Знания-Онтологии-Теории (ЗОНТ-09)

Опыт использования онтологий верхнего уровня при проектировании базы знаний музея оптических технологий

Баландин Е.А.[1], Катков Ю.В. [2], Муромцев Д.И.[3], Починок И.Н.[4]

[1]*Санкт-Петербургский Государственный Электротехнический Университет, ул. Профессора Попова, д. 5, г. Санкт-Петербург, 197376, Россия.*
[2]*Санкт-Петербургский Государственный Электротехнический Университет, ул. Профессора Попова, д. 5, г. Санкт-Петербург, 197376, Россия.*
[3]*Санкт-Петербургский Государственный Университет информационных технологий, механики и оптики, пр. Кронверкский, д.49, г. Санкт-Петербург, 197101, Россия.*
[4]*Санкт-Петербургский Государственный Электротехнический Университет, ул. Профессора Попова, д. 5, г. Санкт-Петербург, 197376, Россия.*

jecklandin@gmail.com, ganqturgon@gmail.com, dimour@mail.ru
ipochinok@gmail.com

Аннотация. *Описан пример применения онтологии верхнего уровня для проектирования базы знаний музея оптики. Проведено сравнение онтологий верхнего уровня, в результате которого в качестве базовой онтологии выбрана DOLCE. Выполнена адаптация DOLCE в контексте предметной области (оптика). Основываясь на существующих методах извлечения знаний и имеющихся инструментальных средствах, предложена методика извлечения знаний, носящая итеративный характер и предполагающая следующие шаги: выбор темы, конспектирование, формализация, рефакторинг и объединение с существующей онтологией. Предложена концепция "срезов знаний", представляющих собой множества связей онтологии и концептов, относящихся к этим связям. Разработана первая версия онтологии предметной области. Текущая версия (включает в себя понятия волновой оптики) запущена на сервере лаборатории интеллектуальных систем ИТМО (http://ailab.ifmo.ru:8081/optics/) и доступна в музее оптических технологий СПбГУ ИТМО.*

Ключевые слова: онтология верхнего уровня, инженерия знаний

1 Введение

В последнее время наблюдается тенденция к увеличению числа проектов, использующих онтологии в качестве основы для структуры информационных ресурсов, систем навигации по порталам, систем поиска. В частности использование онтологий находит широкое применение в системах eLearning [1]. Подобные системы и ресурсы требуют предварительно созданной вручную онтологии, используемой в качестве шаблона. Идея онтологического подхода, его философские и методологические аспекты детально рассмотрены в работе [2]. В данной статье описывается процесс создания и применение онтологии предметной области (domain ontology) на примере задачи создания учебного on-line портала в области оптики. Разработанный авторами прототип инструментария (рабочее название «Онтолинж Вики») использует технологии Wiki [3] (для обеспечения возможности быстрого совместного наполнения контента) и технологии визуального представления онтологий (для наилучшего структурирования и навигации).

На этапе планирования были выделены следующие стадии разработки:

1. Эксперты в предметной области создают онтологию совместно с инженерами по знаниям.
2. Начиная с первых версий, онтология становится структурной основой для портала, средством навигации и языком представления знаний, поэтому изначально был использован компонент для её визуализации в форме интерактивного гиперграфа.
3. С помощью созданного инструмента для каждого концепта онтологии автоматически создается wiki-страница, в которую эксперт вносит информацию об описываемом понятии.
4. Конечные пользователи получают возможность просматривать онтологию, представленную в виде графа, могут осуществлять навигацию и по концептам, и по wiki, а также просматривать содержимое wiki-страницы, соответствующей выбранному понятию.

Так как онтология напрямую влияет на будущую структуру контента, продставляется важным формулировка набора требований, которые позволят избежать возможных ошибок в процессе инженерии знаний и повысить актуальность создаваемой онтологии в данном контексте.

2 Требования к онтологии

Требования к онтологии можно условно разделить на содержательные и формальные. В данной работе набор требований неотделим от учебных задач, стоявших перед авторами. Первый набор требований касается собственно предметной области, а второй – методической стороны вопроса.

Содержательные требования:

- Онтология должна охватывать волновую, геометрическую и квантовую оптику на уровне учебного курса ВУЗа, содержать основные понятия и отношения между ними и предоставлять исторический (например, какие ученые исследовали и описывали те или иные явления), теоретический (что лежит в основе явления, какие свойства необходимы для возникновения явления) и практический (какие приборы и устройства основаны на явлении, где и как оно используется) "срезы" предметной области.
- Онтология должна разрабатываться с учетом её целевой аудитории, куда входят в первую очередь студенты младших курсов, школьники старших классов, а также посетители виртуальной экспозиции музея оптических технологий, и все интересующиеся историей и проблемами оптики. Такая целевая аудитория определяет два основных способа использования онтологии: во-первых, в качестве учебного пособия (информационная поддержка лекционных и практических занятий или самостоятельная работа студента или школьника), во-вторых, в качестве информационного материала к экспозиции оптического музея.
- Глубина проработки оптического контента должна отвечать основной задаче онтологии – формированию целостного взгляда на предметную область у целевой аудитории – посетителей музея оптики. Иными словами, она отражает когнитивный взгляд на оптику не с точки зрения, например, технологического процесса, а с точки зрения понимания студентами и школьниками фундаментальных основ оптике.

Формальные требования:

- Онтология должна быть представлена на каком-либо языке описания онтологий, поддержанном прикладными инструментальными средствами [4].
- Должны быть предусмотрены средства визуального представления онтологии.

В качестве отправной точки создания онтологии предметной области было решено использовать одну из существующих онтологий верхнего уровня.

3 Онтологии верхнего уровня

Использование онтологий верхнего уровня в качестве основы для онтологии предметной области позволяет избежать сложностей и противоречий при определении типа иерархии классов, а также определить базовый глоссарий (тезаурус), в терминах которого могут быть описаны понятия и объекты предметной области.

Существует несколько онтологий верхнего уровня, среди них Cyc (OpenCyc, ResearchCyc), BFO (Basic Formal Ontology), GFO (General Formal Ontology), SUMO (Suggested Upper Merged Ontology), DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), онтология Джона Сова (Sowa's ontology) [5].

Сравнительные характеристики онтологий верхнего уровня, подходящих для использования в качестве основы онтологии оптики, приведены в таблице ниже.

Таблица 1. Сравнение онтологий верхнего уровня.

	Количество классов	**Язык**	**Применения**	**Прочее**
BFO	36	OWL	в основном в медицине (ProteinOntology, Foundational Model of Anatomy (FMA), Gene Ontology (GO))	ориентирована на моделирование «физического мира» и создание онтологий в рамках научных исследований, не содержит свойств.
DOLCE	около 100 классов	KIF, OWL, логика первого порядка	в системах и сервисах на базе веб-технологий	предоставляет ясный когнитивный базис, нацеленный на описание лежащих в основе естественного языка и здравого смысла онтологических категорий [6]
GFO	79 классов	логика первого порядка, OWL	в основном в медицине (Gene Ontology, Celltype Ontology)	
SUMO	около 1000	OWL	в лингвистике, медицине, для представления информации я разных предметных областях	

Для создания онтологии оптики в качестве базиса была выбрана онтология верхнего уровня DOLCE (DOLCE+DnS Ultralite, версия 3.17). Основными причинами выбора DOLCE были, во-первых, её понятность, простота и небольшой размер (в отличие от SUMO, которая избыточна для этой задачи) и, во-вторых, набор свойств, хорошо подходящих для предметной области. Эта версия онтологии является сокращенной версией DOLCE Lite-Plus и отличается от неё упрощенной иерархией классов и, как следствие, более понятными и простыми именами классов и свойств и более простыми временными и пространственными отношениями.

Несмотря на сравнительно небольшое количество классов и свойств в этой версии онтологии, некоторые из них не актуальны для предметной области, поэтому, для избежания избыточности, перед добавлением оптического содержимого DOLCE была адаптирована: удалены не используемые классы и свойства и подкорректированы описания классов.

Например, были удалены все классы, связанные с представлением социальных институтов и их участников, такие как CollectiveAgent (подклассы Community, Group), SocialAgent (подклассы CollectiveAgent, Personification, SocialPerson), Organization; классы, позволяющие описывать рабочие процессы, например, Plan (подклассы Project, Workflow), PlanExecution (подкласс WorkflowExecution).

Иерархия классов DOLCE после редактирования представлена на рисунке 1.

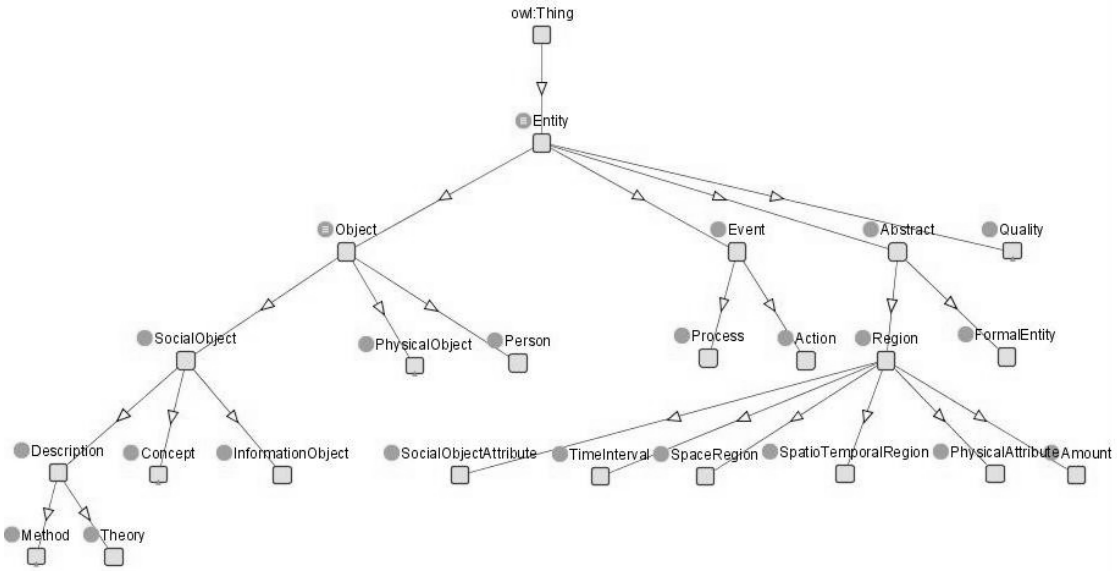

Рис.1. Адаптированная иерархия классов DOLCE (Protege).

После ознакомления с литературой по предметной области и небольшого общения с экспертом, были выделены основные понятий предметной области и распределены по классам верхнего уровня DOLCE (Рис. 2). Были также выделены и добавлены в DOLCE свойства, специфичные для предметной области.

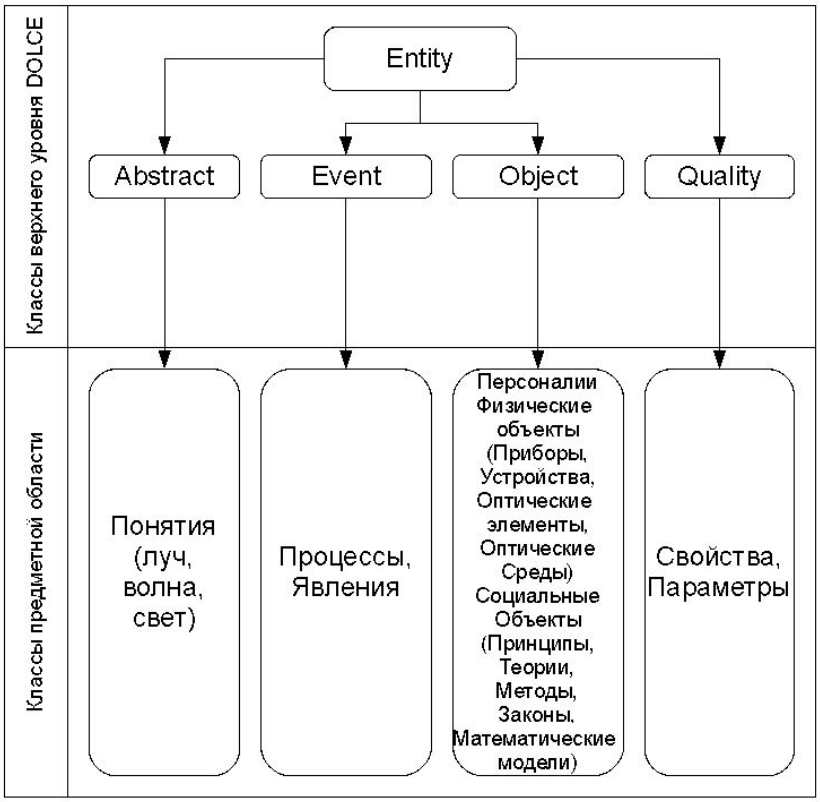

Рис.2. Распределение понятий предметной области по классам верхнего уровня DOLCE.

Выделенные понятия были категоризированны и распределены по уровням абстракции. По мере рассмотрения тем предметной области в процессе извлечения знаний онтология пополняется новыми понятиями.

4 Извлечение знаний

Как известно, извлечение знаний – это процесс, зависящий как от специфики предметной области, так и от области применения полученной формальной структуры. Решение базировать онтологию оптики на онтологии верхнего уровня было продиктовано, в первую очередь, степенью абстрактности многих понятий в предметной области. Это, а также и то, что онтология предназначалась в первую очередь для визуализации, накладывало определенные ограничения на применяемые при инженерии знаний методы.

4.1 Концепция "срезов знаний"

Для упрощения процесса составления онтологии и из соображений наглядности при составлении онтологии была сформулирована концепция так называемых *срезов знаний*. Срезы представляют собой множества связей (в терминологии OWL "объектных свойств", object properties) онтологии и концептов, относящихся к этим связям. Деление онтологической схемы на такие множества должно предоставить возможность пользователю рассматривать тот или иной вопрос с разных точек зрения, например, с точки зрения истории, практического применения, научного обоснования.

Для онтологии были выделены следующие смысловые срезы:

- исторический,
- практический(практическое применение),
- теоретический (например, понятия лежащие в основе какого-нибудь явления).

К *историческому срезу* отнесены такие свойства, как «объяснил», «провел», «открыл», «доказал», «ввел понятие». Этот срез позволяет взглянуть на некоторое явление в его историческом контексте – выяснить какие ученые его изучали, какие опыты они проводили.

К *практическому срезу* отнесены такие свойства, как «основано на», «использует». Этот срез позволяет оценить практическое применение понятий. Например, для явлений – просмотреть какие приборы, установки, практические технологии, на нём основаны.

К *теоретический срезу* отнесены такие свойства, как «есть условие», «является параметром», «является частью». Этот срез позволяет увидеть концепты, с помощью которых рассматриваемое понятие может быть объяснено. Например, в случае интерференции это такие понятия, как когерентность, длина волны.

Одно свойство может входить в несколько срезов. Например, свойство «основано на» входит в практический и в теоретический срез.

В языке OWL срезы реализуются иерархией объектных свойств.

4.2 Используемые модели

Так как составление онтологии представляет собой длительный процесс, а язык OWL не предназначен для интерпретации с естественного языка, существует необходимость в промежуточных моделях. Для выбора этих моделей был использован опыт составления продукционных баз знаний для экспертных систем, а также задействованы современные методики визуального представления знаний. В итоге были выбраны две промежуточные модели: mindmaps и conceptmaps.

Mindmaps

Mindmaps (интеллект-карты, и-карты) используются преимущественно на первом этапе инженерии знаний для быстрого и наглядного представления лекций эксперта. Карты имеют радиальную структуру и позволяют инженеру по знаниям с одной стороны, вести быстрый конспект, а с другой, следить за тем, чтобы эксперт не отвлекался от темы и согласовывать черновые записи. Приемы составления и-карт подробно описаны в [7]. Ниже приведен пример интеллект-карты. Такую

карту можно использовать для объяснения идеи срезов знаний, о которых говорилось ранее.

Рис.3. Пример интеллект-карты.

Conceptmaps

В отличие от интеллект карт, концепт-карты (к-карты) представляют больше возможностей для формализации. В концепт-картах наиболее значительные роли играют:

- расположение концептов и связей между ними,
- цветовое кодирование,
- эмфаза с использованием шрифтов.

Пример концепт карты, написанной при составлении онтологии оптики:

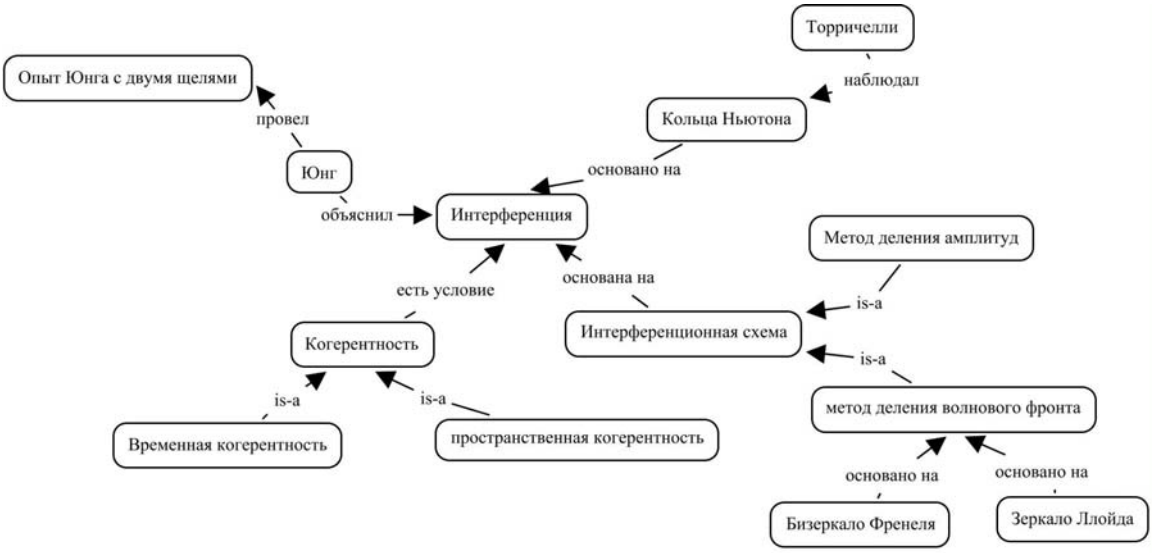

Рис.4. Пример концепт-карты.

Фактически наибольшая часть работы по инженерии знаний состояла в работе с к-картами. Методика составления к-карт описана в [8] и состоит из:

- определения контекста с помощью фокусирующих вопросов,
- выделения базовых понятий области,
- построения связей между концептами области,
- упорядочения графа.

5 Методика извлечения знаний

Составленная методика извлечения знаний носит итеративный характер и предполагает следующие шаги: выбор темы, конспектирование, формализация, рефакторинг и объединение с существующей онтологией. На схеме (Рис.5) показан процесс извлечения с получаемыми промежуточными артефактами.

На первом шаге выбирается центральный концепт, вокруг которого будет строиться дальнейшая работа. Далее, в ходе сеанса инженерии знаний формируется и-карта (mindmap), представляющая структурированные знания. После согласования и-карты с экспертом составляется выполняется формализация, результатом которой является к-карта (concept map), на которой разделены классовые и индивидные концепты, проведены связи между ними. На четвертом шаге выполняется оптимизация к-карт (заменяются синонимичные связи, удаляется избыточность) совместно с экспертом, делится на смысловые срезы, при необходимости происходит его пополнение и исправление – критический обзор. Наконец, результат записывается в документ формата OWL и отображается текущей версией инструмента Онтолинж Вики.

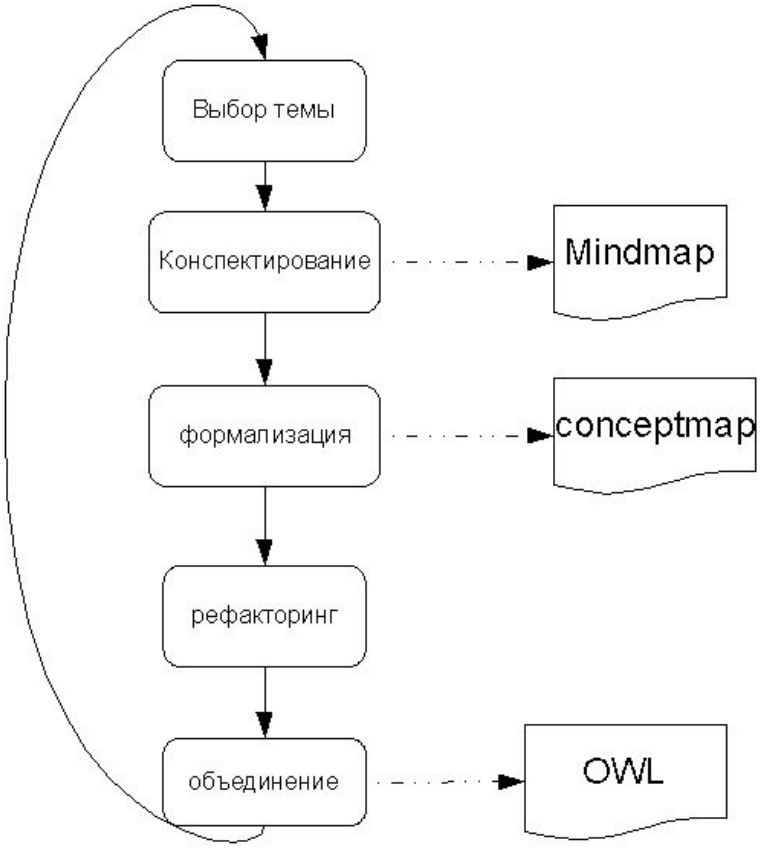

Рис.5. Процесс извлечения знаний.

Описанный процесс производится для наиболее важных концептов онтологии. В случае с оптическим музеем работа велась начиная с оптических явлений – дифракции, интерференции и т.д. Представленный метод извлечения знаний фактически относится к группе методов "интервью" [9] и может использоваться в любых задачах составления онтологий, где эксперт предметной области имеет дело с абстрактными сущностями. Однако в случае сложно формализуемых предметных областей лучше воспользоваться одним из методов опроса (структурированный или с наводящими вопросами).

6 Заключение

В настоящий момент разработана первая версия онтологии предметной области и ведется работа над ее развитием в частности, добавление понятий геометрической и квантовой оптики. Текущая версия (включает в себя понятия волновой оптики) запущена на сервере лаборатории интеллектуальных систем ИТМО (http://ailab.ifmo.ru:8081/optics/) и доступна в музее оптических

технологий СПбГУ ИТМО. Для внесенных в онтологию концептов созданы соответствующие wiki-страницы. Авторы предполагают продолжить развитие описанной методоки и прототипа инструментария до полноценного образовательного ресурса, включая перевод учебных материалов на иностранные языки. Важным достижением также является разработанная технологическая основа, фактически представляющая собой универсальную оболочку on-line баз знаний, которая без особых изменений может быть адаптирована для работы в других предметных областях.

Следует отметить, что уже первое знакомство педагогов школ и преподавателей вузов с разработанной онтологией и построенной на ее основе БЗ вызвало значительный интерес, так как предложенный подход обладает высокой наглядностью, когнитивной эргономичностью, значительной гибкостью программной реализации и масштабируемостью.

Литература

[1] Муромцев Д.И., Горовой В.А. Реализация технологии активного обучения на базе онтологического моделирования. СПб - «Научно-технический вестник СПбГУ ИТМО», 2009, №2(60)/2009. С. 107-114

[2] Гаврилова Т.А., Муромцев Д.И. Интеллектуальные технологии в менеджменте: инструменты и системы: Учебное пособие. - СПб.: "Высшая школа менеджмента", 2008, -488с.

[3] Stafford T., Webb M. What Is a Wiki (and How to Use One for Your Projects). http://www.oreillynet.com/pub/a/network/2006/07/07/what-is-a-wiki.html)

[4] Allemang D., Hedler J.A. Semantic web for the working ontologist modeling in RDF, RDFS and OWL. – Morgan Kaufmann, 2008,- 352c.

[5] Maskardi V., Corde V., Rosso P. A comparison of upper ontologies. //Technical report DISI-TR-06-21, available at www.disi.unige.it/person/MascardiV/Download/DISI-TR-06-21.pdf

[6] Masolo C., Borgo S., Gandemi A., Guarino N., Oltramari A., Schneider L. The WonderWeb Library of Foundational Ontologies. //Preliminary Report, available at http://www.loa-cnr.it/DOLCE.html.

[7] Buzan T. The Mind Map Handbook:The Ultimate Thinking Tool. - HarperCollins Publishers, 2005, -464c.

[8] Novak J. D., Cañas A.J. The Theory Underlying Concept Maps and How to Construct Them. //Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008", available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf.

[9] Осуга С., Саэки Ю., Судзуки Х. Приобретение знаний: Пер. с японского. - М.: "Мир", 1990,-304