Semantic Enhancement of Polar Cyberinfrastructure

Wenwen Li

Assistant Professor, GeoDa Center for Geospatial Analysis and Computation

School of Geographical Sciences and Urban Planning

Arizona State University

06-02-2014

Towards a 'CyberPolar' Community

- Studies about polar regions has attracted great attention:
  - Increasing interest in mining and natural resource exploration
  - Polar regions are key drivers of the Earth climate
  - Both poles are sensitive to human activities and global, environmental, and climate changes
- "President's National Strategy for the Arctic Region"
  - Released in May 2013 by the white house
  - Overarching stewardship objectives:
    - "Make decisions using the best available information"

"Making decisions with best available information"

- Accessibility
  - Given the fact that polar dataset are widely dispersed on the Internet, how to improve their accessibility?
- Discoverability
  - Given a large collection of dataset, how to identify the most relevant data to some research?
- Integratability
  - How to integrate heterogeneous dataset and transform them into useful information?

Building a Service-Oriented Polar CI Portal to Support Sustained Polar Sciences

- Funded by National Science Foundation (NSF)
- Four objectives:
  - Objective 1: Establish a large-scale web crawler to improve the accessibility of Big Polar Data
  - Objective 2: Semantic reasoning to improve the discoverability of relevant data residing in the polar CI portal
  - Objective 3: Monitoring Data and Service Quality to Improve Portal Usability
  - Objective 4:Rich Internet Application (RIA) based Graphic User Interface (GUI)

Outline

- PolarHub
  - Large-scale Web crawler to find polar data
- Semantic Search
  - Ontology-based semantic reasoning
  - Semantic mining for intelligent search
- Polar CI portal
  - Share, integration, visualization and analysis of polar data

Accessibility

Discoverability

Intergratability

Improving accessibility of polar data through PolarHub

http://polar.geodacenter.org/polarhub

W. Li, C. Yang and C. Yang, 2010. An active crawler for discovering geospatial web services and their distribution pattern, *International Journal of Geographic Information Science*, 24(8), 1127-1147.

PolarHub: A Large-scale Web Crawler for Polar Data Discovery

Crawler Dashboard | Search | Miscellaneous Crawlers

Task List

Create | Start/Resume | Pause | Delete | Realtime Crawling | Realtime Statistics | Show All crawl | UploadCSW

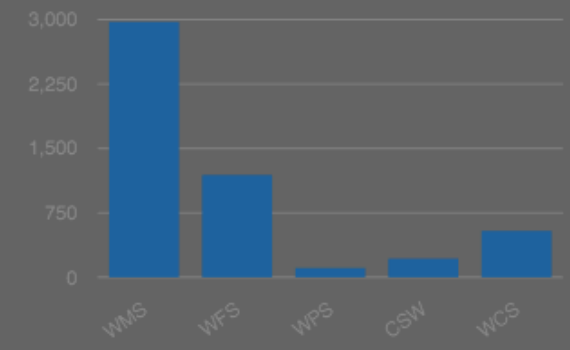
		Item ID	Keywords	User Name	Role	Create Time
1		96	permafrost	guest	1	Wed Mar 19 15:39:43 MST 2014
2		97	snow' 'ice	guest	1	Wed Mar 19 15:40:02 MST 2014
3		98	alaska' 'WMS	guest	1	Wed Mar 19 15:40:14 MST 2014
4		99	antarctic' 'ship	guest	1	Wed Mar 19 15:40:37 MST 2014
5		111	web' 'map' 'service	yifei	1	Fri Mar 28 13:38:42 MST 2014
6		23	web' 'map' 'service' 'WMS' 'Clearinghouse	admin	2	Fri Jan 31 16:08:22 MST 2014

Realtime Statistics

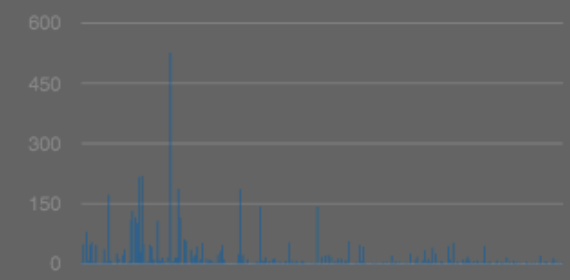
Total Number Of Services 5026
Total Number Of WMS 2973
Total Number Of WFS 1190
Total Number Of WPS 108
Total Number Of CSW 216
Total Number Of WCS 539

AllTask

WXS vs Number of Services



Location vs Number of Services



Geospatial semantic search to improve discoverability

A hybrid approach

Geo-Ontologies

- Definition
  - classes and individuals for representing e.g. geospatial objects, their properties, and mutual relationships.
- Ontology
  - Plays an important role in establishing robust theoretical foundations for GIScience in the future

SWEET 2.0

INSPIRE Ontology

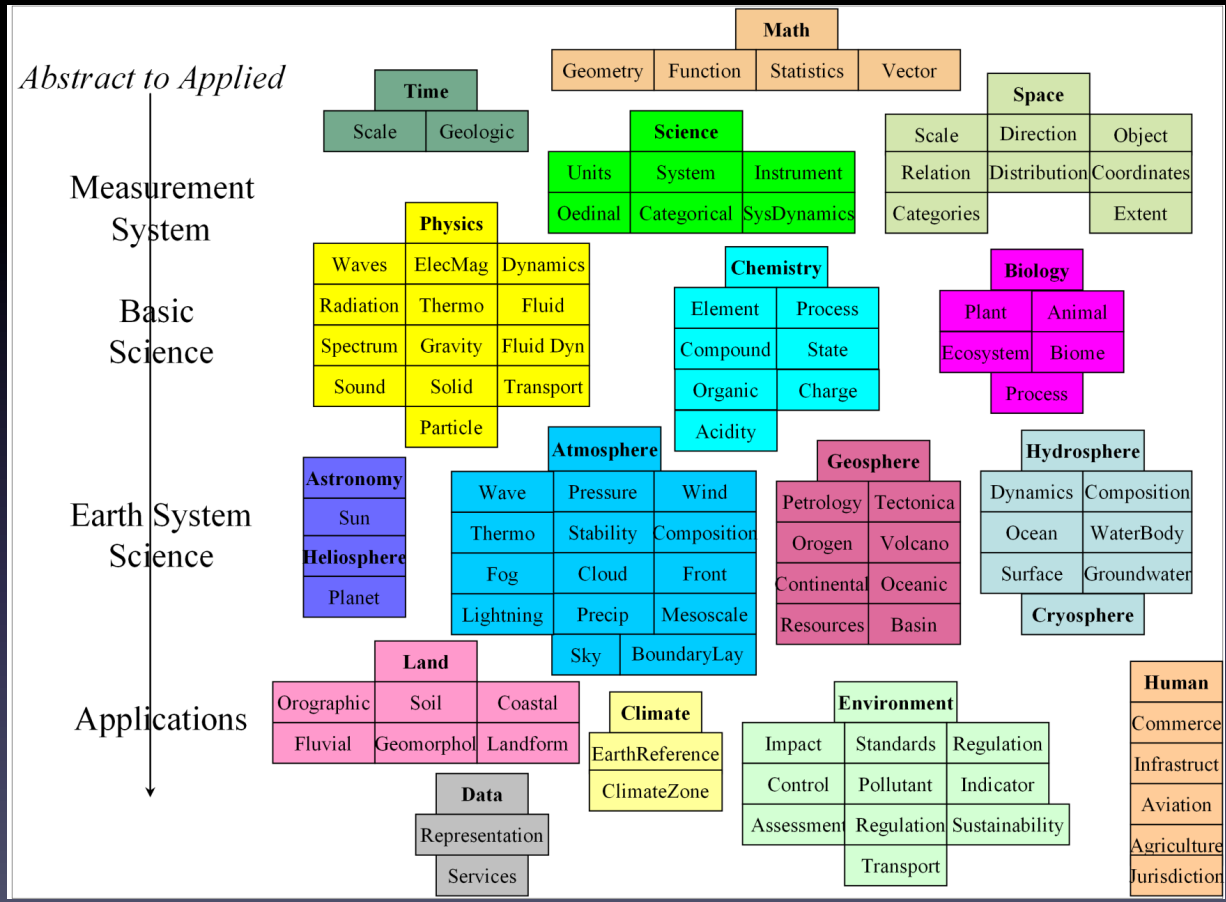
NASA GCMD Science Keyword

Dbpedia Geolinked data

GEMET Thesaurus

CUASHI Ontology

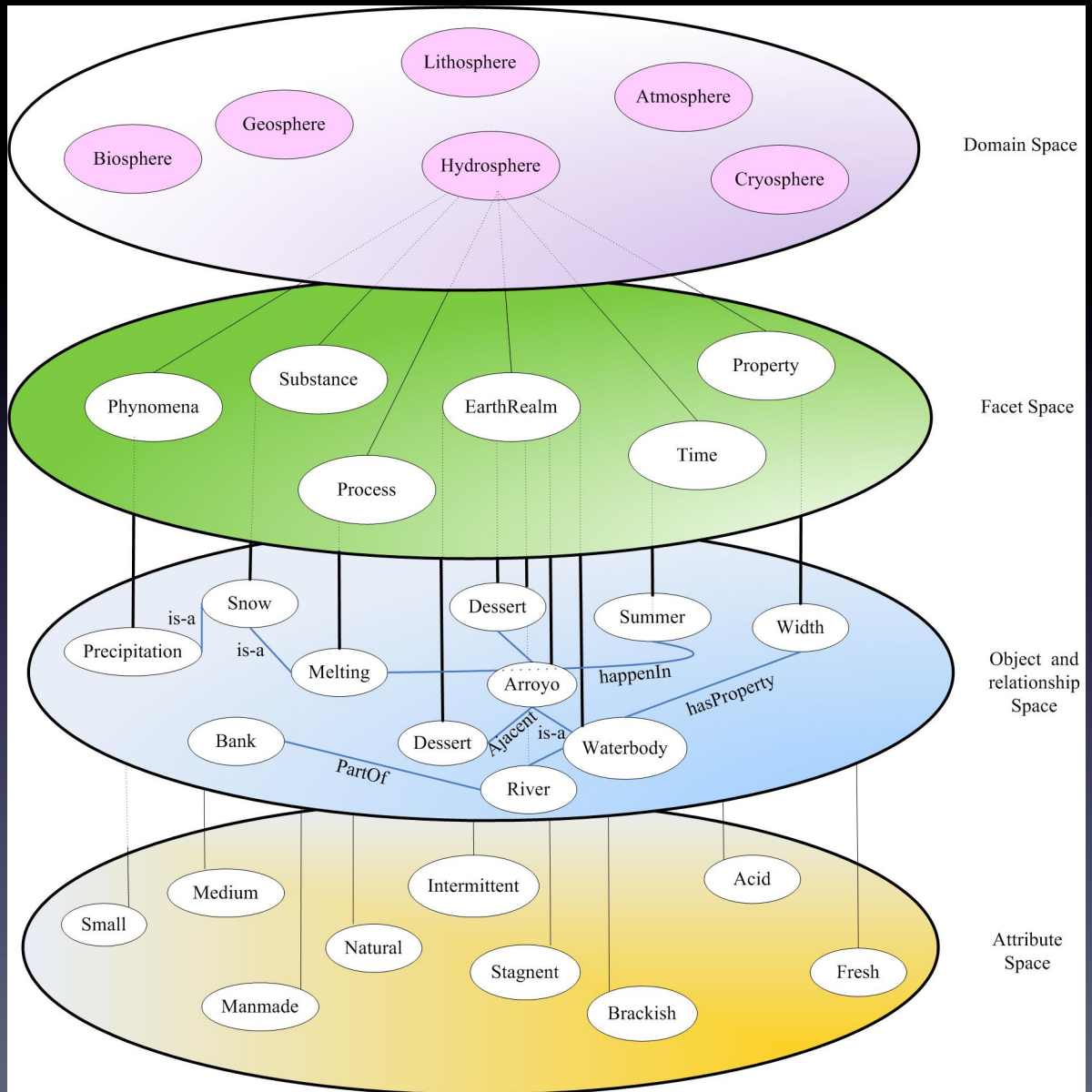
MMI Ontology



SWEET: Semantic Web for Earth and Environmental Terminology

Raskin, R.G. and Pan, M.J., 2005. *Computers & Geosciences*, 31, pp. 1119-1125.

Conceptual Framework for a Domain Ontology

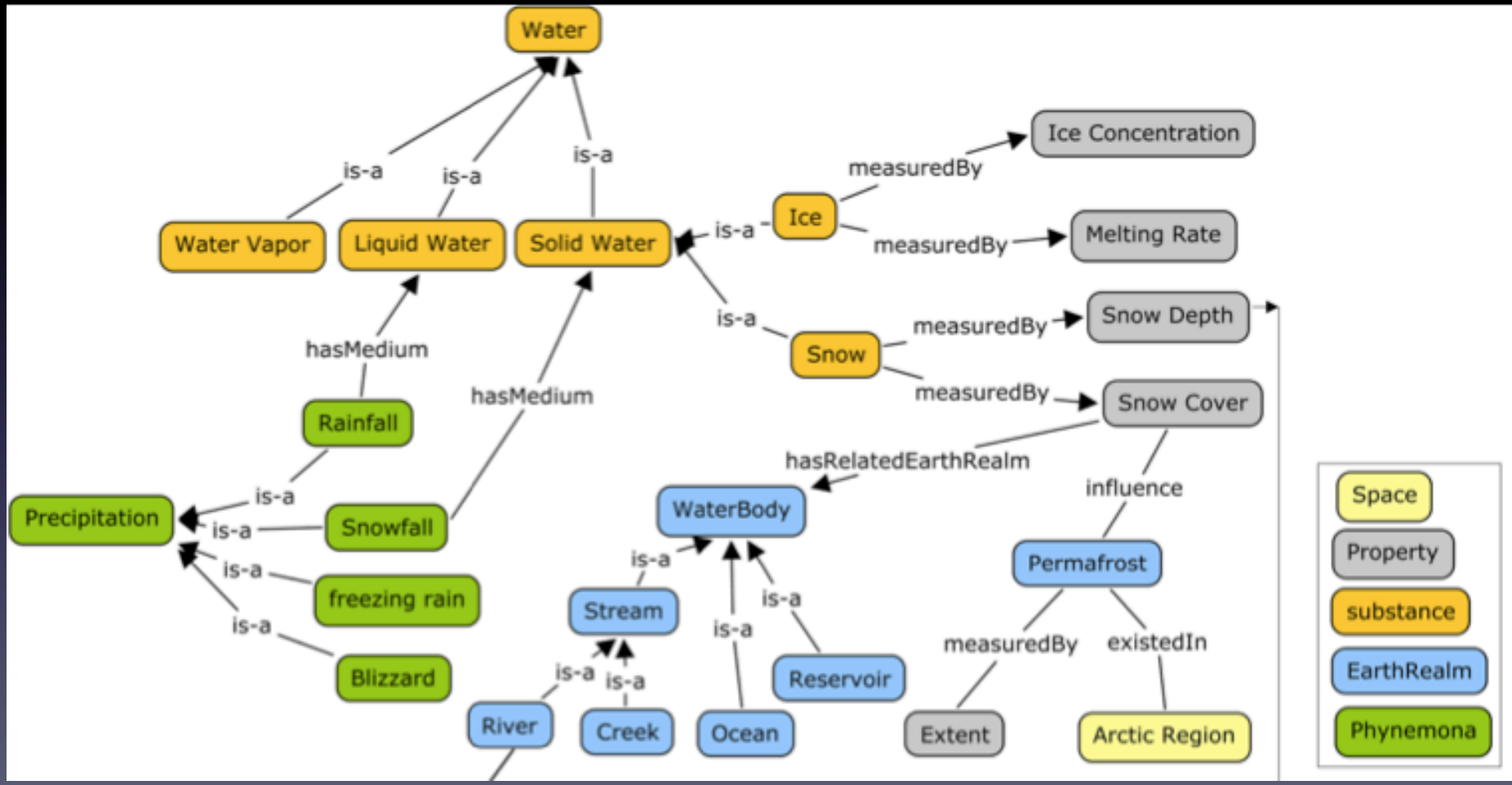


GeoOntology

- ✓ Terminology
- ✓ Attribute Space
- ✓ Value Space

<Subject, Predicate, Object>

An Example of Polar Data Ontology



Logic Reasoning (1)

Case study: "***How does solid water melt influence stream flow in the Arctic Region over the summer time?***"

- Syntax Analysis – Query decomposition
  - Component
    - What- "Solid water"
    - Place - "Arctic"
    - Process – "Change"
    - Time – "Summer"
  - Description logic-based query

Logic Reasoning (1)

- Semantic Analysis

Q1a: **SomeSWClass** $\cap \exists$ *isSubClassesOf* ."Solid Water"

Q1b: (AProperty $\cap \exists$ *isSubClassOf* ."Property")

$\cap$ (AProperty $\cap \exists$ *is* Pr *edicateOf* .SomeSWClass)

$\cap$ (**Parameter** $\cap \exists$ *isObjectOf* .SomeSWClass)

Q1c: **SomeStreamClass** $\cap \exists$ *isSubClassesOf* ."Stream"

Q1d: (Parameter $\cup$ SomeStreamClass).hasData $\cap \forall$ *takePlaceIn*."Arctic" $\forall$ *hasTime*."Summer"

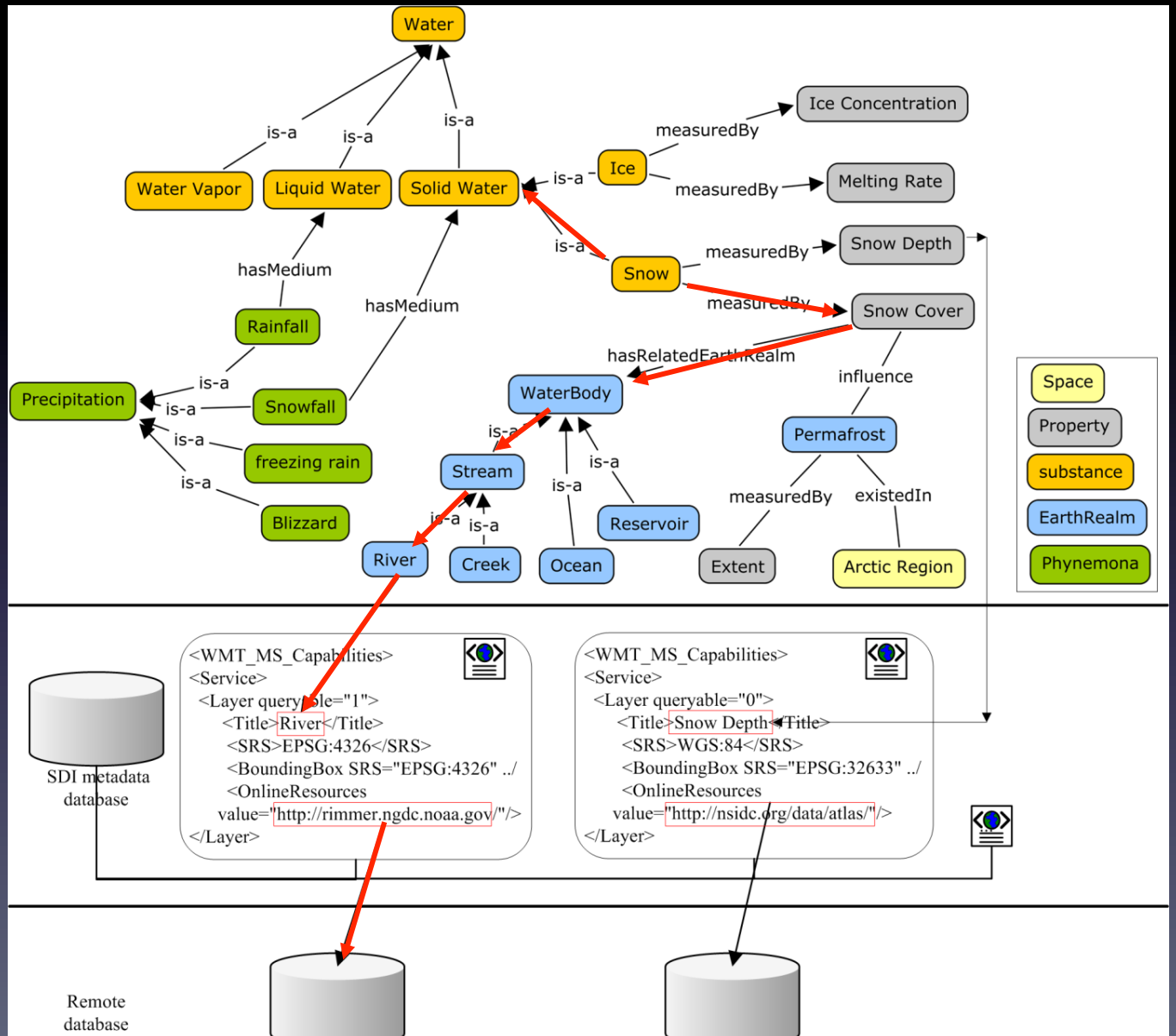
- Q1a and Q1c aim to find <n1, n2…nk> of all of the subclasses and other related terminologies of the given terms.
- Q1b are formed by checking all of the roles that are connected with "SomeSWClass" and the connected predicate is a type of "Property".
- Q1d redirects the query request with the values of desired variables as the keyword to data repository

Logic Reasoning (3)

- Formal Query – Machine language
  - Q1d

```
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:  <http://www.w3.org/2002/07/owl#>
PREFIX PhenomenaNS: <http://localhost/ontology/phenomena.owl#>
PREFIX PropertyNS: <http://localhost/ontology/property.owl#>
PREFIX SubstanceNS: <http://localhost/ontology/substance.owl#>
PREFIX EarthRealmNS: <http://localhost/ontology/earthrealm.owl#>
PREFIX ProcessNS: <http://localhost/ontology/process.owl#>
SELECT *
    WHERE {
    ?Parameter PropertyNS:hasData ?data
    ?data PropertyNS:takePlaceIn 'Arctic'^^xsd:String
    ?data PropertyNS:hasTime 'Summer'^^xsd:String
```

Logic Reasoning to Answer Science Question



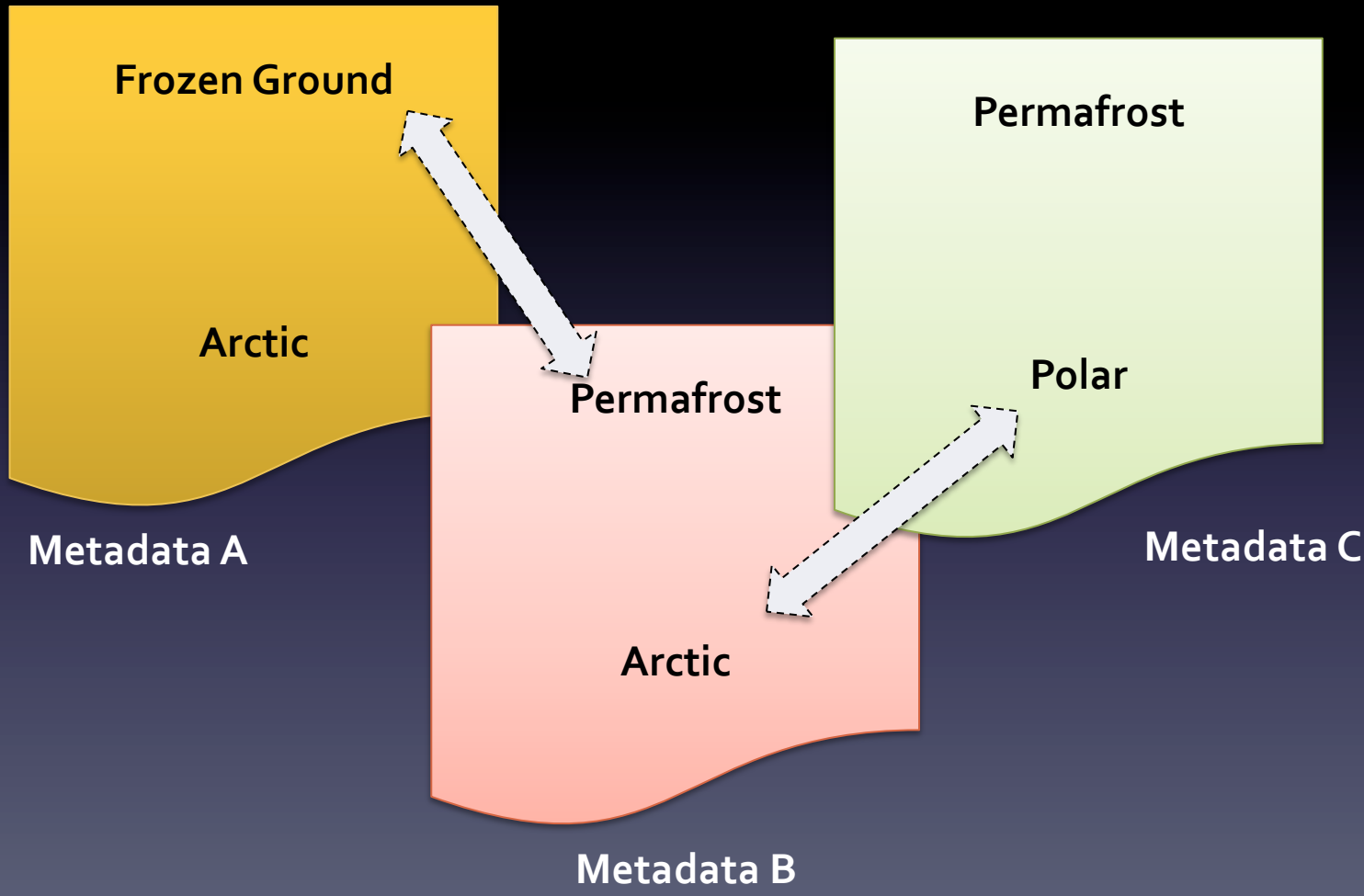
Challenges of the Ontology Approach

- Hard to model spatial relationships
  - Equal, within, touch, disjoint, intersect..
- Hard to build a consensus domain ontology
- Limited spatial reasoning capability

A New Knowledge Mining Approach

- Latent Semantic Analysis
  - Learning knowledge, hidden semantic associations from data
  - Assume co-occurred words are more semantically related
  - Utilize mathematical modeling

Latent Semantic Analysis



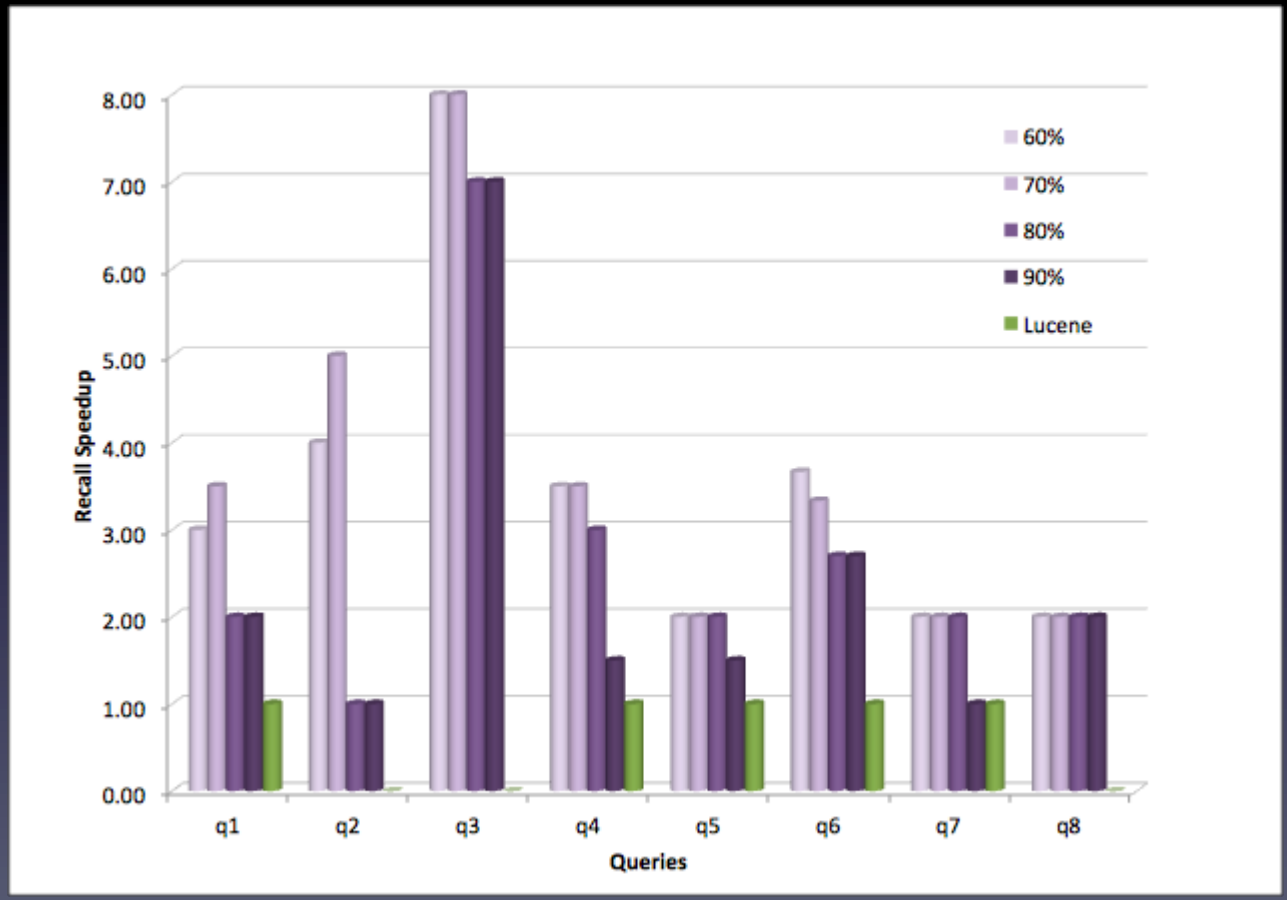
W. Li, M.F. Goodchild and R. Raskin, 2012b. Towards geospatial semantic search: exploiting latent semantic relations in

Queries used

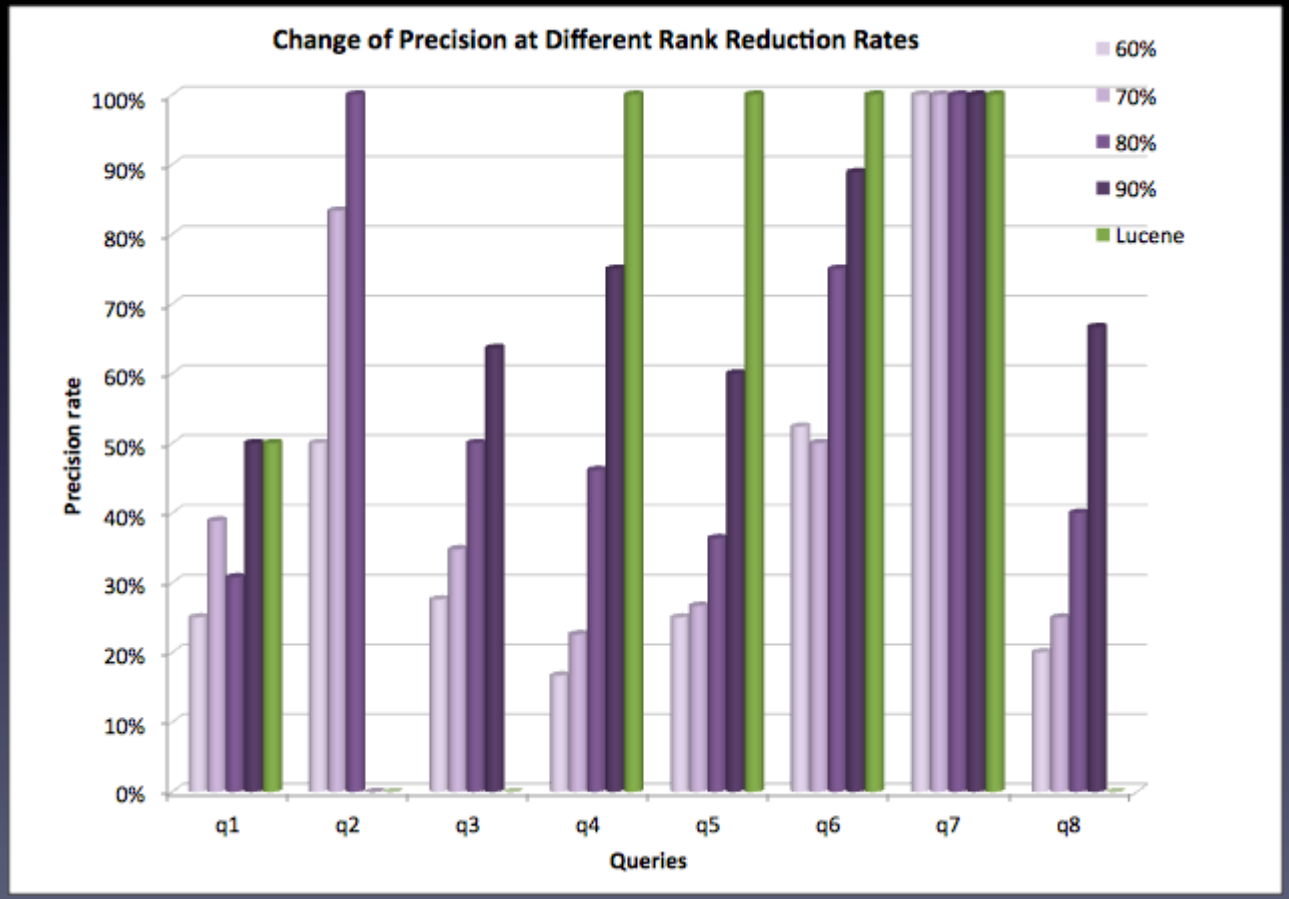
- Q1: offshore hydrocarbon
- Q2: offshore hydrocarbon mining
- Q3: Marine biology data
- Q4: Frozen ground Canada
- Q5: Sediment data Yukon
- Q6: Canada air pilot
- Q7: Earthquake magnitude data Arctic
- Q8: Sea level increase climate

W. Li and V. Bhatia, 2014 (in revision). Toward Polar Cyberinfrastructure: Enabling Semantic Search in Metadata Catalogue to Enhance Polar Data Discovery. Earth Science Informatics

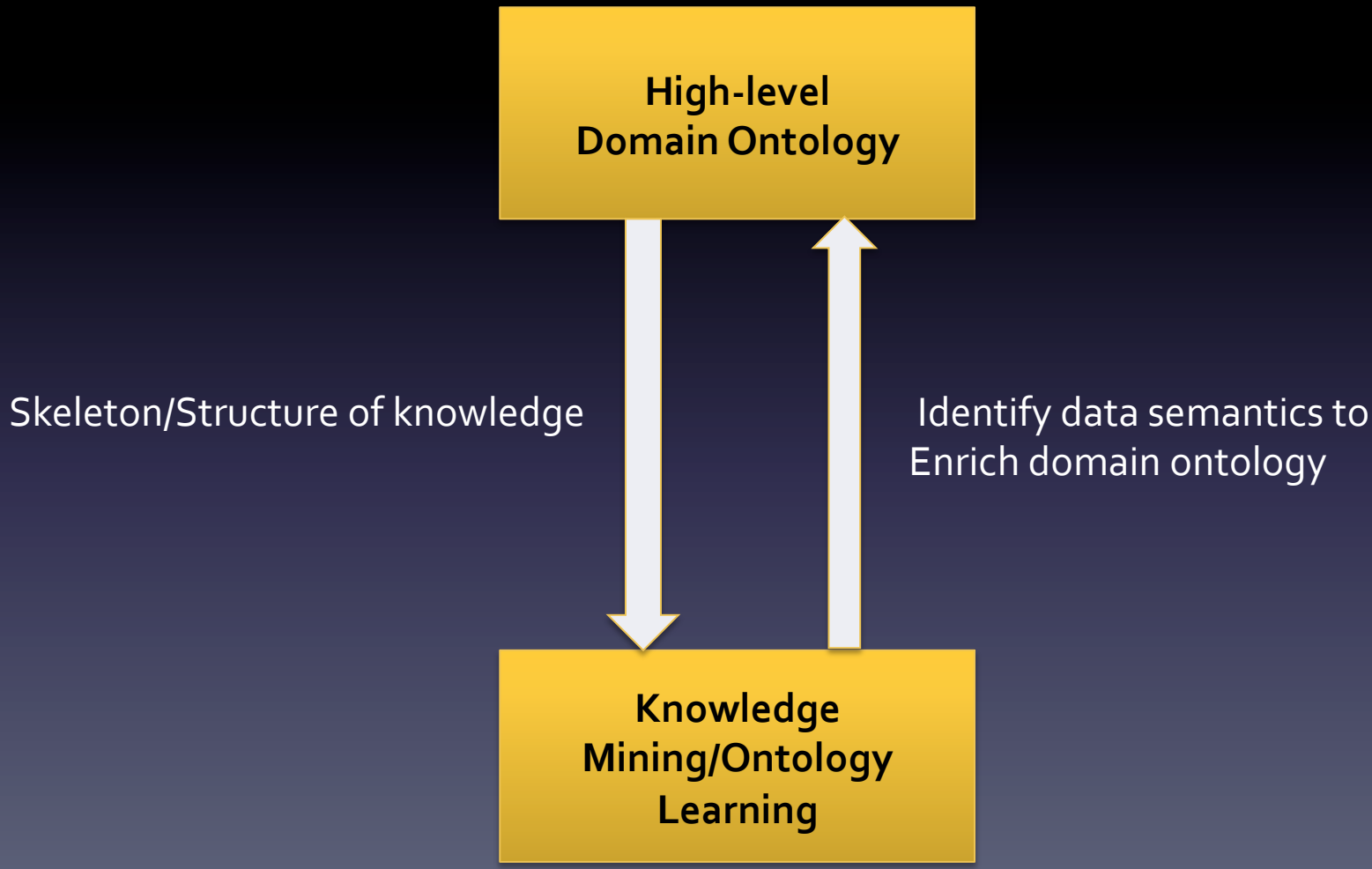
Comparison of Recall



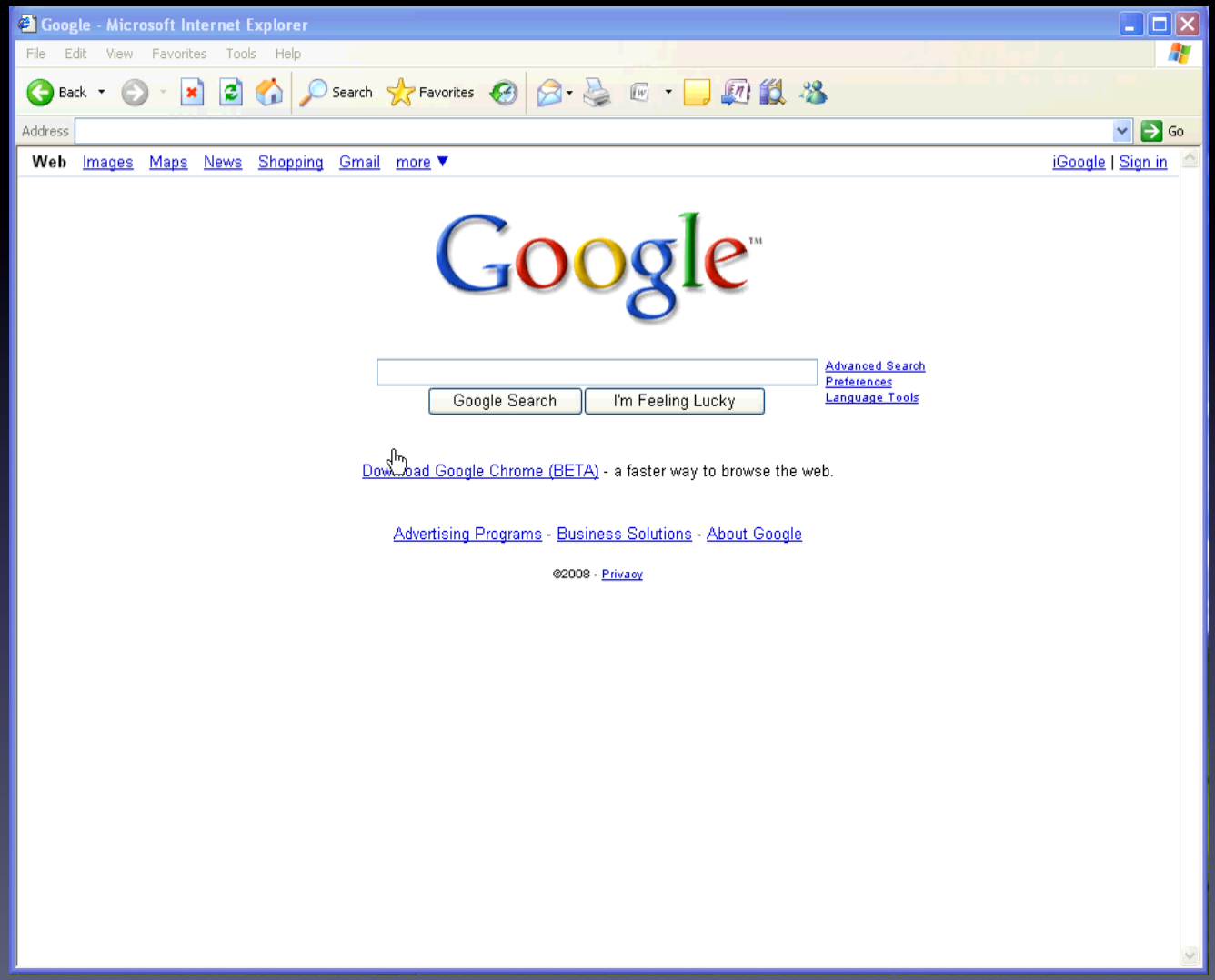
Comparison of Precision



A Bigger Picture on Semantic Search



Demo


Google - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address
Go
Web Images Maps News Shopping Gmail more
iGoogle | Sign in
Google
Advanced Search
Preferences
Language Tools
Google Search
I'm Feeling Lucky
Download Google Chrome (BETA) - a faster way to browse the web.
Advertising Programs - Business Solutions - About Google
©2008 - Privacy

Polar CI Portal to Enhance the Integratability of Polar Data/ Services

Demo

References

- W. Li, Automated Data Discovery, Reasoning and Ranking in Support of Building an Intelligent Geographic Search Engine, Ph.D. Dissertation. George Mason University, August 2010.
- W. Li, C. Yang, D. Nebert, R. Raskin and H. Wu, 2011. Semantic-based service chaining for building a virtual Arctic spatial data infrastructure. Computers & Geosciences. 37(11), 1752-1762.
- W. Li, R. Raskin and M.F. Goodchild, 2012a. Semantic similarity measurement based on knowledge mining: an artificial neural network approach. International Journal of Geographic Information Science, 26(8), 1415-1435.
- W. Li, M.F. Goodchild and R. Raskin, 2012b. Towards geospatial semantic search: exploiting latent semantic relations in geospatial data. International Journal of Digital Earth, DOI: 10.1080/17538947.2012.674561.

Acknowledgement

- National Science Foundation
- Vidit Bhatia, Arizona State University
- Miaomiao Song, Arizona State University

Special Issue Announcement

- "Semantic Interoperability in GIScience" with Transactions of GIS
  - Big Data semantics
  - Geospatial semantic search
  - Ontological and uncertainty issues in geospatial data
  - Provenance and metadata for spatial analytical methods
  - New methods, new technique, and new models to address interoperability issues among heterogeneous GIS data, methods and systems
  - Semantic modeling, categorization and analysis of geographical phenomena
  - Intelligent retrieval, integration, and processing of distributed computing resources
  - Semantic interoperability in physical, environmental and social applications

http://www.public.asu.edu/~wenwenl1/tgis.pdf

Thank you and Questions?