The Potential of Using Knowledge Organising Tools in Collaborative System Development

Lisbeth Björklund, LIBLAB, Department of Computer & Information Science, Linköping University, S-581 83 Linköping

Abstract: The potential use of a collaborative tool for ad-hoc construction of a dynamic thesaurus, which can be used as a meta-database to support collaborative system development, involving people from different disciplines who are supposed to create a common artefact, is discussed. An important feature of this meta-database, is the relations and connections to other kinds of software used in the design process by different actors. Dynamic relations between the source material and different versions of the model must be maintained in order to support decision making based on relevant material from different sources, which might be inspected by several team members simultaneously.

1.0 Background

Working in a field where collaboration between people from different disciplines become more and more the normal way to do things, has revealed these ideas about how to support group knowledge building and information sharing, using modified tools from the area of knowledge organisation. This kind of collaborative work needs support for dynamic descriptions of the ongoing process, relations in space and time between parts of the work. In interdisciplinary groups there is also a need for good communication in order to solve problems relating to the use of different professional languages, agreement on terms to be used, negotiations on meanings etc. Structures should be created which defines the scope of the work, relating the underlying data, which is subgroup-dependent, time-dependent and changing, to the common model of the artefact.

A thesaurus has the basic representation structure for this. What is lacking is dynamic relations, time-stamps, ownership stamps etc. and the possibility to change and enhance the structure ad-hoc. The use of a mark-up language together with a modified thesaurus-engine seems like an appropriate tool to handle this kind of data. An architecture for organisation of heterogeneous, multi-media, time and space dependent information, which could be used in this kind of system, has been proposed and is presented in this volume by Andreas Björklind, who is also a member of the LIBLAB group.

2.0 Knowledge organisation related problems in Computer-Supported Co-operative Work (CSCW)

One of the most critical parts of collaborative work is the communication between the members of the group. In computer based communication, we lose a
lot of the social cues, which seems to help us to understand each other and interpret utterances in the intended way. Without proper communication, it is harder for a group to gain what is called mutual knowledge, knowledge that resides in one or more people is transferred to one or more others, and both sides are aware that they have this common knowledge (Krauss, 1990). The nature of problems that might occur is of course dependent upon many different things as the type of collaboration, whether it is all distributed or if the group has the possibility for face-to-face meetings. Other perspectives might be synchronous or asynchronous work, a homogenous group, what is the task etc. Examples of this kind of discussion can be found in (Beck, 1993), (Minör, 1993) and (Monarch, 1993). I will address some of the problems, which might occur in every group, namely awareness, agreement, authority and retrieval, where knowledge organisation can be, if not the solution, then at least part of it.

2.1 Awareness
Crucial for all sorts of collaborative work, is that every member of the group, at any time, is aware of the status of the work. Computer support for this must not only give the actual status, but also point at the underlying reasons for different decisions and choices of action. This feature can also support new group members who enters the group in the middle of the process (Monarch, 1993) It must also be possible to follow the development over time. This means that the support is both time and space dependent, and very dynamic.

2.2 Agreement
Communication between group-members in order to come to an agreement, must also be supported. The agreement might be very pragmatic, i.e. schedule a meeting, or it might concern the meaning of a concept. Especially in interdisciplinary groups, using different professional languages, the same concept can have very different meanings. The creation and negotiation of conceptual networks has been suggested to support shared meaning in groups (Konda, 1992).

2.3 Authority
Although there are research results indicating that traditional, hierarchical social networks are loosened up in computer-supported communication (e.g. Bikson, 1990), and are replaced by new, more equal patterns, the question about who is the originator of something is still important. It is a question of authority and reliability, which raises the need for ownership stamps.

2.4 Retrieval
In addition to the awareness functionality, it must also be possible to retrieve specific information needed, in an efficient way. Since we are talking about evolving material of different types, recorded on different media, there are special problems, which are not resolved by traditional indexing methods. Again, the support must be flexible, and sensitive for relations in time and space.
3.0 Meta-databases

The concept of meta-databases is not new. In library and information science we have had meta-thesauri and intelligent front-ends to help the end-user to select and search databases for a long time. In other areas, e.g. software development, meta-databases like repositories and data dictionaries are being developed as part of CASE-tools (Computer-Aided Software/Systems Design) to keep control of definitions etc (McGaughey, 1993). Most of these are one-user systems, supporting management of established structures. As the databases grow, both the actual size of single databases and the total number of databases, the need for meta-databases, concerned with description of and guidance to the database world, increases, and in the last years the meta-database have had a revival. Modern meta-databases are designed to be more than just “the sum of the parts”. Great efforts are put into developing meta-databases which reveal the context for different databases, in order to improve the use of the individual database (Hsu, 1991). Our main focus is the meta-data base as support for dynamic, time and space dependent, collaborative work.

3.1 Use of Meta-databases in LIS

I will only briefly discuss two different kinds of meta-databases used in the area of library and information science, the Meta-thesaurus and intelligent front-ends.

3.1.1 Meta-thesaurus

A meta-thesaurus can be seen as a translator or a mapping mechanism between one unifying language and several other languages (natural or formalised). A discussion of meta-thesaurii can be found in e.g. (Rada, 1990). The meta-database here consists of the unifying language, one or more grammars for translation or mapping, and a search engine. This is a very rigid construction, and changes in one language is not necessarily prompting an update of the others.

3.1.2 Bibliographical meta-databases

Bibliographical meta-databases contains information about bibliographical databases. This information can be organised in e.g. an expert system together with knowledge received from intermediaries about how to select data-base(s) according to a given question. In this case the meta-database is working against several data-bases. It can also work against only one data-base, and contain information about that specific data-base together with knowledge about query-formulation etc. A good over-view of both these types of expert systems is to be found in (Drenth, 1991).

3.2 Thesaurus as (part of) a Meta-database Supporting CSCW

In this part I will discuss features and shortages of the concept of thesaurus as a meta-database supporting the problems of awareness, agreement, authority and retrieval, as described above.
3.2.1 Awareness

One of the features of a thesaurus is that it functions as a map describing the domain of knowledge for which it is created. This is a good basis for an awareness facilitating tool. A traditional thesaurus can relate different areas of the map to each other by referencing. What it does not support is description of the evolution, the development over time. To be able to use a thesaurus for this, improvements must be made to the maintenance functions. The question of version handling becomes crucial. A single concept has a history, it has been included at some point, for some reason, it might have been transformed or excluded, there might be several definitions, concurrently or sequentially. But also the whole domain, that is more than just the sum of the parts, has got a history. An awareness facilitator must be able to reason about the parts, and construct a total, which is a short-cut, a picture of the development, at a certain point in time. Awareness also requires control over the relations between the thesaurus and the source material in the same way as described above.

3.2.2 Agreement

Inevitably, in a interdisciplinary group, there will be some concept upon which it is difficult to come to an agreement of its meaning. The traditional way to cope with this in a thesaurus is the USE or UF - used for - relation, together with the scope note apparatus. In the traditional thesaurus this is agreed upon from the start. If we want to use the thesaurus as part of a meta-database related to a variety of dynamic databases, there must be features allowing ad-hoc negotiation and up-dating together with preservation of earlier relations. Tools must be at hand for the group to inspect the thesaurus and the relation to the underlying data in order to make the right decisions. The scope note apparatus must also be enhanced, allowing for different scope notes belonging to different individuals or subgroups, and it must also be time-sensitive. It should be noted that in this kind of collaborative work, a total agreement is not always the ultimate goal. What is important, though, is that everyone is aware that there is a dis-agreement, and why, and what practical solution there is.

3.2.3 Authority

In the case where we use a controlled vocabulary like the traditional thesaurus, there is no need to explicitly describe who is responsible for the choice of terms etc. Usually there is some kind of authoritative committee who has created the thesaurus and maintain it. In the situation described in this paper, where anyone can make changes (probably one would prefer some kind of majority decisions) the question of who has suggested and accomplished a change is of great importance. The question of information authority, whom do we trust to know different things, calls for mechanisms to manage an ownership apparatus. Ownership might be on a personal or a subgroup level, and should penetrate the whole meta-database as well as all the databases underlying it. This gives the possibility to trace the reasons for different peoples ways of actions and decisions, and evaluate them.
3.2.4 Retrieval

Efficient retrieval is of course necessary in a system like this. Every user should be able to retrieve all relevant information, from all the databases, at any time. But as we know, this is not an easy problem to solve working with static, uniform information, like in bibliographic databases. Thesauri are built to facilitate retrieval, and modern search engines are quite good at retrieval in this environment.

The databases in a CSCW environment are more complex than the traditional bibliographic database. Since it is a working environment, the concept of document is not very clear. There is notes, sketches, working papers, video recordings etc. that should be retrievable as soon as their owner is ready to share them with the group. This means that it is necessary to have either very simple indexing strategies for fast user indexing, or proper automatic indexing, either which might later be revised without losing the initial context.

3.3 A tentative architecture for management of dynamic thesaurii

Software for thesaurus management is today a matter of choosing, while a couple of years ago, it was a matter of finding. Reviews can be found in e.g. (Milstead, 1990) and (Ritzler, 1990). Discussions about the future of thesauri and new application areas can be found in e.g. (Schmitz-Esser, 1990) and (Schmitz-Esser, 1991). However, software for thesaurus management are developed to handle a traditional thesaurus. The version handling mechanisms are not developed to support an active history in these tools, and they have no support for ad-hoc, dynamic, co-operative updating and enhancement.

The general „Architecture for Organising Dynamic Information About Space and Time“ described by Björklind in this volume, could be used for thesaurus management in the context of this paper. This architecture is developed for a crisis management system based on a GIS. (GIS = Geographic Information System). The idea is to manage rapidly changing data, from different sources in relation to the geographic location and time. The system is based upon the use of a mark-up language, HyTime, developed for description of time-dependent hypermedia „documents“. A system for mark-up of incoming data (a kind of automatic indexing, based on qualitative reasoning) and a search engine for the HyTime marked data and the geographic data together is to be developed. The main purpose of the system is that it should be possible to ask questions about information concerning a certain geographic region at a specific time.

The spatial dimension in the case of management of a dynamic thesaurus, would not be the geographical position, but the position of the current subject in relation to the rest of the thesaurus. Version handling and visualisation of different relations are covered by the general architecture. What is not explicitly treated is the collaborative inspection and revision of structures, which would be an important part of the dynamic thesaurus.
4.0 Computer support for collaborative design

In the discussions on Computer Support for Collaborative Design many authors addresses the problems that arises when people from technology meet and work together with people from sociology and related domains. (e.g. (Randall, 1993), (Social, 1993)). The important role that communication plays in this process has been studied by Sonnenwald (Sonnenwald, 1992). She models the design of information retrieval systems as an iterative, dynamic and collaborative process. One example of a computer support system for collaborative design, which seems to be supporting the communicative process well, is the n-dim system described in (Monarch, 1993). This system is supposed to support all phases of design, and support e.g. the construction of shared meaning. The n-dim is a stand-alone system, inside which all design work is to be performed. One part of the system is a ad-hoc created conceptual semantic network, where the links connecting different concepts are negotiated by the group members. This is consistent with our approach of a dynamic thesaurus.

Another example of a system supporting group work, is the gIBIS system (Conklin, 1988). Here the main purpose is to support discussion in the group, and come to an understanding of different arguments, more than to resolve a problem or come to an agreement. The organisation principle in this system, is more like a traditional thesaurus, where the types of relations are built into the system a priori as link types, but there is also an „emergency exit“, allowing users to create new link types.

For the type of design team, which I have in mind, the interdisciplinary team, I think that the idea of a meta-database as a support system is more appropriate. All the different peoples involved have their own computer based tools, which they use today, and which are designed for them. The function of the meta-database should be to connect these different sources and to facilitate the communication between the sub-groups and support the co-operative part of the work. By using a tool- and platform independent engine like HyTime, it would be possible to tie together information from CASE tools, qualitative analysis software used to manage data from user studies, information based on different media and so on.

One very important part of this meta-database should be a dynamic thesaurus as described above. Given this, the design team can, at critical points, check out the results from the user studies, compare this with the technical requirement specification, and check the repository for re-usable code written for this kind of problem. They can also back-track to see what brought them there and trace earlier discussions, in which maybe a different vocabulary was used.

Sub-groups, working on their own material, can check the other subgroups material and study this in relation to their own work, and get help from the thesaurus to solve ambiguities that might appear due to the use of different professional languages. (e.g. communication is a concept which is interpreted very different by an engineer and a sociologist.)
The system should also support the team in their group knowledge building in a practical way, by letting the members simultaneously inspect, discuss and change or extend existing structures. As mentioned earlier, the ultimate goal is maybe not to achieve total agreement, but to let all the members come to an understanding of each others views and also record these different meanings in the system.

5.0 Discussion

The focus in this paper is the thesaurus as a dynamic support system for well-defined groups, working rather close together with a specified goal in mind. The feature of allowing a group to define their own structures for knowledge organisation, could also be of interest in wider groups. The LIBLAB group has discussed this in the HYPERCATalog project, where one of the basic ideas is to allow users of bibliographic information services to link together items in an associative way, giving other users hints and tips on literature that might be relevant to them. A group of users should also be able to co-operatively create classifications and/or thesaurii.

It would also be of interest to provide this facility in a wider environment, where researchers work together in loosely coupled groups, like on the Internet. To create and maintain a controlled vocabulary for something as dynamic and diverse as the Internet is probably impossible. The question is, whether such a disparate group as the Internet user community could agree upon principles for how to develop a collaborative thesaurus without ending up in an even worse mess than today.

There are of course many both technical and social problems to be dealt with. One question is whether people will take the time needed to really explore each others views of the domain and the concepts used. This could only be tested in a working environment.

References


