Contextualizing bookmarks:
An approach based on user context to improve organization and retrieval of bookmarks

Hyeon Kyeong Hwang[1]

[1] Doctoral School of Information and Communication Technology,
University of Trento, Via Sommarive, 9
I-38123 Povo (TN), Italy
hyeonkyeong.hwang@unitn.it

Abstract. The ubiquitous nature of the Internet and the connectivity it provides has rendered a vast amount of information accessible on demand. Consequently an increasing number of people are drawn to this new source of knowledge not just out of curiosity but also for their studies or work. Thus it cannot be overstated that being able to manage efficiently and effectively what we find or learn from the web resources is crucial. The most popular way of preserving such knowledge from the web is through bookmarking web pages. Bookmarks, however, would serve little purpose if they cannot be easily found for re-use. In this paper, we examine current bookmarking tools with the focus on their indexing mechanism and discuss the role and importance of user context in information search and retrieval. Then we propose a new bookmarking tool that offers user context information as additional retrieval cue on top of existing semantic cues such as tags, annotation and formal ontology.

Keywords: Personal Knowledge Management, Social Bookmarking, Context-based Information Retrieval, Personal Web Annotation

1   Introduction

Internet and World Wide Web have touched and re-shaped almost all aspects of our everyday life. Knowledge seeking and acquiring is not an exception to this phenomenon: according to Internet Live Stats[1], on average 40,000 search queries are handled by Google[2] for each second, which accumulates to a staggering 3.5 billion searches per day worldwide. A study conducted by Savolainen and Kari in Finland showed that people considered Internet as an important primary information source for self-development purposes owing to "easy accessibility, currency, interactivity, and the broad repertoire of information" [1]. Not surprisingly most web users often revisit information found before. Web user survey conducted by Herder [2] exhibited that 51% of web pages were re-accessed on average. More recently, Teevan et al. analyzed one-year web query of Yahoo's log and discovered that 40% of all queries were for re-accessing visited pages [3], demonstrating that users not only have the need to search for new information but also to keep information they deem "valuable" for re-use at a later time. To cope with this demand, various research efforts have been taken in order to provide a more efficient and meaningful tool for managing knowledge mined from the Internet at the personal level. Most common and popular method is to "bookmark" web pages to organize and access them posthumously. One of such bookmarking tools, which have gained immense popularity in Web 2.0 era, is the social bookmarking. Social bookmarking tools help users to create, manage and share bookmarks

---

[1] www.internetlivestats.com
[2] www.google.com

[4] of Internet resources by indexing web pages by folksonomy, known as "tags". Tags offer personalized classification of bookmarks by permitting users to add labels they find most relevant to the web resources. Nevertheless some researchers have observed that most web users do not use bookmarks to re-access information they have stored [5][6]. More interestingly, using tags[3] does not seem to mitigate the limitation of using folders (i.e., users cannot remember which folder contains information they seek). Several comparative studies done on folders and tags showed no significant difference in retrieval performance [7][8][9]. If neither folders nor tags solve the problem of re-accessing saved bookmarks, what could be done to help users easily to find them when they need? This consequently leads us to examine more closely how human brains manage information, especially its retrieval process of memory. Context[4] seems to play a very important role in bringing back memory in human brains as purported by researchers in the field of Psychology [10][11][12]. Most notably, the theory of context-dependent memory contends that context information is recorded along with target information in such way that target information can be retrieved better when context is presented as cues [11]. Using context in information systems is not new, however. Context has long been recognized as key importance in numerous fields (i.e. computational linguistics, mobile and pervasive computing, automatic image analysis, information retrieval) [13]. Especially in information retrieval, context is crucial because it helps to provide expedient delivery of content relevant to people's information needs [14]. Notwithstanding proven significance of user context in information search and retrieval, context has not been considered as an indexing mechanism of bookmarking tools as of today. To put this missing piece into the puzzle of easier retrieval of bookmarks, we propose a new bookmarking tool that encapsulates user context in which bookmarks are created and uses it for search and retrieval.

The rest of this paper is organized as follows: section 2 presents a review of different state-of-the-art approaches in existing bookmarking tools and use of context in information systems. Section 3 identifies shortcomings of current bookmarking methods and discusses how using context may bridge the gap. Then in section 4, we propose a new approach that exploits context, ontology and user's annotation for a more holistic way of managing personal bookmarks. The conclusion and contribution is provided in section 5.

2   State of the Art

In this section, we review the state-of-art in research fields related to effective management of web information with the focus on search and retrieval: indexing mechanisms of current bookmarking tools and use of context in information retrieval systems.

2.1   Bookmarking in Web 2.0 era

Bookmarks are Uniform Resource Identifiers (URI) of web resources saved for retrieval at a later time, typically through built-in tools provided by various web browsers[5] or 3rd parties. According to an early survey done by Pitkow and Kehoe [15], over 80% of 6,619 web users responded they used bookmarks to locate information found on the web and more recently a survey done in Finland showed 92.4% of respondents used bookmarks in their primary

---

[3] Tags here do not necessarily mean those of folksonomy but are any "keywords" or "labels", both user-defined and otherwise.

[4] Context is a very broad term and its definitions are proliferated. In this paper, we limit the meaning of context to "any information that can be used to characterize the situation of an entity" [17], with the entity being the user and the source of information being both internal and external to the user.

[5] Some browsers call bookmarks as favorites. For simplicity, we will use bookmark(s) as the generic term for functionality that allows saving and retrieving of web resources

browser [6]. Moreover, Abrams et al. [16] discovered that number of bookmarks of a user and time had a positive linear relationship, meaning that the size of bookmark collection increased over time. The results of these surveys indicate that bookmarking is one of the representative activities in which a user attempts to manage knowledge discovered on the web. Having this in mind, we review the current approaches of bookmarking tools in next subsections.

2.1.1 Folksonomy-based approach

Folksonomy, by definition, is a composite term of "folk" and "taxonomy" and a "kind of user creation of metadata" [17]. As opposed to classical taxonomy of folders for resource organization, Folksonomy uses user-defined keywords, "tags", for indexing and retrieval. Tags not only give users unlimited freedom in choosing the keywords that are most significant in their point of view but also minimize the "cost of participation", [4] typically met with formal ontology, because users do not need in-depth knowledge of existing ontology or taxonomy to meaningfully organize resources. Thanks to its easy-of-use and high degree of personalization, tagging is widely popular among users and its implementation can be found in numerous bookmarking tools. The most well known ones are Diigo[6] and Delicious[7]. Delicious was the first tool to offer the tags-based indexing system [18]. Tags are classified non-hierarchically: meaning establishing relations among the tags is not possible. Similarly Diigo is a social bookmarking system that offers a set of "recommended" tags to each saved page, which are based on the web content. Despite its wide acceptance, tags do not come without price. Because of the very reason for its popularity – high degree of arbitrary personalization – the search and retrieval of resources become a challenge. Moreover, a study carried out by Bergman et al. showed that users prefer navigation to search [19]. In fact, both Diigo and Delicious provide users with folder-like functions to organize their tagged resources.

2.1.2 Annotation-based approach

Annotation is the act of adding a metadata to an object. With the increasing availability of Internet, many users started adding annotations to the web resources in various forms. According to the study conducted by Fu et al [20], most users created annotations on web pages in forms of text selection for emphasis and notes or symbols for building association with the main content. These annotations, beyond their physicality, represent user's interpretation of the given content, showing "what" user has found valuable out of the whole. According to Kawase et al., annotations are a more efficient way of re-finding information. Their user experiment showed that users spent the least time in relocating information with the annotation than using search engines or bookmarks [21]. Currently web annotation tools are found mostly in educational and collaborative settings. For instance, Azouaou and Cyrille [22] proposed MEMONOTE, an annotation-based PKM tool for teachers meanwhile Shukla and Chaudhary [23] suggested another for research scholars. Nonetheless, general-purpose tools are also available for web annotations (i.e., A.nnotate[8], Crocodoc[9]), which allow users to highlight, add notes and share their annotations with others. Diigo and Scrible[10] are two most popular bookmarking tools that provide annotation functionalities.

2.2   Context in information systems

With the advent of mobile and pervasive devices, much attention has been given to incorporating context in information systems. Goker et al. argues that context information "provides an important basis for identifying and understanding" the information needs of users [14]. The extent of user context used in systems varies but it can be largely divided into "external" and "internal" [24]. While external context deals with physical environment factors (i.e. location, time, temperature, light, audio, and motion), internal context deals with logical factors implicitly expressed by user's activities (i.e. Google search). Most context-aware systems of today are centered on using external context. Although they have shown a degree of success in some areas (i.e. mobile computing and smart spaces) [25], internal context could play a major role in providing more relevant information to users. As a matter of fact, recent research is geared towards combining both external and internal context [24]. Use of context in bookmarking is still novel. However, a few researchers have conducted experiments with context-based tools to help users re-find information in recent years. YouPivot, proposed by Hailpern et al. [26], brings a new dimension to browsing history search by adding user specific context for easier retrieval of visited web pages. Users can pick a contextually related activity to "pivot" and view web pages visited in the close time frame before and after the pivoted activity. YouPivot also provides a special time-annotation method called "TimeMarks" with which users can mark a specific activity or time as to-be-remembered. Their pilot study showed that users' strong preference to using YouPivot and improved performance in re-finding visited web pages. Nevertheless, YouPivot's context is rather limited: it considers only time and the name of activity (i.e. Facebook). Another tool, "ReFinder", proposed by Tangjian et al. takes a "context memory snapshot" [27] of each saved web page or desktop file. This snapshot includes time, place and concurrent activity and allows users to query web pages and desktop files based on context. As with YouPivot, user study showed positive results: users re-found files and web pages more quickly [27]. Although ReFinder is a powerful tool that allows users to add context not only to web pages but also to local files but it is not without drawbacks: it lacks the breadth and depth of user context by limiting it to location, time (only in date-time format), and activity. Moreover, the activity does not reflect current user activity but is an arbitrary list of to-do or categorization. Still the greatest caveat could be that users must manually insert all context information, which was found to be cumbersome in user study [27].

3   Problem statement

Re-finding the right information can be painstakingly difficult as the sheer amount of information constantly increases. A research group at University of Washington in collaboration with Microsoft Research[11] defines this problem as "Keeping Found Things Found" (KFTF[TM]). According to their study, even though bookmarking is commonly used, most users do not rely on bookmarks to re-find the information they need but resort to other methods [5]. Similar survey done by Aula et al. [6] in Finland validates the findings of [5]: the median value of frequency of use of bookmarks as re-access strategy was "sometimes" despite the fact that an overwhelming 92.4% of the respondents were using bookmarks in their primary browser. The main reason for not using the bookmarks to re-access web pages stems from difficulty in locating the information once the collection of bookmarks becomes large. In a recent study of 236 experienced web users, 79.2% of respondents had more than 50 bookmarks, with average of 220 bookmarks and 29.7 folders per user [6]. As seen in this study, users resort to giving their ever-growing bookmarks a form of structure, typically folders, to aid them in managing and retrieving information. However, "folders can obscure

---

[11] http://kftf.ischool.washington.edu/index.htm

as well as organize" [5]. The classic limitation of using folders is that users tend to forget where to look when they need a bookmark hidden from the naked eye. This is caused by folder name not necessarily reflecting fully the content they include as it assumes one-to-many relationship (i.e., one folder can have many documents). Using tags may alleviate this problem by allowing multiple keywords to be associated with single content. Nevertheless users still seem to experience frustration when re-locating information with tags as revealed in a study geared by Civan et al. [7]. They discovered that folders and tags yielded similar retrieval performance and that some users claimed using multiple categorizations [tags] "could cause confusion, redundancy, and inefficiency" [7]. Moreover, Wetzker et al. [28] conducted analysis of approx. 150 million bookmarks downloaded from Delicious and found out that users tended to use an increasing number of tags per bookmark for better organization and retrieval since few tags offered little distinction of each bookmark, which may hinder search at a later time. As these evidences point, if neither folders nor tags solve the problem of re-accessing saved bookmarks, what could be done to help users easily find them when they need? In 1995, Barreau & Nardy purported that people prefer location-based methods for searching files on desktop [29] and this has been proven to be true also for bookmarks by recent studies. Bergman et al. discovered that direct search was the last resort for many people, after failing to navigate to target information [30] via digital location and a cross-tool (file, email and bookmark). A study of personal information management by Boardman and Sasse revealed that participants demonstrated "a strong preference for browsing over search in all tools" [31]. The reason why people prefer "browsing" (i.e. folders) to "direct search" (i.e. tags) could be found in the way how human brains work. Browsing requires users to select a specific location and then scan content for recognition whereas direct search calls for users to remember the exact query terms. Recognition is much less labor-intensive process than "recall" in memory management, hence one of the basis of Graphic User Interface design is that "see and choose is easier than recall and type" [32]. This thesis, consequently, aims to answer following questions.

- Could user context be offered as additional browsing cue to search and retrieve bookmarks more efficiently and effectively?
- In that case, what kind of user context would prove to be most useful?

In the next section, we propose a prototype system that exploits advantages of both personalized (folksonomy and annotation-driven) and formal (ontology-driven) approaches, under the umbrella of user context.
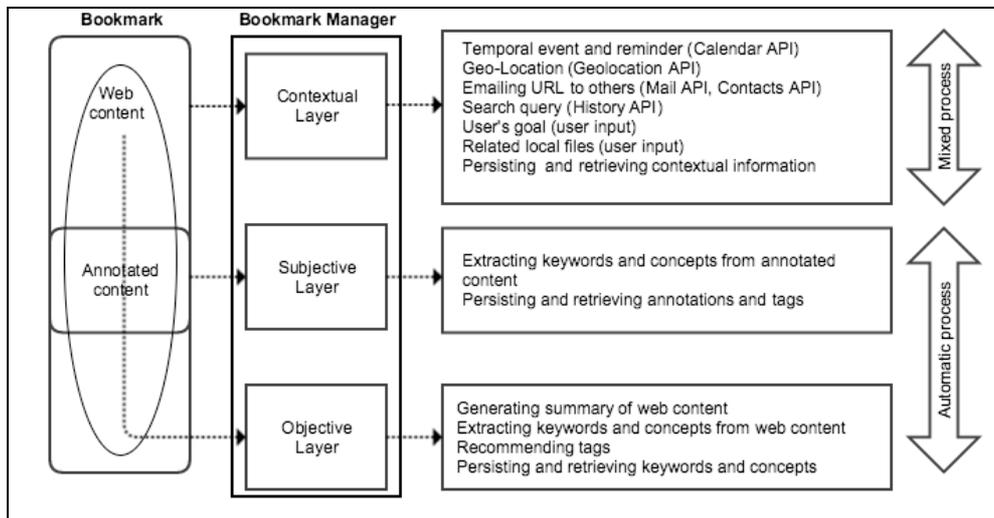

4   Proposed solution

The proposed solution harnesses positive benefits both from existing approaches (annotation, ontology and tags) and adds on contextual cues to index each bookmarked web page. The bookmarking tool will be developed as a Chrome extension application built using HTML5 and Javascript with communication with server over HTTP. When user clicks on the app button from Chrome browser bar to save a new bookmark, he or she will be shown a bookmark dialog box where he or she can view or edit contextual information, add or choose recommended tags. In the main home page, user can search saved bookmarks using various contextual cues, tags or concepts automatically extracted from annotated text or content of saved web pages.

4.1   System Architecture

The main processing of each bookmark is done through the "Bookmark manager" running on the backend. Bookmark manager handles processing and extracting all metadata of each bookmark. This processing consists of three distinct parts (or layers), which work together to ensure easy retrieval of bookmarks in multi-dimensional way. The first part is the contextual

layer. Contextual layer is concerned with automatically extracting and saving the context including external (location and time) and internal (user's intention) factors. We assume that user's intention could be captured by 1) search query used to find the target web page, 2) local file for which the web page is useful or related to, 3) a future event for which the target web page needs to be re-accessed, and 4) user-provided goal for which the target web page is being saved (i.e personal projects, exams or research). Context layer persists and retrieves aggregated context information to be used in search. Context data itself will be modeled using a formal ontology, which makes it possible to do reasoning and find relations among the bookmarks or, going further, to support "similar-context" bookmarks among different users. Apart from context information, users can also add annotations or tags to the web page. These are processed by the other two layers called "Subjective" and "Objective" layers. Essentially, these handle extracting keywords and concepts from the web content using formal ontology extraction tools. The subjective layer, hence annotations, will handle saving of the annotations and the extraction from the annotated text and the objective layer from the whole content of the web page. The extracted concepts and keywords are provided to users as recommended tags and also constitute to creating an individual ontological graph of each user, providing a visualized knowledge map to navigate through. The overall architecture of the system is provided in **Fig. 1** below.

Fig. 1. The overall architecture of proposed bookmarking tool

Proposed bookmarking tool can provide users with powerful browsing-oriented search by any fragment of context, tags or annotations appended to the target page. Additionally it will create a personalized ontological diagram that allows users to view, browse and discern his or her domain-based dynamics of saved knowledge resources.

5.   Conclusion and expected contribution

The new approach put forward in this paper is based on ontology, annotation and context. It not only embraces the advantages of existing ontology and personalization-based approaches but also considers the importance of user's context when a web resource is bookmarked in order to ensure easier and faster retrieval of saved web resources. In fact, the human brains store information along with the contextual cues and they are vital for retrieval, which opens a sea of possibilities for this novel approach. By providing user context information as cue, we plan to discern the significance and impact of context in bookmark retrieval and which

type of user context is most useful for that task, which can bring about substantial changes in personalized information search and retrieval. Furthermore, our proposed tool can also allow users to share their bookmarks with others based on various context information by using ontological context model. That could provide us a great insight into most popular type of information users search and save set in various contexts. To affirm and realize this potential, however, a carefully planned empirical experiment would be indispensable.

References

[1] 1. Savolainen, R. and Kari, J.: Placing the Internet in information source horizons: A study of information seeking by Internet users in the context of self-development. In: Hernon, P., Schwartz, C. (eds.) Library & Information Science Research, vol. 26., pp.415-433. Elsevier (2004)

[2] Herder, E.: Characterizations of user Web revisit behavior. In Proceedings of Workshop on Adaptivity and User Modeling in Interactive Systems (2005)

[3] Teevan, J., Adar, E., Jones, R. and Potts, M.A.S: Information Re-Retrieval: Repeat Queries in Yahoo's Logs. In: SIGIR '07 Proceedings of the 30th Annual International ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 151-158. ACM, New York (2007)

[4] Noll, M., Meinel, C.: Web Search Personalization via Social Bookmarking and Tagging. In: Proceedings of the 6th International Semantic Web Conference (ISWC) & 2nd Asian Semantic Web Conference (ASWC), Pusan, South Korea 2007, LNCS vol.4825, pp. 367-380. Springer, Heidelberg (2007)

[5] Bruce H., Jones, W., and Dumais, S.: Keeping and re-finding information on the web: What do people do and what do they need? In Proceedings of 67th ASIS Annual Meeting (October 2004)

[6] Aula, A., Jhaveri, N., and Kaki, M.: Information Search and Re-access Strategies of Experienced Web Users. In: Proc. 14th Int'l Conf. World Wide Web, pp. 583-592, ACM Press, New York (2005)

[7] Civan A., Jones, W., Klasnja, P, and Bruce H.: Better to Organize Personal Information by Folders Or by Tags?: The Devil is in the Details", In: Proceedings of 68th ASIST Annual meeting, Columbus, OH (2008)

[8] Pak, R., Pautz, S., and Iden, R.: Information organization and retrieval: A comparison of taxonomical and tagging systems. In: Cognitive Technology, vol.12(1), pp. 33-44 (2007)

[9] Ma, S. and Wiedenbeck, S.: File management with hierarchical folders and tags. In: Proc. 27th SIGCHI Conference on Human Factors in Computing Systems (CHI 2009) Extended Abstracts, pp.3745-3750. ACM Press, New Yock (Apr. 2009)

[10] Tulving, E., Donald, T.: Encoding specificity and retrieval processes in episodic memory. In: Psychological Review, vol.80 (5), pp. 352-373. American Psychological Association (1973)

[11] Smith, S. M.: Theoretical principles of context-dependent memory. In: Gruneberg M. M., Morris, P. E. (eds.) Theoretical Aspects of Memory, pp.168. Routledge, London (1994)

[12] Hannon, B., Fergus, I. M.: Encoding Specificity Revisited: The Role of Semantics. In: Canadian Journal of Experimental Psychology, vol. 55 (3), pp. 231- 243. Canadian Psychological Association (2001)

[13] Mylonas, Ph., Vallet, D., Castells P., Fernandez, M, and Avrithis, Y.: Personalized information retrieval based on context and ontological knowledge. In: The Knowledge Engineering Review, vol. 00:0, pp. 1-24, Cambridge University Press, UK (2004)

[14] Goker, A., Myrhaug, H., Bierig R.: Context and Information Retrieval. In: Goker, A., Davies, J. (eds.) Information Retrieval: Searching in the 21st Century, pp. 131-155. Jon Wiley & sons (2009)

[15] Pitkow, James E., Kehoe, Colleen M.: Emerging trends in the WWW user population. In: Communications of the ACM, vol.39 (6), pp. 106-108. ACM Press, New York (1996)

[16] Abrams, D., Baecker, R., Chignell, M.: Information archiving with bookmarks: personal web space construction and organization. In: CHI '98 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 41-48. ACM Press, New York (1998)

[17] Wu, X., Zhang, L., Yu, Y.: Exploring Social Annotations for the Semantic Web. In: WWW '06 Proceedings of the 15th International Conference on World Wide Web, pp. 417-426. ACM Press, New York (2006)

[18] Mathes A.: Folksonomies – cooperative classification and communication through shared metadata", Computer Mediated Communication, vol.47(10), pp. 1-13. International Communication Association (2004)

[19] Bergman, O., Beyth-Marom, R., Nachmias, N., and Whittaker, S.: Improved search engines and navigation preference in personal information management. In: ACM Transactions on Information Systems, vol.26(4), pp.1-24. ACM Press, New York (2008)

[20] Fu, X., Ciszek, T., Marchionini, G., Solomon, P.: Annotating the web: An exploratory study of web users' needs for personal annotation tools. In: 68th Annual Meeting of the American Society for Information Science & Technology (ASIS&T), Charlotte, NC, USA (2005)

[21] Kawase, R., Papadakis, G., Herder, E., and Nejdl, W.: The Impact of Bookmarks and Annotations on Refinding Information. In: HT'10, Proceedings of the 21st ACM conference on Hypertext and hypermedia. pp. 29-34. ACM Press, New York (2010)

[22] Azouaou, F., Desmoulins, C.: MemoNote, an annotation-based personal knowledge management tool for teachers. In: IADIS International Conference, Cognition and Exploratory Learning in Digital Age (ELDA 2006), Barcelona, Spain (2006)

[23] Shukla, A., Chaudhary, B. D.: Annotation based personal knowledge management of research scholars. In: Engineering Education: Innovative Practices and Future Trends (ALCERA), 2012 IEEE International Conference. Kottayam, India (2012)

[24] Prekop, P. and Burnett, M.: Activities, context and ubiquitous computing. In: Special Issue on Ubiquitous Computing Computer Communications, vol. 26(11), pp. 330-334. Elsevier (2003)

[25] Ongenae, F., Duysburgh, P., Verstraete, M., Sulman, N., Bleumers, L., Jacobs, A., Ackaert, A., De Zutter, S., Verstichel, S., De Turck, F.: User-driven design of a context-aware system: an ambient-intelligent nurse call system. In: Proceedings of 6th International Conference on Pervasive Computing Technology for Healthcare and Workshops, pp. 205-210, San Diego, CA, USA (2012)

[26] Halipern, J., Jitkoff, N., Warr, A., Karahalios, K., Sesek, R. and Shkrob, N.: YouPivot: Improving Recall with Contextual Search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, May 07-12, 2011, Vancouver, BC, Canada (2011)

[27] Deng, T., Zhao, L., Wang, H., Liu Q. and Feng, L.: ReFinder: A Context-Based Information Refinding System. In: IEEE Transactions on Knowledge and Data Engineering, vol. 25(9) (September 2013)

[28] Wetzker R., Zimmermann, C., Bauckhage C.: Detecting Trends in Social Bookmarking Systems: A del.icio.us Endeavor. In: Taniar. D. (ed.) Exploring Advances in Interdisciplinary Data Mining and Analytics: New Trends, pp. 34-51. IGI Global (2011)

[29] Barreau, D. and Nardi, B.: Finding and reminding: File organization from the desktop. In: SIGCHI Bulletin, Vol.27(3), pp. 39-43. ACM Press, New York (1995)

[30] Bergman, O., Beyth-Marom R., Machmias, R., and Whittaker, S.: The user-subjective approach: a new direction for PIM systems design. In: Proceedings of 3rd International CHI workshop on Personal Information Management (PIM '08), Florence, Italy (2008)

[31] Boardman, R. and Sasse, M. A.: Stuff Goes into the Computer and Doesn't Come Out – A Cross-tool Study of Personal Information Management. In: CHI '04 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 583-590. ACM, New York (2004)

[32] Johnson, J.: Designing with the Mind in Mind – Simple Guide to Understanding User Interface Design Rules, pp. 109-117. Morgan Kaufmann Publishers, Burlington, USA. Printed in China (2010)