Re-thinking Bookmark Management –
Less Choice Is More Efficient

Siu-Tsen Shen[1] and Stephen D. Prior[2]

[1] Department of Multimedia Design, National Formosa University,
64 Wen-Hua Rd, Hu-Wei 63208, Taiwan
[2] Faculty of Engineering and the Environment, University of Southampton,
Highfield Campus, Southampton SO17 1BJ, United Kingdom
stshen@nfu.edu.tw, s.d.prior@soton.ac.uk

Abstract. This research investigates the role of a Controlled Vocabulary (CV) in next generation bookmark management systems. The search for a more efficient graphical user interface solution to deal with the massive information overload situation faced by most computer users today is a pressing problem. CVs allow categorization of title words and phrases into the appropriate location recognized by the user, so as to facilitate easier information storage and retrieval. The results of this user study involving 152 individuals indicated that there is potential for a well-defined two-tier controlled vocabulary system to assist user categorization, information storage and retrieval in personal information management systems.

Keywords: Bookmark management, web browser, controlled vocabulary, filing system, information storage and retrieval.

1 Introduction to Categorization

Categorization is heavily dependent on the user's cognitive ability to gather and store information to and from the desktop, web browsers, and hierarchical filing systems. However, several studies have pointed out the difficulty of categorization [1-4]. Not only is it difficult to describe each individual's capability in an explicit way, but also user behavior tends to be more flexible than that of a typical hierarchical structure [5-6]. A reason posited for this, is that categorization and filing systems are opaque to the user, because of their austere classification schema that brings with it opacity [7]. Other studies have also shown that categorization schemes are unstable and difficult to define [8-10]. It has been reported that filing systems with their rigorous hierarchical mechanisms, such as 'My Documents' and 'Bookmark' functions have severe usability problems in personal information management [11-15].

1.1 Relevant Work

Malone (1983) conducted interviews with ten office workers into how they organized their desktops, and offered suggestions for the future design of computer-based

information systems. He pointed out that the tasks in finding and reminding desired information are very important to desktop organization. Furthermore, the capability of classifying information relies on an individual's cognitive ability [1]. A key to success is the ability to appropriately label and store files and folders into the categories where they belong. This is crucial in order to be able to retrieve them at a later date.

Research conducted over the last two decades, has attempted to improve the existing filing systems from four different perspectives:

1. Cognitive, memory and contextual aspects.
2. Task management examples in email systems and mobile devices.
3. User adoption and patterns.
4. Time-based example.

Czerwinski and Horvitz (2002) developed two prototypes of the reminding systems (i.e. PersonalStream and PersonalStills) to test users' memory, and it was found that users could memorize the temporal episodes within a maximum of one month time frame. Nevertheless, users could bring back their memory with the help of video reminders, even after a month had elapsed [16].

TaskMaster, an emailing system, includes carrying tasks not only with threaded collections of messages files and drafts, but also with previewing the content of the email messages such as documents and web pages [17]. However, the current computer system is not 'smart' enough to deduce appropriate contextual tasks for users. Moreover, it is possible to make information retrieval more effective, if we could re-examine the associated matters in a proper, well-defined contextual infrastructure.

The great popularity of recent products, such as Apple's iPhone/iPad and the HTC smart phone, with their high level of touch screen resolution shows that the margin between the traditional office-based tools either for real objects or digital ones, and up to date tools is getting smaller in terms of user preference. In other words, the user adoption of the latest technological applications, whether it is high or low, will vary by the different context of use. For example, users will have to adapt to using Facebook and Twitter because of the need for social networking either on their laptops or mobile phones in order to share photos and videos.

Freeman (1997) proposed Lifestreams to support data retrieval tasks. The Lifestreams interface is a time ordered stream of documents that is like a dairy of your life, whereby you could store your documents that you create and others send you [18]. More recent documents appear to be in the front of the screen, while the older ones stay in the back of the screen. It provides a personal collection of saved documents that could be viewed through filtering based on time.

1.2   Controlled Vocabularies as Applied to Information Management

Controlled Vocabularies (CVs) are restricted systems of textual tags and are normally applied for large datasets. An example of the use of controlled vocabularies is the

International Patent Classification (IPC) system, which offers a hierarchical system of language independent representations for the classification of intellectual property. There are eight major subject headings (A-H), into which each patent application has to be classified [19]. In total there are over 70,000 IPC codes for different technical areas (see Table 1). The major object of using controlled vocabularies is to maintain consistency in the description of content objects and to facilitate retrieval.

Table 1. An Example of the International Patent Coding Structure

Sub Heading Code	Description
G	Physics
G02	Optics
G02C	Spectacles, sunglasses or goggles
G02C5	Construction of non-optical parts
G02C5/20	Adjustable, telescopic

Having conducted an extensive search, the authors could not find a single reference to an application of a controlled vocabulary to the design of a web browser user interface. However, there have been several researchers that have explored the use of a controlled vocabulary approach in other types of information retrieval scenarios. Svenonius (1986) investigated the effectiveness of using free-text versus a controlled vocabulary in information retrieval in the field of library science, and suggested that it would depend on the ability of the user's search precision and recall [20]. Butler (2010) implemented a controlled vocabulary on a browser called VectorBase for the mosquito anatomy ontology, and aimed to help with searching data and increase the collection rate via crowd sourcing principles [21].

1.3   The Implementation of Controlled Vocabulary in Design

An initial user survey into bookmarking behavior, conducted prior to this research, showed that just over half the participants used bookmarks to save important websites, and that 75% of these people use up to three sub-levels only [22]. The bookmark facility within all current web browsers was therefore underutilized and the argument that users need and want greater freedom and flexibility to create their own unique file structure was questioned.

The elements in need of improvement are: categorization, optimum levels of sub-folders, ambiguity of the use of vocabulary and contextual user mental models. The authors hypothesize that if the filing system could be less complicated and less strict, it would encourage users to be more willing to organize their information under such a structure. The intention is not to replace the existing file system

structures, but rather to offer a fresh perspective in context, visualization and user interface design.

1.4 Analysis of Typical Daily Life Patterns

The creation of narrative scenarios of user life patterns is a useful iterative process that can shed light on meaning and context. User life patterns were validated with users by a series of usability tests. The authors paraphrased sequences of actions into three different contexts, i.e. regular weekdays, weekends and holidays for most working people in a narrative style.

Table 2 was generated in order to show the connection between the daily activities and the four main CV headings.

Table 2. Connection between daily tasks and CV main headings

Task	Composed of Activity	Related to
Prep for the day	Shower, Shave, Dress, Makeup, Exercise.	Health
Breakfast	Eating, Socializing, Emailing, Reading.	Health, Social
Travel	Driving, Train, Plane, Boat, Bus, Walking.	Travel, Health
Chores/Tasks	Cleaning, Shopping, DIY.	Health, Social
Work	Reading, Writing, Research, Meetings, Physical.	Work, Social
Social/Play/Home	Reading, Watch TV, Play Sport.	Health, Social
Eat/Dinner	Prepare Food, Cooking, Eating, Cleaning.	Health, Social
Sleep	Shower, Reading, Listen to the Radio.	Health

1.5 Social Classification System – Controlled Vocabularies

Based on the analysis of user lifestyle patterns above, the author concluded that a simplified set of four headline categorizations, namely, Work, Social, Travel and Health, could be deployed as a starting point for the user evaluation studies.

Further analysis of mainstream media, such as TV, online social/media websites and daily newspaper publications allowed further sub-headings (categories) to be elucidated as can be seen in Table 3 below.

Based on user feedback and initial results these can be further modified as required. Notice that the last entry is designated as 'Other' to denote the possibility of a catch-all category whereby users could place a website that they were unable to categorize under any other heading.

Table 3. Controlled Vocabulary main headings and their related sub-headings

Work	Social	Travel	Health
Computing and IT	Arts & Entertainment	Car	Alcohol & Drugs
Job Search	Education	Country Information	Arthritis
Leave	Family & Friends	Currency	Baby/Child
Meetings & Presentations	Fashion & Clothes	Flights	Blood Pressure
Office Equipment	Finance & Business	Holidays	Cancer
Pension	Food & Cooking	Insurance	Coughs, Colds and Flu
Promotion	History & Religion	Luggage	Diet/Weight/Fitness
Research	Hobbies & Sport	Maps/Directions	Diabetes
Retirement	Home & Garden	Passport/Visa	Heart Disease
Salary & Allowances	Natural World/Environment	Traffic Reports	Hear/Sight/Smell/Taste/Feel
Tax	News/Politics	Vaccinations	Mental/Depression
Training & Study	Science & Tech.	Weather	Sexual/ Family Planning
Other	*Other*	*Other*	*Other*

2   Formative User Studies

2.1   Methodology

The researchers used a closed card sorting method which included pre- and post-card–sort questionnaires to collect more information from participants before and after the card sort task. Participants therefore had the opportunity to offer good insights that helped to overcome the limitations of online card sorting. This could be especially useful with closed card sorts. The aim of the study was to find patterns of users' mental models and see the degree to which participants agreed on which cards belonged within each category. Closed card sorting enabled not only the understanding of users' interpretation of a Controlled Vocabulary's structure, but also to determine which label should be associated with each stand alone piece of content in the early stages of the design process. The study was launched on September 28, 2012 and closed on October 23, 2012. There were 230 participants, and 152 (66%) of those participants

sorted all 48 cards. Each participant sorted 48 cards into the four predefined main categories: Work, Social, Travel and Health.

2.2   Results and Discussions

The median time taken to complete the survey was 11.9 minutes. For pre-card-sort questionnaires, there were 56% male participants and 44% female participants. In terms of the participants' age, 54% were between 18-25 years old and 18% from 35-45 years old. 57% of the participants were students and 29% were employed. 52% of the participants had a University background and 32% had a Master degree. In terms of computer experience, 19% of the participants had 10-12 years, and 17% of the participants had 8-10 years. Interestingly, 16% of the participants had more than 20 years of using computer experience. 96% of the participants were aware of the Favorites (Bookmarks) Function. The majority of the participants (83%) regularly saved web pages by using the Favorites (Bookmarks) Function. 85% of the participants knew how to organize Favorites (Bookmarks). However, 29% of the participants found it difficult to manage Favorites (Bookmarks). 65% of the participants maintained a maximum of about 50 bookmarks in their Favorites, whilst 86 % of the participants used up to three levels of Favorites (Bookmarks). In terms of revisitation, 65% of the participants experienced problems when they revisited web pages. There were 67% of the participants who used Favorites (Bookmarks) to find previous web pages, whilst 59% of the participants used Google Search.

In terms of the post-card-sort questionnaire, 44% of the participants stated that between 6-17 cards fitted into more than one category. The majority of the participants (73%) found that less than 5 cards didn't fit into any of the four categories. More than half of the participants (57%) considered the four categories appropriate.

The questionnaire enabled participants to suggest alternative group-labeling options via their open-ended comments, i.e. adding more categories like leisure, personal interests, and domestic, some of the items were dependent on what kind of things they were thinking of and therefore it was hard to decide to which category to put them; and some cards could be overlapping to another category.

The popular placements matrix (see Table 4 below) attempts to propose the most popular groups based on each individual card's highest placement score. Each table cell shows the percentage of the participants who sorted that card into the corresponding category.

Table 4. Popular Placements Matrix in Percentages (Closed Card Sort Result)

Category	Health	Social	Travel	Work
Blood Pressure	95	2	1	2
Heart Disease	95	4	0	1
Diet/Weight/Fitness	94	3	1	2
Coughs, Colds and Flu	93	2	3	3
Cancer	92	3	1	3
Diabetes	91	2	3	4

Table 4. (*Continued*)

Category	Health	Social	Travel	Work
Mental/Depression	88	5	3	5
Vaccinations	82	2	10	6
Alcohol & Drugs	82	15	1	3
Hear/Sight/Smell/Taste/Feel	80	11	6	3
Arthritis	78	12	2	9
Food & Cooking	54	31	12	3
Sexual/ Family Planning	48	41	9	3
News/Politics	2	82	2	14
Family & Friends	9	74	15	1
Fashion & Clothes	1	74	17	8
History & Religion	1	64	25	10
Home & Garden	16	62	18	5
Arts & Entertainment	6	59	19	16
Hobbies & Sport	28	52	17	3
Education	3	50	2	45
Baby/Child	43	48	2	7
Maps/Directions	1	4	93	1
Luggage	1	5	91	3
Flights	3	2	90	5
Passport/Visa	3	5	89	3
Holidays	2	12	80	6
Weather	6	12	78	3
Car	2	22	68	7
Country Information	1	33	61	5
Traffic Reports	3	14	58	24
Natural World/Environment	12	30	53	5
Currency	1	25	45	28
Job Search	1	3	1	95
Office Equipment	3	3	3	91
Meetings & Presentations	1	11	2	87
Training & Study	4	8	3	86
Computing and IT	1	12	2	85
Research	1	11	4	84
Salary & Allowances	2	13	2	83
Finance & Business	3	14	3	81
Promotion	1	20	4	75
Retirement	5	15	8	72
Tax	1	34	6	59
Pension	5	29	9	58
Annual Leave	2	9	39	51
Science & Technology	4	44	2	50
Insurance	25	26	20	28

Based on a-priori judgment of the correct associations, there were only five cards that were misplaced into other categories amongst the total of 48 cards, i.e. "Finance & Business", "Science & Technology", "Natural World/Environment", "Food & Cooking", and "Insurance".

81% of the participants chose to group "Finance & Business" under "Work" and only 14% of the participants chose "Social". This could be explained in that the participants regarded this label to be relevant to the pre-defined "Work" group.

50% of the participants chose to group "Science & Technology" under "Work". However, 44% individuals chose "Social", which was close "Work". 53% of the participants categorized "Natural World/Environment" into "Travel", rather than 30% who chose "Social". 54% of the participants placed "Food & Cooking" into "Health", whereas 31% grouped it into "Social".

Probably the most problematic term was "Insurance" with its connotations within the Home, Work, car (Travel) and the person (Health). It was found that 28% of the participants classified "Insurance" under "Work", with 26% of the participant's choosing to categories under "Social", 25% under "Health" and 20% under "Travel".

3 Conclusions

By analyzing a typical users lifestyle, it has been possible to adopt four major controlled vocabulary headings, namely, Work, Social, Travel and Health. Under these major headings there exist twelve sub-headings. With this simplified (two-tier) structure, a closed card sorting user survey was conducted online with 152 participants. The research concluded that the use of these primary and secondary facets can help users reduce confusion and simplify the procedure when they organize their personal information.

Further work will involve developing an intuitive user interface front end and then porting a freely available fully functional browser add-in to the internet to allow a much wider cross section of the user community to test and report back on the usability and functionality of this interface over time.

References

1. Malone, T.: How Do People Organize Their Desks? Implications for the Design of Office Information Systems. ACM Transactions on Information Systems (TOIS) (1983)
2. Lansdale, M.: The Psychology of Personal Information Management. Applied Ergonomics 19(1), 55–66 (1988)
3. Kidd, A.: The Marks are on the Knowledge Worker. In: Proceedings of CHI on Human Factors in Computing Systems. ACM, Boston (1994)
4. Whittaker, S., Sidner, C.: Email Overload: Exploring Personal Information Management of Email. In: Proceedings of CHI 1996. ACM, Vancouver (1996)
5. Eysenck, M.W., Keane, M.T.: Cognitive Psychology: A Student's Handbook, 5th edn. Psychology, Hove (2005)
6. Oren, E.: An Overview of Information Management and Knowledge Work Studies. In: Proceedings of the ISWC Workshop on the Semantic Desktop (2006)
7. Randall, D.W., Bowker, G.C., Star, S.L.: Sorting Things Out: Classification and Its Consequences. Computer Supported Cooperative Work 2001 10(1), 147–153 (2001)

8. Bowker, G.C., Star, S.L.: Knowledge and Infrastructure in International Information Management: Problems of Classification and Coding. In: Bud-Frierman, L. (ed.) Information Acumen: The Understanding and Use of Knowledge in Modern Business, Routledge, London (1994)

9. Trigg, R.H., Blomberg, J., Suchman, L.: Moving Document Collections Online: The Evolution of a Shared Repository. In: Proceedings of the 6th European Conference on Computer-Supported Cooperative Work, Kluwer Academic Publishers, Copenhagen (1999)

10. Bowker, G.C., Star, S.L.: Sorting Things Out: Classification and Its Consequences. Inside Technology. MIT Press, Cambridge (2000)

11. Bowker, G.C., Star, S.L.: Building Information Infrastructures for Social Worlds - The Role of Classifications and Standards. Community Computing and Support Systems, 31–248 (1998)

12. Dourish, P., Edward, W., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., Terry, D., Thornton, J.: Extending Document Management Systems with User-Specific Active Properties. ACM Transactions on Information Systems 18(2), 140–170 (2000)

13. Hearst, M.A.: Clustering versus Faceted Categories for Information Exploration. Communications of the ACM 49(4), 59–61 (2006)

14. Bondarenko, O., Janssen, R.: Documents at Hand: Learning from Paper to Improve Digital Technologies. In: Proceedings of CHI 2005, ACM, Portland (2005)

15. Marsden, G., Cairns, D.E.: Improving the Usability of the Hierarchical File System. In: Proceedings of the 2003 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on Enablement through Technology. South African Institute for Computer Scientists and Information Technologists, Republic of South Africa (2003)

16. Czerwinski, M., Horvitz, E.: An Investigation of Memory for Daily Computing Events. In: Proceedings of HCI 2002, ACM Press (2002)

17. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I.: Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. In: Proceedings of SIGCHI 2003, ACM Press (2003)

18. Freeman, E.T.: The Lifestreams Software Architecture. Faculty of the Graduate School, Yale University (1997)

19. World Intellectual Property Oranization: International Patent Classification (IPC). World Intellectual Property Organization (eds.) World Intellectual Property Organization (2011)

20. Svenonius, E.: Unanswered Questions in the Design of Controlled Vocabularies. Journal of the American Society for Information Science and Technology 37(5), 331–340 (1986)

21. Butler, R.E.: The Design and Development of Vectorbase: A Bioinformatic Resource Center for Invertebrate Vectors of Human Pathogens. In: Biological Sciences, p. 120. University of Notre Dame, Indiana (2010)

22. Shen, S.-T., Prior, S.D.: The Utility of Controlled Vocabularies within Bookmark Management Tasks. In: The 5th International Conference on Advances in Computer-Human Interactions. Xpert Publishing Services, Valencia (2012)