

Structural Analysis of Computer Vocabulary: How do Users Conceptualize the Computer?

Adity Mutsuddi

Introduction

Computers in all forms and sizes are used by people with different levels of computer expertise. Among these users, there are many who do not have any technical knowledge about the computer– they use it to “get something done”. What do these users think about the computer? This research attempts to discover the concepts users have about the computer from statistical analyses of the words they had used when asked to complete some computer tasks and to describe specific technical terms they might have encountered when using computers in their daily life. An understanding of user perceptions of the computer may suggest the level at which users should be approached when assisting them to solve computer-related problems.

Related Work

Words used to tag online resources such as videos, papers and photos [1, 4] have been investigated to understand the structure of social tagging systems on the Internet. Words used to describe interests in user profiles of online journals have been analyzed to understand the structure of user interests [5] and the changes in those interests [3]. In this research, words used by participants during the computer tasks and the exit interview in which they described computer-related terms are analyzed to understand user concepts of the computer.

Data collection

This research is part of a larger project [2] which used design principles to design secure and usable interfaces for users to administer their computers. User studies were performed to examine both the current Windows XP and the newly designed interfaces. The participants were asked to complete four tasks based on a scenario of four family members. The participants had to 1) create accounts for the family members 2) login to visit a website that would show a certificate 3) login to install software 4) configure the system. Joint-exploration¹ [6] was used to elicit the thought process of the participants during the tasks. Upon completion of the tasks, the participants had an exit interview in which they were asked to describe technical terms (e.g. operating system, firewall, network, administrator) in addition to questions related to their tasks.

The vocabularies for this research have been compiled from videotapes of the discussions participants had with the joint-explorer and the exit interview. There were twenty-two (n=22) participants of ages between 25 and 65 years old who used computers regularly and did not hold a degree in a computing related field. Any technical word used by the participants during the tasks and the interview was included in the data. Words like

¹ Joint-exploration utilizes deception to have the participant complete tasks with another “participant”, who is actually a researcher. [2]

“login”, “install”, “user”, “administrator”, and “computer” that were used in the text to describe the tasks were excluded. Furthermore, words like “things” and “stuff” were not included.

Data Description

The data used for statistical analysis consists of 83 unique words, the total number of participants who used a particular word (num_ppl) and the participants who used that word (S2, S3 and so on). The table below shows the structure of the data:

Words	num_ppl	S2	S4	S5	S6	S7
access	17	0	0	1	1	1

There is also information on the contexts under which a word was used.

Data Analysis

The data was analyzed using Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) as implemented in the statistical programming language and environment “R” [7]. The PCA allows the characterization of the major dimensions of variation in the data. The results from the PCA were then used in HCA to identify the groups of users sharing the usage of similar words. The clusters of words that identify groups of users indicate the concepts users have about the computer. This method has been successfully used in [3, 4, 5].

Results

The statistical analysis together with the analysis of the contexts under which the words were used shows the following:

- 1) The five most frequently used words are- software, program, access, virus and security
- 2) For most users, the computer is all about software, program and access to different things (e.g. computer, network, website, software)
- 3) Anything threatening to the computer is a virus
- 4) Some people think of the computer in terms of data- files, documents, internet, information, upload and download

Conclusion

Most users conceptualize the computer in terms of software and access to different things on the computer. Hence, when we are trying to help general users with computer-related problems we should not use any technical terms without explaining their meaning in words that they would understand. When trying to help users to protect their computers from worms or malware we should make sure to differentiate them from a virus.

References

- [1] Marlow, C., Naaman, M., Boyd, D., and Davis, M. 2006. *HT06, tagging paper, taxonomy, Flickr, academic article, to read*. Proceedings of 17th Conference on Hypertext and Hypermedia, 31-40.
- [2] Mutsuddi, A., Jacob, B., Sun, Y., Connelly, K., and Gupta, M. 2007. *A Case Study in using Design Principles for Secure Operating Systems Interfaces*, Under Submission.
- [3] Paolillo, J.; Mercure, S.; and Wright, E. 2005. *The Social Semantics of LiveJournal FOAF: Structure and Change from 2004 to 2005*. Proceedings of the International Semantic Web Conference, Workshop on Semantic Network Analysis; G. Stumme, B. Hoser, C. Schmitz, and H. Alani, eds.
- [4] Paolillo, J. and Penumarthy, S. 2007. *The Social Structure of Tagging Internet Video on del.icio.us*. 40th Annual Hawaii International Conference on System Sciences. Los Alamitos, CA.
- [5] Paolillo, J and Wright, E. 2007. *Social Network Analysis on the Semantic Web: Techniques and Challenges for Visualizing FOAF*. In V. Geroimenko and S. Chen, eds., *Visualizing the Semantic Web*. Berlin: Springer Verlag.
- [6] Sasse, M. A. 1997. *Eliciting and describing user's models of computer systems*. PhD thesis. School of Computer Science. University of Birmingham.
- [7] "R" Statistical Programming Language. <http://www.r-project.org>.