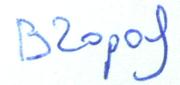


На правах рукописи



Горовой Владимир Андреевич

Методы оценки онтологий для построения порталов знаний

Специальность:

05.13.11 - Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание ученой
степени кандидата технических наук

Санкт-Петербург – 2011

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Санкт-Петербургский государственный политехнический университет».

Научный руководитель

доктор технических наук, профессор
Гаврилова Татьяна Альбертовна

Официальные оппоненты

доктор технических наук, профессор
Котенко Игорь Витальевич
доктор технических наук, профессор
Устинов Сергей Михайлович

Ведущая организация

Институт Системного Анализа РАН

Защита состоится «1» марта 2012 г. в 16 часов на заседании диссертационного совета Д 212.229.18 при федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Санкт-Петербургский государственный политехнический университет»: 195251, Санкт-Петербург, ул. Политехническая, д. 21, 9-й учебный корпус, ауд. 325.

С диссертацией можно ознакомиться в Фундаментальной библиотеке университета.

Автореферат разослан « » 2012 г.

Ученый секретарь
диссертационного совета



Васильев А.Е.

Общая характеристика работы

Актуальность. Одним из эффективных инструментов для решения задачи управления знаниями являются порталы знаний. Они применяются не только для качественного обучения в различных предметных областях, но и в повседневной профессиональной деятельности специалистов для поиска и анализа необходимой информации в огромном объеме накопленных данных.

При использовании портала как инструмента для передачи знаний его эффективность зависит от качества его структуры, т.к. пользователю важно понимание концептуальной модели предметной области: понятий, связей между ними, места каждого понятия в общей картине. При отсутствии должного внимания к структуре портал может со временем превратиться в хаос гипертекста, что существенно затруднит поиск и анализ информации. Несмотря на то, что качество структуры носит отчасти субъективный характер, достижения в области психологии и когнитивной эргономики (области междисциплинарных исследований, изучающей процессы восприятия и понимания интерфейсов, моделей, представлений с точки зрения эргономики) Д. Миллера, М. Вертгеймера, Б.М. Величковского позволяют его оценивать.

Использование онтологий (формальных спецификаций разделяемой концептуальной модели) для описания концептуальной модели предметной области, лежащей в основе портала знаний, позволяет повысить качество структуры портала, что влияет на его эффективность, выраженную в сокращении сроков обучения и сокращении времени поиска и анализа информации. Онтологии в настоящее время широко применяются для решения задач управления знаниями, а также в интеллектуальных системах обучения. Исследованию проблемы формирования онтологий посвящены множество работ известных российских и зарубежных ученых: Т. Грубера, Н. Гуарино, Р. Мизогучи, Д. Хендлера, А.Гангеми, Т.А. Гавриловой, В.Ш. Рубашкина, А.В. Смирнова, Г.С. Осипова.

Построение портала на основе онтологий требует разработки специальной методики, а для современных порталов, описывающих большое количество понятий и связей между ними, необходимы автоматизированные средства оценки качества портала. Этому вопросу уделяется внимание в работах Х.Яо, А.Гангеми, однако в них нет описания методики оценки онтологии с точки зрения качества структуры, а описанные метрики не носят всестороннего характера.

Немаловажным является и то, что порталы, построенные на основе онтологий, могут являться источником структурированных данных для междисциплинарных исследований и для предоставления машинно-читаемых данных различным приложениям. Таким образом, портал может стать частью движения Linked Open Data, объединяющего данные в открытых

форматах представления онтологий OWL и RDF, в которых организации выкладывают в открытый доступ накопленные знания.

Наконец, в портале знаний важно обеспечить не только возможность чтения и поиска информации, но возможность работы над созданием и изменением портала распределенным коллективом разработчиков с помощью wiki-технологии.

Все перечисленное выше делает актуальной тему диссертационной работы, посвященную методике оценки онтологий для построения порталов знаний и методике построения порталов знаний на основе онтологий, использующей wiki-технологии.

Цель и задачи диссертационной работы. Основной целью диссертационной работы является разработка методов оценки качества онтологий для построения порталов знаний.

Задачи, решаемые для достижения данной цели:

1. Разработать классификацию методов оценки онтологий;
2. Разработать методику оценки когнитивной эргономичности онтологий;
3. Предложить методику построения порталов знаний на базе онтологий;
4. Предложить методику интеграции wiki-технологии в порталы знаний на базе онтологий;
5. Разработать архитектуру и реализовать систему построения порталов знаний на основе разработанных моделей и методов.

Объект исследования. Модели и методы структурирования, представления знаний для задач построения порталов знаний.

Методы исследования. Для решения поставленных задач используются методологии проектирования информационных систем, баз знаний, теоретические основы математической статистики, теории графов, а также методы инженерии знаний, методологии онтологического инжиниринга.

Основные научные результаты, выносимые на защиту, и их научная новизна:

1. Впервые предложена классификация существующих методов оценки онтологий на основе следующих критериев: стадии применения, анализируемых объектов, целей и средств анализа. Предложенная классификация снижает трудоемкость принятия решения о выборе метода оценки исходя из специфики задач.

2. Разработана методика оценки когнитивной эргономичности онтологий, основанная на анализе графа онтологии, позволяющая, в частности, оценивать качество онтологий, предназначенных для построения порталов знаний.

Впервые выделены и описаны основные группы метрик, позволяющие частично автоматизировать работу эксперта по оценке онтологий: метрики циклов, метрики Ингве-

Миллера, метрики разнообразия количества связей и типов связей концептов, метрики глубины и ширины, метрики запутанности и ветвистости.

3. Предложена методика построения порталов знаний на базе онтологий, использующая привязку страниц портала к понятиям, описанным в онтологии.

В рамках этой методики разработаны механизмы навигации по portalу, а также не имеющая аналогов реализация визуального представления онтологии в виде гиперболического графа, реализованного на базе решения hypergraph (<http://hypergraph.sourceforge.net>), позволяющая снизить трудоемкость изучения структуры онтологии и навигации по portalу.

4. Разработана методика построения порталов знаний на базе онтологий, использующая wiki-технологии для совместной работы над информационным наполнением portalа.

В рамках этой методики обеспечивается процесс создания portalа знаний, структура которого описана онтологией предметной области, а совместная работа над содержимым ведется при помощи wiki-технологий.

5. Разработана архитектура программного комплекса, позволяющего автоматизировать процесс создания порталов знаний на основе онтологий на базе предложенных выше методов, реализовать которые невозможно при помощи существующих универсальных систем создания порталов.

На базе разработанной архитектуры созданы программные продукты ОНТОЛИНЖ-КАОН и ОНТОЛИНЖ-Wiki.

Практическая значимость работы и внедрение результатов. Основные практические результаты исследования:

1. Предложенные методики могут быть использованы как для автоматизации процесса создания новых порталов, так и для улучшения характеристик уже существующих порталов.
2. Программные комплексы ОНТОЛИНЖ-КАОН и ОНТОЛИНЖ-Wiki позволяют создавать порталы знаний на основе owl-онтологий.

Разработанная система построения порталов знаний на основе онтологий и методика оценки когнитивной эргономичности онтологий используется в portalе знаний для музея оптики при СПбГУ ИТМО. Методика оценки качества онтологий была применена для сравнения двух онтологий в области искусственного интеллекта с целью выбора одной из них в качестве основы учебного курса. Методика создания порталов знаний на основе owl-онтологий и инструментарий ОНТОЛИНЖ-Wiki были использованы для работы над курсовой работой в области операционных систем группой студентов.

Апробация работы. Отдельные элементы диссертации вошли в результаты 5 исследовательских проектов, 3 из которых поддержаны Российским фондом фундаментальных исследований:

1. Разработка моделей и методов ОНТОлогического ИНЖиниринга (ОНТОЛИНЖ), грант РФФИ N 04-01-00466 (2004-2006).
2. Визуальное проектирование корпоративных баз знаний на основе онтологий (ВИП), грант РФФИ №08-07-00062-а (2007-2009).
3. Разработка методологии формирования учебно-дидактических онтологий (МЕДОНТ), грант СПбГУ 16.1.56.2009.
4. Разработка моделей и методов ОНТОлогического ИНЖиниринга в задаче управления знаниями в ВУЗе, грант СПбГУ 16.1.179.2010.
5. Структурирование знаний и Контента МЕтодами группового дизайна онТологий (КОМЕТ), грант РФФИ №11-07-00140-а (2011-2013).

Публикации. Основные результаты диссертации опубликованы в 23 печатных работах (8 статей в журналах, в т.ч. 5 в журналах из перечня ВАК, 1 статья в сборнике статей, 14 тезисов в трудах конференций, в т.ч. 10 международных, 7 из указанных работ опубликованы на английском языке).

СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность проблемы, формулируется цель диссертационной работы, приводится краткое содержание работы, приводятся результаты, выносимые на защиту, отмечается их новизна и практическая значимость.

В первой главе произведен обзор литературы и анализ роли онтологий в управлении знаниями, а также исследование использования порталов как инструмента управления знаниями. Фрагмент интеллект-карты исследований в области онтологического инжиниринга, с обозначенными на ней знаками вопроса направлениями, основное внимание которым уделяется в диссертационной работе, представлен на рис. 1.

Порталы знаний являются одним из самых распространенных инструментов управления знаниями применяемых как в организациях, так и в образовании. Одной из основных тенденций развития порталов знаний является использование wiki-технологии, которая позволяет не только использовать страницы портала в режиме чтения, но и редактировать их, а также создавать новые и удалять существующие, потерявшие свою актуальность страницы. Таким образом, при помощи wiki-технологии порталы знаний могут играть ключевую роль в процессе экстернализации знаний.

Одной из основных проблем использования порталов знаний является плохая структурированность информации на портале, которая делает сложным процесс навигации и

поиска, а также процесс усвоения структуры предметной области. Использование онтологии в качестве основы портала при его создании может помочь в решении этой проблемы, т.к. в этом случае структура портала будет настолько же хороша, насколько и онтология, лежащая в его основе. В этом случае становится критичной оценка качества структуры онтологии. Помимо этого, использование онтологии в качестве основы портала позволяет связать данные портала с данными Linked Data. Это дает возможность использовать открывающиеся структурированные данные для вывода новых знаний и построения приложений Semantic Web, основанных на данных из разных источников, связанных между собой.

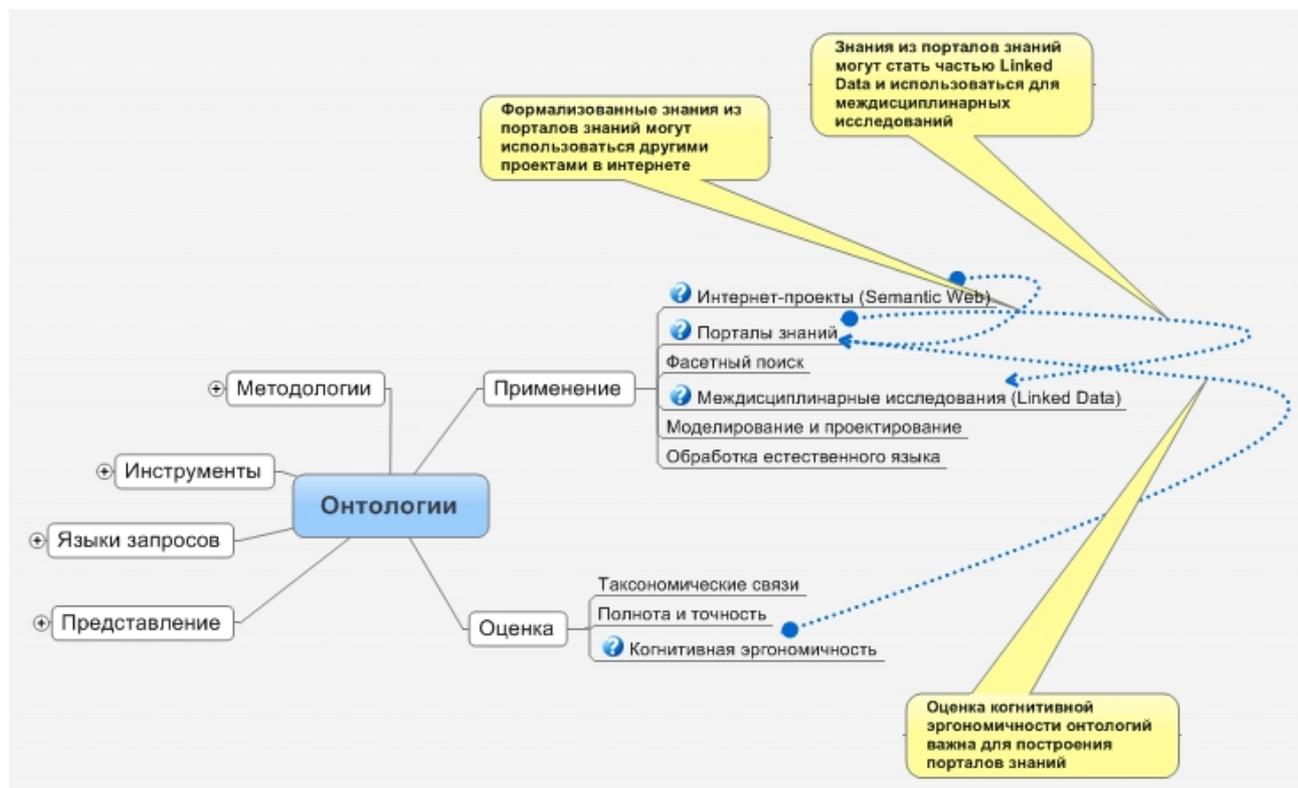


Рис. 1 Исследования в области онтологического инжиниринга

Несмотря на изложенные выше преимущества, существует немного инструментов, поддерживающих разработку порталов знаний на основе онтологий, работающих в стандартных браузерах. Существующие инструменты либо не поддерживают создание порталов на базе OWL-онтологий (KAON Portal, PORTO, OntoWeb), либо позволяют только генерировать web-представление онтологии для навигации (DAML Dumpont). OWL (Web Ontology Language) является рекомендованным стандартом для описания онтологий, в связи с чем критически важна его поддержка в инструментах разработки порталов знаний. Особого внимания заслуживает проект EduOntoWiki, ставший одним из первых инструментов, использовавших преимущества онтологий, как инструмента структурирования знаний и wiki, как эффективного средства для совместной работы. EduOntoWiki однако не лишен таких недостатков, как отсутствие визуального представления

графа онтологии, ограниченная функциональность собственного wiki-движка, отсутствие возможности интеграции в Linked Data.

Целью диссертации является решение представленных выше проблем путем разработки программного комплекса построения порталов знаний на основе онтологий, позволяющего визуализировать онтологии в формате OWL/RDF и создания методов оценки онтологий, положенных в основу порталов.

Во второй главе описана новая методика оценки когнитивной эргономичности онтологий на основе анализа графа. Под когнитивной эргономикой понимается область междисциплинарных исследований, изучающая процессы восприятия и понимания интерфейсов, моделей, представлений с точки зрения эргономики. Начало главы посвящено исследованию существующих методов оценки онтологий и разработке их классификации. Обобщенный взгляд на существующие подходы предлагает разработанная в рамках работы дескриптивная модель классификации методов оценки онтологий (см. рис. 2), позволяющая выбирать подходящие методы в информационных системах, использующих онтологии.

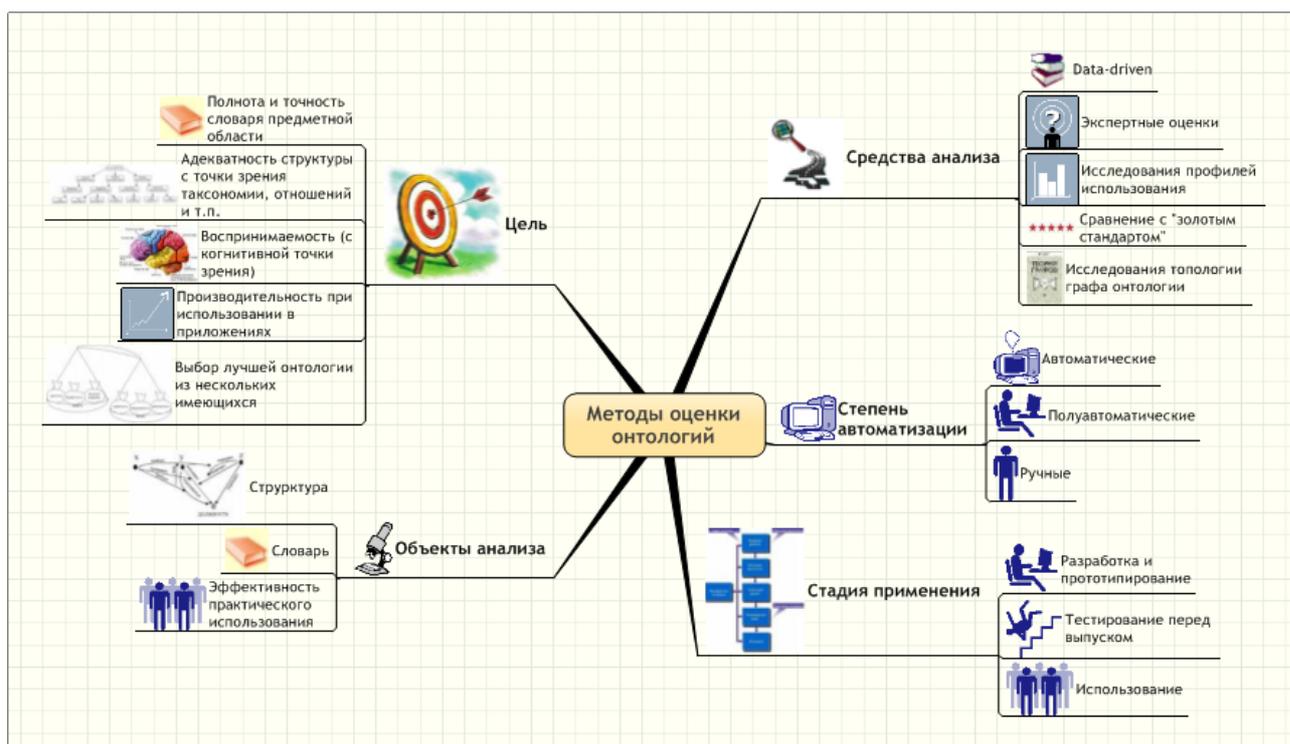


Рис. 2 Модель классификации методов оценки онтологий

С точки зрения данной на рис. 2 классификации предлагаемый в работе метод оценки можно описать следующим образом: целью является оценка воспринимаемости (с когнитивной точки зрения) или выбор лучшей онтологии из нескольких имеющихся, объектом анализа является структура онтологии, средством анализа служит топология графа онтологии, степень автоматизации метода полная, а также полуавтоматическая, т.к. окончательное решение может принимать эксперт на основе посчитанной модели.

Наиболее общей теорией для описания и измерения графических структур является теория графов. В ней определяются понятия, которые, безусловно, являются релевантными, для задачи оценки онтологий, но уровень абстракции теории делает ее непригодной для непосредственного применения. В связи с этим рядом исследователей разработаны подходы к оценке качества на основе топологии графа онтологии, учитывающие особенности этой задачи. Среди работ в этой области можно выделить труды Клещева А.С., Яо Х., Гангеми А. Несмотря на полезность метрик, введенных в этих работах, они покрывают лишь небольшую часть факторов, влияющих на восприятие и способность к запоминанию.

Основные принципы оценки визуальной воспринимаемости и понятности онтологий изложены в работах Гавриловой Т.А. В их основе лежит развитие взглядов Вертгеймера М. в области гештальт-психологии. Он рассматривал все задачи с точки зрения незавершенности или несовершенства структуры. Так основной принцип хорошего гештальта (хорошей формы) или закон прегнантности был сформулирован так: *«Организация любой структуры в природе или в сознании должна быть настолько хороша (регулярна, полна, сбалансирована или симметрична), насколько позволяют существующие условия»*. Среди других когнитивно-перцептивных принципов полезны следующие: закон близости, закон сходства, закон включения В. Келера, закон парсимонии.

Полученная в данной работе модель оценки качества онтологии основана на принципах, изложенных выше, и во многом является их формализацией. Целью применения модели является оценка сбалансированности и воспринимаемости онтологий пользователями.



Рис. 3 Фрагмент интеллект-карты метрик для оценки онтологий с точки зрения когнитивной эргономики. Полная версия приведена в диссертации.

Предлагаемые методы оценки базируются на нескольких группах метрик, на основе измерения которых, можно сделать выводы о качестве онтологий с точки зрения качества и прозрачности передачи формализованных знаний. На рис. 3 показан фрагмент итоговой интеллект-карты метрик, используемых для оценки онтологий с точки зрения когнитивной эргономики. С помощью префикса “*” отмечены метрики, впервые введенные в рамках диссертационного исследования. Префиксом “!” отмечены метрики, вошедшие в набора самых важных метрик на этапе Э8 оценки онтологий, описанном ниже.

Приведем формальное описание одной из групп метрик – метрик Ингве-Миллера. Примем следующие обозначения: g – граф, представляющий онтологию, концепты (классы и экземпляры) онтологии являются вершинами графа, отношения между концептами представлены в виде ребер графа; G – множество всех вершин g , E – множество всех ребер g .

Основная идея метрик Ингве-Миллера в том, что количество связей у одного концепта не должно превышать 7 ± 2 . Будем называть вершинами с нормальной степенью (степень вершины — количество дуг, для которых она является концевой) вершины, у которых суммарное число входящих и исходящих ребер не превышает 9.

- Отношение количества вершин с нормальной степенью по отношению ко всем вершинам. Чем ближе оно к 1, тем лучше онтология с точки зрения когнитивной эргономики.

$$m_{i1} = \frac{N_{v \in GD}}{n_G} \quad (i1)$$

Где n_G — количество вершин графа, $GD = \{v \in G | \deg(v) \leq 9\}$ — множество вершин с нормальной степенью, $N_{v \in GD}$ — количество вершин графа с нормальной степенью.

- Средняя степень вершины графа.

$$m_{i2} = \frac{\sum_{v \in G} \deg(v)}{n_G} = \frac{2 * n_E}{n_G} \quad (i2)$$

Где $\sum_{v \in G} \deg(v)$ — сумма всех степеней вершин графа, n_E — количество ребер.

- Медиана степени вершины графа.

$$m_{i3} = \widehat{\deg}(v) \quad (i3)$$

Где $\widehat{\deg}(v)$ — медиана степени вершины графа (значение степени вершины, которое делит ранжированную совокупность вершин на две равные части: 50 % «нижних» единиц ряда данных будут иметь значение степени вершины не больше, чем медиана, а «верхние» 50 % — значения степени вершины не меньше, чем медиана).

- 90%-line степени вершины графа.

$$m_{i4} = P_{90}(\deg(v)) \quad (i4)$$

Где $P_{90}(\deg(v))$ — 90-я процентиль степени вершины графа (возможное значение степени вершины, которое не превышают степени 90% вершин).

- Дисперсия степени вершины графа.

$$m_{i5} = \frac{\sum_{v \in G} (\deg(v) - \frac{\sum_{v \in G} \deg(v)}{n_G})^2}{n_G - 1} = \frac{\sum_{v \in G} (\deg(v) - \frac{2 * n_E}{n_G})^2}{n_G - 1} \quad (i5)$$

Часть других метрик, разработанных в ходе диссертационного исследования представлена в таблице 1. Подробные описания этих метрик, а также формальные описания всех остальных метрик представлены в диссертационной работе.

Таблица 1. Часть разработанных метрик

Метрика	Формула
f1 (количество вершин, у которых есть и листья и нелистовые вершины в качестве детей, по отношению ко всем количеству вершин у которых есть дети-листья)	$m_{f1} = \frac{N_{v \in S_{LEA \& SIB}}}{N_{v \in S_{LEA}}}$
dtp1 (количество вершин с разными типами исходящих связей по отношению ко всем вершинам графа)	$m_{dtp1} = \frac{N_{v \in VD}}{n_G}$
dtp2 (количество вершин с разными типами входящих связей по отношению ко всем вершинам графа)	$m_{dtp2} = \frac{N_{v \in V\bar{D}}}{n_G}$
d7 (дисперсия глубины)	$m_{d7} = \frac{\sum_j^P (N_{j \in P} - \frac{\sum_j^P N_{j \in P}}{n_{P \subseteq g}})^2}{n_{P \subseteq g} - 1}$
d8 (дисперсия глубины, деленная на среднюю глубину)	$m_{d8} = \frac{\frac{\sum_j^P (N_{j \in P} - \frac{\sum_j^P N_{j \in P}}{n_{P \subseteq g}})^2}{n_{P \subseteq g} - 1}}{\frac{\sum_j^P N_{j \in P}}{n_{P \subseteq g}}}$
f5 (среднее квадратичное отклонение детей-листьев у предпоследних вершин в графе)	$m_{f5} = \frac{\sum_{j \in SIB_{LEA}} (N_{j \in SIB}^{j \in LEA} - \frac{\sum_{j \in SIB_{LEA}} N_{j \in SIB}^{j \in LEA}}{n_{SIB_{LEA}}})^2}{n_{SIB_{LEA}} - 1}$
w5 (максимальное отношение ширины соседних уровней)	$m_{w5} = \frac{N_{l_i \in L}}{N_{l_{i-1} \in L}}$ $\forall k (\frac{N_{l_i \in L}}{N_{l_{i-1} \in L}} \geq \frac{N_{l_k \in L}}{N_{l_{k-1} \in L}})$
w8 (дисперсия отношения ширины соседних уровней)	$m_{w8} = \frac{\sum_{i=2}^{n_{L \subseteq g}} (\frac{N_{l_i \in L}}{N_{l_{i-1} \in L}} - \frac{1}{n_{L \subseteq g} - 1} \sum_{i=2}^{n_{L \subseteq g}} \frac{N_{l_i \in L}}{N_{l_{i-1} \in L}})^2}{n_{L \subseteq g} - 1}$

Основным преимуществом всех разработанных метрик, является возможность их автоматического вычисления на основе анализа графа онтологии. Для автоматизации процесса был разработан инструментарий COAT (Cognitive Ontology AssessmentT), реализованный на языке Java и поддерживающий онтологии в формате OWL.

Успех использования описанных метрик зависит, прежде всего, от эффективной и понятной методики, которая позволит экспертам не запутаться в большом количестве результатов измерений. Ниже описаны основные этапы оценки разработанной методики:

Э1. Этап анализа размера онтологии. Для того чтобы эксперт мог понять, насколько значение некоторых метрик является приемлемым, необходимо сначала получить сведения о размере онтологии. Для этого необходимо посчитать следующие метрики: количество вершин графа, максимальное расстояние от корневого узла, количество листьев дерева онтологии, количество вершин дерева онтологии, у которых есть листья в непосредственных потомках, количество дуг графа онтологии.

Э2. Этап анализа критических ошибок онтологии. Есть такие категории ошибок в онтологиях, при которых дальнейшую оценку проводить бессмысленно до их исправления. К такого рода ошибкам относится наличие циклов в онтологии (метрики $c1$ и $c2$), множественное наследование ($t2$ и другие метрики). Также отдельного рассмотрения заслуживают вершины, имеющие в качестве детей как листья, так и узлы дерева, не являющиеся листьями ($f1$ и другие метрики).

Э3. Этап анализа метрик Ингве-Миллера. Рассматриваются метрики, значения которых указывают на способность человека одновременно воспринимать узел онтологии вместе со всеми узлами, связанными с ним свойствами (метрики $i1, i2, i3, i4, i5$)

Э4. Этап анализа типов связей онтологии. Количество различных типов связей, присутствующих в онтологии, непосредственно влияет на ее восприятие. Негативно отразится на когнитивной эргономичности онтологии может большое количество концептов онтологий с разными типами исходящих связей (метрики $ntp1, ntp2$ и другие)

Э5. Этап анализа метрик глубины. Если различные ветви дерева онтологии имеют разную длину, то это может негативным образом отразиться на воспринимаемости онтологий. С помощью метрик глубины можно понять, насколько равномерно проработана онтология в разных областях ($d8$ и другие метрики).

Э6. Этап анализа метрик ширины. Метрики ширины позволяют проанализировать сбалансированность онтологии на основании изучения различий между разными уровнями иерархии ($w5$ и другие метрики).

Э7. Этап анализа метрик ветвистости. Особое внимание при анализе различных уровней иерархии в дереве онтологии следует уделять последнему, наиболее детальному уровню. На этом этапе анализа можно выделить те вершины графа, у которых он не достаточно проработан ($f5$ и другие метрики).

Э8. Этап принятия решения и выработки рекомендаций. В процессе анализа онтологии по изложенной методике на каждом этапе выявляются потенциальные узкие места в восприятии онтологии. Полученный список концептов и проблем можно эффективно использовать для последующей работы по улучшению качества онтологии. Анализируемые в описанной методике аспекты онтологии, влияющие на качество и скорость ее восприятия

пользователями настолько разнообразны, что выработать некоторый интегральный критерий качества представляется затруднительным. Тем не менее, выделим основные факторы, на основе которых можно будет в некоторых случаях уверенно говорить о том, какая из представленных на рассмотрение онтологий лучше или хуже других с точки зрения когнитивной эргономики: c1, c2, f1, i1, dtp1, dtp2, i5, d8, f5, w5. В тех случаях, когда у одной онтологии все 10 факторов имеют лучшие значения (i1, dtp1, dtp2 – большие; c1, c2, f1, i5, d8, f5, w5 – меньшие), чем у другой, мы можем утверждать, что она лучше с точки зрения когнитивной эргономики, если размер онтологий при этом сопоставим. Если же не все так однозначно, то эксперту необходимо учесть другие метрики и принимать решение исходя из того, какие метрики в данном случае имеют для него больший вес, а какие – меньший.

Предложенная модель также позволяет разработчикам онтологий понимать, улучшается ли их онтология с точки зрения когнитивной эргономики, сравнивая последующие итерации с результатами предыдущих или выбирать лучшую онтологию из набора онтологий, описывающих одну предметную область. Методика актуальна при большом объеме онтологий, т.к. позволяет автоматизировать расчет метрик, а также даже для небольших онтологий при отсутствии эксперта в области качества онтологий. Описанные результаты могут эффективно применяться в системах управления знаниями на предприятиях и в образовательном процессе, и помогать специалистам и преподавателям создавать более качественные онтологии.

В третьей главе предложена методика построения порталов знаний и wiki-порталов на базе онтологий и инструментарий ОНТОЛИНЖ-КАОН и ОНТОЛИНЖ-Wiki, поддерживающий описанную методику. В начале главы описывается технологический цикл разработки и использования онтологий.

Система ОНТОЛИНЖ реализует следующую методику использования онтологий для построения порталов знаний, состоящую из нескольких этапов. **На первом этапе** осуществляется создание экспертом или командой экспертов одной или нескольких онтологий при помощи одного из существующих инструментов для создания и редактирования онтологий. На этой стадии рекомендуется использовать такие хорошо зарекомендовавшие себя инструменты, как TopBraid Composer или Protégé, однако также могут применяться любые другие редакторы, позволяющие сохранять онтологии в общепризнанном формате OWL. В результате этого этапа разрабатываются одна или несколько онтологий.

На втором этапе происходит валидация и оценка построенных онтологий. Наиболее важным аспектом оценки онтологий, используемых для построения порталов знаний, является их оценка с точки зрения когнитивной эргономичности. На этой стадии

рекомендуется применять методику оценки, описанную во второй главе диссертационной работы, а также разработанную классификацию методов оценки для выбора других методов.

На **третьем этапе** осуществляется генерация портала знаний на основе полученных онтологий. Для каждого понятия онтологии, класса, экземпляра или свойства генерируется уникальная страница, к которой также можно привязать информационное содержимое. При этом изменения сделанные в онтологиях, автоматически отражаются на страницах портала. В системе ОНТОЛИНЖ-Wiki на этом этапе частью сгенерированной страницы является wiki-компонент, позволяющий добавлять и редактировать контент портала.

На **четвертом этапе** осуществляется наполнение документов портала содержимым аналогично wiki-статьям. При создании оптического портала наполнение осуществлялось экспертами предметной области, специалистами более узкого профиля, а также активными пользователями из числа студентов-дипломников.

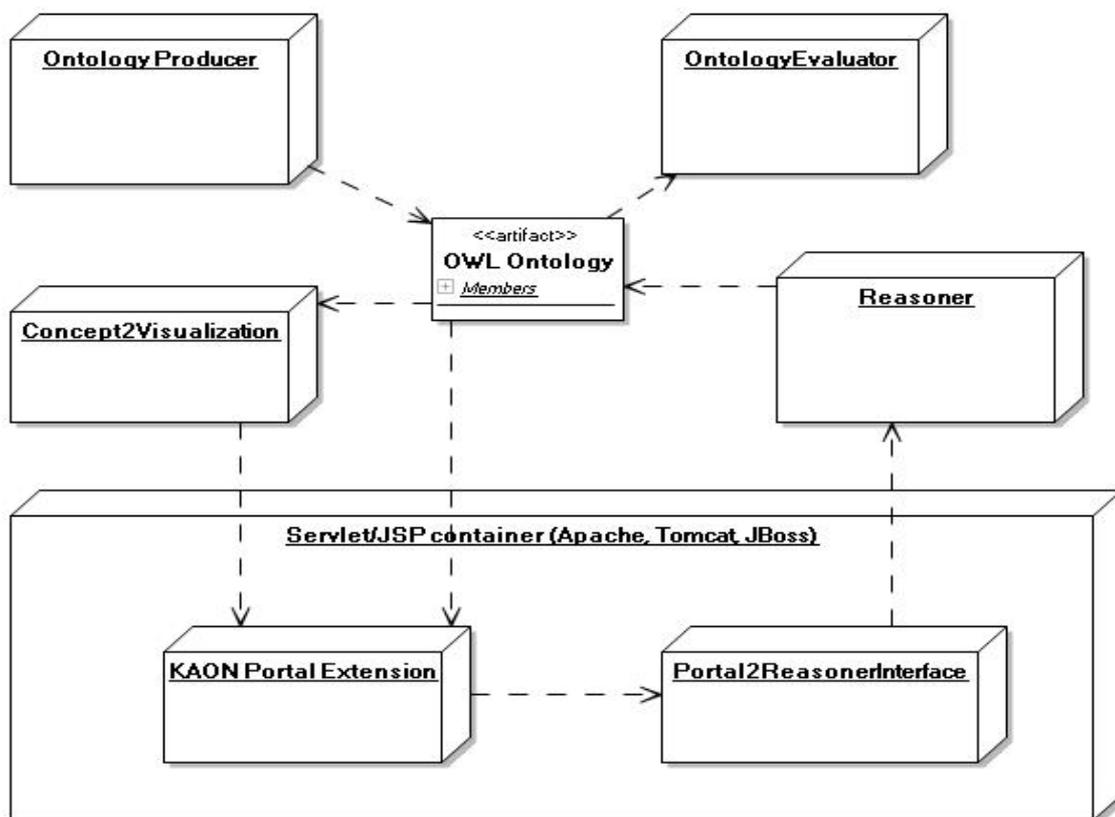


Рис. 4 Архитектура ОНТОЛИНЖ-КАОН

Заключительным этапом является использование портала. Например, портал может применяться для дистанционного обучения, для создания курсовых работ студентами, для передачи знаний по сложным предметным областям.

Архитектура ОНТОЛИНЖ-КАОН представлена на рис. 4. Основные компоненты системы ОНТОЛИНЖ-Wiki и потоки данных между ними и пользователями системы представлены на рис. 5.



Рис. 5 Основные компоненты ОНТОЛИНЖ-Wiki и потоки данных между ними и пользователями системы

В четвертой главе представлены примеры использования разработанной методики оценки качества онтологий. В начале главы описывается процесс создания онтологии в области оптики, разработанной для создания портала знаний на базе инструментария ОНТОЛИНЖ-Wiki для оптического музея при СПбГУ ИТМО. Затем проводится сравнительный анализ двух версий онтологии с помощью разработанной методики. Фрагмент второй версии онтологии оптики показаны на рис. 6.

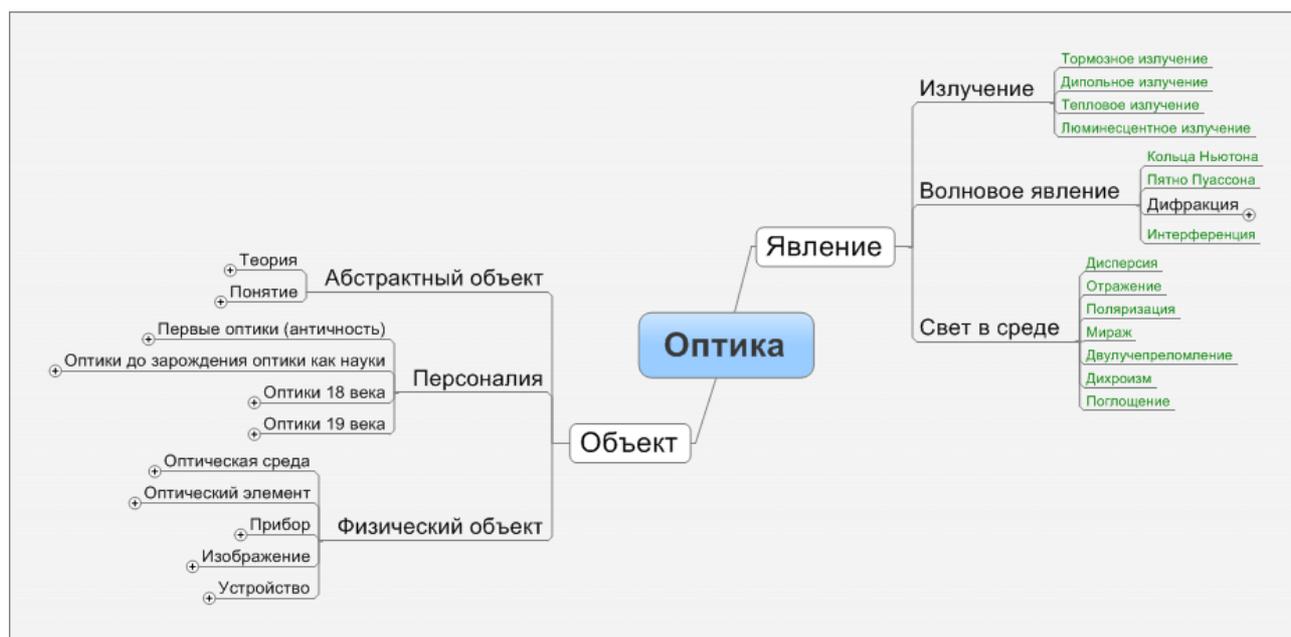


Рис. 6. Фрагмент второй версии онтологии оптики

В таблице 2 показаны значения метрик, применяемых на этапе Э8 разработанной методики оценки. Из нее, можно видеть, что вторая версия онтологии по всем параметрам стала лучше с точки зрения когнитивной эргономики.

Таблица 2. Значения основных метрик для двух версий онтологии оптики

Метрика	Первая версия	Вторая версия
c1 (количество различных циклов в графе)	0	0
c2 (количество вершин, имеющих несколько родителей, деленное на количество вершин в графе)	0	0

f1 (количество вершин, у которых есть и листья и нелистовые вершины в качестве детей, по отношению ко всем количеству вершин у которых есть листья среди детей)	0.29	0.27
i1 (отношение количества вершин с нормальной степенью по отношению ко всем вершинам)	0.94	1.0
i5 (дисперсия степени вершины графа)	3.84	0.60
d8 (дисперсия глубины, деленное на среднюю глубину)	0.391	0.268
f5 (дисперсия детей-листьев у предпоследних вершин в графе)	14.225	4.644
w5 (максимальное отношение ширины соседних уровней)	6.0	4.333

Вторая часть главы посвящена использованию методики для сравнительного анализа двух онтологий в области искусственного интеллекта: онтологии, построенной на базе широко известной классификации ACM и онтологии, построенной по учебнику С. Рассела и П. Норвига «Искусственный интеллект: современный подход». Данные примеры иллюстрируют процесс оценки онтологий и процесс создания онтологий для порталов знаний.

ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ

1. Дан анализ состояния использования онтологий в порталах знаний, основных тенденций развития Semantic Web и методов оценки качества онтологий, в результате которого идентифицированы проблемы существующих систем построения порталов знаний, использующих онтологии, и определены требования к желаемой автоматизированной системе построения порталов знаний на базе онтологий. Результаты анализа позволили сформулировать цель диссертационной работы — разработка методов оценки онтологий для порталов знаний.

2. *Предложена классификация существующих методов оценки онтологий* на основе целей, стадии применения, анализируемых объектов и средств анализа. Разработанная классификация снижает трудоемкость принятия решения о выборе метода оценки исходя из специфики задач, а также позволяет увидеть существующие пробелы в методиках оценки.

3. *Разработана методика оценки когнитивной эргономичности онтологии*, основанная на анализе графа онтологии и построенная на базе основных принципов когнитивной эргономичности, позволяющая, в частности, оценивать качество онтологий, предназначенных для построения порталов знаний.

Впервые выделены и описаны основные группы метрик, позволяющие частично автоматизировать работу эксперта по оценке онтологий с точки зрения когнитивной эргономики: метрики циклов, метрики Ингве-Миллера, метрики разнообразия количества связей и типов связей концептов, метрики глубины и ширины, метрики запутанности и

ветвистости.

Процесс применения методики описан на примере сравнительной оценки двух онтологий в области искусственного интеллекта и оценки онтологии портала оптического музея при СПбГУ ИТМО.

4. Предложена методика построения порталов знаний на базе онтологий, использующая привязку страниц портала к понятиям, описанным в онтологии. В рамках этой методики разработаны механизмы навигации по portalу, а также не имеющая аналогов реализация визуального представления онтологии в виде гиперболического графа, реализованного на базе решения hypergraph (<http://hypergraph.sourceforge.net>), позволяющая снизить трудоемкость изучения структуры онтологии и навигации по portalу.

5. Разработана методика построения порталов знаний на базе онтологий, использующая wiki-технологии для совместной работы над информационным наполнением портала. В рамках этой методики обеспечивается процесс создания портала знаний, структура которого описана онтологией предметной области, а совместная работы над содержимым ведется при помощи wiki-технологий.

6. Разработана архитектура программного комплекса, позволяющего автоматизировать процесс создания порталов знаний на основе онтологий на базе предложенных выше методов, реализовать которые невозможно при помощи существующих универсальных систем создания порталов. На базе разработанной архитектуры созданы программные продукты ОНТОЛИНЖ-КАОН и ОНТОЛИНЖ-Wiki.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Гаврилова Т. А., Кудрявцев Д. В., Горовой В. А. Модели и методы формирования онтологий // Научно-технические ведомости СПбГПУ, 2006. № 4. – С. 21-28.
2. Горовой В.А., Муромцев Д.И. Реализация технологии активного обучения на базе онтологического моделирования // Научно-технический вестник СПбГУ ИТМО, 2009. № 02(60). – С. 107-112.
3. Гаврилова Т.А., Горовой В.А., Болотникова Е.С. Оценка когнитивной эргономичности онтологии на основе анализа графа // Журнал РАН "Искусственный интеллект и принятие решений", 2009. № 03. – С. 33-41.
4. Муромцев Д.И., Горовой В.А., Злобин А.Н., Катков Ю.В., Починок И.Н. Архитектура системы управления знаниями на основе wiki-технологии и интегрированных онтологических моделей // Журнал известия ВУЗов. Приборостроение, 2011. №1. – С. 5-12.
5. Гаврилова Т.А., Горовой В.А., Болотникова Е.С. Об одном методе оценки онтологий // Журнал Известия РАН. Теория и системы управления, 2011. № 3. – С. 98-110.
6. Гаврилова Т.А., Гелеверя Т.Е., Горовой В.А. Онтологии как средство концептуализации web-порталов // Научно-теоретический журнал «Искусственный интеллект», 2002. № 3. – С. 80-86.
7. Гелеверя Т.Е., Горовой В.А., Горовая Д.О. Porto: Средство визуального проектирования web-порталов // Труды 5-й всероссийской объединенной конференции «Технологии информационного общества - Интернет и современное общество», 2002. – С. 82-84.
8. Gavrilova T., Gorovoy V. Ontological Engineering for Corporate Knowledge Portal Design // In "Processes and Foundations for Virtual Organizations", Eds. L. Camarinha Matos and H. Afsarmanesh, Kluwer Academic Publishers, 2003. – pp. 289-296.

9. Гаврилова Т.А., Горовой В.А. Управление знаниями как технология разработки интегрированных интеллектуальных систем // Труды седьмой научно-практической конференции «Реинжиниринг бизнес-процессов на основе современных информационных технологий. Системы управления знаниями». – Москва, 2004. – С. 166-170.
10. Гаврилова Т.А., Горовой В.А. Методологии онтологического инжиниринга // Материалы международной научно-технической конференции «Интеллектуальные и многопроцессорные системы», ИМС'2005. – Дивноморское, Россия, 26 сентября - 1 октября 2005. Т. 2. – С. 405-409.
11. Гаврилова Т.А., Горовой В.А. Архитектура программного комплекса ОНТОЛИНЖ-КАОН // Материалы 12-й международной конференции "Knowledge-Dialog-Solution" – KDS 2006, Варна, Болгария, 20-25 июня 2006. – С. 133-139.
12. Gorovoy V., Gavrilova T. Technology for ontological engineering lifecycle support // International Journal "Information Theories & Applications" Vol.14 / 2007. – pp. 19-25.
13. Горовой В.А. Модель классификации методов оценки онтологий // Материалы 2-й международной молодежной конференции «Искусственный интеллект: философия, методология, инновации». Санкт-Петербург, 15-17 ноября 2007. – С. 307-310.
14. Гаврилова Т.А., Горовой В.А., Злобин А.Н., Катков Ю.В., Кудрявцев Д.В., Малинин А.А., Муромцев Д.И. Интеграция wiki-технологии и онтологического моделирования в задаче управления знаниями предприятия // Труды 11-ой национальной конференции по искусственному интеллекту с международным участием – КИИ 2008, Дубна, Россия, 2008. Т. 3. – С. 360-368.
15. Gavrilova T., Gorovoy V., Petrashen E. Ontology-Based Conceptual Domain Modeling for Educational Portal // Ninth IEEE International Conference on Advanced Learning Technologies ICALT, IEEE Computer Society, Riga, 2009. – pp. 495-496.
16. Gavrilova T., Gorovoy V., Petrashen E. Learning Resources Organization using Ontological Framework // Advances in Web-Based Learning, Lectures Notes in Computer Science 5686, Proc. of the 8-th International conf. on Web-based Learning ICWL 2009, RWTH Aachen University, Springer, 2009. – pp. 158-162.
17. Gavrilova T., Gorovoy V., Petrashen E., Mouromtsev D. Categorizing Knowledge Patterns into Ontological Framework // 17-th Int. Conf. on Conceptual structures ICCS'09, Proc. of "Knowledge and Ontology ELSEWHERE" workshop, University Higher School of Economics, Moscow, 2009. – pp. 22-28.
18. Муромцев Д.И., Горовой В.А., Баландин Е.А., Катков Ю.В., Починок И.Н. Инструментарий онтологического моделирования для совместного выполнения учебно-исследовательских проектов // XVI Всероссийская научно-методическая конференция Телематика 2009. Санкт-Петербург, 2009. – С. 169-171.
19. Gavrilova T., Gorovoy V., Petrashen E. Merging WIKI and ontological approach to e-learning portal design // International book series "Information Science and Computing", Book 11 "Intelligent Engineering", Supplement to International Journal "Information Technologies and Knowledge" Vol. 3, ITNEA, Bulgaria, 2009. – pp. 131-135.
20. Гаврилова Т.А., Горовой В.А., Болотникова Е.С., Голенков В.В. Субъективные метрики оценки онтологий // Материалы Всероссийской конф. с межд. Участием «Знания-Онтологии-Теории» (ЗОНТ-2009), т.1, Институт математики СО РАН, Новосибирск, 2009. – С. 178-186.
21. Болотникова Е.С., Гаврилова Т.А., Горовой В.А. Когнитивная эргономичность онтологий в обучении: об одном методе оценки онтологий // Труды 12-ой национальной конференции по искусственному интеллекту с международным участием КИИ-2010. Т.2.Физматлит, 2010. – С. 281-289.
22. Mouromtsev D., Gorovoy V. Ontology-Based Approach to Build Intelligent E-Learning Systems // Proceedings of the 2010 international conference on e-learning, e-business, enterprise information systems, e-government (EEE Worldcomp'2010), Las Vegas, USA, July 12-15 2010. – pp. 102-107.
23. Горовой В.А., Ивашенко В.П. Использование данных семантического веба поисковыми системами // Труды всероссийской конференции «Управление знаниями и технологиями Semantic Web (KMSW-2010). СПб: СПбГУ ИТМО 2010. – С. 129-135.