

Catalysing the Development of a Conference Workspace

Jukka Huhtamäki, Ossi Nykänen and Jaakko Salonen
Tampere University of Technology, Hypermedia Laboratory,
Korkeakoulunkatu 1, FI-33720 Tampere, Finland
{jukka.huhtamaki, ossi.nykanen, jaakko.salonen}@tut.fi

Abstract. Modern Web introduces new means to support conference information retrieval and, moreover, social tools that enable conference delegates to actively contribute to a conference. Simultaneously, existing case studies show that collecting explicitly represented conference data is possible in real-life settings. From these grounds, methods and tools of component-based information visualisation can be used to process conference data, firstly, in order to create visualisations of the conference data, and, secondly, to produce data for populating interactive conference tools. We claim that when coupled with usage tracking and analysis, the automated creation of visualisations and population data can catalyse the iterative development of a conference workspace. In this article, we present case OPAALS 2008 where formal conference data was successfully harvested and utilised to partly automate the creation workflow of the conference workspace. Moreover, we introduce the methods that were used to gain insight on the dynamics of the workspace usage.

Keywords: information visualisation, online communities, Semantic Web, information modelling, social media

1 Introduction

Most of the conferences and workshops today have a homepage including information about the venue, program, and presenters. Some events offer delegates the possibility to define their profile and to contribute to the conference online. Lately, work for equipping conference delegates with appropriate interactive tools has been done in the field of online conferences.

Traditionally, the role of a conference delegate has been rather passive. Related to *unconferencing*, Winer [1] claims that “[t]he sum of the expertise of the people in the audience is greater than the sum of expertise of the people on stage”. Although this surely is always not the case, we agree that the usefulness of a conference can be added with tools supporting delegates to find their way to see the right talks, to meet other delegates with complementing interests and to contribute to the conference discussion before, during and after the event.

In chemistry, *catalysts* increase the probability of a reaction by reducing the activation energy needed for a reaction to happen. Reducing “activation energy” is important in data-sharing systems [2] and we see that the same analogy is applicable

to conference workspace development. We claim that once the conference data is presented in a form explicit enough to enable automated processing, pipeline-based data processing methods that are commonly used in *information visualisation* can be used to partly automate the creation process of a conference workspace. First, views can be created for visualising the conference data, possibly based on user preferences, and, secondly, conference data can be transformed into formats that can be used to populate conference community and other interactive conference tools. Together with a tracking-based, visualisation-driven feedback loop of the workspace usage, the automation of the workspace creation process can catalyse the iterative and incremental development of a high-end conference workspace: alterations to workspace content or structure can be made quickly as the understanding on the requirements of a workspace improves over time.

In this article, we present work done related to the 2nd International Conference on Open Philosophies for Associative Autopoietic Digital Ecosystems (OPAALS 2008)¹. OPAALS 2008 was organised on 7th and 8th October 2008 in Tampere, Finland by the Hypermedia Laboratory (HLab) of Tampere University of Technology (TUT) in association with Mindtrek Conference² and W3C Finnish Office³. OPAALS 2008 was a relatively small conference with about 50 registered delegates representing fields of, among others, social science, biology, and computer science.

OPAALS 2008 is related to OPAALS Network of Excellence (FP6-034824). The two main objectives of OPAALS are to build a sustainable interdisciplinary research community in the area of *Digital Ecosystems* (DE) and to develop an integrated theoretical foundation for DE research. A significant part of OPAALS is building an *Open Knowledge Space* (OKS) to support the work done in the project. To simplify, we describe OKS as a repository and a model for capturing and managing OPAALS knowledge in an evolutionary manner in the spirit of open source. Currently, *OKS wiki* is the main tool used by the members of OPAALS. As a partner of the OPAALS consortium, HLab is working to develop an information visualisation system for the OKS. As HLab was also responsible for organising OPAALS 2008, the development process of the conference workspace was selected as a real-life pilot scenario for the visualisation system.

The first objective of this article is to describe a semi-automatic process of preparing and setting up a data-driven conference information infrastructure that also supports the task of bootstrapping the related conference community. The second objective is illustrating the pipeline-based mash-up and visualisation technology (reported elsewhere) in this context, analysing the critical factors of technically and socially sound applications.

The rest of the article is organised as follows. In Chapter 2, we review inspiring work related to case OPAALS 2008. The development workflow of OPAALS 2008 is presented in Chapter 3. Chapter 4 includes an informal analysis of our work and Chapter 5 concludes the paper.

¹ <http://matriisi.ee.tut.fi/hypermedia/events/opaals2008>

² <http://www.mindtrek.org/conference>

³ <http://www.w3c.tut.fi/>

2 Background and Existing Work

Let us next briefly review some inspiring work and insights related to (1) conference data modelling, (2) conference author social network visualisation, and (3) the possibilities of using formal conference data to support conference delegates in making the most of the time spent in a conference. Moreover, we introduce *Wille*, a lightweight information visualisation framework that was applied in OPAALS 2008 to automate the processing of the conference data.

In order to demonstrate the real-life usage of Semantic Web technologies and to test the workflows and tools needed to collect semantically rich data in a realistic setting, or to eat the “Semantic Web Dog Food”, conference metadata about people, papers and talks has been created for a number of European and International Semantic Web Conferences (ESWC and ISWC) [3]. The metadata is represented in Resource Description Framework (RDF) format following the *ESWC2006 Conference Ontology*. The ontology combines concepts of the Friend of a Friend (FOAF) vocabulary⁴ to concepts specific to the conference domain. Currently, data representing six academic conferences is available online⁵ both as a browsable repository and via an Application Programming Interface (API) supporting SPARQL Query Language for RDF (SPARQL)⁶. Moreover, a simple geographic map of “Dog Food Organisations”⁷ is provided as an additional mean for navigating the data.

A profound case of academic conference data visualisation was done related to IRIS (Information Systems Research Seminar in Scandinavia) conference [4]. To visualise the evolution of the social network of authors and the research topics of IRIS on a period from 1978 to 2006, data representing articles and their authors was collected manually from web pages, conference proceeding CDs and other sources and inserted into a database. A dynamic network visualisation tool Commetrix was used to visualise the data. The outputs of the work include animations of the evolution of IRIS co-authorship and of the diffusion of terms in IRIS community⁸. In addition to solving the issues related to heterogenous sources of data, manual work was needed to identify the authors whose names or origins were written in various ways. In particular, Scandinavian letters (å, ø, ä, ö, æ, etc.) caused problems for data collectors since author names including such letters are expressed in various ways.

In his winning proposal for the Semantic Conference Design Challenge of the 3rd Annual European Semantic Web Conference, Grimnes [5] nicely sums up many of the needs of a conference delegate: delegates should be able to efficiently browse conference program and tag talks, on basis of which a conference system, where the user data could possibly be initialised with existing profile information in FOAF format, could guide the delegate intelligently through the conference. Today, such intelligent online communities can be built via integrating and tailoring existing community engines and other Web 2.0 development tools.

⁴ <http://www.foaf-project.org/>

⁵ <http://data.semanticweb.org>

⁶ <http://www.w3.org/TR/rdf-sparql-query/>

⁷ <http://data.semanticweb.org/organization/>

⁸ <http://chacini.cs.tu-berlin.de/Commetrix/iris/iris.html>

The underlying objective of *information visualisation* is to serve as an amplifier of the cognition of a user through expressive views giving insight on a certain phenomena represented by the data [6]. In order to support the development of information visualisations for the Open Knowledge Space and other sources of explicit or heterogenous information, information visualisation framework Wille is being developed by the Hypermedia Laboratory [7]. A *visualisation designer* can use Wille to build information processing pipelines that connect to different sources of data, process the data through series of queries and transformations and create stand-alone visualisations to *visualisation users*. Alternatively, Wille can be applied to produce data for existing information visualisation tools including Vizster (see [8]). An example of applying Wille is a social network visualisation of OKS wiki contributors where Wille is used to access and process OKS wiki data to create a representation of wiki editing history in a format supported by Vizster [9].

Two main versions of Wille exist, Wille Phase 1 and Wille Phase 2. Due to the early stage of development of Wille Phase 2, the batch-processing based Wille Phase 1 was used for conference data processing and visualisation creation.

3 Development of OPAALS 2008 Workspace

The OPAALS 2008 development workflow is depicted in Fig. 1. Conference data is collected and represented in an explicit format enabling the automation of workspace creation. Moreover, the usage of the workspace is tracked and visualised to gain insight on workspace dynamics.

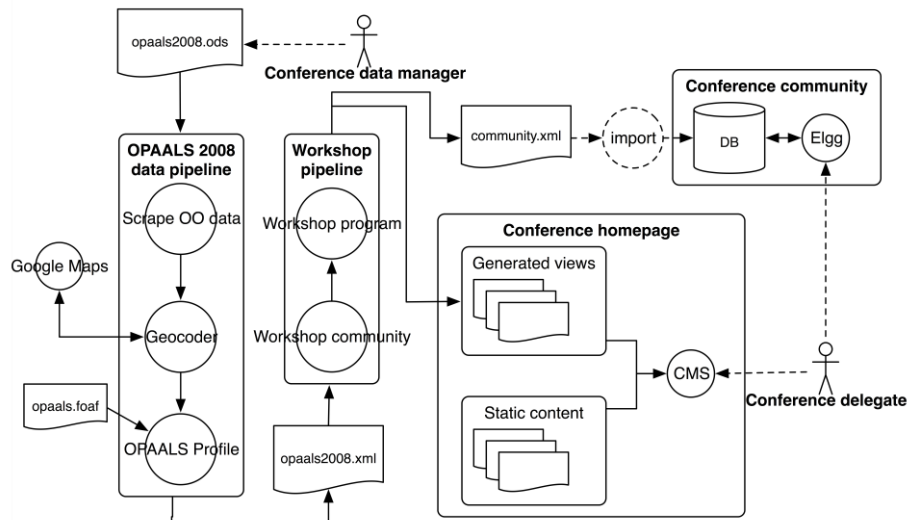


Fig. 1. A diagram presenting the workflow of OPAALS 2008 workspace development

A detailed description of the workflow follows.

3.1 Collecting and Modelling OPAALS 2008 Data

Traditional conference management tools were used to collect OPAALS 2008 data. Articles were submitted through an online open source conference management system OpenConf⁹ that was installed for OPAALS 2008. Conference delegates registered via Mindtrek registration system. The registration data was delivered to the OPAALS 2008 team every now and then via email. Since OpenConf (we used version 2.01) does not provide a separate management process of camera-ready articles, authors submitted the final versions of articles via email in Portable Document Format (PDF). The files were named according to a uniform naming scheme to enable link generation.

OPAALS 2008 conference data is composed of *article data* (id, author(s), title, session, keywords, abstract, full article in PDF format), *delegate profile information* (id, name, conference community username, email, organisation, location [city, country] and a list of interests) and *session information* (id, title, type, chair, start time, end time). Moreover, for members of OPAALS, additional profile information (image, additional list of interest) was collected from the OKS. For this, OPAALS username was added to the conference data to connect the OPAALS member identities to respective OPAALS 2008 delegate identities.

In order to avoid time-consuming software development, a (human) *conference data manager* conducted the aggregation of conference data, an effort resulting to an Open Office Calc spreadsheet (opaals2008.ods in Fig. 1). A few iterations were made to collect sufficient data that fully supported the creation of the needed visualisations.

Although a human actor was needed to manage data collection and aggregation, efforts were made to smoothen the process. For exporting article information from the submission management system via a Web interface, a data API capable of accessing the OpenConf database and exporting the data in a machine-readable format was implemented. Moreover, a set of Wille pipelines were developed for processing the data: Two previously developed pipelines, one for getting OPAALS member data in canonical FOAF format (opaals.foaf in Fig. 1) (see [9] for more information) and another for fetching OPAALS member images were re-used. In addition, a dedicated *OPAALS 2008 data* pipeline was developed for picking the spreadsheet content from the Open Office Calc file (which is, effectively, a zip file including spreadsheet content in XML format), scraping the data load, geocoding delegate origins with Google Maps API¹⁰, adding pre-fetched OPAALS member profile information and, finally, creating an Extensible Markup Language (XML) representation of the conference data (opaals.xml in Fig. 1).

3.2 OPAALS 2008 Workspace

The OPAALS 2008 conference workspace¹¹ is designed to support conference delegates to (1) prepare for the conference, (2) work effectively during the conference and (3) follow-up discussion and initiatives after the physical event. The workspace is

⁹ <http://www.openconf.com/>

¹⁰ <http://maps.google.com>

¹¹ <http://matriisi.ee.tut.fi/hypermedia/events/opaals2008/>

two-folded: a traditional conference homepage is spiced up with conference data visualisations and complemented with a dedicated conference community driven by an open source social network engine Elgg¹².

The following visualisations are included in the workspace. *Interactive program* is a densely hyperlinked view to the conference program. *Presenter list* introduces all the presenters appearing in OPAALS 2008. *Article list* enumerates all the articles published in OPAALS 2008 proceedings. The full version of each article is linked to Article list in PDF as well as the related conference presentations that were delivered to the OPAALS 2008 team. A *keyword cloud*, a tag cloud (see e.g. [10]) based on a histogram of conference article keywords is provided to support browsing the articles by their subject. Finally, *Presenter map*, a traditional map-based mash-up utilising Google Maps, shows the whereabouts of each presenter giving a talk in OPAALS 2008.

The interlinking of the different views of the workspace is tight, thus providing flexible means to browse the conference contents. Moreover, conference sessions are linked to discussions running in the conference community and author information to their user profiles in the community. A simple Content Management System (CMS) was built to serve the static content as well as the generated views within a page template providing a uniform navigation scheme and general look and feel to the conference homepage. The CMS does not implement user session management or authentication, thus creating e.g. a personal tagging system for program was not possible.

A general *Workshop pipeline* was built to produce the needed visualisations and population data files. The pipeline is composed of components *Workshop program* and *Workshop community*, first creating the Interactive program, the Delegate list, the Article list and marker data for the Presenter list (markers represent information in a Google Map) and second creating a data file (community.xml in Fig. 1) used to populate the conference community. A set of tailored XML vocabularies are used to represent the data within the pipelines. Most of the data processing is implemented as Extensible Stylesheet Language Transformations (XSLT). The visualisations are based simply on Hypertext Markup Language (HTML) and Cascading Stylesheets (CSS). In addition, Presenter Map uses Google Maps basic technologies, Keyhole Markup Language (KML) and the Javascript API of Google Maps.

A mildly tailored version of Elgg (version 0.9) was used to run the conference community. A bash script was implemented to create a community account for each delegate, to populate the community profiles with collected data and to deliver the account information to delegates via email. To minimise the efforts of delegates and to enable linking from the Presenter list to the conference community, the conference data manager defined the community usernames manually into the conference data.

In addition to the online conference workspace, a stand-alone version of the workspace was created through parameterisation of the Workshop pipeline and disseminated as a CD-ROM. Finally, a manually edited version of OPAALS 2008 proceedings in PDF was included both into the conference homepage and the CD-ROM.

¹² <http://elgg.org/>

3.3 Gaining Insight on Workspace Dynamics

During the development of the OPAALS 2008 workspace, feedback from different visualisations and features of the conference workspace was collected from the members of the OPAALS 2008 team in face-to-face meetings and, in addition, from social network usefulness experts through informal discussions and reviews. Moreover, in order to provide the workspace development team with concrete data on the workspace usage before, during and after the conference, we installed Google Analytics¹³, an online usage tracking and visualisation tool, for collecting and visualising the usage data. In addition, a tailored log collector was implemented to the conference community platform by utilising the extension mechanism of Elgg.

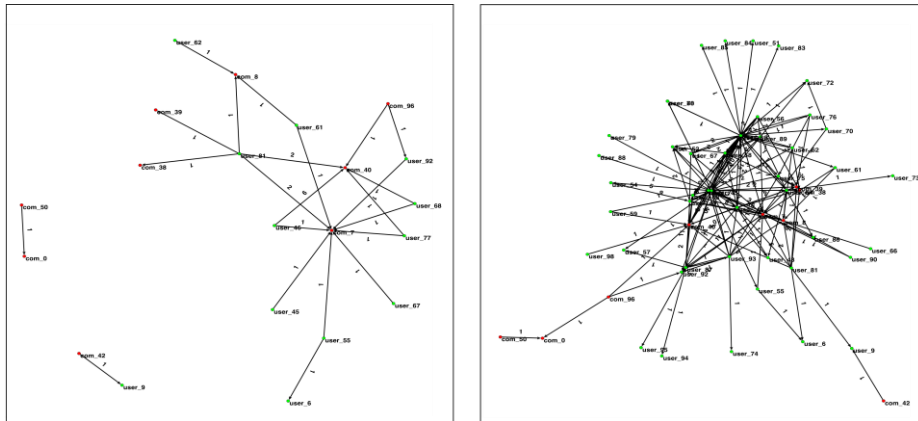


Fig. 2. A visualisation of the evolution of the conference delegate social network (two frames)

Google Analytics enables us to gain insight on the usage of the workspace in a general level. Moreover, on basis of the community log information, we are e.g. able to generate a visualisation of the evolution of the social structure of the community. Two excerpts of the development are shown in Fig. 2. Nodes represent the community members and the theme groups in the conference community and lines represent the interconnections between them. The graph representation of the community is created with a social network analysis tool Pajek¹⁴. In OPAALS 2008, the profile views constitute the majority of the community activity, thus making the social network data less significant. Similar means can, however, be used to create social network visualisations in conferences where online interaction is more diverse. Moreover, animations of social network evolution can be created from the log data e.g. with SoNIa, a tool for visualising dynamic, longitudinal network data¹⁵.

The information derived from usage tracking and visualisation can be used to reflect the dynamics of the workspace. For instance, we decided to separate the

¹³ <http://www.google.com/analytics/>

¹⁴ <http://pajek.imfm.si/>

¹⁵ <http://www.stanford.edu/group/sonia/>

traditional, static homepage of the conference from the more interactive workspace. This meant that we had two versions of the conference program, one with hyperlinks to different parts of the workspace and another without. In the period of September 15, 2008 to February 15, 2009, the unlinked version of the program was viewed twice as much as the interactive version. We suspect that this may have had a major effect on the popularity of the other parts of the workspace. In the future, we would create only one program that would initially be a static one and would be replaced by the interactive version once enough data for creating one is available.

4 Discussion and Informal Analysis

Case OPAALS 2008 serves as an example of the potential of data driven development of online workspaces. We have showed that when sufficient technical expertise is available in a conference team, explicitly represented conference data can be collected with traditional conference management software, spreadsheets and other state-of-the-art tools and, moreover, a useful workspace can be created from this data for supporting the conference delegates.

From the end-user point-of-view, an optimal solution for a conference workspace would be achieved through a fully integrated system supporting submission management, registration, collaborative tagging, discussions and communication, networking and conference data visualisation. Since developing such a tailored system is not always possible and application needs vary, a lightweight alternative created by utilising explicitly represented conference data, component-based information visualisation pipelines and contemporary mash-up and community technologies provides a valid option.

An automated publishing process has several advantages in addition to enabling an iterative, data-driven development of a conference workspace. Examples include the possibility to create a consistent Uniform Resource Identifier (URI) space for the workspace to enable deep linking of conference artefacts and third-party created complementing metadata, following the principles of design for reuse [11]. Programmatically created views to the conference data are, due to their predictable structure, both browsable (by humans) and crawlable (by machines), thus forming a repository of digital conference artefacts. Moreover, it is possible to refresh the format of the data views e.g. by adding support for Microformats¹⁶ for representing the conference schedule and the profile information of article presenters. In general, the automation introduces means to conference source data normalisation, thus adding to the data quality by preventing errors in data replication and insuring the referential integrity of links within the data and the generated views.

The dynamics of the workspace usage can be better understood through tracking and visualisation of the usage providing insight on the popularity of individual views, the navigation paths and even the structure of the social network of the delegates. While Google Analytics provides useful general views to workspace usage, the usage data collected with Google Analytics is still quite imprecise. Importantly, the users

¹⁶ <http://microformats.org/>

identity remains unknown, thus we are not able to connect users e.g. to their demographic information. Using only Google Analytics for the public part of the conference workspace hinders us to follow the paths of individual conference community users outside the community where a built-in tracking mechanism does enable user-specific tracking.

Several efforts can be made to further add to the usefulness of a conference workspace. Using a temporary community for a conference does not encourage the use. Moreover, system-specific authentication credentials hinders the use of the community. Applying OpenID¹⁷ or some other general authentication scheme would enable the use of existing digital identities in such a short-lived community. Also, more advanced conference data can be collected. Article references, for example, can be formalised e.g. with Zotero¹⁸. A general identification scheme for authors and articles is, however, needed in addition to a consistent reference representation scheme to enable advanced computing on basis of the reference information.

5 Concluding Remarks

Complemented with component-based data processing pipelines and extensive export and import mechanisms of individual conference tools, explicit representation of conference data provides means to decouple the unique, invaluable data from the individual conference tools that tend to change on yearly basis as more prominent technologies emerge. Data processing pipelines can be developed to create the needed visualisations views directly and to generate data files in formats suitable for feeding various visualisation players and social platforms. Iterative development is supported through the possibility to add new views to existing data by implementing new publishing routines and to refine the workspace by tailoring the existing ones.

Once collected, a formal conference data set can be utilised in the future e.g. by aggregating the data with other existing sets of data or by creating new representations of the data in appropriate formats. However, to ensure the option to use a certain data format, the characteristics of the format should be taken into account so that all the necessary pieces of information are collected.

The system developed for OPAALS 2008 is a combination of a publishing framework, an information visualisation system and a conference management system. Following the presented approach in a series of future conferences, we are able to analyse and also actively steer the development of workspaces for academic conferences and other communities. The OPAALS 2008 conference workspace serves as a firm basis for the development of future workspaces. Work on tailoring existing components and pipelines and creating new ones is, nevertheless, needed when new workspaces are created, insisting on experience in utilising Wille.

In order to fully understand the dynamics of the conference workspace and thinking of the workspace users, the methods of user-centered design and qualitative research methods should be used to complement the measures and figures created through quantitative means. However, usage visualisation surely serves as means to

¹⁷ <http://openid.net/>

¹⁸ <http://www.zotero.org/>

facilitate the discussion on workspace development both for the archive of OPAALS 2008 as well as for future conferences.

To enable the emergence of a conference ecosystem composed of individual services capable of communicating dynamically to each other, APIs are needed for data access and easily machine-readable formats for data interpretation. Shared policies and practices, however, ultimately define whether such a system will emerge.

Acknowledgments. The authors would like to express their gratitude to Thumas Miilumäki for creating the social network visualisations with Pajek, Arto Liukkonen for tailoring the conference community and Matti Haapaniemi for the development of early prototypes of workshop data visualisation. Moreover, we appreciate the help and feedback from the OPAALS community.

References

1. Winer, D.: What is an unconference? (March 5, 2006), <http://scripting.wordpress.com/2006/03/05/what-is-an-unconference/>
2. Smith, K., Seligman, L., Swarup, V.: Everybody Share: The Challenge of Data-Sharing Systems. *Computer*, 41(9), 54--61 (2008)
3. Möller, K., Heath, T., Handschuh, S., Domingue, J.: Recipes for Semantic Web Dog Food — The ESWC and ISWC Metadata Projects. In: *The Semantic Web, Lecture Notes in Computer Science*, Vol. 4825, pp. 802--815. Springer, Berlin/Heidelberg, Germany (2008)
4. Molka-Danielsen, J., Trier, M., Slykh, V., Bobrik, A., Nurminen, M.: IRIS (1978-2006) Historical Reflection through Visual Analysis. In: *IRIS30*. Tampere, Finland (2007)
5. Grimnes, G.: Semantic Conference Program. In: *3rd European Semantic Web Conference*. Budva, Montenegro (2006), <http://www.eswc2006.org/technologies/designchallenge/semantic-conference-program.pdf>
6. Ware, C.: *Information Visualization: Perception for Design*. Elsevier, San Francisco (2004)
7. Nykänen, O., Salonen, J., Haapaniemi, M., Huhtamäki, J.: A Visualisation System for a Peer-to-Peer Information Space. In: *2nd International OPAALS Conference on Digital Ecosystems*, pp. 76--85. Tampere University of Technology, Tampere, Finland (2008)
8. Heer, J., boyd, d.: Vizster: Visualizing Online Social Networks. In: *2005 IEEE Symposium on Information Visualization (INFOVIS'05)*, pp. 33--40. IEEE Computer Society, Washington, DC (2005)
9. Huhtamäki, J.: Community visualisations in Open Knowledge Space: Uncovering rabbit holes in a digital ecosystem. In: *1st OPAALS workshop*. opaals.org (2007)
10. Hearst, M. A., Rosner, D.: Tag Clouds: Data Analysis Tool or Social Signaller? In: *41st Annual Hawaii International Conference on System Sciences*, p. 160. IEEE Computer Society, Washington, DC (2008)
11. Vinoski, S.: Serendipitous Reuse. *IEEE Internet Computing*, 12(1), 84--87 (2008)