Ontologies and Folksonomies: False Friends

Chris Welty
IBM Research

False Friends

- Words that appear on the surface to be the same but are different
- Red
  - Spanish: network
  - English: color
- Embarrass/Embarazar
  - Spanish: To impregnate
  - English: To make someone feel stupid
  - Pen slogan

    "It won't leak in your pocket and embarrass you"

Outline

- Brief history of classification
- Brief history of ontology
- Briefer history of folksonomies
- Clearing up some confusions
- From False Friends to Friendship

Classification – Part I

- Dates back to Aristotle (Categories – 350 BC)
  - All western science proceeds from this effort
- *Subject* Classification roughly 400 years old
- Used in libraries to organize books
  - For *librarians* to place/find books (not users)
  - Special training required (you're meant to ask)
  - Still today: *primary* subject heading for placement of physical objects
  - Still today: mostly linear organization
- US system (LOC)
  - Similar to Dewey Decimal
  - based on Thomas Jefferson's personal library organization
  - Based on Didier&D'Alembert *Encyclopedia*
  - Based on Bacon's Tree of Knowledge

Classification – Part II

- Subject classification makes intuitive sense
  - But classifications are in the eye of the beholder
  - 20 people, 20 hierarchies, hundreds of classifications per item
  - Librarian-based approaches don't scale
- COLON system
  - Single set of keywords (facets)
  - Any combination legal
  - The conceptual origin of existing folksonomy systems

Ontology

- Dates back to Aristotle (Categories)
  - Metaphysics (after physics)
  - Ontologia (study of being) >400 years old
- Generally taken to refer to the study of "What kinds of things exist and how they are related"
- *Formal* Ontology
  - Remove human perception (objective)
- A Basic Formal Ontology
  - Events, Objects
  - Objects *participate* in Events

Neo-Ontology

- Drawn into AI in early 80s [McCarthy, 1980]
- A specification of a conceptualization [Gruber, 1993]
- Any logical theory [Walter, 2005]
- Anything expressed in OWL [W3C, 2003]
- A *description* of the data
  - A database schema
  - An object model
  - A UML diagram

Neo-neo Ontology

- Anything to do with reasoning?

Folksonomies

- The result of a social tagging process
  - From scale & data structure emerges
- Built by communities
  - Decentralize the classification problem
- Requires some “training”
  - But clearly more focused
  - Individuals have more control
  - No “one right” way
- Unproven, but popular
- Some want to call them ontologies
  - Why?

A Common Misconception

Not all partial orders are subclass

- Mereological relations (part of)
- Spatial relations (containedIn, connected)
- Temporal ordering (before, after)
- etc…

Ontology Defined

- ~~A specification of a conceptualization~~ [Gruber, 1993]
- ~~Any logical theory~~ [Walter, 2005]
- ~~Anything expressed in OWL~~ [W3C, 2003]
- ~~A *description* of the data~~
- Objective description of the kinds of entities there are and how they are related. Good Ontologies provide:
  - **Meaning**
  - Organization
  - Taxonomy
  - Agreement*
  - Common Understanding
  - Vocabulary
  - Connection to the "real world"

Classification and Ontology

- All ontologies include some form of classification
- Not all classifications are ontologies

- Ontologies are an *objective* description of the kinds of entities there are and how they are related
- Classifications organize/group things according to *subjective* criteria

A (Formal) Ontology

- Universal
  - Relation
- Particular
  - Substance
  - Quantity
  - Quality
  - Location
  - Time
  - Action

Subject Taxonomy

- Universal
  - Relation
- Particular
- History
  - Quantity
  - Quality
  - Location
- Art
- Politics

Philosophy of Science

History

History

History

Italian Military History

Entity-ness

About-ness

Ontology

- Ontologies describe entities in the world
  - They do NOT describe classes
- Taxonomies in an ontology are typically strict trees, with disjointness
- Classes in ontologies tend to be *rigid*
  - Compare being a person to being a scientist
- Membership is not subjective
- Identity is crucial

Useful distinction

- Ontologies describe the *invariant structure* of a domain
  - What are the fundamential types of things in the domain?
- IMDB: web pages, movies, people
  - Not: action, drama, horror, science fiction
- So what is the ontology of subjects?

Subjects

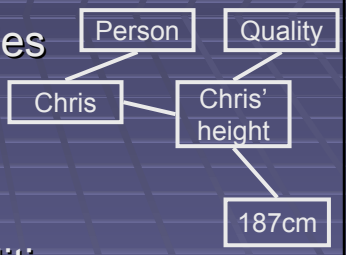
- Subjects are sometimes hierarchical
- More commonly *faceted*
- Subjects are personal
- Not just for books
- Particulars can be subjects
- Subjects can be related in many ways
- General topics

Subjects and Topics

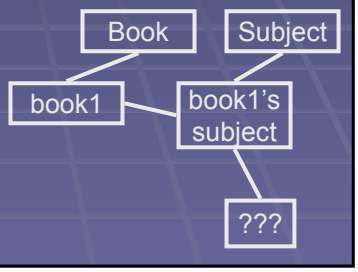
- …Aristotle again
- Topic
  - A place to go for information
  - Common and Special
- Subject
  - “To throw under” (Fr.)
  - A thing that is under, at a topic
- Both words (in English) have a spatial connotation
- Are subjects spatial?

Qualities

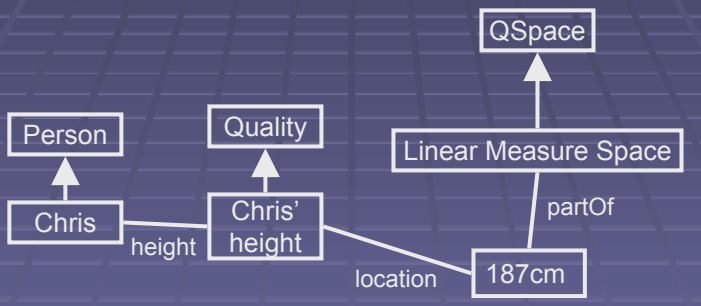
- Qualities are dependent entities
  - E.g. “The height of Chris”



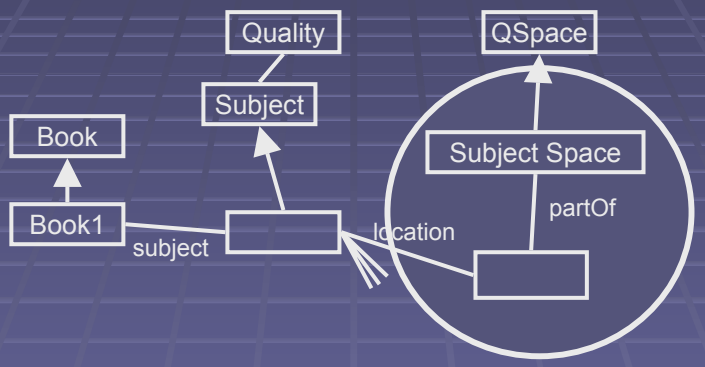
- Subjects can be seen as qualities
  - E.g. the subject of this book
  - But what is the “value”?



Quality Spaces



Subject Spaces?



What is this?

Subject Spaces

- A subject space is a multidimesional space whose axes are words (tags)
- Points in the space are bag of words vectors
- Certain axes (tags) may be highly correlated indicating containment, overlap, nearness
- Regions of the space represent general topics
- Provides a flexible, dynamic open way to assign and retrieve subjects
- This is the folksonomy

Subject Ontology

- Subjects are spatial
  - “Subjects are places themselves…” (Dewey)
- Books, documents, etc. are *placed* there
- The relationships are mereotopological, not taxonomic
  - Contains, overlaps, borders, near, far
  - Forcing a taxonomy actually limits the value of f-omy
- The space of subjects is n-dimensional
  - Can be assigned with tags
  - The *space* is not an ontology, it is a topology
  - This is the folksonomy - topological not ontological