

## «Онтологии и тезаурусы»

Авторы: Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д.

### Краткая аннотация

Предлагаемый курс направлен на формирование базовых знаний об онтологиях и тезаурусах и практических навыков по проектированию и применению онтологий при разработке компонентов интеллектуального программного обеспечения. Курс знакомит студентов с основными понятиями области представления знаний, примерами лучшего опыта в разработке систем основанных на онтологиях и тезаурусах, описывает задачи, решаемые с их помощью, а также набор инструментальных средств проектирования и представления онтологий и информационно-поисковых тезаурусов.

Центральное место в курсе отводится тематике информационного поиска. Подробно рассматриваются как традиционные методы поиска: по ключевым словам, основанные на техниках двоичного поиска, ранжированного поиска и вероятностных моделях, так и подходы к улучшению качества поиска с помощью тезаурусов и онтологий. В части лекций, посвященных проектированию онтологий, наравне с описанием классических и современных методологий внимание уделяется разбору примеров реальных проектов.

Данный курс разработан в Российском научно-образовательном центре по лингвистике им. И.А.Бодуэна де Куртенэ в рамках программы создания серии инновационных курсов. Разные его фрагменты апробировались в курсах, читавшихся в Казанском государственном университете:

1. В.Д.Соловьев на ф-те ВМК осенью 2005 г. "Обработка естественного языка on-line"
2. В.Д.Соловьев на ф-те ВМК осенью 2004 г. "Информационный поиск, категоризация текстов, автоматическое резюмирование"

и в Московском государственном университете:

1. Н.В.Лукашевич на филологическом факультете весной 2003 г. "Автоматическая обработка больших текстовых коллекций"
2. Н.В.Лукашевич на филологическом факультете осенью 2003 г. "Лингвистические онтологии для автоматической обработки текстов"
3. Б.В.Добров на факультете ВМиК осенью 2004 г. "Интеллектуальные информационные технологии (знания и машинное обучение в информационном поиске)"
4. Н.В.Лукашевич на факультете ВМиК весной 2003 г. "Онтологии и автоматическая обработка текстов".

Данный курс также основан на материале лекций, прочитанных Н.В.Лукашевич на Казанских школах по компьютерной лингвистике в 2001-2004 гг.

Предлагаемый курс стал победителем открытого конкурса учебных курсов в области разработки программного обеспечения, организованного компанией Microsoft в 2006 году.

Авторы надеются, что курс способствует привлечению талантливой молодежи в область информационного поиска и появлению специализированных средств и библиотек для применения онтологий в этой сфере.

Базовые требования к слушателям курса ограничиваются знаниями по математике и компьютерным наукам в объеме программы начальных курсов университетов.

## Программа курса

1. *Определение понятий онтология, концепт, отношение, аксиома. (2 часа)*

1.1. Определение понятий: онтология, концепт, отношение, аксиомы.

2. *Типы онтологий: верхнего уровня, предметных областей, прикладные онтологии. Лексические онтологии. Примеры онтологий (онтология вин и пищи). (4 часа)*

2.1. Типы онтологий: верхнего уровня, предметных областей, прикладных онтологий. Лексические онтологии.

2.2. Примеры онтологий (онтология вин и пищи)

3. *Назначение онтологий. Задачи, решаемые с помощью онтологий и тезаурусов (информационный поиск, интеграция гетерогенных источников данных, SemanticWeb). (4 часа)*

3.1. Назначение онтологий. Информационный поиск.

3.2. Назначение онтологий. Интеграция разнородных источников данных. SemanticWeb.

4. *Онтологии верхнего уровня: назначение, отличительные черты, решаемые задачи (примеры проектов - CYC, SUMO, Sowa's ontology). (4 часа)*

4.1. Онтологии SUMO и Sowa's ontology

4.2. Онтология CYC

5. *Онтологии предметных областей и прикладные онтологии: назначение, отличительные черты, решаемые задачи (примеры проектов). (4 часа)*

5.1. Онтологии предметных областей и прикладные онтологии. Онтология в области документации в сфере культурного наследия: CIDOC CRM

5.2. Онтологии товаров и услуг

6. *Языки описания онтологий. Основные синтаксические структуры: классы, отношения, аксиомы. Примеры: RDF, OWL. (4 часа)*

6.1. Архитектура метаданных WWW. Язык RDF.

6.2. Языки представления онтологий: RDFS, OWL. Язык запросов SPARQL.

7. *Инструментальные средства проектирования онтологий. Protege. (2 часа)*

7.1. Редакторы онтологий.

8. *WordNet. Описание. Усовершенствования. Приложения WordNet. RusNet. (6 часов)*

8.1. WordNet. Описание ресурса

8.2. WordNet. Приложения

8.3. WordNet. Проблемы

9. *Тезаурусы. Основные принципы разработки, создания и использования традиционных информационно-поисковых тезаурусов. Примеры тезаурусов. (2 часа)*

9.1. Тезаурусы. Основные принципы разработки, создания и использования традиционных информационно-поисковых тезаурусов. Примеры тезаурусов.

*10. Информационно-поисковые тезаурусы в условиях сверхбольших электронных коллекций и автоматической обработки текстов. Тезаурус для автоматического концептуального индексирования как особый вид тезауруса. (4 часа)*

10.1. Информационно-поисковые тезаурусы в условиях сверхбольших электронных коллекций и автоматической обработки текстов. Тезаурус для автоматического концептуального индексирования как особый вид тезауруса

10.2. Технология автоматической рубрикации текстов с использованием тезауруса для автоматического концептуального индексирования

Рекомендуемая литература:

1. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. - С-Пб.: Питер, 2000. - 384 с.
2. Овдей О.М., Проскудина Г.Ю. Обзор инструментов инженерии онтологий // Электронные библиотеки – Москва: Институт развития информационного общества, т.7 вып.4, 2004. – Электронный журнал, посвященный созданию и использованию электронных библиотек. - (Рус.). - URL: <http://www.elbib/>.
3. Разработка онтологий 101: руководство по созданию Вашей первой онтологии. - [ifets.ieee.org/russian/depository/ontology101\\_rus.doc](http://ifets.ieee.org/russian/depository/ontology101_rus.doc)
4. <http://www.w3c.org/TR/2004/REC-owl-guide-20040210/> (в Интернете доступно русскоязычное описание языка OWL)
5. George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller Introduction to WordNet: An On-line Lexical Database.
6. George A. Miller Nouns in WordNet: A Lexical Inheritance System.
7. Lars Marius Garshol Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all. (<http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>)
8. Topic Maps – A Standard For Information Organisation (<http://www.techquila.com/topicmaps.html>)
9. Berners-Lee, T., Hendler J., & Lassila, O. (2001). The semantic web. Scientific American, 284(5), 34-43.
10. N. Fridman Noy and C.D. Hafner: The State of the Art in Ontology Design, AI Magazine, 18(3):53---74, 1997. (<http://www.aaai.org/Library/Magazine/Vol18/18-03/Papers/AIMag18-03-005.pdf>)
11. Uschold M., Gruninger M. Ontologies: Principles, Methods and Applications. In *Knowledge Engineering Review* 11(2), 1996, pp. 93–155.
12. Gruber T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *KSL-93-04, Knowledge Systems Laboratory, Stanford University, 1993.*
13. Miller, George A., Christiane Fellbaum, Judy Kegl and Katherine J. Miller. WordNet: an electronic lexical reference system based on theories of lexical memory. In: *Revue quebecoise de linguistique* 17 (2), 1988, pp. 181 - 213.

Полезные ссылки:

<http://www.dialog-21.ru/>

<http://swoogle.umbc.edu/>

<http://www.w3.org/2001/sw/WebOnt/>.

<http://wonderweb.semanticweb.org/>.

<http://www.w3.org/2001/sw/BestPractices/>.

<http://www.xml.com/pub/a/2002/11/06/ontologies.html>.