# Increasing the Usability of Knowledge Descriptions by Imposing Constraints on the Knowledge Representation

Harald Kjellin with Theresia O. Neve & Hillevi Sundholm
*Stockholm University, Dept. of Computer and Systems Science*
*Forum 100, 114 32, Kista, Sweden, hk@dsv.su.se*

**Abstract**. It is well known that stored knowledge is not always reused to the extent that was planned by the system designers, even when it can be shown that the represented knowledge has a sufficient quality. We propose constraints to be used when representing knowledge, and we also present a study that shows that the constraints increased the accessibility of knowledge when it was used in a knowledge network. This because users had no difficulties with moving between Web-pages written by different authors since the basic structure of the pages conformed to a standard.

**Keywords**: Constraints, Knowledge representation, Knowledge reuse

## 1. Introduction

Much of what is written within the area of Knowledge Reuse concerns the transfer of knowledge from one group of people to another group of people. There are mainly two different approaches to how this transfer is carried out:

1. A communication process is designed to support knowledge transfer in individual contacts between people. For example in face to face contacts where people can teach each other by talking to each other or showing each other how things should be done. Nancy Dixon (2000) describes how this approach can be divided into various types of knowledge transfer.
2. Knowledge is converted into well structured explicit representations that is stored in some electronic media in order to be reused by another unknown user. The knowledge could, for instance, be stored on a web-page in an intranet. The storing and reusing of knowledge is then adjusted to some kind of system architecture. Borghoff and Pareschi (1998) and Tiwana (2000) describe a number of methods for how to implement systems of storing and retrieving various types of knowledge.

In this paper we will only discuss the second approach as presented above, i.e. how knowledge can be stored in some electronic media, and we will only investigate the points four and five of the knowledge acquisition processes listed below, i.e. how the knowledge is structured and represented:

1. Finding available knowledge sources
2. Elicitation of knowledge from these sources
3. Transforming the knowledge into explicit knowledge
4. Structuring the knowledge
5. Representing the knowledge into formalised descriptions
6. Indexing and storing the knowledge
7. Creating routines for updating/maintaining the knowledge, in the knowledge base

The focus is on how the knowledge can be processed to make it suited for being stored on web-pages, in an intranet knowledge network, in order to facilitate the reuse of the stored knowledge. We will discuss how the knowledge structure affects the usability, and how the proposed constraints on this structure makes reuse easier.

The constraints are related to the structure of the knowledge and not to the quality of it. It is always necessary to have a high quality of the knowledge to be reused, but even when it has a high quality, there is still a risk that the users (1) cannot find the knowledge from their own intuitive search requests, (2) do not understand the terminology that is used, (3) cannot understand the perspective of the author, and (4) cannot browse efficiently due to an abundance of unstructured text.

Problems with navigating and searching occur when the knowledge is not indexed in a way that corresponds to the search request. Another problem occurs when the knowledge is described in a way that can only be understood by people already familiar with the area of knowledge. We have tested the constraints in a number of tests that show the constraints applied on the knowledge representation facilitate the search and reuse of retrieved knowledge. The constraints can be applied in the construction of a knowledge management system or an eLearning system. The constraints are especially useful when the knowledge acquisition is distributed among a large number of knowledge contributors. In the experiments to be described they were tested in groups of ten to 160 people.

## 2. The proposed constraints

For each of the constraints below we first present theories motivating the use of the constraint when creating a web-based knowledge network and then we continue with descriptions of how the constraint is applied:

**Use domain-independent terms**

Information is usually situation dependent (Jonassen, et al. 1995). A result of this is that most information also includes a description of the background context of the information. The context in itself is there to help the reader interpret the information but in many any cases it may instead prevent an end-user from finding the information. The context may be misleading if the end-user is searching for the information as it is applied in a different context. In such a case the differences between the users context and the described context confuses the user.

This raises the question of the importance of the context when describing information. The hypothesis here is that context-free descriptions have advantages over descriptions with context when reusing knowledge. We assume that it is possible to replace the context with generalisations of the concepts. The generalisations are the core structure of the descriptions and in this way they can be used to index the knowledge. The generalisations need to be described both in the stored knowledge and in the search request. Bringsjord (1998) shows that humans often constrain the search space with general categories when they search for information. Computers often do the opposite and search from specific keywords, which may produce less accurate results when searching for complex information that is more than the combinations of specific keywords (Blair & Maron 1985).

We propose that headers and general concepts in web-pages of represented knowledge in the knowledge network may only consist of domain-independent terms. This constraint helps an author to create an information structure that can be reused in related domains. A domain-independent term is a combination of words that may be found in any type of information area. The opposite is a combination of words that only exists in some specific domains. The term “java-programming” is an example of a domain-specific term. The constraint instructs the people who writes the information to use only domain-independent terms on all general levels of the description. This means that headers and concepts in the beginning of lines should not contain any domain-specific terms (see examples further on).

**Use only explicit relationships**

Whenever a written text contains complex structures, there is a risk that the user cannot see how the parts of the structure are linked together and thus he does not get a good overview of the structure. To help the reader, a graphical presentation or various levels of headers may be used to assist the reader in understanding the structure of the information. However, the added structure, on how the parts are related, may not be explicit enough. We may find a line symbolising a relationship between two modules in a graph without any name/label on the line, which would tell the reader what type of relationship it is. In such cases, the author probably assumes that it is evident what kind of relationship it is. This is often an erroneous assumption. A frequent error when describing complex knowledge is that the authors tend to forget that the readers do not have all the background information to understand how the parts are related. The readers must be informed about the type of the relationships and not only that some kind of a relationship exists. A similar error often occurs when a list of statements is presented. The author presents a list of terms and assumes that the reader knows the context well enough to understand what the list itself symbolises, e.g. (1) a number of steps (2) parts of an object or an area (3) subclasses of a super-class or 4) any other kind of assumed structure on the list of terms. Mackworth (1977) has shown the necessity of using explicit relations in descriptions. Without the explicit relations, it is difficult to get the whole picture and it is difficult to navigate in the structure. On the other hand, when we use explicit relationships, we can introduce a structure without having to explain the structure outside the structure itself.

The proposal is a recommendation to use as few general relation names as possible. Information schemas often contain relation names like: {isa = a an object is a subclass of another object, has_part = one objects has other objects as a parts, steps = one activity is carried out as a number of sub-activities, by_ascending_valueX = the parts are sorted in ascending order according to some value X which could be an economic value or the size or weight of something}. In the experiments described further on, we have found out that it is possible to create useful information descriptions by using a small amount of relation names. It is possible to limit the amount to less than ten relation names. In one experiment, a group of students used 35 relation names were used, but we claim that it usually is possible to use less than ten. The less relation names we use, the easier it is to study complex structures, since the reader get to know the relation names so well that the structure is automatically interpreted.

**Simple names or maximum two sentences of text**

Experiments with humans “working memory” have shown that most people can remember at most between five and nine independent variables. We propose one constraint that guides the user to use as short, domain specific explanations as possible. This constraint is assumed to make it easier for the user to make a rough estimation whether the specific information contains what s/he is looking for. The constraint is an instruction to the user to either write only the domain specific name itself or else limit the text to two sentences.

**The hierarchy of web-pages must be flat**

When there are a large number of knowledge contributors that provide web-pages to be used in a knowledge network on the web, problems may occur when linking all contributions. In one of the experiment to be described further on there were, for instance, 160 subjects that contributed web-pages. A problem is that some people may want to contribute web-pages containing many hierarchical levels while others do not. This may create an asymmetry in the knowledge network. In the first experiments, it was discovered that the following constraints solved these problems:

1. The entire knowledge network may not contain more than three levels of html-pages.
2. Each knowledge contributor may only provide pages with links to a maximum of one sub-level.

**A standard syntax**

A simplified standard syntax for all pages of stored knowledge makes it much easier for the user to navigate in a hierarchical structure. Shneiderman (1998) points out that if web-pages do not adjust to a simple standard layout then the complexity in the layout will prevent the user from finding information when searching for it. He also showed that if the syntax of the layout varies between the different pages of a site this will prevent the user from finding information. Another advantage with a standard syntax is that it facilitates machine readings of web documents for agents, crawlers or programs for automatic indexing. The syntax may be used in various areas but since it was designed for reusing knowledge on the web it has only been tested in web-applications.

Below is an illustration of how the proposed syntax constraint can be utilised to create a hierarchy of relationships between headers and concepts:

1. Use a standard type of headers and levels of headers. Such standards can be found in most well known editors like MS-Word, WordPerfect and similar well known editors.
2. The explicit relations should be in *italics*, and should only be used between the nodes.
3. Tabulations emphasise the hierarchic level of the nodes.
4. All general concepts in the network should have a bold font to show it is a general domain independent term. Headers and the first term in a line are considered as being general concepts.
5. All domain specific terms could only be placed as normal text directly after the nodes.

**Any Header**

- *Relation nameA:*
    - **Node1.** A name
    - **Node2.** Text containing a domain specific description of the Node2.
    - **Node3.**
        - *Relation name B:*
            - **Node4.** A domain specific description of the Node4.
            - **Node5.** A name.
- *Relation name C:*
    - **Node6.**
    - **Node7.**

Any kind of information that does not belong to the formalised knowledge like: pictures, tables or stories in free text could be described outside the dot-notation syntax exemplified above. The *Relation nameA:* is an explicit link between **Any Header** and **Node1, Node2** and **Node3**. *The Relation name B:* is an explicit link between **Node3** and the two nodes: **Node4** and **Node5.**

The following example illustrates how the syntax is used for describing a company:

**A company that deals with automation**

- *Features:*
    - **Name**. The Automation & C.
    - **Size**. 167 employees.
    - **Type.** The kind of organization above described, is a typical matrix-organization. To each division in which the company is organized, specific tasks and rules are given in order to carry out the various projects.
- *Aim:*
    - **Company goals**. Automation & C. wants to develop new software and use the already existent one in order to reduce the time production in the manufacturing process as much as possible.
    - **Advantages**. The advantages that the use of Automation &C. provides are evident: it can lead to a significant improvement in the production costs and therefore become more competitive.
- *Departments:*
    - **Service & Marketing department**.
        - *Tasks:*
            - **Customer relations.** Keep and handle the relationships with already existing customers.
            - **New customers.**
        - *Details*
            - **Contracts.** It stipulates the various contracts.
            - **Payment.** Defining the way of payment.
            - **Penalty clauses.**
            - **Delivery dates.**
            - **Data collecting.** Collect all the data about the customers: their needs, their complaints, and their suggestions.
            - **Client recruitment.** Try to contact prospective clients.
            - **Market surveys**. Make some market surveys in order to forecast future marketing trends and needs.
    - **Development Department.**
        - *Tasks:*
            - **Guidelines.** In this department defines all the general guidelines given from the marketing division to reach two different aims.
                - *Aim:*
                    - **Customer requirements.** Satisfy specific requirement of the customer .
                    - **Innovations.** Carry on a policy of innovation.
    - **Testing Department.**
        - *Tasks:*
            - **Tests.** In this department are made some tests to verify the feasibility of what has been projected.
    - **Administrative Department.**

……continued from previous page:

**The division of the work**

- *Requirements:*
  - **A meticulous activity of the Service and Marketing Department**. It is very important to point out the future market trends; innovation in this kind of business area is fundamental;
  - **A short time to market**. If the company market new and innovative products before its competitors, it is easier for it to get more customers and leading in the business area
  - **A good service post-purchasing**.
- *Rules*
  - **One responsible person**. There is a person responsible for the communication among the various departments.
  - **Task documentation**. At the beginning of every month, a document in which all the tasks and the duties of every department and employee has to be issued.
  - **Periodical meetings**. There are periodical meetings in which the results performed from the various divisions have shown.
  - **Report at end of the day**. At the end of every work every employee has to make a report in which it explain in a clear manner all the theoretical information used to carry out its task, the step followed to reach it, the difficulties that he has met, his suggestions and so on.
  - **Distribute**. All these information has to be collected in a congruent and useful manner in a sort of database that can be used from the all the employees, in according to their needs.
  - **Training meetings**. There are some periodical training meetings in which all the employees of the Development Department are brought up on the new discoveries in this field

## 3. Method

The research has been performed in the following way: (1) Establishment of requirements on the semantic that is used, on the standard syntax, and on the generality of terms in a hierarchy, (2) Conduction of empirical studies acquiring and representing knowledge and tests on the represented knowledge, and (3) Evaluation of the results from the empirical studies.

The criteria for selecting organisations for empirical data was: (1) The employees in the organisation must be available and not occupied in work that prevents them from participating, (2) They must be interested in participating in the experiments in order to be able to carry out the tests without continuous sanctions from the management, (3) It should be realistic to believe that they were carriers of some type of complex knowledge about how work can be carried out that they are willing to share in the experiments and that it is relevant to reuse in a knowledge network.

We found: (1) A company with 98 technical consultants. The employees were willing to participate in a project for testing the constraints, (2) A society, where 14 of the free lancing consultants were willing to test the use of the constraints, (3) Four university courses where the constraints could be tested in assignments, involving altogether 310 students participating in the four courses. In all cases the tests concerned the application of the constraints on knowledge descriptions on a Web-site.

Three types of empirical data was acquired from:

- inquiries about the difficulty of using the syntax.
- tests concerning how difficult it was to find a specific piece of knowledge.
- a test measuring if the user had understood the essence of what they had read on the Web-site.

## 4. Collecting and evaluating test data

Most of the tests were carried out to see if the participants could create knowledge and reuse stored knowledge to solve problems. The tests were combined with inquiries where the subjects answered questions about their opinions about the use of the constraints. The aim was to continuously refine the constraints and test/verify new versions of the syntax. This can be classified as a prototyping approach. We present how the results from the tests were interpreted for each of the following evaluation criterion:

**The syntax should be easy to learn**
It should be easy to introduce and explain the logic of the syntax. The seminars had a short introduction for about 15 minutes, where a typical introduction would contain five minutes for introducing the constraints, five minutes for describing examples and five minutes with instructions for how the knowledge elicitation could proceed.

In the very first experiments there were many complaints about difficulties with understanding what was expected from the participants to do. By and by, the constraints were modified and better instructions and examples were constructed. The two major lessons learned were that the relation names must be extensive and the hierarchy should consist of levels of sub-headers rather than too many levels of nodes below a header. When the last version of the syntax was proposed there were few questions concerning how it should be applied.

**It should be easy to create the knowledge representations**
We compared the time it took for the students to create knowledge descriptions with the time it took them to create any other kind of formal knowledge structure.

We could not detect any difference when the subjects could choose another way of structuring the knowledge than the proposed one. The conclusion from this was that the proposed constraints do not add any considerable difficulty to the work of describing knowledge.

**It should be easy to navigate in the network**
In the last versions of the constraints almost all subjects reported that the standard syntax was easy to learn and helped the students in navigating in the knowledge network. Here we measured the number of mouse-clicks students needed to reach a certain piece of information.

By observing six students when they searched for information, an average of five mouse-clicks seemed to be enough to find answers to three questions when the student searched among 55 descriptions of knowledge. The mouse-clicks were counted while the students were searching. This included an estimation of the use of the scrollbar as one mouse-click.

**People should be able to solve problems by using the knowledge representations**
In some of the tests an explicit use of analogies were used to support subjects in a problem solving process. The task was to describe an abstract theory with help of a concrete analogy or example. The analogy was used to hint at how the problem could be solved.

A majority of subjects claimed that the analogies or examples were more useful when they were represented in exactly the same structure and syntax as the presented solution. Therefore we assumed that the constraints could support the type of reasoning that is often referred to as Case Based Reasoning (Leake & Plaza 1997), which can be applied in the following two steps:

1. Matching a new problem with similar stored problems, where each stored problem is associated with a solution to the problem.
2. Applying the solutions of the stored problems to the new problem.

The mapping of knowledge to a new solution was supported by the structure of the analogy being identical with the structure of the theory.

## 5. Conclusions

The study shows that the standard syntax increased the accessibility of knowledge. This because users had no difficulties with moving between Web-pages written by different authors since the basic structure of the pages conformed to a standard. The domain independent terms allowed users to read knowledge from an area they were unfamiliar with. Finally, the search was easier when the user could impose general constraints on an area before the user had any knowledge of the possible specific constraints that could be made.

### *References*

Borghoff U M and Pareschi R, (1998) eds. *Information Technology for Knowledge Management*, Berlin, Heidelberg, New York: Springer-Verlag.

Bringsjord S. (1998) Domain-Independent Abstract Mediating States Are A Function Of Education, *PSYCOLOQUY* 9(85).

Blair D C and Maron M E, (1985) An evaluation of retrieval effectiveness for a full-text document-retrieval system ACM Press, New York, NY, USA.

Dixon N M, (2000) *Common Knowledge. How Companies Thrive by Sharing What They Know* Boston: Harvard Business School Press.

Jonassen D, et al. (1995) "Constructivism and Computer-Mediated Communication in Distance Education". In, *The American Journal of Distance Education*. Vol. 9, No. 2.

Leake D, and Plaza E, eds. (1997). *Case-Based Reasoning Research and Development: Proceedings of the Second International Conference on Case-Based Reasoning*. Springer Verlag.

Mackworth, A K, (1977) Consistency in networks of relations. *Artifical Intelligence*, 8:99--118.

Shneiderman B, (1998) *Designing the User Interface: Strategies for Effective Human-Computer-Interaction*. Addison-Wesley, third edition.

Tiwana A, (2000 ) *The Knowledge Mangement Toolkit*, ISBN:0130128538, Publisher: Prentice Hall.