

УДК 004.822

ПОДХОД К КОЛЛЕКТИВНОЙ РАЗРАБОТКЕ ОНТОЛОГИЙ

А.В. Невидимов, И.А. Бессмертный

Рассмотрен подход к коллективной разработке онтологий на основе идеологии систем контроля версий с центральным хранилищем. Предложены решения для проблем обеспечения обратной совместимости онтологий, а также для решения конфликтов вносимых экспертами правок.

Ключевые слова: онтологии, совместная разработка, системы контроля версий.

При разработке информационных систем все более актуальными становятся семантические технологии [Л]. Структурирование знаний в виде онтологий позволяет уменьшать трудозатраты на обработку информации и принятие решений практически во всех предметных областях. Для получения максимально качественной онтологии той или иной предметной области разумным представляется задействовать не одного, а нескольких экспертов. В этой связи возникает задача обеспечения распределенной коллективной работы над онтологиями, и ее решение сопровождается следующими сложностями:

- разные эксперты могут иметь разные точки зрения на предметную область, что приводит к различиям в онтологиях этой предметной области, созданных разными экспертами. Это приводит к необходимости обеспечить слияние онтологий;
- параллельно с разработкой онтологии может разрабатываться программное обеспечение, использующее эту онтологию в качестве источника знаний. Прикладные программы должны использовать именно ту версию разрабатываемой онтологии, с которой они совместимы.

В настоящей работе предлагается подход, основанный на идеологии систем контроля версий с центральным репозиторием, таких как Git, Mercurial и SVN. Приведем список ключевых правок онтологий в таком подходе с их описанием и особенностями.

- Добавление новых триплетов вида <Сущность–субъект, Предикат, Сущность–объект> не нарушает обратной совместимости онтологии с ее предыдущей версией, так как все множество фактов, которые могли быть получены из старой версии онтологии, могут быть получены и из новой. Важным является также необходимость определять случаи параллельного добавления разными экспертами схожих сущностей для исключения дублирования данных в онтологии.
- Удаление старых триплетов в некоторых случаях может привести к нарушению обратной совместимости. Факт наличия или отсутствия обратной совместимости предлагается определять путем попытки логического вывода всех удаленных триплетов из новой онтологии, и если хотя бы один триплет не может быть выведен, то считается выявленным нарушение обратной совместимости. Удаление триплетов может также привести к нарушению валидности правок, вносимых другим экспертом.
- Удаление элементов из множества сущностей или множества отношений всегда приводит к нарушению обратной совместимости и может привести к нарушению валидности правок, вносимых другим экспертом.

В работе предлагается алгоритм коллективной работы над онтологиями, позволяющий избежать возникновения конфликтов и противоречий:

- Перед внесением в основную ветку онтологии набора правок проверить на совместимость с текущим состоянием онтологии;
- Если результирующая онтология несовместима с исходной, сформировать конфликтный список и предложить автору правок внести соответствующие изменения;
- Если автор вводит новую сущность, выполнить поиск в онтологии схожих сущностей с помощью вычисления расстояния Левенштейна между их названиями или на основе данных ресурса WordNet о синонимичности понятий. Сформировать список схожих понятий и вернуть автору для корректировки правок;
- Проверить обратную совместимость новой версии онтологии. Если в данной ветви онтологии есть нарушения обратной совместимости, опубликовать новый URL, являющийся точкой доступа SPARQL-endpoint, что обеспечит прикладным программам возможность пользоваться последней версией онтологии, с которой они совместимы.

Предложенный алгоритм в настоящее время проходит апробацию в проекте создания онтологии для корпоративной информационной системы, обеспечивающей работу торговых площадок.

[Л]. Бессмертный И.А. Семантическая паутина и искусственный интеллект // Научно-технический вестник СПбГУ ИТМО. – 2009. – № 6 (64). – С. 77–83.

Невидимов Александр Валентинович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, alex.nevidimov@gmail.com

Бессмертный Игорь Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, igor_bessmertny@hotmail.com