



Noesis: Is it a library with built-in searching or a search engine with a built-in library?

Citation

Peter Suber, Noesis: Is it a library with built-in searching or a search engine with a built-in library?, Campus Technology, March 2, 2002.

Published Version

<http://campustechnology.com/articles/2002/03/noesis-is-it-a-library-with-builtin-searching-or-a-search-engine-with-a-builtin-library.aspx>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4728880>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Noesis: Is it a library with built-in searching or a search engine with a built-in library?

- By Peter Suber
- 03/04/02

Every discipline has a rapidly growing body of literature on the Web. Many hard-working volunteers in every field have built Web directories of this literature. Some have even built discipline-specific search engines. As the scholarly content on the Web grows, life gets more and more difficult for these directory and search engine editors. Think about the problems they face. They must try to cover the field, or their own topic within the field, comprehensively. They must distinguish worthy literature from unworthy. They must discover new sites within a reasonable time and add them if they are worthy. They must fix or delete dead links. The directory editors must organize their contents to help users navigate. If they can, they should offer searching, not only of the links and their annotations, but of the full-text files to which they point. Finally, they must use methods that scale up as the relevant body of literature continues to grow. Methods that worked five years ago when the Web was small no longer work today.



N'esis (n.esis.evansville.edu) is an online library and search engine for the field of philosophy that solves these problems. Moreover, the software enabling it to solve them is transferable to any other discipline.

I'm one of the two co-editors of N'esis. My partner, Tony Beavers, deserves the credit for envisioning and implementing the features of this powerful software. In what follows, I can make immodest claims for N'esis because I'm praising Tony.

N'esis Today

N'esis has a board of topic editors, each with a different specialization within the field. The topic editors are responsible for monitoring their corners of the field for old, new, and worthy content. The N'esis software gives them a Web form for adding sites, which is much easier than writing HTML code or sending e-mail to another human editor who then writes HTML code. (N'esis also gathers new content by inviting user submissions, which are evaluated by the editors.) Topic editors may organize their topic area according to the sub-topics of their choice. Users can browse or search the entire N'esis collection or any sub-collection produced by an individual editor. By dividing the labor among the editors, an

entire discipline can be covered comprehensively and kept up-to-date. If one editor has too large a topic to cover adequately, then we only have to divide the topic and add another editor.

Gateway Selection Filters

N'esis uses several kinds of peer review to identify and recommend worthy sites. The first is at the gateway, when editors use their professional judgment to decide what deserves to be included. In addition to the criteria invoked in the gateway decisions, N'esis currently requires (with a few exceptions) that the texts be written by Ph.D.s. As we'll soon see, N'esis supports other, higher kinds of quality control that sort out the better from the worse among the texts that make it into the collection.

Adjustable-Scope Searching

Searching is the glory of N'esis. Because N'esis stores all its texts in a database, it can index them for searching much more quickly than a traditional search engine can crawl a series of Web sites. For the same reason, it can fine-tune the construction of the index. Traditional searchable collections only support all-or-nothing searching: if a file contains the search string, then a link to the file appears on the hit list, and otherwise not.

But N'esis is an adjustable-scope search engine. Users can search the whole collection, any sub-collection created by a topic editor, the collection of works by a given author, the collection of works from a given journal or set of journals, or the custom collection created by the user. N'esis also classifies its texts by genre (essays, reviews, course syllabi, and so on) and lets users filter any search by genre. Finally, editors only need to collect links to desirable texts; N'esis will automatically provide fulltext searching of those texts.

Adjustable-scope searching allows users to add another layer of peer review to their research. If you trust the peer review judgments made by the editors of journals A, B, and C, then you can set the scope of N'esis to search just those journals.

When updating its search index, N'esis automatically purges dead links. The next version of the software will put dead links in a special offline graveyard for post-mortem analysis. Most of the time, dead links mean that content has been moved, not deleted. With a little effort, the new location can be found and the link revived.

N'esis Tomorrow

The version of N'esis now online is 2.0. N'esis 3.0 will have two key features that we've already proved to work, so it's not premature to sketch here how they could enhance research.

I said that in 2.0, users could create a custom collection to help organize and search a subset of the master collection. A custom collection could contain texts relevant to a course,

a dissertation, or an essay.

In N'esis 3.0, user control over custom collections is set free to flourish. The first key feature in 3.0 is that users can create as many custom collections as they want. That might mean one for each course, each essay, each research interest. By default, all N'esis collections are public, so the collections you make for your courses can be used by your students. Each collection has a unique URL, making it easy to tell your students where to look.

At first only N'esis-approved editors will have the authority to add new items to the master collection—i.e., to make the gateway decisions about relevance and worth. Other N'esis users will only be able to make custom collections from the items in the master collection.

But eventually all users will be able to make N'esis collections from any content anywhere on the Web. We can give up the gateway control because N'esis will contain other, more effective forms of peer review and quality control.

Collection Building

The second key feature in N'esis 3.0 is that users can "adopt" collections built by other users. If you build a collection on Plato, and another scholar builds one on Aristotle, then I could start a collection on Greek philosophy by adopting both of these pre-existing collections (see Figure 1). You retain control over your Plato collection and update it whenever and however you like. When you do, my collection subsuming it is automatically updated.

Figure 1: Adopting small collections into larger ones



Figure 2: Dragging items into Noesis collections



This allows a team of scholars to divide the labor of covering a large subject like Greek philosophy. The final collection on Greek philosophy can be searched as a whole by users who don't know and don't care how it is constituted. The sub-collection on Plato is a bona fide, separately searchable N'esis collection that might in turn have adopted smaller sub-collections.

The result is that editors of large collections can make their collections comprehensive and up-to-date without monitoring the whole field themselves, and can make their collections authoritative without being experts in every sub-topic. If N'esis 2.0 is about searching, N'esis 3.0 is about modularity and cooperation in making collections worth searching.

If the editors of an online journal took control over the N'esis collection of its articles, then they could decide its internal structure—e.g., sub-collections for research articles, review articles, letters, sub-collections by year, and so on. They could also put the N'esis search box for their collection on their journal's Web page. Other N'esis users could adopt their

collection when building larger disciplinary collections.

Expert Quality Controls

N'esis users can become editors or peers for the purpose of peer review. When they build a custom collection, they are endorsing the texts they choose to include. The result can be an online journal, encyclopedia, "virtual reference shelf " for a course, or full-text bibliography for an evolving essay.

What's important to other N'esis users is not just that you've built a collection on a certain topic, but that you've done so with certain standards. If your collections are miscellaneous or heedless of quality, others will tend not to adopt them. If I trust your judgment about Descartes or Kant, then I might adopt your Descartes collection into my larger collection on epistemology. If I decide later that someone else has a better Descartes collection, I can adopt it too or I can remove yours and add the new one. Adopting one collection into another (see Figure 1) uses the same drag and drop interface as adding URLs to a collection, shown in Figure 2.

N'esis 3.0 will start with a digital library of philosophy, inherited from N'esis 2. 0 and supplemented by the index of Hippias (hippias.evansville.edu) and my Guide to Philosophy on the Internet (www.earlham.edu/~peters/philinks.htm). Tony and I will nurture N'esis libraries in a few other disciplines, such as ancient history, religion, and law, and encourage volunteers to use the software to build collections in any discipline. N'esis collections will be adoptable into other collections regardless of where they reside on the Internet.

The most scalable way to build large, long-term, up-to-date, authoritative digital libraries that cover entire disciplines is to let individual experts build individual N'esis collections on the topics of their expertise. Other users can yoke these together in any combination. From this natural N'esis activity will emerge strong collections on Greek philosophy and epistemology, for example. These can become components for larger collections that cover all of philosophy. Researchers on a multidisciplinary topic, such as racism, could build the collections they need, for example, by dragging together collections on economics, sociology, politics, and law.

Emergent Peer Review

Finally, N'esis can harness user activity to create what we call emergent quality control or emergent peer review. If a certain article has been adopted by three collections, then it has three endorsements. This is a start, but only a start, because it only counts votes without weighing them. If one of the collections adopting the article has itself been adopted by other collections, then the collection editor is not just an endorser but an endorsed endorser. If the author of works in the N'esis master collection is also the editor of one or more custom collections, then endorsements of the author's works can increase the weight of the author's endorsements.

In both of these ways, and in many others, we can start to weigh votes and find the works most endorsed by the most endorsed endorsers. There's no reason not to put some of these

parameters under user control and allow them to turn a "quality knob" in order to get fewer hits of higher quality or more hits of mixed quality. Once the structure is in place, "quality zooming" will be at every researcher's fingertips, including novice researchers who need it most.

N'esis can use this information to create special collections on its own, say, the cream of the crop on Plato, as determined by collective user activity but not by any individual user. It can also use it to sort search hits. When you search for Plato, you can sort hits by relevance, date, or by N'esis-determined quality.

N'esis collections not only carry several kinds of built-in quality controls; they also solve the problem of information overload. When users search N'esis collections on relevant topics rather than turn to general search engines, they will find only relevant hits and no false positives. Searching a Greek philosophy collection for "Plato" will return hits about the philosopher and none about the software or the town in Illinois.

Tony and I are committed to making N'esis available to ordinary academic users (collection builders and collection searchers) free of charge. If we ever charge for it, we'll charge users who want advanced features or businesses that want a commercial version to manage their proprietary information. The first purpose of revenue will be to subsidize the free N'esis services.

N'esis and the Free Online Scholarship Movement

Like most search engines, N'esis can only link to texts that are freely available on the Web. It can't see texts behind passwords accessible only to paying subscribers of a journal or database. Like most search engines, we don't see this as a limitation. However, our rationale for not seeing this as a limitation differs from that of most other search engines. We don't aspire to comprehensive coverage of the general Web. We aspire to be one of the premier tools for organizing and searching free online content, especially academic content. We aspire to be such a useful tool that content now in print or now online behind passwords has one more reason to move into the free online sector where it can be picked up, organized, and made visible by N'esis.

There is a growing movement to publish scientific and scholarly literature, especially journal articles and preprints, on the Internet and to make them available to readers free of charge. The movement is fueled in part by the exorbitant and rapidly rising costs of print journals, partly by the unprecedented opportunity for virtually cost-free worldwide dissemination afforded by the Internet, and partly by the venerable tradition in which scientists and scholars write journal articles and preprints without expectation of payment.

Progress occurs in this movement whenever a journal makes its contents freely available online or a university creates a free online archive for the research articles by its faculty. Making these collections free and online solves part of the problem. The rest of the problem is to find what you need in these separate collections. Each is separately searchable, of course. But researchers shouldn't have to run separate searches at separate archives, let alone learn which ones are likely to contain literature relevant to their research interests.

Standards and Cross-Archive Searching

One evolving strategy is a cross-archive search engine. The name explains the technology. As long as the separate archives conform to a basic standard—in this case, a metadata standard from the Open Archives Initiative (OAI)—then these special search engines can search all cooperating archives as if they were a single, grand archive. You needn't know which archives exist, where they are, or what they contain. As new ones come online, they are incorporated seamlessly, scaling up and supporting the division of labor in maintaining archives with different topical or regional specializations.

N'esis already d'es this—with N'esis collections. When a future version of N'esis can read OAI-compliant archives as if they were N'esis collections, and when OAI-compliant search engines can read N'esis collections, then N'esis will be a powerful force for accelerating the free online scholarship movement.

N'esis wants to be a simple but highly flexible tool for building, maintaining, and searching large collections of texts. But it wants to serve this function above all for free texts, to support free online archives and attract new content to them. A print journal may have many reasons for not migrating to the Web. But if one reason is that it will not be well indexed or visible to scholars in the field, N'esis will answer that worry.

If a group of experts wants to create a new, peer-reviewed journal on the Internet, then making it a N'esis collection will be by far the fastest and easiest way to do so, and the most convenient for readers and researchers. In both these ways, we want N'esis to create incentives to enlarge the body of online scholarship available to readers without charge.

About the Author

Peter Suber is a professor of philosophy at Earlham College, editor of the Free Online Scholarship Newsletter, and co-editor of Noesis.