OPAALS PROJECT

Contract n° IST-034824

WP6: Socio-Constructivism and Language

D6.11 - View to the evolution of the OKS from 50,000 feet

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

Contract Number: IST-034824

Project Acronym: OPAALS

Deliverable N°: D6.11

Due date: May 2010

Delivery Date: May 2010

Short Description: This deliverable presents and discusses a visualisation application that provides a bird's eye view to the selected OPAALS OKS data obtained during the project.

Author: Ossi Nykänen (TUT), Frauke Zeller (UniKassel)

Partners contributed: TUT, UniKassel

Made available to: All partners

Versioning		
Version	**Date**	**Name, organization**
0.7	01/05/2010	Ossi Nykänen, TUT
0.8	17/05/2010	Ossi Nykänen, TUT & Frauke Zeller, UniKassel
0.9	24/05/2010	Frauke Zeller, UniKassel
1.0	25/05/2010	Ossi Nykänen, TUT

Quality check

Internal Reviewers: Antonella Passani, T6-Eco; Thomas Kurz, SUAS

Dependencies:

Achievements*	A novel visualisation application that demonstrates information gathering and analysis tasks in the context of an evolutionary knowledge space.
Work Packages	This work is based on the evolution research from WP6 and the technical visualisation system developed in WP10.
Partners	TUT, UniKassel
Domains	N/A
Targets	All partners
Publications*	An extended version of the WP6 results will be submitted as a journal article by TUT and UniKassel.
PhD Students*	N/A
Outstanding features*	N/A
Disciplinary domains of authors*	Computer Science, Linguistics, Social Science

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

Executive Summary

The main contribution of this deliverable lies in the concrete visualisation application and using it for raising issues considering the evolutionary components in the underlying processes and data of the Opaals OKS. At the same time, this work also serves as a foundation in future research and publications, in order to identify and synthesize common characteristics of evolutionary knowledge spaces, in order to propose a general analysis and validation criteria of taking evolution explicitly into account in knowledge space design and control.

More concretely, the visualisation application provides a tangible view to the OKS, providing a specific, concrete use case and a reference framework for discussing the Opaals OKS and collaborative knowledge spaces (online) in general.

Considering the abstract scope of this deliverable, a view to the evolution of the OKS from 50,000 feet, it is sufficient to say that the underlying aspects through which evolution is demonstrated include ***processes***, ***data***, and ***language***. This three-fold approach represents a novel interdisciplinary way to address the complex nature of knowledge evolution in open knowledge spaces. Whereas the three aspects are interdependently connected, language can be seen here as a means to analyse in depth first results/insights from the visualisation. This is also based on the notion that knowledge is processed information, and that the collaborative act of processing information is always based on or carried out by means of language/communication. Moreover, the language stance also allows to take into consideration the social dimension of knowledge, i.e. integrating relevant facts of the community that inhabits the knowledge space, for example different disciplinary backgrounds.

Finally, the visualisations and data processing tools can also help to identify and visualise the sensemaking processes as well as failures in sensemaking processes. On the other hand, the visualisation results can be enriched significantly in terms of its descriptive nature. By drawing upon prior works in the field of language evolution, the visualisations can also be used for normative results regarding for example recommendations for visualisation processes and techniques in general and in the field of knowledge processes visualisation.

Table of Contents

1. Introduction

In this deliverable we present a visualisation application that depicts the OPAALS OKS "from 50,000 feet".

The deliverable includes two parts: 1) This deliverable document, and 2) the Java-based software application that is the actual implementation of the visualisation application. (On-line documentation of the underlying visualisation system and the related software components are available at http://wiki.tut.fi/Wille.)

The visualisation application provides a tangible view to the OKS, providing a specific, concrete use case and a reference framework for discussing the Opaals OKS and collaborative knowledge spaces (online) in general. It also illustrates several challenges in knowledge space analysis from an evolution analysis point of view.

The motivation for this work is largely due to our previous analysis considering the OKS evolution (please see the references at the end of this document). In particular, by concretising the OKS with respect to the selected case of a given data set, the discussion of evolution becomes more tangible. Indeed, from a modelling point of view, the quantitative and the qualitative aspects of talking about knowledge and data become apparent. In brief, while qualitative discussion is needed to understand an OKS and its evolution, in a large distributed knowledge space the quantitative aspect pretty much underlines the analysis. Note that for purposes of this article, we simply define OPAALS OKS as a system of accessible formal and informal information contributed during the OPAALS project. This means that our main interests lies in the evolution of knowledge (carried by information), and not in the concrete development stages of the software itself. This means that we are taking a mainly socio-determined perspective (as in SCOT – the Social Construction of Technology, see e.g. Bijker, Hughes and Pinch, 1987), whereas the technological determinism certainly is also present regarding for example the actual content of the analysed data, and of course the visualisation processing techniques *per se*.

In this setting, we find adopting the basic terminology from statistics helpful: Since consumers of knowledge almost never have access to all data, they must try to draw appropriate conclusions from the impartial data. As a consequence, we may nicely conceptualise this activity as generalising the conclusions drawn using the available "knowledge sample" to the as a whole inaccessible "knowledge population". Using this simple framework, we may promptly point out where acknowledging evolution is needed, when drawing conclusions from the knowledge sample. As suspected, the visualisation application provides an insightful illustration of this line of thinking.

The main contribution of this deliverable lies in the concrete application and using it for raising issues considering the evolutionary components in the underlying processes and data. We plan to use this work as a foundation in future research and publications, in order to identify and synthesize common characteristics of evolutionary knowledge spaces, in order to propose a general analysis and validation criteria of taking evolution explicitly into account in knowledge space design and control. Given that the field of online knowledge communities – either in academia or commercial areas – is becoming increasingly relevant, this work will also be interesting for research and development projects outside OPAALS.

The work is based on previous OPAALS results, namely efforts within the domains of the WP6 and WP10. In brief, this work includes investigating the epistemological, metaphorical and ontological foundations of knowledge and establishing the foundation of visualisation systems (see e.g. Bräuer, Steinicke & Zeller, 2009); Lapteva & Zeller, 2008; Lapteva & Peukert, 2008; Lapteva, Peukert & Zeller, 2009; Nykänen, Mannio, Huhtamäki & Salonen 2007; Nykänen, Salonen, Haapaniemi, &

Huhtamäki 2008; Zeller, Lapteva & Crone, 2007). In addition, the data foundation perspective has been investigated in terms of specific modelling languages, in order to establish a concrete computer-supported framework for understanding large volumes of data (Nykänen 2009a; Nykänen 2009b).

In general, our work falls into the domain of data mining and visualisation. This allows us to build upon the well-known concepts such as data pre-processing and normalisation, visual mapping, and interactive exploration patterns (see Ware, 2004; Telea 2008; Geroimenko & Chen, 2006).

The deliverable is organised as follows: In Section 2, we consider the data foundation of the visualisation application that is then presented in terms of a simple visual analysis case study in Section 3. Section 4 outlines the factor of analysing the observed structures and Section 5 finally concludes the article with discussion and further remarks.

2. Source Data

According to its definition, the OPAALS OKS provides the socio-technical environment that will be used to create, manage and make knowledge available for the OPAALS community. For purposes of this study, we shall next identify a subset of the OKS data, focusing the data that is available consortium-wide.

The architecture of the visualisation application is based on a component-based pipeline processing architecture (Nykänen, Salonen, Haapaniemi, & Huhtamäki 2008). Because of the complexity of the data sources the visualisation application is effectively implemented as a pull data processing pipeline. In brief, the data is requested from several sources, pre-processed, suitably transformed, and finally normalised into a single database representation suitable for interactive viewing.

In an ideal setting this process would be fully automated and the visualisation tools indeed support this. However, in practice, this study includes data that effectively required manual processing, due to inconsistencies in archive locations, data structures, and/or semantics. As a consequence, some manual efforts were required to compile and clean the data.

In principle, automating this data processing is of course possible. However, this essentially requires that all source documents are strictly machine readable with agreed semantics. For purposes of the particular OKS from 50,000 feet application, data was after pre-processing organised into tabular form, including the fields type, date, title, author, link, and pageData.

In brief, the data included in the visualisation application includes community wiki pages, community-wide emails (the "Opaals-All" mailing list), reported project deliverables, and reported publications. However, because of the limited availability and the practical challenge in retrieving the data, all conceivable project data could not be included. In particular, the data includes final versions of wiki articles between June 2006 and March 2010, emails between June 2006 and March 2010, deliverables between 2006 and 2009, and finally publications between 2006 and 2009.

The data is by no means complete. While the wiki data demonstrated the most complete data entity, it suffers from a technical change of platform in mid 2009. Further, emails do not include e.g. private messages between researchers. Publication and deliverable data were compiled and cleaned by hand and include only partial metadata, i.e. lacking the actual data content. Finally, for practical or privacy purposes, the data do not include e.g. personal or group-specific data, versioned file repositories, or study-specific experimental data at all. As suspected, informal knowledge was not accessible at all, including e.g. face-to-face and group discussions.

We believe that in a complex knowledge space environment, this setting is quite typical: While the enabling technological framework in principle provides a common collaboration medium and repository for all data, such as the Web or a Peer-to-peer infrastructure, actual tools, data instances, and semantics are fragmented or poorly accessible as a whole. A "better" data organisation would require much stricter rules and policies for managing data, something many research communities might not be ready for. Moreover, collaborative knowledge creation that spans different disciplines, needs to be rather open instead of restricting it with a overly meticulous set of rules and policies. This notion takes into account the different disciplinary cultures and modes of knowledge production, that can vary significantly among the participating communities/disciplines. Hence, a pre-defined strict set of rules and policies could easily hinder certain communities to participate in the knowledge creation, which would in turn jeopardise the whole collaborative knowledge creation endeavor, hence the development of a successful and sustainable OKS.

In turn, the down-to-earth approach also highlights the several meanings to the term "Knowledge Space". While a knowledge space may intuitively include all conceivable (explicit or even tacit) data, practical applications have reasonable access only to a fraction of the externalised data.

It turns out that it is very useful to characterise the relationship of the data e.g. accessible in a particular visualisation application and the potentially available "full" knowledge space in terms of statistics. In this terminology, the accessible subset of all data is called a *knowledge sample*, while the all conceivable data denotes the *knowledge population*. In brief, we usually study the sample in order to understand the population, and to draw general conclusions.

However, without fixing representations, it is of course very difficult to define the "statistical unit" for knowledge. In the case of the visualisation application, the approach was very pragmatic: (Knowledge) sample data was *itemised* so that each knowledge item (or unit) was represented as an array of semantic properties.

Since we focus on event or publication kind of data in our case study, knowledge items are essentially reduced to resources described with metadata. In brief, the deliverables and publications included only item metadata. Wiki articles included links to real wiki content, and emails included a snapshot of the text body of the messages (as pageData).

After processing, our sample data includes total 1598 items with 13 properties. When represented as an uncompressed file (including markup), this comprises about 6.0 megabytes of textual data. (A default zip compression reduces the size to roughly 1.6 megabytes.)

In addition to the actual sample data, visual mappings and decorations are also needed to support visualisation. In our case this requires specifying a visual 2D visualisation model on which the items are projected. The item projection was kept as simple as possible: The X coordinate was computed by projecting items on a timeline, and the Y coordinate was computed by sorting items alphanumerically by title, starting from the top. The items were decorated by a simple calendar grid, with a specific slot for items with unknown dates.

In general, the applicable projection methods obviously strongly depend on the visualisation metaphor and the available item properties. Also, visualising e.g. taxonomy or rule-based knowledge would require introducing a more complex visualisation model and an associated intuitive metaphor. In applications where sensory representations or maximal affordance are sought after, the metaphors should be intuitive and familiar to the end-users.

3. Visualisation Application

Once the sample data is properly itemised and associated with decorative graphics, it can be explored in an interactive, general-purpose visualisation application called the *Databox View*. (The underlying Databox view application and the required data processing systems have been developed and reported in WP10 in Deliverable 10.18.)

In brief, we may identify two main user roles for the Databox View application. End-users work with data pre-processed by others. End-users may use the application to simply navigate within data, or to analyse data-driven or visual hypotheses using the search and filtering tools. The second main user group is developers who select and pre-process data so that it can be depicted using the Databox View application.

Figure 1 presents a global view to the sample data. In brief, items are associated with coloured icons and text labels on a 2D plane. Other item properties may be examined by viewing item information.

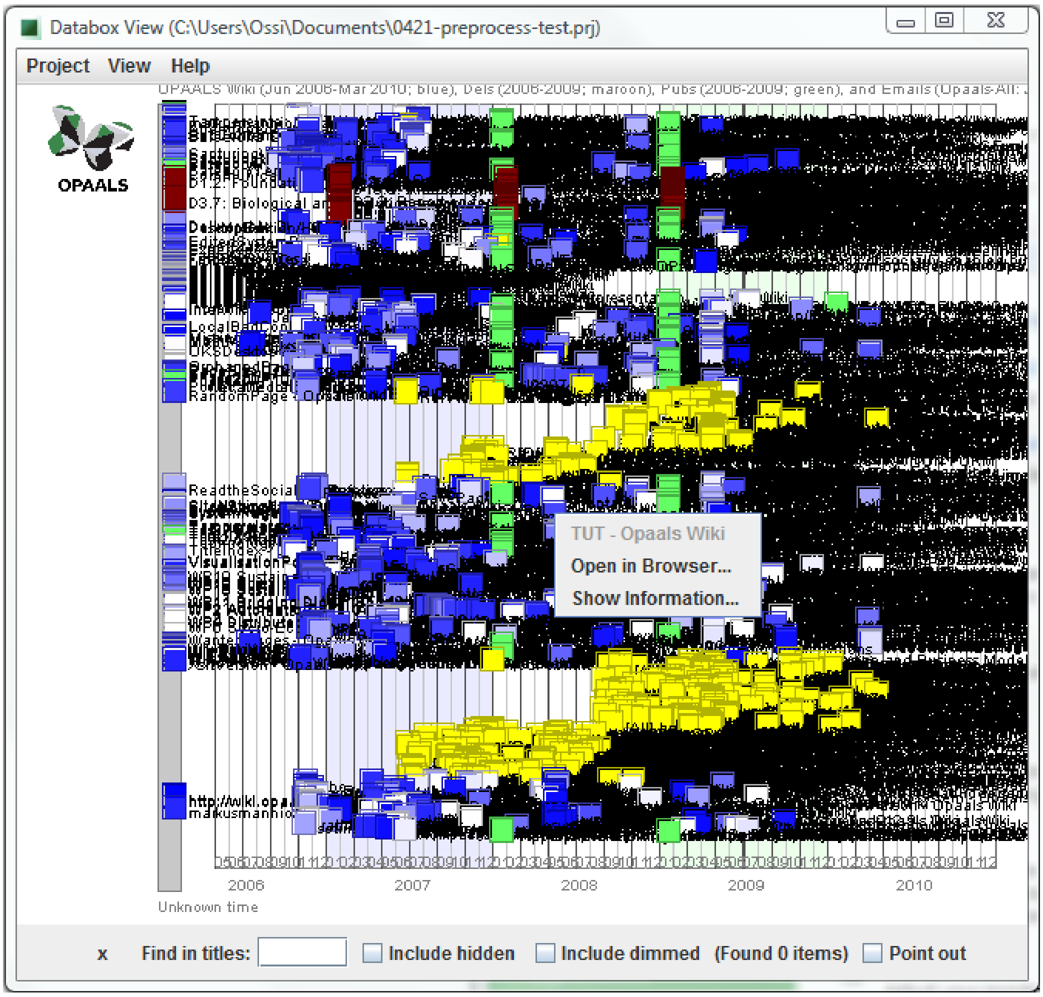

Figure 1. Databox View to the OPAALS OKS

With the help of the interactive application, users can navigate within the item data by zooming and panning (see Figure 2 below), filter and highlight data, and open particular items for further examination in the desktop Web browser.

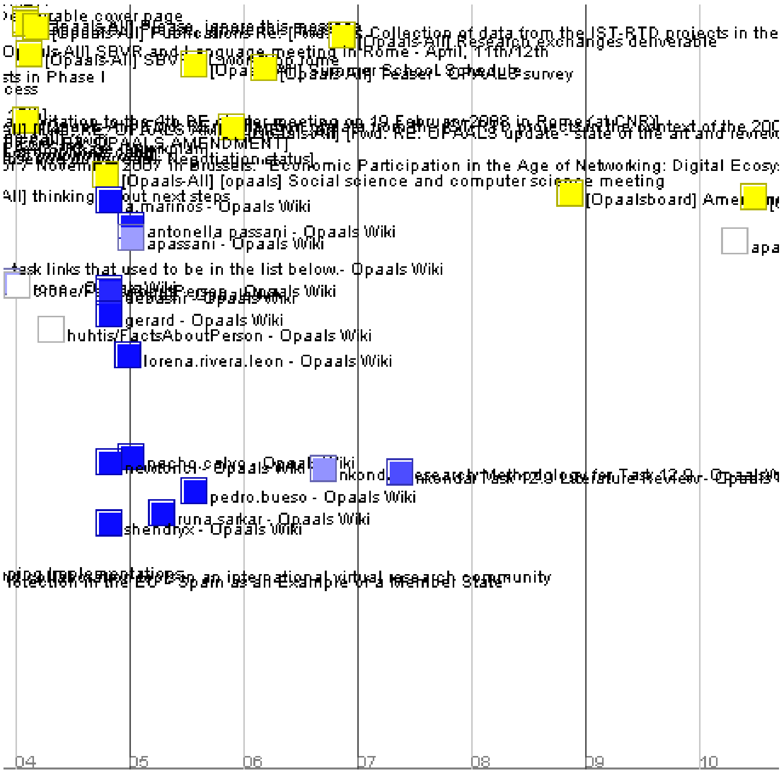

Figure 2. A close-up view for some early 2008 data

The fact that the visualisation may represent also data content, and not only metadata, provides a nice setting for non-trivial explorative analysis. Figure 3 (below) illustrates the usage of specific terms in emails between mid 2007 and early 2010 (n=480), highlighting specific aspects of the data.

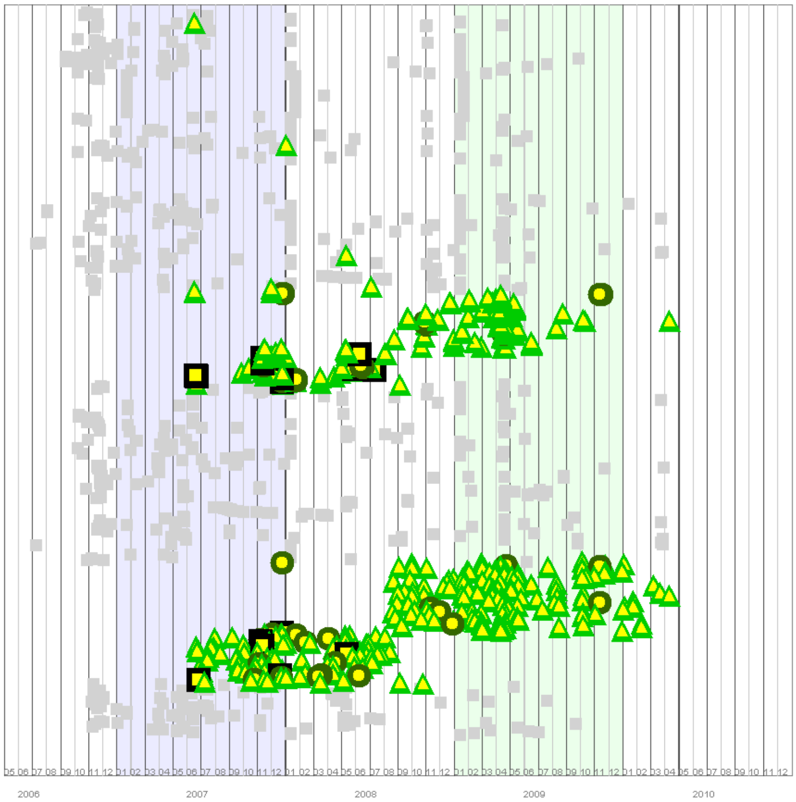

Figure 3. A trend of terminology in emails

The view is computed as follows: First, all item text labels are hidden for clarity and all items besides emails (by type property) are dimmed (presented as small light gray boxes). The email items are then highlighted and depicted as light green triangles. The email items whose title or actual content includes the word "[Ee]cosystem" (the title or pageData property match with the regular expression) are emphasised and depicted as darker ovals. Finally, email items whose content in addition includes the word "[Aa]pplication" are highlighted as boxes with thick black border.

As suspected, the application provides a rich basement for other enquires as well.

4. Analysis

In principle, the visualisation application alone helps information gathering and preliminary analysis activities in several ways (Ware, 2004):

- To comprehend huge amounts of data
- To perceive emergent properties of data
- To validate data
- To associate large and small-scale structures
- To support hypothesis formulation

It is now intriguing to ask what the directly affordable OPAALS view, solely based on sample data tells us.

For instance, one can immediately observe that ecosystems have been a continuous topic in the Opaals-All discussions (the oval items in Figure 3). However, the term application does not appear in this Opaals-All email context after July 2008 (the black thick boxes).

4.1. Elements of Evolution

While this setting is intentionally naive from e.g. a linguistic and semantic point of view, it nicely illustrates our limited access to the knowledge space population. Several plausible explanations of several categories exist.

From a content perspective, two kinds of main assumptions might be drawn:

1. *Evolution of activities.* For instance, after having identified applications, the researchers went on with more theoretical research.
2. *Evolution of language.* For instance, once a general consensus on research directions had been established, it became unnecessary or even redundant to speak about applications, and the general term was either dropped or encoded differently.

These potential explanations are quite intuitive. In particular, the language evolution aspect is nontrivial when the time span of activities is considered.

However, acknowledging that the sample might not capture the knowledge space faithfully, other kinds of hypotheses should also be considered, namely:

3. *Non-representative sample.* For instance, discussion of applications actually increased over time but took place in a medium that was not included in the knowledge sample (such as private or group-specific communications).
4. *Evolution of data semantics.* We do not actually have a clear picture of the process at all because the formal encoding of item properties changed over time on the level of semantic schemata.

While these points are quite challenging to prove/disprove, they effectively underline the more content-oriented analysis above (items 1 and 2). Similar concerns arise, e.g. in discovery or in visualisation applications that seek systematic methods for data validation.

Finally, considering the abstract scope of this deliverable, a view to the evolution of the OKS from 50,000 feet, it is sufficient to say that the underlying aspects through which evolution is demonstrated include ***processes***, ***data***, and ***language***. However, since the language aspect is often neglected, we shall next elaborate it a bit.

4.2. Knowledge Production and Linguistic Processes

The OPAALS community integrates a wide set of different disciplinary perspectives that come in different disciplinary languages, theories, and scientific policies. This diversity holds also true for any trans- or interdisciplinary community that comes together for collaborative knowledge production. Therefore, an important question is firstly how knowledge can be produced collaboratively by crossing disciplinary borders, and secondly how those evolutionary knowledge processes look like and can be explained.

Weick's theoretical analyses of organisations supports our twofold strategy as to combining data visualisation with theoretical interpolations, but also connects the systemic view on communities and their behaviours in OKSs with a evolutionary language perspective. Weick (1995) uses the concept of a 'sensemaking' organisation, where the act of sensemaking among the organisation's members is an ongoing and central process, as it is "the primary site where meanings materialize that inform and constrain identity in action (Mills 2003: 35, cited in Weick, Sutcliffe & Obstfeld 2005: 409). Weick, Sutcliffe and Obstfeld conclude that therefore sensemaking is an issue of language, talk, and communication. This means that language (in general) is a core means for the development of organisations, which also holds true for scientific communities and their interaction/collaboration in OKSs. Striving towards a better understanding of the nature and underlying mechanisms of any OKS, it can be stated from a linguistic point of view that "the mechanisms of language change and variation provide an important input for any system development in the context of its dynamic character, self-organisational aspects and evolution" (Lapteva, Peukert & Zeller: 2009).

This means on the one hand that any interaction, particularly interdisciplinary interaction and collaboration, integrates a certain degree of language change and variation. The reason for this is that the different disciplinary vocabularies and communicative patterns need to be adapted to a 'joint' community language in order to be able to collaborate (based on the premise that human collaboration always depends on communication – be it in natural or formal languages). On the other hand, these language changes (or evolutionary language notion) represent the *conditio sine qua non* for the development of an organisation. Connecting this to Weick, the act of making sense of an organisation's acting (i.e. in the process of knowledge production), could only contribute to the procedural character of the shaping and functioning of a community/organisation, if it is done in a social context and not on a discrete or isolated individual basis. These sensemaking processes can be related for example to more abstract or immaterial aspects of an organisation, such as policies, informal rules, or identity building, or to concrete aspects such as technology usage.

Sensemaking processes also refer to the negotiation of theoretical strategies (research agenda), which means the community needs to find a consensus upon the epistemological framework (i.e. which theories and paradigms are eligible, and which are not) of its collaborative knowledge production. Both processes are complex and significant, however the second implies that the sensemaking process needs to integrate a negotiation of meaning, too. This means that in order to discuss different theoretical frameworks, differing meanings of the same concepts or key terms, such as 'network', need to be negotiated among the various disciplines and stakeholders that are part of the community and OKS.

In deliverable D6.7 we have already pointed out the distinct associations and meanings of key terms within the OPAALS community. Whereas the distinctions were then made according to the different disciplines, it might also be interesting to looks at the temporal dimension of sensemaking and negotiations regarding the emergence of a joint language. Figure 4 therefore presents the results of a longitudinal survey conducted within the OPAALS community, specifically the results from the word-associations questionnaire items (participants had to list 3 to 4 descriptions that they associate with certain pre-defined key-terms).

Thus, re-considering the above depicted finding that the term "application" does not appear anymore in the Opaals-All email context after July 2008, can be enriched by the temporal development of the associations and understandings of the term OKS. 'OKS' can be taken as a synonym for a specific set of tools and applications for collaboration and knowledge production.
As we can see, there are interesting variations regarding the frequency of specific associations over time. Whereas the first wave (conducted in Winter 06/07) shows a remarkable concurrence of the associations "sharing", "openness", and "free", and knowledge being the association with the highest percentage, the results of the second wave (conducted in Winter 07/08) show a rather different and less homogenous picture. Waves three and four were conducted in Summer 2008 and Winter 08/09, and present an increasing stabilisation of the main associations. Looking at the results of the fourth wave in particular, we can see a clear saturation of joint sensemaking and meaning negotiation, with a two-fold set of associations: 'Collaboration', 'sharing', and 'knowledge' represent the set with the highest scores, and in itself a semantic logic as to 'sharing' and 'collaboration' do have a certain semantic proximity, and 'knowledge' representing the connotative object. Likewise, the second set also depicts linguistic proximity with the terms "free" and "open".

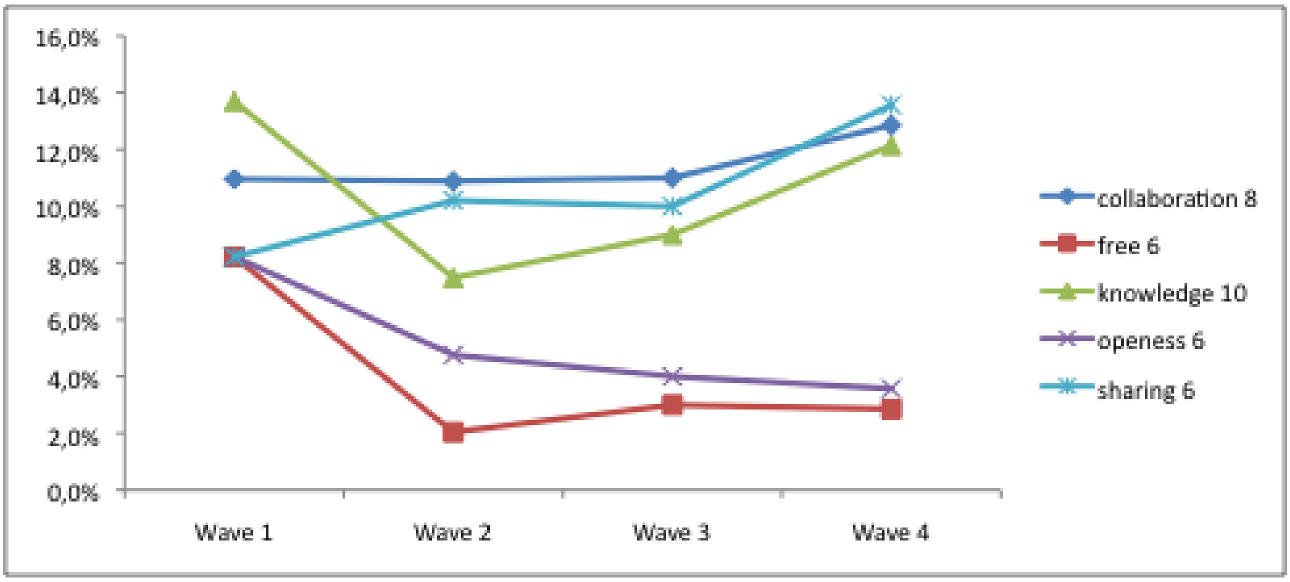

Figure 4. OKS associations

Hence, the absence of the term 'application' in the Opaals-All email context after July 2008 can be explained by a semantic saturation or maybe even a consensus regarding the main application environment that had been achieved after Summer 2008. This also corresponds to our first and second assumption.

5. Conclusion

In this deliverable, we have presented a visualisation application that depicts the OPAALS OKS "from 50,000 feet". The visualisation application provides a tangible view to the OKS and illustrates challenges in knowledge space analysis from evolution analysis point of view.

In the forthcoming articles, we will continue our research by analysing the various factors of evolution, in order to provide of framework for explicating the role of evolution in understanding large-scale knowledge spaces. In brief, we believe that the basic elements of understanding evolution include ***data*** (as the factual/material foundation), ***processes*** (as the organisational foundation), and ***language*** (as the fundamental media of knowledge in the first place). Our working hypothesis is that the elements of evolution depicted in Section 4.1 play a significant role in this setting.

It is interesting to observe that the visualisations and data processing tools can also help to (perhaps indirectly) identify and visualise the sensemaking processes as well as failures in sensemaking processes. On the other hand, the visualisation results can be enriched significantly in terms of its descriptive nature. By drawing upon prior works in the field of language evolution, the visualisations can also be used for normative results regarding for example recommendations for visualisation processes and techniques in general and in the field of knowledge processes visualisation.

Acknowledgements

This work has been kindly supported by the partners involved in WP6 and WP10 work. The authors wish to especially thank their colleagues, Jukka Huhtamäki, Jaakko Salonen, Oxana Lapteva, Marco Bräuer, and Ingmar Steinicke for their contributions.

References

Bijker, W. E., Hughes, T. P., & Pinch, T. J. (1987). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press.

Bräuer, M., Steinicke, I., & Zeller, F. (2009). *Self-reflection of community building, communication and collaboration processes in the NoE*. Report for the European Commission, RP6 Information-Society-Technology IST-2005-034824.

Geroimenko, V. & Chen, C. (Eds.) (2006). Visualizing the Semantic Web. XML-Based Internet and Information Visualization. Springer, USA.

Lapteva, O. & Zeller, F. (2008). *Metaphorological tool-kit*. Report for the European Commission, RP6 Information-Society-Technology IST-2005-034824.

Lapteva, O. & Peukert, H. (2008). *First results from a showcase model of an applied ontology and taxonomic hierarchy*. Report for the European Commission, RP6 Information-Society-Technology IST-2005-034824.

Lapteva, O., Peukert, H. & Zeller, F. (2009). *Models of the Evolutionary Framework*. Report for the European Commission, RP6 Information-Society-Technology IST-2005-034824.

Mills, J. H. (2003). *Making Sense of Organizational Change*. Routledge.

Nykänen, O., Mannio, M., Huhtamäki, J. & Salonen, J. (2007). A Socio-Technical Framework for Visualising an Open Knowledge Space. In: Isaías, P., Baptista Nunes, M. & Barroso, J. (eds.) Proceedings of the International IADIS WWW/Internet 2007 Conference, 5-8 October, Vila Real, Portugal (pp. 137-144). IADIS Press.

Nykänen, O., Salonen, J., Haapaniemi, M., & Huhtamäki, J. (2008). A Visualisation System for a Peer-to-Peer Information Space. Proceedings of OPAALS 2008, 7-8 October 2008, Tampere, Finland. pp. 76-86. A local copy is available at http://matriisi.ee.tut.fi/hypermedia/julkaisut/20080807-nykanen-etal-p2pvis.pdf (PDF 0.6 MB)

Nykänen, O. (2009a). Understanding Data via an RRS in RDF/XML. In IADIS International Conference Applied Computing 2009, Rome (Italy), Nov. 19-21, 2009, Vol. I, ISBN 978-972-8924-97-3.

Nykänen, O. (2009b). Semantic Web for Evolutionary Peer-to-Peer Knowledge Space. In Birkenbihl, K., Quesada-Ruiz, E., & Priesca-Balbin, P. (Eds.) Monograph: Universal, Ubiquitous and Intelligent Web, UPGRADE, The European Journal for the Informatics Professional, Vol. X, Issue No. 1, February 2009, ISSN 1684-5285, CEPIS & Novática. Available at http://www.upgrade-cepis.org/issues/2009/1/upgrade-vol-X-1.html

Telea, A.C. (2008). Data Visualization: Principles and Practice. A K Peters, USA.

Ware, C. (2004). Information Visualization (Second Edition): Perception for Design, Elsevier, USA.

Weick, K. (1995). *Sensemaking in Organizations*. Sage.

Weick, K., Sutcliffe, K.M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science 16*(4).

Zeller, F., Lapteva, O., Crone, A. (2007). *Studies/papers on discourse organisation of epistemic cultures – theoretical and methodological analyses and practical interpolation*. Report for the European Commission, RP6 Information-Society-Technology IST-2005-034824.