SYMBERED
– a Symbol-Concept Editing Tool

Mats Lundälv, Katarina Mühlenbock, Bengt Farre, Annika Brännström

DART, Regional habilitering, The Queen Silvia Children's Hospital
Kruthusgatan 17, SE-411 04 Göteborg, Sweden
Email: mats.lundalv@vgregion.se

Abstract

The aim of the Nordic SYMBERED project - funded by NUH (the Nordic Development Centre for Rehabilitation Technology) - is to develop a user friendly editing tool that makes use of concept coding to produce web pages with flexible graphical symbol support targeted towards people with Augmentative and Alternative Communication (AAC) needs.

Documents produced with the editing tool will be in XML/XHTML format, well suited for publishing on the Internet. These documents will then contain natural language text, such as Swedish or English. Some, or all, of the words in the text will be marked with a concept code defining its meaning. The coded words/concepts may then easily be represented by alternative kinds of graphical symbols and by additional text representations in alternative languages. Thus, within one web document created by the author with the SYMBERED tool, one symbol language can easily be swapped for another. This means that a Bliss and a PCS symbol user can each have his/her preferred kind of symbol support.

The SYMBERED editing tool will initially support a limited vocabulary in four to five Nordic languages plus English, and three to four symbol systems, with built-in extensibility to cover more languages and symbol systems.

1. Background

Many persons with multiple cognitive and/or motor disabilities use augmentative alternative communication (AAC) to compensate for problems with accessing standard means of spoken or written communication. The AAC-systems may utilise gestures, signs, pictures, graphical symbols, text and sound, or combinations of these, to support communication. They are implemented in manual/low-tech solutions, as well as in high-tech software and hardware systems, including speech synthesis support etc. Standard as well as dedicated assistive technologies are rapidly developing in capability and quality to support these needs, but the lack of co-ordination between different technologies, tools and symbol systems inhibits further development and more wide-spread provision of adequate AAC support and material in the networked society.

This problem has been addressed in the newly completed European IST-project WWAAC - World Wide Augmentative and Alternative Communication (2001-2004), in which a major goal was to develop and implement some suggested standards for syntax and semantic encoding and tools for conversion between symbols and text. The primary achievement of the WWAAC-project was the Concept Coding Framework (CCF) (Concept Coding Framework & Interest Group – 2004-2006). It includes a suggested open technical design of, and APIs for, the creation of a universal set of concept codes, based on standard lexical ontologies. It also specifies methods for linking up to existing and emerging concept and representational resources, such as symbol libraries and databases. It is based on state-of-the-art language technologies developed within the w3c and SemanticWeb activities, such as RDF, OWL, XML and XHTML. WordNet (Miller, G.A. Et al - 2005) is currently used as the main ontological base.

A number of other initiatives, projects and products are striving to promote AAC support in web based information services, such as *Widgit Communicate: WebWide (2005-2006)*, *Papunet* (2001-2006), *Navigabile* (2003-2006) and others. There is a need for, and emerging interest in, co-ordinating the technologies for these efforts based on the above mentioned general web standards. The CCF is a sincere attempt in this direction. However, the suggested standards and APIs need to be further refined and implemented in tools and template components, making them available for practical testing and inclusion in products and services.

2. The SYMBERED Project - Overview

The SYMBERED project will deliver a symbol-concept editing tool to help web content providers in delivering symbol supported web material in a flexible and reasonably efficient way. This is currently implemented in two versions: One in PHP – MySQL, based on the tools used for the Papunet website, and another more general version written in Java. The Java version can interact with ontology data both in the MySQL database (or other dadabases), or in XML/RDF based data sources.

The SYMBERED project is depending on the on-going further development of the CCF providing the underlying resources for the web content editing tools being developed. These tools, on the other hand, provide a first full scale application of, and test bed for, the CCF technology. The SYMBERED project group is therefore working in close co-operation with the CCF efforts in the UK and elsewhere, and with Andy Judson et al at the dept of Applied Computing, University of Dundee in Scotland in particular. The relations between the different aspects of the work is depicted below in Fig. 1.

The SYMBERED editing tool will be tested through the publishing of a number of plain text books/stories with concept/symbol support. These documents will first be published on the SYMBERED project website (www.symbolnet.org 2005-2006), and later on the Papunet site and possibly others. The Java plugin components will be offered for integration in other web authoring and content management systems (CMS:s) .
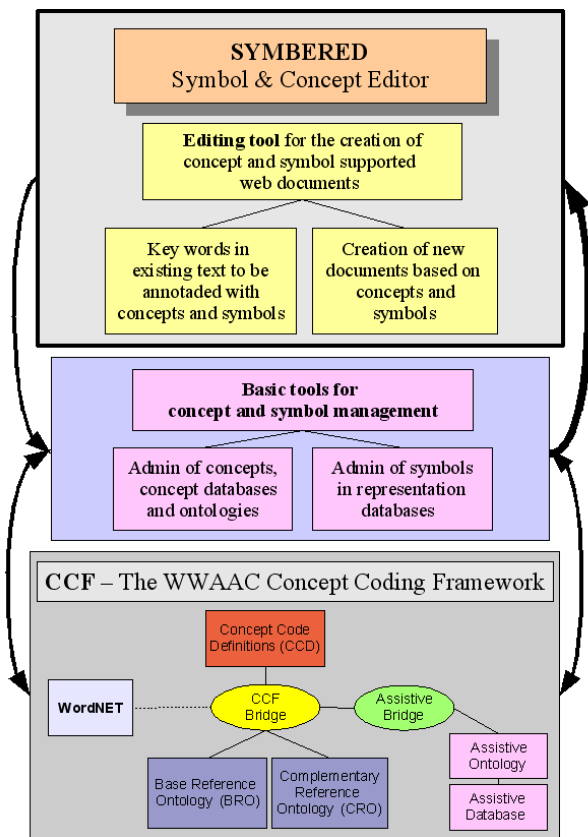
Fig. 1: SYMBERED - CCF Relations

2.1. A View of Concept Coded Content

The following two figures show possible views of a short concept coded English text. It's displayed with two different settings for language and symbol preferences, in the SYMBERED-Papunet web content environment:
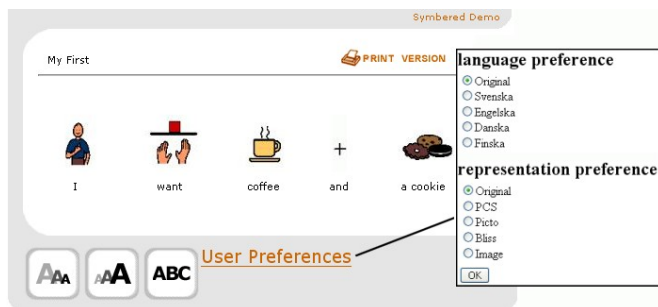

Fig. 2: English text + PCS symbol representation – the original format in which it has been created.
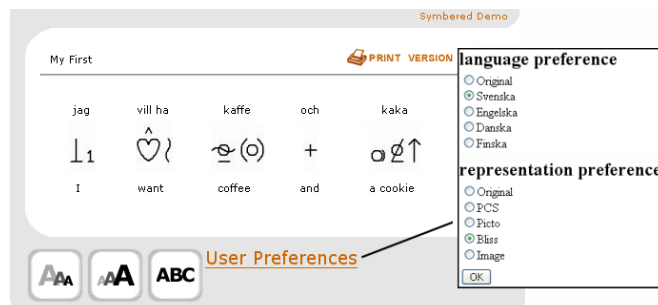

Fig. 3: The same text displayed with a selected preference of Blissymbols and additional Swedish text

Each main word in the sentence has been annotated with a concept code. Each concept code is linked to a range of alternative representations for the underlying concept. A complementary original symbol representation is selected for each original word/concept during edit-time. When the reader views the content, an alternative symbol representation may be chosen in the preferences. An additional alternative text representation may also be selected. As in the figures 2 and 3 above, words and symbols may be re-represented :

Fig. 4: Detail - re-representation of content

The selected text and symbol representations for each concept are vertically aligned using the recommended w3c Ruby Annotation (2001) standard format.

2.2. A view of the SYMBERED-Papunet Editor for Concept Coded Stories

The following sequence of images show how a short text sentence is annotated with concept codes in the SYMBERED-Papunet Story Editor, and then presented with PCS symbol and a second language support:

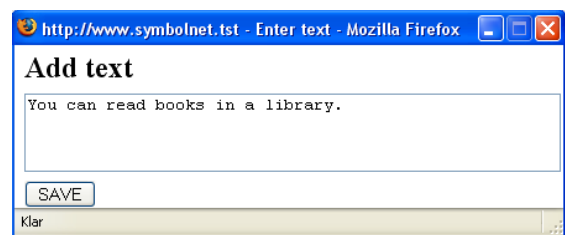Fig. 5: First a New Story entry of the "Concept Story" type is created ...

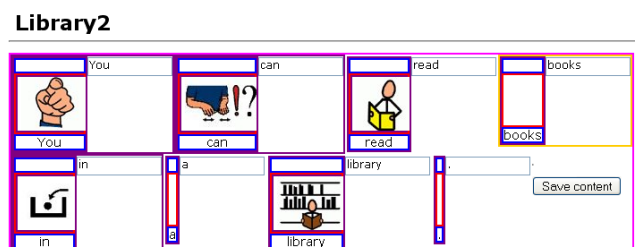Fig. 6: Then the firs part of the story text is entered ...

Fig. 7: The editor looks up possible concept and symbol matches for each word ...

As there is no direct match for the word "books" (in Fig.7 above) this entry is selected for editing via the Editor Concept Browser:
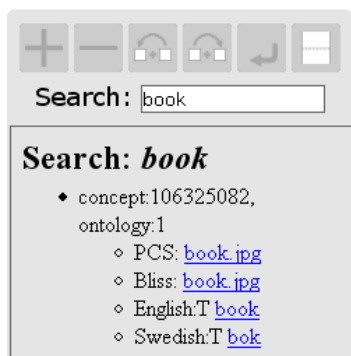

Fig. 8: The word "books" is transferred to the browser, and when the plural 's' is removed matches are found ...

After some minor additional editing the resulting story may appear like this to the reader (depending on pref:s) :
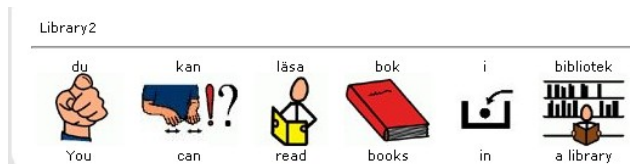

Fig. 9: English text with PCS symbols and additional Swedish text

3. SYMBERED-CCF Technology & Tools

We'll take a somewhat closer look at the technologies and tools behind the above examples: first the underpinning technologies from the SemanticWeb, the CCF, the Java Plugin Framework (JPF), and the Papunet CMS environment; then we'll present the different CCF and SYMBERED tools a bit more in detail.

3.1. Underpinning Technologies

A central ambition for the current SYMBERED development, as well as for the preceding CCF work, within and ouutside of the WWAAC project, has been to adhere to existing and emerging open standards. We have set out to develop concept coded language and symbol support for AAC users and others as applications of, and extensions to, these standard technologies. We have tried to avoid separate and isolated solution as far as possible.

3.1.1. Semantic Web Technologies

The Semantic Web initiative offers a set of standards (RDF, RDFS and OWL) for the representation and exchange of information. The primary strength of OWL/RDF is support for inference and classification, flexibility of representation, handling of semi-structured data, and support for information integration and reuse.

The CCF and SYMBERED builds heavily on the Semantic Web initiative, and uses OWL, RDF, XML, XHTML, and Ruby annotations to both define the datasets for the content and to present the content with as much compatibility as possible with existing and emerging Internet technologies (web browsers, mail systems etc.).

3.1.2. CCF – The Concept Coding Framework

CCF is an openly defined structure to handle concept based language material in a standardised way. It has been designed with a special focus on representing language concepts in parallel with text (in many different languages), graphic symbols (from several different systems) and other alternative representations.

The main purpose of the CCF is to maintain a published restrictive vocabulary for AAC users that can act as a common base for communication between different systems – openly defined or proprietary.
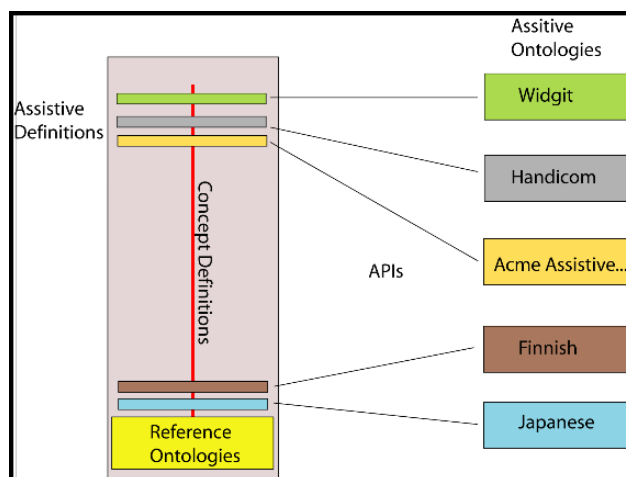

Fig. 10: The CCF - Spine view

The CCF was designed as an Semantic Web application. It is based on two (or more) so called Reference Ontologies in OWL/RDF to be able to infer meanings with proper search modules. The BRO (Base Reference Ontology) is derived from WordNet. It is complemented by the CRO (Complementary Reference Ontology), mainly for function words not contained in WordNet. (Fig. 10 above and Fig. 11 below)
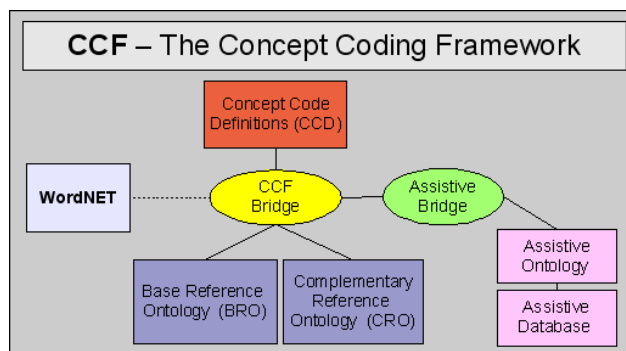

Fig. 11: The Concept Coding Framework CCF and its ontologies

The CCF API allows applications to encode or convert words into semantically enhanced content. It provides a central bridging functionality that connects the different ontologies with each other. This enables other developers to link up their ontologies to be accessible through the CCF.

The structure of the CCF is outlined in Fig. 10 above. For more details, refer to the Concept Coding Framework & Interest Group (2004-2006) and to Judson, A. Hine, N. Lundälv, M. Farre, B. (2005)

3.1.3. The Java Plugin Framework

JPF provides a runtime engine that dynamically discovers and loads "plug-ins". A plug-in is a structured component that describes itself to JPF using a "manifest". JPF maintains a registry of available plug-ins and the functions they provide (via extension points (where it can be extended) and extensions). Plug-ins are not loaded until they are called. (*JPF – a Java Plugin Framework*)

3.1.4. Papunet – PHP - MySQL

Papunet is a Finnish web site aimed at users with limited reading skills and/or in need of graphic symbol support as an alternative.

The Papunet site is built on a dedicated CMS (content management system) implemented in PHP with a MySQL database as a backend for content storage .

In Papunet, words and phrases are connected through a word lookup, binding them to an image or graphical symbol such as Bliss or PCS. Each word-to-symbol binding in a story is unique. This does not allow users to choose between different alternative symbol-sets or language representations for a story.

3.2. SYMBERED Design & Implementation

The general ambition has been to base the design and implementation on clearly defined API contracts and current Semantic Web recommendations. All to achieve a transparent and easily extensible design and implementation. This is more realistic for the more general and long-term Java plugin development. For the short-term SYMBERED-Papunet implementation a great deal of adjustments to the existing environment have been necessary. However, the MySQL database and Java plugin components are tuned for future co-existance.

3.2.1. The SYMBERED-Papunet Demonstrator

The demonstrator was implemented in PHP using MySQL as a backend for data storage of ontologies and content. CCF libraries were reimplemented in PHP and then integrated into the Papunet CMS (content management system).

Parts of the CMS itself have been substantially rewritten for the demonstrator to use the concept codes from CCF as the content encoding and also to add multilingual support.

- The PHP code and MySQL database has been substantially restructured and complemented to include concept ontology and representational data and links for the supported languages and symbol systems - in accordance to the CCF definitions

- In the Papunet CMS there are two main interfaces; presentation and administration. This had to be extended to implement a preference system to allow the user to choose presentation of the content according to needs and preferences.

- For the reader/user the CMS can now present the content with support of symbols and languages that are selected by the user via the preferences settings.

- The presentation view of the symbered was changed to use Ruby annotation for the symbolic view instead of tables, and support was added for recognizing the primary languages of the browser. This is used to give a language specific view of the interface of the demonstrator.

- The administration view contains both content management such as editing of folder structure and creating/deleting/editing of the contained stories. It also provides access to the CCF management tools.

- In the editor words are connected to concepts in the semi-automated process. One of the available symbol representations that is related to each concept is selected and bound to the word in the contents presentation. If the user chooses to see what the author has actually written (text and/or symbols) a user preference of 'Original' is selected. Other settings override or complement the original.

3.2.2. General Purpose Java Plugin Components

The java implementation builds on the fact that api-contract can be fully supported in a plugin framework, in which the full version of the SYMBERED symbol-concept editor is currently being implemented.

The plugins are divided into several groups; core, connector, text processing and application layer plugins.

The central core plugin contains the CCF-API definition, and establishes the API-contracts.

One of the core plugins implements the CCF-API and ontology lookup algorithms. A strength of this API plugin is that it contains a registry of ontologies that has been used to store a lexicon.

The other plugins in the core group implement the CCF-API and act as connectors to let other plugins look like if they were in fact CCF-API plugin.

One of these connector plugins is the WordNet Adapter Plugin that enables word searches directly into WordNet via its API. The WordNet Adapter plugin lets the synsets look like concepts and connects them with the central concept codes.

Other planned plugins are the text-processing plugins, that can process the text and use the CCF-API to look up words in dictionaries, lexicons and ontologies.

One of these plugins will contain the morphological processor to pre-process the words that are to be looked up in the lexicons.

Other plugins are just interface plugins to make external libraries usable by plugins.

The current word search is a plugin that will use the morphological processor plugin in later versions. The current word search plugin implements a very simple text-processing api.

The application layer plugins include the concept-browser and CCF administration plugins among others.

The SYMBERED symbol-concept editor itself acts as an application layer plugin. This means that the editor can also be integrated in other products that uses the JPF or some other compatible plugin framework.

It also enables future extensions such as semantically encoded sentences and phrases. The first such extension is beeing developed and is a part of the search tool and does morphological processing. Other such plugins will be the semantic frame handler to encode Frame Semantics into the content (The Berkeley FrameNet project, 2004-2006).

The following Fig. 12 depicts a component of the open CCF plugin framework that is under development, and in which the SYMBERED editing components will slot in.

The plugin framework is currently based on the JPF opensource resource (*JPF – a Java Plugin Framework*).
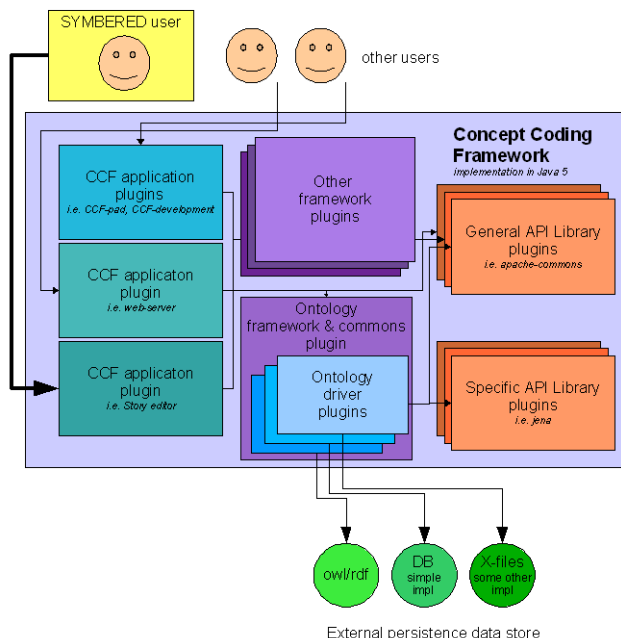

Fig. 12: CCF-SYMBERED Java Plugin Components

The core CCF algorithms are based heavily on the abilities of Jena, which is a Java framework for building Semantic Web applications (Jena – A Semantic Web Framework for Java). Jena provides a programmatic environment for RDF, RDFS and OWL, including a rule-based inference engine.

The CCF core plugin is an adapter to Jena, which implements the CCF OWL/RDF transformations.

The CCF Concept Browser (Fig. 13 below) uses the plugins to discover the ontologies that are available and makes them available for searches and inferences.
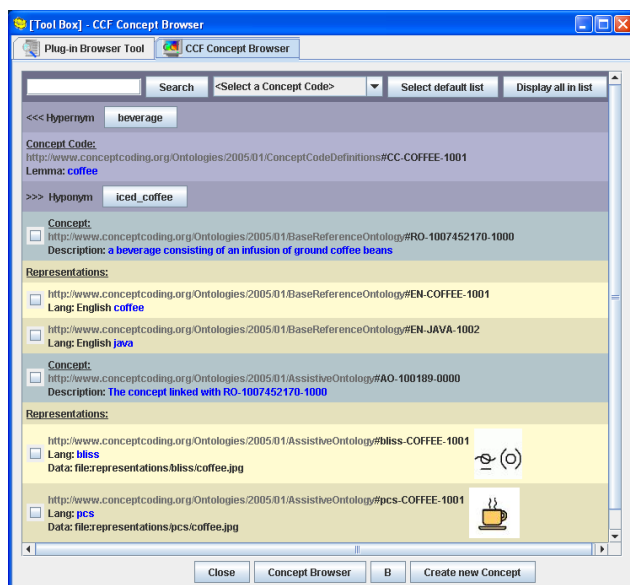

Fig. 13: The SYMBERED - CCF Concept Browser – a JPF plugin component

4. Limitations and Future Tasks

- The CCF still lacks a well balanced core vocabulary of reasonable size. This means that the SYMBERED project will have to build up the concept code databases ad hoc according to the need of the concrete content that is produced. **Future task:** To complement the CCF with a CCD database of 5000-12000 concept codes including the necessary reference ontologies for this.
- The range of alternative representation types is currently limited to a couple of symbol systems and a few languages. **Future task:** To incrementally integrate more symbol systems, sign languages (supportive signing) and natural languages.
- The semi-manual word-concept look-up of THE SYMBERED editor is now relying on manual disambiguation. **Future task:** To add gradually more sofisticated automatic disambiguation functionality to minimise the need of manual processing for correct word-concept binding.
- The first plugin tools for building up and managing the content of the reference ontologies and the conceptcode definitions repository are still under development and primitive. **Future task:** To complement and further refine the range of CCF based plugin tools. This should eventually also include facilities for linguistic support and multimodal representation of the full range of educational material, including E-books etc.

Research tasks:

- To investigate optimal levels of concept-symbol refinement for different user needs: Symbol concepts are often used in a way reflecting rather coarse or fuzzy concept definitions (implicitly or explicitly). This is generally functional, as symbol users usually have access to a quite limited set of symbols, which may in this way be used to a wider range of meanings depending on the context.
How may this be reflected in the CCF in terms of semantic links between a concept code and a range of more or less related fine grained concepts and concept definitions?
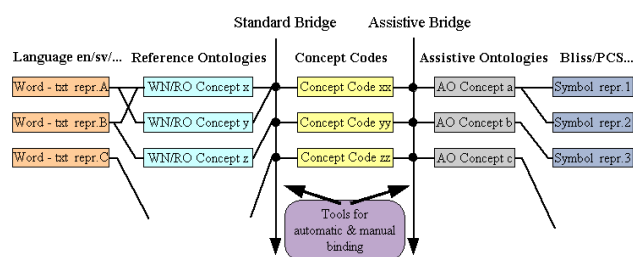

*Fig. 14: Possible word-concept-conceptcode-symbol relations ...*

- How may links to research and development results in relation to the EuroWordnet, GlobalWordnet, Parole and other initiatives be establish to greatly enhance the multi-lingual aspects of the CCF work?
- How may other linguistic research initiatives benefit from the multi-modal potential of the CCF technologies and resources?

5. Summary

The SYMBERED project will result in a first full-scale CCF based application in the form of a web content autoring tool and environment for more flexible language and symbol support.

The impact of these will initially be limited by the still fairly rudimentary ontology and representational resources within the CCF itself.

However, a first set of CCF plugin components and tools are now being developed alongside with the SYMBERED project. These will provide a first platform for starting a gradually speeded up formation of a more substantial infrastructure for the provision of concept coded multi-modal language support for a range of user needs.

SYMBERED Editing Tool
for Concept Codes
... on the web
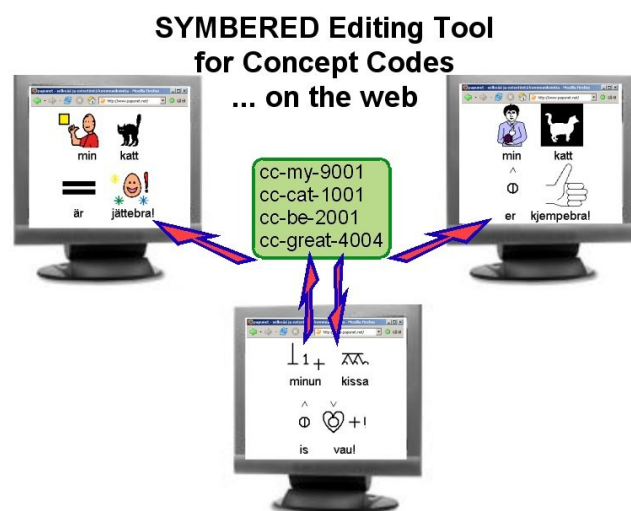
6. References

WWAAC - World Wide Augmentative and Alternative Communication (2001-2004), a European Union IST-project. Website: www.wwaac.org

Concept Coding Framework & Interest Group (2004-2006). Website: http://www.conceptcoding.org

Widgit Communicate: Webwide (2005-2006). Website: http://www.widgit.com/products/webwide/index.htm

The Finnish Papunet website (2001-2006) - with lots of symbol supported and plain text content. Website: http://www.papunet.net

The Italian Navigabile project (2003-2006). Website: http://www.navigabile.it/NavigAbileSP/English.aspx

The Nordic SYMBERED project and demo website (2005-2006). Website: http://www.symbolnet.org

Judson, A. Hine, N. Lundälv, M. Farre, B. (2005) Empowering Disabled Users through the Semantic Web - The Concept Coding Framework an application of the Semantic Web, *WEBIST, Miami, Florida, May, 2005*

Ruby Annotation, W3C Recommendation (31 May 2001). Website: http://www.w3.org/TR/ruby/

Miller, G.A. Et al (2005). WordNet, a lexical database for the English language, Princeton Univ. Website: http://wordnet.princeton.edu

JPF – a Java Plugin Framework. Website: http://jpf.sourceforge.net

The Berkeley FrameNet project (2004-2006). Website: http://wordnet.princeton.edu

Jena – A Semantic Web Framework for Java. Website: http://jena.sourceforge.net/