

Trends, Fashions, Patterns, Norms, Conventions . . . and Hypertext Too

Einat Amitay

Department of Computing, Division of Information and Communication Sciences, Macquarie University, NSW, 2109, Australia, and Commonwealth Scientific and Industrial Research Organisation, CSIRO Mathematical & Information Sciences, Locked bag 17, North Ryde, NSW, 1670, Australia. E-mail: einat@ics.mq.edu.au

This article describes the theoretical approach behind the InCommonSense system. This approach makes use of writing conventions on the Web. The theory behind InCommonSense is based on research findings from the fields of linguistics, psychology, HCI, sociology, and information retrieval. The theoretical background for finding and reusing conventions in hypertext is discussed, and the possibilities of extending the system to improve existing hypertext systems and creating new futuristic hypertext meanings (in particular on the Web) are examined.

Introduction

Among the many new things I have learned in my undergraduate years there was one thing that led me to carry on studying. This thing is a short definition of the term *language*. It was made by my professor for the history of the English language, and was an excuse for every shift and turn that occurred in English. "Language," he said, "is maximum communication with minimal effort." Because English mutates every other day, this mantra was repeated very often. So often, in fact, that it became part of my understanding of the world.

This definition breaks the problem of language evolution into two smaller questions: first, what is the communicative goal of the language change; and second, what would be its cost in terms of cognitive and cultural effort. In this article I will address the latter—how people try to reduce the cognitive and cultural cost of a recent language change. The language change this article studies is hypertext writing on the Web.

The cost of language change is often reduced through the development of conventions. Conventions are agreements between people about how they should act in certain situations. In fact, language is a set of conventions. Although today we refer to grammar and spelling checkers to make sure we write properly, both spelling and grammar are

language conventions that evolved from repeated mistakes. Conventions can also be found in vocabulary choice, styles of speech and writing, styles of layout design, arrangement of ideas, and so on. Some of these language conventions are not defined formally; however, most people would detect immediately when they are violated.

Hypertext is a recent small change in the use of language. During the last decade there have been many studies claiming that there is great cognitive effort in reading and comprehending this form of text, creating the "lost in hyperspace" phenomenon. This article, however, is about how people try to facilitate this effort through forming new conventions of writing in hypertext. Some of these conventions are adopted from the paper form of text, and some of these conventions are new and still shifting and improving. The hypertext writing that this article will focus on is found on the Web, which is where most people interact with hypertext today.

This article will first outline the theory behind the formation of language conventions in general, then it will go on to show where and how conventions are evolving in the community of people writing hypertext on the Web. Finally, it will demonstrate how these conventions can be used to augment and shift the meaning of the already published hypertexts. This demonstration is part of a system called InCommonSense, which reuses particular hypertext language conventions to retrieve and assemble people's annotations and descriptions on the Web.

What's a Language Convention

People have different notions of the meaning of the word *convention*. Some would find it to be a loose agreement, and some might call it a convention only if it exists for a long time and even defined in written guidelines. To level the ground first, let us look at the definition of the term *language convention* as it is defined in the context of translation.

Toury (1995) defines language conventions as socio-cultural constraints that stretch along a continuum between two extremes: general, relatively absolute rules on the one hand, and pure idiosyncrasies on the other. Some of these conventions are stronger, more rule-like, while others are weaker, almost idiosyncratic. The conventions (or what some people refer to as norms) are floating on this continuum, changing through time and social necessity: "thus, mere whims may catch on and become more and more normative, and norms can gain so much validity that, for all practical purposes, they become as binding as rules; or the other way around, of course."

In a more recent article Toury (1998) surveys the work of Davis (1994), in the context of sociology and social anthropology, to claim that "social creativity" and "agreements about actions" are always negotiated, with or without the intervention of language, and that this process, which requires some time, results "in the establishment of social conventions, according to which members of the group will behave when they find themselves under particular circumstances." He then goes on to claim that "conventions are a necessary outcome of any striving for social order, as well as a means for its attainment and maintenance. At the same time, they are not specific and binding enough to serve as guidelines for (and/or a mechanism for the assessment of) instances of behaviour and their products."

From this definition it can be understood that the meaning of language convention is very flexible, and depends on the community in which it is created. In the case of the very young Web community, the conventions are probably still shifting from the weaker side of the continuum to the more binding side of it. The following section shows trends and patterns already detected in this new community. Some of these fashions and patterns are language related, and some of them are more abstract and depend on the technological context of the Web.

Conventions in Hypertext

This section is a small survey of studies and research that look at emerging patterns in the on-line community. Some of these studies are based on empirical evidence and experimentation, while others are based on personal observations and educated predictions. From these studies emerges an image of a new social order that has formed within itself patterns, codes, and fashions. Erickson (Erickson, 1996) describes this as the transformation of the Web into social hypertext. He refers to the patterns of behavior as fashions, where the community defines what is acceptable and what is not. In his conclusion he predicts that "if [the Web] turns out to be a medium that allows the construction, negotiation, and propagation of the styles of appearance we refer to as fashion, its impact may be profound indeed."

Patterns of social behavior is also intrinsic to the work of Donath (1995). She is trying to simulate, in an electronic environment, the visual cues or patterns of activity we detect when a crowd is choosing to move this way or the

other in the real world: "the ebb and flow of rush hour traffic or the swift appearance of umbrellas at the onset of a rain-shower . . . the sea of business suits streaming from a commuter train . . . the bright t-shirts and sun-glasses of tourists circling a historic site." Her system finds social patterns in the electronic community and makes them visible. Users can view the structure of the electronic community and also perceive temporal patterns created by electronic interactions between other members of the community.

In the last decade or so, users of the electronic world have become aware of these electronic social currents and by clicking their way through the texts, have developed likes and dislikes. Authors are cutting and pasting their preferred patterns into their Web space and deleting portions that are less fashionable, not readable or poorly designed.

This semisocial sifting created new common taste as Nielsen (1998a) observed at the beginning of 1998: Web users become more and more conservative in their preferences of Web designs. He identifies "an increasing reluctance among users to accept innovations in Web design. The prevailing attitude is to request designs that are similar to everything else people see on the Web . . . The Web is establishing expectations for narrative flow and user options and users want pages to fit within these expectations." Nielsen suggests that the reason for this emerging public taste is that users navigate frequently between pages on different sites and that "the entire corpus of the Web constitutes a single interwoven user experience rather than a set of separate publications that are accessed one at a time."

In the same year, Bernstein (1998) suggests that there are patterns in hypertext linking. He identifies and names patterns from observation and makes a brave and a seemingly correct claim: "The reader's experience of many complex hypertexts is not one of chaotic disorder, even though we cannot yet describe that structure concisely; the problem is not that the hypertexts lack structure but rather that we lack words to describe it . . . I do not argue that the observed structural patterns are uniquely desirable, that superior patterns cannot be devised, or indeed that the writers of these hypertexts meant to use these patterns at all. I do propose that by considering these patterns, or patterns like them, writers and editors may be led to more thoughtful, systematic, and sophisticated designs."

To continue this interesting trend Huberman, Pirolli, Pitkow, and Lukose (1998) discovered through extensive empirical studies of different user communities, that there are strong statistical regularities of Web user surfing patterns. They claim that discovering patterns within the electronic community is important "because the sheer reach and structural complexity of the Web makes it an ecology of knowledge, with relationships, information 'food chains,' and dynamic interactions that could soon become as rich as, if not richer than, many natural ecosystems."

Because most of the above studies describe the Web as a whole and the patterns detected are fairly general, it has

been decided to focus in this work on the emergence of a within-node convention: how people choose to position links within paragraphs. This shifts the focus from the hyperspace as a tangle of links to the nodes as points of departure. To study patterns in hypertext writing there is a need to understand the unique process in the creation of Web hypertext. The next section describes this process by introducing the new reader–author role on the Web, where the reader becomes an author and the author becomes a reader at the same time.

The Reader and Author Roles and Their Convergence

The on-line community is made out of readers and authors. Each reader may choose to become an author. At any given time authors may choose to change their writing according to something that they have seen and read somewhere else. The role of the reader–author is cyclic, they can constantly improve their writing, linking it to other texts, or they may change their texts altogether, leaving no trace of what was there before. This dynamic process is taking place on the Web, communities of readers–authors merge and divide, links are made between ideas to form new ideas, and new social structures are created. From this interaction, like from any other human interaction, social codes are being formed. These codes can be codes of behavior (e.g., copyrights, moral codes) (Lessig, 1999), stylistic codes (e.g., promotional writing vs. critical writing), technical agreements (e.g., plug-ins that can be used by any browser, fonts that are large enough to read), etc.

The idea that when people write they apply their knowledge as readers is not new. Bizzell (1982), in the context of literacy and composition research, suggests to combine two theoretical approaches to writing: the authoring inner-cognitive processes, where cognitive and human perception are in control, and the social context of language use, where world knowledge and sense of context rule. In this sense, Web authors write for and in a discourse community that has conventional ways of interpreting and understanding the particular written discourse used. Thus, when they write on the Web they employ not only their knowledge about the medium and its audience, but also their experience as online readers.

This merge between the author and reader roles on-line are exemplified by Hill, Hollan, Wroblewski, and McCandless (1992). They describe a collaborative interface to help edit and read online documents shared by a group of users. They suggest that the “entwined concepts of authorship and readership are changing . . . to accommodate their emerging computational form.” This is claimed in their work to provide a motivation for observed reader/author behavior changes occurring when reading and authoring activities shift from the private domain to the semipublic domain.

The next sections will discuss the cognitive and historic reasons for the emergence of paragraph writing conventions in general. Then we will bring evidence for the existence

of these conventions in Web hypertext. Finally, it will be illustrated how most of these on-line conventions recently became written guidelines.

Reasons for the Emergence of Paragraph Conventions

This article focuses on writing within single nodes of Web hypertext. However, to compare Web hypertext conventions to the already existing conventions in writing on paper, we will now look at research conducted in relation to paper reading and writing conventions. Because this work is related to text segments in hypertext we will specifically look at conventions related to the rhetorical structure and physical layout of paragraphs.

Kieras (1978) defines the structure of the conventional paragraph as having a topic sentence first, followed by connected, coherent sentences. In his research he found that when the conventional paragraph structure was not kept, subjects were choosing fewer topical proposition as the best title; performed the task with longer reading times, and had lower recall of passage content. From these findings he concludes that the role of paragraph structure conventions is to minimize memory load. Goldman, Saul, and Cote (1995) suggest that readers use conventional paragraph structure as a heuristic for facilitating the reading process, while the interpretation of the paragraph content relies on semantic information in the text which interacts with prior knowledge.

Note that these findings do not suggest that the conventional paragraph arrangement is the only arrangement possible. However, they do support the claim that people look for familiar patterns in the writing to reduce the effort of looking for the important bits of information in the text.

Not only do people rely on paragraph arrangement to facilitate the understanding of the text, they also form a mental relation between the ideas presented and the overall layout of the text. To reinforce this claim Stark (1988) found that the position and presence of paragraph cues affect what ideas in the text were considered to be important. Kieras (1981) also found that two-topic passages were processed differently from the usual one-topic passages in prose memory studies.

According to Aikman and O’Hear (1997), people have been aware of the conventional paragraph arrangement for at least 500 years. They studied 15th- to 19th-century American authors of scientific, social, humour, philosophy, and popular prose, to claim that they all used main idea statements and placed them in the first, second, or last sentences in a paragraph. So, if people practice paragraph conventions for 500 years, would they continue to practice them when paper is replaced by screen?

The Role of Paragraph Conventions in Hypertext

In the last decade we have seen many new hypertext systems. Some of those systems resemble what we now see

on the Web, but most of them are designed for specific tasks and within defined contexts. Much of the research looks at non-Web hypertext systems to suggest how to improve the interaction with the hypertext they present. In this section we will try to bring together findings from both types of systems: hypertext on the Web and hypertext off the Web to examine the role of the paragraph convention.

Again, in many of the studies of hypertext, researchers put an emphasis on the changing roles of author and reader. They point at the fading borderline between these two roles. Foltz (1996) talks about comprehension and coherence mechanisms in hypertext. He surveys the role of the narrative schema of a text, pointing out that "placing a text in a well known narrative schema can improve the comprehension of that text." Foltz predicts that "as hypertexts become more accepted and widespread, writers of hypertext may develop standard rhetorical styles. Readers who are then familiar with those rhetorical styles can use that knowledge to help in their structuring of the information in an effective manner."

To maximize the coherence of the text, to facilitate comprehension of hypertext, and to support familiar reading strategies, Foltz suggests that authors should provide an appropriate context for readers. Such context would be a clear text structure that will aid the reader in finding coherent paths within the hyperspace. He also suggests providing good labels for the nodes because some readers tend to use a label-following strategy, "labels that clearly indicate the role of a particular node, will help the reader in successful guessing of the appropriate coherent [navigation] path." It is also suggested that authors should provide a structure similar to that of a standard rhetorical structure in linear text, so that readers will be able to reuse their familiarity with this structure and exploit it to their advantage without adding to their cognitive effort of comprehending the text.

Dillon and Vaughan (1997) contradict the claims of Foltz by writing that there is no need to base digital documents on the forms that evolved in nondigital documents. However, they recognize the need of users to hold on to something familiar that matches their experience from reading on paper. To satisfy this, they study the influence of structural-shape (layout and sequencing of information) on readers' perception of texts. They define shape as "a property of information that is conveyed both by physical form and information content."

Dillon and Vaughan look at readers' ability to identify meaning associated with perceptual cues in paragraphs of text. In several studies of digital and nondigital documents they bring evidence to the claim that physical properties of information structure are tightly linked with conceptual properties, and that readers who are familiar with these properties seem to be aware of them and to respond to many of them automatically. The shape of the text seems to be a useful pragmatic method for creating regularities in the discourse "that are followed by both creator and user in an ideal sense." Their study concludes with a statement that the

shape of text and its meaning is shared by a discourse community and is guided by sociocultural forces.

The connection between text layout on the screen and its understanding is also made by Grabinger and Amedeo (1985). In the context of text layout on screen, they asked participants to rate models of computer-generated text on the perceived ease with which it could be read and studied. They found that participants were guided in their judgments by the overall structure, simplicity, and spaciousness of the documents.

These findings seem to confirm that people employ their knowledge from writing on, and reading from paper in the context of the new medium. So far, the only difference found was related to the spaciousness of the text on the screen.

To strengthen these claims, Morkes and Nielsen (1997) have written guidelines for writing on-line texts on the Web based on extensive user studies. Guidelines are based on their findings that some users read only the first sentence of each paragraph, and that most users do not read the text at all, they just scan it. The guidelines are not new, and it seems that nothing much has changed: (a) use topic sentences; (b) place topic sentence at the beginning of the paragraph; (c) limit each paragraph to one main idea; (d) text should be short (or at least broken up); (e) users like summaries; (f) simple and informal writing is preferred.

In other columns Nielsen (1997, 1998a, 1998b) expands these guidelines to add another aspect to the traditional structuring of the paragraph. This new aspect is focusing on the anchor placement and naming. Anchors, according to his findings, enhance users' understanding of the text by being highlighted keywords that are distinctive from other words in the text. Anchors also improve comprehension if placed at the beginning of the paragraph or list items.

However, hypertext anchors are not usually complete sentences but largely short noun-phrases (Amitay, 1997). This finding leads us to consider anchors at the beginning of a paragraph to be titles or paragraph headings that were found to facilitate recall of unfamiliar topics (Lorch & Lorch, 1996). To take this even farther, in Web hypertext writing the anchor at the beginning of a paragraph might replace the topic sentence altogether. This transformation of the topic sentence into an anchor might seem only a slight one, but in fact, its visual characteristics cause people to detect and reproduce the pattern more easily. As we will later demonstrate, this pattern is so easily spotted that it can be retrieved automatically. This claim is supported by studies of Web hypertext documents of Chakrabarti, Dom, Gibson, Kleinberg, Raghavan, and Rajagopalan (1998), which indicate that there are more occurrences of relevant topic-related terms within and after the anchor than before it.

The use of language in Web hypertext was also found to be slightly different. Language use patterns were examined in a study of a 1,000 personal Web pages (Amitay, 1997). It was found that active verbs and simple, factual language are used in most personal home pages, and that vocabulary choice is limited to simple, unsophisticated words. Findings

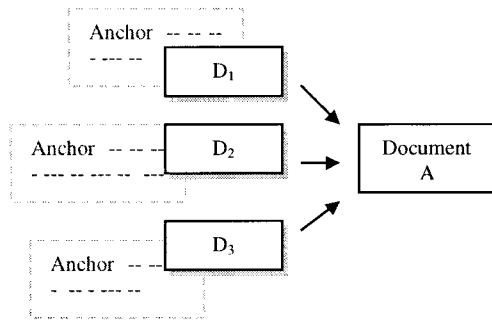


FIG. 1. Relations that InCommonSense looks for.

supporting these claims were recently published by Haas and Grams (2000). In a similar context, the context of e-mail writing, Rice (1997) analyzed 200 samples of e-mails to find rhetorical patterns. His findings show writers prefer simple coordinate sentences, brief paragraphs, and active verbs.

The change is more evident now. The convention of layout or shape of text is preserved from the paper form, and so are the placement of the topic sentence and the single main idea. However, the topic sentence itself might be replaced by a linked anchor, the language is more simple and factual, and because of screen limitations paragraphs seem to be shorter. If the reader is not convinced yet that there are such changes, the next section provides evidence to the existence of the above conventions, that by now became guidelines.

The Convention Becoming a Guideline

As we have seen so far, trends become conventions and conventions become rules quite rapidly. It is not surprising then that by the middle of 1999 the W3C published written guidelines on how to write online paragraphs (Chisholm, Vanderheiden, & Jacobs, 1999). Although there is no bibliographic reference in the guidelines to indicate where this understanding of on-line writing emerged from, it is quite reasonable to assume that it emerged from common social sense and convention. Chisholm et al. enumerate techniques for helping readers comprehend a Web page or a site:

- (1) Strive for clear and accurate headings and link descriptions.
- (2) State the topic of the sentence or paragraph at the beginning of the sentence or paragraph.
- (3) Limit each paragraph to one main idea.
- (4) Avoid slang, jargon, and specialized meanings of familiar words.
- (5) Favor words that are commonly used.
- (6) Use active rather than passive verbs.
- (7) Avoid complex sentence structures.

InCommonSense—A Proof of Concept

The claims in this article, that people have adopted a convention of writing paragraphs on the Web, can actually

be tested. This section presents a system that uses these regularities to automatically detect and reuse anchor–paragraph patterns. InCommonSense is a system that takes advantage of the paragraph convention found in Web hypertext. It automatically extracts annotations, ideas, descriptions, and other scribbles that people write about other texts, and assembles them. The vision behind InCommonSense is similar to the one described by Landow (1991) in his chapter “Reconfiguring the Author.” He envisions another form of linking through an information retrieval system that would automatically extract related information: “Other forms of linking will permit automatic data gathering, so that lists of relevant publications or current statements about [my document] created after I had completed [it] would automatically become available (p. 85).

The information used by InCommonSense to identify such statements is implicit, and the only guide for extracting this information is the convention of online paragraph writing and the convention of positioning anchors in them. The following paragraphs describe the system in brief. A more comprehensive technical description is given in Amitay (2000).

Overview of the System

The InCommonSense system takes a Web document A and looks for other documents D_i that point to it. Then the system assumes that the segments of texts found in the proximity of the anchors in D_i that point to A are related to A. The relations between the documents found by the system is shown in Figure 1. The following sections explain the system and details its components.

Using Paragraph Conventions in InCommonSense

The underlying approach in collecting information from arbitrary Web pages is that there is a pattern in the way people describe and link to other documents. This pattern is found in the way people write and annotate in hypertext. As shown before, the conventional paragraph writing has mod-

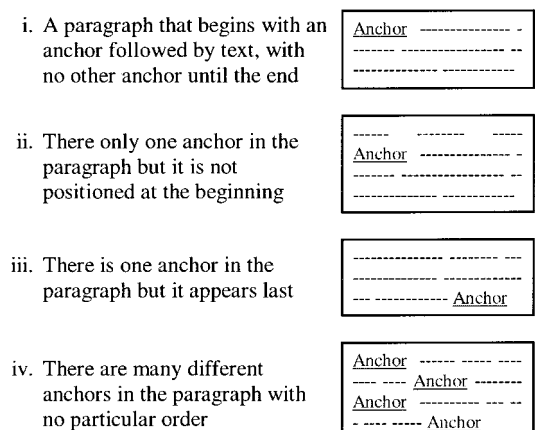


FIG. 2. Anchor positioning patterns within a paragraph of text.

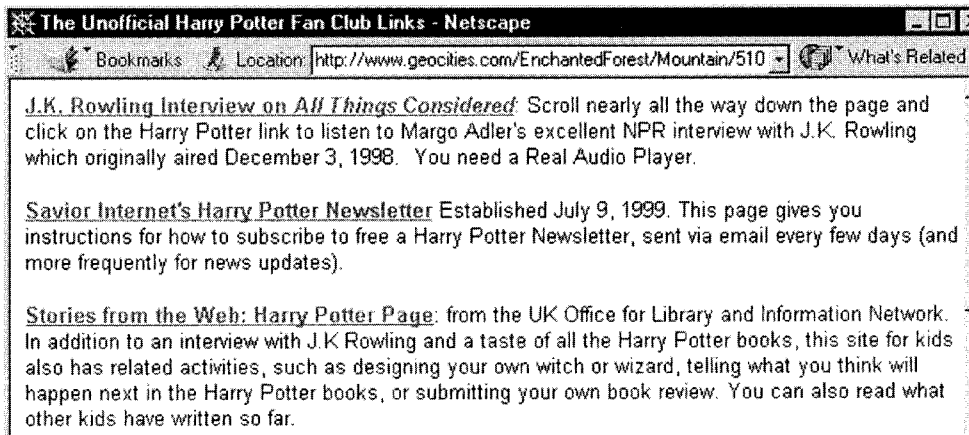


FIG. 3. Anchor positioning at the beginning of the paragraph.

ified itself in Web hypertext writing, and is now more visibly distinct than in its paper version.

The reason for this better visibility is the use of explicit markup. Thus, an anchor is marked in the code and is visible to the user as a distinct entity, in color, font type, and/or underlining. These cues are meant to show the exact place of the link, but they also guide the reading of the paragraph (Nielsen, 1998a; Ricardo, 1998).

On the Web there are different patterns of linking within the limits of a paragraph. These patterns can generally be viewed as four distinct patterns: The first pattern is where the paragraph begins with an anchor followed by text, with no other anchor until the end. The second pattern is where there is only one anchor in the paragraph but it is not in the beginning. The third pattern is where there is one anchor but it appears last, usually with anchors such as: “click here,” “more information,” “full story,” etc. The fourth pattern is where there are many different anchors in the paragraph with no particular order. Figure 2 shows these relations between anchor and paragraph arrangements.

By following the Web hypertext paragraph convention carefully, we can almost always predict that the first pattern, where a paragraph begins with an anchor followed by text, can be interpreted as a topic anchor followed by text that relate to it. This form of writing is practised by most of the on-line directories, and is becoming more widespread everyday. In fact, in an empirical experiment with 250,000 Web documents we have detected this pattern to appear more often than any of the other patterns, and in many cases even 50% more than any other anchor–paragraph pattern. Figure 3 is an example for such arrangements.

This positioning of anchor at the beginning of the paragraph can be seen in other languages also. Figure 4 shows the same arrangement in Hebrew, which is a language written right to left. The convention is the same only the direction of the paragraph is different. Note that the anchor position is the right corner of the text, where the paragraph actually begins.

As mentioned by Aikman and O’Hear (1997), main idea statements can be placed last in a paragraph. However, on



FIG. 4. Anchor–text arrangement in Hebrew.

URL: <http://www.lycos.com/help/Web-anatomy.html>
 TITLE: LycosHelp - Web Anatomy
 ANCHOR: World Wide Web FAQ
 TEXT: Thomas Boutell's (and others') answers to some of the most frequently asked questions about the Web. And it's available in several languages.

URL: <http://www.hwg.org/resources/faqs/>
 TITLE: Frequently Asked Questions Index - The HTML Writers Guild
 ANCHOR: World Wide Web FAQ
 TEXT: This may be the most definitive FAQ ever - provided by Thomas Boutell. Check here first.

URL: <http://riceinfo.rice.edu/about/>
 TITLE: About RiceInfo and the World-Wide Web at Rice University
 ANCHOR: WWW Frequently Asked Questions (FAQ)
 TEXT: from comp.infosystems.www

FIG. 5. Snippets relating to <http://www.boutell.com/faq/>.

the Web, it seems that anchors that are positioned at the end of the paragraph are less topic related and more repetitive, iconic, cues like “more information,” “full story,” “click here,” etc. Many on-line newspapers were using this anchor arrangement but are now shifting to the anchor-first arrangement. The reason for this is probably the repetitive nature of these anchor labels that are not informative, and might be interfering with the reading patterns described earlier.

Across Pages Examples

InCommonSense collects information about specific documents by querying commercial Web search engines for links to the document specified (query of the type—“link: URL”). Then the system fetches the pages and analyzes them for paragraph markup cues (any markup line break). InCommonSense is looking for segments of text that have empty spaces before and after the text, and also segments of

text that are marked as different entities (list items, table cells, text indents, etc.). This analysis helps the system to detect visual cues like paragraph shapes and bulleted lists that readers might consider to be individual or stand-alone paragraphs of text.

If the anchor linking to the required page is followed by text, then both text and anchor are retrieved. For example, the snippets of text in Figure 5 relate to the document found in <http://www.boutell.com/faq/>

It is interesting to see how different people relate to the same document from different perspectives. As in Figure 6, the FDA site is a source for food and safety information, but it is also a source for information about vegetarian foods.

Currently, InCommonSense processes up to 220 documents relating to a single URL in one run (using Google, HotBot, AltaVista, and Infoseek). There is no limit to the number of documents processed except for the limits that the commercial search engines pose: The Web is reflected through the search engines used by InCommonSense, which means that the documents processed are only the ones that are detected by the search.

Summary and Conclusions

In this article we have traced the emergence of writing conventions in Web hypertext. A system, InCommonSense, that uses this information to retrieve snippets of text about related documents is also described. The conventions this article describes are only a small portion of many that exist. As professed by many of the articles cited, these conventions are useful in many contexts. It is easy to imagine the usefulness of InCommonSense: It can be used as a basis for an annotation system where users share annotations without formatting them for a specific system. It can also be used as an authoring tool for enhancing and updating documents. Automatic directory listings can use it for adding more

URL: <http://www.ai.org/safetynet/HEALTH/>
 TITLE: Health Safety
 ANCHOR: U.S. Food and Drug Administration
 TEXT: has extensive information and databases available on foods, human drugs, biologics, animal drugs, cosmetics, medical devices, and children & tobacco; and it has a search engine. A section for consumers is accessible from the bottom of the home page, and it contains an online version of FDA Consumer, the official magazine of the FDA.

URL: <http://www.vegetariantimes.com/links/>
 TITLE: Vegetarian Links
 ANCHOR: FDA Homepage
 TEXT: This online source to the U.S. Food and Drug Administration includes topics such as news, foods, human drugs, animal drugs and cosmetics--all of which have relevant applications to vegetarians and green thinkers alike.

FIG. 6. Snippets about the FDA site retrieved with InCommonSense.

opinions to their lists. The snippets can be used for improving categorization tasks such as the ones presented by Zamir and Etzioni (1999) and Attardi, DiMarco, and Salvi (1998).

The system can also be used to augment the hypertext writing by adding a link to parallel pages that connect to the same node. This feature is actually creating new hypertexts from already completed hypertexts, similar to the idea presented by Landow (1991).

In this work InCommonSense is used for improving search interfaces, supplying users with descriptions of documents as search results, similar to the ones presented by directory listings such as Yahoo!. This is done by adding another layer to the system which ranks the retrieved texts. The ranking is based on results from two on-line experiments that tested more than 700 Web users' opinions about what they consider to be a good textual description of a Web document. Because this article is about the theory behind the system and not about the system itself, we will not expand on this issue. More details about the InCommonSense system and its applications are available on the Web and in Amitay (2000).

Links Mentioned

Google—www.google.com; HotBot—www.hotbot.com; AltaVista—www.altavista.com; Infoseek—www.infoseek.com; Yahoo!—www.yahoo.com; InCommonSense—<http://www.ics.mq.edu.au/~einat/incommonsense/>.

References

Aikman, C.C., & O'Hear, M.F. (1997). Main idea: Writers have always used it. *Journal of Adolescent and Adult Literacy*, 41(3), 190–195.

Amitay, E. (1997). Hypertext—The importance of being different. MSc Dissertation, Centre for Cognitive Science, Edinburgh University, Scotland. Also Technical Report No. HCRC/RP-94.

Amitay, E. (2000). InCommonSense—Rethinking Web search results. New York: IEEE International Conference on Multimedia and Expo (ICME 2000).

Attardi, G., Di Marco, S., & Salvi, D. (1998). Categorisation by context. *Journal of Universal Computer Science*, 4(9), 719–736.

Bernstein, M. (1998). Patterns of hypertext. In *Proceedings of the ninth ACM conference on Hypertext and hypermedia: Hypertext98* (pp. 21–29).

Bizzell, P. (1982). Cognition, convention, and certainty: What we need to know about writing. *PRE/TEXT*, 3(3), 213–243. [Reprinted in Bizzell, P. (1992). *Academic discourse and critical consciousness*. Pittsburgh: University of Pittsburgh Press; and in Vitanza, V. (Ed.). (1993). *PRE/TEXT: The first decade*. Pittsburgh: University of Pittsburgh Press.]

Chakrabarti, S., Dom, B., Gibson, D., Kleinberg, J., Raghavan, P., & Rajagopalan, S. (1998). Automatic resource compilation by analysing hyperlink structure and associated text. *Proceedings of the 7th International World Wide Web Conference*, Brisbane; also in a special issue of the journal *Computer Networks and ISDN Systems*, 30(1–7).

Chisholm, W., Vanderheiden, G., & Jacobs, I. (Eds.). (1999). *Techniques for Web content accessibility guidelines 1.0 W3C NOTE 5-May-1999* <http://www.w3.org/TR/WAI-WEBCONTENT-TECHS/>.

Davis, J. (1994). Social creativity. In C.M. Hann (Ed.), *When history accelerates: Essays on rapid social change, complexity and creativity* (pp. 95–110). London: The Athlone Press.

Dillon, A., & Vaughan, M. (1997). "It's the journey and the destination": Shape and the emergent property of genre in evaluating digital documents. *New Review of Multimedia and Hypermedia*, 3, 91–106.

Donath, J.S. (1995). Visual who: Animating the affinities and activities of an electronic community. In *proceedings of ACM Multimedia 95*.

Erickson, T. (1996). The World Wide Web as social hypertext. *Communications of the ACM*, 39(1), 15–17.

Foltz, P.W. (1996) Comprehension, coherence and strategies in hypertext and linear text. In J.-F. Rouet, J.J. Levonen, A.P. Dillon, & R. J. Spiro (Eds.), *Hypertext and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldman, S.R., Saul, E.U., & Cote N. (1995). Paragraphing, reader, and task effects on discourse comprehension. *Discourse Processes*, 20(3), 273–305.

Grabinger, R.S., & Amedeo, D. (1985). CRT text layout: Prominent layout variables. *Annual Convention of the Association for Educational Communications and Technology*, Anaheim, CA, January 17–23.

Haas, S.W., & Grams, E.S. (2000). Readers, authors, and page structure: A discussion of four questions arising from a content analysis of Web pages. *Journal of the American Society for Information Science*, 51(2), 181–192.

Hill, W.C., Hollan, J.D, Wroblewski, D., & McCandless, T. (1992). Edit wear and read wear. In *proceedings of CHI 92* (pp. 3–9).

Huberman, B.A., Pirolli, P.L.T., Pitkow, J.E., & Lukose, R.M. (1998). Strong regularities in World Wide Web surfing. *Science*, 280(5360), 95–98.

Kieras, D.E. (1978). Good and bad structure in simple paragraphs: Effects on apparent theme, reading time, and recall. *Journal of Verbal Learning and Verbal Behavior*, 17(1), 13–28.

Kieras, D.E. (1981). Topicalization effects in cued recall of technical prose. *Memory and Cognition*, 9(6), 541–549.

Landow, G.P. (1991). *Hypertext: The convergence of contemporary critical theory and technology*. Baltimore, MD; Johns Hopkins University Press.

Lessig, L. (1999). The law of the horse: What cyberlaw might teach. *Harvard Law Review*, 113(12), 501–546.

Lorch, R.F., Jr., & Lorch, E.P. (1996). Effects of headings on text recall and summarization. *Contemporary Educational Psychology*, 21(3), 261–278.

Morkes J., & Nielsen J. (1997). Concise, SCANNABLE, and objective: How to write for the Web. Available online at <http://www.useit.com/papers/webwriting/writing.html>

Nielsen, J. (1997). How users read on the Web. In *Alertbox for October 1, 1997*.

Nielsen, J. (1998a). The Increasing Conservatism of Web users. In *Alertbox for March 22, 1998*.

Nielsen, J. (1998b). Microcontent: How to write headlines, page titles, and subject lines. *Alertbox for September 6, 1998*.

Ricardo, F.J. (1998). Stalking the paratext: Speculations on hypertext links as a second order text. In *Proceedings of the ninth ACM conference on Hypertext and hypermedia: Hypertext98* (pp. 142–151).

Rice, R.P. (1997). An analysis of stylistic variables in electronic mail. *Journal of Business and Technical Communication*, 11(1), 5–23.

Stark, H.A. (1988). What do paragraph markings do? *Discourse Processes*, 11(3), 275–303.

Toury, G. (1995). The nature and role of norms in translation. In *Descriptive translation studies and beyond* (pp. 53–69). Amsterdam: John Benjamins.

Toury, G. (1998). A handful of paragraphs on "translation" and "norms." In C. Schäffner (Ed.), *Translation and norms*. Clevedon etc.: multilingual matters, 1998 (pp. 10–32). [also available as Vol 5, Nos 1&2 of *Current Issues in Language & Society*].

Zamir, O., & Etzioni, O. (1999). Grouper: A dynamic clustering interface to Web search results. In *Proceedings of the 8th World Wide Web conference* (pp. 283–296).