

# Ontology for Organizing Information

## A Practical Introduction

Adam M. Goldstein, PhD, MSLIS

[z\\_californianus@shiftingbalance.org](mailto:z_californianus@shiftingbalance.org)

<http://shiftingbalance.org>

Assistant Professor of Philosophy, Iona College  
<http://www.iona.edu/faculty/agoldstein>

Associate Editor, Darwin Manuscripts Project  
<http://darwin.amnh.org>

LIS 663 Pratt Institute Guest Presentation  
6 December 2010

Ontology: What, Why, and Who?

What is an Ontology?

[slide #1]

## Where Are We, and What's Next?

- 1 Ontology: What, Why, and Who?
  - What is an Ontology?
  - Why use an Ontology?
  - Who can Ontologies be Useful to?
- 2 A Brief History
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery

Ontology: What, Why, and Who?

What is an Ontology?

[slide #2]

## Ontology: A Concise Description

### Ontology

A classification scheme that organizes things by representing them as being the **kinds of things that they are**, and by representing the **kinds of relationships** those kinds of things can have to one another.

- “Ontology,” like “metadata,” “digital library,” and “Open source” is not consistently used.
- So, “ontology” has become almost meaningless.
- Many people feel strongly about how the term should be understood.
- In this philosopher’s humble opinion, getting worked up about the issue is a waste of time.

Ontology: What, Why, and Who?

What is an Ontology?

[slide #3]

## Examples that may be familiar

Netflix Categorizes films by genre, themes, aspects of film plots and characters.

Amazon.com Categorizes books by subject and author.

Facebook.com Describes users in terms of social roles and relationships, places and groups associated with the user, preferences (films, books, etc).

## Examples that may not be familiar

**Clinical Medicine** Describe injury and illness using anatomy, physiology.

**Medical researchers** Describe chemical substances used in pharmaceuticals in terms of their physical properties; coordinate with clinicians to describe drug effects.

**Biologists** Describe subjects of research using species names, names of organs, bones, systems of organs, the names of genes.

**Historians, archivists** Index digital texts by subjects, places, people, concepts, events, relationships among them.

## LoC Subject Headings: Not an Ontology

### LOC Subject Headings

Adams, Abigail, 1744-1818-Marriage.

Adams, Abigail, 1744-1818-Drama.

Adams, Abigail, 1744-1818-Fiction.

- “Marriage” and the like are not represented as kinds of Abigail Adams, which does not make sense.
- Each subject heading is a topic works are *about*, but there is no information about the relationships of the subjects to one another.
- Each heading is particular to Abigail Adams: no means of grouping works about the marriage of Abigail Adams to those about marriages of others.

## Gruber’s Famous “Conceptualization” Definition

### Gruber’s Definition of “ontology”

A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them (*Logical Foundations of AI*, Genesereth, M. R., & Nilsson, N. J., 1987, Morgan Kaufmann Publishers). . . . An **ontology** is an explicit specification of a conceptualization.

## Where Are We, and What’s Next?

- 1 Ontology: What, Why, and Who?
  - What is an Ontology?
  - Why use an Ontology?
  - Who can Ontologies be Useful to?
- 2 A Brief History
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery

## Ontologies are excellent for. . .

- Promoting information discovery in a complex data set, even if that data set is unfamiliar to the researcher;
- Identifying patterns otherwise impossible to detect;
- Integrating your web pages with others, on the Semantic Web;
- Automatically classifying your data using machine reasoning; and
- Automating decision making, also using machine reasoning.

## Where Are We, and What's Next?

- 1 Ontology: What, Why, and Who?
  - What is an Ontology?
  - Why use an Ontology?
  - Who can Ontologies be Useful to?
- 2 A Brief History
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery

## Ontologies are of use to. . .

- Librarians and library administrators—either as researchers or catalogers, or as decision-makers evaluating IT decisions.
- Information Architects, professional or “accidental.”
- Researchers wanting to learn about a large heterogeneous data- or text-base;
- Online retailers, social networking sites, organizations (business, not-for-profit, government) integrating data across departments, offices, projects, other organizations.

## Where Are We, and What's Next?

- 1 Ontology: What, Why, and Who?
- 2 A Brief History
  - Roots in Knowledge Management, AI
  - The Semantic Web
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery

## Early Projects

- Ontology is derived from DARPA sponsored projects in the 1980's and 1990's.
- Recognizing the increased quantity, diversity of information being produced and stored, the intention was to build robots to search large databases.
- If the database describes things and their relationships with one another, the robot can reason, respond to the environment, or to user queries.
- John Gruber is well-known for important early work.

## Where Are We, and What's Next?

- 1 Ontology: What, Why, and Who?
- 2 A Brief History
  - Roots in Knowledge Management, AI
  - The Semantic Web
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery

## Design for Sharing and Reasoning on the Internet

- A 2001 paper in *Scientific American* articulates a model of metadata for web pages designed for machine reading.
- The markup, together with web-accessible databases, forms an ontologically-organized data store at each site.
- If various sites use the same ontology, or a framework for interoperability with other ontologies, they can interact by automated "intelligent agents."
- The agents interact with one another in response to the data they encounter at each site, inhabiting a virtual world consisting of digital entities representing real ones, and bearing the same relationship to one another.

## Where Are We, and What's Next?

- 1 Ontology: What, Why, and Who?
- 2 A Brief History
- 3 Ontological Engineering
  - Philosophical Background to Ontology
  - Design Considerations
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery

## The Study of What Is

- Ontology is the discipline of philosophy devoted to the study of being.
  - “Why is there something rather than nothing?”
  - “What are the most basic constituents of the universe?”
  - “What does science tell us about what there is?”
- Some ontologies (in our sense, not the philosophical sense) aim to describe *everything there is*.
- Most ontology users will not, and should not, worry about whether the objects described by their ontologies are real in any deeper metaphysical sense.
- Be pragmatic: Adopt the most useful description of the domain of interest.

## Where Are We, and What’s Next?

- 1 Ontology: What, Why, and Who?
- 2 A Brief History
- 3 Ontological Engineering
  - Philosophical Background to Ontology
  - Design Considerations
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery

## There Are Many Perspectives on the World . . .

- Objects can be described at different levels of “granularity.”
  - A giant boulder.
  - Bit of mineral substances crushed under the weight of Earth, heated to extreme temperatures, etc.
  - Atoms that fuse together because of their chemical structure to form bits of mineral substances.
  - Where does it ever end, if it ends at all . . .
- Different backgrounds, purposes among ontology engineers and users can result in different ontologies describing the same events, things, relationships.
  - Consider: Differences among ontologies most useful for Fox News, Greenpeace, an oil corporation.

## Where Are We, and What’s Next?

- 1 Ontology: What, Why, and Who?
- 2 A Brief History
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
  - Overview
  - A Medical Example
- 5 Ontology for Mining & Information Discovery

## Ontologies Make Sharing & Coordinated Action Possible Online

- Organizations that want to share data and interact with other organizations using the web agree on a standard for the content, implementation of ontologies they will use.
- Intelligent agents shuttle between the various linked sites.
- They read and process data they find at each site.
- By way of scripts that they can execute over the web, they query databases at each organization, and add information to them as well.

## Where Are We, and What's Next?

- 1 Ontology: What, Why, and Who?
- 2 A Brief History
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
  - Overview
  - A Medical Example
- 5 Ontology for Mining & Information Discovery

## Creating and Processing the Patient Record

- A patient is treated for a broken leg. The diagnosis is entered into an electronic patient record.
- The record is processed by an automated machine reasoner.
- The reasoner has access to a database of diagnoses, medical specialities, and treatments.
- It determines that a broken leg is treated by an orthopedist, and that follow-up X-rays are required three weeks after the injury.

## Exchange Among Agents Using Structured Data

- The hospital's machine reasoner ("agent") contacts the patient's insurance company, transmitting the information just described.
- Shared data structures make it possible for the two reasoners to communicate—contrast with an email message or unstructured keyword list.
- The insurance company's agent queries the company databases.
- The agent identifies orthopedists, radiologists based on location, availability, areas of expertise, the patient's preferences (if known).
- This information is transmitted to the patient, who can select from among the proposed choices.

## Where Are We, and What's Next?

- 1 Ontology: What, Why, and Who?
- 2 A Brief History
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery
  - Overview
  - Data Mining: Identifying Patterns
  - Information Discovery: Finding Relevant Information

## Lost in a Sea of data and Text?

- The explosive growth of the amount of information results, paradoxically, in a reduction in access to it.
- The barriers to access:
  - Heterogeneity Linguistic, cultural, diversity of media, historical, diversity almost equal to that of the entire human species.
  - Volume Even a large group of people organized solely for the purpose of reading, listening to, watching all the works available could never complete the task in a lifetime.
  - Usual suspects Synonyms, changes in the meaning of terms, no authority control, the “who moved my cheese” problem.

## Where Are We, and What's Next?

- 1 Ontology: What, Why, and Who?
- 2 A Brief History
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery
  - Overview
  - Data Mining: Identifying Patterns
  - Information Discovery: Finding Relevant Information

## Digitized Works Crawling with Intelligent Agents

- OCR technology, digitization of scientific, archival data, makes an enormous quantity of data accessible to machine reading.
- Using natural language processing, texts can be automatically categorized ontologically.
  - This is astonishing, but true, thankfully.
- Intelligent agents can be programmed to identify patterns.
- Exploring the information organized by the ontology as they would if they were physically real, they can learn about the things represented.
- Think about the medical example.

## Where Are We, and What's Next?

- 1 Ontology: What, Why, and Who?
- 2 A Brief History
- 3 Ontological Engineering
- 4 Ontology on the Semantic Web
- 5 Ontology for Mining & Information Discovery
  - Overview
  - Data Mining: Identifying Patterns
  - Information Discovery: Finding Relevant Information



## Learning by Browsing & Eclectic Searching

- The ontology represents things and relationships among them as they are in the real world.
- A human “intelligent agent” can explore an ontology category by category, learning about the subject matter as he or she does so.
- Relationships between categories show how the latter are connected, how the things are structured relative to one another.

Film library example Organize films by genre; country of origin; plot; characters' personalities, gender, profession; film stock, production company.

- Really, not so different from how someone can use subject headings to learn about a subject.

## References

-  BERNERS-LEE, TIM, JAMES HENDLER, and ORA LASSILA. 2001. The semantic web. *Scientific American* 284: 34 –43.
-  GRUBER, T.R. 1995. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies* 43: 907–928.