

Гибкая подсистема визуализации онтологии и информационного наполнения порталов знаний на протяжении их жизненного цикла*

© З.В. Апанович¹, П.С. Винокуров¹, Т.А. Кислицина²

¹Институт систем информатики им. А.П. Ершова СО РАН, г. Новосибирск

²Новосибирский государственный университет
apanovich@iis.nsk.su

Аннотация

Рассматривается компонента визуализации наполнения информационных порталов, основанных на онтологиях. В этой компоненте реализован принцип декомпозиции, на основе отношений. Пользователь имеет возможность выбирать одно или несколько отношений, просматривать классы или объекты, связанные выбранными отношениями, а также выбирать методы визуализации, учитывающие типы конкретных отношений, и комбинации отношений разного типа.

1 Введение

Процесс разработки онтологии и информационного наполнения веб-портала является весьма длительным и трудоемким и требует усилий большого коллектива разработчиков. Многие данные вводятся вручную, что потенциально опасно ошибками ввода, которые нелегко обнаружить, просматривая одну за другой текстовые формы. По мере своего развития информационные порталы накапливают все больше разнородной информации и становятся бесценным источником информации, пригодной для всевозможных форм анализа. Поэтому возникает задача анализа развивающегося информационного наполнения с целью выявления различных закономерностей. Так, например, в случае информационного портала, посвященного некоторому научному направлению, весьма актуальной является задача наукометрического анализа его наполнения, в частности, выявления шаблонов сотрудничества между различными научными сообществами.

Общепризнанным инструментом, обеспечивающим понимание больших объемов абстрактной информации, является визуализация информации с применением графовых моделей [1, 4, 6]. Онтоло-

гия, составляющая основу информационного портала, может быть представлена в виде графа, вершины которого изображают сущности, такие, как классы, объекты и атрибуты онтологии, а ребра изображают отношения между этими сущностями. Просматривая изображение различных подграфов, генерируемых по описанию данной онтологии и информационного наполнения, разработчик системы может обнаружить как ошибки данных, возникшие при ручном вводе информации, так и ошибки проектирования, незаметные при просмотре текстовых форм.

Следует отметить, что в настоящее время имеется значительное количество работ, посвященных визуализации онтологий. Эти работы развиваются в двух базовых направлениях. Во-первых, постоянно возникают новые методы визуализации и навигации для графов общего вида. Затем эти методы начинают применяться для визуализации различных предметных областей, в том числе и онтологий. Примером движения по этому пути является использование для визуализации онтологий таких известных программ общего назначения, как GraphViz [7]. С другой стороны, онтологии имеют четко определенную структуру, которая должна приниматься во внимание при визуализации. Поэтому такие известные редакторы онтологий, как Protege и OntoEdit [11, 15], используют специализированные методы визуализации. Достаточно полный обзор методов визуализации онтологий представлен в работе [12].

Несмотря на значительное количество существующих методов визуализации онтологий и баз знаний, эта проблема далека от своего окончательного решения. Это объясняется сложностью и большим объемом визуализируемой информации, а также тем, что каждая предметная область имеет особенности, требующие специализированных методов визуализации. Одним из важнейших условий успешной визуализации является разработка методов, зависящих от семантики визуализируемых связей и объектов.

Существенной особенностью информационного наполнения порталов знаний является наличие большого количества информационных объектов и онтологических отношений. Тестовые данные, с

Труды 12^й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2010, Казань, Россия, 2010

которыми осуществлялась большая часть наших экспериментов, содержали четыре тысячи онтологических объектов, связанных пятнадцатью тысячами онтологических отношений. Для обеспечения понимаемости таких объемов информации необходимо предоставить пользователю гибкие средства декомпозиции информационного наполнения. Средства декомпозиции должны позволять, с одной стороны, выбирать осмысленные фрагменты информации, объем которых не превышает лимитов их понимаемости. С другой стороны, при декомпозиции не должно происходить потери существенной информации. В основу нашего подхода была положена декомпозиция информационного наполнения, основанная на отношениях. Пользователь имеет возможность выбирать одно или несколько отношений, просматривать классы или объекты, связанные выбранными отношениями, а также выбирать методы визуализации, учитывающие типы конкретных отношений и комбинаций отношений разного типа.

2 Пользовательский интерфейс компоненты визуализации

Компонента визуализации информационного наполнения принимает входные данные в виде двух xml-файлов, первый из которых (ontology.xml) содержит информацию о классах и отношениях классов, а второй файл (data.xml) содержит информацию о конкретных объектах и их отношениях. Эти данные преобразуются во внутренний формат, после чего на экране появляется окно управления визуализацией. Использование xml-файлов не является существенным ограничением. В ходе работы проводились также эксперименты с данными, представленными в RDF-формате.

Структура пользовательского интерфейса подчинена основной задаче: дать пользователю как можно больше возможностей анализа информационного наполнения портала. С этой целью были реализованы следующие решения:

- пользовательский интерфейс состоит из набора закладок, соответствующих основным «средам» онтологии и информационного наполнения портала;
- пользователь имеет возможность выбирать методы визуализации, учитывающие типы конкретных отношений и комбинаций отношений разного типа;
- имеется широкий набор возможностей взаимодействия с информационным наполнением, позволяющий выбирать нужные подграфы указанием интересующих отношений между классами или объектами, а также содержащий набор геометрических трансформаций изображения выбранных подграфов;
- по усмотрению пользователя на экране могут быть размещены названия выбранных вершин, отношений, значений атрибутов;

- реализованы дополнительные окна описаний атрибутов вершин, выбираемых в окне визуализации.

Общий вид панели управления визуализацией показан на рис. 1. В верхней части окна визуализации находится семь закладок, соответствующих основным режимам визуализации: **Классы**, **Отношения**, **Ассоциативность и вложенность**, **Отношения между объектами**, **Отношения между классами**, **Размеры классов**, **Дерево классов**.

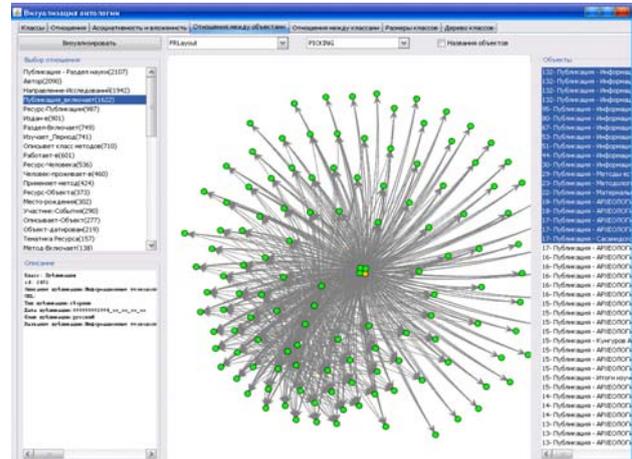


Рис. 1. Общий вид панели визуализации

Закладки **Дерево классов** и **Размеры классов** соответствуют двум различным режимам визуализации отношения наследования между классами.

Закладка **Отношения между классами** позволяет получить изображение любого выбранного подмножества отношений между классами.

Закладки **Отношения между объектами** и **Ассоциативность и вложенность** позволяют просматривать изображения любого выбранного подмножества отношений между объектами и комбинаций ассоциативных отношений с отношениями вложенности.

Закладки **Классы** и **Отношения** являются вспомогательными и используются для настройки параметров визуализации.

Далее будет показано, каким образом указанные режимы помогают пользователю разобраться в структуре портала знаний и его информационном наполнении.

3 Изображение связей между классами онтологии

Знакомство с информационным наполнением портала предлагается начинать с изображения отношений наследования между классами. Одним из стандартных способов представления этой информации являются списки с отступами, как это показано на рис. 2.

Такой способ изображения имеет несколько недостатков. Во-первых, для того чтобы получить глобальную информацию об онтологии, необходи-

мо раскрыть все уровни этого иерархического списка, но при таком раскрытии список очень быстро становится необозримым. Во-вторых, список с отступами не дает никакой информации о прочих отношениях, существующих между элементами информационного наполнения. Поэтому в нашей подсистеме визуализации реализовано еще несколько дополнительных возможностей для изображения отношения наследования. Пример изображения отношений наследования при помощи радиального алгоритма показан на рис. 3.

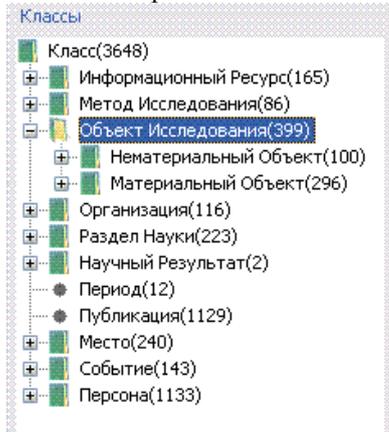


Рис. 2. Изображение отношения наследования классов в виде списка с отступами

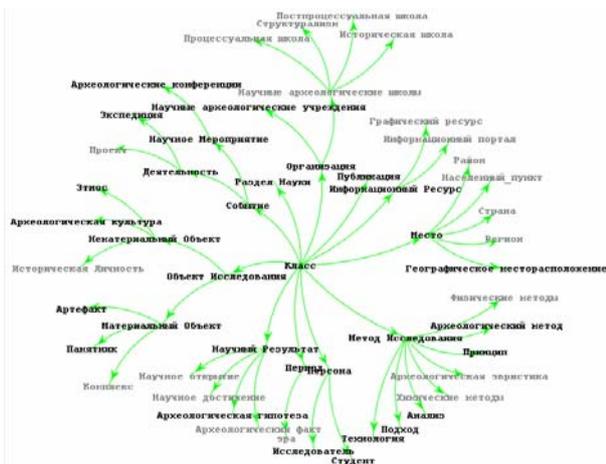


Рис. 3. Изображение отношения наследования при помощи радиального алгоритма

Такой способ изображения позволяет пользователю визуально оценить сбалансированность онтологии. Кроме этого, названия пустых классов, в которых еще нет объектов, высвечиваются более светлым цветом. Это сделано для того, чтобы привлечь внимание пользователя к таким классам и дать ему возможность подредактировать онтологию.

В подсистеме визуализации реализована также версия радиального алгоритма, в которой размер секторного сегмента, изображающего каждый класс, пропорционален количеству объектов в этом классе (рис. 4).

Помимо отношения наследования в онтологии информационного портала используются отноше-

ния партономии, большой спектр транзитивных отношений (к которым, в частности, относится отношение включения), а также большое разнообразие ассоциативных отношений. Возможность просмотра ассоциативных отношений между классами в дополнение к уже изображенному отношению наследования является весьма важной для понимания структуры портала. В настоящий момент эта возможность предоставляется пользователю в двух режимах.

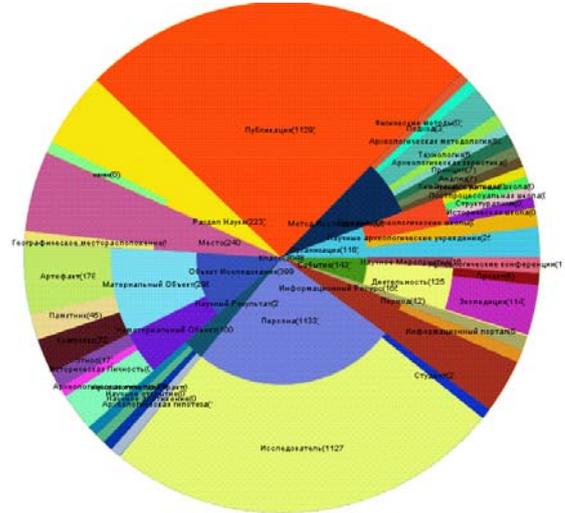


Рис. 4. Изображение относительного количества объектов в каждом классе

В режиме «Наследование классов» достаточно выбрать при помощи мыши произвольный класс на изображении всей онтологии, и в результате этого действия к изображению будут добавлены ребра с метками, изображающие ассоциативные связи выбранного класса, как это показано на рис. 5. Одновременно на панели «Связи», расположенной слева от окна визуализации, будут высвечены названия этих отношений с указанием количества экземпляров данного отношения.

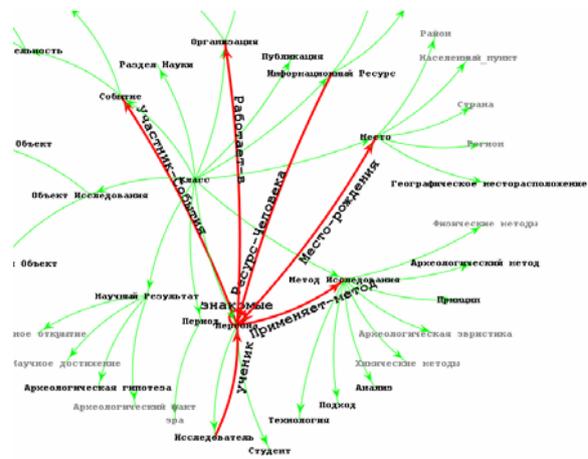


Рис. 5. Иерархия классов с высвеченными ассоциативными отношениями класса **Персона**

выбрать при помощи мыши интересующие отношения на диаграмме классов и перейти в режим визуализации отношений между объектами. Можно также выбрать нужные отношения из списка имеющихся отношений, перечисленных в компоненте выбора отношений. При выборе мышью одного или нескольких отношений из предлагаемого списка отношений справа появляется список объектов, связанных выбранным отношением. Из списка объектов можно выбрать произвольное подмножество и заказать его изображение. Программа разделяет граф связей объектов на компоненты связности, а затем применяет к ним один из имеющихся алгоритмов размещения. В настоящий момент для визуализации имеющихся данных используются алгоритмы Kamada-Kawai [10] и Fruchterman-Reingold [8].

Визуализация объектов и связей между ними оказалась весьма удобным инструментом, позволяющим быстро идентифицировать как ошибки, возникающие при ручном вводе информационного наполнения портала, так и ошибки, возникшие на этапе проектирования онтологии. Часто признаком ошибочного ввода данных является наличие циклов в графе связей объектов, а также наличие петель, то есть ребер, соединяющих вершину с самой собой. Такие ошибки обнаруживаются при визуализации отношений между объектами одного и того же класса (петля на графе ассоциативных отношений между классами). Например, при визуализации отношения *Исторически следует* для объектов класса **Период** был обнаружен цикл, связывающий объекты «Нижний палеолит», «Средний палеолит» и «Верхний палеолит» (рис. 8).

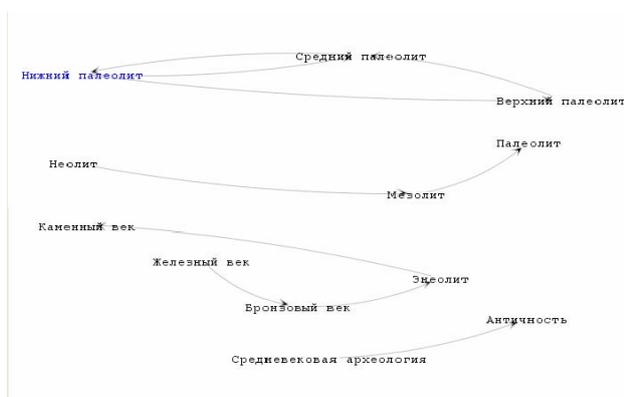


Рис. 8. Наличие циклов в подграфе объектов типа **Период**, связанных отношением *Исторически следует*

В настоящий момент единственным способом обнаружения подобных ошибок является просмотр всех имеющихся отношений по одному, а также в комбинации с отношениями вложенности, что может оказаться весьма длительной и трудоемкой проблемой. Поэтому желательно проанализировать заранее семантику имеющихся в онтологии отношений и выделить те средства анализа, которые позволяли бы автоматически определять коррект-

ность ввода данных. Имеющиеся на данный момент методы визуализации позволяют, по крайней мере, сформулировать критерии для такой автоматической проверки.

5 Изображение комбинации отношения вложенности и ассоциативных отношений

Другой важной особенностью нашей компоненты визуализации является возможность комбинированного изображения отношений. Для этого введен режим визуализации «**Ассоциативность и вложенность**». Дело в том, что в онтологиях порталов знаний имеется большое количество ассоциативных отношений, связывающих объекты, которые являются частью иерархии, индуцированной отношением вложенности (партономии). Например, имеется большая иерархия объектов из класса **Метод исследования**, связанных отношением *Метод Включает*. В то же время объекты класса **Метод исследования** связаны ассоциативным отношением *Применяет-Метод* с объектами класса **Исследователь**. Отношения партономии обычно достаточно просто обнаружить визуально на уровне изображения отношений между классами. Эти отношения (но не только они) изображаются ребром-петлей. Если визуализировать соответствующий граф без учета отношения вложенности, изображение может оказаться весьма запутанным, как это показано на рис. 9. Эта запутанность частично связана с тем, что отношение вложенности (в данном случае *Метод Включает*) изображается при помощи ребер, соединяющих вершины-объекты. Если изобразить факт наличия отношения вложенности между объектами при помощи геометрической вложенности вершин-объектов, рисунок упростится благодаря исчезновению существенного количества ребер. Понимаемость структуры отношений при этом только улучшается, как это можно видеть на рис. 10.

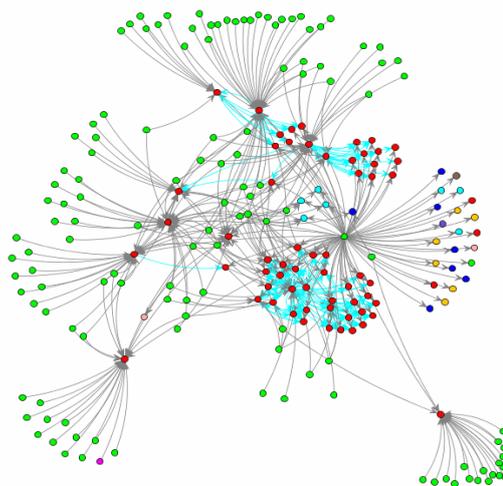


Рис. 9. Совместное изображение отношения партономии и одного из ассоциативных отношений при помощи силового алгоритма

Например, на рис. 11 изображено отношение соавторства между научными сотрудниками, работающими в разных городах. В качестве основы изображения используется круговое изображение дерева, соответствующего отношению *Место-включает*. Маленькие черные кружочки изображают географические объекты, такие, как страны, города, поселки и т. д. Прямолинейные ребра между этими объектами соответствуют отношению вложенности, т. е. ребро, соединяющее объект Россия с объектом Иркутск, соответствует факту, что Иркутск находится в России. Маленькие светлые кружочки изображают исследователей, а прямолинейные ребра, соединяющие исследователей с географическими объектами, соответствуют тому, что исследователь проживает в указанном месте. Светлые криволинейные шлейфы изображают отношения соавторства между исследователями из разных городов, а более тонкие и более темные шлейфы показывают отношения сотрудничества одного выбранного исследователя. Таким образом, можно видеть, что выбранный для анализа исследователь (Холюшкин) имеет соавторов в Новосибирске, Санкт-Петербурге, Красноярске и др., но не имеет соавторов в Москве. Это же изображение показывает и некоторые недоработки при введении данных. Например, все российские города располагаются на окружности, центром которой является вершина с названием Россия. Но город Москва расположен по периметру другой окружности. Значит, в базе данных не хватает информации о том, что Москва находится в России. Таким же способом можно исследовать зависимость отношения сотрудничества между исследователями, работающими в разных разделах науки, в различных научных организациях, применяющих разные методы исследования, и т. д.

Заключение

Подсистема визуального анализа информационного наполнения порталов знаний опробована на примере тестовых данных, описывающих знания по археологии, компьютерной лингвистике, а также сайта кафедры Программирования ММФ, основанного на онтологии. Опытная эксплуатация подсистемы показала ее полезность при работе с реальными данными, она действительно позволяет обнаруживать некоторые недоработки и ошибки ручного ввода в предоставленных данных, которые весьма затруднительно обнаружить в текстовом представлении знаний, а также при помощи стандартных методов навигации по большим графам. Разработанная подсистема визуализации может быть полезна как на этапе разработки онтологии информационного портала, так и в процессе всего жизненного цикла портала, упрощая тестирование и анализ информационного наполнения портала знаний.

Эксперименты с визуализацией информационного наполнения портала знаний показали необходимость дальнейшего развития подсистемы визуализации в нескольких направлениях. Во-первых, для

эффективной визуализации классов, содержащих большое количество объектов и отношений, необходимо дополнить существующие алгоритмы методами кластеризации и размещения кластеров. Во-вторых, необходимо исследовать типы отношений, имеющих в онтологии, и определить наиболее эффективные методы визуализации как для определенных типов отношений, так и для различных типов подграфов, выделяемых при анализе информационного наполнения портала знаний. И, в третьих, для эффективной работы с информационным наполнением портала необходима разработка специализированных методов анализа и верификации этого наполнения. При разработке первой версии подсистемы визуализации использовалась свободно распространяемая библиотека классов Java, называемая JUNG [8].

Литература

- [1] Апанович З.В. Методы навигации при визуализации графов// Вестник НГУ. – 2008. – Т. 6, Вып. 3. – С. 35-47.
- [2] Загоруйко Ю.А., Боровикова О.И., Холюшкин Ю.П. Построение предметной онтологии для археологического портала научных знаний// Информационные технологии в гуманитарных исследованиях. – 2006. – № 10.
- [3] Холюшкин Ю.П., Гражданников Е.Д. Системная классификация археологической науки (элементарное введение в науковедение). – Новосибирск, 2000. – 58 с.
- [4] Apanovich Z.V., Vinokurov P.S., Elagin V.A. An approach to visualization of knowledge portal content// Bulletin of NCC. – 2009. – Issue 29. – P. 17-32.
- [5] Barabasi A.-L. The origin of bursts and heavy tails in human dynamics// Nature. – 2005. – No 435. – P. 207-211.
- [6] Di Battista G., Eades P., Tamassia R., Tollis I.G. Algorithms for drawing graphs: an annotated bibliography // Computational Geometry, Theory and Applications. – 1994. – No 4. – P. 235-282.
- [7] Ellson J., Gansner E.R., Koutsofios L., North S., Woodhull G. Graphviz. Open source graph drawing tools proceedings//Graph Drawing. – 2002. – P. 483-484.
- [8] Fruchterman T.M.J., Reingold E.M. Graph drawing by force-directed placement software //Practice and Experience. – 1991. – V. 21 (11). – P. 1129-1164.
- [9] Holten D. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data // Transactions on Visualization and Computer Graphics. – 2006. – V. 5 (12). – P. 741-748.
- [10] Kamada T., Kawai S. An algorithm for drawing general undirected graphs// Information Processing Letters. – 1989. – V. 31. – P. 7-15.
- [11] Katifori A., Torou E., Halatsis C., Lepouras G., Vassilakis C. A comparative study of four ontology visualization techniques in Protege: experiment se-

- tup and preliminary results//Proc. of the Conf. on Information Visualization. – 2006. – P. 417-423.
- [12] Katifori A., Halatsis C., Lepouras G., Vassilakis C., Giannopoulou E. Ontology visualization methods – a survey//ACM Comput. Surv. – 2007. – V. 39 (4).
- [13] Madahain J.O., Fisher D., Smyth P., White Sc., Boey Y-B. Analysis and visualization of network data using Jung// J. of Statistical Software, VV(II).
- [14] Newman M.E.J., Girvan M. Finding and evaluating community structure in networks// Physical Review, E. – 2004. – P. 69.
- [15] Sure Y., Erdmann M., Angele J., Staab S., Studer R., Wenke D. OntoEdit: collaborative ontology development for the Semantic Web// Proc. of the first Int. Semantic Web Conference, June 9 – 12 2002, Sardinia, Italia.
- [16] Wuchty S. Jones B., Uzzi B. The increasing dominance of teams in production of knowledge// Science Express. – 2007. – No 5827 (316). – P. 1036-1039.

A flexible subsystem of ontology and content of knowledge portals visualization throughout their life cycle

Z.V. Apanovich, P.S. Vinokurov, T.A. Kislicina

The process of development of an ontology-based knowledge portal and creation of its content is time-consuming and labor-intensive. The life cycle of such portals is sufficiently long and they collect plenty of valuable information. This paper describes a subsystem for visual analysis of content of ontology-based knowledge portals during their life cycle. This analysis is based on a flexible decomposition strategy. Browsing of the content of a knowledge portal is organized as a multilevel stepwise process. Appropriate placement algorithms are used at each step of this process. They take into account certain types of ontological relations and their combinations.

* Работа выполнена при финансовой поддержке РФФИ (проект 09-0700400) и РАН (проект 2/12)

Благодарности. Авторы выражают благодарность Ю.А. Загоруйко и С.В. Булгакову за предоставленные тестовые данные в xml-формате