Application of theories of indexical cognition to a Web-based workspace

Brian Fisher
Media and Graphics
Interdisciplinary Centre
University of British Columbia
Vancouver, BC, V6T 1Z4
fisher@cs.ubc.ca

John Dill
School of Engineering Science
Simon Fraser University
Burnaby, BC, V5A 1S6
dill@cs.sfu.ca

Abstract

The World Wide Web and its browsers have created a large and changing information space that is intended to be accessible to users with a minimal level of technical expertise. Web information is said to exist in a "Web-space" that presumably carries with it some (but perhaps not all) of the characteristics of a physical space. This metaphor of moving through a space contrasts with the prevalent "desktop metaphor" of tools specialized for specific tasks. The less constrained tasks and changing information on the Web calls for a correspondingly flexible information visualization approach.The fluidity of our proposed interface requires us to rely less on visual semantics (e.g. icons) in our design, and more on basic aspects of human cognitive architecture: spatial cognition, visual attention and indexical cognition (FINST) models.

Introduction

Traditionally, interface designers have had as their goal the creation of software "tools' that would allow users to perform particular tasks easily and effectively. Methods and theories from Cognitive Science and Psychology that deal with the conscious performance of mental tasks (e.g. GOMS and keystroke models, task analysis, protocol analysis) and visuomotor tasks (e.g. Fitts Law, SR Compatibility) were adapted to create core methodologies in the study of human-computer interaction [1].

Adaptation of theories and methods from Cognitive Science in interface design has been highly selective. The emphasis on semantic similarities (e.g. of a word processor with a typewriter, or a visual calculator with a paper spreadsheet) has reduced the need for design methods that rely upon basic characteristics of human attention and perception.

Thus, findings in perceptual cognition have played a largely supportive role. Mapping the functions of the software onto real-world tasks is made easier when the stylized icons are perceptually similar to real-world objects and dissimilar to each other. Findings from perceptual research have been employed to help designers to accomplish this. However the way in which these icons are interpreted by the perceptual system is not central to the process of choosing and implementing the visual metaphor. Instead, they typically take the form of general guidelines [1], such as limiting short-term memory load to less than seven items, maintaining SR compatibility, and limiting the use of attention-demanding stimuli such as flashing icons. Indeed, some respected HCI researchers have even suggested that deeper theories from Cognitive Psychology are of little use in software design [2].

Challenges of the Web

Designers of an interface that complements the fluid and ambiguous nature of Web-based materials would find it difficult to create the sorts of semantic analogues that characterize the desktop metaphor. Unlike data and applications on a personal computer, Web information:

- May change in an unpredictable way, and at any time
- May move to a new location on the Web
- May be distributed in a variety of formats (e.g. JPEG, GIF, TIFF) that are logically equivalent to the user
- May be actively generated by an interactive application or consist of static html
- May be generated locally (e.g. Java) or server-side (e.g. CGI). This may or may not be important to the user.

Mapping these characteristics onto a static interface such as a desktop would create expectations of constancy that would not be kept.

Accordingly, we set out to create a fluid interface environment, an “anti-desktop” if you will, that was nevertheless understandable and navigable. Despite changes in the size, positions, and characteristics of objects in the interface, users must be provided the perceptual support needed to function within the environment. This requires a new set of design parameters based in the cognitive architecture of human performance in complex environments.

Spatial cognition in fluid environments

In a complex environment, multiple objects and events must be individuated prior to perceptual and cognitive processing. Each object will have a position in space and may have a variety of visual, auditory, and tactile characteristics associated with it. These characteristics are parsed from the environment as a “feature cluster” and processed to give rise to a mental representation (called an “object file” [3]) for each important object in view. Object files contain not only the current characteristics of the object, but also much of its history.

The process of forming this mental representation requires a certain amount of time, probably in the range of 100s of milliseconds. In a dynamic environment, changes in the location or other characteristics[4,5] of an object during the time course of processing would make the process of forming a correct mental representation more difficult. In order to “keep track” of a given object, some low-level mechanism must maintain the continuity of processing of a changing object.

It seems that there is a significant amount of interaction between cognitive and perceptual processes that is required to function in a rich perceptual environment. Yet, the nature of the divisions in the flow of information and control that make up the human cognitive architecture requires that cognition have limited access to the sensory processes that create these representations[6,7]. In order to understand operator performance in rich sensory environments, analysis of users' conscious performance of mental operations associated with their tasks must somehow be integrated with an understanding of their perceptual and motor processes. This requires a fast, low-level mechanism for linking thought processes with specific events in the perceptual world.

The minimal mechanism for this is a set of spatial indexes, pointers, or attentional tokens that serves to link mental operations (ranging from simple visual routines such as item collinearity to complex conceptual structures) and specific perceptual events. These pointers were first used in computer vision applications by Pylyshyn and colleagues [8, 9,10]. These attentional tokens were given the name FINST, for FINger of INSTantiation. They have subsequently been used to explain human perceptual and cognitive processing in a wide variety of domains [11].

According to the FINST hypothesis, there are a limited number of attentional tokens (FINSTs) that constitute a fundamental bottleneck in human processing. Just as short-term memory limitations (i.e. 7 +- 2 items [12]) and focal attention (i.e. a “spotlight” of attention) limit our ability to perform certain tasks, the number of spatial indices can be a determining factor in our ability to parse complex displays. Current thinking suggests that at most six items in a display can be simultaneous individuated by FINSTs for cognitive processing. These items receive preferential processing in a number of ways that will be described below.

According to this hypothesis, items in a complex display will fall into three separate categories:

- Attended items-- Items that occur in a spatially contiguous region that the operator is currently attending to. Processing in this region follows the traditional “spotlight” metaphor of endogenous attention, where the level of processing (grossly defined) varies inversely with the size of the area attended.

- Indexed items-- Up to six display items can be individuated and indexed by having a FINST assigned to them. FINSTs are usually assigned in a bottom-up manner based upon salient display events such as the onset of a new item. FINSTed items gain a number of specific processing advantages: They are available (potentially in parallel) as arguments for simple perceptual routines such as collinearity, conjunctive search, subitizing, and for rapid selection for focal attentive processing. These items generate unique mental representations called object files [3] that allow multiple characteristics (i.e. perceptual features) to be monitored in parallel and maintained through time (i.e. as an object-bound history in short-term memory). Unlike attended items, FINSTed items can be distributed across the screen in any configuration (i.e. they do not need to fall within a contiguous region to receive enhanced processing).

- Background items-- The remainder of the display receives very limited processing. New items that appear at unindexed locations may draw an index if their onset is salient; however if onset occurs during a saccade or a screen blink the item itself is unlikely to be noticed [13]. Changes in existing objects are also likely to go unnoticed, and will in all cases be responded to more slowly than similar changes in FINSTed items. Finally shifting attention to these items is slower than attending FINSTed items, and the time to required to attend to them will be roughly proportional to their distance from the current focus of attention [14].

In tests with human subjects, display items that are FINSTed have the potential to be accessed and acted upon by cognitive processes in parallel, with higher priority than unFINSTed items. We can use this model to predict the interaction of dynamic display events and cognitive processes, and to design displays that are optimized for the particular mental processes.

A concentration on the nature of the linkages between perceptual events and cognitive structures substantially alters the way we think about mental representations. Given this mechanism, we can derive models of information processing that rely upon the perceptual world to provide much of its own representation [15,16]. The theoretical issues involved are beyond the scope of this paper, but a forthcoming book by Pylyshyn (to be published by Ablex) should provide sufficient background for interested readers.

An interesting repercussion of the FINST hypothesis deals with the way in which we think about and remember physical spaces. Our introspections tell us that the way in which we think about and remember objects in space (e.g. the arrangement of furniture in our living room, or the path we take to work each day) is qualitatively different from the way we think about and remember facts or experiences we have had. This difference is born out by a great deal of research in spatial cognition [17].

The nature of that difference, however, may not be intuitively obvious. While we typically feel that our memory of events in space is itself spatial (e.g. we "visualize" our living room in a way that is very similar to actually seeing it), this intuition is inconsistent with the modularity hypothesis and the evidence that supports it. Instead, it is suggested that cognition is sensitive to qualitative spatial relationships between objects (e.g. "inside of", "above" beside") that are the products of simple "visual routines" [18].

Mental models of Web space

We set out to devise a graphical representation of a fluid and complex data space (in this case the World Wide Web) that takes into account what the FINST hypothesis says about how we understand complex displays. The first question that we set out to answer was: How should this map differ from a static (paper) map or a physical space? Physical space has metric characteristics (i.e. the distances between items can be measured to any arbitrary level of accuracy).There is no obvious analog to metric distance in Web space, yet we do think of the Web a space of sorts.

According to the FINST hypothesis, the metric characteristics of events in physical space are used to parse visual scenes, but receive limited processing at the cognitive level. Processing spatial relationships takes place at the perceptual stage, and FINSTs serve to individuate a small number of salient events for cognitive processing. The alternative view would render the FINST mechanism unnecessary, as events could be individuated at the cognitive level by simply remembering their spatial coordinates.

At the cognitive level, we hypothesize that spatial relationships are preserved in a qualitative sense, but metric positions of events are not readily available. Individuation of objects and events may be spatially derived at the perceptual level, but are communicated to cognitive processes by way of pointers (FINSTs) that do not specifically pass on metric information.

Web users may indeed think of the Web as a spatial entity in the sense that individual Web pages can be individuated (i.e. they can be thought to be in different places), and exist in some relationship to other pages (e.g. they are on the same server), but without the metric characteristics of real space.

If the FINST hypothesis is correct, it should help us to design representations of data spaces that are tailored to the way in which we process, think about, and remember events in complex scenes. We can extrapolate from these theories to generate some design recommendations:

First, the number of important display items would be kept small, to avoid exceeding the number of FINSTs available (approximately 6). this would insure that each display item could be indexed. This in turn would make it likely that each item could be tracked across eye movements or display changes, that it would maintain an individual object-history representation, and that changes in its display characteristics would be noticed quickly.

Second, since metric distances are relatively unimportant, changes in the display configuration could be allowed insofar as they do not interfere with the more qualitative spatial relationships (e.g "inside-of", "on-top-of") that are hypothesized to be preserved in memory. This allows for limited alterations in the size and layout of display items when circumstances warrant (for example, when more screen real-estate is needed).

Third, if continuity of processing is important, we must avoid display transformations that disrupt the ability of FINSTs to "stick" to their object: rapid onsets of irrelevant items should be avoided if possible, and indexed items should move smoothly to new positions rather than abruptly shifting location.

Patterns of use of the Web

In order to apply our theoretical perspective to the design of an application, it must be focused by observational and experimental studies of Web use. Studies of Web use report patterns of behaviour that support the idea that users are often interested in assembling and revisiting a limited number of pages--what Saul Greenberg calls a "recurrent system" [19]. According to research in U.Calgary's Grouplab [20]:

- About 30% of all logged navigation activity involve use of the Back button
- About 60% (s.d. = 9%) of all page visits are to pages a person has been to before
- A list of 6 or so URLs just visited contain the majority of pages a person will visit next
- Frequency of visits is a poor indicator: most people only access very few pages frequently and regularly.
- Users do not exhibit strongly repeatable linear pattern of visits; while they tend to revisit recent pages, they do so in a different sequence

The recurrent system model of Web use postulates that users are primarily interested in creating sets of small numbers of familiar sites, bringing in new information from other sites as it is needed. If we consider the finding in #3 above as a rough estimate, we find that most of their referents could in principle fall within the number of FINSTed display items.

CZWeb: a flexible map of Web space

CZWeb (figure 2) is a tool for capturing and representing activities in the Web. CZWeb runs alongside the browser, generating a representation of visited sites.

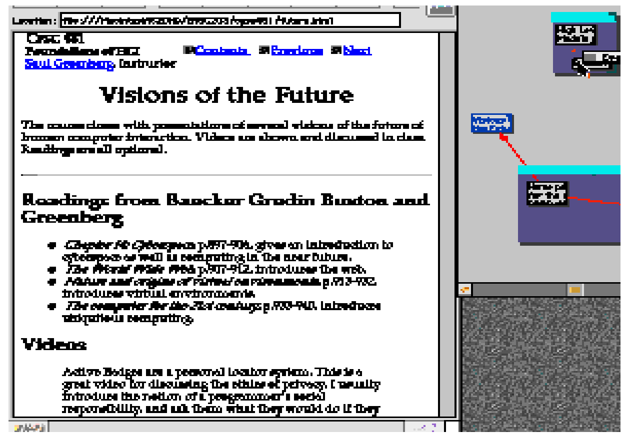

Figure 1: CZWeb and Netscape

In order to minimize the number of objects in the screen, CZweb uses a heirachial graph representation. Individual pages loaded into Netscape are represented as page nodes on the CZWeb display. One or more related page nodes can be grouped into cluster nodes (Figure 2).

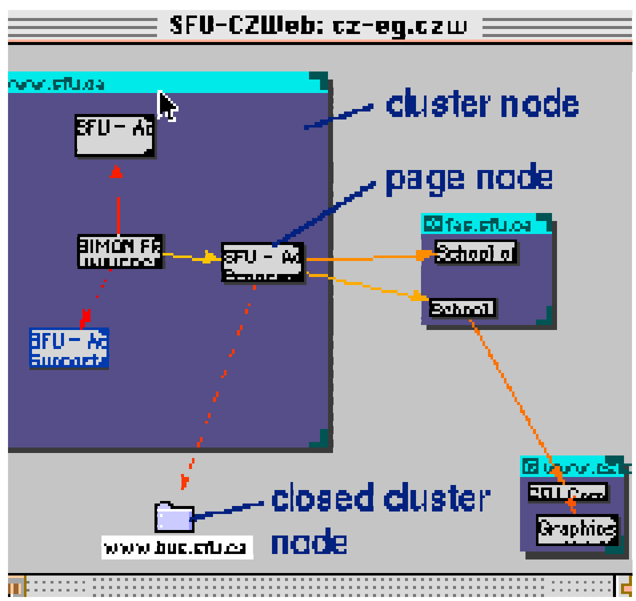

Figure 2: CZWeb display items

The default is to generate a cluster node for each Web server visited, and to display the pages served by that server within its cluster node.

As new pages are added, CZWeb's Continuous Zoom display algorithm [21] sequentially changes the position of existing pages, then reduces their size, and finally collapses smaller clusters into icons to free up real estate.

Since metric distances and item scale are hypothesized to be of less importance for memory than display (and hence FINST) continuity, these changes are predicted to be well tolerated by users.

CZWeb does maintain qualitative spatial relationships (e.g "inside-of", "on-top-of") that are critical for spatial cognition and which often carry relationship information needed for situation understanding and decision-making. Preserving the relative configuration of elements may also help the user to cognitively "chunk" individual elements to further aid the decision-making process [22].

CZweb page nodes are active, and the user can reload a page in Netscape by clicking on its icon in CZWeb. The display is also configurable by the user-- they can move items, increase or reduce their size, close cluster nodes, create new clusters and populate them with pages dragged from other clusters. This enables them to create customized representations of important sites in their preferred organization and spatial layout. These can be saved to disk and mailed to other users.

User tests

User tests of CZWeb cannot in themselves confirm the effectiveness of the manipulations without appropriate control applications. However high user satisfaction ratings and observations of usage in laboratory contexts support our usability claims in a general sense. These findings were reported in more detail in a previous publication[23]. In summary:

- 12 university students acted as subjects
- Their task was to answer questions about the 96 Olympic sailing races
- Information came from a local Web site on the races
- Subjects were asked to think aloud and were videotaped as they solved the problem
- Subjects answered a pre-test and post-test questionnaire

The results of the questionnaire showed that users found CZWeb generally easy to use ($p < .05$), and useful for moving from site to site ($p < .00101$). Compared to Netscape navigation features (back/forward, history) Subjects rated CZWeb as more useful for helping to understand Web organization ($p < .0001$), moving to less recently visited sites($p < .0001$), and knowing where they are in the Web and avoiding going to the wrong site ($p < .0001$).

CZWeb as a environment for knowledge manipulation

CZWeb was originally envisioned as a Web mapping and bookmarking tool that used spatial transformations to more effectively use limited screen real estate. More recently, the concept was extended to allow users to annotate page and cluster nodes, an idea suggested by Steven Forth of DNA Multimedia.

The effect of this change in the interface is to allow users to iteratively alter the CZWeb representation to reflect their changing understanding of the information. By re-annotating page nodes, moving page nodes to clusters, annotating clusters, creating new clusters and clusters of clusters, and moving them to new locations, users create visual analogues of their changing understanding/ interpretation of the material.

The effect of this is to create a "visual thinking space' where changes in the relative location and clustering of items carries qualitative information, while more precise information can be carried in the annotations. User interaction with CZWeb is similar to that of concept mapping tools, with the exception that the objects they manipulate combine user knowledge with chunks and clusters of media found on the Web.

Decision evolution using CZWeb

Klein and colleagues [24] have argued that human cognitive abilities limit mental models to a maximum of three active agents (e.g. moving parts) and six discrete steps. This limitation constrains the effectiveness of formal inferential methods that are observed in laboratory tasks and taught in formal training programs. Klein's research shows that skilled decision-makers rely instead on their trained perceptual abilities to support an understanding of the situation.

Traditional decision evolution environments are typically large (e.g. War Room) spaces where models of a situation are manipulated in a series of "what-if" analyses, downloading cognitive information to a physical representations [25]. Creating a similar "thinking space" on a small computer screen requires detail-in-context spatial transformations that maximize the effectiveness of limited screen "real estate". Pylyshyn's FINST theory of indexical perception provides us with the basis for transforming the display in such a way as to minimize the impact on the perceptual/ cognitive linkages (i.e. FINSTs and object files) that support the decision-making process.

The application of the FINST visual cognition theory to display design is new, and may lead to other useful applications. We have begun work on several of these in collaboration with industrial partners ThoughtShare Inc. and Virtual Learning Environments Inc. We are also exploring the impact on the design of air traffic control systems with colleagues from HRL Labs.

As a collaborative environment, CZWeb allow users to reference specific materials on the Web directly, and their incorporation in online conference can support multimedia-based collaboration and training. Because discussions takes place "over top of the web sites", comments can be changed without consequence to the web site, and the contents of the site can be altered without damaging its association with the user's annotations.

Other uses for the CZWeb format might allow them to be placed on the server to be downloaded and modified by users. These may contain fixed icons such as landmarks, trademarks, etc. Search Engines can also serve results in the form of CZWebs, for reduced server load and increased usability.

Technical:

CZWeb currently requires a PowerMac computer with Netscape Navigator 3 or better. A PC-compatible commercial application based upon CZWeb is under development by ThoughtShare Communications Inc.

Acknowledgments

This research was supported by Hughes Raytheon Labs, the Institute for Robotics and Intelligent Systems (Human Machine Interface-1 project) Network of Centres of Excellence, and the Telelearning Network of Centres of Excellence. The authors would particularly like to acknowledge the input and support of Zenon Pylyshyn and colleagues at the University of Western Ontario and Rutgers University.

References

1. Shneiderman, B. *Designing the User Interface* Addison Wesley, Reading, MA, 1998.

2. Landauer, T. K. Let's get real: A position paper on the role of Cognitive Psychology in the design of humanly useful and usable systems. in J.M. Carroll (Ed.) *Designing Interaction*. Cambridge, Cambridge MA 1991.

3. Treisman, A. and DeShepper, B. Object tokens, attention, and visual memory. In T. Inui and J. McClelland (Eds.) Attention and Performance XVI: Information Integration in Perception and Communication, Cambridge, MA: MIT Press, 199. pp.15-46.

4. Fisher, B.D. *Integration of visual and auditory information in the perception of speech events.* Doctoral Dissertation, University of California at Santa Cruz. 1991

5. Fisher, B.D. and Pylyshyn, Z.W. The cognitive architecture of bimodal event perception: A commentary and addendum to Radeau. *Current Psychology of Cognition* 13:1. pp. 92-96.

6. Pylyshyn, Z. W. *Computation and Cognition*. MIT Press, Cambridge MA 1984.

7. Pylyshyn, Z.W. The role of cognitive architecture in theories of cognition. In K. VanLehn (Ed.), *Architectures for Intelligence*. Hillsdale: Lawrence Erlbaum Associates Inc. 1991.

8. Pylyshyn, Z.W., Elcock, E.W., Marmor, M. and Sander, P. A system for perceptual-motor based reasoning. (Tech. Rep. no. 42) London Ontario: Dept. of Computer Science, University of Western Ontario.1978.

9. Pylyshyn, Z.W. The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. Cognition, 32, 1989 pp 65-97.

10. Pylyshyn, Z.W. "Here" and "There" in the visual field. In Z.W. Pylyshyn (Ed.), Computational Processes in Human Vision: An Interdisciplinary Perspective. Norwood, N.J.: Ablex Publishing. 1988.

11. Pylyshyn, Z., Burkell, J. Fisher, B. Sears, C. Schmidt, W. Trick, L.Multiple parallel access in visual attention. Canadian Journal of Experimental Psychology 48:2, 1993, pp.260-283

12. Miller, G. The magical number seven plus or minus two: Some limitations on our ability to process information *Psychological Science* 63, 1953, pp. 81-97.

13. Rensink, R.A., O'Regan, J.K., & Clark, J.J. Failure to Notice Scene Changes made during Image Flicker. Talk presented at Cambridge Basic Research. June 26, 1995 (abstract).

14. Tsal, Y. Movements of attention across the visual field. Journal of Experimental Psychology: Human Perception and Performance, 5, 1983 pp. 523-530.

15. Clark, A. *Being There: Putting Body, Brain, and World Together Again* MIT Press Cambridge, MA 1996

16. Suchman, L. *Plans and Situated Actions: The problem of Human-Machine Communication*. Cambridge, Cambridge MA 1987.

17. Pylyshyn, Zenon W. The imagery debate: Analogue media versus tacit knowledge. *Psychological Review* 88(1) 1981. pp. 16-45.

18. Ullman, S.Visual routines. *Cognition*, 18, 1984 pp. 97-159.

19. Greenberg, S. The Computer User As Toolsmith: The Use, Reuse and Organization of Computer-Based Tools (Cambridge Series on Human-Computer Interaction, No 6) 1993

20. Tauscher, L. and Greenberg, S. (1997). Revisitation Patterns in World Wide Web Navigation. In Proceedings of the ACM SIGCHI'97 Conference on Human Factors in Computing Systems, Atlanta, Georgia, March 22-27, ACM Press.

21. Dill, J. Bartram, L., Ho, A., and Henigman, F. A continuously variable zoom for navigating large hierarchical networks. *Proceedings of the 1994 Conference on Systems, Man and Cybernetics*, pages 386-390, 1994.

22. Ballard, D.H., Hayhoe, M.M., & Pook, P.K. Deictic codes for the embodiment of cognition. *Behavioural and Brain Sciences* (in press).

23. Fisher, B.D., Agelidis, M., Dill, J., Tan, P., Collaud, G., & Jones, C. CZWeb: Fish-eye views for visualizing the World Wide Web. In M.J. Smith, G. Salvendy & R. J. Koubek *Design of Computing Systems: Social and Ergonomic Considerations: Volume 2, Advances in Human Factors/Ergonomics,* 21B. pp 719-722, Elsevier. Amsterdam 1997

24. Klein, G. Sources of Power: How People Make Decisions. MIT Press Cambridge, MA 1998

25. Hutchins, E. *Cognition in the Wild*. MIT Press Cambridge, MA 1995