



STILL NOT BLEEDING

Robert Filman • Lockheed Martin • filman@mcc.com

This month the Spider seeks out stuff on the cutting edge.

THE LIBRARY WEB

When the World Wide Web isn't being compared to a super highway, it's cited as a library. The Spider first visits a pair of sites that emphasize their relation to a classical athenaeum.

Alexa • www.alexacom.com

The ancient library at Alexandria, Egypt, attempted to assemble all the world's knowledge in one place. Alexa attempts to compile and preserve the entire World Wide Web—or at least its text. (In a properly anti-Orwellian mode, you can't even completely bury your mistakes anymore. Publish, say, a not so well-researched article on, say, nerve gas use in Vietnam, and before you can withdraw it, there it is, preserved on Alexa for all the world to retrieve.)

Alexa bills itself as a "free advertising-supported Web navigation service." You download the software (I tried Alexa 2.0 for Internet Explorer on NT), which gives you a toolbar at the bottom of your browser window. The toolbar is divided into sections, showing the relative popularity of the site you're visiting, suggestions for similar sites (weighted, I fear, towards the advertisers who support Alexa), archived copies of pages that are miss-

ing or 404'd, and single-button access to several reference volumes.

I tried the system by visiting a music club (bmgmusic.com). Alexa was kind enough to offer me both single-button access to a competitor (Amazon) and the ability, at the click of the mouse, to bring up an old (and doubtless nonfunctional) copy of BMG's home page. I liked the easy access to the reference volumes, though a touch of effort could get my browser window configured the same way. Most of the time, the recommended alternative pages are useless, though every now and then they've got something worth visiting. The relative popularity statistics are good only for a laugh, as the most popular place I seem to visit is Alexa. Sometimes the page archive and similar page references are useful, though they are hardly impartial recommendations.

I guess I'll leave it installed for a while. 🐜 🐜 🐜

IBM Patent Server • www.patents.ibm.com/ibm.html

Alexa seeks to catalog the chaos of everything. IBM's Almaden Research Center has a catalog of the certifiably novel and undeniably sequential—US patents of the last 27 years. This effort was intended both to show how multimedia information can be treated as

database elements and to provide a useful tool for IBM engineers. In a magnanimous gesture, IBM makes the patent server freely available (even without advertising).

The patent server allows search by patent number, very simple Boolean searches, and searches relative to the descriptive fields of a patent. The database includes over 1.7 terabytes of text and images. The patent server's novel technology includes encryption and compression (so no one will see you transferring those big patent image files across the Internet and figure out what you're working on). A quick try of the server turned up 418 patents mentioning "Internet" while the Boolean search for "wide" and "web" timed out.

While you're there, check out the obscure patent gallery (<http://www.patents.ibm.com/gallery.html>), which features reader-nominated unusual patents such as number 4887543, "Unforgettable umbrella method" (a tag that goes on your key ring) and number 4978948, "Combined earthquake sensor and night light" (the light is projected through a suspended crystal, which produces diffraction patterns on earthquakes, the passing of large trucks, or people walking around the room). 🐜 🐜 🐜 🐜

Very good, but particularly for a database tool, I'd like to see a richer search engine.

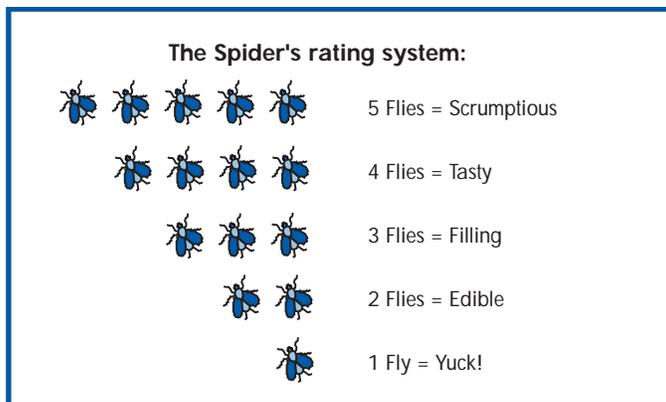
WHAT A CHARACTER

In the virtual library, who will replace Marion, Madame Librarian? Search engines are pretty stiff. The Spider next visits a pair of sites that offer a more lively alternative.

Extempo • www.extempo.com/

In the mid-seventies, Ken Colby and his students at the Stanford AI Lab wrote "Parry," a simulation of a paranoid schizophrenic having a conversation with a therapist.

Parry had several notable features. First, it did parsing by pattern matching. At a time of rich theories on how to understand (written) natural language based on elaborate parsing mechanisms and complex underlying representations, Parry's language



understanding bordered on keyword lookup. Another feature was that Parry's internal state ("emotions") effected (or affected, if you prefer) its responses. For example, if you cursed Parry, he became angry and less cooperative. And third, the project clearly had no economic impact. Who needs a simulated paranoid schizophrenic when a real one is so inexpensive?

Well, at least no economic impact until now.

The Spider visited Extempo, a Santa Clara startup that "develops interactive Imp characters for applications in e-commerce, entertainment, and corporate communications and training." These characters can be used for Web-based tours, advertisements, or help-desks—the theory being that some class of users will respond better to a system that includes an anthropomorphic visualization than they will to pure text. (The theory also skirts neatly around the problem of people expecting more understanding from an anthropomorphic representation than it delivers.)

The site includes a good description of the underlying pattern/action technology. Like Parry, the characters are meant to have emotions (and vary their responses by their emotions). Unlike Parry (this is 1998, after all), elaborate graphics are used to display the characters. Like Parry, the parsing still seems to be based on simple keyword- and pattern-matching routines. Unlike Parry, the authors don't seem to have spent a few years accumulating the actual patterns people use.

The Extempo characters are supposed to be able to talk about life in general (like a bartender), but I had a

pretty low success rate in getting them to respond appropriately to simple variants of straightforward, on-topic comments. Parry convinced many people he was paranoid. The characters at Extempo get a diagnosis of "receptive language impaired." 

Needs work.

Neuromedia Inc. • www.neurostudio.com

Want to do this stuff at home? Neuromedia will sell you their NeuroStudio tools for creating your own animated, interactive characters.

The Spider gave the Neuromedia robots a spin. Unlike Extempo, they talked, but also unlike Extempo, there was no evidence of emotional state. The robots could answer elementary questions about the company but were far more ignorant about the underlying technology than Extempo's. As a shining example of Neuromedia success, their robot informed me that "GovCon built a bot named Julie-1 with NeuroStudio. Julie-1 reduced e-mail traffic on their site by 80 percent in its first month." While this may seem convincing evidence of the technology's value, I have known companies that achieved similar results merely by switching mail programs. (More seriously, another part of the site informs us that Julie-1 has reduced e-mail requests for customer support by 80 percent. I'm not sure if this means that the customers found what they were looking for with the bot that they were not finding with the FAQ, or if it means they gave up in frustration. I'd also be curious as to how the bot compares to a simple search engine.) 

Calls for Papers

IEEE Software wants you! Actually, we want articles by you on

- ◆ Linux
- ◆ Fixing the Software Labor Shortage
- ◆ Software Development in the Face of Rapid Technology Change
- ◆ Process Diversity
- ◆ Cross-Pollinating Disciplines
- ◆ Licensing & Certification of Software
- ◆ Life in an Internet World: Software Security
- ◆ What Needs to be Done to Define Software Engineering as a Profession
- ◆ Software Development for Embedded Consumer Systems
- ◆ Software Estimation: Recent Advances
- ◆ Critical Success Factors for Software Projects
- ◆ We are also interested in papers on general software development topics.

These topics fall within the scope we've defined so far. If you think we'd be interested in other topics, by all means tell us!

To submit an article or to become a reviewer, contact Angie Su at asu@computer.org. Articles must not exceed 5,400 words including figures and tables, which count for 200 words each.

Find out specific types of articles we're looking for at:

[http://www.computer.org/
software/genres.htm](http://www.computer.org/software/genres.htm)

IEEE SOFTWARE