In: Zimmermann, Harald H.; Schramm, Volker (Hg.): Knowledge Management und Kommunikationssysteme, Workflow Management, Multimedia, Knowledge Transfer. Proceedings des 6. Internationalen Symposiums für Informationswissenschaft (ISI 1998), Prag, 3. – 7. November 1998. Konstanz: UVK Verlagsgesellschaft mbH, 1998. S. 296 – 305

Long-Term Value Adding in an Open Category Network: An Informal Social Approach Towards Relating Conceptual Order Systems on the Internet

Alexander Sigel

German National Research Center for Information Technology,
Schloss Birlinghoven, D-53754 Sankt Augustin, sigel@gmd.de[1]

Contents

1. Introduction and Problem Statement
2. Suggested Approach and Underlying Assumptions
3. From Examples Towards a Support Tool
4. Limited Implementation and Experiences with a Predecessor
5. Related Work
6. Discussion
7. References

Summary

Discusses a special case of the problem of comparability and compatibility of conceptual order systems which is of practical relevance when Knowledge Management deals with items from internet sources described with metadata from various order systems. Presents examples from a project which developed an integrated working environment for mediators in the domain of business-to-business contact information brokering. Sketches a tool which could assist in searching/browsing within and across category systems and in gradually capturing and improving knowledge about the interrelations between the order systems. Argues that it is not feasible to start tackling the problem by centralised and authoritative mapping or a stronger formalisation. Suggests instead that users should be better supported in their normal workflow such that they could co-operate on improved mappings. Claims that it would be worthwile to extend the very limited implementation to test if so much lack of control would really lead to acceptable quality and convergence. (Author)

[1] Now with InformationsZentrum Sozialwissenschaften, Lennéstraße 30, D-53113 Bonn, sigel@bonn.iz-soz.de, where all correspondence should be sent to.

1. Introduction and Problem Statement

This paper asks: What could reasonably be done to improve the current situation in which a plethora of different and ever changing conceptual order systems of mediocre quality is used for categorisation on the internet and, even worse, in which one cannot expect a consensus about their mapping because of many independent players and multiple, potentially conflicting views?

This problem is of practical relevance e.g. when in a Knowledge Management System items from sources are decentrally described with metadata, or when information brokers want to find companies by industries across several internet sources.

During the development of the bizzyB[2] prototype of an integrated working environment for information mediators in the domain of business-to-business contact information brokering (EC project COBRA[3]), it turned out that the old problem of the compatibility of order systems showed up in a new disguise, namely in the specification of appropriate categories in search profiles for web robots which query several networked information sources. Imagine a broker who has to search across yellow pages, business directories, buyers' guides, company listings by region or industry, etc., in order to find companies within a certain industry. As some of these sources actively compete on their categorisation schemata and deny the usage of standard trade classifications and business nomenclatures, and typically no mapping is available between a reference system and one of these idiosyncratic systems, the broker has to use a different schema for almost each source. Together with language barriers this turns out to be quite cumbersome.

2. Suggested Approach and Underlying Assumptions

In contrast to more formal and prescriptive approaches, it is proposed to start with the provision of an open category network environment in which a group of users can gradually build up added value by relating items from various category systems over time. At the same time, this would be an environment for human group filtering. The integration of this component into the users' normal workflow eases motivation problems, as the added value can be shared and reused in the sense of Knowledge Management.

Given the pressing need to immediately support users to expand category sets in order to improve their search specifications, and given an enormous amount of dynamic and highly incompatible order systems, one should not strive for in-depth conceptual analysis, but rather distribute the problem soon to more people, i.e. actively engage the user base, and analyse later.

---

[2] bizzyB (Business-to-business Information Brokering with Categories) is a registered GMD trademark.

[3] within the ACTS domain. The acronym stands for Common Open Brokerage Architecture. Short information about the project: http://zeus.gmd.de/cobra/

We assume that:

- a useful knowledge structure will evolve because the brokers co-operate in professional working situations on capturing, accumulating and sharing knowledge about categories and relations;
- because of the immediate availability of an evolving knowledge structure, an improvement in metadata descriptions, in profile specifications, or in the improvement process for this knowledge structure itself is expected;
- formal approaches are more appropriate in later stages when a minimal and more stable knowledge base has been set up.

3. From Examples Towards a Support Tool

This section discusses some useful functionalities and illustrates them with selected categories describing *rubber-related products/services* from several category systems, and some of their possible interrelations. Fig. 1 shows a typical set of category systems used to describe company activities or products/services of companies:

a) Internationally Standardised Classifications
ATECO/NACE SIC/NAICS Harmonised System NATO
b) Buyers' Guides / Company Directories
„Wer liefert was?" (Germany) Yahoo! Branchenbuch (Classified Company Directory) (Germany)
c) National Yellow Pages
Pagine Gialle (Italy) Gelbe Seiten (Germany)

Fig. 1: Selected category systems in the domain of company information

Fig. 2 names two examples of machine-readable sources which can be used to provide the system with initial mappings, before users contribute additional mappings:

Official Concordances, intellectually compiled
SIC<->NAICS
Co-occurences of categories, statistically computed
ATECO<->SIC From a national DB describing companies both with ATECO and SIC codes, from about 55.000 categories co-occurences above the threshold of 4 co-occurences were selected

Fig. 2: Selected machine-readable mappings

To give an impression of the difficulty of the mapping task, selected rubber categories from some category systems are shown in Fig. 3a)-h). One can easily

see that the construction of category sets is generally not based on firm principles, and that there are no facets.

a) ATECO/NACE
D::Manufacturing
  DG::Manufacture of chemicals, chemical products and man-made fibres
    24.17::Manufacture of synthetic rubber in primary forms
  DH::Manufacture of rubber and plastic products
    25::Manufacture of rubber and plastic products
    25.1::Manufacture of rubber products
      25.11::Manufacture of rubber tyres and tubes
      25.12::Retreading and rebuilding of rubber tyres
      25.13::Manufacture of other rubber products

b) SIC
3069800::Other Rubber Goods, N.E.C.
3069814::Rubber Coated Fabrics Garment Fabrics
3069841::Rubber Clothing Industrial Rubber Gloves

c) Harmonised System
Section VII, Chapter 40: Rubber and articles thereof
  4001::Natural rubber, balata, gutta-percha, guayule, chicle and
        similar natural gums, in primary forms or in plates, sheets or
        strip:
    4001.10.0000::Natural rubber latex, whether or not prevulcanized
                  Natural rubber in other forms:
    4001.21.0000::Natural rubber in smoked sheets
    4001.22.0000::Technically specified natural rubber (TSNR)
    4001.29.0000::Other (Natural rubber in other forms, nesoi)
    4001.30.0000::Balata, gutta-percha, guayule, chicle and similar natural
                  gums...
  4002::Synthetic rubber & factice from oils or mixed with natural rubber in primary form
        or sheets, strips [...]
  4003::Reclaimed rubber in primary forms, plates, sheets, strips [...]
  4016::Other vulcanized rubber articles (not hard rubber) floor coverings,
        mats, erasers, gaskets, washers, seals, inflatables, containers, handles

d) NATO
2630::Tires, Solid and Cushion.
      Includes Rubber Track Laying Treads.
3620::Rubber and Plastics Working Machinery.
      Includes Plasticators; Presses; Synthetic Rubber Working Machinery;
        Thread Extruding Machinery; Vulcanizing Machinery.
      Excludes Protective Covering Laminating Presses;
        Tire Maintenance and Repair Equipment.

e) Wer liefert was? (134 categories!)
Alterungsschutzmittel für Gummi
Anfahrrampenpuffer aus Gummi
Aufvulkanisierbare Gummi-Etiketten [...]
Kautschukschläuche [...]
Gummi-Dämmatten
Gummi-Farbkonzentrate
Gummi-Fussbodenbelag
Gummi-Kunststoff-Metallassemblieren im Lohn
Gummi-Kunststoff-Verbindungen nach Kundenwunsch
Gummi-Metall-Dichtungen [...]
Treibriemen aus Gummi
Verschleissteile für die Gummi-, Kunststoff- und Ölmühlenindustrie
Wandschutzverkleidungen aus Gummi
Zweikomponentenformteile, Gummi-

f) Yahoo!-Branchenbuch
Gummi- und Asbestindustrie
Gummi- und Kunststoffverarbeitungsmaschinen: Hersteller [...]
Gummischläuche: Hersteller
Gummischuhe: Hersteller
Gummiwaren: Handel
Gummiwaren: Hersteller
Kunststoffwaren und Gummiwaren Industrie
Nahrungs-, Genuß-, Chemie-, Gummi- und Kunst-
Technische Gummiwaren: Hersteller

g) Yellow Pages Italy (PagineGialle, Seat)
83386 Bellows metal rubber fabric and plastic
16432 Business agents and representatives chemicals rubber and plastics
76181 Coatings rubber and plastic industrial
89161 Fabrics plasticized and rubberized
65071 Floor rubber linoleum and plastic [...]
92351 Hose rubber [...]
84706 Rubber moulding
41311 Rubber natural and synthetic
41091 Rubber packing and gaskets manufacturers and wholesale
41256 Rubber processing and vulcanization machinery
41201 Rubber products miscellaneous manufacturers and wholesale
83376 Rubber products miscellaneous retail

h) Yellow Pages Germany (Gelbe Seiten, Telekom)
Gummi u. Gummiwaren
Gummiauskleidungen
Gummiformartikel
Gummihandschuhe
Gummimetallverbindungen
Gummischläuche
Gummiwalzen

Fig. 3a)-h): Selected rubber-related categories from some category systems
Known mappings from intellectually compiled or statistically computed sources could provide users with initial links (cf. Fig. 4), e.g. to other versions of the same category system (here: translations, a)), and to suggestions for categories from other category systems (b), c)):

a) Show translations for ATECO 25.1: (intellectual mapping between multiple versions within one category system)	
ATECO	
25.1::Herstellung von Gummiwaren (DE)	
25.1::Fabbricazione di articoli in gomma (IT)	
25.1::Industrie du caoutchouc (FR)	
25.1::Fabricación de productos de caucho (ES)	
b) Show partial mappings for SIC 3069*: (intellectual n:m mapping between two category systems)	
SIC	NAICS
3069::Fabricated Rubber Products, NEC	-
3069::(partial) Rubberized Fabric	31332::Fabric Coating Mills (partial)
3069::(partial) Rubber Resilient Floor Covering	326192::Resilient Floor Covering Manufacturing (partial)
3069::(partial) Other	326299::All Other Rubber
c) Show heuristic suggestions for categories related to ATECO 25.11: (statistical 1:n mapping between two category systems)	
SIC	
15 times co-occuring with 3069000::articoli in gomma n.a.c.	
5 times co-occuring with 2822000::gomme sintetiche (chimiche) e altri elastomeri sintetici vulcanizzabili	
5 times co-occuring with 3011000::pneumatici e camere d'aria	

Fig. 4: Suggestions for ATECO expansion using the initial mappings

Much more interesting are support functions which allow users to contribute added value over time. Then one broker could e.g. look for suggestions of related categories which were made by one of her colleagues during search profile specification. As an example, Fig. 5 depicts what user U5 may see after searching for user-contributed suggestions to extend the category „Gummi-schläuche" (rubber hoses).

A tool with a uniform interface is needed which supports searching for categories (and their interrelations) as well as convenient browsing in them, regardless of the origin of the categories. Furthermore and most important, the group of co-operating brokers is strategically interested in improving this category network with their own relations and categories. As in different contexts views on the „right“ mapping might differ, and one often cannot judge upon the correctness, the system should allow for conflicting definitions and not exert too much control. In order to be used the system should be rather simple, especially no semantic definition of categories at all is preferred over a logics approach.

User-defined Category	Relation Type	Type of Suggestion	Source of Suggestion	Category System:: Category
Gummi-schläuche	probably-similar-extension	indirect from co-usage within the same search profile of one user	user U1 date D1	Wer liefert was?:: Kautschukschläuche
	probably-similar-extension	indirect from co-usage within the same search profile of one user	user U2 date D2	Yellow Pages Italy:: 92351::Hose Rubber
	similar-to	direct from user input	user U3 date D3	Yahoo!:: Gummischläuche: Hersteller
	similar-to	direct from user input	user U4 date D4	Yellow Pages Germany:: Gummischläuche

Fig. 5: Example of user-contributed suggestions for category mappings between multiple category systems

4. Limited Implementation and Experiences with a Predecessor

bizzyB is a prototype of an integrated information environment to support human brokerage. The system is implemented with Java, Javascript, Perl, an RDBMS, and HTML. A detailed description of the main ideas of the system and where and how it allows value-adding can be found in (Sigel, Rockenberg & Klemke 1998). (Koenemann & Thomas 1998) provide details about how this system is embedded in the activities of GMD’s working group on agent-based information brokering.

In the initial analysis phase of the project, information brokers in the Economic Information Center of the Milan Chambers of Commerce voiced need for a tool to support browsing/searching in a variety of category systems, as they most often searched manually in code books. This requirement was extended to a tool to create and share knowledge about categories and their interrelations. bizzyB’s component „WebCatNet“ realises this in limited form. The final prototype was evaluated by the brokers.

Unfortunately, the first quite restricted evaluation with the limited prototype does not yet allow us to draw firm conclusions whether a system with the expanded functionality as sketched above would work. This makes us even more curious to see if so much lack of control would really lead to acceptable quality and convergence. Therefore we suggest that it would be worthwile to extend the implementation and to test the assumptions on a larger scale.

5. Related Work

How to achieve compatibility of order systems has been heavily discussed in the specialty of Knowledge Organisation (cf. e.g. a research seminar devoted to that topic, ISKO *et al.* 1996). (van der Walt 1998) analysed and compared the quality of order systems on the internet and provided simple principle-based suggestions to improve some, e.g. Yahoo! This could be applied to other category systems, too.

In the beginning of the project it was checked if formal approaches (using knowledge representation and logics) could help. One can e.g. achieve semantic integration with ontologies (Vickery 1997, Gruber 1993, Stanford Ontolingua Server o.J.) which may be defined using propositional or description logics.

One source of early inspiration for this work were discussions with Soergel about his open SemWeb proposal (Soergel 1997), as well as with Soergel and Fischer about GMD's formal TerminologyFramework system, implemented in Smalltalk (Fischer 1998, Fischer, Möhr & Rostek 1996, Fischer 1995, Rostek & Fischer 1988).

The idea of learning an evolving conceptual structure from user queries over time can already be found in the work of (Jüttner & Güntzer 1988). They tried to learn thesaurus relations from user queries to an OPAC, and employed explicit feedback to ensure quality.

6. Discussion

It is not claimed that an interactive environment for co-operatively relating heterogeneous categories was something very new. But new is the situation where a proliferation of many intentionally independent categorisation schemata is used to organise internet data, and that they are so dynamic that standardisation and central efforts inevitably fail. Furthermore, a long-during intellectual effort to formalise the categories' semantics is out of reach until contributions in a social setting will have provided some baseline structure. This paper claims that the combination of the following issues has not been encountered before:

- The application to a large number of business categories from several category systems (which is of rising importance to e-commerce);
- The focus on value-adding across several category systems;
- The highly interactive groupwork environment feasable via network and its effects on knowledge management;
- The focus on order systems on the internet (while compatibility is usually discussed for rather stable universal order systems, cf. e.g. (Dahlberg 1996)), and

- An open social, not a standardising or formal approach.

During the project it became clear that appropriate support for category structures - although not originally planned - was a core requirement, and that information science offers interesting contributions in this area.

With only preliminary work carried out, it remains an open question whether in practice formal approaches or informal collaborative environments are more appropriate to improve mappings between conceptual systems. In the author's view, after a period of value-adding with the latter type of systems, the results could be used within the former approach. Formal approaches were not employed because the categories downloaded from the web sites were not even associated with definitions but are only text strings, and the data volume was rather high. However, support for category search and browsing was immediately in need, and there were users available who could contribute to the improvement of the resulting network. These users could not handle logical definitions but preferred an informal environment.

Advantages of a social approach to the compatibility problem are:

- Less time is needed to achieve a useful structure because no consensus is necessary.
- Subject specialists can immediately bring in the vocabulary they work with and are more directly involved.
- Even conflicting views are allowed.
- Already in the beginning phase of Knowledge Management (in which structures are still sought after) added value can be created.

Disadvantages are:

- As there is almost no formal semantics, one cannot support inferences or e.g. consistency checks.
- There remain doubts about the user motivation, the reward system and the critical mass of „good" relations.
- It is not guaranteed but only expected that the structure will converge or be of sufficient quality.
- There are not yet enough experiences with so much lack of control.

7. References

[Dahlberg 1996]
Dahlberg, Ingetraut: *Library Catalogs in the Internet: Switching for Future Subject Access*, Green, R. (Ed.): Knowledge Organization and Change, Proc. 4th Int. ISKO Conf., Washington, DC, 1996, Frankfurt/M.: INDEKS, 1996, p. 155-164.
[Fischer 1998]
Fischer, Dietrich H.: *From Thesauri towards Ontologies?*, Structures and Relations in Knowledge Organization, Proc. 5th Int. ISKO Conf., Lille, France, 1998, Würzburg: Ergon, 1998, p. 18-30.
[Fischer, Möhr & Rostek 1996]
Fischer, Dietrich H., Möhr, Wiebke & Rostek, Lothar: *A modular, object-oriented and generic approach for building terminology maintenance systems*, Proc. TKE'96, Frankfurt/M.: Indeks, 1996, p. 245-258.

[Fischer 1995]
Fischer, Dietrich H.: *An executable specification language for thesauri and their operational semantics* (Abstract), Learning and Knowledge: Structural aspects, quantitative methods and new applications, Proc. 19th Annual Conf. of the Gesellschaft für Klassifikation, Univ. Basel, 1995.
[Gruber 1993]
Gruber, T.R.: *Toward principles for the design of ontologies used for knowledge sharing*, Guarino, N. & Poli, R. (eds.): Formal Ontology in Conceptual Analysis and Knowledge Representation, Kluwer, 1993 (= Stanford Knowledge Systems Laboratory Report KSL-93-04). Also online: <http://ksl-web.stanford.edu/knowledge-sharing/papers/onto-design.ps>
[ISKO *et al.* 1996]
International Society for Knowledge Organisation / Polish Librarian Association / Society for Professional Information: Compatibility and Integration of Order Systems. Research Seminar Proceedings of the TIP/ISKO Meeting, Warsaw, Poland, 1995, Warszawa: Wydaw. SBP, 1996.
[Jüttner & Güntzer 1988]
Jüttner, Gerald & Güntzer, Ulrich: Methoden der künstlichen Intelligenz für Information Retrieval, München: K.G. Saur, 1988.
[Koenemann & Thomas 1998]
Koenemann, Jürgen & Thomas, Christoph G.: *Agent-Supported Information Brokering*, KI - Künstliche Intelligenz 3/98 (=Themenheft Intelligente Informationsagenten).
[Rostek & Fischer 1988]
Rostek, Lothar & Fischer, Dietrich H.: *Objektorientierte Modellierung eines Thesaurus auf der Basis eines Frame-Systems mit graphischer Benutzerschnittstelle*, Nachrichten für Dokumentation 39, 1988, p. 217-226.
[Sigel, Rockenberg & Klemke 1998]
Sigel, Alexander, Rockenberg, Anja & Klemke, Roland: *Brokering von Firmeninformationen mit bizzyB*, Proc. Symposium Information Brokering, Leipzig, 1998, Wiesbaden: Vieweg & Sohn, 1998, p. 129-137.
[Soergel 1997]
Soergel, Dagobert: *SemWeb: Proposal for an open, multifunctional, multilingual system for integrated access to knowledge about concepts and terminology: Exploration and development of the concept,* expanded version (1997) of a paper published in Advances in Knowledge Organization, Vol. 5, Frankfurt/M.: Indeks, 1996, p. 165-173.
[Stanford Ontolingua Server o.J.]
<http://ontolingua.stanford.edu/frame-editor>
[van der Walt 1998]
van der Walt, Marthinus S.: *The Structure of Classification Schemes Used in Internet Search Engines*, Structures and Relations in Knowledge Organization, Proc. 5th Int. ISKO Conf., Lille, France, 1998, Würzburg: Ergon, 1998, p. 379-387.
[Vickery 1997]
Vickery, B.C.: *Ontologies*, Journal of Information Science, 23(4), 1997, p. 277-286.