PIM Meets Web 2.0

Moira C. Norrie

Institute for Information Systems, ETH Zurich
CH-8092 Zurich, Switzerland
norrie@inf.ethz.ch

Abstract. Web 2.0 refers to a new generation of web applications designed to support collaboration and the sharing of user-generated content. These applications are increasingly being used, not just to share personal information, but also to manage it. For example, a user might use Facebook to manage their photos and personal contacts, a networking site such as LinkedIn to manage professional contacts and various project Wiki sites to manage and share information about publications and presentations. As a result, personal data and its management become fragmented, not only across desktop applications, but also between desktop applications and various Web 2.0 applications. We look at personal information management (PIM) issues in the realm of Web 2.0, showing how the respective communities might profit from each other.

1 Introduction

The term Web 2.0 has been adopted to refer to a new generation of web applications specifically designed to support collaboration and the sharing of user-generated content [1]. Applications commonly classified under Web 2.0 include social networking sites such as Facebook, sites to share and manage multimedia content such as YouTube and sites that support collaborative authoring such as Wikipedia.

Web 2.0 applications are increasingly being used not just to share personal information, but also to manage it. For example, a user might use Facebook to manage personal contacts and photos, networking sites such as LinkedIn to manage professional contacts and various project Wiki sites to manage information about publications and presentations. As a result, personal data and its management becomes fragmented, not only across desktop applications, but also between desktop applications and various Web 2.0 applications.

We propose that there should be a clear separation of concerns between *publishing data* and *managing data* with the former being the task of Web 2.0 applications and the latter the task of personal information management (PIM) systems. Further, the PIM system should provide an *integrated solution* to the management of all forms of personal information management, whether related to social or professional activities of the user, and it should also be responsible for controlling where, when and how information is published to Web 2.0 applications.

Q. Li et al. (Eds.): ER 2008, LNCS 5231, pp. 15–25, 2008.
© Springer-Verlag Berlin Heidelberg 2008

At the same time, we believe that the developers of PIM systems can learn valuable lessons from the popularity of Web 2.0 applications when it comes to designing systems for the management of personal data. Sites such as Facebook provide simple, intuitive interfaces along with a plug-and-play architecture that allows users to easily select and combine applications. Further, users can even create and share their own applications.

In this paper, we examine personal information management issues in the realm of Web 2.0, showing how the respective communities might profit from each other. We start by examining some of the data management issues related to Web 2.0 applications in Sect. 2. In Sect. 3, we then discuss the recent renewal of interest in PIM systems within the research community and outline the main approaches proposed in various research projects. Following on from these discussions, we present an architecture designed to develop an integrated solution to data management for PIM and Web 2.0 in Sect. 4 and outline our on-going work in this area. Concluding remarks are given in Sect. 5.

2 Data Management for Web 2.0

As mentioned previously, Web 2.0 applications include social networking sites such as Facebook, Xing and LinkedIn, sites to share and manage photos and videos such as Flickr and YouTube, and sites that support collaborative authoring such as project Wikis. While Web 2.0 does not define a particular technology, it is commonly associated with a number of technologies that can support the forms of interaction, collaboration and information sharing characteristic of these applications. For example, Asynchronous JavaScript and XML (AJAX) increases the interactivity and responsiveness of web pages important in many Web 2.0 applications. AJAX toolkits support the development of the required JavaScript and are available for most web scripting languages such as PHP and ASP.NET. To support the development of Web 2.0 applications, Google Web Toolkit (GWT) can be used to transform Java-based applications into AJAX applications.

The term Rich Internet Application (RIA) introduced by Macromedia in 2002 to describe web applications with the same level of interactivity as desktop applications is often used in relationship to Web 2.0 applications. In the early days of web applications, Java Applets were proposed as a technology to support highly interactive applications by downloading Java applications to allow client-side processing. This even included systems where components of a DBMS were downloaded onto the client to improve user interaction [2]. However, later, Java Applets tended to be abandoned in favour of Java Servlets and server-side processing due to various problems such as browser variability, security restrictions and latency. Now that web technologies are more mature, the vision of desktop-style applications being accessible over the web and within browsers is more realistic. Major software companies such as Adobe, Microsoft, Google and Sun Microsystems are all developing tools to make this vision a reality. Examples of technologies that have been developed or extended to support RIA

are DHTML, Adobe Flash, Microsoft Silverlight and JavaFX. Based on these technologies, a number of RIA development frameworks have been proposed such as Adobe Flex, Microsoft Popfly and the open-source project OpenLaszlo.

While RIA technologies are designed to support rich interaction which is certainly a characteristic of Web 2.0 applications, they do not specifically support other characteristics such as user participation and collaboration. A Wiki is software that supports collaborative authoring of web sites and the term dates back to 1994 with the emergence of WikiWikiWeb. Wikis have been widely adopted with the best known application being Wikipedia. They are often used nowadays to support research and commercial projects, enabling members of a project team to easily upload and share documents as well as collaboratively authoring design documents and articles.

Another feature of many Web 2.0 applications is the ability to reuse content from existing web sites, often integrating it to provide new or value-added services. Users can create their own applications by combining data from existing web applications through a notion of web mashups[1]. The content is usually generated by RSS or Atom web feeds, screen scraping or public programming interfaces. A common example is to combine data tagged with location information, for example hotels, with Google Maps. Various tools have been developed to allow users to easily combine data from web feeds to create their own mashups, e.g. Yahoo Pipes, Microsoft Popfly and the Google Mashup Editor.

Research within the database community related to Web 2.0, tends to focus mainly on issues of *data integration* of which mashups are one example. A recent joint effort by the University of Illinois and the University of Wisconsin is a project to develop a software platform to set-up and support on-line communities [3]. As a first step, they have developed a community portal DBLife [4] for the database research community that will serve as a driving application for their research. The DBLife systems monitors more than 900 data sources, extracting and integrating data about people, events, publications etc. relevant to the database community. A major research issue that they want to address is how to ensure data quality and part of the proposed solution is the encouragement of user participation.

Having outlined the key ideas and technologies characteristic of Web 2.0 applications, we now turn to consider related research in the field of web engineering. The long-term goal of the web engineering research community is to develop technologies, tools and methods to support the systematic design, development, deployment and maintenance of high quality web applications. This is a major challenge in a field in which new technologies and tools are constantly emerging, but a major influence has been the promotion of model-based approaches. Leading research efforts in this field include WebML [5], Hera [6], WSDM [7], OOHDM [8], OO-H [9], SiteLang [10] and UWE [11]. WebML and Hera stand out as model-based approaches which feature comprehensive implementation platforms. As the requirements of web applications have evolved to deal with features such as multi-channel access, context-awareness and mobility, researchers have

[1] http://www.programmableweb.com

addressed how to adapt and extend their models and methods to support these. WebML, in particular, has been enriched several times, using its built-in support for extensibility to introduce additional concepts for the definition of business workflows, web services and context-aware adaptation [12,13]. With the rapid growth in interest in RIA and Web 2.0, a current topic of research within the web engineering communtiy is how to support the design and development of RIA and Web 2.0 applications.

We note that, with few exceptions, research projects in all areas related to Web 2.0 tend to be based on existing data management platforms and there has been little consideration given to how databases could play a more central role in managing all forms of data that define a web application and supporting both the development and operation of a web site. Further, user studies related to Web 2.0 applications have tended to focus on the social networking or collaboration aspects rather than on issues of personal information management. If anything, these issues have only been considered at the level of *data integration* rather than *data management*. Yet, anyone who is a regular user of Web 2.0 applications will be well aware of the rich variety of personal information being managed by these applications and the fact that application support for managing all sorts of data ranging from contacts to photo albums often makes it much more convenient to use a Web 2.0 application than desktop applications. A key advantage of using a Web 2.0 application such as Facebook is the fact that all these applications are integrated in a single, portal-like interface. Also, simple tagging mechanisms allow links to be easily created across applications, for example, between a contact and a photo.

On the down side, personal data often ends up being replicated and fragmented. For example, some personal contacts may be managed using Facebook, while professional contacts are managed using a site such as LinkedIn or Xing. At the same time, a desktop application such as Microsoft Outlook may be used to manage more general contact information including contacts who are not registered on Web 2.0 sites. Photos may be stored on a desktop PC, with subsets uploaded to Web 2.0 sites such as Facebook. Information about publications may be published on one or more project web sites and also personal web sites.

We therefore feel that studies should be undertaken to find out more about *how* and *why* users are managing personal data using Web 2.0 applications. This should include examining the problems of replication and fragmentation of data across Web 2.0 applications as well as between desktop and Web 2.0 applications. Based on these studies, new data management solutions should be developed that will allow personal information to be managed in a convenient, integrated manner and published to Web 2.0 applications as and when required.

3 PIM Systems

Although personal information management (PIM) is a topic that has long been of interest to the research community, particularly with respect to possible replacements for the desktop paradigm, there has been a recent renewal in interest

as seen by the series of PIM workshops started in 2005[2]. The workshops are inter-disciplinary, bringing together researchers from various domains including human-computer interaction, information retrieval and databases.

The basic model of managing personal data has changed little over the past decades. Essentially, today's PIM solutions are based on the file system and desktop applications. One problem is the fact that personal information is typically managed by different applications and often stored in different places, making it difficult to handle data uniformly and integrate it in interesting ways. This problem has been referred to as *information fragmentation* [14] or *information compartmentalisation* [15].

The most radical approach is to consider replacing the file system as the basic model underlying PIM with a different model that allows information to be managed and shared in more flexible ways. For example, in the Presto system [16], they developed a notion of shareable document spaces to replace the file/folder means of hiearchically classifying documents within personal spaces. Documents could be freely tagged with properties that could then be used to classify and retrieve documents. One of the major drawbacks of such an approach is the problem of migrating existing data and applications. If applications are to take advantage of the flexibility that new PIM models offer, then they have to be re-designed.

With the dramatic increase in the volume of personal data typically stored by users, researchers in the information retrieval and database communities have become interested in trying to adapt their technologies to the problems of retrieving and processing information stored as personal data. In both cases, they typically build tools on top of existing file systems and applications that can allow data to be extracted and integrated from various documents to meet a user's information needs. For example, in the position paper by Franklin, Halevy and Maier [17], they propose a notion of *dataspace systems* where traditional database technologies such as metadata management, indexing and query processing can be used alongside traditional file systems and applications to support the administration, discovery and enhancement of personal data. This is the approach that has, for example, been adopted in the iMemex system [18].

Both of the above approaches have had limited success to date. One reason for this is that both approaches typically require major efforts in the reengineering of applications or ways of user working. Therefore while they tend to be of theoretical interest, they have had little impact in the everyday use of computers. In the meantime, the development of Web 2.0 applications has caused a dramatic shift in personal information management that has almost gone without remark in the research community. Many users are increasingly shifting away from traditional desktop applications for managing all of their personal information and instead are using Web 2.0 applications. This applies to professional as well as social information since people are increasingly using Web 2.0 applications such as Wikis and community portals not only as a basis for collaboration, but also

[2] Information about these workshops, papers and report can be found at http://pim.ischool.washington.edu/

to manage information about publications, articles of interest, bookmarks etc. Also, messaging supported in systems such as Facebook and community portals is now often being used to support asynchronous communication rather than email systems. There are a number of reasons for this trend away from some desktop applications to Web 2.0 applications. One is the nature of Web 2.0 applications to empower the users as information providers and promote information sharing. Thus, a user does not need to create and manage the contact details of friends and colleagues as they do this themselves. By each user providing a small amount of information, the combined effect is a vast information space.

We believe another reason is the very nature of Web 2.0 applications and their portal-style interfaces as discussed in the previous section. Sites such as Facebook provide an integrated solution to the management of all sorts of data through a very simple, intuitive style of interface. While a core set of applications are provided to manage basic information such as contacts, messages, photo albums etc., it is simple for users to install other applications of interest and even to write their own applications. Facebook now offers several thousand applications[3]. This plug-and-play style typical of many Web 2.0 applications makes it easy for users to customise their site in terms of the types of information stored and published, their own visibility, the level of information sharing and also the layout. In addition, Facebook provides a rich networked information space by automatically generating links between information items and applications based on social networks as well as explicit links created by users through image tagging etc. Last but not least, Facebook offers awareness information about the activities of users through status messages and news feeds.

Given the overwhelming success of Web 2.0 applications, we believe that the PIM community could benefit from trying to understand the reasons behind their success and possibly adopting the Web 2.0 paradigm in the design of future PIM systems.

4 Integrating PIM and Web 2.0

Our goal is to provide improved, integrated PIM solutions based on the Web 2.0 paradigm that will at the same time support the publishing and sharing of data through Web 2.0 applications. The information architecture that we aim for is shown in Fig. 1. Each user manages their personal information through an instance of PIM 2.0, a personal information management portal, and users have control over how and when this information is published to one or more Web 2.0 applications. Further, since Web 2.0 applications are about the sharing of user-generated content, it is possible for users to have data published on Web 2.0 applications by other users automatically imported into their own personal information space.

PIM 2.0 has a plug-and-play architecture that allows users to select and even develop their own information management components as and when required.

[3] Facebook listed more than 17'600 in February 2008.

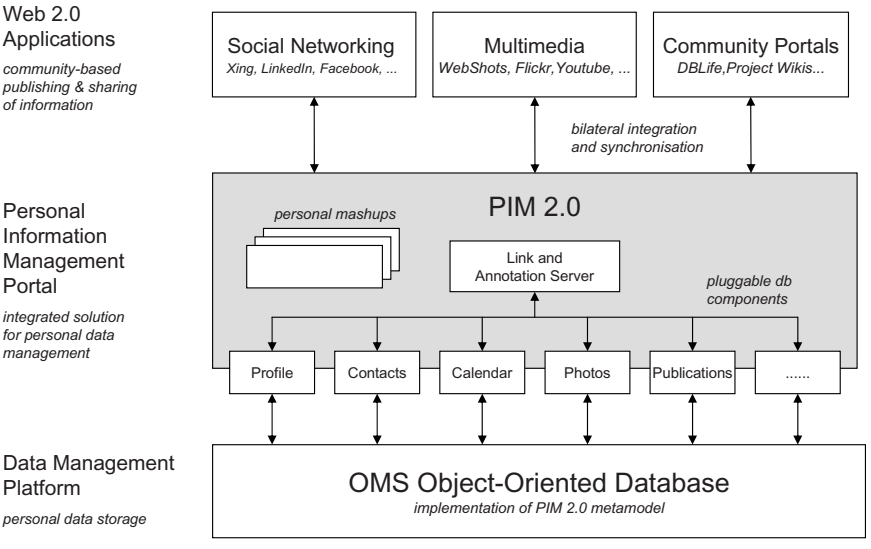

Fig. 1. Information Architecture

We aim to make these *database components* rather than *services* since we want to achieve tight integration at the level of data management to enable us to leverage as much as possible of the database functionality and semantics within PIM 2.0. Also, it should be possible to create links between objects in different database components and to create mashups by integrating data from one or more components. We therefore introduce a general link and annotation server as well as the concept of personal mashups in PIM 2.0.

The concept of plug-and-play architectures at the database level is something that has received little attention to date within the research community. A lot of emphasis has been placed recently on service-oriented architectures and specifically the use of web services, but this is more suited to integration and orchestration at higher levels, especially in heterogeneous environments. We want to be able to integrate components *within* the database in order that we could, for example, introduce constraints and triggers over these components as well as executing queries over them. This in turn would enable the integration of data from different components required for personal mashups to be performed within the database. Currently, we are in the process of formulating precisely a notion of a database component and designing an architecture and mechanism to support this concept. Also, since users should be able to, not only select components, but also develop their own components and personal mashups, we need to investigate how this can best be supported through declarative languages and graphical tools.

To provide improved PIM systems, it is important that the underlying data management platform is based on a semantic data model. Specifically, it should be able to support rich classification structures, versions, constraints, triggers

and associations as well as a declarative query language. The ability to support multiple classification is particularly useful as a basic means of specifying which information objects should be published to which Web 2.0 applications. We are using the OMS Avon system [19] as a data management platform for PIM 2.0 since it supports these concepts. OMS Avon provides a semantic data management layer on top of the object database engine db4o[4]. In OMS Avon, all data—application, metadata and system objects—are handled uniformly and the system is bootstrapped from a core metamodel. This provides the basis for its flexibility in being able to integrate new concepts required to meet the demands of emerging domains as has been done previously for web engineering [20], peer-to-peer data management [21] and context-awareness [22,23]. The implementation of the database will be based on a PIM 2.0 metamodel which in turn will take into account the database component concept under development.

An important part of the architecture is the mechanism used to support the various forms of integration and synchronisation required. On the one hand, there needs to be some form of data synchronisation between PIM 2.0 and the Web 2.0 applications that will be the basis behind the publishing of personal information in the Web 2.0 applications. Thus changes to the data in PIM 2.0 should propagate to all Web 2.0 applications that are registered as using that data. We may also want bilateral synchronisation which means that it should also be possible to propagate changes to data in the Web 2.0 applications to PIM 2.0. An example of this would be propagating changes to the contacts information in PIM 2.0 if the corresponding data has been updated in the Web 2.0 application. On the other hand, there also needs to be integration and synchronisation of data within PIM 2.0 across database components. For example, a personal mashup application may integrate data from two or more database components.

The PIM 2.0 architecture that we propose exemplifies the various forms of data integration and synchronisation that are found in many forms of modern distributed information systems, especially those based on web technologies. Therefore it is important to develop general mechanisms that are flexible enough to meet these requirements and can be customised to specific settings. We want to use this project to investigate how we can achieve a general model and associated mechanisms for data integration and synchronisation that can be applied both within object databases and between object databases and external data sources. We therefore propose to investigate how we can generalise and extend the generic proxy mechanism that we recently developed for the integration and synchronisation of OMS Avon databases [24] with external data sources to these more general architectures.

An advantage of the generic proxy approach is that it allows the details of how and when synchronisation takes place to be customised through proxy processes. Also, it supports integration at the database level, which again means that we can leverage database fucntionality and semantics. The generic proxy mechanism was developed for object-oriented databases and we will need to consider how the concept can be adapted and extended to cater for situations where the

[4] http://www.db4o.com

data sources are non-OODBMS and possibly heterogeneous. In particular, we need to investigate in detail how we can interface with Web 2.0 applications to achieve bilateral synchronisation. Another key issue is how to ensure that the mechanisms are efficient.

The PIM 2.0 project is in its first phases and there are many open issues. An initial prototype that allows data stored and managed in a personal data space to be published to one or more Web 2.0 applications has already been developed [25]. In the next stage, we will implement a second prototype based on the concept of database components and the plug-and-play architecture. In addition to the issues mentioned above, an important aspect of the project will be the means for users to specify how and where data should be published. Currently we are developing a simple language that can be used to specify the necessary data mappings and also modes of synchronisation. Later, we will design and experiment with various tools to allow these to be specified graphically.

Alongside the technical work, we plan to carry out various user studies. These will cover the use of Web 2.0 applications for personal information management as well as evaluations on the system and tools that we will develop.

5 Conclusions

We have discussed the issue of personal information management in the realm of Web 2.0 and how the problem of information fragmentation has now extended beyond the desktop. We make the case for an architecture that supports a clear separation of concerns between the management of data and the publishing of data. The proposed system PIM 2.0 provides an integrated solution for personal information management based on the Web 2.0 paradigm of a portal with a plug-and-play architecture. The publishing of data to Web 2.0 applications is controlled through a bilateral synchronisation mechanism that also offers the possible automatic importation of data published by other users into a personal information space. Central to the plug-and-play architecture is a notion of database components that allow personal information spaces to be constructed in a modular way.

The concepts presented in the paper are still under discussion and the PIM 2.0 system is in the early stages of design and implementation. However, we are optimistic that significant advances in PIM systems can be achieved by learning from the success of Web 2.0.

Acknowledgements

Many members of the Global Information Systems group at ETH Zurich have contributed to the ideas expressed in this paper. Special thanks are due to Michael Grossniklaus and Stefania Leone who are leading the PIM 2.0 project and Martin Schnyder who has implemented a first prototype.

References

1. O'Reilly, T.: What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software (2005), http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html
2. Norrie, M.C.: Client-Server Database Architectures for the Web. In: Masunga, Y., Spaccapietra, S. (eds.) Advances in Multimedia and Databases for the New Century, A Swiss/Japanese Perspective (2000)
3. Doan, A., Ramakrishnan, R., Chen, F., DeRose, P., Lee, Y., McCann, R., Sayyadian, M., Shen, W.: Community Information Managemnt. Data Engineering Bulletin 29(1) (2006)
4. DeRose, P., Shen, W., Chen, F., Lee, Y., Burdick, D., Doan, A., Ramakrishnan, R.: DBLife: A Community Information Management Platform for the Database Research Community. In: Proceedings of CIDR 2007 (Demo) (2007)
5. Ceri, S., Fraternali, P., Bongio, A., Brambilla, M., Comai, S., Matera, M.: Designing Data-Intensive Web Applications. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers Inc., San Francisco (2002)
6. Frăsincar, F., Barna, P., Houben, G.J., Fiala, Z.: Adaptation and Reuse in Designing Web Information Systems. In: Proceedings of International Conference on Information Technology: Coding and Computing, Las Vegas, NV, USA, April 2-4 (2004)
7. Casteleyn, S., De Troyer, O., Brockmans, S.: Design Time Support for Adaptive Behavior in Web Sites. In: Proceedings of ACM Symposium on Applied Computing, Melbourne, FL, USA, March 9-12 (2003)
8. Rossi, G., Schwabe, D., Guimarães, R.: Designing Personalized Web Applications. In: Proceedings of International World Wide Web Conference, Hong Kong, China, May 1-5 (2001)
9. Garrigós, I., Casteleyn, S., Gómez, J.: A Structured Approach to Personalize Websites Using the OO-H Personalization Framework. In: Proc. Asia Pacific Web Conf., Shanghai, China (2005)
10. Schewe, K.D., Thalheim, B.: Reasoning About Web Information Systems Using Story Algebras. In: Proceedings of East-European Conference on Advances in Databases and Information Systems, Budapest, Hungary, September 22-25, 2004, pp. 54–66 (2004)
11. Koch, N.: Software Engineering for Adaptive Hypermedia Systems. PhD thesis, Ludwig-Maximilians-University Munich, Munich, Germany (2001)
12. Ceri, S., Daniel, F., Matera, M., Facca, F.M.: Model-driven Development of Context-Aware Web Applications. ACM Transactions on Internet Technology 7(2) (2007)
13. Ceri, S., Daniel, F., Facca, F.M., Matera, M.: Model-driven Engineering of Active Context-Awareness.In: World Wide Web (2007)
14. Tungara, M., Pyla, P., Sampat, M., Perez-Quinones, M.: Defragmenting Information using the Syncables Framework. In: SIGIR Workshop on Personal Information Management (2006)
15. Bellotti, V., Smith, J.: Informing the Design of an Information Management System with Iterative Fieldwork. In: Proceedings of the Conference on Designing Interactive Systems (DIS 2000) (2000)
16. Dourish, P., Edwards, W., LaMarca, A., Salisbury, M.: Presto: An Experimental Architecture for Fluid Interactive Document Spaces. ACM Transactions on Computer Human Interaction 6(2) (1999)

17. Franklin, M., Halevy, A., Maier, D.: From Databases to Dataspaces: A New Abstraction for Information Management. ACM SIGMOD Record (December 2005)
18. Dittrich, J.P., Blunschi, L., Färber, M., Girard, O.R., Karakashian, S.K., Salles, M.A.V.: From Personal Desktops to Personal Dataspaces: A Report on Building the iMeMex Personal Dataspace Management System. In: BTW 2007 (2007)
19. Norrie, M.C., Grossniklaus, M., Decurtins, C., de Spindler, A., Vancea, A., Leone, S.: Semantic Data Management for db4o. In: Proceedings of 1st International Conference on Object Databases (ICOODB 2008), Berlin, Germany (March 2008)
20. Grossniklaus, M., Norrie, M.C., Signer, B., Weibel, N.: Producing Interactive Paper Documents based on Multi-Channel Content Publishing. In: Proceedings of International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution, Barcelona, Spain, November 28-30 (2007)
21. Norrie, M.C., Palinginis, A.: A Modelling Approach to the Realisation of Modular Information Spaces. In: Bressan, S., Chaudhri, A.B., Li Lee, M., Yu, J.X., Lacroix, Z. (eds.) CAiSE 2002 and VLDB 2002. LNCS, vol. 2590. Springer, Heidelberg (2003)
22. Grossniklaus, M., Norrie, M.C.: Using Object Variants to Support Context-Aware Interactions. In: Proceedings of International Workshop on Adaptation and Evolution in Web Systems Engineering, Como, Italy, July 19 (2007)
23. Norrie, M.C., Signer, B., Grossniklaus, M., Belotti, R., Decurtins, C., Weibel, N.: Context-Aware Platform for Mobile Data Management. Wireless Networks 13(6) (2007)
24. Vancea, A., Grossniklaus, M., Norrie, M.C.: Database-Driven Web Mashups. In: Proceedings 8th International Conference on Web Engineering (ICWE 2008), New York, USA (July 2008)
25. Leone, S., Grossniklaus, M., Norrie, M.C.: Architecture for Integrating Desktop and Web 2.0 Data Management. In: Proceedings 7th International Workshop on Web-Oriented Software Technologies (IWWOST 2008), New York, USA (July 2008)