Ubiquity and Common Sense for the Web

Leona F. Fass

P.O. Box 2914, Carmel CA 93921
lff4 [AT] cornell [DOT] edu

Abstract

Web access is ubiquitous in our community, giving us opportunity to observe it *via* our background in theoretical behavioral modeling. Our goal is to help advance the developing Web and, from the perspective of Web Intelligence and Security, identify areas where policy and design may improve. Issues we consider include adaptation of learning techniques, user education, and problems of security, privacy and trust. Observing everyday peoples' Real World access can assist in creating an optimal Web. With ubiquitous access, designers and developers need apply common sense.

Keywords: Web Intelligence, ubiquitous computing, security, HCI, learning.

1. Introduction

Issues of Web Intelligence and Security are of interest to us as a computer scientist and observer of everyday peoples' computer access. We have the opportunity to survey computer usage in "Real World" environments within our community. There local non-computer scientists access public and quasi-public machines for communication, information, and services on the Web. But our community is also a popular destination for tourists with varying levels of technical skills. They, too, may access many of our local public networks and machines, anonymously.

Such surroundings present us with a wide variety of computer-related activities unlike our own, that we would not ordinarily see professionally. This convinces us to adapt our theoretical research to assess Web successes and identify Web-related problems to-be-resolved. Thus we may influence access policy and suggest emphases for designers and developers who might improve the Web. In an environment of ubiquitous access, all need to apply common sense.

2. A Theoretician Views the Web

Web development includes aspects of human-computer interaction (HCI), automated learning and software design. When considering the Web, we see how our research results relate to each of these areas. We have relevant theoretical background in behavioral modeling, knowledge representation and formal system design that primed our interest in the developing Web. Our work is grounded in theory, but we always seek applications to practice.

In our original theoretical research we analyzed behaviors and devised techniques to construct behavioral models. Our results characterized information or behavior already observed, but also could be predictive of future events. The models constructed were optimized, relative to established criteria. We next developed analysis techniques for testing potential models, and adapted this work to practical problems to assist in detecting defects and flaws. Devising software and system potential models and testing them was just such a problem. The goal was to repair defects and correct errors, dynamically creating a model closer to the optimal one (e.g., see [5]).

The Internet and Web are dynamic and adaptable, and can always change and improve. Even independent of network issues, this is so when considering software aspects. That persons of varying skills and aptitudes have access adds an HCI and a learning aspect to the dynamic processes. By observing, developers can learn from humans to improve systems and services, the Web can learn from humans, humans can learn from the Web, and Web-dwelling software agents can learn from each other.

When accessing the current Web, or observing others do so, our tendency is to see how processes may be improved. Our aim is to suggest features that may have been overlooked, since complex behaviors and systems cannot be completely specified. Another aim is to detect defects to be corrected, since predeployment testing can *never* be adequate. Our experience differs from that of most theoreticians or developers, for we typically observe Real World *non*-computer scientists and analyze *their* processes and difficulties when accessing the Internet and the Web.

3. Ubiquitous Local Access

There are computing environments throughout our community where we interact with the Web, access deployed services, and observe others doing so. Policies and equipment may vary from place to place, but anyone seeking a Web connection can find it.

One location in which we have observed and participated in Web access is ***the library in the village where we reside***. There residents and visitors may obtain time-slots to use machines. Staff who assist them are trained librarians or clerks, but not computer scientists. We have observed [6] that librarians using the Web are trained differently than computer scientists and do not use computers the way we do. Wilder [22] notes that librarians are experts in reading and writing, and should not be expected to provide technological education. In our community, that is what they often must do. There are "techie" consultants for staff, not for patrons. The result is an evolving usage policy that never can meet the needs of patrons, staff, techies, and administration simultaneously. There also have been "surprises" for developers when deployed services have been accessed on these public machines. As library access has become less constrained (anonymous computer use is now possible) actual Web access has become more so. At times there has been monitoring and blocking of email, Web sites and services, in a constant battle against misuse (intentional or otherwise). The library relaxed its precautions and began charging an access fee (waived for some) to defray the costs of the inevitable e-problems brought in by the public (and the staff) [9]. The experiment was short-lived; at this time the library has returned to blocking attachments, some forms capability and some sites. We learn alot about Web Intelligence and Security by watching the general public using library equipment. One thing learned is that a satisfactory balance between security and privacy has yet to be achieved.

We also have had access to the computer laboratory in a nearby ***senior citizens activity center*** where computer use is heavily applications-oriented and very different from our own. There is much interest in computing, and many grandparents are taking their first "computer lessons" so they can learn to communicate electronically with family and friends. A 96-year-old neighbor takes classes there, for she likes to "keep current" in this information society. An Interest Group meets sporadically, to discuss innovations in computing and the Web. E.g., we were present at an online banking demonstration there. When several people tried it in that quasi-public lab (where access doesn't require "secret" passwords) we concluded that they trusted the Web too much. Instructors and lab monitors are volunteers, so knowledge can be spotty. E.g., one monitor claiming he's never been on the Internet oversees a lab network of Internet-connected, Web-enabled machines. There *is* some damage control: professionals clear the system twice weekly, removing users' files and Web refuse.

Local businesses throughout our community have Web access, some for public use and some for employees. Observations there relate to issues of security, privacy, trust and a need for common sense. E.g., we had a notable experience at a computer rental place, believing we were sending emails to colleagues. We later learned a previous renter had altered the system so all email was routed to his home, out-of-state. In another example, a "local" was convinced that a Web site would increase her business. She took a course at the senior citizens center, created the site with the assistance of a non-techie business partner, publicized it in newspaper ads, and was soon inundated with pornography. She disbanded the site (and dissolved the partnership). Lastly, a gallery director needed a Web site to inform customers and accept orders online, with an Internet connection at her place of work. She permitted the teenage child of a fired co-worker to check email on her machine. It was soon overwhelmed with spam (and 256 "bulk mails" in one sitting!) and spim that persisted after the teenager had logged off. She can't use the system anymore, nor can anyone else. But, we have yet to find a local business that does not have an Internet and Web connection. Even an Asian antiquities shop engages in ecommerce and needs information and services from the Web. Most businesses have the common sense to protect their installations from reconfiguring tourists. They enlist the services of skilled Web masters who know about firewalls, and they delete questionable emails when they arrive. And, unlike the library, which *must* give access to residents, they won't permit walk-in game-players to tamper with their networks and sites. Hence they use the Web quite successfully.

A travel agent friend, seeking to help us book a trip from Carmel CA to the JCIS held in Atlantic City NJ, carefully read the routing that the airline Web site produced. She saw the only possible automated routings resulted in travel for 29 hours, so that was *not* how we planned the trip. Sometimes good business, savvy computing and common sense *all* are knowing that the Web is something one can't always trust.

4. Applying Some Common Sense

Anyone needing or wanting Web access in our community can find it, in locations such as those described. As a computer scientist with theoretical background in behavioral modeling, we observe this activity to see how processes might improve. From the perspective of Web Intelligence and Security we identify areas where both users and developers can use some common sense. We review some relevant observations, and related work of other researchers.

A little knowledge is a dangerous thing. Years ago, "computer anxiety" protected systems and stored information; only a few had access by virtue of their expertise. Now "everyone" uses computers and the Web, without needing to understand how they work.

More than one senior citizen computer lab student has reconfigured or disabled the local network when trying to use the printer [7], or destroyed instructors' files on the "protected" hard disk. More than one public machine user (such as the computer renter who redirected emails) has tampered with equipment because he knows how to do it at home and forgot *this* set-up isn't his! Even without malice, much harm can be done. We recommend restricted access policies on public systems, and extended education and training of potential users. Then they may develop better computer and Web sense. We also advise that developers pay more attention to everyday peoples' computer and Web-related activities. What is seen among the self-selected populations of university and industrial research environments is very different from what is seen in the Real World. There are lots of inadvertent spimmers out there, and worse.

Developers must understand different environments. E.g., we'd like to see sites and services recognize if they are being accessed from public machines so they may take precautions. Techies at an online brokerage house were quite surprised when we alerted them that cookies they'd planted on the library system were revealing private account information to all subsequent users (described in [10]). A local system administrator can only do so much. Web designers and developers must assist too.

Experts can learn from non-experts. *Computer scientists* may think in keywords, logical connectives and links, but most people think in Natural Language (NL). In [6] we described learning to use NL search engine queries from our library staff. Cercone *et al* [4] suggest lifting NL techniques to become Web learning techniques, e.g., to improve Recommender Systems. Customer service Web agents that interact in NL may improve peoples' experiences, giving a business a competitive edge [2]. Agent-agent interactions described by [21] can be modeled on human language games. Furthermore, by observing the links activated by everyday human users [20] suggests semantics will emerge, to assist in creating Semantic Web links. Mining and analyzing everyday user behaviors can help us build Adaptive User Interfaces [18, 23]. Designers and developers need not force everyday people to think about the Web and use it the way the few of *them* do. It is common sense that the Web should be designed to think and act the way everyday people do. This includes providing proactive protections from well-meaning peoples' inadvertent errors.

Real experts make mistakes too. In the "big picture" specifications of complex systems can never be complete, nor can testing ever be adequate. The important thing is to detect anomalies and flaws so that the Web and Web services can be improved. In the "smaller picture" even mundane "applied behavior" can be riddled with mistakes. E.g., the techie consultants at our library frequently leave rebooted equipment set on "autocomplete", posing security and privacy problems to unknowing users. When we delivered [7] we and all "expert" users learned, after the fact, the conference Internet connection had similar security and privacy flaws. And, numerous Web-experts at a previous JCIS, booking hotel rooms on the Web, received emailed confirmations and trusted them. But they arrived and found rooms non-existent and confirmations bogus. Even for experts, common sense is to always double-check.

There's a risk-reward factor to ubiquitous access. Our own bank convinced us to try online bill-paying (despite that Interest Group demo), for we were satisfied the site and the bank offered sufficient protections. Thus far the rewards in saved time and money have outweighed the risks (using "private" machines). As indicated in [12], cost-effectiveness and indispensability go hand-in-hand. Conversely, we took a risk and made numerous unsuccessful attempts to register online for a conference, using one of our credit cards. The site and that bank protected us after the registration failed, telephoning immediately to see if the botched attempts indicated fraud. We would deal with that company again.

Often security must trump privacy. Our library presented a security problem [9] that we now update. A librarian observed "furtive" user behavior and suspected the user of sending email (prohibited at that time). As she approached she saw his behavior indicated an attempt to use library equipment to purchase a gun online. His privacy violated, the user fled. But the library monitored the system and presented police with evidence indicating the purchase was attempted with a stolen credit card. The police identified the culprit and alerted the gun company, which assured them the sale would not go through. The librarian recently told us she believes she should not have interfered with the user. We believe the site should have taken proactive steps, "red-tagged" the attempted purchase, and notified authorities.

Developers and business can learn from theoreticians. In our theoretical research (e.g. [5,8]) we found models producing desired behaviors by decomposing goal behaviors into subgoals and determining relationships among them. Components of fulfilling models corresponded to the behavioral components. When the goal changed we could dynamically adapt the model. Similarly, Web agent tasks, decomposed into subtasks, are effected by integrated AI techniques [2]. Learning linkage patterns (relationships among subgoals) produces rules to focus search engine results [1]. Collaborative filtering adaptively acquires preferences [16], learning Web content Recommender Systems should deliver. Learning Web mining techniques [15] aids automatic composition of Web services, crucial for the Semantic

Web [17]. Agent systems dynamically alter configurations to maximize dependability [11]. Coalitions can collectively fulfill tasks, with subtasks fulfilled by components, dynamically modifiable when alternates are identified [19,8]. Business processes may choreograph Web services for automation [14], arranging components to interact properly for adaptability.

Humans are still needed in the loop. Service agents described by [2] use NL to interact with online customers, but defer to human representatives when responses aren't satisfactory (and customer patience seems ready to run out). The airline automated routing of an absurd itinerary to JCIS in NJ implies the need of a human expert (to suggest flying to Philadelphia and taking a limo or train). Just our travel experiences [8] convince us that Recommender Systems need common sense components to allow for the "unexpected". And, perhaps, to defer to humans.

There are Web-related problems Web developers will never solve. Non-technical problems of information trustworthiness and behavioral ethics are among them. As we write, the week's news was topped with two relevant stories. One decried the level of dishonesty in online auctions. The other celebrated the first anniversary of legislation against spam, noting its apparent increase during the year. Education and certification processes may decrease such problems. But these problems will never go away completely.

5. Conclusions

We agree with [12] that ubiquity of access can make the Internet (and Web) essential to most people, but we disagree with the [12,13] view that economic factors still define the "digital divide". In our Real World, training and education may define the divide [7], as may a "need to know" culture. "All knowledge" may soon be digitized [3], but we have plenty of Ivy-educated friends who never use the Internet or Web, and plenty of Real World clerk friends who do.

By observing the wide variety of computer and Web users in our community, we have identified some issues of Web Intelligence and Security that we hope may influence access policies and assist developers and designers in improving the Web. When these involve matters of security, privacy and trust, we concur with [12] that ubiquity has a "dark side".

In our community computer, network and Web access are ubiquitous, in the sense that anyone seeking access may gain it easily. This brings with it a need for common sense, both to hone acceptable access policy and to develop processes that are of use to everyday people in the Real World.

6. References

[1] S. Altingovde and O. Ulusoy, "Exploiting Interclass Rules for Focused Crawling", in [23], pp. 66-73.

[2] M. Barbuceanu, M.S. Fox, L. Hong, Y. Lallemant and Z. Zhang, "Building Agents to Serve Customers", *AI Magazine*, Fall 2004, pp. 47-60.

[3] S. Carson and J.R. Young, "Google will Digitize and Search Millions of Books from 5 Top Research Libraries", *Chron. Higher Ed.*, Jan. 7, 2005, pp. A37-40.

[4] N.Cercone, L.Hou, V.Keselj, A.An, K. Naruedomkul and X. Hu, "From Computational Intelligence to Web Intelli-gence", *IEEE Comp*, Nov. 2002, pp.72-76.

[5] L.F. Fass, "Determining Software Models that are Less Incorrect", *Model Based Validation of Intelligence*, Stanford Univ., March 2001, AAAI Press, SS01-04, pp. 113-116.

[6] L.F. Fass, "Can We Improve Web Access in the Real World?", Statement of Interest in [17], pp. xv-xvi.

[7] L.F. Fass, "The 'Digital Divide' Just Isn't What It Used to Be", in *Proc. of the Int. Conf. on Software Engineering IFIP Workshop on Bridging the Gaps Between SE and HCI*, Edinburgh Scotland, May 2004, The IEE W1L: pp.83-87.

[8] L.F. Fass, "An Automata-Theoretic View of Agent Coalitions", in [19]: pp. 18-21.

[9] L.F. Fass, "Small Steps and Giant Leaps toward Homeland Security", in *AI Technologies for Homeland Security*, Stanford Univ., March 2005, AAAI Press, SS05-01, pp. 133-136.

[10] T. Finin and B. Grosof, (Eds), *AI for Electronic Commerce*, Orlando FL, July 1999, AAAI Press, WS99-01.

[11] M. Greaves, V. Stavridou-Coleman and R Laddaga (Editors), "Dependable Agent Systems", *IEEE Intell Sys*, Sept./Oct. 2004, pp. 19-70.

[12] D.L. Hoffman, T.P. Novak and P. Venlatesh, "Has the Internet become Indispensable?", *Comm. of the ACM*, July 2004, pp. 37-42.

[13] L.J. Jackson, A. Von Eye, G. Barbatis, F. Bocca, H.E. Fitzgerald and Y. Zhao, "The Impact of the Internet on the Other Side of the Digital Divide", *Comm. of the ACM*, July 2004, pp. 43-47.

[14] N.Leavitt, "Are Web Services Finally Ready to Deliver?", *IEEE Comp*, Nov. 2004, pp.14-18.

[15] B. Liu, R. Grossman and Y. Zhai, "Mining Web Pages for Data Records", in [23], pp. 49-55.

[16] D. Pavlov, E. Manavoglu, C.L. Giles and D. M. Pennock, "Collaborative Filtering with Maximum Entropy", in [23], pp. 40-48.

[17] A. Pease, R. Fikes, and J. Hendler (Editors), *Ontologies and the Semantic Web*, Edmonton AB, July 2002, AAAI Press, WS02-11.

[18] S. Rogers and W.Iba (Eds),. *Adaptive User Interfaces*, Stanford Univ., March 2000, AAAI Press, SS00-01.

[19] L-K Soh and J.E. Anderson (Eds), *Forming and Maintaining Coalitions & Teams in Adaptive Multiagent Systems*, San Jose CA, July 2004, AAAI Press, WS04-06.

[20] S. Staab (Ed), "Emergent Semantics", *IEEE Intell Sys*, Jan./Feb. 2002, pp. 78-86.

[21] L. Steels, "Language Games for Emergent Semantics" in [20], pp. 83-85.

[22] S. Wilder, "Information Literacy Makes All the Wrong Assumptions", *Chron Higher Ed*, Jan 7, 2005, p. B13.

[23] Q. Yang, C.A. Knoblock and X. Wu (Editors), "Mining Actionable Knowledge on the Web", *IEEE Intell Sys*, Nov./Dec. 2004, pp. 30-73.