

Domain model enhanced search

- A comparison of taxonomy, thesaurus and ontology

Master Thesis
Katharina Schwarz
20.07.2005

Agenda

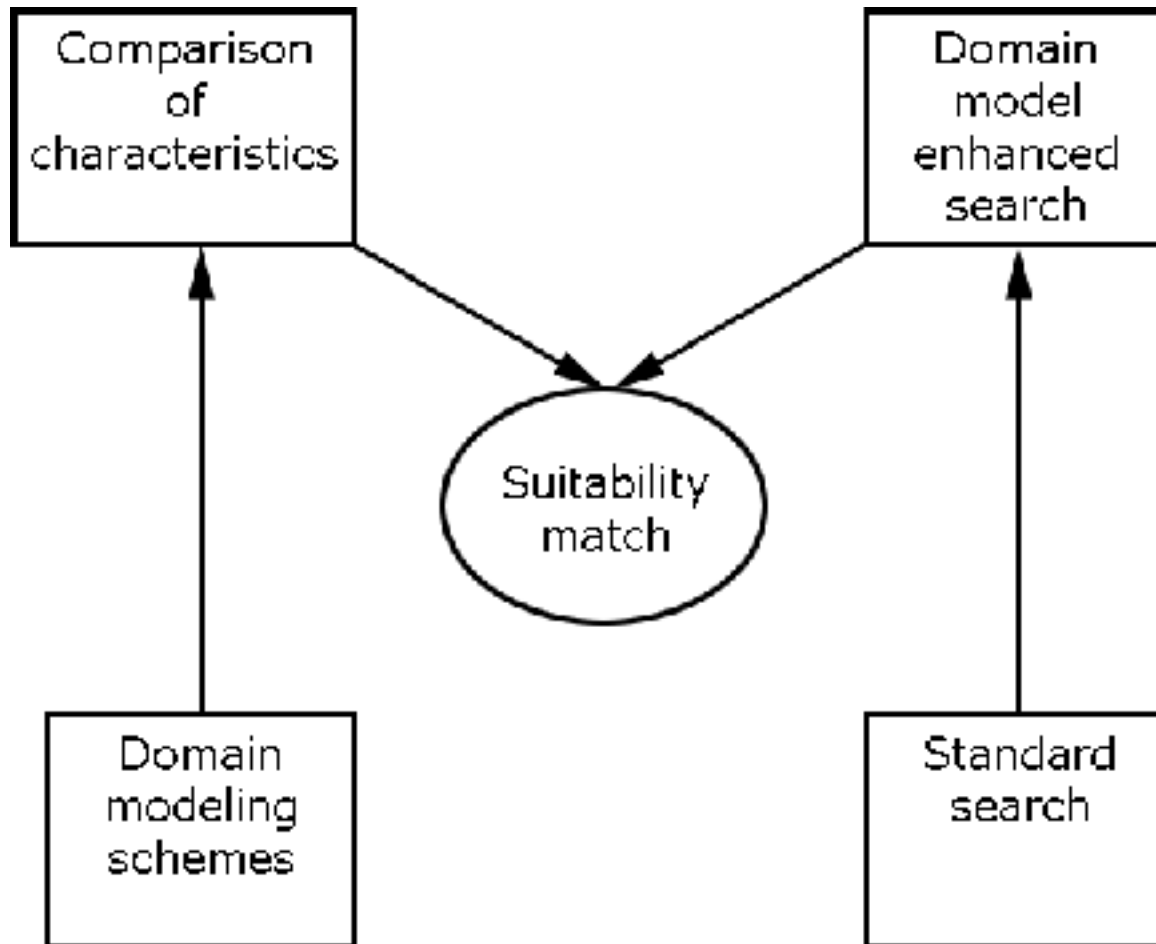
- Objectives
- Domain models
- Domain model enhanced search
- Suitability match

Objectives

How are domain models used to enhance search?

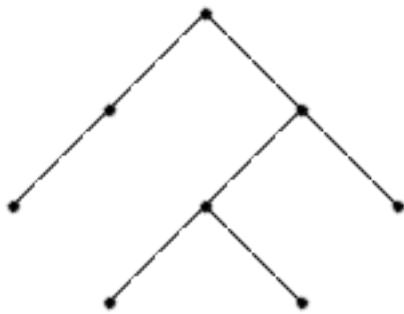
For which problems is each of these domain models especially suited?

Objectives

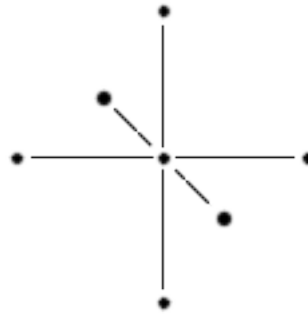


Domain modeling schemes

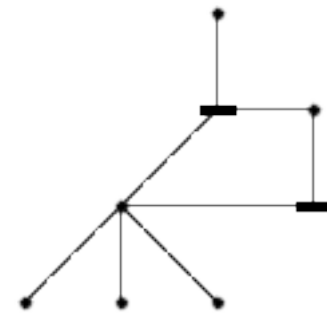
Taxonomy



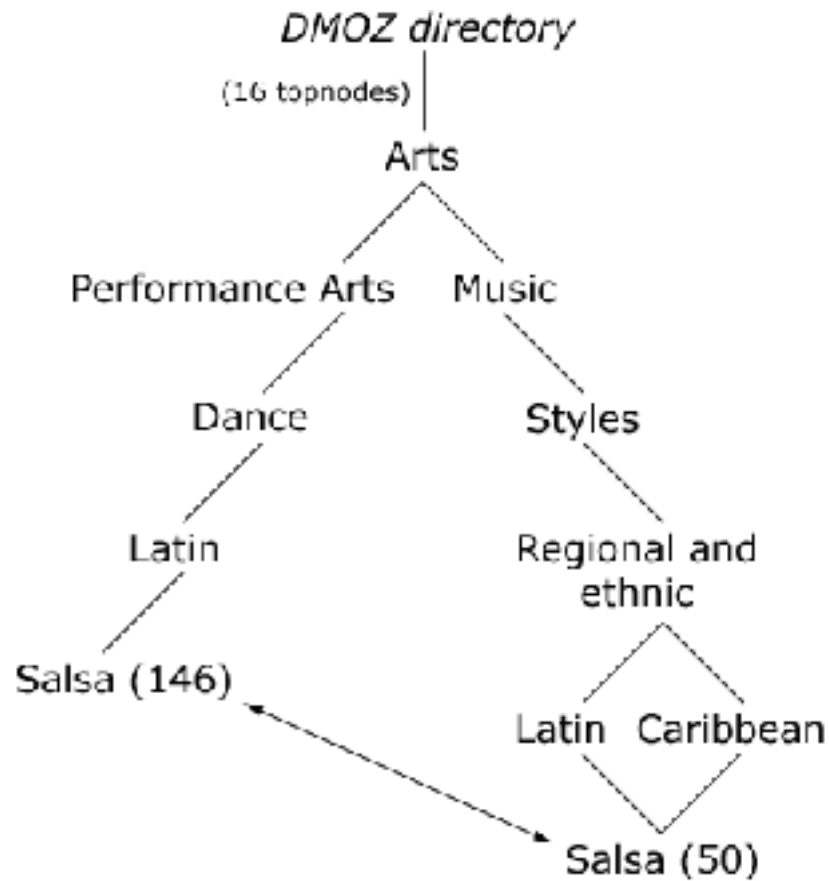
Thesaurus



Ontology



Taxonomy



Thesaurus

The screenshot shows the Visual Thesaurus application interface. The main window displays a network diagram centered on the word "matrix". The diagram consists of a central node "matrix" connected to several other nodes. The nodes and their connections are as follows:

- "matrix" is connected to "array" via a dashed line labeled "is a type of".
- "matrix" is connected to "enclosure" via a solid line.
- "matrix" is connected to "matrix enclosure" via a solid line.
- "matrix" is connected to "ground substance" via a solid line.
- "matrix" is connected to "intercellular substance" via a solid line.
- "matrix" is connected to "mould" via a solid line.
- "matrix" is connected to "old" via a solid line.
- "matrix" is connected to "cast" via a solid line.

On the right side of the application, there is a list of related terms categorized by part of speech:

- NOUNS**:
 - an angular array of elements (or entries) set out by rows and columns
 - an enclosure with in which something originates or develops (from the Latin for womb)
- ADJECTIVES**
- VERBS**

The application interface includes a search bar with "matrix" entered, a "LOOK UP" button, and various navigation and utility buttons like "BACK", "FORWARD", "PRINT", "EMAIL", and "HELP".

Thesaurus

SALSA

N SALSA ON TWO
BT MUSIC
BT DANCE
RT LATIN
RT MEXICAN
K CUBA

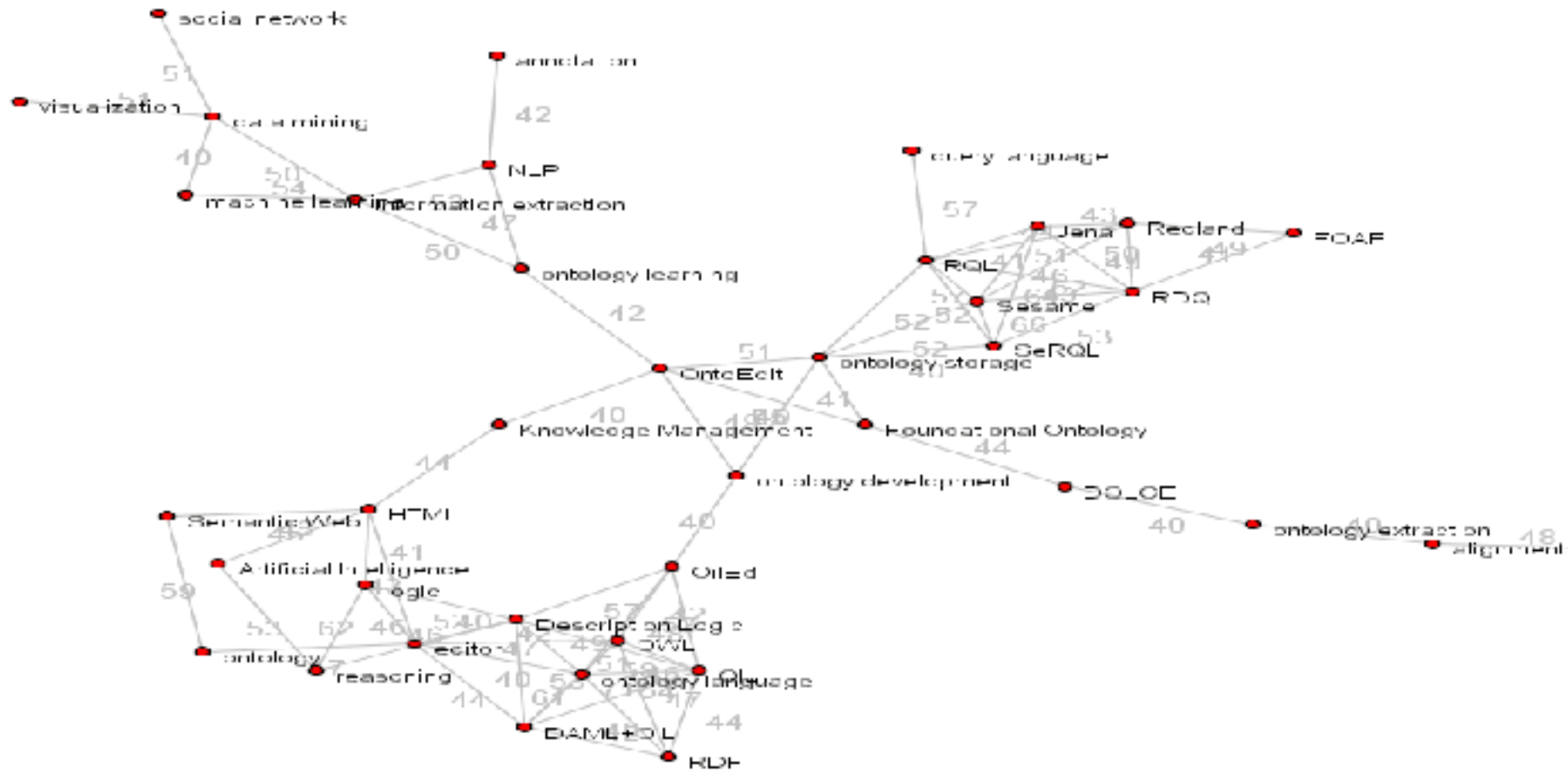
SALSA ON TWO

BT SALSA
UF SALSA NEW YORK STYLE

SALSA NEW YORK STYLE

USE SALSA ON TWO

Ontology



Ontology

```
<owl:Class rdf:ID="salsa">
```

```
<rdfs:comment>the class of the music and dancing  
style salsa.</rdfs:comment>
```

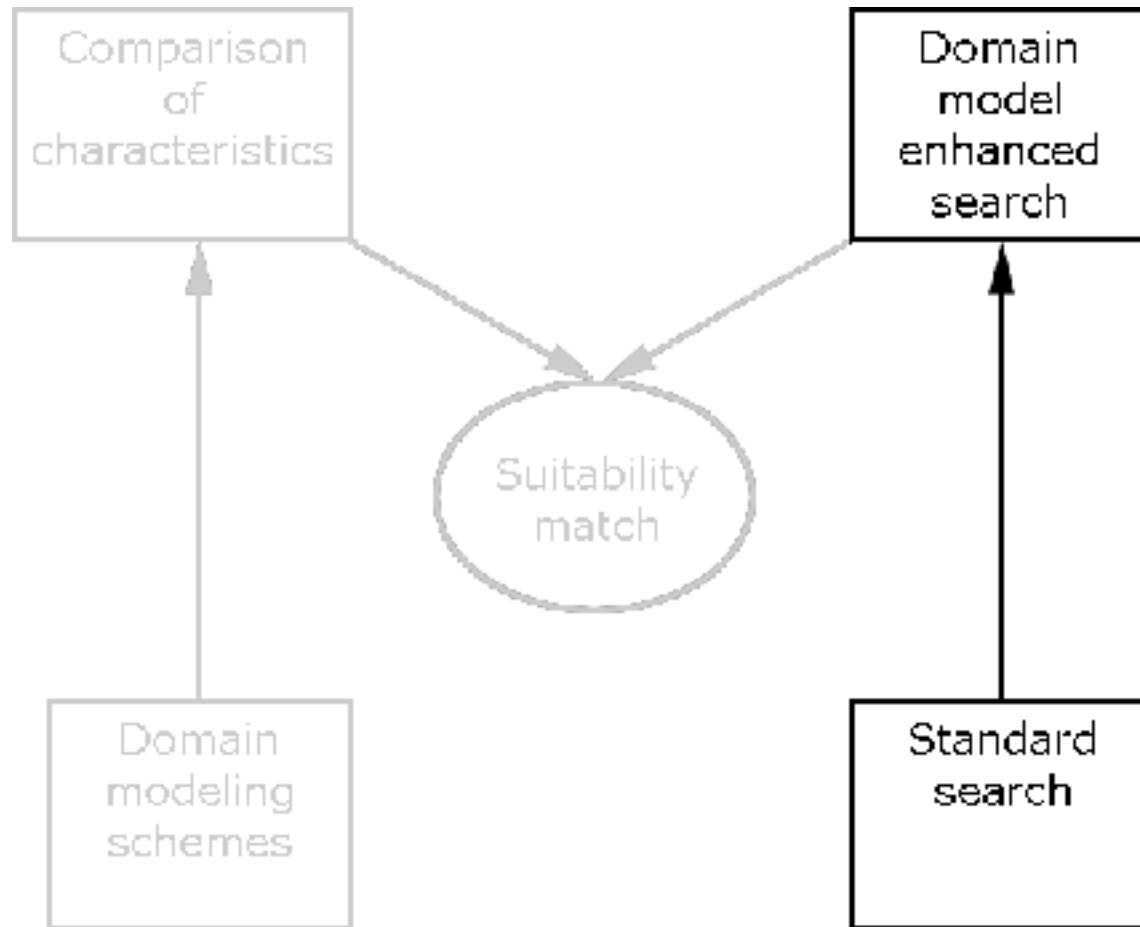
```
<rdfs:label>salsa</rdfs:label>
```

```
<rdfs:subclassOf rdf:resource="#music"/>
```

```
<rdfs:subclassOf rdf:resource="#dancing"/>
```

```
</owl:Class>
```

Search



Problems

- Ambiguity of language, e.g.
 - Homonyms (Salsa = style of dancing, Salsa = dipsauce)
 - Synonyms (Salsa new york style = Salsa on two)
 - Differing conceptual models (Salsa in Europe vs Salsa in Latin America)
- Low recall and precision

Standard search

Compensate lack of knowledge about meaning of words with statistical analysis

Requirements

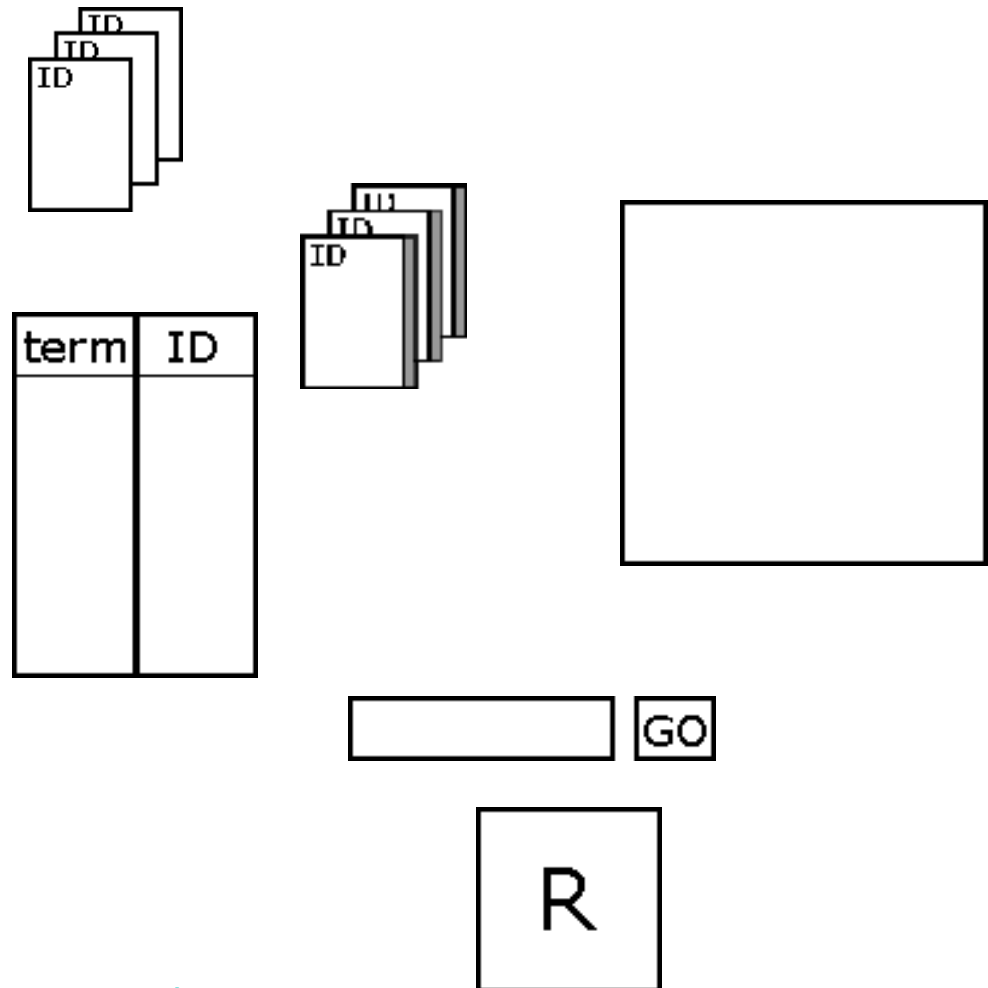
- Requirements of resources
 - computer readable text
 - At least a page of text each
 - Many resources, ca 10.000
 - Heterogeneous language
- Requirements of users
 - Know what they are looking for
 - Know domain well
 - Share same vocabulary as used in resources

Domain model enhanced search

Add context of terms to clarify meaning so that search is more effective

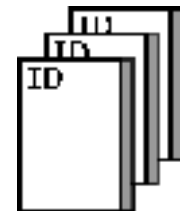
Elements of the search process

- Resources
- Metadata
- Search engine
- Index
- User Interface
- Result set



Resources and metadata

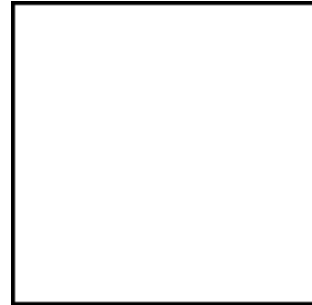
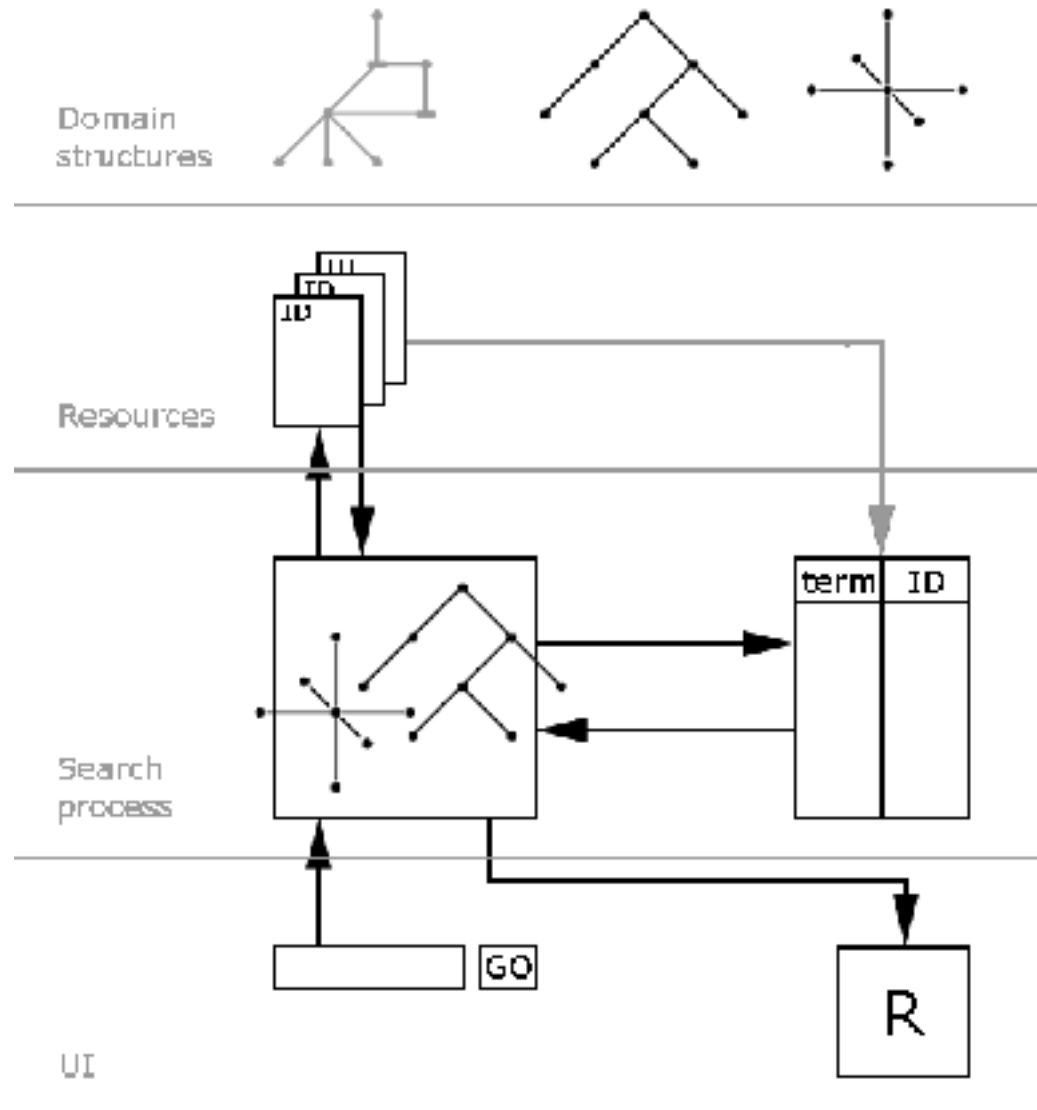
Associate resources with concepts from the domain model
→ explicitly define their topic



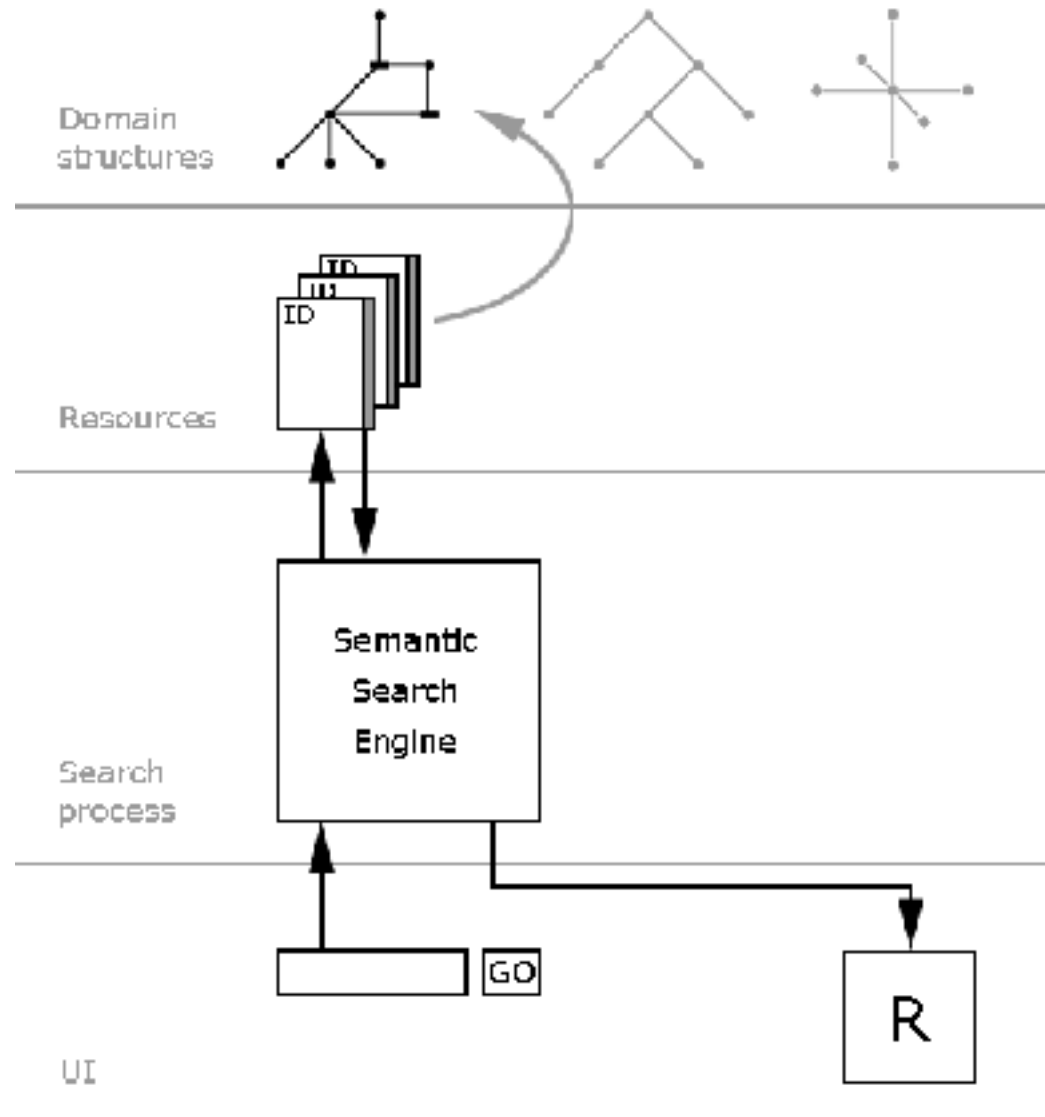
Benefits:

- Describe non-textual resources with text, make them searchable like text-based resources
- Model the same domain from the perspective of different target groups and map them to each other → bridge the semantic gap
- Describe resources in distributed repositories according to the same domain model → search in all with the same query

Search engine

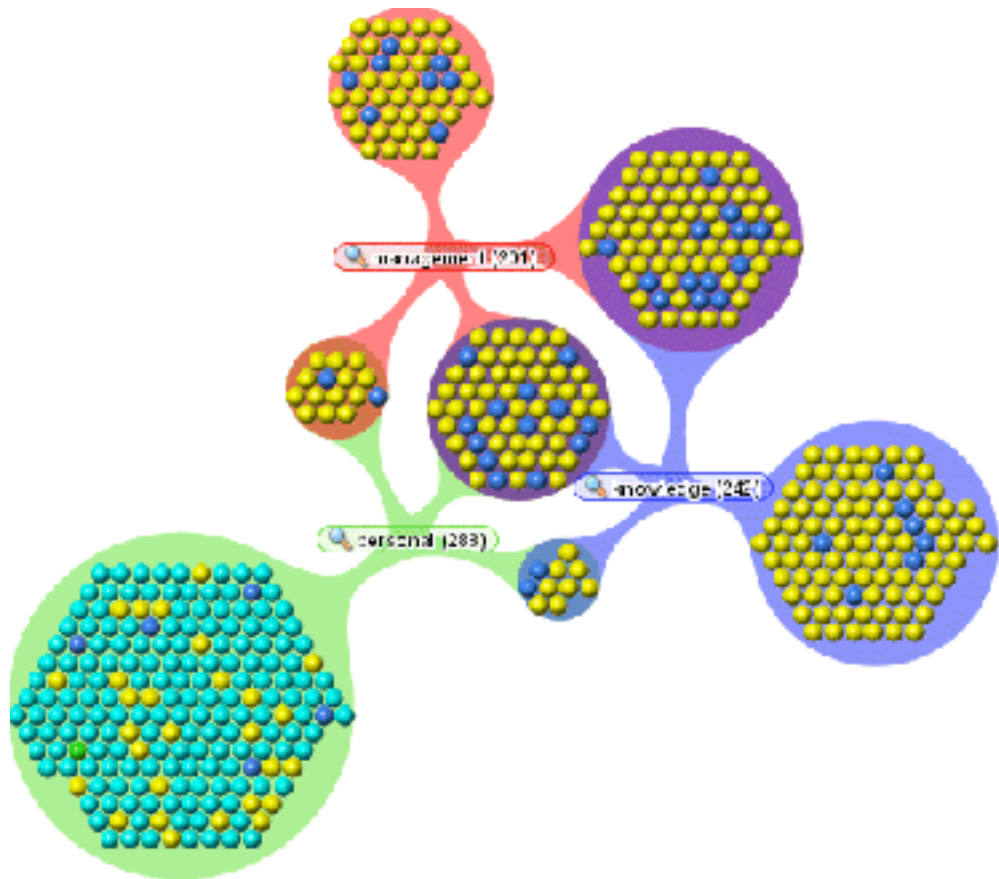


Search engine and index



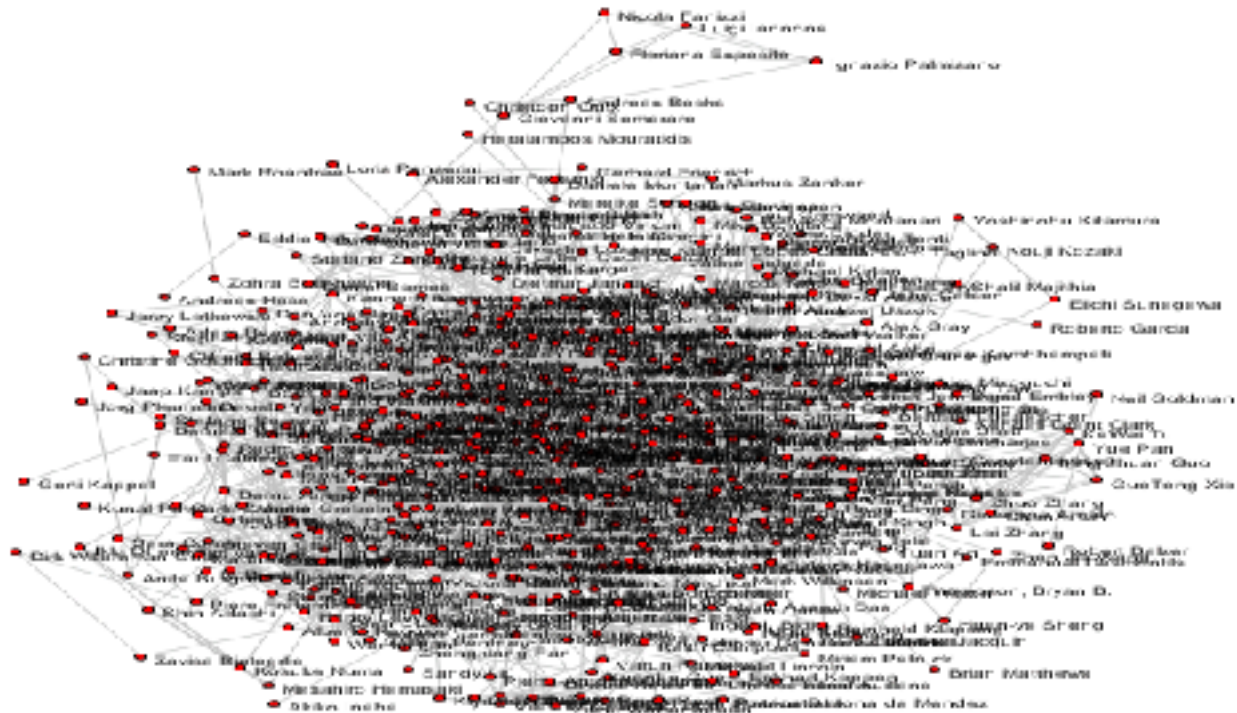
Result set

R




Result set

R



User interface

 open directory project

[about dmoz](#) [suggest URL](#) | [help](#) | [link](#) | [editor login](#)

[advanced](#)

[Arts](#)
Movies, Television, Music ..

[Business](#)
Jobs, Real Estate, Investing ..

[Computers](#)
Internet, Software, Hardware ..

[Games](#)
Video Games, RPGs, Gambling ..

[Health](#)
Fitness, Medicine, Alternative...

[Home](#)
Family, Consumers, Cooking ..

[Kids and Teens](#)
Arts, School Time, Teen Life ..

[News](#)
Media, Newspapers, Weather ..

[Recreation](#)
Travel, Food, Outdoors, Events...

[Reference](#)
Maps, Education, Libraries ..

[Regional](#)
US, Canada, UK, Europe ..

[Science](#)
Biology, Psychology, Physics ..

[Shopping](#)
Autos, Clothing, Kids ..


[Society](#)
People, Religion, Issues ..

[Sports](#)
Baseball, Soccer, Basketball ..

[World](#)
Deutsch, Español, Français, Italiano, Japanese, Nederlands, Polski, Dansk, Svenska..

Help build the largest human-edited directory of the web

Copyright © 1998-2010 Netscape



over 4 million sites - 67,745 editors - over 690,000 categories

User interface

- Little domain knowledge
- Unfamiliar with query language
- Loosely constrained information need

Disadvantages:

- No overview of information space
- Only vertical navigation, or total confusion
- Repeating groups

User interface

Faceted navigation:

Describe different facets (dimensions/ views/
perspectives) of each resource

Users combine any facets they want in the search query

SEARCH all items within current results

Refine your search further within these categories:

Media: [all](#) > [Painting](#) ([group results](#))[0](#) (1)**Date:** [all](#) > [17th century](#) ([group results](#))[1650 - 1659](#) (1)**Location:** [all](#) > [Europe](#) > [Holland](#)**Occupations:** [all](#) > [Worker](#) > [hunter](#) ([group results](#))**Animals and Plants:** [all](#) > [Birds](#) > [crow](#) ([group results](#))**Heaven and Earth:** ([group results](#))**Built Places:** [all](#) > [Part of Building](#) > [entrance](#) ([group results](#))[Arch](#) (1)**Objects:** ([group results](#))**Themes:** [all](#) > [music, writing, and sport](#) > [sport](#) > [archery](#) ([group results](#))**Shapes, Colors, and Scenes:** ([group results](#))**Artists:** [all](#) > [Bol, Ferdinand, 1616 - 1680](#) ([group results](#))These terms define your current search. Click the to remove a term.**Animals and Plants:** [birds](#) > [crow](#) **Artists:** [Bol, Ferdinand, 1616 - 1680](#) **Built Places:** [Part of Building](#) > [entrance](#) **Date:** [17th century](#) **Location:** [Europe](#) > [Holland](#) **Media:** [Painting](#) **Occupations:** [Worker](#) > [hunter](#) **Themes:** [music, writing, and sport](#) > [sport](#) > [archery](#) **1 result** (11 grouped)

The Crowning of
Bol
1663

Drawbacks

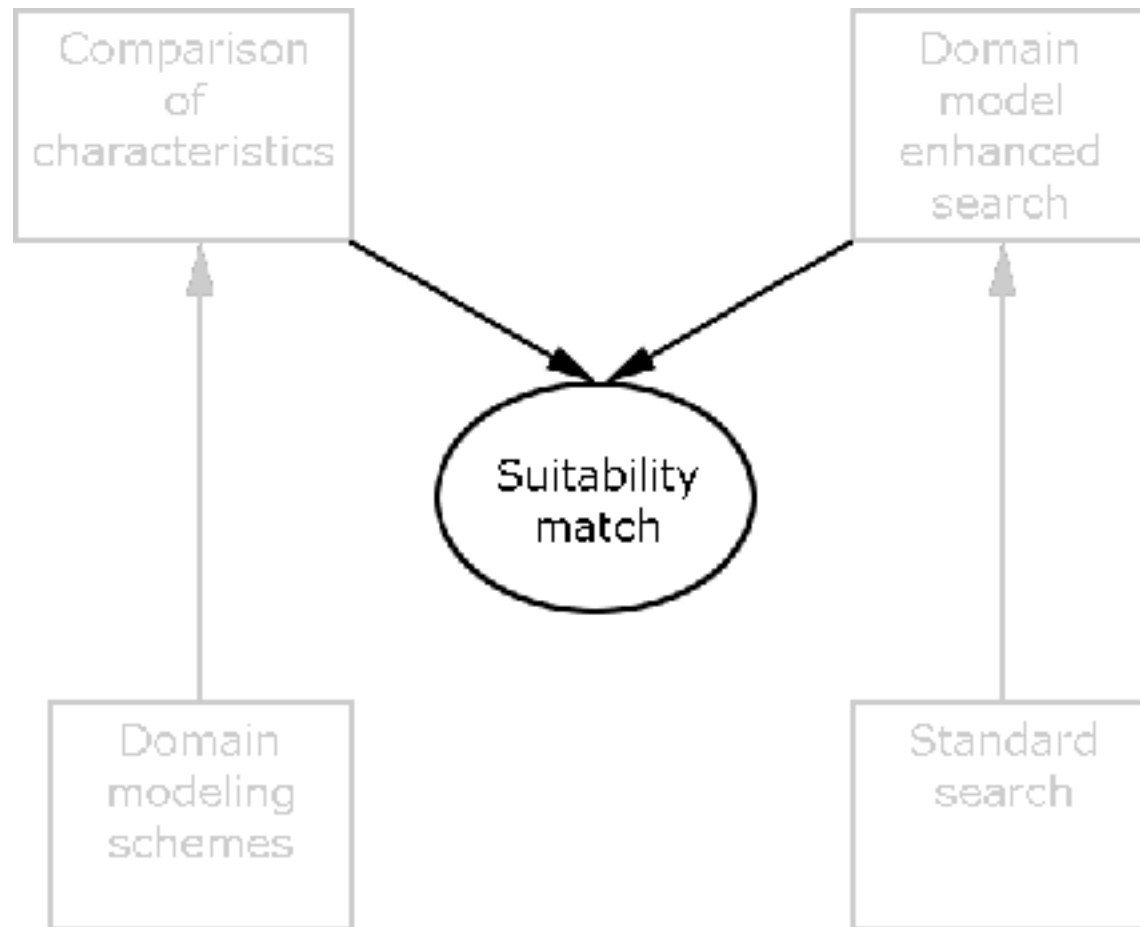
Domain model:

- Modeling
- Maintenance

Metadata:

- Annotation
- Consistency
- Maintenance

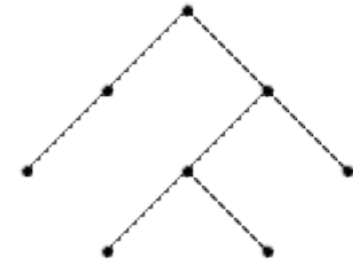
Suitability match



Suitability match

Each domain model has specific areas of application for which it is suited best

Taxonomy - suitable



User interface

Navigation – usability guidelines affect shape of taxonomy

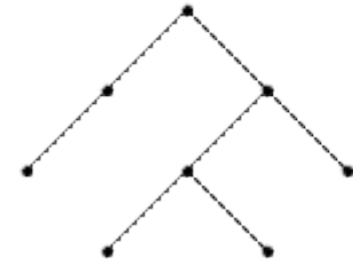
Resources and metadata

Classification

Relate diverging conceptual models from different target groups to each other

→ Several taxonomies, one for each target group

Taxonomy - unsuitable



Modeling

- Precise modeling of a domain
 - Too little expressivity
 - Complex domain
 - Dimensions of domain result in repeating groups
- Use in combination with top-level ontology (facets)

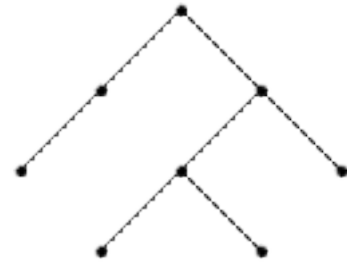
Application

Interoperability

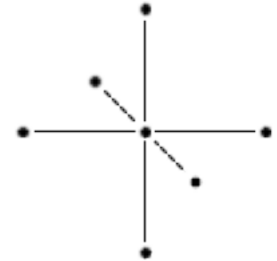
- No standard for modeling or notation

Taxonomy – benefits

- Much tool support
- Comparatively quick modeling result



Thesaurus - suitable



Modeling

Large amount of terms

Resources and metadata

Classification

Search engine

Translation to rules to specify search query (query expansion/contraction)

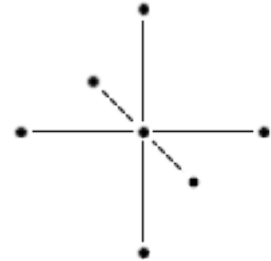
Thesaurus - unsuitable

Result set

Visualization

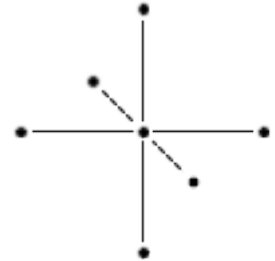
User interface

Visualization



Thesaurus - benefits

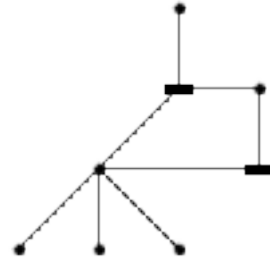
- Old, established craft
- Reuse of existing thesauri
- High interoperability through standardization (semi-automatically translate to rules, RDFS, OWL etc)



Ontology - suitable

Modeling

Precise modeling of a domain, describing domain knowledge



Resources

Both unstructured and structured resources.

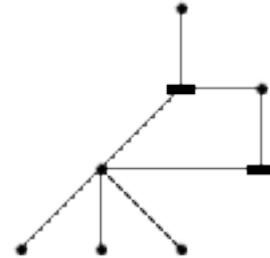
Resources and metadata

Classification of resources required.

Ontology - suitable

Search engine

Inferencing based on machine readable notation of ontology and annotations



Result set

Visualization

User interface

Visualization, faceted navigation

Application

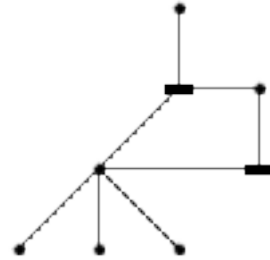
Interoperability, standards

Ontology - unsuitable

Modeling and application

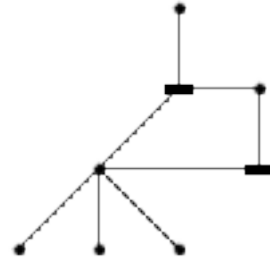
Quick solution

- Every resource has to be annotated and described in RDF
- Modeling requires a lot of effort



Ontology - benefits

- Revealing relationships that weren't modeled explicitly based on inferences
- Resources can't disappear in hyperspace because all are annotated
- Momentum in research community, development of tools and applications



Thank you for your attention

Questions?

Other Ontology Representation Languages

- OWL – Web Ontology Language
 - data type properties,
 - cardinality,
 - more detailed range and domain specifications,
 - types of properties (transitive, symmetric, functional),
 - boolean combinations and enumerations
- Topic Maps
 - Topics
 - Associations
 - Occurrences

Modeling constructs III

Shared understanding of meaning assumed in taxonomy and thesaurus

Ontology makes meaning explicit:

- Specify meaning of concepts with constraints that narrow down the range of possible interpretations
- The more precise, the less ambiguous

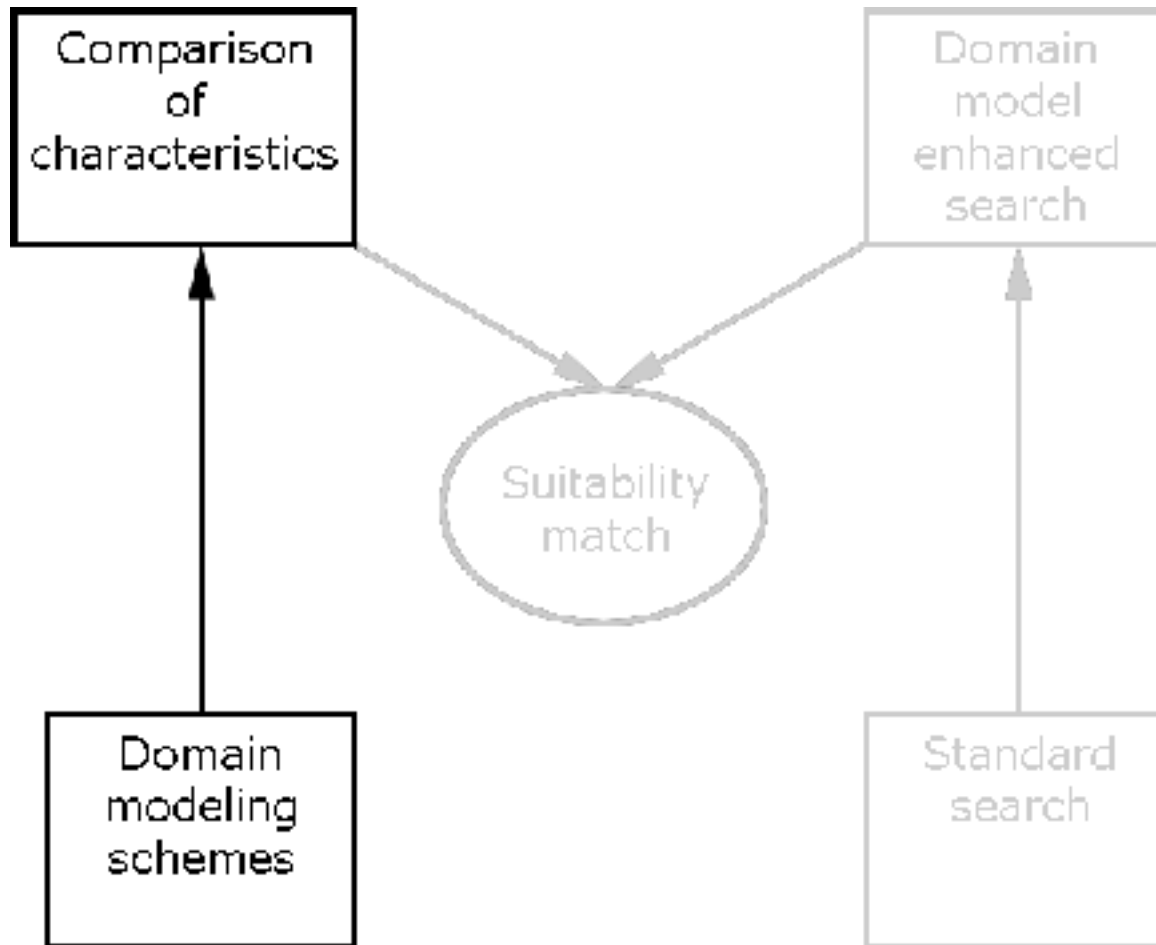
Quality of ontology:

Good → precise, correct constraints

Ok → less specific constraints, more room for misinterpretation, danger of false agreement when mapping

Bad → wrong constraints

Comparison of domain modeling schemes



Comparison properties

- Background
- Standard
- Modeling constructs
- Application

Background

Taxonomy	Thesaurus	Ontology
Natural Sciences, e.g. biology, chemistry	Library Sciences	Metaphysics, Artificial Intelligence, Knowledge Engineering

Standard

	Taxonomy	Thesaurus	Ontology
Modeling	none	ISO 2788, BS 5723, ANSI/NISO Z39/19	Some methodologies, but no official standard
Notational	Graphical hierarchy	Thesaurus symbols (BT, NT, RT, UF, USE)	Focus on Semantic Web technology: W3C recommendations (e.g. RDF Schema, OWL)

Modeling constructs I

	Taxonomy	Thesaurus	Ontology (RDF Schema)
Concepts	both concepts and terms	mainly terms	concepts
Relationships	basically hierarchical, but all types of relationships are modeled using the same notation	untyped hierarchical, associative and equivalence	typed hierarchical and associative, defined as properties

Modeling constructs II

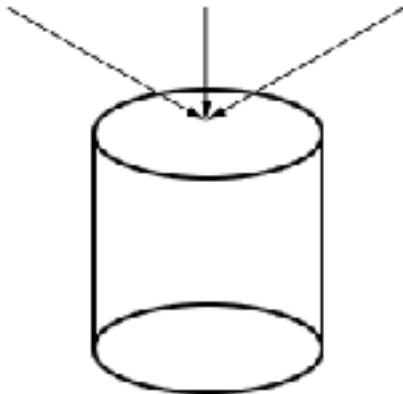
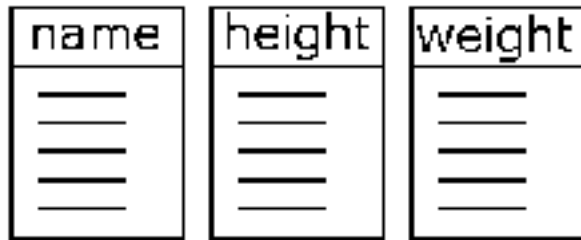
	Taxonomy	Thesaurus	Ontology (RDF Schema)
Properties	none	if required, they can be described in scope notes	Relationship properties and restricting properties
Poly-hierarchy	void if possible	es	es
Coordination	pre	mainly post	pre

Application

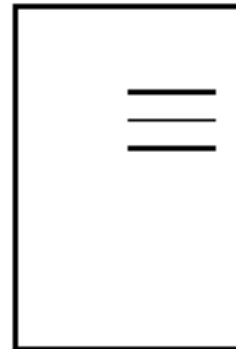
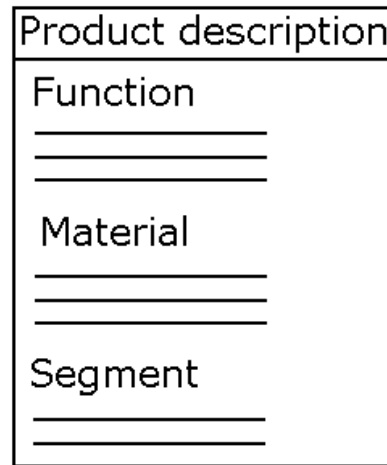
Taxonomy	Thesaurus	Ontology
classification, navigation, search, visualization	classification, navigation, search	classification, navigation, search, visualization, automated reasoning

Resources

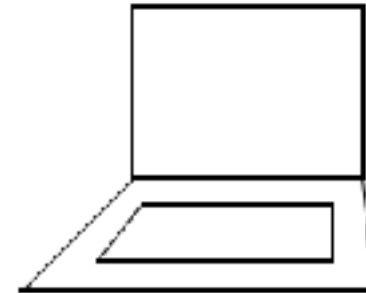
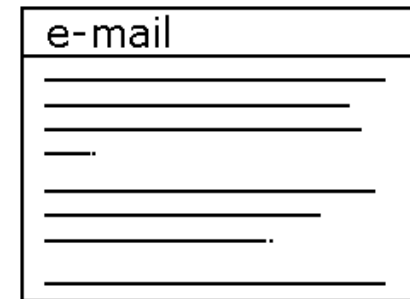
Structured



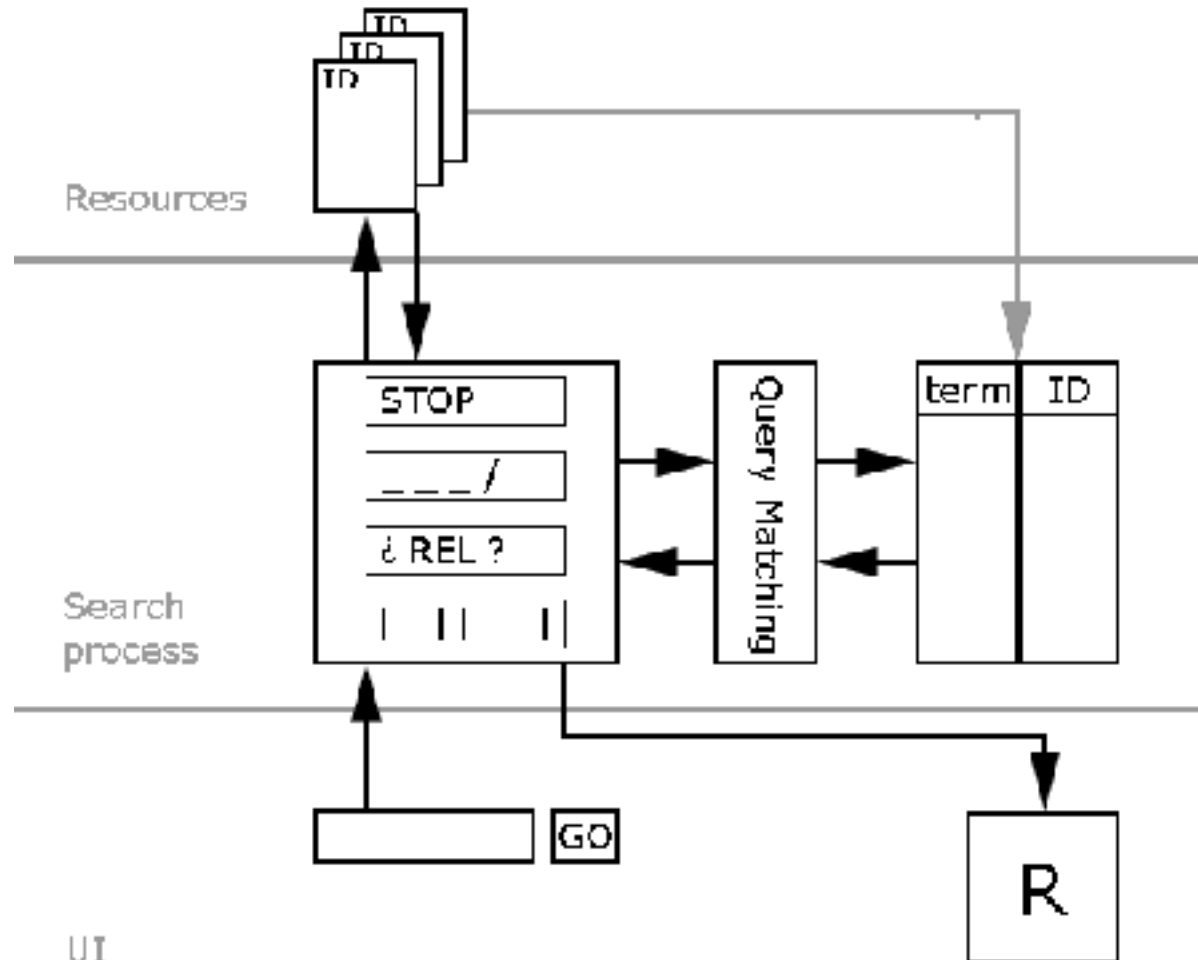
Semi-structured



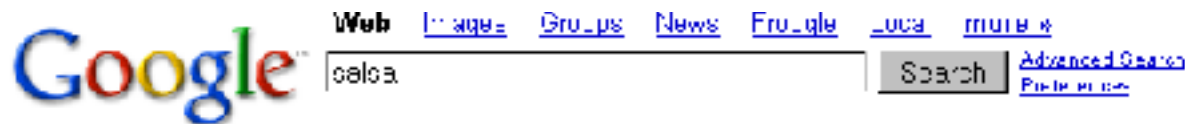
Unstructured



Search engine and index



Result set



Web

[Salsaweb.com - the World's Largest Salsa Magazine, the World's ...](#)

Reviews and interviews with **Salsa's** top musicians

[www.salsaweb.com/ - 78k - 17 Jul 2005 - Cached - Similar pages](#)

[Salsa Cycles](#)

Product index, salsa recipes, catering requests, mission statement and contact information

[www.salsacycle.com/ - 14k - 17 Jul 2005 - Cached - Similar pages](#)

[Just Salsa Magazine ~ Find Salsa Music & Dance, Salsa History ...](#)

Just **Salsa** - A Web Magazine Dedicated to Latin Music, Dance, and Culture. Find **Salsa** Music, **Salsa** Dance, **Salsa** Clubs, **Salsa** History, **Salsa** Photos, **Salsa** ...

[www.justsalsa.com/ - 40k - 17 Jul 2005 - Cached - Similar pages](#)

[Salsa Recipes](#)

474 **salsa** recipes and growing! "It's not hard yet" Hot **Salsa**; Hot 75th Ranger **Salsa** (Quick and Easy); Adobo Herb **Salsa**; Ancho **Salsa**; Apple Black Bean **Salsa** ...

[www.panix.com/~cay/cookbook/salsa.html - 27k - Cached - Similar pages](#)

[Salsa Recipes](#)

Over 100 **Salsa** recipes from mild to very, very hot.

[www.panix.com/~cay/cookbook/salsa/ - 20k - 17 Jul 2005 - Cached - Similar pages](#)

[PLANET SALSA, from the ORIGINS of CLAVE to MILLENNIUM MAMBO, Salsa ...](#)

Result set

The screenshot shows the Clusty search engine interface. At the top, there are navigation tabs for 'Web+', 'News', 'Images', 'Shopping', 'Wikipedia', 'Jobs', and 'Customize!'. The search bar contains the word 'salsa'. Below the search bar, there is a 'Cluster by:' dropdown menu set to 'Topics'. A sidebar on the left lists various clusters for 'salsa' (2/1): Salsa Dancing (52), Recipes (13), Music (41), Latin (29), Salsa Videos (14), Sauces (14), Salsa Congress (15), Merengue (3), Casino, Rueda (7), and Cato Salsa Experience (7). The main content area displays the top 244 results of at least 6,941,009 retrieved for the query 'salsa'. The results are numbered 1 through 4, each with a title, a brief description, and a URL. The titles are: 1. Salsa4Fun Salsa & Rueda, 2. Hot Salsa Dance Fitness, 3. Salsaweb.com The World's Largest Salsa Magazine, and 4. Just Salsa Magazine ~ Find: Salsa Music & Dance, Salsa History, Salsa ...

Clusty

Web+ News Images Shopping Wikipedia Jobs Customize!

salsa

Cluster by: Topics

salsa (2/1)

- Salsa Dancing (52)
- Recipes (13)
- Music (41)
- Latin (29)
- Salsa Videos (14)
- Sauces (14)
- Salsa Congress (15)
- Merengue (3)
- Casino, Rueda (7)
- Cato Salsa Experience (7)

more | all clusters

Tell us what you think.
clusty@clusty.com

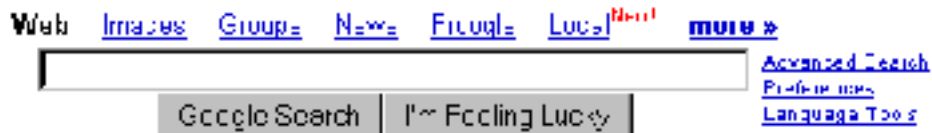
Open in new window

Top 244 results of at least **6,941,009** retrieved for the query [salsa](#) [Details](#)

- [Salsa4Fun Salsa & Rueda](#)
diy. curiousen in Adam & Zabethed, and ill's workshop of pride eedn! www.salsa4fun.com
- [Hot Salsa Dance Fitness](#)
Learn the sexy, hot dance moves for fun, fitness and better health! www.hot-salsa-dance-fitness.com
- [Salsaweb.com The World's Largest Salsa Magazine, the World's Largest ...](#)
Salsa salsa music salsa dance salsa dancing salsa entertainment salsa videos salsa dvd's salsa world salsa event salsa world events salsa ...
www.salsaweb.com - [cache] - WebCite, MIIU, MIIU Search, MIIU, GgalBest, AskJeeves, LookSmart
- [salsacycles2006](#)
Salsa salsacycles hikes bicycle bicycles parts components high-end forks fusp=nsn's=atp=sts lth's Nin=I a= C
www.salsacycles.com Bobby ...
www.salsacycles.com - [cache] - GigzElas, MSN, MSN Search, Open Directory LookSmart
- [PLANET SALSA, From the ORIGINE of CLAVE to MILLENNIUM MAMBO. Salsa ...](#)
... salsa news, mambo, Latin dance and music interviews, listings of salsa music teachers, salsa dance clubs and
University of Salsa ...
www.planetsalsa.com - [cache] - MIIU, MIIU Search, GgalBest, AskJeeves, Vols=hit
- [Just Salsa Magazine ~ Find: Salsa Music & Dance, Salsa History, Salsa ...](#)

User interface

Choose vocabulary
Advanced search options



Know WHAT they are searching for
Know domain well
Know query language