят применение во многих отраслях, требующих извлечения новых знаний из данных и представления скрытых в данных закономерностей в сжатом виде с целью их использования для прогнозирования и принятия решений [1].

Сдерживающими факторами широкого внедрения технологий интеллектуального анализа данных являются:

- рост объемов подлежащих обработке и анализу данных;
- неравномерность в распределении по времени потребностей в использовании систем интеллектуального анализа данных и необходимых для них вычислительных ресурсов;
- необходимость участия экспертов на этапе построения и сопровождения моделей;
- неразвитость инфраструктуры для организации совместной работы.

Целью настоящей работы является выработка принципов построения и реализация облачного сервиса интеллектуального анализа и обработки данных в соответствии с моделью «программное обеспечение как услуга» [2].

Облачный сервис интеллектуального анализа данных должен обладать следующими основными функциональными возможностями:

- редактор потоковых сценариев анализа данных [3], позволяющий визуально запрограммировать в виде графа последовательность шагов решения задачи;
- облачное хранилище наборов данных и сценариев;
- возможность расширения набора базовых элементов сценария пользовательскими модулями;