

Webinator as an indexing and retrieval tool for creating vertical search portals on network hubs



Ecological Internet (EI) maintains up-to-date climate, forests and environment portals that serve more than 35,000 visitors a day. By implementing Thunderstone's Webinator, EI enables its website users to search the indexed content of five million URLs and quickly retrieve the desired information.

Why Ecological Internet?

Having earned his B.A. degree in Political Science at Marquette University, Glen R. Barry joined the Peace Corps and went to Papua New Guinea—where he fell in love with the rainforest while witnessing the tragedy of their very extensive destruction for the sake of making cardboard boxes and other such stuff.

According to him, “During my Peace Corps service in Papua New Guinea from about 1990 I became an early adopter of the Internet and began looking seriously at how networking technologies could be used to facilitate environmental conservation. In the early days of the Internet I was struck by the fact that communication between people anywhere in the world could be used to spread information that would lead to better resource management decisions and better conservation decisions.”

After returning from the Peace Corps he completed an M.S. degree in Conservation Biology and Sustainable Development, as well as a Ph. D. in Land Resources, both from the University of Wisconsin-Madison. His primary research revolved around the creation and maintenance of environmental web portals such as Forests.org—which became one of the first 10,000 web sites on the Internet. Dr. Barry's Ph.D. dissertation was entitled *Global Forests and the Internet: Assessing the Reach and Usefulness of the Forest Conservation Portal*.

In 1999 he decided to add search capabilities to Forests.org, while also launching a climate site and an environmental sustainability site.

Customized Search Engine for Web Sites

Dr. Barry explained, “We wanted to be able to make our own customized search engine. We preferred an off-the-shelf solution that we could easily install to crawl, index, search and retrieve content from more than 4,000 reviewed scientific-content sites of interest to our target audience of

conservation professionals. I remember searching on the Internet and finding a huge list of spidering and robot software that had about a hundred products on it. A lot of them were ‘open source,’ with little snippets of code. I was more concerned with having a fully implemented product that does what you need it to do. I wasn’t interested in doing an open source sort of thing. Where do you go for technical support in those situations? Going through the list, most of them weren’t fully implemented packages. Many of them were free, but the amount of time that a small organization would need to spend getting them operational would have offset any cost benefits. There were a few other options, but they were going to be much more expensive than Webinator.

“At that time our entire budget was like fifteen thousand dollars a year (even now it’s only about seventy-five thousand dollars a year in mostly \$25 - \$100 donations.) So, we’re a really small organization. We chose Webinator. I think our initial license with Thunderstone was eight thousand dollars, which was a major purchase for us. It was a big deal. We were trying to do something that hadn’t been done before. We had a vision that we wanted to create a specialized search engine on forests content, on climate change content and on water conservation content. The whole purchasing and installation process was straightforward. And Webinator was very, very stable. It just ran. I’m using it on a Windows platform. My operating system is Windows.

“We wanted to walk about four thousand sites we were feeding, and then we also

wanted to do off-site pages. Here’s where I think ‘customized search’ is so good. Not only are we getting the content of the reviewed four thousand sites that I as a scientist have identified, but also each of those sites has links to other sites that are included in our index. So, you have some synergy where you find unexpected things at other good sites. Webinator is a really well thought-out product that has a lot of different tools built into it. It’s a full-functioning web indexing and retrieval package. You can even include or exclude specified external links. For instance, we don’t want Green Peace’s online store and merchandise in our search engine.”

Network “Hubs” to Support Environmental Professionals

Ecological Internet (EI) does not directly focus on the general audience that’s looking for fluffy pictures of panda bears. There are other web sites that do that very well. EI’s target audience is primarily conservation professionals who need information retrieval tools and who seek useful data to factually support their own work. These people tend to be already highly motivated on the issues, and what they get from Ecological Internet are practical tools to do their work better.

Dr. Barry had been employed in the biology department at the University of Wisconsin as their ‘bioinformatics person’ until he left several years ago to run Denmark, Wisconsin-based Ecological Internet, Inc. (<http://www.ecologicalinternet.org>) on a full-time basis.

“There’s a whole branch of science, network science, that over the last decade has studied how diseases spread or how the Internet’s organized in a “hub” design comprised of nodes with disproportionately high numbers of links to them. It’s like the whole Kevin Bacon ‘six degrees of separation.’ We’re all networked, and there are hubs. The Internet is a good demonstration of a lot of these networks. What we tried to do with Ecological Internet was to make a network hub on climate change, a network hub on forests, etc. where all of the best content is linked, indexed and made available in support of intelligent activities to protect the environment. Part of this is awareness, but it’s awareness with a purpose to actually achieving something. There is reason to be hopeful. The forces of ignorance and corruption are ominous, but we have new tools – like Webinator – that we’ve never had before,” said Dr. Barry.

He continued, “I went up there to Thunderstone’s headquarters in Cleveland, Ohio to participate in a Webinator training program two years ago. I had already been using the product for six years. During this whole time I think that the Thunderstone Software team has always been very responsive. I don’t know of any other comparable product that brings full-text customized search to non-profits at a reasonable price. We wholeheartedly support Thunderstone and would recommend the Webinator search platform highly.”

Ecological Internet (EI) now maintains up-to-date climate, forests and environment portals that serve more than 35,000 visitors a day. By implementing Webinator, EI enables its website users to search the

indexed content of five million URLs and quickly retrieve the desired information.

The nonprofits’ conservation portals currently include:

EcoEarth.Info (<http://www.ecoearth.info>)
ClimateArk.org (<http://www.climateark.org>)
Rainforestportal.org
(<http://www.rainforestportal.org>)
WaterConserve.org
(<http://www.waterconserve.org>)
Forests.org (<http://www.forests.org>)
OceanConserve.org
(<http://www.oceanconserve.org>)
My.EchoEarth.Info (<http://my.ecoearth.info>)

Webinator

Thunderstone's Webinator Web Index & Retrieval System allows a website administrator to easily create and provide a high quality retrieval interface to collections of HTML documents. Webinator serves as an example of the type of applications that can be built around Thunderstone's Taxis RDBMS and Web Script.

About Thunderstone

Thunderstone Software LLC (<http://www.thunderstone.com>) pioneered simultaneous searching of both structured and unstructured data with the Taxis relational database optimized for full text search. Since 1981 Thunderstone has continued to develop its global reputation as provider of the world’s most powerful, scalable and flexible enterprise search solutions.