

Understanding What Personal Information Items Make Categorization Difficult

Kyong Eun Oh
Simmons College
300 The Fenway
Boston, MA 02115
Kyongeun.oh@simmons.edu

Nicholas J. Belkin
Rutgers University
4 Huntington Street
New Brunswick, NJ 08901
belkin@rutgers.edu

ABSTRACT

This study examined personal information items that are difficult to categorize, and how people deal with these information items. For this research study, 18 participants were asked to keep a diary over a week and record diary entries whenever they decided to save or organize any electronic personal information items. Then, two post-diary semi-structured interviews were conducted to ask questions about how they organized their information items and why they made such decisions based on the diary they kept. Interviews were analyzed by using a grounded theory approach. The results show that the information items that are ambiguous or anomalous are difficult to categorize. Participants dealt with ambiguous or anomalous information items by categorizing them into one of their existing categories, placing them into a miscellaneous category or revising the existing organizational structure. Since personal information items that are difficult to categorize directly influence the time and effort needed to organize personal information, it is important to understand what kind of personal information items makes personal information organization difficult. In particular, the results of this research study have practical implications in developing interfaces and applications that help organizing personal information items effectively. This study also helps us further understand categorization, information organizing behavior, and personal information management.

Keywords

Information organization, categorization, personal information management, information organizing behavior

77th ASIS&T Annual Meeting, October 31- November 4, 2014, Seattle, WA, USA.

Copyright is retained by the author(s).

INTRODUCTION

In our daily lives, we often categorize objects into categories. Personal information items, which individuals keep for personal use either directly or indirectly, are one of the things we categorize in our everyday life (Jones, 2007; Whittaker, 2011). However, when we categorize personal information items, often the time and the effort it takes to organize them are not really the same for different items. Some information items do not require much time and effort to make organizational decisions, while other information items take longer time and more effort to make decisions about where to categorize them. This study explored personal information items that are difficult to categorize, and how individuals handle these information items to categorize them.

CHARACTERISTICS OF A CATEGORY

Categorization has been studied in many different fields including philosophy, psychology, and linguistics (Langridge, 1992). Here, we highlight a few core theories on characteristics of categories. First of all, Aristotle (2007[B.C.350]) stated that a category has a clear boundary and each member in a category has a concrete defining property that determines its membership. However, this view was challenged, especially by Wittgenstein (1953) who asserted that the members in a category have *family resemblances*, which are some similarities shared by the members instead of definite defining properties (p. 32). According to Wittgenstein's perspective, an object can be organized into multiple categories. In a similar vein, Langridge (1992) stated that even the simplest objects have many different aspects so that they can be categorized in a variety of ways. By conducting a number of experiments, Rosch (1978) found that there is a *prototype* in a category, which is the clearest member in a category based on people's judgments of goodness of membership in a category (p. 36). In fact, she and her colleagues found that people's judgments about the membership of an object in a category is not exactly the same, and people make this judgment much faster when an object is more prototypical (Rosch, Simpson, & Miller, 1976). These studies provide good background knowledge in comprehending the

characteristics of categories, and support the fact that certain information items are easier or more difficult to categorize than other information items. Understanding what personal information items are difficult to categorize and how people handle them are important because it is directly related to our daily lives, and can support devising strategies or tools that can help solving the problem. However, there appear to be no empirical studies that have focused explicitly on this problem.

METHODS

Participants

Eighteen participants in this study were recruited from academics including professors, graduate students and undergraduate students in a social science field in an institution of higher education in the United States.

Procedures

To collect data, participants were asked to keep a diary of whenever they decided to save or organize a personal information item they received in electronic form in a given template over a week. Then, two post-diary semi-structured interviews were conducted, one at the end of the week, and the second three weeks later. The first asked how they categorized each information item in the diary, and why they made those decisions; the second asked if they had changed any of the decisions. In particular, whether it was easy or difficult to make each decision and the reasons for easiness or difficulty were asked. The interviews were recorded, transcribed, and analyzed based on the grounded theory approach (Strauss & Corbin, 1990).

RESULTS AND DISCUSSION

Personal information items that were difficult to categorize could be grouped into two different types: (1) Ambiguous information, which can be placed into multiple categories; and (2) anomalous information, which does not fit into the existing organizational structure. Explanations about each type of information and how participants dealt with them are described as follows.

Ambiguous Information

The results showed that participants found it difficult to categorize when an information item can be categorized into multiple categories. While explaining about why certain information item was hard to categorize, participants often mentioned facing difficulties in deciding into which category they should place an information item when it could be placed into multiple categories. For instance, participant 13 (P13) said, "I had to think about where to put that. Because I do have another thing [folder] called 'Articles'. And it must be under 'Dissertation'." P20 also reported a similar difficulty by saying that "It is summer work. It's not technically class work, but it is part of the summer work that I am doing for school this semester", showing challenges when information can belong to more than one category. This problem derives from the fact that

an information item has various aspects (Langridge, 1992; Lansdale, 1988), which allows an information item to be categorized in many different ways. In addition, often categories are not mutually exclusive (Langridge, 1992; Lansdale, 1988) which makes it possible for an information item to be placed into several categories. Thus, ambiguous information items that can be categorized into multiple places were difficult to categorize.

Anomalous Information

The results of the analysis also showed that when participants encountered information items that could not be categorized into one of the existing categories, they found it challenging to categorize them. While explaining about an information item that was harder to categorize than other items, P9 said, "They don't fit into a folder I already have". Similarly, participants said, "It's just not associated with any class or research project that I currently have" (P22), and "They don't have a natural place in any of my existing files" (P25), showing difficulties in categorizing information items that do not fit into the existing organizational structure. In particular, this happened when participant had a new task or new type of information items. For instance, P21 said, "I didn't have anything – that was the first thing I had for it", while explaining the reason why it was difficult to make an organizational decision. This also happened when the possible future use of an information item was unknown. For example, while describing about challenging situations in categorizing personal information, P22 responded, "This happens often when somebody sends you something interesting, and you know that it is interesting, but you can't decide if you'll ever use that or not. And if you will, then where and how?" Thus, anomalous information items that cannot be placed into one of the existing categories were found to be hard to categorize.

Dealing with Ambiguous or Anomalous Information

There were three different ways of dealing with ambiguous or anomalous information items when categorizing them. First, participants handled these information items by categorizing them into an existing category by regarding them as members of one of the categories. For example, P20 said, "At first I wasn't sure where is best to put it, and I ended up putting it here", indicating that the participant treated it as one of the members of a category although the membership was not as clear as other information items. Second, participants decided not to categorize them into the existing categorical structure but instead, placed them into a miscellaneous category. For instance, P9 said, "They don't fit into a folder I already have, and I just like save them into 'Documents' folder". Similarly, participants said, "I have like a Miscellaneous folder that anything that I don't have a folder for, I kind of throw in there" (P21), and "Things that I find at some point interesting or useful but can't decide whether to go under teaching some subject or research some topic ends up there" (P22). Third, participants dealt with these information items by revising the existing

organizational structure. For example, participants said, “So it was work-related and I saved it in a folder that I already had, but you can see I changed the name of the folder” (P21). Similarly, P14 responded that “You have a new conference, you create a new folder” (P14), indicating changes made to the existing folder structure to categorize anomalous information item.

CONCLUSION

This study examined personal information items that are more difficult to categorize than other information items. In addition, how participants handled these information items was investigated. The results show that ambiguous information items that can be placed into several categories or anomalous information items that do not fit into the existing organizational structure are harder to categorize than other information items. In addition, participants dealt with these information items by categorizing them into one of their existing folders, placing them into a miscellaneous category, or changing the organizational structure. Understanding what personal information items require more time and effort to organize them as well as the way people handle these information items is directly related to developing interfaces and applications that support individuals’ personal information organization. In addition, the results of this study deepen our knowledge of categorization, information organizing behavior, and personal information management.

REFERENCES

- Aristotle. (2007). *Categories* (D.M. Edghill, Trans.). Retrieved August 25, 2011 from <http://etext.library.adelaide.edu.au/a/aristotle/categories/> (Original work written B.C. 350).
- Jones, W. (2007). Personal Information Management. *Annual Review of Information Science and Technology*, 41, 453-504.
- Langridge, D. (1992). *Classification: Its Kinds, Elements, Systems and Applications*. London: Bowker-Saur.
- Lansdale, M. (1988). The psychology of personal information management. *Applied Ergonomics*, 19(1), 55-66.
- Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. Lloyd (Ed.), *Cognition and Categorization* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rosch, E., Simpson, C., & Miller, R. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 491-502.
- Strauss, A., & Corbin, J. M. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage Publications, Inc.: Thousand Oaks, CA.
- Whittaker, S. (2011). Personal information management: From information consumption to curation. *Annual Review of Information Science and Technology*, 45, 3-62.
- Wittgenstein, L., (1953). *Philosophical Investigations* (G.E.M. Anscombe, Trans.). New York, NY: Macmillan (Original work published 1945).