Statement of Research and Teaching Interests*

Filippo Menczer
Department of Management Sciences
The University of Iowa
`http://dollar.biz.uiowa.edu/~fil`

February 2002

Intelligent Web mining agents

I am interested in the study and development of intelligent, adaptive, personalized, topic-driven crawlers to help search engines overcome their current limitations in scaling with the dynamic Web. My research attempts to close the loop from user queries back to crawling, which will allow the crawling process to be informed by the interests and preferences of the users.

Understanding the variety of information cues that coexist in a complex environment such as the Web is key to building the next generation of Internet search tools. In the current generation of search engines, lexical and linkage clues cues are used as independent, static resources to estimate the meaning of a page and its relationship with a query. In the next generation we must exploit the relationships between the various topologies of the Web. I have started to investigate the correlation between lexical, linkage, and semantic Web topologies with promising results as a major focus of my current research.

In trying to integrate linkage and lexical cues for managing the Web in a scalable fashion, I focus on distributed multi-agent search algorithms. In particular, populations of adaptive crawlers called *InfoSpiders* browse networked information environments online in search of pages relevant to the user, by traversing links in an autonomous and intelligent fashion. Each crawler adapts to the spatial and temporal regularities of its local context thanks to a combination of machine learning techniques inspired by ecological models: evolutionary adaptation with local selection, reinforcement learning, and selective query expansion by internalization of environmental signals.

My group is evaluating such algorithms inspired by multi-agent ecologies against other crawling algorithms in a framework developed to ensure use of equal resources and to benchmark algorithmic complexity. This way crawler performance is assessed using quality/cost analysis and the scalability of the algorithms can be tested by varying their available resources.

Ultimately I am interested in integrating these smart crawlers with search engines. Adaptive crawlers can be employed to replace existing blind robots for indexing, to perform add-on query-driven crawls whose results are to enhance the index, and to build indexes for topical search engines. The integration of adaptive topic-driven crawlers with search engines will allow to amortize the costs of query-driven crawls and lead to improvements in scalability, coverage, and recency, thus helping users to find relevant information as soon as it appears.

*For references and collaborators please see CV or `http://dollar.biz.uiowa.edu/~fil/papers.html`

In the longer term I am interested in the study of mobile crawlers that will migrate to remote servers and execute near the data. I envision the use of peer-to-peer protocols to enable agent collaboration, mediation and referral. This line of a research will be funded over the next five years by an NSF CAREER award.[1]

Other research interests

In the area of evolutionary computation, I have developed a class of *Evolutionary Local Selection Algorithms* (ELSA) inspired by density dependent fitness models in evolutionary ecology. This approach leads to more exploratory search behaviors than traditional genetic algorithms, and therefore is suitable to efficiently cover a diverse set of objects in a search space. I am interested in applications of such models to data mining problems, particularly for feature selection where the search occurs over the combinatorial space of feature sets to obtain solutions that are not only accurate but also parsimonious and easily interpretable. My algorithm outperformed the state of the art in both supervised and unsupervised learning tasks with multiple conflicting objectives to be optimized, where a diverse set of compromise solutions should be presented to the user. I want to extend these methods to evolve ensembles of classifiers that are optimally accurate and efficient.

Recently I have begun to study applications of machine learning algorithms to *electronic commerce,* with the goal of empowering consumers with respect to personalization, pro-activity, and privacy. I have led a team of researchers and students across three institutions in the development of a prototype personal shopping assistant. This agent observes the users while shopping and learns their preferences with respect to various features that characterize shopping items. It also proactively remembers the users' requests and autonomously monitors vendor sites for new items that might match the users' needs and preferences. Finally, The assistant protects users' privacy by means of pseudonymity, IP anonymizing, and trusted filtering. We have successfully tested the learning component through user studies with a large number of subjects and found that achieving pseudonymity through the use of personae also behooves successful classification.

In addition to studying applications of algorithms inspired by natural systems to real-world problems, I am interested in the use of computational models to study complex social and ecological systems. As a Fellow-at-Large of the Santa Fe Institute, I have been motivated by *computational economics* approaches and in particular by the use of agent based models to attack socio-economic problems. One project on which I am working attempts to build endogenous evolutionary models to study the phenomenon of referral networks in labor markets. In our agent based model, agents maximize their employment satisfaction by allocating resources to build friendship networks and to adjust search intensity. An evolutionary local selection algorithm maintains a diverse population of strategies. The emerging network topologies display mixtures of regularity and randomness, as in *small-world* networks. We have studied the trade-offs that evolved referral networks achieve between the global efficiency of the market and the individual robustness of the agents. I intend to further pursue this line of research to explore issues of inequality and segregation, with an eye toward the implications of social and economic policies.

Finally, I remain interested in artificial life models and applications to problems in evolutionary ecology. Years ago I developed *Latent Energy Environments* (LEE), an open source simulation package to model the behaviors of natural populations interacting in environments of controlled complexity. I still maintain and distribute LEE and have published several papers on its use to explore a number of questions on adaptive behavior. Recently I have become interested in applications of LEE models to study foraging behaviors in multi-species marine ecosystems. For example, I have shown that the vertical movement patterns of various oceanic predator species can be accurately predicted by a single model in which behaviors coevolve in a shared biotic environment, without consideration of physiological constraints.

[1] Grant No. IIS 01-33124

Teaching

I have taught introductory and advanced graduate and undergraduate courses spanning a wide range of core CS/IS areas (databases, networks, AI, HCI) as well as more advanced topics close to my research interests (intelligent information agents, electronic commerce, complex adaptive systems). Several of these courses were newly developed. This portfolio has left me with a solid set of teaching skills and much satisfaction. I received six departmental certificates of teaching excellence and one university instructional improvement award; however the most rewarding experience is when a former student stops by my office and says to me "Your class was hard but what I learned is now giving me an edge — Thank you."

The main points of my teaching philosophy are:

- Stimulate students, encourage their originality, challenge them to excel.
- Encourage team work and critical thinking. Teach students by giving them tools to use in problem solving.
- Prepare students for life-long learning. Help them identify proper resources to "figure things out."
- Respect students and their capabilities. Maintain high expectations. Teaching is not a popularity contest.
- Use technology in the classroom cautiously. It should facilitate learning, not side-step it.

I encourage active learning in the classroom with technology. For example, when teaching the HTTP protocol, I use telnet to send a URL request to a Web server and look at the response; suddenly a boring lecture becomes fun as students observe how the protocol works in their every-day life.

Another example of active learning with technology is the *OAMulator*,[2] a Web based resource I developed to support the teaching of instruction set architecture, assembly languages, memory, addressing, high level programming, and compilation. The tool is based on a simple, virtual CPU architecture. A compiler allows to take programs written in a special programming language and transform them into assembly. An assembler/emulator allows to interpret and execute assembly code produced by the compiler or written directly by students. The OAMulator is targeted at non-CS majors who take introductory courses in information technology or information systems. Such students are normally exposed to concepts of computer hardware and software, but it is difficult for them to make the connection between the two. The OAMulator takes the mystery out of CPU architecture by letting students gain confidence with the concepts of compilers and binary execution. The Web based deployment allows teachers to work on examples in class and students to work on problems in a convenient fashion and with rewarding interaction.

[2] `http://dollar.biz.uiowa.edu/~fil/OAM/`