OPAALS PROJECT

Contract n° IST-034824

WP6: Socio-Constructivism and Language

Del6.7 – Metaphorological Tool Kit: implementation and results

Project funded by the European Community under the "Information Society Technology" Programme

Contract Number: IST-034824

Project Acronym: OPAALS

Deliverable N°: 6.7

Due date: M30

Delivery Date: March 2009

Short Description:

This deliverable provides the description of the Metaphorological Tool Kit and some of the studies that use ideas, approaches and tools provided by this platform. Generally, it covers three areas of research within the OPAALS community and Digital Ecosystems: metaphors as a result of semantic richness, ambiguity and change; MTK as a platform for collection of data and computational (linguistic) analysis of textual information; and metaphors as one of the mechanisms in HCI design and implementation.

Author: Oxana Lapteva

Partners contributed:

Made available to: Consortium

Versioning		
Version	**Date**	**Name, organization**
V1	01.09.08	Oxana Lapteva, UniKassel
V2	15.12.08	Oxana Lapteva, UniKassel
V3	06.03.09	Oxana Lapteva, UniKassel

Quality check

Internal Reviewers: Ossi Nykänen, Raimund Eder

Dependences:

Achievements*	Accomplished work: – Analysis of metaphors in OPAALS (in collaboration with social scientists working on qualitative and quantitative research). The "word associations" results from the questionnaire have been used for analysis of metaphors and their usage in the OPAALS community. On the other hand, social scientists use statistical methods from the Metaphorological Tool Kit for their analysis. Additionally to the questionnaire data, collection of texts (wiki pages made available by Tampere team, deliverables, and newsletters) have been used for terminological work and natural language analysis. – Testing of tools for text analysis, identifying possible drawbacks in respect to the SMEs and DEs. – Implementation of the preprocessing stage in text data analysis. – Program development for transforming results of the dependency grammar parser (made with Stanford parser) to the SBVR-components (such as fact types). Work in progress: - Development of the web interface for the Metaphorological Tool Kit. Data of analysis, results, access to the tools, and suggestions for further work will be provided to the consortium.
Work Packages	<u>WP10:</u> – OPAALS Questionnaire: analysis of "word associations" has been done with the statistical tools from MTK. Furthermore, the results of collocation analysis have been provided for integration into quantitative and qualitative research (T10.9). Future contributions involve the further analysis of textual data, which will be provided for further evaluations to the social scientists. – Data Visualisation: Semiotic analysis of a visual sign helps in the development of visualisation systems as it provides basic requirements for it's understanding. Future contributions involve the research on the dependency grammar for extraction of ontological categories, expanding the list of tools for the analysis of textual data, and collaborative work with TUT regarding methodologies for analysis and visualisation of unstructured textural data. <u>WP9:</u> Statistical results of the "word associations" have been used within this work package. <u>WP2:</u> The extraction of fact types (SBVR) based on the constituency grammar: this approach developed by UniKassel team is already integrated and used in Sepiax Web Editor. The algorithm for extraction of SBVR-components (such as fact types) based on the dependency grammar will be shortly provided. This method will improve the quality

	of correct extraction of fact types from the natural language sentences in the Sepiax Web Editor. Future contributions involve the possible modifications of the parser. WP5: Research on metaphors and their use for HCI design helps developers of user interfaces (e.g. TI, IPTI) to create systems that facilitate learning, orientation and understanding from the user point of view. WP6: Metaphors are one of the steering mechanisms of language evolution, i.e. its change and variation. The research on metaphors provides useful insights for the task T6.6.
Partners	SUAS, TUT, TI, IPTI
Domains	Linguistics: the role of linguistic analysis, especially constituency and dependency grammar approaches Computer Science: analysis and testing of tools dealing with analysis of unstructured textual information, work on development of some methods (tokenizer, filtering, normalisation) Human-computer interaction: role of metaphors in HCI environments and visualisation systems, first view on Sironta and Guigoh from the lense of metaphors.
Targets	OPAALS researchers, Scientific communities, HCI-developers
Publications*	1. Lapteva, O. & Peukert, H. (2007). Linguistic Dimension of Terms: towards community's common vocabulary and concepts. *Paper presented at OPAALS Conference 2007*. November 26-27 2007, Rom, Italy, 2. Lapteva, O. & Zeller, F. (2008). Digital Ecosystem: How metaphors become real environments. *Paper presented at Seventh International Conference on Researching and Applying Metaphor (RaAM7)*. May 29-31 2008, Caceres, Spain. 3. Peukert, H. (2009). Discourse as unlimited semiosis in Digital Ecosystems. Accepted for publication in European Journal for Semiotic Studies.
PhD Students*	Oxana Lapteva – terminological work and analysis of metaphorical use of concepts in OPAALS – testing of tools for text analysis – working on preprocessing methods of analysis – research on the dependency grammar and its role to the formal representation of information – research on metaphors in the HCI design and visualisation systems
Outstanding features*	Work on the parsing techniques for the transforming natural language sentences into their formal representations (e.g. SBVR). Applying the linguistic view and analysis to different scientific domains (social science and computer science), e.g. questionnaire development

	and analysis, analysis of textual information, metaphors in the design of HCI and visualisation systems
Disciplinary domains of authors*	Oxana Lapteva (Computational Linguistics)

The information marked with an asterisk (*) is provided in order to address Recommendation n. 4 from the Year 2 review report

Contents

List of Figures

List of Tables

1 Executive Summary

The Metaphorological Tool Kit (MTK) as a research platform provides new views and insights to the problems that different domains and disciplines are facing in their research regarding Open Philosophies for Associative Autopoietic Digital Ecosystems (OPAALS), Small and Medium Size Enterprise (SME) and Digital Ecosystem (DE). Three main directions become visible in this deliverable: linguistic analysis (e.g. metaphors, terminological work, methods and tools of natural language processing), social research of communities (integration of linguistic analysis into qualitative and quantitative research), and the collaborative research in computer science/engineering domain (e.g. formal representation of information based on linguistic analysis, metaphorical approaches in HCI and visualisation systems).

2 Scope of this deliverable

Collaboration between different domains is always an intricate issue and subject to many discussions. This deliverable provides concrete examples and results of successful collaboration between researchers working in linguistics, social science and computer science. Furthermore, it provides new ideas, questions and problems for further discussions and research. The scope of this deliverable covers the following areas:

- Natural language as a steering mechanism of communication and knowledge representation in DE
  Our investigations are related to metaphors as a result of semantic richness, ambiguity, and change (linguistics), their role in organisational life and discourse (linguistics and social science), and the aspects of linguistic analysis at the boundary between natural and formal languages (linguistics and computer science). Collaborations with different partners in OPAALS are also mirrored in the MTK.

- Computational analysis of textual data
  The integration of different methods and tools from a linguistic and computational perspective has been initiated within this work (collaboration with TUT and SUAS). Within MTK, the UniKassel team provides different methods dealing with core activities of the OPAALS Description of Work (DoW), that is Natural Language Processing (NLP), formal representations of information, and visualisation examples.

- Design of HCI
  This is a new direction of research within MTK that UniKassel discovered through collaborations with TI and IPTI. Therefore, this platform aims to provide a slightly different view and some suggestions to the problems of user interface design (e.g. Open Knowledge Space (OKS)). Furthermore, in the context of DE and SMEs, the visualisation systems are of high importance. Based on the input from TUT, who provided the Wille visualisation environment for testing purposes, we analysed this system trough a "communication" lense, i.e. natural language processing.

3 Introduction

In deliverable D6.3 different views on metaphor have been presented as well as a MTK has been proposed. MTK aims to provide insights of natural language processes and metaphors regarding their role in OPAALS and DE. Furthermore, it serves as an analytic framework involving different approaches for the analysis of textual data that can be applied to any digital environment with natural language processes.

The proposed ideas and tools are used in several tasks within the project (e.g. T6.6, T10.9, T2.1). Some of the approaches and studies provide important directions for further research in both domains, social science and computer science (e.g. T2.1, T2.5, T5.5, T10.5). Close collaboration with social scientists within tasks T6.1, T10.9, D9.7 aims to integrate qualitative and quantitative research methods. The research on metaphors provides new views on the problems of organisational communication and discourse as well as helps to understand the dynamics of communities. For example, the section "Word Associations" in the online questionnaire had been collaboratively developed in order to provide a platform for research in different directions:

- the use of relevant concepts in the OPAALS community
- qualitative research that provides social scientists with valuable information about knowledge flow and change over time in communities
- how the use of linguistic concepts and metaphors influences the communication processes among partners

Additionally, some of the tools presented in MTK have been used for statistical analysis, evaluation and visualisation of questionnaire results (see D10.9, D9.7 for more details).

Another area of applicability is the analysis of textual data (natural language) in DE, serving both as a means for social science related research (see listing above), and as an important part of computational research, developing software applications for DE. For example, MTK provides an important input for SUAS working on the Sepiax Web Editor (Semantics of Business Vocabulary and Business Rules (SBVR)). Furthermore, several methods open a promising collaboration (which has already been initiated) with TUT in two directions: visualisation of textual information and ontological representation of information. Moreover, we propose the idea of using metaphors in design and development of tools, platforms and other types of environments in DE. This work provides new valuable insights into research

in the area of HCI that might be of importance for partners working on OKS design and implementation (e.g. TI, IPTI). Finally, working on the theory of DE from the metaphorical perspective would help the OPAALS community to describe this concept, identify and specify components of it, understand their functions and interaction flow. The aim of the MTK is not only to provide the research on metaphors from the linguistic point of view. It aims to create a platform for an interdisciplinary collaboration within the OPAALS community.

This deliverable is structured as follows: in Section 4, we are going to present the notion of MTK and its structure. Section 5 describes some of the case studies and projects that are already using MTK. Finally, in Section 6 we discuss the possible outlooks of expanding and using MTK. The significance of research related to natural language and metaphors in the context of the OPAALS community and DE is illustrated through the deliverable.

4 Metaphorological Tool Kit (MTK): goals and structure

MTK covers three areas of research within the OPAALS community and DE. First, we deal with metaphors as one example for natural language regarding semantic richness, ambiguity and change. In the context of the OPAALS community we observe the formation of new terms, shifting the meanings of existing concepts, and expanding the scope of possible meanings of the same concept. According to Knowles and Moon (2006), metaphor is a "basic process in the formation of words and word meanings. Concepts and meanings are lexicalised, or expressed in words, through metaphor. Similarly, the names of many new concepts or devices are metaphorical or extended uses of pre-existing words: for example, computer terms such as web, bug, and virus"(p. 4). The same dynamics exist within SME, for example the invention of new products is often done through metaphors. The aspects of semantic analysis involving metaphors are crucial in the DE since the communication between SMEs through natural language requires avoiding its ambiguity and herefrom resulting potential misunderstandings.

Second, we use MTK as a platform for collection of data and computational (linguistic) analysis of textual information. The analysis does not only apply to the metaphors in text. MTK is important for other areas of OPAALS research such as visualisation of information search and retrieval, and others (see Sections 4.2, 5.2, and 5.3).

Third, the design and implementation of platforms, software, tools and other environments are often based on metaphors. For example, the iconic desktop design of many platforms (with Apple as its originator, one famous example being the dustbin) represents visual metaphors and has been a cornerstone of interface design (see also works in computational semiotics and interface design, such as Souza (2005); Nadin (1997); Winograd, Bennett, Young, and Hartfield (1996)). Therefore, we hold that our investigations provide useful input for developing user interfaces for SMEs and for OKS within the OPAALS community.

4.1 Metaphors in OPAALS and Digital Ecosystems

Metaphors represent our way of thinking about specific concepts. In the context of DE this aspect is of high relevance. Understanding metaphorical mapping would help to prevent misunderstanding at different levels: interpersonal, organisational, scientific and cultural. "Metaphor is important because of its functions - explaining, clarifying, describing, expression,

evaluating, entertaining" (Knowles & Moon, 2006, p. 4-5).

Metaphors as a part of daily communication open a problem of semantic ambiguity. One of the important aspects of metaphor research in the context of OPAALS and DE is the gap between natural language and formal languages. Metaphors impede this process of merging and bringing closer together the natural and formal languages. In digital environments, people might face a problem of the necessity to translate their ideas, problems or solutions into code machines can understand. How can we bridge this gap? Can we at all, given the overall challenging problem of handling natural language input in unrestricted contexts by means of binary structures, i.e. automated/computational processes? It is important to identify and define the problems we are facing here. One of the major difficulties is the ambiguity in natural language on the one side and the requirements for precision, unambiguity and comprehension on the other - formal - side. In this case, linguistic analysis could reduce the dimensions of complexity and ambiguity of natural language mirrored in concepts, phrases, sentences, and even paragraphs. MTK provides some applications that can be helpful in further analysis and integration into different applications.

We hope, this work on metaphor from a combined linguistic and computational perspective will help social science researchers to look at organisations (SMEs), communities, and cultures through the lense of natural language interaction to understand their structure, functionality, and interactional aspects (see deliverable D6.3 for more detail). Furthermore, metaphors are one of the steering mechanisms of natural language change and variation. This aspect is not only important for the investigations within task T6.6 "Evolutionary Framework for Language". In the context of DE and SME, this aspect is crucial considering, for example, aspects such as the invention of new products and the creation of a name for them, the shift/change of meaning of a concept, or the expanding the meanings of existing words (e.g. memory, program, disk, and many others).

4.2 Collection of tools

MTK provides a collection of tools for analysis of textual data [1], i.e natural language in non-restricted contexts. Due to the complexity and unstructured character of such data, different approaches can be used depending on the goal of analysis. Several methods have been tested: preprocessing, concordance, collocation, frequency analysis, parsing techniques, term and keyword extraction.

[1]See Appendix 6 for the list of tools and their short description.

4.2.1 Preprocessing

In every application dealing with unstructured textual data, the **preprocessing** stage is essential. In MTK, we use tokenization, filtering and normalisation methods. There is a variety of tools dealing with tokenization. However, in respect to SMEs, we see the weakness of them in identifying prices as tokens. For example, "200$" is usually identified as two separate tokens. Therefore, we developed the tokenizer considering this issue. In order to keep control over the preprocessing steps, we are implementing filtering and normalization steps as well. Figure 1 illustrates the exemplary output of preprocessing involving tokenizer (with identifying prices as one token), filtering and normalization. The original text sample is *"While a student at the University of Texas at Austin in 1984, Michael Dell founded the company as PC's Limited with capital of $1000. In 1985, the company produced the first computer of its own design, the Turbo PC, sold for US$795."* (taken from http://en.wikipedia.org/wiki/Dell). During the filtering process, the nonspe-

Figure 1: *Preprocessing of data*

cific words are removed (in information retrieval, they are called "stopwords") and the normalisation process involving low-case conversion has been accomplished. After the preprocessing step, different techniques can be applied to the results (e.g., keyword extraction, extraction of specific patterns based on the part-of-speech tagging, filtering specific information such as email address, phone number, text mining methods, etc.)

4.2.2 Concordance

Concordances provide the way of looking at a pattern (word, phrase) within the local context. For this purpose we use the tool provided by the Text Anal-

ysis Portal for Research (TAPoR)[2]. It uses Keyword In Context (KWIC)[3] type of concordancing, i.e. displays a list of lines with the pattern in the middle (Figure 2).

a specific session dedicated to this WP - integration of	**digital ecosystem**	infrastructure. Results from other WPs, especially 2, 3, 4
/ www. comp. hkbu. edu. hk/ ~comp2221/ I-know-UML. ps] 29-05-07	**Digital Ecosystem**	adoption problems. A software-social perspective. Some questions, no answers:
research outline and set out a conceptual framework for studying	**digital ecosystem**	governance from the point of view of innovation, sustainability
a specific session dedicated to this WP - integration of	**digital ecosystem**	infrastructure. Results from other WPs, especially 2, 3, 4
we could never do without the first, which in a	**digital ecosystem**	(" digital" here is meant in the narrow sense
external monitoring environment: By means of the Indian DEAL (	**Digital Ecosystem**	for Agriculture and Rural Livelihood) project platform for the
? language as the analogue to the DNA in a	**digital ecosystem**	. And as in nature, a digital ecosystem grows by
DNA in a digital ecosystem . And as in nature, a	**digital ecosystem**	grows by replicating and evolving its DNA, that is
of language must be established and remain intact as the	**digital ecosystem**	grows, in order to guarantee the compatibility with new
impact on relationships between different SMEs and on the overall	**digital ecosystem**	. In this page we will summarize the main outcomes
their business focus and typical organization; ? ? ?	**Digital Ecosystem**	services and software technologies - applications running on top
also the aspects related to accountability and trust for the	**digital ecosystem**	will be considered, with the objective of facilitating behavioural
who have integrated and shared novel IT services using the	**digital ecosystem**	infrastructure. Fieldwork will be comprised of micro-level investigation into
The overall aim of this research is to develop a	**digital ecosystem**	governance taxonomy around which debates on core issues can
within a highly dynamic and purely distributed environment of a	**Digital EcoSystem**	[/ EcoSystem] . But at the same time,
define models for distributed identity, accountability and trust for the	**digital ecosystem**	that facilitates behavioural and economic collaborations. Activities in this
identity and trust which act as an input into the	**digital ecosystem**	infrastructure deployment in WP5. The integration and deployment will
ideas We would like to outline a set of high-level	**digital ecosystem**	scenarios [/ WP4_Distributed_Accountability_Identity_And_Trust/ Scenarios] we believe could potentially
months. WP4 : D4. 1 Distributed Identity Model for the	**Digital Ecosystem**	Javier Noguera Techideas Llacuna, 162 08018 ? ? ?
suffering a digital divide. Experiment rural community networks into	**digital ecosystem**	for the socio-economic development in India. Tasks T7. 1-
trend to extend traditional disciplines into new disciplines through the	**Digital Ecosystem**	paradigm. The goal of the conference is to attract
LSE) This task sets out a conceptual framework for studying	**digital ecosystem**	governance from the point of view of innovation, sustainability

Figure 2: *Concordance of the word "digital ecosystem": OPAALS wiki corpus*

Concordances are less suitable for further computational analysis involving search and extraction of specific patterns. However, this method is widely used for qualitative analysis, for instance, analysis of metaphor use in text corpora is often based on the results of concordancing. Evaluating the context, where a specific concept occurs, is also useful for social science related research on discourse. For example, together with social science methodologies in content analysis, aiming also at the analysis of framing, the MTK

[2] http://taporware.mcmaster.ca/ taporware/.

[3] Another type of concordancing is Keyword Out of Context (KWOC), which displays the pattern separate from a context(sentence or paragraph).

could be adopted and hence a framework for semi-automated analyses of any descriptive texts (such as company descriptions, service descriptions, etc.) from SMEs within the DE could be developed.

4.2.3 Collocation

Collocation analysis aims to "identify words that occur together significantly more often than one would expect by pure chance" (Mason, 2000, p. 213). Consider the following example: we want to find out the context where the word "network" occurs. Instead of reading a large collection of data, we use the collocational analysis. The idea is to find collocates (words) "which occur within a certain distance from the node word". The evaluation of the candidates is usually based on the frequency of their occurrence in the defined environment (other methods of evaluating the significance of collocates are *t-score* and *mutual information*[4]). A defined environment is the fixed span of a specified number of tokens to the right and left side of the node (i.e. the patterns we wish to analyse). For example, we might wish to find collocates that are within the distance of 10 tokens from the specified node "network". The result of collocational analysis that we applied to the OPAALS wiki-corpus is shown in Figure 3. It illustrates the collocates of word "network" based on the frequency of co-occurrence. The

Figure 3: *Collocational analysis of the pattern "network": OPAALS wiki corpus*

collocation approach is used in task T10.9 for a comparative analysis of the concepts used in the OPAALS community (e.g. questionnaire data on "word associations" versus the collocations of the concepts in the wiki corpus).

[4]The description of these methods can be found in Appendix 6.

4.2.4 Parsing

Parsing is a process of "using a grammar to assign a syntactic analysis to a string of words, a lattice of word hypotheses output by a speech recognizer" (Carroll, 2003, p. 233). In MTK, we use two types of grammar: constituency and dependency. The idea of the constituency approach is that "groups of words may behave as a single unit or phrase, called constituent" (Jurafsky & Martin, 2000, p. 324). In this case, the parser is able to identify these units and provide a hierarchical structure of a sentence. For instance, consider the sentence "I prefer a morning flight". An exemplary result of parsing this sentence is shown in Figure 4.

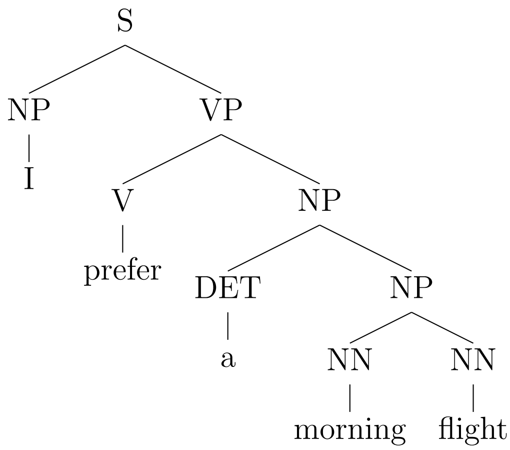


Figure 4: *Parsing with constituency grammar*

Figure 5 illustrates the result of applying the dependency grammar concept to this sentence. The list of dependency relations is provided in Appendix

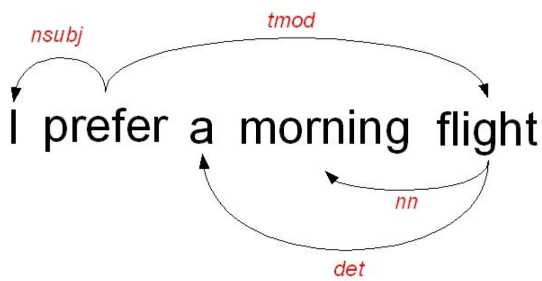


Figure 5: *Parsing with dependency grammar*

6. The parsing techniques are useful in a variety of applications dealing with textual data. In respect to the OPAALS research we focus on using parsing

techniques for specific tasks. In Phase I, the UniKassel team showed the importance and necessity of syntactic analysis in NLP systems in general and for formal representations such as SBVR, in particular. Furthermore, an algorithm for extracting SBVR fact types from the parsed sentence[5] was provided to the SUAS-team working on the development of the Sepiax Web Editor (see deliverable D2.2 for more detail). Further research of parsing approaches led us to the idea of testing a different method, the dependency grammar. The detailed description of this approach is provided in deliverable D6.4. We argue that the dependency grammar could be successfully applied not only to the Sepiax Web Editor, but that it could also serve as an important component of systems dealing with the transformation from natural language to its formal representation. Generally speaking, we can say that the formal representation of natural language sentences (e.g. SBVR and ontologies) requires first the identification of concepts (actors, subject, object of action, etc.) and relations among them (actions, manipulations, etc.). Verbs in natural language are items that comprehend the relations between concepts. Since dependency grammar is verb-oriented, i.e. the verb is defined as a head of the sentence, it seems to be a promising approach. Some of investigations in this direction have been done by Fundel, Küffner, and Zimmer (2007). The authors integrated the dependency grammar into the system that deals with the extraction of ontological relations from text. Within the MTK we apply the Stanford parser to identify both, the hierarchical structure of the sentence and the dependencies between words in a sentence. Furthermore, we develop a program that extracts the SBVR-relevant components such as fact types from the output dependencies. First tests show that our parsing technique using dependency grammar improves the quality of extraction entities relevant for the formal representations. This work is going to be used in the Sepiax Web Editor developed by SUAS. Additionally, we see a great potential of further research in this direction in the context of ontologies. Therefore, this part of MTK is also relevant and interesting for partners working on knowledge representation systems, semantic web and ontologies[6]. Likewise, in the course of the planned collaboration in Phase III, the partners' feedback will provide a valuable input for UniKassel on further work on the subject of natural language processing.

[5]The parsing was carried out with the Stanford parser available at http://nlp.stanford.edu/software/lex-parser.shtml.

[6]Collaboration with Tampere team.

4.2.5 Statistical analysis

At the current stage of research, we use two kinds of **statistical analysis**: frequency analysis of extracted items and C-value analysis. Frequency analysis is one of the common methods used in computational linguistics. We use it in collocations, extraction of concepts, visualising the most common words used in text, on a web site, or in the collected corpus. This method has been used by the analysis of the word associations provided by the OPAALS members in the online questionnaire. An exemplary output is shown in Figure 6[7].

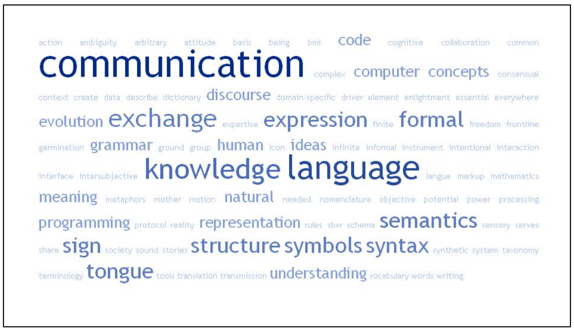


Figure 6: *Frequency analysis and visualisation: word associations of the term "Language"*

C-value method aims to extract multi-word terms in a text. The detailed description of this method has been already provided in deliverable D6.3. In MTK we use the TerMine tool[8] for the purposes of automatic term recognition and extraction.

4.2.6 Term extraction

Additionally to the TerMine tool that recognises terms based primarily on the statistical analysis method (C-value), we use the TermExtractor environment for identifying terms, concepts and applying metaphor analysis to the extracted collection. In the context of SME as well as scientific communities such as OPAALS, extraction of terms represents an important input for research related to different tasks, for instance, domain ontology construction and knowledge management. The advantage of the TermExtractor is that it allows the identification of domain-relevant terms.

[7] OPAALS questionnaire.

[8] Available at http://www.nactem.ac.uk/software/termine/.

4.3 Metaphors and user interfaces in Digital Ecosystems

One crucial issue in DE is the design and implementation of appropriate user interfaces. As already mentioned above, the metaphorical approach within the field of HCI receives special attention from researchers in computer science and engineering domains. An important advantage of this approach is that it "seeks to increase the initial familiarity of actions, procedures and concepts by making them similar to actions, procedures and concepts that are already known" (Carroll, Mack, & Kellogg, 1988, p. 67). There are three main directions of research on metaphors in HCI (Carroll et al., 1988):

- Operational analysis focusing on learning and performance effects of metaphors in a user interface.
- Structural analysis dealing with the definition, recognition and representation of metaphors as well as their mappings from source to target domains.
- Pragmatic analysis dealing with context analysis. For instance, the question *how a specified metaphor operates in real HCI environments* is crucial for the overall design of a system.

One of the successful interface metaphors in the design of digital environments is the desktop metaphor. Table 1 illustrates further examples of applications and metaphors associated with them.

An important aspect in DE and SMEs is the question of how to design and implement HCI environments that support the work and have a positive effect in organisational life. Morgan (2006) suggests to conceptualise the dynamics, structure and functions of an organisation in terms of metaphor. Several attempts have been already made in the area of HCI, for example the notion of a "digital office" where all information is processed electronically. Preece and colleagues (Preece et al., 1994) argue that "metaphors can be used to present a coherent image of the whole system or to deal with specific functions or parts of the system" (p.456).

Furthermore, the right choice of visualisation metaphors is crucial in the representation of communicative and functional components of a user interface. The most common type of visual metaphors are icons. In addition, the visualisation of information and extracted patterns is important in DE as well. This is done through applying different kinds of graphical representation, for instance, bar charts, tree diagrams, and many others. These graphical representations "derive much of their power to inform and enlighten

Application area	Metaphor	Familiar knowledge
Operating environment	The desktop	Office tasks, file management
Object-oriented environments	Physical world	Real-world behaviour
Hypertext	Notecards	Flexible organisation of structured text
Learning environments	Travel	Tours, guides, navigation
File storage	Piles	Categorising objects in terms of urgency, projects and so on
Computer-supported cooperative work	Multi-agents	Travel agents, butlers and other serving roles

Table 1: *Applications and Metaphors (from Preece et al. (1994))*

through the use of graphical (or visual) metaphors" (Risch, 2008, p. 1). The visual metaphors map a visual pattern of a particular type (tree, chart, diagram, network, etc.) to some aspects of the information of interest. Two components of visualisation metaphors are of high importance:

- the subject of representation (what do we want to represent), and
- the method or type of representation (how do we represent).

We argue that metaphorical view and analysis can provide useful insights and helps to create visualisations that users can understand just by looking at[9]. Visual metaphors, be an icon representing a specific component or function in a system or a graphical representation of information, are all created through signs (i.e. constituting natural languages). As V. L. Averbukh (2001) mentioned, "to choose a metaphor is to choose a sign system that will be used for visualisation" (p. 230). In the context of visualisation systems, we suggest to use the semiotic view of a sign (Peirce, 1931-1958). Looking at the visualistions in a Peircean tradition helps to identify three components of any sign in design of HCI environments (see Figure 7). The parts of the Peircean sign system build an ensemble through the functions

[9]Considering the scope of this deliverable, an extensive discussion as to the importance and role of visualisation techniques in DEs and user interface design, cannot be provided here. Please refer to (Ware, 2004; Card, Mackinlay, & Shneiderman, 1999; Zhang, 2008); and more specifically TUT's work on visualisation and IPTI's expertise in perception.

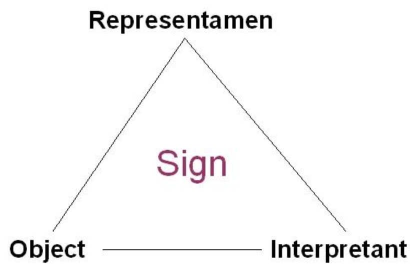


Figure 7: *Peirce's Triadic Sign Model*

of representation, communication and signification. The semiotic approach can be applied to different parts of user interface design, such as icons of interface components (icon of a chat program, email icon) or specific functions (e.g. "cut", "print", "save"), or a development of visualisation systems where a combination of signs requires a more complex analysis. In any case, the interpretation of the visual sign(s) is a vital aspect in the design of HCIs.

5 MTK: case studies and related projects

MTK not only provides a collection of data (corpora containing OPAALS wiki pages, deliverables, newsletters) and tools for analysis of textual data, it serves as a platform for different case studies, projects and the collection of ideas. Currently, the following directions of research can be identified:

- OPAALS concepts, their metaphorical use and influence on communication within the OPAALS community
- Text analysis of OPAALS data
- Role of linguistic analysis in formal representations
- Analysis of OPAALS' user interfaces and visualisation systems

5.1 Case Study A: OPAALS' concepts and metaphors

Scientific languages are very rich of metaphors. Another aspect is that researchers try to explain complex problems and subjects more easily. An attractive mechanism of doing so is a metaphor. At the current stage of research, metaphors are difficult to automatically identify and extract. Nevertheless, it is an interesting representation of complexity in natural language, on the one hand, and serves as a means for the creation of new knowledge and mapping to existing knowledge, on the other. In OPAALS, especially at the beginning of the project, we observed in collaboration with social scientists as to how metaphor can mislead the understanding of the same concept of different scientists and how it can help to understand the meaning of it.

This research was carried out in order to see whether "OPAALS' language(s)" is rich of metaphors, what metaphors are interpreted differently depending on the scientific domain, and gain insights as to how metaphors influence the communication processes in the community. One important restriction was made during our study: we focus on terms and concepts that are used metaphorically. The reason for this is that in the OPAALS community terminology and concepts play a vital role. They transfer important ideas, new or modified meanings and are important termini of the community.

We gained the first insights into concepts and their metaphorical use by means of the OPAALS questionnaire, a long-term study in which the OPAALS consortium participated. The participants were asked to provide their associations to several concepts that are of relevance in the research on DE: knowledge, network, language, digital ecosystem, community, col-

laboration[10]. A qualitative analysis of the information submitted in three questionnaire waves showed two main concepts that are used metaphorically depending on the domain of research. These concepts are "Digital Ecosystem (DE)" and "Network". The concept of "Digital Ecosystem (DE)" is seen in the lights of Virtual Environment, Biological Environment, Social Organism, and Network. This information is relevant regarding the development of a theory of DE, identifying functions of its existence, development, change, and interactions with other environments[11].

Figure 8 represents results related to the concept of "network". Here, in

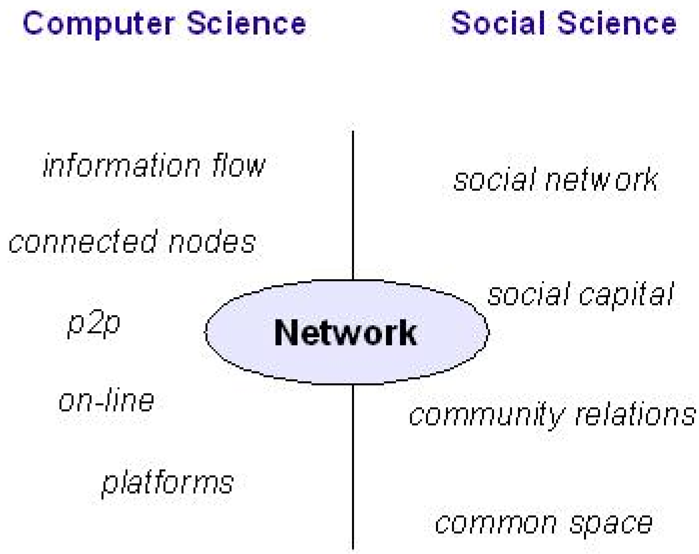


Figure 8: *"Network"-metaphor in OPAALS*

collaboration with social scientists we were able to identify differences in the usage of this term. This was achieved through qualitative analysis of results that have been gained based on the frequency-based method. It is interesting to note, that the word "peer" in the concept "peer-to-peer network" is used explicitly in the Computer Science domain. However, it is "social" in its nature. According to (Gilmour et al., 2002), peer is defined as:

- member of the nobility
- person of the same status, age, etc.

[10]see Deliverable D6.1; Braeuer, Crone, Dürrenberg, Lapteva, and Zeller (2008); Braeuer, Crone, Dürrenberg, and Zeller (2008) for more details.

[11]see Deliverable D12.1.

The term "network" is used differently in scientific domains. An interesting example of metaphoric influence on communication was observed during one of the OPAALS' meetings in Phase I. Participants from social and computer science domains referred to different "meanings" (or, in metaphor terms, mappings from source domain to the target) of the same concept "identity", which caused misinterpretation, miscommunication and misunderstanding.

The "word associations" in the questionnaire study were the first step in our analysis. After a quantitative and qualitative analysis of the results, the UniKassel team decided that it would be beneficial to expand the study of concepts and their metaphorical use in terms of including the overall "OPAALS corpus" (i.e OPAALS wiki pages, deliverables, newsletters, and description of work). The procedure contained several steps:

R	Term	Acronym	Weight
☐	digital ecosystem		0.965
☐	knowledge space		0.872
☐	social network		0.853
☐	research interest		0.825
☐	research activity		0.805
☐	social science		0.797
☐	business model		0.791
☐	community member		0.786
☐	community building		0.783
☐	concept map		0.780
☐	sustainable community		0.770
☐	knowledge creation		0.769
☐	feedback loop		0.769
☐	content creation		0.763
☐	group work		0.763
☐	code generation		0.762
☐	discussion paper		0.756
☐	plenary session		0.752
☐	software development		0.751
☐	file repository		0.744

Figure 9: *Terms from OPAALS wiki*

- Building a collection or corpus of OPAALS data
  For this purpose we collected different kinds of data: official documentation in form of deliverables, description of work and OPAALS wiki data (in connection with TUT).

- Data Preparation
  The data collection was filtered, cleaned and processed to the ASCII format for further analysis.

- Analysis of Data
  The collection was analysed with the TermExtractor tool. The output

was the terms and concepts used in corpora. An example is illustrated in Figure 9.

- Qualitative analysis of extracted terms
  Qualitative analysis has shown that the majority of terms and concepts was used metaphorically. Furthermore, our analysis proved that the metaphorical use of the same concepts varies depending on different research domains, such as the terms "network" and "identity". A variety of multi-word terms extracted from OPAALS data are metaphors in its nature. Table 2 shows a few exemplary multi-word terms from OPAALS and their corresponding domains found in WordNet.

Multi-word term	**Domain Word 1**	**Domain Word 2**
Social Network	Animal	Electronics
Evolutionary Framework	Biology	Modelling
Community network	Ecology	Electronics
Digital Ecosystem	Electronics	Biology

Table 2: *WordNet domains of OPAALS metaphors*

In Digital Ecosystems, we must be aware not only of the existence, but also of the dynamic change of semantic meaning closely related to language evolution. The process of semantic change is often caused through the creation of new metaphors or change/shift of metaphorical meaning in a specific domain.

5.2 Case Study B: Text Analysis

The amount of textual information in digital environments is growing. Therefore, the analysis, pattern extraction, search and retrieval of relevant information is significant in our research. Furthermore, the bridge between natural language and formal language is crucial in DE.

Within MTK we did several analyses of tools and their applications based on OPAALS data (deliverables, wiki). The description is provided in Section 4. In this chapter we are going to describe some results of text analyses and examples of possible visualisations.

A crucial part of processing textual information is the pre-processing step. Depending not only on how good this step is implemented but also what methods are necessary for a given problem, the pre-processing step is

usually a combination of different applications. Almost in every application we need to apply the tokenization in order to identify words (or sentences, or phrases). On the first sight, this process seems to be very trivial and easy. For example, when we tokenize the sentence "The company sells computers", it seems to be no problem for our tokenizer to correctly identify all the tokens: "the", "company", "sells", "computers". However, the standard tokenizer would fail to do the tokenization for the following sentence: "The company sells computers for 2.000$". The standard tokenizer is not able to identify "2.000$" as one token. In business environments such little issues would cause wrong analysis from the very beginning influencing the final results. The output of the preprocessing step involving the tokenizer we developed to meet the specific demands of this project is presented in Section 4.2.1, Figure 1.

Within MTK, we applied different analysis tools to three types of collections:

- Comments from the Review 2008
- OPAALS wiki corpus
- OPAALS deliverable corpus
- OPAALS newsletters

The results of the analysis will be shortly published online (in the Web interface of MTK), thus the data can be used by other partners for further analysis or visualisation.

One of the first methods we applied was the discovery of the most frequent words used in each corpus. The procedure is illustrated in Figure 10. Additionally, the C-value method and TermExtractor have been used. Based

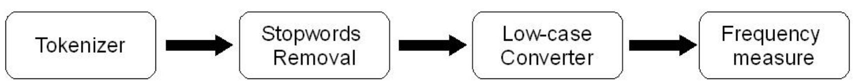


Figure 10: *Discovery of frequent words*

on the results of statistical analysis, we selected several terms that frequently occur in our corpus and applied collocational analysis to them. For example, one of the common terms used in the documentation of the reviewers' comments regarding OPAALS' work in Phase II is "integration". The expanded output of the collocational analysis for this pattern is illustrated in Figure 11.

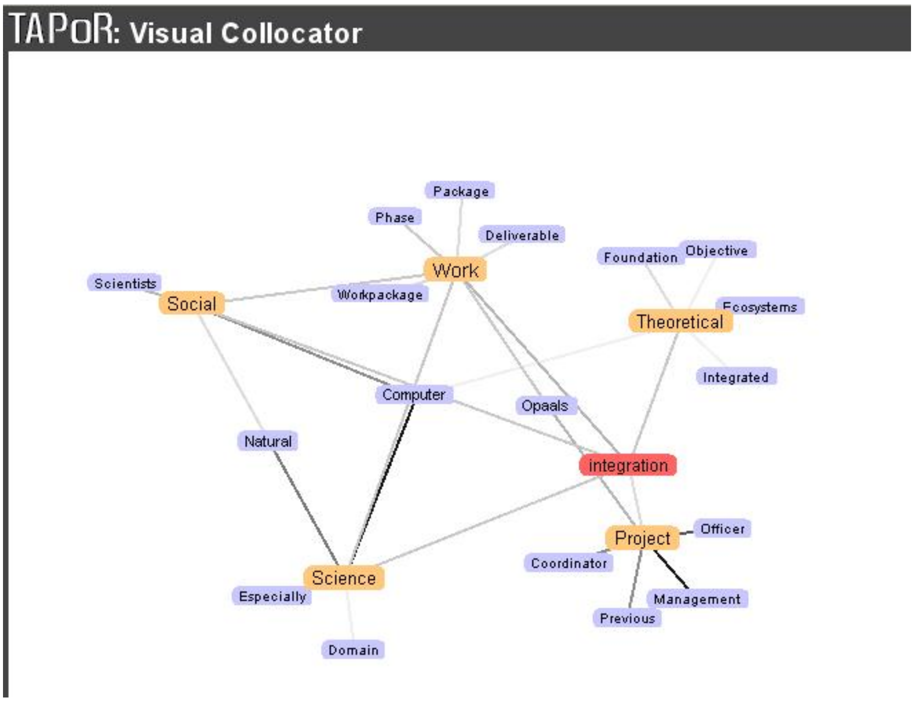


Figure 11: *Collocations of the pattern "integration"*

Collocatoinal analysis of the word "integration" identified five lexical units that are strongly correlated in the text: science, project, theoretical, work and social. However, these words alone do not provide a lot of information. Further expansion of collocations produces the visual network providing interesting results. For example, the word "computer" is a connected item between words "social", "theoretical", and "science". This indeed reflects one of the important discussions related to the importance of collaboration between domains.

Text analysis methods aim to reduce the complexity and dimensionality of unstructured textual information in order to provide a platform for quantitative and qualitative analysis of data. In the example of the review's comments the textual data indeed came in a certain structured form, however given the high amount of input, this collocation served as a tool to support the decision making processes that resulted from the review. Furthermore, different visualisation techniques can help to illustrate a variety of patterns and characteristics of data.

5.3 Case Study C: MTK and Formal Languages

In this section, we provide an example of how MTK can be used in relation to formal languages. We introduce some aspects of extracting information necessary for SBVR from sentences written in natural language.

In the context of natural language, the core items important for SBVR are the verb and its relation to the subject/actor and object. Identifying the verb using a constituency approach is possible. However, some cases such as passive constructions might cause problems. Furthermore, identifying the correct subject and object often fails with constituency grammar, when the sentences are longer or the subject appears after the verb. Dependency grammar focuses on the verb identification and the dependencies between different parts of the sentence. In the MTK, we built a program, that uses the results produced by the Dependency Grammar (Stanford Parser) and extracts verbs, subjects, and objects that are in some relation to this verb (= head word). First tests have shown that even the passive constructions such as "The products have been bought by the company" have been processed correctly. The exemplary result is presented in Figure 12. This method will be provided to SUAS in order to integrate it to Sepiax Web Editor and to improve the quality of extraction of SBVR fact types from a sentence written in natural language.

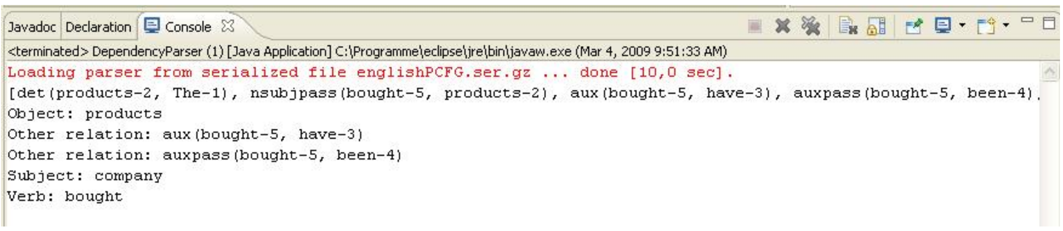


Figure 12: *Dependency parser: extraction of relations in a sentence*

The dependency grammar approach can be used not only for SBVR-purposes. We see a great potential of its using in research on ontologies. Some general ideas have been already presented in deliverable D6.4. An important element of ontologies are the relations between concepts. Since the verbs are the bearers of relations, their correct extraction would help in automating the process of ontology buildings from the unstructured textual data. For this purposes, further investigations of existing relations in dependency grammar as well as testing different sentences and sentence structures with a dependency parser are necessary in order to improve the quality of a desired output (SBVR fact types, ontology relationships or concepts).

For the purposes of ontology building, the TermExtractor tool can be

used for the extraction of concepts from a collection of textual data. This tool finds terminological candidates based on the extraction of terminological structures such as, for instance, compounds (community building), adjective-noun (digital ecosystem), noun-preposition-noun sequences (description of work, grammar of metaphors, exchange of goods, character of money) and calculations of five values Domain Relevance, Domain Consensus, Term Cohesion, Artificial Frequency, and Term Weight for each candidate[12]. In the MTK, we applied the TermExtractor tool to the OPAALS Description of Work (Figure 13).

The qualitative analysis of the results shows the succeseful identification

Term	Term Weight	Domain Relevance	Domain Consensus	Lexical Cohesion	Artificial Frequency
digital ecosystem	0.917	1.000	0.794	1.000	1.000
business model	0.824	1.000	0.887	0.331	0.361
computer science	0.804	1.000	0.793	0.516	0.352
evolutionary framework	0.797	1.000	0.808	0.536	0.199
community network	0.792	1.000	0.833	0.275	0.312
code generation	0.775	1.000	0.755	0.487	0.244
feasibility study	0.774	1.000	0.817	0.412	0.058
semantic web	0.762	1.000	0.761	0.477	0.104
natural science	0.760	1.000	0.860	0.082	0.081
formal language	0.759	1.000	0.807	0.234	0.131
article publication	0.752	1.000	0.807	0.227	0.063
online open journal	0.749	1.000	0.801	0.228	0.058
research activity	0.744	1.000	0.732	0.155	0.361
summer school	0.744	1.000	0.689	0.618	0.067
social science	0.742	1.000	0.711	0.323	0.253
automatic code generation	0.736	1.000	0.697	0.399	0.180
software engineering	0.735	1.000	0.732	0.218	0.203
epistemological basis	0.728	1.000	0.672	0.534	0.058
cooperative article publication	0.725	1.000	0.750	0.201	0.054
community currency	0.724	1.000	0.749	0.127	0.117
distributed accountability	0.721	1.000	0.689	0.311	0.144
integrated project	0.715	1.000	0.709	0.135	0.176
dissemination material	0.711	1.000	0.684	0.307	0.072

Figure 13: *Results of TermExtractor applied to the Description of Work*

of the important terminological concepts in the text corpus. The extracted terms mirror the interdisciplinary character of the OPAALS community and its "knowledge", for example code generation, formal language, social science, community currency, natural science, cell metabolism, gene expression, and many others.

5.4 Case Study D: Metaphor and design of HCI

Research on metaphors does not only occur in respect to natural language, i.e. linguistics. It is widely used in a variety of areas such as organisational communication, human-computer interactions, software engineering and many others. In digital environments, the user-interface metaphors are

[12]Further information to the TermExtractor can be found at http://lcl2.uniroma1.it/termextractor/help/help.jsp.

very common, even if the computer scientists and engineers working on design and implementation of software, tools, and platforms are not really aware of using them. We can find such metaphors in almost all graphical user-interfaces. In this case study, we investigate metaphors from two perspectives: human-computer interface design and visualisation. The interface metaphor can be described as "the basic idea of likening between interactive objects and model objects of the application domain. Its role is to promote the best understanding of semantics of interaction, and also to determine the visual representation of dialog objects and a set of user manipulations with them" (V. Averbukh et al., 2008, p. 59).

V. Averbukh et al. (2008) define the visualization metaphor as "a map establishing the correspondence between concepts and objects of the application domain under modeling and a system of some similarities and analogies. This map generates a set of views and a set of methods for communication with visual objects" (p. 60).

In this case study, we analysed two examples of collaborative environments: Sironta and Guigoh. They use different development approaches, not only from an implementational perspective, but also from the point of view of design. And the design of both systems is of course influenced by metaphor usage.

The role and influence of metaphors become inevitable to take into account regarding the design of digital environments, due to the need to describe new phenomena (tools, software, environments), their structure and functionality in a familiar way for the user. The vocabulary used to describe and represent objects of such phenomena is often selected based on different kinds of similarity with the real world objects, structures or functions. Table 3 illustrates some examples of metaphors in digital environments.

Source	**Metaphor object**	**Target**
Container of a card-catalogue	File	File of data
Schematic diagram of an instrument or an electronic device	Block Diagram	diagram representing the modular structure of the program
A piece of furniture for organising documents	Desktop	Environment for organising and storing
Paper containing written information	Document	Digital representation of information

Table 3: *Interface metaphors in digital environments*

Our study of interface and visualisation metaphors in the context of the OPAALS project is based on three parts:

- An analysis of Guigoh and Sironta through the lense of metaphor
- An analysis of a visualisation system in OPAALS
- A collection of references to the subject of "Metaphors and Design of HCI"

5.4.1 Interface metaphor: Guigoh and Sironta

Within MTK we analyse user interfaces for human communication and collaboration in digital environments. Our investigations focus on the broader role of metaphor and to understand its influence on building collaborative platforms. Building user interfaces based on the metaphorical concepts is a common practice among engineers. Just consider the usage of concepts such as window, desktop, menu, trash in digital environments. Some authors (Heckel, 1991; Schön, 1993) argue that metaphor implies three aspects of user interface design: familiarisation, transportation, and invention.

Familiarisation processes reflect the idea of making products (software, tools, platforms) easier to understand. For example, Sironta uses a "room metaphor". This is a convenient way of introducing the notion of working environments supporting collaboration and communication, that usually take place in rooms. Another important aspect the "room" metaphor considers is the security issue. As Cikic, Jeschke, Lehmann-Grube, and Sablatnig (2008) mentioned, "a virtual room concept reduces information traffic and makes the necessary traffic safer in an intuitive way"(p.1). The aspect of familiarisation can be illustrated by chat programs in Guigoh and Sironta. They are based on the knowledge of functionality in real environments. The user expects particular capabilities that are usually similar to the ones in the physical world. The chat metaphor is closely related to the physical room: we can enter the chat like we enter a room, we might talk to the people that are in the room, and of course we can leave the room. In other words, the metaphors describe different kinds of information regarding digital environments: what the product/tool/software can do, the mechanics of how the tasks are accomplished (for example, the user needs to enter his or her email interface or room in order to write, send, and receive mails). There is a variety of research dealing with the design of user interfaces and environments based on the room metaphor (Greenberg & Roseman, 2003; Travers, 1989; Fitzpatrick, Kaplan, & Mansfield, 1996; Savidis & Stephanidis, 1995; Fahlén, Brown, Ståhl, & Carlsson, 1993; Herlea & Greenberg, 1998). One of

the core advantages is that such systems "help to enhance distributed work and sharing of underlying data as well as of derived results that may be sensitive"(Cikic et al., 2008, p. 2). Metaphors in both environments, Guigoh and Sironta involve specific objects that exist in the physical world as well: for example, files and documents. The source domain of them roots in our daily life and working experience (at home, at work). The users in digital environment apply this knowledge about the source domain and intuitively manipulate the corresponding objects in a similar way as they would do it in their physical environment: the users might move, group, split, send, receive, share, etc. the specific objects. In summary, metaphorical approaches to the design of HCI systems provide a convenient way of familiarity-based learning, in which users use their "real-world" knowledge about physical objects and functions while using a virtual system.

The role of ***transportation*** is the forming of a "personal experience framework" of metaphor usage and applying this knowledge to new platforms, user interfaces, and other kinds of virtual environments. One good example is the path metaphor that can be partially found in Guigoh. The path metaphor represents either the history of navigation (such as commonly used in web browsers through allowing the user go back to previous stage(s)) or the hierarchical structure of the environment by providing the path of the current directory (such as used in Windows Explorer). Since the users already learned to use the path metaphor, based on their experience with other programs, tools and platforms, it is a convenient way to introduce this metaphor allowing user-friendly navigation. By applying knowledge about the source domains, computer users reason that they can navigate through data the way they navigate through space: by searching, entering, leaving, exploring, backtracking, following, etc.

The ***invention*** aspect reflects the creativity of the designer coming up with new ideas. This is an important part of the design allowing users to see problems in a new way. However, we would suggest to be careful with the choice and usage of such metaphors. The risk is that the users might not understand the meaning of it (also taking into account the different cultural backgrounds of the OPAALS community or of social network communities, for example). Therefore, a careful research involving organisational analysis, discourse, communication and other aspects of "organisational life" needs to be done in advance.

In interaction processes, humans predominantly use conversation. Therefore, the conversation metaphor is one of most important forms in design. When we talk about the software for communication and collaboration, most people end up with the basic examples: chat, instant messaging, VoIP. But taking into consideration the complexity of human interaction (exchange

of information, information coupled with social consideration, variability in form and manner), we can conclude that metaphors for social computing are underdeveloped and lack in depth and creativity.

For identification and selection of appropriate metaphors for user-interface design, the categories introduced by Lakoff and Johnson (1980) can be useful. The authors distinguish three kinds of metaphors: orientational, ontological, and structural. These types of metaphors are present not only in human languages, but are also integrated in user-interfaces. Orientational metaphors are very important and useful components in user-interfaces due to their ability to facilitate the navigation through different parts of a system. The advantage of using orientational metaphors lies in the fact that they are closely related to our physical experience. Some of the widely used examples are "up", "down", "left", or "right" (often displayed as icons).

In both testing environments, we did not find a lot of orientational metaphors. One example in Sironta is the "exit room" - metaphor. It is a part of a more abstract room-metaphor introducing the notion of entering and leaving the space. In Guigoh, the orientational metaphor is partially presented in form of navigating through personal networks ("next" and "previous" signs). In DE, this type might be important not only for navigation purposes ("left", "right"), but also for the quantification, for instance, the associations with "more" and "less", or "up" and "down"(Figure 14)[13].

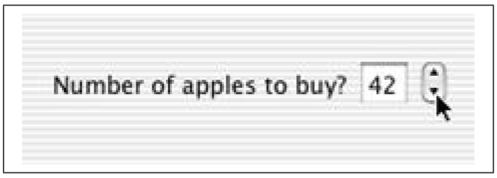


Figure 14: *Orientational metaphor for quantification*

"An ontological user-interface metaphor is one that identifies a system concept with a basic category of existence in the physical world, such as substance, object, container or entity" (Barr et al., 2005). The room-metaphor used in Sironta introduces several characteristics that are familiar from our physical world:

- The room is not only an object and space, it is also a container for other objects (e.g. documents and files).
- The restriction of access, i.e. entering the room.

[13]Example taken from (Barr, Khaled, Noble, & Biddle, 2005).

- The restriction of context, i.e. documents inside a room are separated from others
- Communication with users that are in the same room

The structural metaphors differ from ontological metaphors in fact that they are more specific in the choice of the target. Examples of structural metaphors that we found in both Guigoh and Sironta are: "THE COLLECTION OF DATA IS A PAGE" and "THE COLLECTION OF DATA IS A DOCUMENT". Documents are objects in our physical world. Since we can do several things with them in our real life, it would be convenient to do it in a virtual world as well. We can write a document, put it in a folder, mail it to others, throw it into the trashcan, etc. These are just some of the exemplary actions the user might expect in digital environments such Sironta and Guigoh. A strong metaphorical view/approach would not only help to identify important functions of a user interface, it would also help to determine malfunctioning mappings of actions.

Besides different kinds of metaphors and their role in HCI, supporting relationships among users in digital environments is an important design issue. In the context of the OPAALS-community, SMEs and DEs, two categories of relations, personal and working, must be considered (see Gabarro (1987) for more information on the categories). These categories are supported in both environments, Guigoh and Sironta. However, looking at these environments through a metaphor lense, we can identify a major influence in each of them. In Guigoh, the personal and social relationships come into foreground. Sironta, through implementing the room metaphor, brings the working relationships into foreground.

In DE, it is important to recognise the role of metaphors in developing user interfaces, software and other environments. Guigoh and Sironta have a good starting point as they support (at least to some extend) both kinds of user relationships: personal/social and working. However, in order to make the software more understandable, intuitive, and easy to use, it is necessary to consciously involve metaphorical concepts into the design and implementation processes.

5.4.2 Visualization metaphor

In the context of DE, we identify two areas of visualisation metaphor:

- visualisation of a system and its components
- visualisation of data

The sign denotes a unit of what is represented (the object), how it is represented (the representamen), and its possible interpretation(s). These elements are the ground of any sign system. In the context of visualisation, the study of the sign (semiotics) helps to understand different characteristics of the object or function one needs to represent. The triadic model introduced by Peirce (1931-1958) contains three components:

- The representamen - the form a sign takes
- The object to which this sign refers
- The interpretant depicting possible interpretations

Semiotic analysis guides the development of appropriate signs in user interface design (e.g. icons representing objects, system components or even functions). Furthermore, it helps to identify the ambiguity of interpretations. Figure 15 illustrates two possible interpretations in HCI based on the Peircean triadic model[14]. The same sign can be interpreted differently. In

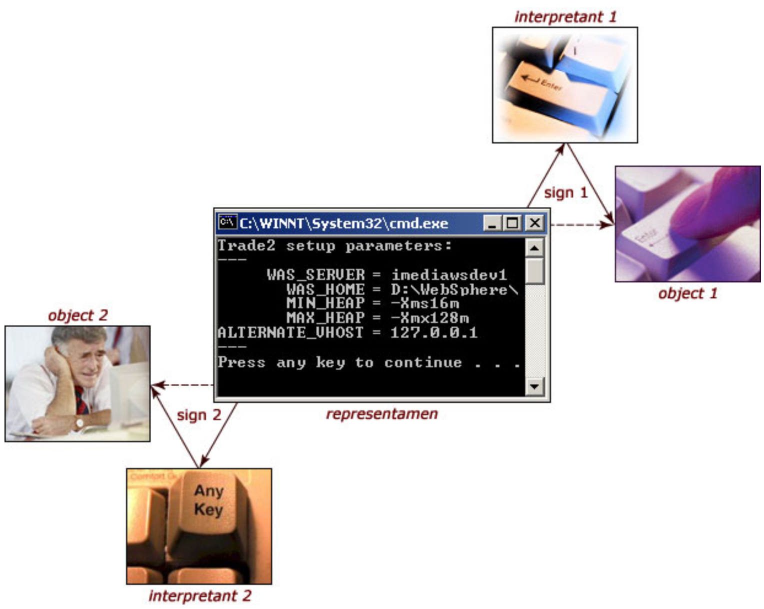


Figure 15: *Semiotics in HCI*

one case the user might think of the "Enter"-key. It is the interpretation of this sign (interpretant). The object is the "Enter"-key the user has to press. On the other hand, the same sign can cause a situation where the user might

[14]Picture taken from http://www-cse.ucsd.edu/ goguen/courses/271/tutorial/peirce.php.

think of the "Any" key. This interpretation leads to the confusion because he or she looks for the key called "Any" (object) on the keyboard.

Looking through the lense of semiotic analysis, there are three types of representation regarding visualisation metaphors[15]:

- The iconic type is defined by the relationship of resemblance.
- The indexical approach can be defined as a way of representing information relating to a specific context (e.g. spatial, temporal)
- The symbolic type refers to the representation of an arbitrary object or function

In HCI environments, icons are one of the common signs to represent an object or a function. They can be iconic, indexical, or symbolic depending on their representation role and on their perception. Figure 16 illustrates icons used in Sironta.

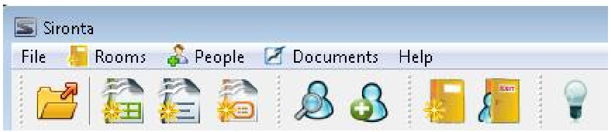


Figure 16: *Icons in Sironta-environment*

Depending on the user's prior knowledge and experience, the understanding of icons may vary. For example, the "lamp"-icon on the right side of the figure might require some further explanation for some of the users. A similar example can be found in Guigoh environment (Figure 17), where the conference-icon is not fully self-explanatory.

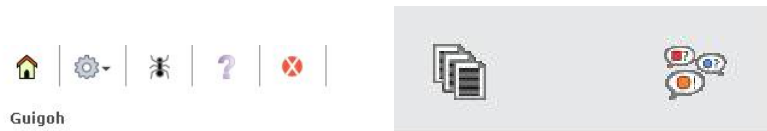


Figure 17: *Icons in Guigoh-environment*

Semiotic research of visualisation systems provides insights into strategies of intuitive image mapping. Consider an example of "Trash" (Figure 18). In digital environment, this object refers to a place where the user can dispose things he or she does not need. This object is represented iconically: it looks like a trashbin one has in the office or at home. Additionally, it is represented

[15] Three types of signs in semiotics (Peirce, 1931-1958).

Figure 18: *Object "Trash"*

symbolically due to the actions(object's functions) the user can apply to it (e.g., delete file = throw a document away; restore file = get the document out from the trashbin).

In conclusion, we would identify three basic characteristics of icons that need to be taken into account while designing a HCI design:

- Icons should be recognisable
- Icons should be memorable
- Icons should be discriminable

In the context of data visualisation, the correlation between two aspects plays a significant role and often requires a "metaphorical point of view": the choice of a visualisation type (graph, network, chart, etc.) and the pattern(s) or specific characteristics of data we want to visualise. This mapping between the information that has to be illustrated to the user and signs used for its representation is a metaphor in its nature. The choice of a "good" visualisation metaphor strongly depends on the goal of a system. If, for instance, a statistical comparison of data is a primary goal of a visualisation system, the diagram can be one of the choices of illustration (Figure 19)[16].

Figure 20 provides a different example of visualisation where the focus lies in the context (e.g. what pages have been edited by a selected person) and not in the statistical information.

In DE, the visualisation system is a type of communication between users and interfaces. This interaction can be improved through dynamic visualisation metaphors. For instance, the visualisation in Figure 20 could be modified by the user through the following actions: change the focus of attention (choose one website and display all users who edited it). Working with visualisation of textual data (keywords, most common terms, collocates of a given pattern, and others) can be implemented dynamically as well. The

[16]Examples are taken from the Wille visualisation environment developed by the Tampere team.

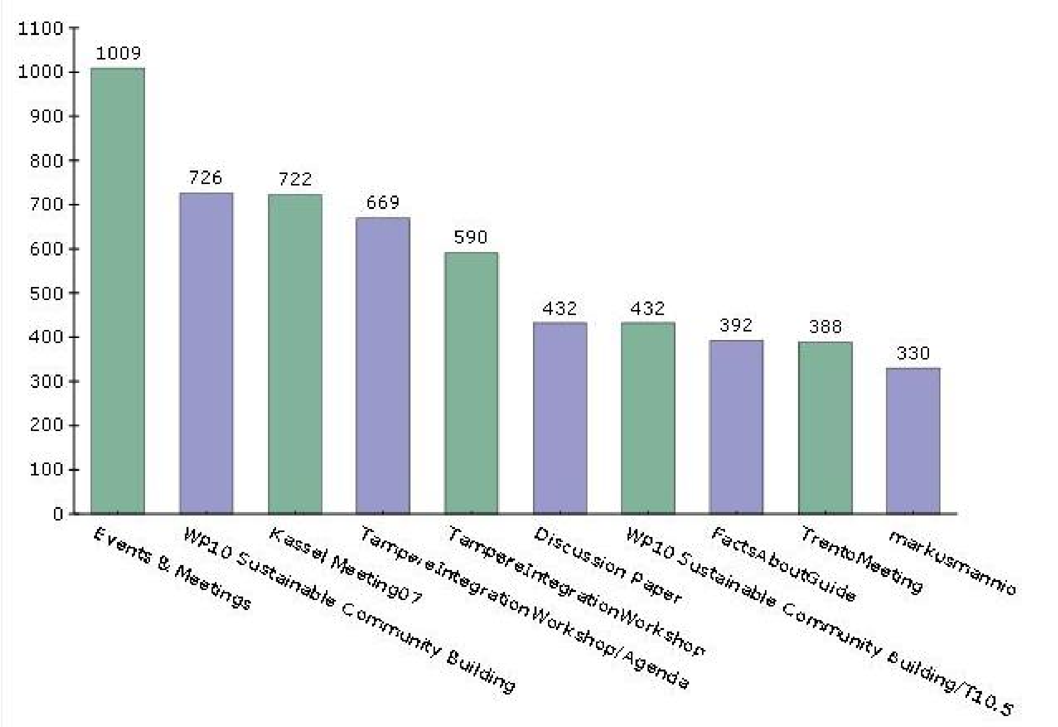


Figure 19: *Visualisation of statistical information*

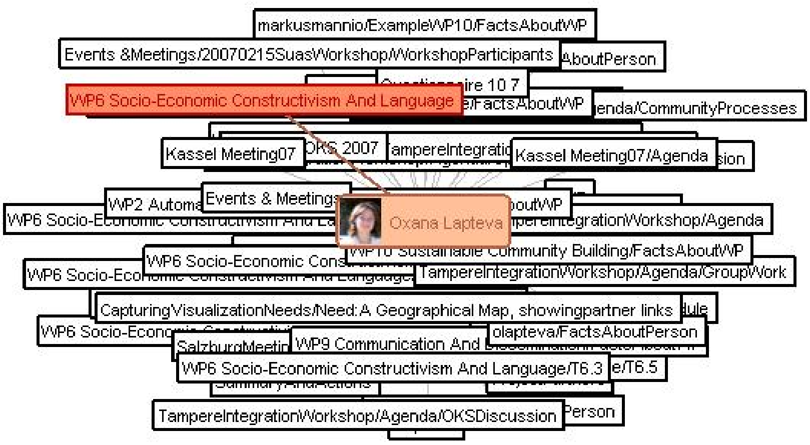
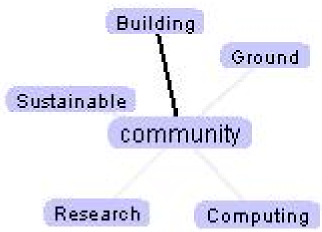


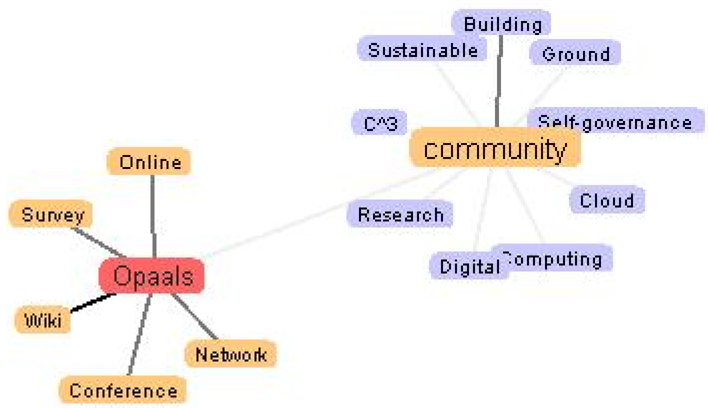
Figure 20: *Visualisation of context*

user might be able to navigate through the environment by selecting an item, or the possibility to change the number of visualised items. For example, the visualisation of collocates for a given pattern (e.g. "community") displays five items by default (Figure 21(a)). The user can increment this number when he or she double-clicks on the pattern (Figure 21(b)). Additionally, the user might expand the network of collocates by selecting one of them. Then the collocation analysis will be applied to the selected item, for example, the word "opaals" (Figure 21(c)).



(a) Default display

(b) After increasing the number of collocates



(c) Additional collocates for the word "opaals"

Figure 21: *OPAALS newsletter visualisation*

When working with metaphors in the context of HCI design and visualisation, it is important to provide a system or taxonomy for the process of searching, selection, design and implementation. It is not enough (and would be a poorly strategy) to only choose a couple of metaphors and implement them without a sound design strategy. In case of implementing an interface design strategy based on metaphorical concepts, the hierarchical approach (from an abstract, general idea to the specific objects and functions) seems to be reasonable. V. Averbukh et al. (2008) suggest to use the following hierarchy (p.72):

- A global metaphor that represents the general idea of an application (e.g. the world as an office)
- A basic metaphor of an application (e.g. desktop metaphor, a landscape, a city, a room)
- A local metaphor for specialisation of an application (e.g. a card file metaphor)
- A metaphor for widgets and separate operations

In visualisation systems we might consider the role of the sign not only in the process of transmitting information, but also for intuitive dynamic interaction between the user and the digital environment. Furthermore, semiotic analysis of signs used in visualisation provides more detailed information about their "meaning" and function. Finally, a systematic approach of applying metaphors to the user interface design and visualisation helps the facilitating processes for learning, orientation, and understanding. The role of semiotics in HCI and DE is twofold: first, it can provide useful insights about user interfaces (what they represent, what they refer to, and how they are interpreted), and second, it can be used as a platform for the analysis of a system design in a variety of real contexts. The semiotic approach can be applied not only for providing input and guidelines of the design, but also for creating models for design optimisation of user interfaces.

6 Conclusion

The MTK described in this deliverable aims to provide a platform for further research within and outside the OPAALS community, and at the same time functioning as an important "outreaching" means for OPAALS' research into the wider DE field (i.e. regarding future interactions with SMEs). In this deliverable, we identified several directions of investigation. It can be used for social science research on communities, organisational communication and discourse. Furthermore, analysis of textual data would enrich the qualitative and quantitative research.

The computer science domain can also profit from this research by applying to their work some of the ideas presented here. For example, the design of user interfaces and visualisation systems, the analysis, extraction, and visualisation of textual data, natural and formal languages and their existence in DE, language change and variation[17].

Finally, by means of the MTK we intend to support and deepen the collaboration between different disciplines. The next steps of our work in the context of MTK include:

- MTK web interface for accessing data, methods, tools, and projects/case studies.
- Continuation of research on dependency grammar and formal languages. This approach needs to be tested on a variety of sentences in order to improve the quality of the developed algorithm for SBVR. Additionally, we are interested in investigations of the dependency parser for the extraction of ontological relations in texts.
- Metaphors and HCI: the metaphorical approach to the design of user interfaces and visualisation systems opens a new direction of collaboration between social science and computer science domains.

[17] Task 6.6.

References

Averbukh, V., Bakhterev, M., Baydalin, A., Gorbashevskiy, D. Y., Ismagilov, D., Kazantsev, A., et al. (2008). Searching and analysis of interface and visualization metaphors. In K. Asai (Ed.), *Human-computer interaction: New developments* (p. 49-84). Vienna, Austria: InTech Education and Publishing.

Averbukh, V. L. (2001). Visualization metaphors. *Programming and Computer Software*, *27*(5), 227-237.

Barr, P., Khaled, R., Noble, J., & Biddle, R. (2005). A taxonomic analysis of user-interface metaphors in the microsoft office project gallery. In *Auic '05: Proceedings of the sixth australasian conference on user interface* (p. 109-117). Darlinghurst, Australia, Australia: Australian Computer Society, Inc.

Braeuer, M., Crone, A., Dürrenberg, C., Lapteva, O., & Zeller, F. (2008). Appropriateness of communication and collaboration tools in an international virtual research community. In O. Nykänen, J. Huhtamäki, J. Salonen, S. Pohjolainen, & K. Silius (Eds.), *Proceedings of the 2nd international opaals conference on digital ecosystems.* Tampere, Finland: OPAALS 2008.

Braeuer, M., Crone, A., Dürrenberg, C., & Zeller, F. (2008). Governance in the opaals community. In O. Nykänen, J. Huhtamäki, J. Salonen, S. Pohjolainen, & K. Silius (Eds.), *Proceedings of the 2nd international opaals conference on digital ecosystems.* Tampere, Finland: OPAALS 2008.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Carroll, J. (2003). Parsing. In R. Mitkov (Ed.), *The oxford handbook of computational linguistics* (p. 233-248). Oxford: Oxford University Press.

Carroll, J., Mack, R. L., & Kellogg, W. A. (1988). Interface metaphors and user interface design. In M. Helander (Ed.), *Handbook of human-computer interaction* (p. 67-81). Amsterdam: Elsevier Science Publishers B.V.

Church, K., & Hanks, P. (1990). Word association norms: mutual information and lexicography. *Computational Linguistics*, *16*(1), 22-29.

Cikic, S., Jeschke, S., Lehmann-Grube, F., & Sablatnig, J. (2008). Using a room metaphor for e-forensic working environments. In *Proceedings of the 1st international conference on forensic applications and techniques in telecommunications, information, and multimedia and workshop.* Brussels, Belgium: ICST.

Fahlén, L. E., Brown, C. G., Ståhl, O., & Carlsson, C. (1993). A space based model for user interaction in shared synthetic environments. In *Chi '93: Proceedings of the interact '93 and chi '93 conference on human factors in computing systems* (p. 43-48). New York, NY, USA: ACM.

Fitzpatrick, G., Kaplan, S., & Mansfield, T. (1996). Physical spaces, virtual places and social worlds: a study of work in the virtual. In *Cscw '96: Proceedings of the 1996 acm conference on computer supported cooperative work* (p. 334-343). New York, NY, USA: ACM.

Fundel, K., Küffner, R., & Zimmer, R. (2007). Relex - relation extraction using dependency parse trees. *Bioinformatics*, *23*(3), 365-371.

Gabarro, J. J. (1987). The development of working relationships. In J. W. Lorsch (Ed.), *Handbook of organizational behavior* (p. 172-189). Englewood Cliffs, NJ: Prentice-Hall.

Gilmour, L., Collins, C., Cross, R., Mackie, W., Holmes, A., & Weber, P. (Eds.). (2002). *Collins paperback dictionary and thesaurus.* Glasgow, UK: HarperCollins Publishers.

Greenberg, S., & Roseman, M. (2003). Using a room metaphor to ease transitions in groupware. In M. Ackermann, V. Pipek, & V. Wulff (Eds.), *Beyond knowledge management: Sharing expertise* (p. 203-248). Cambridge, MA: MIT Press.

Heckel, P. (1991). *The elements of friendly software design.* San Francisco, CA: Sybex.

Herlea, D. E., & Greenberg, S. (1998). Using a groupware space for distributed requirements engineering. In *Wetice '98: Proceedings of the 7th workshop on enabling technologies* (p. 57-62). Washington, DC, USA: IEEE Computer Society.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: an introduction to natural language processing, computational linguistics and speech recognition.* New Jersey: Prentice Hall PTR.

Knowles, M., & Moon, R. (2006). *Introducing metaphor.* New York: Routledge Taylor and Francis.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago, USA: University of Chicago Press.

Lapteva, O., & Peukert, H. (2007). *Linguistic dimension of terms: towards community's common vocabulary and concepts.* Paper presented at OPAALS Conference, Rom 2007.

Lapteva, O., & Zeller, F. (2008). *Digital ecosystem: How metaphors become real environments.* Paper presented at Seventh International Conference on Researching and Applying Metaphor (RaAM7).

Manning, C. D., & Schütze, H. (2003). *Foundations of statistical natural language processing.* Cambridge, MA: The MIT Press.

Marneffe, M.-C. de, MacCartney, B., & Manning, C. D. (2006). *Generating typed dependency parses from phrase structure parses.*

Mason, O. (2000). *Programming for corpus linguistics: how to do text analysis with java*. Edinburgh: Edinburgh University Press.

Matsumoto, Y. (2003). Lexical knowledge acquisition. In R. Mitkov (Ed.), *The oxford handbook of computational linguistics* (p. 395-413). Oxford: Oxford University Press.

Morgan, G. (2006). *Images of organization.* Thousand Oaks, CA: Sage Publications.

Nadin, M. (1997). *The civilization of illiteracy.* Dresden: Dresden University Press.

Peirce, C. S. (1931-1958). *Collected papers of c.s.peirce.* Cambridge, MA: Harvard University Press.

Peukert, H. (2009). *Discourse as unlimited semiosis in digital ecosystems.* Accepted for publication in European Journal for Semiotic Studies.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-computer interaction.* Wokingham, UK: Addison-Wesley Publishing Company.

Risch, J. S. (2008). *On the role of metaphor in information visualization.* Available from `http://arxiv.org/abs/0809.0884v1`

Savidis, A., & Stephanidis, C. (1995). Building non-visual interaction through the development of the rooms metaphor. In *Chi '95: Conference companion on human factors in computing systems* (p. 244-245). New York, NY, USA: ACM.

Schön, D. A. (1993). *Generative metaphor: a perspective on problem-setting in social policy* (A. Ortony, Ed.). New York, NY: Cambridge University Press.

Souza, D. C. (2005). *The semiotic engineering of human-computer interaction.* MIT Press.

Travers, M. (1989). A visual representation for knowledge structures. In *2nd acm conference on hypertext* (pp. 147–158). Pittsburgh, PA: ACM.

Ware, C. (2004). *Information visualization: Perception for design.* Morgan Kaufmann.

Winograd, T., Bennett, J., Young, L. D., & Hartfield, B. (Eds.). (1996). *Bringing design to software.* Addison Wesley.

Zhang, J. (2008). *Visualization for information retrieval.* Springer.

Appendix A. Significance of collocates

A.1 T-score

This method works with the mean and variance of measurements. It "looks at the difference between the observed and expected means, scaled by the variance of the data" (Manning & Schütze, 2003, p. 163). Mathematically, the t statistic is defined as

$$t = \frac{\bar{\chi} - \mu}{\sqrt{(\frac{s^2}{N})}}$$

where $\bar{\chi}$ is the sample mean, s^2 is the sample variance, N is the sample size, and μ is the mean of the distribution (Manning & Schütze, 2003).

A.2 Mutual information

This method of evaluating the significance of collocates (measure of co-occurrence) was introduced by Church and Hanks (1990). It is defined as

$$MI(x, y) = log\frac{P(x, y)}{P(x)P(y)} = log\frac{\frac{freq(x,y)}{N}}{\frac{freq(x)}{N}\frac{freq(y)}{N}} = log\frac{freq(x, y)N}{freq(x)freq(y)}$$

where $P(x, y)$ is the co-occurrence probability and $freq(x, y)$ - the co-occurrence frequency of two events x and y, $P(x)$ and $P(y)$ are independent probabilities for each event x and y, $freq(x)$ is an independent frequency of x and N is a total number of events (Matsumoto, 2003).

Appendix B. English stopwords

A
a, about, above, across, after, afterwards, again, against, all, almost, alone, along, already, also, although, always, am, among, amongst, amoungst, amount, an, and, another, any, anyhow, anyone, anything, anyway, anywhere, are, around, as, at
B
back, be, became, because, become, becomes, becoming, been, before, beforehand, behind, being, below, beside, besides, between, beyond, bill, both, bottom, but, by
C
call, can, cannot, cant, co, computer, con, could, couldnt, cry
D
de, describe, detail, do, done, down, due, during
E
each, eg, eight, either, eleven, else, elsewhere, empty, enough, etc, even, ever, every, everyone, everything, everywhere, except
F
few, fifteen, fify, fill, find, fire, first, five, for, former, formerly, forty, found, four, from, front, full, further
G
get, give, go
H
had, has, hasnt, have, he, hence, her, here, hereafter, hereby, herein, hereupon, hers, herself, him, himself, his, how, however, hundred
I
i, ie, if, in, inc, indeed, interest, into, is, it, its, itself
K
keep
L
last, latter, latterly, least, less, ltd
M
made, many, may, me, meanwhile, might, mill, mine, more, moreover, most, mostly, move, much, must, my, myself
N
name, namely, neither, never, nevertheless, next, nine, no, nobody, none, noone, nor, not, nothing, now, nowhere
O
of, off, often, on, once, one, only, onto, or, other, others, otherwise, our, ours, ourselves, out, over, own

P
part, per, perhaps, please, put, rather, re
S
same, see, seem, seemed, seeming, seems, serious, several, she, should, show, side, since, sincere, six, sixty, so, some, somehow, someone, something, sometime, sometimes, somewhere, still, such, system
T
take, ten, than, that, the, their, them, themselves, then, thence, there, thereafter, thereby, therefore, therein, thereupon, these, they, thick, thin, third, this, those, though, three, through, throughout, thru, thus, to, together, too, top, toward, towards, twelve, twenty, two
U
un, under, until, up, upon, us
V
very, via
W
was, we, well, were, what, whatever, when, whence, whenever, where, whereafter, whereas, whereby, wherein, whereupon, wherever, whether, which, while, whither, who, whoever, whole, whom, whose, why, will, with, within, without, would
Y
yet, you, your, yours, yourself, yourselves

Appendix C. Dependency relations

The list of dependencies existing in the Stanford Parser (see Marneffe, MacCartney, and Manning (2006) for more details):
dep - dependent
aux - auxiliary
auxpass - passive auxiliary
cop - copula
arg - argument
agent - agent
comp - complement
acomp - adjectival complement
attr - attributive
ccomp - clausal complement with internal subject
xcomp - clausal complement with external subject
compl - complementizer
obj - object
dobj - direct object
iobj - indirect object
pobj - object of preposition
mark - marker (word introducing an advcl)
rel - relative (word introducing a rcmod)
subj - subject
nsubj - nominal subject
nsubjpass - passive nominal subject
csubj - clausal subject
csubjpass - passive clausal subject
cc - coordination
conj - conjunct
expl - expletive (expletive "there")
mod - modifier
abbrev - abbreviation modifier
amod - adjectival modifier
appos - appositional modifier
advcl - adverbial clause modifier
purpcl - purpose clause modifier
det - determiner
predet - predeterminer
preconj - preconjunct
infmod - infinitival modifier
partmod - participial modifier

advmod - adverbial modifier
neg - negation modifier
rcmod - relative clause modifier
quantmod - quantifier modifier
tmod - temporal modifier
measure - measure-phrase modifier
nn - noun compound modifier
num - numeric modifier
number - element of compound number
prep - prepositional modifier
poss - possession modifier
possessive - possessive modifier ('s)
prt - phrasal verb particle
parataxis - parataxis
punct - punctuation
ref - referent
sdep - semantic dependent
xsubj - controlling subject

Appendix D. Tools

List of the text analysis tools used in the MTK:

- **Tokenizer** identifies the basic units from the sequence of characters (word, sentence, paragraph). Due to the specific needs in business environments, UniKassel team developed a tokenizer (see section 4.2.1). This program will be shortly available to the consortium through the web interface of the MTK.

- **Filtering** is the process of reduction the complexity of textual data. It is developed by UniKassel team and will be available to the consortium through the web interface of the MTK.

- **Concordance** is a part of the TAPoRware tools developed at the McMaster University with support from the Canada Foundation for Innovation. Free acces through the web interface of TAPoR is provided: http://taporware.mcmaster.ca/ taporware.

- **Collocation** is a part of the TAPoRware tools developed at the McMaster University with support from the Canada Foundation for Innovation. Free acces is available through the web interface of TAPoR.

- **Parsing**. A natural language parser (constituency and dependency grammar) is developed by the Stanford Natural Language Processing Group. Free download is available at http://nlp.stanford.edu/software/lex-parser.shtml.

- **Extracting of concepts and relations** (for SBVR and ontologies). This is a program based on the Stanford parser that extracts relevant concepts and relations between them based on the grammatical structure of a sentence. It is developed by UniKassel and will be available to the consortium through the web interface of the MTK.

- **C-Value** is a method within the TerMine tool developed at the National Centre for Text Mining (NaCTeM). The web demonstration system (free access) is available at NaCTeM web site: http://www.nactem.ac.uk/software/termine/. Unrestricted access to the NaCTeM services has to be requested by the developers.

- **TermExtractor** is a free software package for Terminology Extraction developed by the Linguistic Computing Laboratory (LCL) at the University of Roma. Full version allows one to upload a corpus of documents for a maximum of 100 MB and set the options of the terminology

extraction process. Free access is available through the LCL web site: http://lcl2.uniroma1.it/termextractor/. In order to use the full version of TermExtractor one must register (both registration and full version are free).