Generalizable Semantic Relations for Knowledge Representation

CTF 07 Conference
Düsseldorf, August 2007

Katrin Weller, Isabella Peters, Wolfgang G. Stock
Institute for Language and Information,
Dep. of Information Science,
Heinrich-Heine-University Düsseldorf
Outline

1. Knowledge Representation in Information Science
2. Classical Relations in Knowledge Representation Methods
3. Syntagmatic Relations in Folksonomies
4. Construction of Relations in Ontologies
Practical Background:
Content Indexing and Knowledge Representation

Knowledge representation from information science’s perspective:

- General aim: Retrieval of (relevant) information and embedding in current work flows.
- Solution: Indexing – Documents are provided with descriptive metadata.
- Methods of content indexing and information retrieval act together.

→ Knowledge representations are constructed for practical aims.
Practical Background:
Content Indexing and Knowledge Representation

Building knowledge representation models
Development of classification schemes, thesauri, rules for abstracting, etc.

Information retrieval
Search with controlled vocabularies, relevance judgments via given abstracts or metadata.

Content Indexing of Documents
E.g. classification, annotation, abstracting.
Methods for knowledge representation

- Classifications
- Thesauri
- Controlled keyword indexing (Schlagwortmethode)

- Methods differ in complexity (e.g. how concepts can be interrelated).
- Some methods have been standardized, for some there are national or international norms (e.g. DIN 32705 for classification systems).
Classification System:
Example – IPC, International Patent Classification

http://www.wipo.int/classifications/fulltext/new_ipc/ipcen.html
Retrieval Without Controlled Vocabulary

Example: Google

Example: Flickr

Examples from www.google.de and www.flickr.com
Retrieval Without Controlled Vocabulary

Basic aim of information retrieval: „find what I mean, not what I say“.

Example I: Searching for photos of Düsseldorf, with search term „Düsseldorf“. Not retrieved are:

Examples from www.flickr.com
Example II: Search for "bank" retrieves:

- German "Bank" = bench
- District Bank in London

Examples from www.flickr.com
Practical Use of Controlled Vocabularies

• The terminology of a domain of knowledge is represented as a structured model of concepts and relations.

• Bundling of synonyms, separation and explanation of homonyms.

• Reduction of varieties. A generally valid and consistent way of indexing is enabled. Unification of users’ and indexers’ vocabulary.

• A search query can be refined and expanded with additional query terms (query expansion or query reformulation).

• Search results can be clustered according to contexts.
Relations in Knowledge Representation

- **Paradigmatic Relations**: fixed, rigidly coupled concept relations within controlled vocabularies. Example: “vehicle” and “bicycle” as hierarchy within a classification scheme.

- **Syntagmatic Relations** originate merely in the actual co-occurrence of terms within a certain setting.

- **Generalizable Relations** are paradigmatic relations that can be meaningfully used in all (or in most of all) general or domain-specific knowledge representation and organization models.
Research on Relations

- There are detailed and highly-professional theoretical reflections on relationships from fields such as linguistics, philosophy, information science, knowledge engineering, artificial intelligence studies and related disciplines.

- Also to be considered: properties, attributes-value pairs, slots and fillers...

New Approach:
→ Examination of existing knowledge representation approaches.
→ Application of detected relations to practical tasks.
In contrast to rich theoretical research, only a small number of paradigmatic semantic relations is currently differentiated and practically applied in classical methods of knowledge representation:

**Classical Types of Relations**

- Relation of equivalence
- Hierarchical Relations
  - Hyponymy
  - Meronymy
- Associative relations (all other relations)
Classical Relations Used in Knowledge Representation Systems

Relation of equivalence

- Synonymy (Relation between different names for the same concept)
- Quasi-Synonyms (Relation between concepts with similar meaning)
- Genidentity (Relation between concepts, whose meaning has changed slightly in the course of time (Leningrad - St. Petersburg))

- Using this relations helps to unify a controlled vocabulary and to increase recall in information retrieval.
Hierarchical Relations
- Relations between (upper) concepts and their sub-concepts.
- Basis framework for most concept models – dominate current systems.

Hyponymy
- Logical view. Specification of features; „is a kind of“, „is_a“.
- kind-of-relation, taxonomic relation, taxonomy
- Examples: mammal – cat; vehicle – car.

Meronymy
- Objective view; „is a part of“.
Classical Relations Used in Knowledge Representation Systems

**Associative Relations**

They are unspecified connections of concepts that can have any kind of relation – except synonyms and hierarchical relations.

Proposals for specification:

- contrasts (antonymy)
- cause – effect (causality)
- producer – product (genetic relation)
- material – product
- predecessor – successor (succession)
- sender – recipient (transmission)
- etc.
Classical Relations Used in Knowledge Representation Systems

Complexity in structure

Thesaurus
Classification
Controlled Keywords

Hyponymy, meronymy, equivalents & associations
Hierarchy & equivalents
Equivalents & associations

Extend of captured knowledge domain
Use of Relations: Query-Expansion

Search queries can be expanded by adding concepts related to the initial search terms to improve search results.

Example:
Search for a stud farm in the German region “Rhein-Erft-Kreis”.

```
region
  Rhein-Erft-Kreis

cities
  Bergheim
  Kerpen

districts
  Quadrat-Ichendorf
  Niederaußem
```

![Image of horses racing](https://example.com/horse-race-image.jpg)
Caution! Expansions may include only one path, if the relations are not transitive.

Transitivity

- **Example**: Tottenham \textit{is\_part\_of} London; London \textit{is\_part\_of} UK. And also: Tottenham \textit{is\_part\_of} UK.
- Classical **counter-example**: Nose \textit{is\_part\_of} Professor; Professor \textit{is\_part\_of} University. But not: Nose \textit{is\_part\_of} University.

Possibly solution: specification of meronymy.
Decomposition of Classical Relations

Example: Meronymy

Meronymy (part-whole)

Decomposition of structures
- Geograph. subunit - geograph. unit
- Element - Collection
- Unit - Organization
- Component-Complex

Structure-independent composition
- Piece - Whole
- Phase - Activity
- Segment - Event
- Part - Object
- Portion - Compound
New Methods for Knowledge Representation

Current Trends

• Web 2.0
  Knowledge Representation with **Folksonomies**.
  Users take over the indexing of web content.

• Semantic Web
  Knowledge Representation with **Ontologies**.
  Higher complexity, formal structuring.

Examples from www.flickr.com and http://protege.stanford.edu
New Methods for Knowledge Representation

Complexity in structure

Trend: enhancement

Extend of captured knowledge domain

Hyponymy, meronymy, equivalents & specified associations

Hierarchy & equivalents

Equivalents & associations (no paradigmatic relations)

Ontology

Thesaurus

Classification

Controlled Keywords

Folksonomy
Folksonomies

Content indexing 2.0

- Users have begun to organize web content (e.g. photos, bookmarks).

- Tags are added to documents to describe their content.

- The process is called (social) tagging and resembles uncontrolled keyword indexing.

- The whole collection of tags within one platform is called folksonomy.

- Tags can then be used for searching.
Relations in Folksonomies?

• No controlled vocabulary, no explicit relations.

• Like in plain texts, relations in folksonomies are purely syntagmatic.

But:
• Users’ Terminology is captured and can provide insights to their actual language use for indexing/searching.

• Relations may be „hidden“ in the combined assignment of several tags. → Further research is needed.
Example I

(geographic) Hierarchy

Synonyms Is_A

Tags
- London
- England
- London Eye
- eye
- ferris wheel
- wheel
- observation wheel
- Bloody high up!!!
- sunset
- silhouette
- ABigFave
- BlueRibbonWinner

Photos and tags from www.flickr.com
Example II

Location - (geographic) hierarchy with region

Equivalent: football - soccer

Name of Stadium – changes over time

Related colours

Rival

Photos and tags from www.flickr.com
Example III

- del.icio.us: Bookmarks can be tagged by different users: tag clouds.
Evaluation of Social Tags

• Systems, where documents are tagged by several users, tend to cover synonyms and spelling variants. We can also find kinds of hierarchies.

• Often more than one tag is assigned to each document, so that interrelations may be exploited.

• Differences can be noted between tags referring to pictures and documents/bookmarks.

• Tags also include many personal associations, most of them cannot be used for our purpose (e. g. me, to do, my dog, last year, fantastic, readthis).
Ontologies

- Ontologies as new knowledge representation systems consist of concepts, instances, relations that connect concepts and specify properties of instances.

- Ontologies can be represented with formal ontology languages such as OWL.

- Ontologies shall enable the semantic annotation and integration of information in a Semantic Web.

- Currently there are no rules or guidelines regarding the use of relations in ontology engineering.
Elements of Ontologies

Classes: hierarchical editor, rules for class membership can be added. Instances (individuals) are assigned to classes.

Excerpt from Generations Ontology: http://www.co-ode.org/ontologies/
Elements of Ontologies

- Additional relations can be chosen to model the domain.

- Object and datatype relations.

- Relations may be specified:
  - Domain and range
  - Functional?
  - Symmetric?
  - Transitive?
  - Inverse counterpart
  - Cardinalities
Example

Relations in Ontologies

- Most dominant: hierarchical is_a, instance_of; also part_of.

- But: Self-defined relations should be used to capture the domain appropriately.

- Many relations are domain-specific.

- Question: Are there new relations that apply to most or many domains of interest (and are these used in current ontologies)?
Example: Domain Specific Relations

**DOLCE**: a Descriptive Ontology for Linguistic and Cognitive Engineering.

Section Social Units.
Example II

UMLS – Unified Medical Language System

Set of non-hierarchical relationships, grouped into five major categories:
- physically related to
- spatially related to
- temporally related to
- functionally related to
- conceptually related to

Suggestions for Generalizable Relations

Properties for General Use?

• Location: geographical information (including hierarchies), locations in comparison to other objects (inclusion, adjacency)...

• Perception: color, shape, taste, sound, haptics, surface, texture, smell...

• Measures: Length, weight, height, density, temperature, time (duration), material? ...

• Functions, actions, task, behavior, processes, events, motions: e.g. develops_from, used_for, used_by, produced_by, transformations, motions...
Conclusions

- Current methods of knowledge representation for information organization and retrieval use only limited types of relations.

- Differentiated relations that can be explicitly formalized in ontologies or may be inherently hidden in folksonomies.

- More relations will be collected, clustered and characterized.

- Newly identified general-use relations will than have to be evaluated for their use in practical applications.
References I


References II