Outline

- Introduction to Ontologies
  - Definition
  - Web Ontology Language (OWL)
- Ontology Generation from Tag Spaces
  - The Problem
  - Tag Meta-Data
  - Ontology Clusters
- Ontology-Based Federation of Data
  - Data Types
  - System Architecture
Introduction to Ontologies
Ontology

- Definition
  - An ontology is a specification of conceptualization

- Practical Used as a form of Knowledge Representation
  - Semantic Web, Software Engineering, Artificial Intelligence, Information Architecture

- Taxonomy
  - A Simple Ontology

- Differences between a classification and an ontology
  - The richness of information available
Elements of an Ontology

- **Individuals**
  - Instances

- **Ground level component of an ontology**

- **Concrete Object**
  - People, Animal, Planets

- **Abstract Object**
  - Numbers, Words
• Classes (Concepts)
  • Abstract groups, sets, or collection of objects
  • Contain individuals, other classes or combination of both
  • Person: the class of all People
Elements of an Ontology 3

- **Attributes**
  - At least a Name and a Value
  - A complex data type

- **Example**
  - **Object**: a Ford Explorer
  - **Attributes**:
    - **Name**: Ford Explorer
    - **Number of Doors**: 4
    - **Engine**: {4.0L, 4.6L}
    - **Transmission**: 6 speed

- **4-Wheel Drive**
- **Ford Explorer**
Elements of an Ontology 4

- **Relationships**
  - Important type of relation is the subsumption:
    - is-subclass-of, the converse of is-a, is-type-of, is-subclass-of
  - **Example**
    - **Object:** Ford Branco
    - **Attribute:** Successor: Ford Explorer

![Diagram](image)
Why Ontologies?

- Sharing common understanding of the structure of information among people or software agents
- Reusing of domain knowledge
- Making domain assumptions explicit
- Separating domain knowledge from the operational knowledge
- Analyzing domain knowledge
Ontology Languages

- **Web Ontology Language (OWL)**
  - A formal language used to encode the ontology
  - To process the content of information instead of presenting information
  - Supported by
    - XML: provides a surface syntax for structured documents (no semantic constraints)
    - XML Schema: Restricting the structure of XML document
    - RDF: A data model for objects and relations between them
    - RDFS: A vocabulary for describing properties and classes of RDF resources
OWL Sublanguages

- **OWL Lite**
  - Support a classification hierarchy and simple constraints
  - A quick migration path for thesauri and other taxonomies
  - Lower formal complexity

- **OWL DL**
  - Support the maximum expressiveness while retaining computational completeness and decidability
  - Including all OWL language constructs, Using only under certain restrictions

- **OWL Full**
  - Maximum expressiveness and the syntactic freedom of RDF with no computational guarantees
Published Ontologies

- Dublin Core
  - A simple ontology for documents and publishing
- WordNet
  - Lexical reference system
- Gene
  - Ontology for genomics
- SBO
  - Systems Biology Ontology for computational model in biology
- LinkBase
  - A formal representation of the biomedical domain, founded on BFO (Basic Formal Ontology)
- FOAF Friend-of-a-Friend
Ontology Generation from Tag Spaces
Tags

- A relevant keyword or term associated with or assigned to a piece of information

- Describes the item and enabling keyword-based classification of information it is applied to

- is usually chosen informally and personally by the author/creator or the consumer of the item
Problem: searching the TagSpace

How would you tag this?

How would you search for it?

Tags: Ikura, Uni, Ebi, Sushi, Nigiri, Japanese food, lunch in Tokyo, Ezobafun-uni, Kitamurashiuni, Murasakiuni, Akazaebi, Tenagaebi, etc.
Problem: exploring the TagSpace

This is a tag cloud - a list of tags where size reflects popularity.
sort: alphabetically | by frequency

design  blog  software  reference  music  programming
news  tools  web  shopping  webdesign  web2.0  art  linux
howto  ajax  css  video  mac  javascript  games  google
travel  photography  technology  humor  politics  internet  science  tech
windows  search  free  tutorial  media  opensource  books  fun  blogs  funny
security  java  apple  business  flash  hardware  firefox  php  development  tips
entertainment  research  history  education  productivity  inspiration  computer  cool  daily
osx  community  graphics  toread  microsoft  audio  tv  rss  home  photo  mp3  radio
maps  computers  sports  food  personal  social  culture  health  diy  movies  magazine
lifehacks  finance  ruby  gadgets  photos  mobile  email  reviews  tutorials  geek  network
download  html  language  p2p  marketing  freeware  interesting  images  webdev  mozilla  privacy
wiki  game  hosting  misc  online  kids  article  bittorrent  downloads  illustration  seo  photoshop
electronics  python  gid

(red tags are tags you share with everyone else)
“Tag Relations improve searchability and exploration.”

Similar Tags
- Spelling and morphology
  - macos <-> mac_os <-> mac os; tagging <-> tags <-> tagged,
- Synonyms:
  - macos <-> tiger; films <-> movies; new york <-> nyc;
- Related:
  - cooking <-> recipes, software development <-> programming,

Tag Groups or Subtags
- Location -> san francisco, london, new york, etc.
- Food -> sushi, sashimi, pizza, etc.
- Programming -> html, java, css, etc.
1. Get Meta Data
2. Build Tag Relation Graphs
Compute Similarity

Similarity measures

- Matching: $|A \cap B|
- Dice: $\frac{2|A \cap B|}{|A| + |B|}$
- Jaccard: $\frac{|A \cap B|}{|A \cup B|}$
- Overlap: $\frac{|A \cap B|}{\min(|A|, |B|)}$
- Cosine: $\frac{|A \cap B|}{\sqrt{|A| \times |B|}}$
Ontology-Based Federation of Data
The Problem

- Representation of Geo-scientific Data
  - Different Data Sources
  - Different Data Representations
  - Different Data Types

- Facilitate the use of this data
Proposed Architecture

Scientist Portal

Filtered Information, Services

Delivery Assembly

Information Filtering

User Models

Domain Ontology

Source Reconciliation

Ontologies for Information Sources Connected To Domain Ontology

Sources: INSAR GPS, etc.
The Data

- Global Positioning System (GPS)
The Data

- Interferometric Synthetic Aperture Radar
  - a.k.a InSAR
QuakeTables

Software Architecture

Web Services & Web Feature Services

InSAR → GPS → EQ Fault → OWL

Mapping Software

DB

Internet
Questions ...?