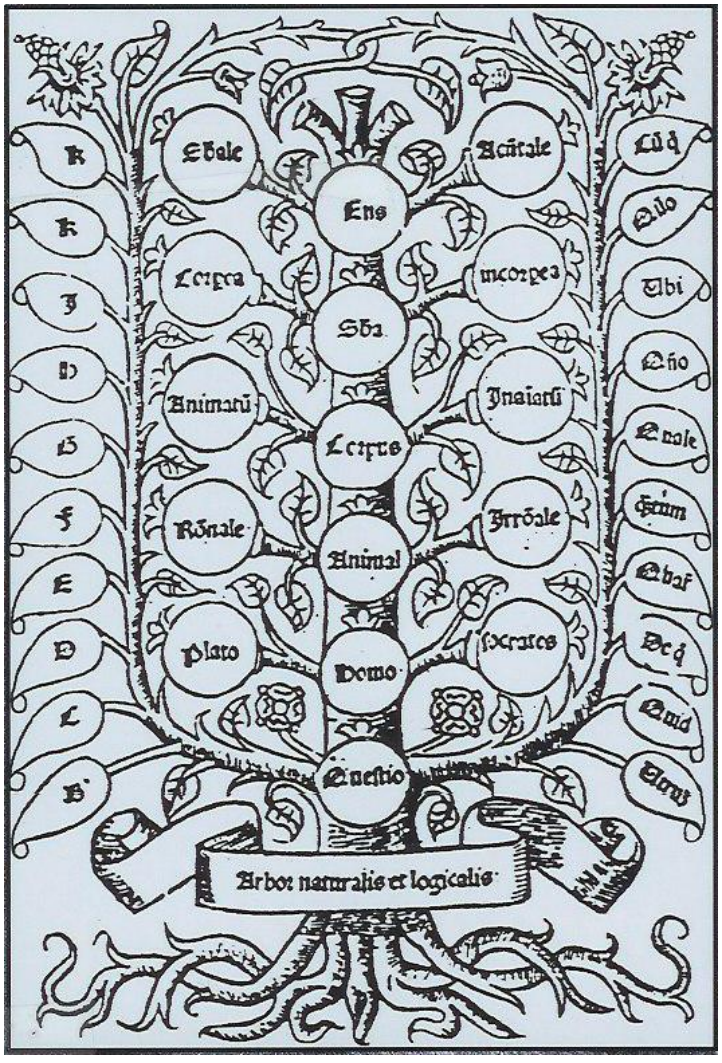
# INFSCI 2180
# Knowledge Representation and the Semantic Web

## Syllabus




**Marek J. Druzdzel**
**School of Information Sciences**
**University of Pittsburgh**
**Spring 2012 (12–2)**

## OVERVIEW:

*(knowledge representation as a discipline) "has not yet changed the world ... ... to realize its full potential it must be linked into a single global system."*
— Berners-Lee, T., Hendler, J. & Lassila, O., 2001*, The semantic web, Scientific American,* May, pp.35-43.

*"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation."* — Tim Berners-Lee

This course focuses on a fundamental problem faced by those who process information: Capturing and representation of knowledge in such a way that it is amenable to being processed by computer programs. While attempts to achieve this objective have been made in several disciplines, knowledge representation has been addressed directly in computer and information sciences by the discipline known as Artificial Intelligence. Researchers in artificial intelligence have developed a collection of approaches to representing and processing knowledge. Knowledge representation has achieved much interest in the framework of the effort to create the semantic net, i.e., a group of methods and technologies to allow machines to understand the meaning of information on the World Wide Web so that they can process this information directly or indirectly.

This course provides an introductory treatment of the problem of knowledge representation. We will review the following approaches:

- propositional and first-order logic
- ontologies
- semantic networks
- production rule systems
- Bayesian networks

The intended participants of the course are students who want to learn more about knowledge representation in the context of the efforts to create the semantic web. While you will not become proficient in any of these methods (an entire courses can be devoted to each of these), this course will give you a general overview and will lay foundations for your future studies.

As you might have already experienced by now, being an engineer or a scientist requires intelligence, independent, creative thinking, and most of all commitment to hard working. This course reinforces this. The material is not really difficult, but you will have to invest quality time in order to master it. The workload in this class will be moderately heavy, but I believe that you will find it interesting and important. I require your commitment, doing the readings, coming to classes, and being their active participant. In return, I promise that you will have fun and you will learn useful skills.

## YOUR RESOURCES:

**The course:**

Name : INFSCI 2180
Knowledge Representation and Semantic Web
CRN : 21036
Credits : 3.0
WWW : http://www.pitt.edu/~druzdzel/infsci2180.html

**The instructor:**

Marek J. Druzdzel
associate professor, School of Information Sciences and Intelligent Systems Program, University of Pittsburgh

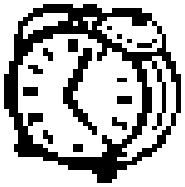
Office : 2B10 IS Building (Decision Systems Laboratory)
Email : *marek@sis.pitt.edu*
Phone : (412) 624-9432 (office, voice mail)
FAX : (412) 624-2788
WWW : http://www.pitt.edu/~druzdzel

**Teaching Assistant:**

Monir Sharker
doctoral student, School of Information Sciences

Office : 407 IS Building (Geoinformatics Laboratory)
Phone : 412-624-8858
Email : *mhs37@pitt.edu*

**Meeting times and locations:**



Classes (406 IS Building):
Tuesdays, 3pm-5:50pm (break 4:20pm-4:35pm)
Marek's office hours (2B13 SIS Building):
Mondays, 5:00-6:00pm or by appointment
Monir's office hours (407 SIS Building):
Tuesdays, 11:00-2:00pm or by appointment

**Your colleagues:**

Name: _____
Phone: _____
Email: _____
Name: _____
Phone: _____
Email: _____
Name: _____
Phone: _____
Email: _____

# THE COURSE:

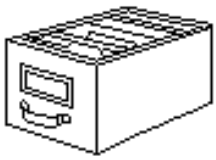
**Objectives:**

The primary objective of this course is to make you acquainted with a set of tools for representing knowledge in information systems and how these relate to the recent efforts of building the semantic web. I expect that you will learn:

- How to relate data to information, knowledge, and possibly wisdom ☺
- How to represent knowledge using such tools as propositional and first order logic, semantic networks, production rules, and Bayesian networks
- How to understand and possibly employ knowledge representation tools in the framework of the semantic web

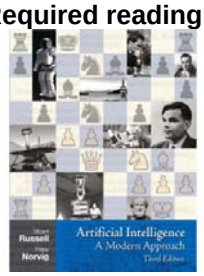
Finally, being successful in the course should contribute to the development of your academic self-esteem.

**Prerequisites:**



All students in the course are expected to have completed elementary programming courses and know elementary logic. I believe that it will be helpful for you to have completed at least one course in cognitive psychology. You will need elements of probability theory, although these will be quite basic and we will review them as needed. You will also be expected to use electronic mail on a regular basis. The most important prerequisite of all, however, is your interest in the course, motivation and commitment to learning.

**Required reading:**

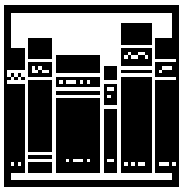



The principal textbook for the course is:

Stuart Russell & Peter Norvig *"Artificial Intelligence: A Modern Approach."* Third Edition, Prentice Hall, 2009, ISBN 0-136-04259-7 (current price between $90.70 and $116.75),

which is arguably the most comprehensive and modern introduction to Artificial Intelligence, and quite certainly the most popular one. It is available in the bookstore. I believe that you will enjoy reading it. Because of the limited focus of this course within the area of artificial intelligence, we will be using only selected chapters and sections from this textbook. The textbook will be supplemented with other readings (listed in this syllabus, although this list may be extended as needed), with one copy of each on reserve in the SIS library (3rd floor SIS Building).

**Supplementary readings:**

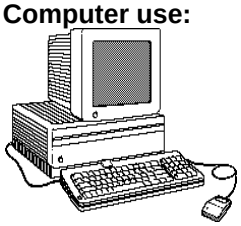


There are many good books and articles on knowledge representation in the context of artificial intelligence and there are also several books on semantic web in case you want to read more. I recommend an introductory book on the topic of the Semantic Web, available electronically for University of Pittsburgh affiliates:

Grigoris Antoniou, Frank van Harmelen. "A Semantic Web Primer", 2nd Edition. The MIT Press, 2008. ISBN 0262012421.

## WORK REQUIREMENTS:

**Computer use:**

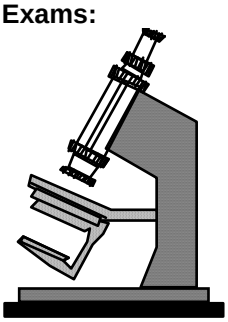


You will be expected to use a text processor or a typesetting program to write your documents and use software for some of the class assignments. Most of our communication will be electronic and you will be expected to use electronic mail on a daily basis.

**Assignments:**

There will be five homework assignments that will help you to practice the material covered in class and will help me to identify those parts of the material that you have difficulties with. Assignments will usually be done in groups of two or three students, formed during the class meetings. The assignments have to be turned in on time, and all members of the group are responsible for meeting the deadline.

**Exams:**

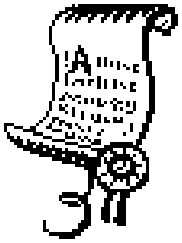


There will be one midterm exam and one comprehensive final exam, both open book.

**Time load:**

To help you with planning your semester, I would like to give you an idea of the minimal workload in this course. Expect to spend about six hours quality time outside of class for every class meeting. I estimate that you will need about three hours to do the readings and three hours (on the average) to do the assignments. If you keep up with readings and do the assignments well, you should not need much extra time to prepare for the exams. The actual load will vary, of course, depending on your background and preparation. If, for example, your knowledge of elementary college mathematics is rusty, you may need more time.

**Grading:**



Your final grade for the course will be determined as follows:

| | | |
|---|---|---|
| Assignments | : | 40% |
| Midterm exam | : | 30% |
| Final exam | : | 30% |

On the top of this all, you can obtain up to 10% of the total score for in-class participation.

# COURSE SCHEDULE

## PART I: INTRODUCTION

*In addition to organizing ourselves, the first class will be an introduction to the problem of knowledge representation, as posed and developed in Artificial Intelligence and applied to the idea of World Wide Web 3, a.k.a. the Semantic Web.*

| | |
|---|---|
| January 10 | Getting to know each other; organization and overview of the course.<br>Big picture (artificial intelligence, The Semantic Web).<br>Data, information, knowledge.<br>Motivation for knowledge representation<br>The Cyc project Google talk by Doug Lenat:<br>http://video.google.com/videoplay?docid=-7704388615049492068. |

## PART II: THE LOGIC-BASED APPROACHES TO KNOWLEDGE REPRESENTATION

*This six-class block covers the "classical" artificial intelligence approaches to knowledge representation and reasoning, based on logic. This is still the predominant approach in the current vision of the Semantic Web.*

| | |
|---|---|
| January 17 | *[Readings: R&N, Sections 7.1-7.6]*<br>Propositional logic. |
| January 24 | *[Readings: R&N, Sections 7.5-7.6, 12.6 (negation as failure)]*<br>Propositional logic: Horn clauses. |
| January 31 | ***** Homework assignment 1 due (propositional logic) *****<br>*[Readings: R&N Chapter 8]*<br>First order logic. |
| February 7 | *[Readings: R&N Chapter 9]*<br>Inference in first order logic.<br>Forward and backward chaining, resolution, production systems. |
| February 14 | *[Readings: R&N Sections 12.1-12-4]*<br>Knowledge representation I: Ontological engineering;<br>Categories and objects; Events; Mental events and mental objects. |
| February 21 | ***** Homework assignment 2 due (first order logic) *****<br>*[Readings: R&N Sections 12.5-12-8]*<br>Knowledge representation II: Semantic networks<br>Description logics; Default logics and truth maintenance systems. |
| February 28 | ***** Homework assignment 3 due (logic-based knowledge representations) *****<br>***** MIDTERM EXAM ***** |
| March 5-11 | Spring recess |

**PART III: THE SEMANTIC WEB**

*In this part of the course, we will look at the main developments in the Semantic Web and tie these to the material that we have covered so far.*

| | |
|---|---|
| March 13 | *[Readings: A&H]*<br>Introduction to the Semantic Web. |
| March 20 | *[Readings: A&H]*<br>The Semantic Web: the XML language. |
| March 27 | *[Readings: A&H]*<br>The Semantic Web: the RDF and OWL languages.<br>Two guest lectures:<br>http://video.google.com/videoplay?docid=-3705570531446755444#<br>http://videolectures.net/koml04_harmelen_o/ |

**PART IV: THE PROBABILISTIC APPROACH**

*The logical approaches to knowledge representation are great whenever there is no uncertainty present. However, they are weak in processing uncertainty, which is predominant in all aspects of real world. The last block of classes focuses on the probabilistic approaches, indicated by Russell and Norvig as the core of the "Modern Approach." In addition to processing of uncertainty, we will look at representation of causality, the fundamental way of storing knowledge about the World in human mind, and decision making. Quite possibly, probabilistic methods are going to play an increasingly important role in the future in all knowledge representation, including the Semantic Web.*

| | |
|---|---|
| April 3 | *[Readings: R&N, Chapters 13,14; **GeNIe** on-line help]*<br>Modeling uncertainty. Bayesian networks<br>Elements of inference in Bayesian networks. |
| April 10 | ***** Homework assignment 4 due (Bayesian networks) *****<br>*[Readings: R&N, Chapter 15]*<br>Time-dependent domains.<br>Causality & causal discovery. |
| April 17 | ***** Homework assignment 5 due (learning/causal discovery) *****<br>*[Readings: R&N, Section 14.6, Chapter 16]*<br>Bayesian networks and first-order logic.<br>Do the numbers matter? Does the structure matter?<br>Decision making under uncertainty. |
| April 24 | ***** FINAL EXAM ***** |

**Sources of readings:**

[R&N] Stuart Russell & Peter Norvig *"Artificial Intelligence: A Modern Approach."* Third Edition, Prentice Hall, 2009.

[**GeNIe** on-line manual] On-line manual of **GeNIe**, a decision modeling environment developed at the Decision Systems Laboratory and available at http://genie.sis.pitt.edu/. Most up-to-date, Wiki version of the documentation is available at http://genie.sis.pitt.edu/wiki/Main_Page.

[A&H] Grigoris Antoniou, Frank van Harmelen. *"A Semantic Web Primer"*, 2nd Edition. The MIT Press, 2008.

## LIST OF HOMEWORK ASSIGNMENTS:

Please see the course schedule for due dates. The assignments can be done in groups. Please list the names of the group members on the first page when submitting the solutions.

**Assignment 1: Propositional Logic**

1. Russell & Norvig, exercise 7.12
2. Russell & Norvig, exercise 7.14
3. Russell & Norvig, exercise 7.18

**Assignment 2: First Order Logic**

1. Russell & Norvig, exercise 8.23
2. Russell & Norvig, exercise 8.24
3. Russell & Norvig, exercise 9.10
4. Russell & Norvig, exercise 9.23

**Assignment 3: Logic-based Knowledge Representation**

1. Russell & Norvig, exercise 12.7
2. Russell & Norvig, exercise 12.8
3. Russell & Norvig, exercise 12.9
4. Russell & Norvig, exercise 12.14

**Assignment 4: Bayesian networks**

*(Submit your solution as a **GeNIe** model with comments through the CourseWeb)*

Construct a simple diagnostic Bayesian network model, consisting of at least 10 variables for a domain of your choice. Subject your model to a test, consisting of three real or realistic cases and describe your observations/conclusions. Example diagnostic domains: Diagnosing a car, a disease or a small set of diseases, diagnosing a colleague whether he/she is addicted to something (e.g., computer games), detecting bank or credit card fraud, etc.

**Assignment 5: Learning/Causal Discovery**

*(Submit your solutions as **GeNIe** models with comments through the CourseWeb)*

Learn Bayesian network models for each of the two following data sets, originating from the Irvine Machine Leaning Repository (http://archive.ics.uci.edu/ml/):

1. Iris (http://archive.ics.uci.edu/ml/datasets/Iris)
2. Congressional Voting Records (http://archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records)

(For your convenience, I have placed both data sets in **GeNIe** format on the CourseWeb.)
The networks should help in classification of iris flowers (1) and congressmen (2) based on flower properties and voting record, respectively. Please clean up and discretize the data, if needed, enter prior knowledge (you can read more about the data sets in the descriptions placed at UCI Repository), and see whether you can make causal sense of the constructed models. Make sure to try different learning parameters and algorithms.