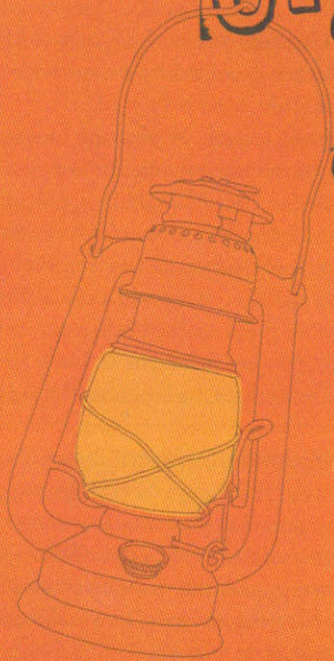University of Washington professor and famed internet data miner Oren Etzioni has been dubbed the godfather of search for his uncanny ability to find jewels amid the slag heap of cyberspace.

UNEARTHING THE INTERNET'S SECRETS

Professor and entrepreneur Oren Etzioni is already a data-mining legend. But he's just getting started.

BY JOE FOLLANSBEE | PHOTOGRAPHY BY RICK DAHMS

IN HIS OFFICE on the fifth floor of the University of Washington's computer science building, Oren Etzioni is distracted by a flash of light outside his window.

The new towers of downtown Bellevue are reflecting the orange glare of a setting winter sun, and the twinkling headlights of homebound commuters are snaking down the 520 bridge toward the Montlake neighborhood. "They're like jewels!" the 43-year-old scientist says distantly, then apologizes to his visitor for veering off task.

Etzioni knows a jewel when he sees one. The college teacher dubbed by the media an "honorary professor of webology" and the "godfather of search" has spent his working life as a mining engineer in the virtual sierras of the internet, unearthing ways to extract diamonds of data from the almost boundless slag heap of irrelevance that, for most people, is the World Wide Web. And it's not just a matter of solving an academic puzzle. Etzioni is that rare hybrid of inventor and entrepreneur who appears to honestly relish his life in academia and isn't particularly interested in the trappings of the private sector.

"He chooses things that will yield results that are more than just something to noodle on," says Mike Cafarella, one of Etzioni's current graduate students. Matt McIlwain, a partner and managing director of Madrona Venture Partners, who sits on corporate boards with Etzioni, attributes a "God-given intuition" to the scientist's business sense and timing. "He has a talent for sensing market opportunities."

A number of top venture capitalists see plenty of opportunity in Etzioni's latest major project, Farecast.com. The site attempts to forecast the price of airline tickets weeks in advance, advising frequent fliers whether to buy now or later, promising to save them an average of $40 a ticket. Investors have already ponied up nearly $21 million to support the firm's development.

Even though Etzioni picks his research projects carefully, with an eye toward their application in the world of commerce, money doesn't appear to be the driving force behind his work. Instead, he spends the bulk of his time teaching, publishing award-winning academic papers and working with young people. Student Cafarella recalled how Etzioni politely broke off a conversation with a star UW professor to take care of an advising appointment.

"I was impressed by that," Cafarella says. "His research and his teaching are his first priorities."

While Etzioni may prefer the calmer atmosphere of the college campus to the fast-paced world of the internet marketplace, the New York City native is said to be as comfortable in the corporate board room as he is on the university campus.

For his part, Etzioni says he enjoys playing on both sides of the internet street: pure research and commercial application. They are two sides of the same coin, in his view, though the environments are different.

"One world is like playing poker. It's dramatic and it can involve high stakes," he says of the two worlds he inhabits. "The other world is like bridge, much more calm and cerebral."

SEARCH PIONEER

ETZIONI WAS destined for the research world almost from his birth in New York in 1964. When Oren was a year old, his family moved to Israel. When he returned with his parents to Manhattan at age 14, he brought back a mild Israeli accent, which still gives an exotic flavor to his speech.

He has an open, square face, a broad, toothy smile and gray eyes. Bits of a similar gray fleck his flyaway hair. Both his parents were professors of sociology, and his father noted his son's interest in math. "But he was worried that one day I would go into pure mathematics," Etzioni remembers. "He wanted me to do something that was more applied."

In 1982, Etzioni enrolled at Harvard University, becoming the first Harvard student to major in computer science. Nine years later, he received his doctorate from Carnegie Mellon University in Pittsburgh, joining the University of Washington's faculty a month later as an assistant professor of computer science.

Etzioni focused his academic interest in artificial intelligence (AI), which, at the time, was limited to teaching robots how to think and behave like humans. But Etzioni saw how AI could be applied to software. To describe his ideas, he coined the term "softbot," which refers to intelligent agents that use software tools and services on a person's behalf.

The notion of softbots attracted both venture capital and attention from UW's technology transfer program, which licensed Etzioni's creation to a company dubbed Netbot. Reluctant to leave his university job, Etzioni dedicated a day a week to the business. But he grew more excited about the possibility of being a major influence in people's lives. "I was doing more than writing technical papers," he recalls.

Netbot developed one of the first comparison shopping engines on the internet, scouring a variety of shopping sites for products and then telling users which online store offered the best deal. Before Netbot, you had to visit sites on your own, much like walking around a mall and getting hungry and exhausted in the process. The portal Excite@Home purchased Netbot for $35 million in October 1997, making Etzioni, as founder, a rich man. Unfortunately, Excite@Home soon became one of the casualties of the dot-com bust. It went bankrupt almost exactly four years later.

METACRAWLER BREAKTHROUGH

ALONG THE WAY, Etzioni and his students took on the web's fundamental problem, one that still challenges engineers today: How do you find the information you require?

In 1994, search engines across the internet were experimental, and searched sites up or down depending on the whims of their undergraduate or graduate student operators. Erik Selberg, one of Etzioni's students at the time, wondered how to check automatically which engines were running. Almost as an afterthought, he and his teacher also sought a way to generate better search results.

The project became Metacrawler, which queried all of the search engines at once, finding which ones were online and returning the best results from the group. Metacrawler put the UW on the map as a leader in search technology. Selberg remembers the intoxicating experience of getting 10,000 users a day (which was a lot in those early years) along with their feedback.

"People in Australia debugged it overnight and told us what was wrong with it, sometimes with very colorful language," he recalls. "That really opened Oren's eyes to making technologies that have

One world is like playing poker. It's dramatic and it can involve high stakes, the other world is like bridge, much more calm and cerebral.

OREN ETZIONI on the difference between commercial application and pure internet research.

a real impact on people."

Metacrawler started a journey not only into the research problem of search, but also into the commercial possibilities. By raising the quality of search results, Metacrawler caught the eye of the management team at Go2Net, an early Seattle-area internet startup. Licensing the Metacrawler technology from the UW, Go2Net's co-founder and CEO Russell Horowitz hired Selberg as "director of search" and Etzioni as chief technology officer. Etzioni took a leave of absence from his UW post and worked full time at Go2Net for a year before returning to the university. The company elected Etzioni to its board of directors as well, which put him in closer touch with the venture capital community.

During his time as an entrepreneur at Netbot, and later on the executive team at Go2Net, Etzioni says he learned much about running a business. Though the lessons may seem obvious to seasoned business owners, Etzioni the academic experienced firsthand the value of having a top executive who can see the big picture.

"Technology is only one part of the equation," he says. "The most important person is the CEO, because he sets the strategy and makes sure the business model works."

Infospace acquired Go2Net in 2000, and Selberg eventually moved on to MSN Search, where he works today as a researcher. Though Etzioni is often credited with jump-starting search as a technology, Selberg says Metacrawler's impact on search in the long run was limited. The pioneering tool was soon eclipsed by Yahoo and, later, by Google.

"People just don't see Metacrawler that much anymore," Selberg says, although it's still running at metacrawler.com.

Colleagues describe Etzioni as having a "God-given intuition" that has resulted in an excellent sense of business timing.

USING THE NEW KID IN TOWN

SINCE HIS Metacrawler days, Etzioni has become something of a search star: He's been quoted by National Public Radio, *The New York Times*, *The Wall Street Journal*, *Forbes* and *Wired*. Several companies have paid Etzioni for search advice, namely Google, Microsoft, Vivismo and the search industry's most recent star, Zillow.

Etzioni admires Google, saying it avoids many of the mistakes of dot-com companies of the 1990s, staying focused on its core technology and doing whatever it takes to keep its lead. "They're running hard and denying competitors' oxygen by hiring all the engineers available," he explains.

Still, he cautions that no company is immune to the next new thing. In fact, Etzioni has argued that today's search engines are relying on a 40-year-old paradigm called "information retrieval." Etzioni wrote in 2005 of a "new kid in town called information extraction," which he described as "a scalpel that takes documents and deftly excises specific facts in a document (e.g., addresses, phone numbers, people's names and much more) and returns them."

How does this manifest itself in the real world? Currently, most of today's search engines rely on retrieval. Let's say you want to find all the pages that mention your market-dominating widget. Search on Google with your product's name and several thousand links turn up. You could fiddle with the search terms and cut the results to several hundred pages. You now have to read these pages one by one to find the most relevant information, such as whether a vendor's description of your product fits your marketing plan.

A search engine built on information extraction techniques, also known as "data mining," could save you time and energy by reading those pages for you and displaying the relevant answers on a single, personalized page. Maybe the page would show only how vendors are describing your product (along with a phone number and e-mail address so you could respond).

HARNESSING DATA

THIS DATA-MINING concept is the idea driving KnowItAll, one of Etzioni's high-profile research projects.

"It won't provide perspective and explanation," Etzioni says. "That's for humans. Its real strength is that it pulls information from a lot of

different documents." KnowItAll has extracted more than one billion facts from web pages since its inception in 2003.

Another project, Opine, works in a similar way on customer reviews of products and services. Displaying only relevant opinions, good and bad, could influence a consumer's decision to buy a product or suggest ways for a manufacturer to position a product. "When's there's lots of reviews, you see a trend," Etzioni says.

Data mining is not new. Large companies have been using the technique for decades to forecast buying patterns, for example, by sorting through mountains of proprietary sales data. What's new is the use of this technology on websites. Google, for example, applies the idea to city-based searches. But current technological limits on information storage capacity and processing power hold back these technologies in cyberspace. That's likely to change in the future, Etzioni says. In 2005, he wrote that advances in computer technology "will support disruptive information extraction capabilities that we are only now beginning to envision."

Web-based applications using data mining are already appearing, and they are bringing the ultimate power of data mining—the ability to forecast—to the average consumer. In 2003, Etzioni and his collaborators launched Hamlet, the precursor to what is now known as Farecast.

Farecast, which looks to be Etzioni's biggest commercial success yet, combs through the millions of plane tickets offered for sale in the United States and forecasts their prices days or weeks down the road. Select a departure date and arrival date, along with the airport at each end of your trip, and Farecast will tell you the level of confidence it has in its prediction as well as the likely price of a ticket.

For example, early in February, Farecast offered a 77-percent confidence rating that a roundtrip ticket from SeaTac Airport to JFK International Airport for a trip lasting four days would rise in price. The site recommended buying the ticket immediately, listing potential carriers and prices. Farecast might also predict that a price will drop, advising passengers to wait.

Other Seattle-based companies using similar ideas to forecast the future include Zillow, which mines residential property data to project a house's potential sales price, and Inrix, a new transportation product that gathers traffic data to predict the likelihood of a freeway jam-up.

The business case for Farecast goes far beyond consumers. If a prospective passenger buys earlier rather than later, an airline can better predict how a particular flight or route will generate revenue as the departure date gets closer. This glimpse into the future could suggest ways to juggle a schedule to lower costs or increase a flight's profit. And not only would Farecast wrap its service around an ad-based revenue model, it could sell the service to other websites.

The wide application in both consumer and business-to-business markets is generating huge interest. It also earned a finalist's spot as 2006 Consumer Product of the Year in the annual Washington Software Alliance industry achievement awards.

But Etzioni is taking his time with Farecast. "We are not in a rush to sell it," he says. "We want to grow it and give people a chance to use it."

Developing products with practical uses is one of Etzioni's gifts, say business partners and past colleagues. He understands the intersection between the obscure near magic of computer engineering and the daily needs of Joe Sixpack as well as Jane CEO.

"The insights he comes up with in a board meeting are far more than what you'd expect for a technologist," says McIlwain of Madrona Venture Partners, an investor in Farecast. "Farecast is a brilliant thing and I'm shocked someone hasn't bought it yet," adds Selberg, Etzioni's former student.

ETZIONI'S NEW PROJECTS

LIKE MOST RESEARCHERS, Etzioni's eyes light up when you ask him about his latest projects.

Never short of ideas, he is currently working on a long list of technologies that may click with the business world, as well. Along with KnowItAll and Opine, there's TextRunner, which indexes simple facts it discovers on the web to make possible lightning-fast searches on questions such as "What did Thomas Edison invent?" Another project, Panimages, knows the difference between an image search for "Granada" in English (an island in the Caribbean) and "granada" in Spanish (a pomegranate).

He's also had a couple of wacky ideas, such as the suggestion he made in a *New York Times* op-ed piece to fight e-mail spam by hitting the reply button for every unwanted message, in other words, spamming the spammers. "Let's send a clear message to the spammers: stop spamming or taste your own medicine," he wrote.

No one knows whether any of these new product ideas will be accepted in the academic world or widely adopted in the business arena. But his work as a teacher and his ability to perceive new market opportunities for established technologies will draw students and venture capitalists alike to his modest office at the UW for some time. When the interested come to visit, he'll point out the jewels among the merely shiny fool's gold. In the process, he may just start a new company and influence the way people use the web.

"The reason I do what I do is to have an impact," Etzioni says. "And there's more than one way to have an impact." 🆂🅱

Etzioni's Greatest Hits

THE FOLLOWING IS A LIST OF PROJECTS THAT SEARCH GURU OREN ETZIONI HAS PURSUED SINCE JOINING THE UNIVERSITY OF WASHINGTON:

Internet Softbot (1991) – Etzioni started the internet Softbot project about the time he joined the faculty of the University of Washington as an associate professor of computer science. The project applied artificial intelligence research previously focused on robots to software.

Metacrawler (1994) – Metcrawler was one of the first commercialized internet search engines. It was purchased in 1995 by local startup, Go2Net, which was then purchased by Infospace in 2000.

Netbot (1996) – From Etzioni's "softbot" research came Netbot, the first major online comparison-shopping tool, Jango. The company was purchased by Excite@Home in 1997 for $35 million.

Clearforest (1998) – Etzioni co-founded Clearforest, an Israeli company working on text-mining solutions for business.

KnowItAll (2003) – KnowItAll is an experimental search technology that extracts information from the web and presents it in a customized format.

Hamlet/Farecast (2003) – First called Hamlet, Farecast launched in 2003. The website forecasts the prices of airlines tickets.

Turing Center (2005) – Etzioni founded and directs this UW multidisciplinary research center, which is focused on web search and data mining.

Opine (2005) – Another experimental tool, Opine mines customer reviews of products and services, displaying relevant opinions, good and bad, on a customized page.

TextRunner (2006) – The experimental Textrunner indexes simple facts it discovers on the web to make possible lightning fast searches on questions such as "What did Thomas Edison invent?"

Panimages (2006) – Another of Etzioni's research projects, Panimages performs image searches that are sensitive to the nuances of language and can distinguish between common homonyms. —*J.F.*