Nuopponen, Anita (2010). Transmission of knowledge about concept relations in terminological vocabularies. In: *Terminology and Knowledge Engineering Conference 2010. Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges*, 376-395, Eds. Ú. Bhreathnach & F. de Barra-Cusack. Dublin: Fiontar, Dublin City University. [Note: This is not a copy of the print]

Transmitting knowledge about concept relations in terminological vocabularies

Anita Nuopponen
University of Vaasa

Key words: concept relation, concept system, terminology work, terminological vocabulary, definition

1. Introduction

The paper discusses preliminary findings from an ongoing research, which studies how terminological vocabularies published in different formats can convey and utilize information on relations between concepts and concept systems. The results discussed in this paper are from a pilot study where an existing terminological vocabulary from the field of Internet telephony was taken as research material and its contents were analysed.

Questions that were asked are: What are the means of expressing concept relations in (traditional) terminological vocabularies? What type of concept relations are expressed in the entries? Is there a need for defining new types of concept relations for terminology work? These questions can be answered here only partly and further vocabularies and special fields are under investigation. In this paper the, the findings are restricted to the analysis of Internet Telephony Vocabulary[1]. The second question will be thus narrowed down to what type of concept relations are expressed in ICT related vocabularies. The results may, however give some general information that could be applied in other technical fields, too. During the last twenty years my students have been working on concept analysis projects on various fields and we have frequently come across problems with structuring concept systems in (information) technological fields. This has been one of the reasons for starting the analysis of vocabularies with this field.

The third and fourth questions are connected to another ongoing research project, too. I have been step by step going through a wider classification of concept relation types created and described in Nuopponen 1994 (see Enclosure 1 for the latest version of the classification). Earlier, I have taken a more qualitative and theory-driven approach when developing the

[1] See: http://www.tsk.fi/tiedostot/pdf/internetpuhelusanasto.pdf

concept relation classification. The study described here is an effort to empirically test the classification, find gaps in it and possibly enhance it.

The meaning of concept relations and concept systems for terminology and ontology work has been emphasized lately and various projects are searching for the ideal classification of concept relations. E.g. Maroto and Alcina (2009: 232) are "developing a dictionary of ceramics terminology that allows the user to make queries based on the meaning and not only through the lemma". They are not satisfied with computer tools for terminology management and electronic dictionaries that they feel "largely fail to take into account the systematicity of specialized knowledge, because the conceptual information is reflected in a despersed way in the definitions in natural language". Also for a human analyst the scattered information on concept relations in compact definitions and additional explanations cause trouble as could be seen when analysing definitions in my material. Maroto and Alcina take as their hypothesis that if the concept relations can be made explicit and represented in a more structured way, it will also be possible to formalize the special field knowledge in order to enable "onomasiological queries using knowledge about relationships". (Maroto & Alcina 2009: 233)

The need for developing the theoretical underpinnings of terminology work has been clearly realized in research and development projects such as Maroto's and Alcina's. Instead of vast databases and ontologies the focus in this paper, however, lies on traditional terminological vocabularies. They are restricted to a certain field or its subfield and are compiled by working groups consisting of experts of the field and a terminologist. In addition to or instead of a printed book, these vocabularies are nowadays often made available in one or another electronic form (e.g. html, pdf or database). In terminological projects like this, a lot of the extraction of material is done manually and terminological analysis relies on expert knowledge of the group members.

2. Material and method

The material consists of the Internet telephony vocabulary by the Finnish Terminology Centre TSK. The vocabulary is a traditional printed vocabulary but it is available also in pdf-format online and its entries are included in a term bank (http://www.tsk/tepa/). The vocabulary is targeted to all consumers interested in Internet calls and covers "concepts that the consumer will encounter when buying equipment and services for making Internet calls". In addition, the vocabulary is meant to help the work of translators, journalists and information officers.

Because of the target group, according to the foreword, emphasis on technical details is avoided. In addition to facilitating the communication via Internet calls, the vocabulary aims at harmonizing the concepts and terms of the field. The vocabulary is a product of a terminological project where terminological principles and concept system analysis have been utilized in all the phases. All the terminology projects where TSK is involved, more or less follow the same methods, which are described e.g. in Terminology Guide by Heidi Suonuuti.

The vocabulary contains a four page introduction to terminological vocabularies including e.g. descriptions of the parts of a term entry, definitions of the main types of concept relations. Same types of conventions are followed in all of TSK's vocabularies. The vocabulary could be taken as an example of good practices in terminology work. In the next phase of my project, it will be interesting to take as material vocabularies and glossaries where terminological guidelines have not been followed. I wanted to start, however, with something that has already been found functional, because in this way I can concentrate on my main questions instead of being distracted by ill-fitted or dysfunctional concept systems.

At this stage, the material was restricted with one vocabulary, because all the terminological entries including definitions and explanations were analysed, all types of relations found in them were identified and registered manually. Concept relation information embedded in definitions and explanations, entries (terms, definitions, and explanations) were extracted according to the clues in definitions and explanations. Relations were marked with expressions taken from the text, i.e. knowledge probes (e.g. Weilgaard 2000), e.g. <x enables y>. Knowledge probes were associated to the concept relations in the classification found in the Enclosure 1.

The vocabulary contains 77 terminological entries. Each entry consists of an entry number, primary term and its synonyms, equivalents including synonyms in Swedish and in English, definitions and explanations. Explanations give more information on the entry concept or its term(s). In addition, the vocabulary has seven diagrams and an alphabetical index for each language.

3. Means of expressing concept relations in a terminological vocabulary

There are several ways in which traditional printed terminological vocabularies can express relations between concepts. Seppälä (2006) from TSK for instance, lists "textual definitions

and notes as well as concept diagrams" as means of conveying information on concept relations. Furthermore, order and numbering of terminological entries are sometimes used to tell something about the relations of the concepts in a printed vocabulary. In the digital versions hyperlinks are serving this function, too. In the following, these different means are discussed with examples from the material starting with the most visible ones: order and numbering (3.1) as well as diagrams (3.2). After that relations in definitions and explanations (3.3) are discussed and finally further issues relevant to the digital publishing forms are dealt with (3.4).

3.1 Order and numbering

The use of concept systems or a thematic order for organizing the entries makes traditional printed vocabularies versatile sources of knowledge. I have often used e.g. TSK's set of systematic vocabularies just to find information on the field in question. The information is in a very concise form and concepts that belong together are arranged near each other. When looking for definitions or equivalents, it is not necessary to know the right linguistic form of the term in order to find the concept. This applies also for digital pdf-versions of the printed vocabularies.

Instead of organizing the whole vocabulary according to an overall concept system, the Internet telephony vocabulary has been divided into 7 chapters, which can be found in the table of contents both in the printed vocabulary and its pdf version:

1. General concepts (10 entries)
2. Devices needed for Internet calls (9)
3. Telecommunication network devices (8)
4. Internet call software (10)
5. Information networks (18)
6. Communication protocols (8)
7. Data security (4)

Each chapter has 4 to 18 entries the average being 10 entries per chapter. This already makes it easy to view through the themes when searching an entry without knowing the right term. Inside the chapters, the entries are organized thematically, or in some cases also according to fragments of concept systems. Concepts are organized so that those closest follow each other, e.g. 21 *modem*, 22 *cable modem* and 23 *xDSL modem*. In TSK's vocabularies, consecutive numbering is used throughout the whole vocabulary, in this case from 1 to 77. A more

systematic way would be numbering 1, 1.1, 1.2, 1.2.1 etc., where the levels show relationships between concepts. I have been using this method with the students in small vocabularies[2]. The problem however is how to distinguish different relation types from each other. Some vocabulary compilers organize the concepts inside the thematic chapters in alphabetical order thus breaking up potential concept systems or their fragments.

When the entries are transferred into a data base, this kind of context and medium-specific "embedded information" is lost when the entries are removed from the original context and the user sees only one entry at time isolated from all the other entries that were created in the same project. So is the case with TSK's TEPA term bank, where the terminological entries of the analyzed vocabulary have been uploaded to. Numbering is relevant in paper or pdf versions, but is left out in the term bank. Instead, the name of the vocabulary is given (see figure 1).

3.3 Diagrams

In terminology work and in terminological principles in general, visualization of concept systems has since Eugen Wüster's times been a central element both when analysing concepts and terms, and when presenting the results in terminological products.

Seppälä remarks that in TSK's terminological (printed) vocabularies, for readability reasons there is no single concept system diagram that covers all the concepts in the vocabulary. Instead, she says, there are several smaller ones covering subfields of the vocabulary or parts of them. Some concepts appear in several diagrams because they may have different kind of relations to different concepts. (Seppälä 2008: 15) This is also true in the *Internet telephony vocabulary*, where there are seven concept diagrams covering most of the concept entries that are included in the vocabulary (81%).

1. General concepts
2. Devices needed for Internet calls
3. Telecommunication network devices
4. Internet call software
5. Information networks
6. Communication protocol
7. Addresses

[2] See: http://lipas.uwasa.fi/termino/WasaTerm/sanastot/

The diagrams are almost identical with the chapters, except that the 6th chapter has two diagrams while the last chapter with 4 entries does not have any (*firewall, antivirus software, spam over instant messaging, spam over Internet telephony*). In addition to these 4 concepts 10 more concepts are excluded from the diagrams (*file transfer, user profile, user account, Bluetooth, incoming channel, return channel, data transfer rate, shared bandwidth, delay, port*).

Table 1: Amount of concepts in diagrams

Concepts of the same chapter	63
Concept from other chapters	12
Concept not in the vocabulary	17
Concept nodes marked with "Other"	17
Total nr of nodes	109
Concepts in vocabulary, but not in diagrams	14

Concepts from other chapters have been borrowed 12 times why these concepts appeared in more than one diagram (e.g. *telecommunications network, Internet*). In addition to the entries in the chapter, the diagrams include several vocabulary external concepts (17) to help to tie together the concepts that are selected to be defined in the vocabulary. Almost all the "supplementary concepts" in diagrams are used to form more complete generic concept systems – either mono- or multidimensional ones. These concepts have been left out of the vocabulary probably because they have not been considered belonging to the core concepts of the domain. They are either too general (e.g. *call, computer, telephone, user*) or too specific and somewhat beyond what is relevant to Internet telephony (e.g. *client-server network, narrow band network* [English equivalents AN]). Some of these concepts appear in the definitions as superordinate concepts even though they are not defined in the vocabulary nor given any equivalents in other languages (e.g. *packet switching protocol* [Engl. AN]). This will be discussed also later on. All the diagrams also included altogether 17 open ended classifications were marked with "other", the function of which here is to fill in gaps in generic concept systems and show that there are more co-ordinate concepts, but those are not relevant in this context.

In addition to the terms, the concept nodes in the diagrams include also entry numbers and definitions. In many cases, the definitions are needed in the diagram, because it is otherwise

difficult to understand the connection between the concepts, especially in case of associative relations that can be anything (else than generic or partitive).

The concept system diagrams are included in both online versions of Internet Telephony Vocabulary. In the term bank, each concept that appears in a diagram has a link in its entry to the diagram. However, the diagrams are not clickable, which would add their usability.

3.2 Definitions and explanations

In terminological vocabularies as well as in other glossaries and dictionaries with definitions, much of the concept system information is hidden in the definitions, explanations and notes. Especially this is true if a vocabulary is organized alphabetically and does not even have diagrams to illustrate the connections between the concepts. These kinds of products are useful only if one has a search term already in mind.

In the *Internet telephony vocabulary* each of the 77 terminological entries include a definition and most of them (90%) have explanations. Explanations are used in this vocabulary to give more information on the entry concept or its term(s); see the example below (the definition and explanation are translations by AN):

> 38
> **vertaisverkko**
> sv icke-hierarkiskt nät n; serverlöst nät n; jämlikhetsnät n
> en peer-to-peer network; P2P network
> *information network*, where all computers can function both as servers and as clients
> Skype is an example of *VoIP software* that works in a peer-to-peer network.

The length of the definitions is one sentence on one or two lines while explanations are up to ten lines and may consist of several sentences. A lot of information on concept relations and concept systems is embedded in definitions and explanations. In the example above, a visible marker for concept relations is italics ("relation marker"). The terms for concepts that are defined elsewhere in the vocabulary are marked with italics. In the both electronic versions they are appear as hyperlinks. In the example we can see that a superordinate concept in a generic concept system for *peer-to-peer network* is *information network* and the explanation expresses an associative relation between *peer-to-peer network* and *VoIP software* which might be preliminary interpreted as locative relation according to the classification in Enclosure 1.

Nuopponen, Anita (2010). Transmission of knowledge about concept relations in terminological vocabularies. In: *Terminology and Knowledge Engineering Conference 2010. Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges*, 376-395, Eds. Ú. Bhreathnach & F. de Barra-Cusack. Dublin: Fiontar, Dublin City University. [Note: This is not a copy of the print]

Table 2: Relation markers in italics in the entries

	With relations	With no relations	Total nr (%)	Total nr of relations	Average nr relations (all)	Average nr relations (with relations)
Entries	76 (99,99%)	1 (0,01%	77 (100%)	161	2,09 (N=77)	2,11 (N=76)
Definitions	55 (71,4%)	22 (28,6%)	77 (100%)	76	0,99 (N=77)	1,38 (N=55)
Explanations	45 (72,9%)	19 (27,1%)	70 (100%)	85	1,21 (N=70)	1,89 (N=45)

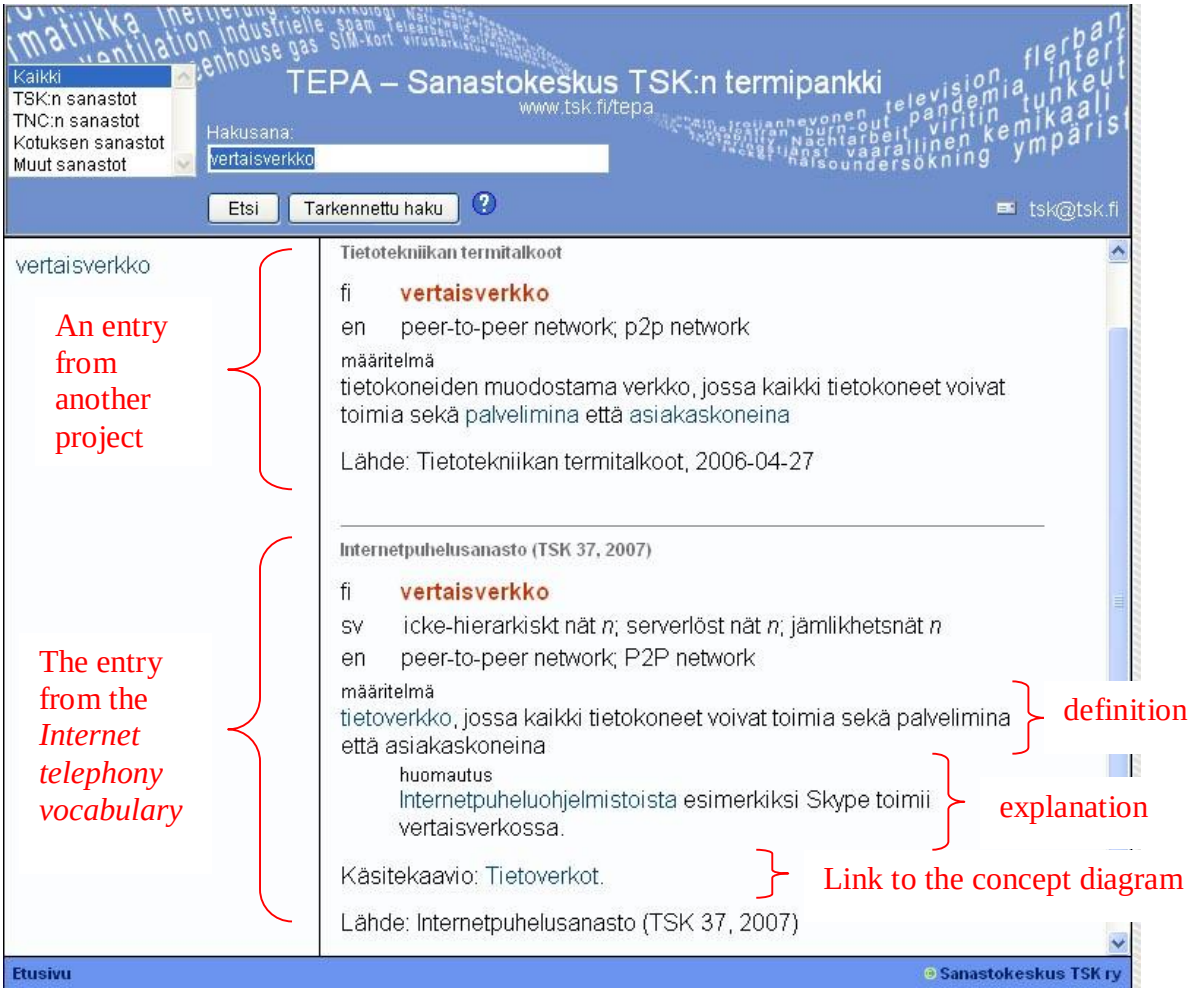
Table 2 accounts for how many relation markers appear in entries as a whole and in definitions and explanations. Only one entry does not have any of marked connections to other entries (*antivirus software*). In average, there are about 2 of them in each entry. About the same proportion of definitions and explanations contain relation markers (71-73%). In average the number of relation markers is higher in explanations than in definitions. This can also be explained by the length of explanations and also by the fact that their function clearly is to convey various relations that the concept has to other concepts. I have counted only those concepts that are marked with italics in the texts, but there is generally much more information in the explanations – in fact, sometimes certain related concepts appear to be explained in them.

3.4 Hyperlinks

There are several possibilities to publish a terminological product in a digital form. An easy and less costly way to publish a vocabulary is to print it as a pdf file and upload it on a website. Pdf is a searchable document but it can be made a bit more user-friendlier by converting the relation markers in italics (see 3.3) as hyperlinks. The same goes for the table of contents as well as for indexes, where all terms including synonyms are listed in an alphabetical order. For a printed systematically organized vocabulary an index is necessary but the search function in the pdf reader makes it partly superfluous if the user knows the term s/he is looking for. The vocabulary under investigation uses hyperlinks through out the text and seems- to be a model case also in this respect. Also in the term bank entries include hyperlinks in the definitions and explanations (see Figure 1).

The concept system diagrams are included in both digital versions. In the term bank, each concept that appears in a diagram has a link in its entry to the diagram. However, the diagrams are not clickable – in neither digital version, which would have added their usability.

Figure 1: Terminology entries for peer-to-peer network in Tepa term bank



In the example *peer-to-peer network* from TEPA term bank (see Figure 1) there is another entry for the same concept from another project. The definitions are almost the same, only the linked relationships to other concepts are different depending on the delimitation of each project. In the printed vocabulary concepts are organized sequentially, but in the term bank this connection is lost.

4. Types of concept relations in diagrams and definitions

The question about what types of concept relations that are expressed in terminological vocabularies is restricted here to concept relations in the field of ICT because of the material. Only three types of concept relations – generic, partitive and associative relation –are utilized in TSK's vocabularies, which applies also to the *Internet Telephony Vocabulary*. In the

beginning of all their vocabularies these three relation types are explained and some examples are given. For each of them, there is a different way to illustrate them in diagrams. In addition, a fourth way of marking relationships appears in the diagrams, but is not explaned (other relations). Generic relations are the largest group in the diagrams (see Table 3), alltogether 60% of all relations in the diagrammes. I have counted also the nodes marked with "other" (generic "other") referring to open-ended classifications, but separated them in their own group in Table 3. Associative relations form the second biggest group of relations (29%) in this vocabulary.

Table 3: Types of relations in diagrams

generic	generic "other"	partitive relations	associative relations	other relations	total
50	18	5	33	8	114
44%	16%	4%	29%	7%	100%

As we saw above, definitions and explanations have an important role in transmitting knowledge on concept relations. However, the type of the relation is not always as clear as in the diagrams even though Seppälä (2008) says "The systematic structure of definitions and the formulation of notes tell whether the relation is generic, partitive or associative". All the definitions in the vocabulary are following the formula of traditional intensional definition, where the superordinate concept and the characteristics that distinguish the concept from other related concepts are given. However, only 21% (16) concepts do have a genus proximum, i.e. their nearest generic superordinate concept, defined in the vocabulary.

The remaining 79% (63) are defined with the help of either their genus proximum that is not included in the vocabulary, e.g. *signalling protocol*, or more frequently with a very general concept on a high abstraction level, e.g. *whole* (6), *device* (5), *call* (4), *function* (3). The small amount of generic relations in the definitions and the use of general concept instead of genus proximum signals that the definitions include other types of relations, after all 71.4% (55) of definitions included some kind of relation to another concept in the vocabulary.

The amount of partitive relations is low both in diagrams (5) and definition and explanations (3). In the definitions and explanations, there are, however, several relations which are have with connection to do and might be classified as partitive relation, e.g. *<x connects g to h>*, *<x combines q>*, *<x is connected to z>*, *<x is connected to g>*, *<z, q, h can be connected to x>*. A more thorough analysis is needed before this classification is clear, because in some

cases also instrumental relation (tool and its use), enhancement relation (something that can be connected to something else without its being a part of the other), or some other relation could be relevant.

Most of the relations expressed in definitions and explanations are associative. Also Seppälä (2008) points out that the text often "illuminates more precisely the nature of the associative relation". Associative relations have been classified in several ways in the literature. A detailed classification of associative relations is presented in Nuopponen 1994, where concept relations are divided into logical (syn. generic) and ontological concept relations. This classification regards both partitive and various types of associative relations as ontological concept relations. A later updated version of the classification is in the Enclosure 1. At least following types of the relations in it appear in the definitions:

Table 4: Different relation types in the material

property relation	< x has property z>
locative relation	<x is in y>
ownership relation	<x belongs to y>
rank relation	<x is different from y>
representational relation	<x signifies y>

The most compressed style in definition writing is not the best source for extracting relation types – not for terminological software or for a human. Many definitions include quite complicated relationships between two or more concepts. One of the most frequent types are instrumental relations (eg. *<x is used for doing y>, <x is used with z for transfer of q>*). Related to instrumental relations is a relation type I sofar call "enabling relations":

- x enables y
- x may enable y
- x enables y through z
- x enables y from z to q
- x enables y to do z with q by using r
- x enables y between z and other in q

In addition, relations extracted from the definitions and explanations confirm my expectations on further need to clarify possible additions and modifications in my previous classifications. I will continue working with these relation types by using more material and present a classification in the near future.

5. Discussion

For this paper I wanted to take a closer look at a systematically compiled terminological vocabulary from information technology in order to find out how concept relations and systems are treated in this kind of terminological products. One of the questions asked here was what the means of expressing concept relations in (traditional) terminological vocabularies are. There are several ways to transmit this information, especially systematic or thematic ordering, numbering of entries and diagrams illustrating concept systems. Also references between entries are effective methods for conveying concept relations. In terminological projects, concept analysis produces much more information than is conveyed by the finished terminological product, as Seppälä states. She points out that lot of core knowledge of the field is collected and thoroughly analysed, and finally condensed in precise definitions. In an ideal case these definitions should build together a knowledge network of the field – not only a network of definitions. (Seppälä 2008: 15)

In a printed or a simple digital version of printed vocabularies, it is easy to add relation markers or links from one entry to another without specifying the type of the relation. The information embedded in definitions and explanations, however, is not always that explicit and the type of the relationships may remain unclear. For a vocabulary user this may not be a problem in the same way it would be for ontology work as described by Seppälä (2008: 15) who states that when merging terminological vocabularies and ontologies a lot of the embedded relation knowledge in terminological entries has to be made explicit and more relation types introduced in terminology work. Despite precise information on concepts in the case of terminological vocabularies, it would be necessary, she points out, to add information on attributes and relation types that have not been adopted in the final terminological product.

As to the *Internet telephony vocabulary* and its online versions, they manage to transmit knowledge on concept relations in multiple ways. However, one problem seemed to be that the relationships are not always retraceable backwards: if the relation markers or linking lead from an entry to another, e.g. from subordinate to superordinate concept, there may not be any link or hint back to the first entry. This is something that can be easily found by following hyperlinks in the digital versions, but it could also be something that should be considered when writing definitions and explanations.

As to the types of concept relations expressed in the entries, it could be said that at least in the material, associative relations are quite abundant in the definitions and explanations. In this type of very restricted vocabulary, generic relations do not play as big role as could be anticipated. After all, the vocabulary does not cover wider classifications of things. Instead, various types of associative relations seem to connect the concepts more tightly and the definitions and explanations form a complex network of knowledge. Is there a need for defining new types of concept relations for terminology work? After going through all the definitions in the material and comparing them with different relation types in my earlier classifications of ontological concept relations, I must admit that many of the relation types in the field of information and communication technology proved to be a challenge. New relation types have to be added in the relation classification. This pilot study convinced me to continue with this type of concepts and concept relations.

There are certain problems when publishing the same products in different formats: what works fine in one medium does not necessarily work in another. I believe that a more precise set of tools for concept analysis would make the definitions and explanations more consistent. If more of the knowledge on concept relations and systems is utilized, terminological vocabularies – in print or digital form – can be made even better and versatile information sources.

Anita Nuopponen, atn@uwasa.fi

References

Internetpuhelusanasto [Internet Telephony Vocabulary] (2007). Helsinki: Sanastokeskus TSK ry. Available online at [http://www.tsk.fi/tiedostot/pdf/internetpuhelusanasto.pdf] (accessed 14 May 2010)

Maroto and Alcina (2009). 'Formal description of conceptual relationships with a view to implementing them in the ontology editor Protégé'. In: *Terminology*, Volume 15, Number 2, 2009 , 232-257.

Nuopponen, Anita (1994). *Begreppssystem för terminologisk analys*. [Concept systems for terminological analysis] Vaasa: University of Vaasa.

Nuopponen, Anita (2005). Concept Relations v2. An update of a concept relation classification. In: *Terminology and Content Development*, 127-138. Edited by Bodil Nistrup Madsen & Hanne Erdman Thomsen. TKE 2005 7th International Conference on Terminology and Knowledge Engineering . Association for Terminology and Knowledge Transfer.

Seppälä, Katri (2006). Ontological terminology work. EAFT & Nordterm workshop: Terminology, Concept Modelling and Ontology - Which approach for which problem?

Nuopponen, Anita (2010). Transmission of knowledge about concept relations in terminological vocabularies. In: *Terminology and Knowledge Engineering Conference 2010. Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges*, 376-395, Eds. Ú. Bhreathnach & F. de Barra-Cusack. Dublin: Fiontar, Dublin City University. [Note: This is not a copy of the print]

Available online at [http://www.eaft-aet.net/fileadmin/files/VAKKI/katri_seppala.ppt] (accessed 14 May 2010)

Seppälä, Katri (2008). Sanastotyötä, mutta missä muodossa? In: Terminfo 4/2008. The English citations originate from an English abstract available at [http://www.tsk.fi/tsk/node/424#sanastotyota] (accessed 14 May 2010)

Suonuuti, Heidi (1996). Terminology Guide. Nordterm 8. Helsinki: The Finnish Centre for Technical Terminology.

Weilgaard Christensen, Lotte (2000). Danske verber som knowledge probes i terminologisk korpusarbejde. In: Nuopponen, Anita & Bertha Toft & Johan Myking (eds.). I terminologins tjänst. Festskrift till Heribert Picht, 233-276. Vaasa: University of Vaasa. Available also at [http://lipas.uwasa.fi/comm/publications/picht/] (accessed 14 May 2010)

Nuopponen, Anita (2010). Transmission of knowledge about concept relations in terminological vocabularies. In: *Terminology and Knowledge Engineering Conference 2010. Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges*, 376-395, Eds. Ú. Bhreathnach & F. de Barra-Cusack. Dublin: Fiontar, Dublin City University. [Note: This is not a copy of the print]

Enclosure 1.