
О ПРИМЕНЕНИИ ОНТОЛОГИЙ ДЛЯ ПРЕДСТАВЛЕНИЯ ПРОГРАММ УЧЕБНЫХ КУРСОВ

Евгений Еремин

Abstract: Publication describes the experience in application of ontology technique to structuring of educational materials. Several topics of physics were formalized by means of Protégé software tool. Some principal problems in building of knowledge structure were found, so the discussion may interest not only ontology users, but also the developers of ontology tools.

Keywords: ontology, program, education, course, Protégé, knowledge structure, objects, inheritance, classes.

ACM Classification Keywords: K.3.1 Computer Uses in Education – Computer-assisted instruction; E.1 Data Structures – Trees; E.2 Data Storage Representations – Object representation; I.2.4 Knowledge Representation Formalisms and Methods – Frames and scripts; I.2.6 Learning – Knowledge acquisition.

Введение

В последние годы для структурирования накопленных человеком обширных знаний и представления их в машинной форме все более широко используются **онтологии** – формальные описания терминов предметной области и отношений между ними. Данный метод компьютерного представления информации уже сейчас имеет некоторые вполне законченные прикладные применения в самых разнообразных областях. Не менее важны и перспективные теоретические разработки, направленные на обеспечение семантически корректного поиска в документах (включая подбор информации сетевыми агентами) и автоматическое формирование онтологий по имеющимся ресурсам. Конкретные примеры создания онтологий и их практического использования неоднократно описывались в литературе [1-4].

Будучи своеобразным общим словарем понятий, онтологии существенно облегчают взаимопонимание людей при совместном использовании информации. К тому же представление данных в форме онтологий не просто обеспечивает возможность их автоматической обработки, но и позволяет людям наиболее наглядно и отчетливо сформулировать свои знания по выбранной тематике.

Последнее обстоятельство с точки зрения образования представляет огромный интерес. Обучение, являясь процессом целенаправленной передачи знаний, вообще может служить весьма естественной областью для использования онтологий (см., например, [3]). В частности, от успехов в структурировании знаний и представлении их в машинном виде во многом зависит построение эффективных автоматизированных обучающих систем.

В данной работе делается попытка применить онтологии к структурированию материала некоторых тем учебного курса. Традиционно многие примеры подобных онтологий разрабатываются для языков программирования (см., например, публикации [5, 6]), где все получается легко и изящно, поскольку языки программирования есть конструкция искусственная и, следовательно, хорошо структурированная. В [5] даже ставится вопрос о гармоничности создаваемой онтологии. Однако в других предметных областях все выглядит не так гладко, о чем свидетельствуют приведенные в [7, 8] результаты анализа онтологий учебных курсов из различных дисциплин.

В качестве структурируемого материала в работе используются отдельные разделы курса физики (для простоты выбран школьный учебник для старших классов). Хотя тематика взята исходя из личных предпочтений автора и его базового образования, есть данные в пользу того, что физика – это хорошее поле для применения онтологий. С рассматриваемых позиций дисциплина интересна тем, что объективным образом отражает сложность и взаимосвязь природных явлений и, кроме того, структурирование базовых знаний в учебном курсе физики во многом определяет стратегию обучения [9, 10]. Вследствие важности систематизации физических знаний, в статье [10] для подготовки учителей предложен специальный учебный курс, где ставится цель «не давать дополнительные знания по физике, но упорядочить то, что уже изучено».

Для создания и ведения учебных онтологий использовалось программное обеспечение **Protégé** версии 3.1.1 (Protégé-2005). Его выбор во многом определялся известностью данного пакета, а также наличием доступных и довольно подробных описаний [11-13, 4].

Постановка проблемы

Рассматриваемая в данной публикации проблема возникла весьма естественным образом из повседневной педагогической практики. Как известно, к любому преподаваемому курсу существует программа – некий перечень тем и вопросов, которые подлежат изучению (в России программы курсов носят характер официального документа, хотя и не являются догмой). В случае, когда программа написана для давно преподаваемого курса, поддержанного целым набором проверенных временем учебников разных авторов (скажем, физика или геометрия), ее воплощение в жизнь особой сложности не представляет. Более того, как свидетельствует опыт, изменения в таких программах происходят не так часто и, как правило, носят несущественный характер. С современными бурно развивающимися компьютерными дисциплинами дело обстоит по-другому: программы курсов периодически меняются, отсюда учебники часто неполны и не успевают своевременно отражать происходящие в науке изменения, опыт преподавания вновь возникающих разделов отсутствует. В результате вопросы, которые попали в программу при ее очередном обновлении, незнакомы учителю, и он не всегда четко представляет себе, где искать ответы на них. В таких случаях хочется видеть хотя бы две-три фразы, расшифровывающие вопрос, и список рекомендуемых по ним источников информации. Итак, возникает вполне естественное желание иметь возможность получить дополнительную информацию по любому заинтересовавшему учителя пункту программы, т.е., пользуясь педагогической терминологией, к каждому вопросу должен быть добавлен краткий конспект (аннотация) материала.

Одним из возможных решений проблемы с помощью компьютера является построение онтологии учебного курса, которая содержит в себе комментарии и библиографические ссылки по каждому вопросу. В качестве дополнительного преимущества онтологического подхода (по сравнению с «бумажным» решением проблемы) мы получим связи между отдельными элементами курса, что весьма полезно при организации своевременного повторения и планировании последовательности изложения материала.

Как уже говорилось выше, для создания онтологии был выбран курс физики, причем в качестве основы учебного материала взят стабильный школьный учебник физики [14]. Вопросы, связанные с возможностью иного изложения материала сейчас рассматривать не будем. Пользуясь терминологией из работ по онтологиям [6, 8], мы хотим построить онтологию учебного курса, но не соответствующей научной области, причем в упрощенном варианте – по одному учебнику.

Принципы построения онтологии

Учитывая некоторую неоднозначность терминологии в области онтологий, сначала перечислим основные термины, которые будут использоваться в обсуждении. Будем базироваться на той системе категорий, которая принята в выбранном программном средстве Protégé.

Онтологии состоят из классов, слотов и экземпляров. **Классы** (classes) описывают отдельные понятия предметной области, а **экземпляры** класса (instances) представляют собой конкретные примеры реализации объектов данного класса. **Слоты** (slots) – свойства и атрибуты, применяются как классам, так и их экземплярам. Например, в качестве классов учебного материала можно предложить выделить *понятие* или *закон*. Характерным слотом *понятия* может служить его *определение*, а для *закона* стоит предусмотреть слот для записи его *математического выражения*. Экземплярами *понятия* могут являться *электрический ток* или *заряд*, а *законов* – *закон Ома* или *закон Джоуля-Ленца*.

Заметим, что в Protégé между перечисленными категориями существуют строго определенные связи [12].

От класса могут быть порождены новые классы, а также конкретные экземпляры этого класса (некоторые классы, имеющие статус **абстрактных**, не допускают порождения экземпляров). Экземпляры являются конечными узлами иерархии, т.е. принципиально не могут иметь наследников. Существует (и в Protégé поддерживается) **множественное наследование классов**, когда класс происходит от нескольких родительских классов и при этом наследует все их слоты. В отличие от классов, экземпляр обязательно порождается от одного класса.

Слоты самостоятельны и напрямую не принадлежат какому-либо классу или экземпляру (это в принципе позволяет использовать один и тот же слот в разных ветвях иерархии). Они бывают двух типов – **собственные слоты** (own slots) и **слоты шаблона** (template slots). Первые являются собственностью данного класса или экземпляра и могут иметь индивидуальное значение. Вторыми может обладать только класс, причем все слоты шаблона передаются по наследству. Слоты шаблона по своей сути не могут иметь значений, пока не будут переданы конкретному экземпляру, где они, став собственными его слотами, приобретут свойство быть заполненными. Слоты могут дополняться только к классу, а экземпляры классов лишь способны их наследовать.

С каждым классом жестко связана интерфейсная **форма** (form), применяя которую пользователь будет заносить необходимые значения в слоты. Формы легко редактируются, что позволяет придавать им удобный для ввода вид.

В дальнейшем в ходе изложения тех трудностей, которые возникли при построении онтологии учебного материала по физике, станет понятно, зачем необходимо столь подробное описание системы связей между компонентами онтологии.

Реализация онтологии, возникающие трудности

Уже предварительный анализ предметной области для рассматриваемой нами задачи показывает, что для ее решения потребуется как минимум три разновидности классов:

- типы учебных материалов, которые определяют, какая именно информация хранится для каждой из разновидностей вопросов программы (например, для описательного материала или для закона);
- фундаментальные понятия, относящиеся непосредственно ко всему курсу физики и применяемые во всех его разделах (физическая величина, единица измерения, система единиц и т.п.);

- категории, непосредственно входящие в состав конкретного раздела физики (в качестве пробного материала был выбран раздел «Электричество и магнетизм»).

Наличие нескольких уровней классов, логически тесно связанных между собой, уже само по себе представляет определенную трудность. Существенная взаимосвязь некоторых физических понятий еще более осложняет ситуацию, поскольку она не всегда хорошо укладывается в иерархическую структуру. Приведем пример из области фундаментальных категорий физики. Для измерения *физических величин* используется *система единиц* измерения. Эта *система единиц* состоит из совокупности отдельных *единиц измерения*, причем для каждой *системы единиц* их набор конкретен. Некоторые *единицы измерения* могут входить в разные *системы единиц*. В то же время, каждая *физическая величина* может измеряться несколькими *единицами измерения*, а выбранная *единица измерения*, в свою очередь, зависит от *системы единиц*. Наконец, сам процесс измерения заключается в сравнении измеряемой *физической величины* с некоторой эталонной *физической величиной*, принимаемой за *единицу измерения*. Подчеркнем, что обойтись без перечисленных понятий никак нельзя, поскольку при изучении любой физической величины всегда возникает разговор о единицах ее измерения.

Трудно не согласиться с авторами [7], которые выделили классификацию учебного контента в виде строго иерархической структуры в качестве одной из главных трудностей. Причина трудности, по-видимому, принципиальна и лежит в природе самого изучаемого материала. Как отмечается в [15, 3], можно выделить пять типов отношений между терминами: часть/целое («part-whole»), например, бампер и автомобиль; сочетание («collocation»), например, слова в предложении; смысловая связь («paradigmatic relations»), например, Солнце и Солнечная система; синонимы и антонимы. В классической книге Г. Буча [16] дается несколько иной перечень основных типов отношений между классами: общее/частное («is-a») – роза есть цветок, целое/часть («part of») – лепесток есть часть розы и семантические, смысловые отношения, ассоциации (розы и свечи – и то, и другое можно использовать для украшения стола). Очевидно, что, несмотря на разницу классификаций, имеющаяся в них обеих семантическая (смысловая) связь хуже всего формализуется в виде однозначного дерева по сравнению со всеми остальными.

Анализ дерева категорий курса также показывает, что иерархия понятий не всегда задает последовательность их изучения. Например, в рассматриваемом разделе закон Ома для участка цепи принято изучать раньше, чем аналогичный закон для полной цепи, хотя с точки зрения соотношения понятий первый есть часть последнего. Если вспомнить, что в познании существует два противоположных пути – индукция и дедукция, то некоторая ограниченность иерархии современных онтологий (по крайней мере, применительно к учебному курсу) станет еще отчетливой.

Другая принципиальная трудность в создании учебной онтологии вытекает из «разнородности» любого реально изучаемого материала. В одном параграфе учебника могут рядом обсуждаться и определение физической величины, и некоторый закон, в который она входит, а также описываться применения этого закона в деятельности людей. В результате построить единую онтологию для всех перечисленных фрагментов оказывается не так просто.

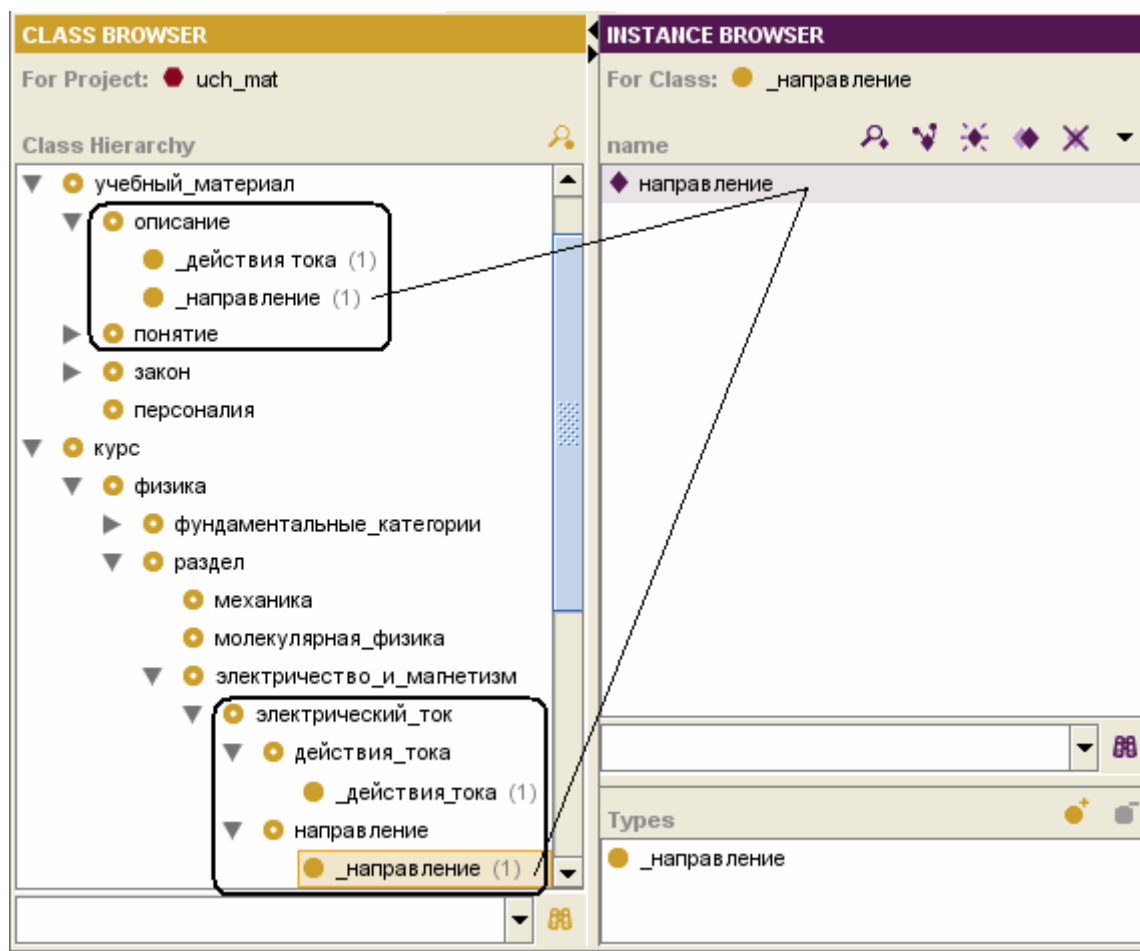


Рис. 1

Снова обратимся к конкретному примеру. На рис. 1 приведен небольшой фрагмент разрабатываемой онтологии темы «Электрический ток».

В верхней части онтологии мы видим класс *учебный_материал*, подклассы которого (*описание*, *понятие*, *закон* и т.п.) соответствуют элементам учебных текстов по физике, образующим любую конкретную тему. Каждый тип материала имеет определенные слоты, которые согласно общим принципам онтологий, наследуются. Например, характерным слотом для *понятия* служит *определение* (поскольку слоты отображаются только для выделенного класса, в правом окне на рис. 2 упомянутый слот не виден).

Обратимся теперь к классу *электрический_ток*, находящемуся в нижней части рисунка. По своей сути это *понятие*, так что с помощью множественного наследования ему надо бы добавить слоты, которые соответствуют данному виду учебного материала. Зато следующие два подкласса – *действия_тока* и *направление*, есть *описания*, следовательно, им слот *определение* из суперкласса совсем не нужен. В то же время, отказаться от его наследования подклассы не могут. Противоречие, таким образом, проявляется в том, что помимо средств расширения своей структуры, класс должен иметь также средства ее ограничения [16].

Возможное решение, не выходящее за рамки заложенной в систему Protégé модели знаний, состоит в том, чтобы добавлять необходимые слоты от подклассов *учебного материала* на самом последнем этапе, т.е. непосредственно перед созданием конкретного экземпляра класса. В результате, например, у класса *направление* приходится создавать дополнительный подкласс *_направление*, который не участвует в общей иерархии, и уже к нему применять принцип множественного наследования, добавляя

суперкласс *описание*. В итоге *направление* (своеобразный класс-«смеситель» для необходимых слотов) получает собственный характерный для данного материала набор слотов, который никак не повлияет на процесс наследования в связи с отсутствием классов-потомков. Зато именно от него будет порожден конкретный экземпляр учебного материала со всеми необходимыми свойствами.

Дополнительные классы, о которых говорилось выше, имеют на иерархической схеме рис. 1 отличное от остальных графическое обозначение – полностью закрашенный кружок. Так в Protégé обозначаются **конкретные** классы, от которых разрешено создавать экземпляры. Альтернативные классы, наследниками которых могут быть только классы, называются **абстрактными** и обозначаются на схеме кружком с незакрашенным центром. В свете принятой модели описанная разница в обозначениях понятна, поскольку экземпляры создаются исключительно от тех классов, в которых предварительно подготовлен необходимый суммарный набор слотов.

Подобный способ множественного наследования был заложен в язык Flavors (цитируется по [16]): небольшие классы, не предназначенные для порождения экземпляров, примешивались к другим классам, обеспечивая их более сложную структуру. Описанный прием называется **созданием примесей** (mixin). В нашей онтологии роль примеси выполняют подклассы, образованные от класса *учебный_материал*.

В рамках Protégé формально существует еще один способ получения отличающихся наборов слотов для различных типов учебного материала. Он состоит в том, чтобы наследовать абсолютно все слоты, а ненужные в каждой форме конкретного класса делать невидимыми. Подобное решение проблемы не только малопривлекательно с теоретической точки зрения, но и крайне трудоемко на практике.

Поскольку описанные трудности, связанные с необходимостью механизмов ограничения наследования, носят принципиальный характер, возможно, стоит подумать об усовершенствовании самих систем ведения онтологий. Можно предложить несколько направлений усовершенствования.

Прежде всего, с теоретической точки зрения весьма привлекательно выглядит предложение о введении роли (role) слотов, подобно тому, как это сделано для абстрактных и конкретных классов. Учитывая, что в Protégé слоты уже имеют два вида (template slot и own slot), легко формально обобщить модель наследования, введя третий тип слотов, назвав его, например, private slot. Как видно из нижеследующей таблицы, этот новый тип слота, в отличие от own slot, будет передаваться экземплярам класса, но не будет наследоваться в порождаемые подклассы.

тип слота	наследование классом	наследование экземпляром
<i>private?</i>	–	+
own	+	–
template	+	+

Возможно, такое обобщение будет противоречить модели знаний, принятой в Protégé. Как вариант можно предложить упомянутый выше прием «одноразового подмешивания» свойств другого класса без включения их в шаблон наследования.

Еще упомянем о распространенных в ООП **скрытых** (частных) свойствах, которые хотя и наследуются, но подклассами-потомками «не видны». Нечто похожее можно сделать и для слотов, но при большом количестве скрытых (ненужных) слотов, рабочие таблицы рискуют стать труднообозримыми. К тому же, как и в описанном выше случае «сокрытия» на форме ненужных слотов, постоянное изменение статуса слотов в каждом классе большой онтологии потребует слишком много манипуляций.

Наконец, можно попытаться реализовать вариантный шаблон слотов – аналог вариантной записи в Паскале, когда состав полей зависит от состояния некоторого узлового поля (в нашем случае – от типа учебного материала).

Организация связей материала

При работе с учебным материалом большое значение имеет возможность взглянуть на связанные с ним вопросы. Пакет Protégé представляет весьма простое и удобное средство для организации таких связей.

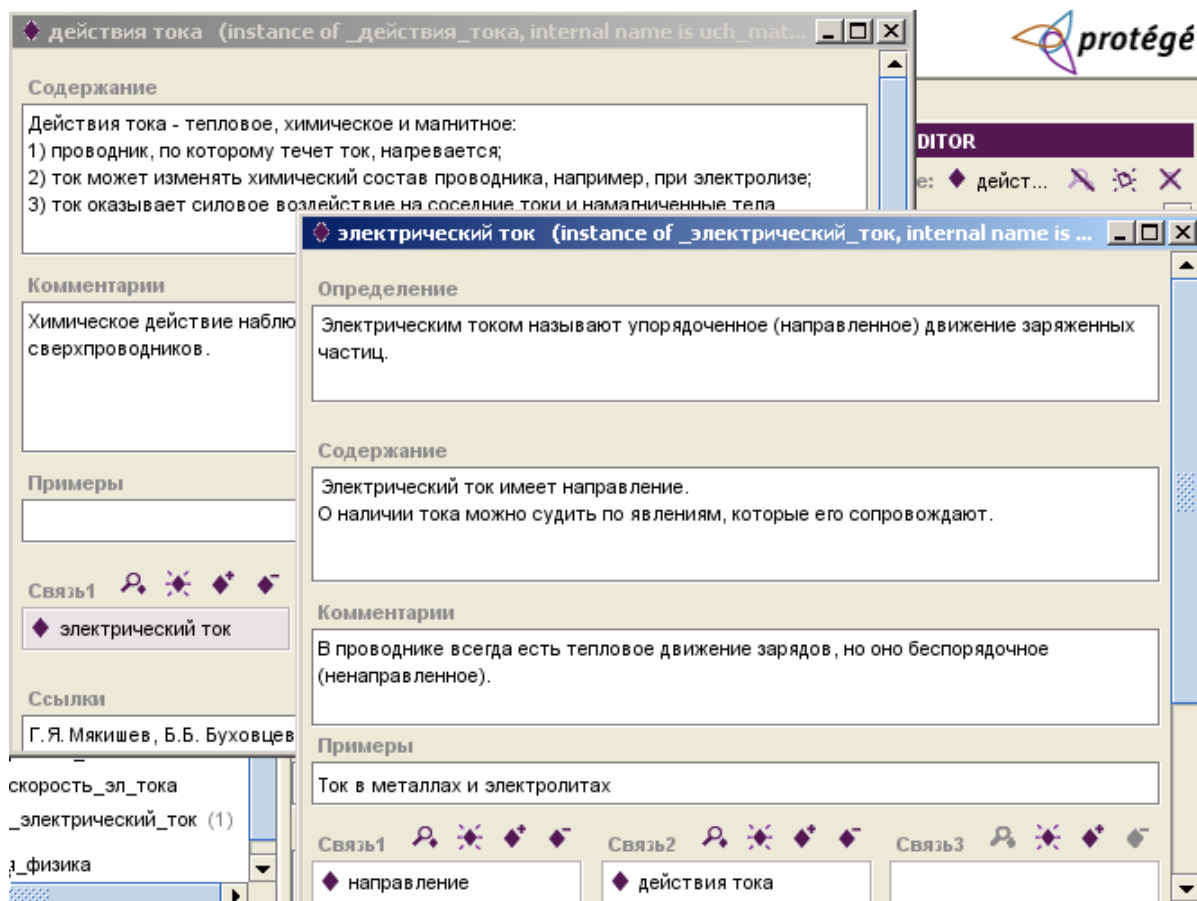


Рис. 2

Для решения рассматриваемой проблемы воспользуемся тем, что слот в качестве типа своего значения допускает использование экземпляров класса. В результате, помещенный на форму с материалом, такой слот становится ссылкой на другой материал: напомним, что в терминах онтологий конкретные фрагменты учебного материала являются экземплярами классов, так что речь идет как раз о связи между экземплярами разных классов. Имея на форме слоты указанного типа, мы можем просмотреть связанные с ними экземпляры классов при помощи двойного щелчка мыши. Полученная реализация по смыслу очень похожа на гиперссылку, которая, будучи выбранной, открывает свой ресурс в новом окне браузера.

На рис. 2 приведен пример формы, на которой отображается вопрос «Действия тока». В располагающемся сверху окне был открыт связанный с ним материал под названием «Электрический ток»; для его вызова использовался одноименный слот-ссылка под именем *связь1*.

Приведенные на рис. 2 формы дают хорошее представление о возможном пути решения поставленной нами задачи о расшифровке содержания вопросов учебной программы. С этих позиций особое значение имеет поле со ссылками на литературу по вызванному вопросу, располагающееся в нижней части формы.

Перспективы продолжения работы

В настоящее время заложена общая основа онтологии и выстроена иерархия классов для нескольких тем курса физики. Некоторый интерес с теоретической точки зрения может представлять продолжение работы по построению онтологий для нескольких разделов физики, поскольку при реализации связей между разделами могут возникнуть некоторые новые специфические проблемы.

В перспективе также планируется проделать подобную работу для нескольких учебных дисциплин с целью сравнения полученных онтологий. Есть предположения, что результаты позволят сделать какие-то объективные выводы о сложности материала отдельных курсов (и их разделов), а также видах организации их внутренней структуры.

Ссылки

- [1] A. Gómez-Pérez, M. Fernández-López, O. Corcho. *Ontological Engineering with Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer-Verlag, London, 2004.
- [2] A. Maedche. *Ontology Learning for Semantic Web*. Kluwer Academic Publishers, Boston, 2002.
- [3] J. Milam. *Ontologies in High Education*. In: *Knowledge Management and Higher Education: A Critical Analysis*. Ed. Metcalfe A.S., Idea Group Publishing, Hershey, PA, 2005, pp. 34-62.
- [4] N.F. Noy, D.L. McGuinness. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001. http://protege.stanford.edu/publications/ontology_development/ontology101.html
- [5] S. Sosnovsky, T. Gavrilova. *Development of Educational Ontology for C-programming*. In: *Information Theories & Applications*, 2005, 13 (4), pp.303-308.
- [6] K. Olsevicova. *Topic Maps e-Learning Portal Development*. In: *The Electronic Journal of e-Learning*, 2006, 4 (1), pp 59-66, available online at www.ejel.org
- [7] D. Dicheva, C. Dichev. *Authoring Educational Topic Maps: Can We Make It Easier?* In: *Proceedings of ICALT 2005*, pp. 216-219.
- [8] D. Dicheva, C. Dichev. *Confronting Some Ontology-building Problems in Educational Topic Map Authoring*. In: *Proceedings of the Workshop on Applications of Semantic Web in E-Learning (SW-EL@AH'06)*, in conjunction with AH 2006, pp. 55-64.
- [9] N.W. Brickhouse. *Teachers' beliefs about the nature of science and their relationship to classroom practice*. In: *Journal of Teacher Education*, 1990, 41 (3), pp. 53-62.
- [10] I.T. Koponen, T. Mäntylä., J. Lavonen. *Challenges of Web-based Education in Physic Teachers' Training*. In: *Proceedings of ICTE 2002*, pp. 291-295.
- [11] The Protégé Project. <http://protege.stanford.edu>
- [12] N.F. Noy, R.W. Ferguson, M.A. Musen. *The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility*. In: *Proceedings of EKAW 2000, Lecture Notes in Artificial Intelligence*, no. 1937, Springer-Verlag, Berlin, pp.17-32.
- [13] N.F. Noy, M. Sintek, S. Decker, M. Crubézy, R.W. Ferguson, M.A. Musen. *Creating Semantic Web Contents with Protégé-2000*. In: *IEEE Intelligent Systems*, 2001, 16 (2), pp. 60-71.

-
-
- [14] Г.Я. Мякишев, Б.Б. Буховцев, Н.Н. Сотский. Физика: Учебник для 10 класса общеобразовательных учреждений. Просвещение, Москва, 2002.
- [15] A. Papadopoulos. Meaningful Search: Why PET Scanners are not about Cats & Dogs. Convera, Carlsbad, CA, 2003.
- [16] Г. Буч. Объектно-ориентированный анализ и проектирование с примерами на C++. Бином, Москва; Невский диалект, Санкт-Петербург, 2000 (G. Booch. Object-Oriented Analysis and Design with Applications. Addison-Wesley Publishing Company, 1993).

Author's Information

Evgeny A. Eremin - Perm State Pedagogical University, 614990, Sibirskaya st., 24, Perm, Russia;
e-mail: eremin@pspu.ac.ru