Noesis and the Encyclopedic Internet Vision

Anthony F. Beavers
University of Evansville

Abstract. *Noesis* is an Internet search engine dedicated to mapping the profession
of philosophy online. In this paper, I recount the history of the project's develop-
ment since 1998 and discuss the role it may play in representing philosophy opti-
mally, adequately, fairly, and accessibly. Unlike many other representations of
philosophy, *Noesis* is dynamic in the sense that it constantly changes and inclu-
sive in the sense that it lets the profession speak for itself about what philosophy
is, how it is practiced, and why it is important. In this paper, I explain how *Noesis*
is dynamic and inclusive. I close by suggesting why such a communitarian repre-
sentation of the profession is both timely and necessary.

The question that frames this discussion of *Noesis* is no small one: that of how
best to "represent" philosophy. First off, the questions of what it means to repre-
sent something and indeed what a representation is are fiercely debated. Second,
we would be hard pressed to reach a consensus, even among professional phi-
losophers, about what philosophy is, much less about its role in any collective
body of knowledge and wisdom, why it is important to study, and so on. Assum-
ing anything about either representation or philosophy could land one in philoso-
phical quagmire. Yet, we do "represent philosophy" in various ways, mostly
through the publishing mechanisms of the journal and book business, through the
structure of our professional associations and conferences, through several initia-
tives to build compilations of philosophy in the form of textbooks, encyclopedias,
bibliographies, etc., and through the partitioning of the discipline that gives struc-
ture to curricula and determines what gets taught and learned. The fact that we
already represent philosophy, but often do so uncritically by adopting these
mechanisms inherited in history, is sufficient reason to address this question.
More significantly, since the way that philosophy is represented determines in
part what philosophy is, both in theory and in practice, and what it will become,
the question is not only important, but also pressing, and more so given that
emerging technologies are providing us with many choices that directly impact its
direction, for better or worse. Nonetheless, if we take it that the *goal* of represent-

ing philosophy, however we might wish to define those terms, is to render the discipline of philosophy transparent as it is, that is, to represent the profession optimally, adequately, fairly, and accessibly, then emerging information technology may well provide what we need to respond to that goal without having to settle the questions raised above and without having to resort to artificial canons cut off from the whole by people who must, of necessity, always make decisions about what constitutes philosophy and what does not.

Philosophy is what philosophers do when they practice it. While this definition initially looks empty as such, it nonetheless provides a useful parameter for determining the outer most limits for what a representation of philosophy must include. To fill it out, we must only then determine who the philosophers are and when it is that they are doing philosophy. While answering these questions requires stipulating criteria as well, perhaps it is fair to define philosophers as those who have been certified as such by the institutions of philosophy who have it as their task precisely to make this determination. As for when philosophers are practicing philosophy, this too is defined by the institutional context in which we work. I will refer to this general notion of letting the profession as a whole come together to represent philosophy as "the openness of Noesis," and I offer it here as a technological antidote to the human dangers of "closing in on philosophy." The risks of the latter are apparent whenever some philosophers take it as their task to tell other philosophers that what they are doing is not philosophy, and that, for reasons of protecting the integrity of our discipline, they should be excluded from the project of helping to guide humanity toward "wisdom" and "truth," however they might be defined.

In what follows, I will attempt to describe how Noesis embodies and facilitates this communitarian notion of representing philosophy. I will not promise that Noesis can put an end to debates about what should and should not count as the content of our discipline, though it will, no doubt, help to provide us with real, concrete evidence of our actual practices, as opposed to mere impressions and anecdotes, and add some truth to the defining debates that continue to haunt it. I will begin here with a little history of the Noesis project and then describe its current and future goals. I will end this paper with how the representation of philosophy achievable by Noesis can help to ground these and other fundamental debates about our discipline.

1. Representing Philosophy: Considerations and Constraints

Noesis can foster a communitarian approach to representing philosophy with what, in principle, is a rather basic notion made possible because of the affordances offered by the Internet and because so much philosophy is currently available online. It consists of specifying which parts of the Internet will be searched

when a search request is executed. This is accomplished technically by defining the search space for a particular search in philosophy. This strategy has two complementary components: it simultaneously determines the relevance of resources, and it provides a way to filter resources qualitatively. How this strategy works will become apparent below where I will offer a brief history of the project and the work leading up to it. Before doing so, however, it will prove useful to offer a few thoughts about search engines more generally, since it is quite easy to miss the full import of what they are doing and, hence, what they can be made to do.

A list of print resources on a particular topic is a bibliography, and it remains so when it is put online. However, when each item in that list is hyperlinked to an actual resource, something important has changed. From the perspective of the user, that list has now become more like a library than a bibliography (Suber, 2002). When such a list is compiled by hand, it reflects the agenda and the judgment of the person who compiled it. But such a list can also be compiled by software agents in answer to a variety of criteria, some of which can be determined by software designers and some by users. A search engine is precisely this kind of tool. Taking a search return set as an automated kind of the list just described, a search engine dynamically creates a momentary "library" to suit the immediate needs of a user. It does so through the interaction of three separate and distinct components. A search return set is 1) a subset of a larger dataset, here, the search space, 2) organized and processed by a program or set of programs, 3) in response to a search request submitted by a user. Each of these components can be controlled independently to produce these dynamic library-like lists.

The first can be controlled by the determination of the search space as indicated above. If the larger dataset from which a search return set is culled is limited in its topical domain, so too will the return set be limited. In the case of Google, the search space is (almost) the entire World Wide Web. In the case of Noesis, the search space is, in principle, just the sections of it dedicated to professional philosophy. The second component can be controlled, of course, by what the programs are made to do. Elements controllable by such programs include things like sort order, processing diacriticals, preparation of any accompanying information that goes with the items returned in a set, such as excerpts from the documents, file size, etc., management of Boolean operators, and so on. Most of the challenging work of search engine design falls in this second category. Google's famous PageRank algorithm is a good example. (The current release of Noesis is built on the backbone of Google, thankfully sparing us from much of the hard work, at least for the time being.) Finally, the third component is controllable by the way that users formulate search requests. Though seasoned search engine users might think that forming search requests is intuitive, log file analysis shows that many users are not quite on target with the task and need to be helped with interactive, navigational tools like topic trees and dynamic visualizations (see Hölscher &

Strube, 2000; Spink, Wolfram, Jansen & Saracevic, 2000; Jansen, Spink, & Saracevic, 2000).

The primary focus of Noesis is on the first component, that of determining the search space for professional philosophy; the task of helping users with their searching is now sufficiently addressed by the Indiana Philosophy Ontology Project (InPhO) (Buckner, et. al., 2010). Even though it is a separate project, a tight affiliation between Noesis and InPhO will add to the utility of both projects. I will say more about this affiliation later in this paper. In the meantime, the following brief history will shed additional light on these preliminary comments and provide the context for our goals with the next release of Noesis.

1.1 A Little Pre-History

In 1994, before the arrival of the search engine, Internet users could go to the WWW Virtual Library at CERN in Switzerland, where they would be greeted with a note announcing that the Internet "had no top," making it clear that they were truly in the midst of a network. Navigating the Internet was not as easy as today, but it could be done, because people who maintained pages on Plato, for instance, would hand link them to other relevant pages, allowing users to pass from page to page in something of an organized way.

Between 1994 and 1996, the exponential increase in the number of pages online started to make it too much of a chore for an individual to keep comprehensive lists of resources, and that task was increasingly left to people or communities of people who were dedicated to particular topic areas. Ross Scaife's "Diotima: Materials for the Study of Women and Gender in the Ancient World" and Pedar Foss's "Romarch: Roman Art and Archaeology" were two examples well-known to scholars of the ancient world. This explosion of resources necessitated the search engine, and by 1996 they had already carved out a place online. With their arrival came a variety of questions and concerns, often centered around the fact that almost anyone anywhere could put anything online, leaving Internet readers in the precarious position of having to determine for themselves what was worthwhile and reliable. This problem was exacerbated by the fact that with the search engine users now would most often find a webpage directly rather than passing through any identifying homepage that might have a statement of a document's source and authorship. The first search engines were no help in the matter either. (Google still is not, but Google Scholar is a rather remarkable step in the right direction.)

In 1996, a search for the term "Plato" on one of the well-known, early leaders among search engines, AltaVista, returned 44,000 "hits", including references early in the return set to a few software packages, an ale in Ireland, a consulting firm, a small town in Illinois, the Spanish word for "plate," and a story about the

"Lizard of Oz." Even when the resources did concern the Plato that we philosophers know and love, they were of mixed quality and included articles written by high school students, bizarre accounts of Plato that would probably never survive any peer-review process in the academy, and a few scholarly accounts. Clearly, the situation was such that any teacher who sent her students to AltaVista to find resources for a paper was deserving of what she was going to get, and those of us working on professional scholarship saw little in these early search engines to warrant their academic use.

At the same time, to make a step toward advancing the scholarly use of the Internet, Hiten Sonpal and I developed what would be our first "Limited Area Search Engine" (LASE) based on the principle that if we could control the shape of a search space, we could create a search engine with a kind of built-in filter to regulate relevance and quality. "Argos: Limited Area Search of the Ancient and Medieval Internet" was developed as our first implementation of this strategy. It worked by crawling a set of associate sites and everything to which they linked. These associate sites were web indices, like Diotima and Romarch mentioned above, carefully selected on the basis of their academic nature and the quality of their links. Eight other sites were selected as well, including Chuck Jones's ABZU index dedicated to ancient near Eastern art and archaeology at the Oriental Institute in Chicago and Gregory Crane's "Perseus Project" at Tufts University. The collective effect of searching these sites and the pages selected by their editors was that control over what could show up in one of Argos' search return sets was passed to the editors of the associate sites. Their actions directly determined content, and so Argos could claim a kind of peer review; indeed, it was the first search engine online that could make this claim.

The improvements in the quality and relevance of items in a search return set were instantly noticeable. On our first run of Argos, a search for "Plato" turned up around 300 "hits," but all of them referred to pages that talked about the Plato from ancient Greece, and all of them were academic resources with the proper authorial credentials. The search return sets from Argos were so clean, in fact, that we could use them as an encyclopedic index to support further research for those using another website, "Exploring Ancient World Cultures" (EAWC), that I designed to supplement the World Cultures Program at the University of Evansville. Strategically placed links from EAWC were hyperlinked directly into Argos. On our page for the Roman Empire, for instance, we linked directly from terms like "Rome," "Vergil," etc. to Argos, and when a user clicked these links, Argos would respond with a library of relevant and quality academic resources. This was certainly a step in the right direction, and so the Argos model was duplicated for the discipline of philosophy in a LASE we named Hippias and put online under the direction of Peter Suber of Earlham College. A similar version was also pro-

duced to power searches in law as part of Bernard Hibbitt's Jurist website at the University of Pittsburgh.

Argos did nothing to help with browsing and could not thereby respond to the third factor mentioned above, namely, helping users to find resources by formulating the right kind of search requests. But a close affiliation with the associate sites meant that users could browse resources topically by visiting each of the associates independently and capitalizing on the structure of the individual associate sites.

By 1998, the work of maintaining the associate sites was starting to become a clear burden. Pages were moved frequently or removed from the Internet altogether, meaning that the associates were frequently needing to be updated to cope with broken links, and the work to do this was escalating, even though the Argos-style search engines would report broken links back to their associates. This problem, compounded by the escalating growth of the Internet between 1996 and 1998 (ISC Domain Survey), made it clear that the Argos model could at best be a temporary solution. The additional need to standardize the search return item format and other metadata components was also becoming clear in the absence of any widespread agreement on tagging standards. Attempts like the Dublin Core Metadata Initiative promised such standards, even as early as 1995, but it was clear by 1998 that it would be some time before anyone could claim to have established anything standard. (It is still apparent today that no consensus has been reached on how to use the title tag that is a regular feature of html.) Thus, in 1998, we turned away from the Argos model to the first Noesis prototypes, continuing our research in the newly formed Internet Applications Laboratory (IALab) at the University of Evansville.

1.2 The Early Noesis Models

The first version of Noesis included its own database, effectively removing the associate editors from the equation. Each page that Noesis would search was individually selected and manually catalogued in a data structure that would allow for easy updating. It may seem silly today to think that this ever could have worked. But it worked then, and the reasoning seemed sound in that context. It took about ten minutes to hand catalog a link, much shorter than the time it took to process a library book. Furthermore, unlike in library cataloging for manuscripts in print, a process that would have to be repeated in each library that would hold the item, once an item was catalogued in Noesis, it was catalogued for the whole world. So, the trade offs made the process seem worthwhile. By 2001, Noesis consisted of around 100,000 hand-catalogued links on philosophy, and the project was already showing signs that the apparent success of online scholarship meant that even this solution could not continue indefinitely.

There were other issues as well. Eliminating the associate editors from the equation raised once again the problem of how to limit scope and determine quality. The absence of the associates also made it clear that somehow Noesis was going to have to assist users with their search and browse practices. We attempted solutions to both of these, with greater success on the former than the latter, but with nothing definitive. The 'failure' (if it is fair to use that term in this experimental context) of both is instructive, and hence worth recounting here.

The relevance/quality problem was handled by imposing a criterion on what we would include in the index. In order to qualify, a document had to be written by someone holding an advanced degree in philosophy or a related field, or the document had to be published in a peer-reviewed initiative. The rationale was simple, and we still rely on a version of it in today's Noesis. Agreeing up front that some graduate programs are better than others in preparing philosophy scholars and teachers (though without also conceding that the most established scholars are necessarily the best at such preparation), institutions offering advanced degrees already represent a credentialing system for the discipline, and, consequently, the attainment of such a degree says something concrete and definitive about the competence of an individual to work in the field. Though this kind of credentialing in no way employs the same kind of standards of peer-review used by journals, which suit purposes of their own, it does do something that the journals cannot do, namely, it allows for some form of meaningful dissemination that is freed from the judgments of just a few people.

For all that can be said in its favor, standard peer-review remains controversial. (See Shatz, 2004, for arguments for and against.) It usually consists of the judgment of just three people, two reviewers and an editor. While it is true that attempts to publish something in another journal, failing acceptance in a first, widens this pool to six, and so on, if a document is accepted on the first go round, it is because just two or three people thought it worthwhile. While it is true that the strategy of including documents in Noesis based on their authorial provenance rather than individual document review also falls back on the judgments of people, the cross-section of evaluators is much more diverse and wide-spread in the credentialing of people than in the credentialing of documents. Thus, while documents appearing in Noesis were not necessarily peer-reviewed, they nonetheless did bear a mark of quality that was determined by the practice of the profession itself. (I say "not necessarily peer-reviewed" here because many of the items that did turn up in search returns were reprints of things that were peer-reviewed or part of a peer-reviewed electronic initiative.)

I do not want to say that "in theory" the procedures used by the first Noesis should have worked. They did work, but so much depends on getting the language just right and communicating precisely what we were trying to do. There were, in other words, complaints, most often by those not holding the advanced degrees

we required for representation in Noesis. We were sometimes accused of academic elitism for excluding those not professionally trained from the dataset and who might nonetheless have contributions to make. Unlike in traditional peer-review, we did not verify documents individually without considering authorship. This fact made it all too easy to miss something truly important. Non-professionals need access to publishing venues to protect us from becoming too inbred as a profession, or so the criticism went. (In principle, blind review is not subject to this criticism, but one may fairly wonder just how often amateur philosophers are actually published in professional journals, and given that Noesis was not intended to be the sole mechanism of dissemination for the profession, the objection is a bit of a non-starter in the first place.)

More pressing and legitimate complaints challenged a lack of a clear and stated criterion for deciding what was a field "related to philosophy." Linguists, computer scientists, psychologists, literary critics and political scientists at one time or another suggested that their individual work was sufficiently philosophical to warrant inclusion in the dataset, and we did our best to evaluate each on a case by case basis. Some were gratified by our decisions, others frustrated. But the criticism in general was a fair one, even though how best to answer to the need for a stated criterion remains problematic. Determining what relates to philosophy depends on having a pretty clear notion of what philosophy is. Then as now, we try to steer clear of making precisely this determination so as to allow Noesis to be open, but at some point, lines must be drawn. If we were to let everything in, then Noesis would no longer be dedicated to philosophy and its utility would be limited for learning more about the discipline. The question is a matter of determining precisely where to draw these lines, and to draw them flexibly. I will say more about these matters later.

As for the issue of helping users with their search and browse practices, the first Noesis engine allowed a team of associate topic editors to log in, develop a taxonomy on a particular area of philosophy to which they had been assigned, and move individual resources from our broader dataset into the proper part of their taxonomy. Users could then search various subsections of the Noesis search space or browse resources topically. These search and browse mechanisms worked well, but the associate editors did not. Scholars are busy people, and in the absence of clear incentives, it was simply too much to ask them to log in regularly to manage their portion of the tree. But this hard lesson was instructive, and it has been a lesson learned by several people trying to develop interactive Web 2.0 applications. It is simply not the case that if designers build it, people will use it. In fact, it turns out to be quite difficult to get them to use it. A sure sign that many in the Internet community have been stung by this lesson is that it has now become an interesting question of why Facebook is so successful. (At the time of this writing, Facebook reports more than 175 million active users. See http://www.facebook.com/press/

info.php?statistics for the latest statistics. For evidence of concern over such success, readers can peruse the search return set for "Facebook Success" on Google.) One solution to the problem of user engagement is to develop mechanisms that translate the actions that users normally perform in managing their own online affairs into meaningful information. Both Google and Amazon follow this approach. So, too, will the new version of Noesis.

Between the version of Noesis just described and the version that went online in 2006, a few other possible models were discussed, and one even made it to a prototype. The 3.0 version was based on parallel distributed processing using a client/server model (popular at the time with networks like Napster and Gnutella) in which both clients and servers participated in the needed computer processing. In simple terms, the 3.0 version worked by tracking an individual's link library, a special bookmark file reserved for communication with a central server. In turn, the server would cross-correlate all of the individual link libraries, spot authorities and partition philosophy based on the collective actions of individual users. This approach was begun partly in response to the inactivity of the associate editors mentioned above. Our thinking was that if we could get enough people to interact with the server, we could use emergent methods for determining expertise and would not need to fall back on the professional credentialing discussed above. Amazon uses similar methods in the way that it recommends books and rates reviewers (Linden, Jacobi, & Benson, 1998). Initial results were promising but the practical problems of maintaining software that would be installed on any number of computers across the Internet were will beyond the resources of the small undergraduate laboratory responsible for the prototype.

Late in 2006, Google announced procedures allowing individuals to build Custom Search Engines (CSEs) as part of their Co-op platform. These CSEs work very much according to the principles that governed the LASEs described above. By restricting the scope of a search space to whatever criterion a designer might want to use, individual or "custom" search engines could easily be created that would search a subset of the Internet mapped by Google. This possibility invited us to rethink how best to approach the problem of search space construction for academic disciplines, and immediately we set to work addressing the problem geographically, according to where resources are stored, rather than by considering them topically. Partnership with Google also simplified much of the hard work involved in running crawlers, compiling searchable data structures, keeping these structures up to date, and so on. At the same time, the Indiana Philosophy Ontology Project (InPhO) was being developed to provide a workable solution to the problem of helping users browse and search for philosophy resources online. Together, these initiatives go along way in making the Noesis concept easy to implement while increasing the range of affordances that we can offer. The remain-

der of this section of this paper will be dedicated to their contributions to Noesis and our goals that result from interaction with them.

1.3 Partitioning the Profession: Fixed, Custom and Dynamic
Search Space Construction

The Google CSE works, as just noted, by allowing search engine designers to specify a search space that is a subset of the overall space covered by Google. Designers may do so by logging into a web-based interface or by uploading files in a few different formats, including XML. Designers may also specify individual files they wish to have searched, or they may use wildcard characters to specify whole directories. Of course, Google cannot search everything in the named directory, since it does not have complete access to its contents. It can only search resources it can find, namely, those hyperlinked into the web and picked up by Google's web crawlers. Additionally, designers can store their search space definitions in Google's databases, or, more important for our purposes, they may pass an address for a search space definition file to Google at the same time that an individual search request is passed. This last option means that designers can build a search engine on the fly to reflect a variety of user-determined decisions. There are other factors that can be manipulated as well, but wildcard specification of directories and the ability to pass search space definitions to Google at run-time are most important for present purposes. I will discuss each in turn in the context of explaining what they afford Noesis and how we plan to use them.

Wildcard specification of a directory (and with it any subdirectories) invites a geographical solution to the question of how best to isolate the regions of the Internet to search for limited area access to academic philosophy. In a few words, the matter is now one of determining where one is likely to find philosophy and search there. Though this might at first sound like a tall order, it really is not, since professional philosophy shows up (for the most part) in a limited variety of places; one must simply isolate the places where philosophers are likely to turn up in their professional capacities. Professional associations, conferences, departments at universities and colleges, and the journals and reference works where they publish are great candidates, along with pre-print archives, individual faculty websites, and online forums like mailing list archives and blogs, though one may fairly wonder about the utility of indexing the last two. (But, then again, what justification is there to leave them out, given that they can be included or excluded in a search according to the wishes of the user?) Assuming that we have a complete list of all associations, conferences, departments, journals, faculty web pages and so on—our goal over time—users of Noesis can search the entire reach of professional philosophy in a single pass, or they may search any one of these regions independently. Even in the minimal version now online, for instance, users may

easily select reference works and search both the *Stanford Encyclopedia of Philosophy* and the *Internet Encyclopedia of Philosophy* simultaneously, or search all and only the online philosophy journals indexed by Noesis.

Limited area searching still provides the mechanism for control over relevance and quality, as with our first versions, but wildcard searching minimizes the problem of keeping records up to date, since it hands editorial control over the contents of Noesis to the profession of philosophers itself. Once we index a journal, for instance, we no longer need to track individual items, since they will automatically be picked up by Google as they turn up online (provided that they are appropriately linked to other pages picked up by Google) and will already fall under the wildcards in our search space dataset. Since the entire search space covered by Noesis is managed by professional philosophers or their delegates (as may be the case with some professional associations), all of the individual resources in Noesis are collectively determined by the body of professional philosopher who, in the process of maintaining their own websites, implicitly add to, edit and delete from the contents available through Noesis. Without any additional work beyond maintaining their own websites, *they* are therefore the real editors of Noesis.

At these outer horizons, we have then a representation of the discipline that is inclusive of everyone and everything in philosophy insofar as philosophy itself is collectively defined by the practices of the profession already in place. Of course, at this point, this representation is in the form of a search space definition only, hardly useful for much beyond fixing the quality and relevance of search return sets. We can get more out of this representation by affiliation with the InPhO project to be discussed momentarily.

To help the following discussion, I will use the term "fixed" as opposed to "custom" and "dynamic" search space construction to designate the most basic partitioning of the profession that Noesis provides. This fixed space is partitioned according to the geography of online philosophical resources; it is determined by *where* resources are located and not by considering their particular contents. However, in the next release of Noesis, users will have the ability to log in and create their own partitions by selecting from among the individual items mapped by Noesis. In other words, they will be able to customize their own philosophy search engines built from the professional resources included in our main dataset. This service will allow users to mix and match resources according to their own criteria, selecting any combination of individual journals, departments, reference works, etc. Since they can also create an engine for just their own site by selecting only it, Noesis can offer users the ability to increase or decrease the scope of their search. For the purposes of clarity, I will refer to these mechanisms as custom, *variable scope*, search engines.

To understand the power of variable scope searching, it might be helpful to imagine a search space as consisting of four concentric circles for a moment, with

the inner circle representing the search space of a particular website and the outer circle representing all of the area covered by Google. From the center outward, the second circle here could be a custom search space established by the editor of the site represented in the inner circle, and the third circle out could be all of the space mapped by Noesis. (See Figure 1 below.) To use an example, if the site in question were the *Stanford Encyclopedia of Philosophy*, the second region would include the SEP plus all of the individually selected journals and archives chosen by its editor. Users starting with a search box on the SEP site would then have the option of deciding the scope of their search at runtime and could easily search the SEP, or a subsection of the Internet defined by the editorial choices of SEP administrators, or all of Noesis, or all of Google. Thus, custom search space construction affords further refinement of the quality and relevance of a particular search space. At the same time, it provides Noesis with evidence to help determine more precisely the quality and relevance of the individual resources it maps.
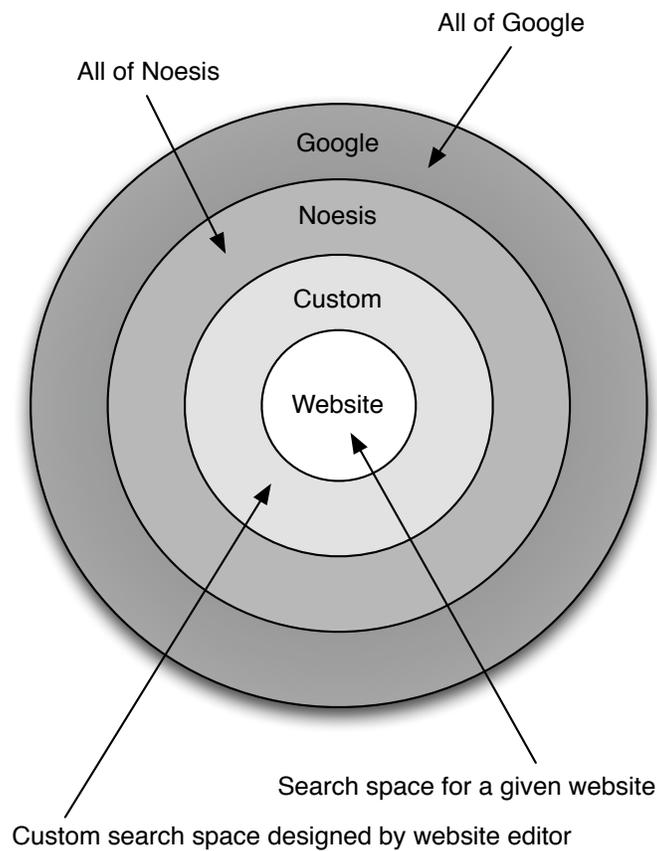
Figure 1. Custom, variable-scope search space definition.

Dynamic search space construction refers to search spaces determined automatically by semantically sensitive search mechanisms. The general idea here was hinted at above, where I discussed the role that *Argos* played with *Exploring Ancient World Cultures* (EAWC), the online undergraduate textbook supplement mentioned earlier. Previously, I noted that the finely tuned search return sets from Argos were sufficient to provide encyclopedic access to quality online resources for the users of EAWC. A hyperlink on the word "Rome," for instance, would pass a search request directly to Argos, which would respond, in turn, with a list of academic resources answering precisely to that request. The idea behind dynamic search space construction with Noesis is similar, though more refined than in our earlier attempts. So that the reader is fairly apprised of the state of things, our plans here are theoretical (that is, untried) at this point, and so, even though I am confident that we can make them work, I might rightfully be charged with talking about 'vaporware' in describing them.

Theoretically, then, using the referring URL information that is passed with a particular hyperlink, Noesis will be able to retrieve the document from which a linked search request originated, analyze it, if it has not already done so, and then use that information to generate a search space definition dynamically suited to the context in which the search request is embedded. Thus, for instance, users reading a paper on Fodor's representational theory of mind could encounter a hyperlinked search request on the term "representation." When clicked, Noesis would respond with a search for the term "representation" across a space dynamically suited to resources in the philosophy of mind and, perhaps more narrowly, responding to the representational theory of mind, in particular. These dynamic search spaces could be combined with the idea of variable scope searching described above to allow users to configure a setting to determine the "range of context" for a given search, that is, in rough terms, a percentage of the degree of overlap needed for the dynamic mechanisms to include an item in a search return set. To make things easier, rather than having individuals mark these hyperlinks by hand, we would prefer to provide a mechanism that creates an overlay for a particular document that automatically links the proper terminology to Noesis. These ideas on dynamic search space construction are speculative, as I said, and to quote Herbert Simon, "In the computer field, the moment of truth is a running program; all else is prophecy" (1965, xv). So, we will have to see. But the matter is on the experimental agenda for Noesis.

1.4 Inphorming Noesis and Expanding the Scope of Its Inphormation

Earlier I mentioned the need to include in a search engine some mechanism to help users with their search and browse habits. I also noted that the existing ver-

sion of Noesis represents philosophy in the form of a search space definition, which has an admittedly limited utility. The Indiana Philosophy Ontology Project (InPhO) discussed elsewhere in this volume provides a partial solution to the need for a taxonomic index that does justice to the organic nature of the profession while adding additional utility and power to Noesis.

InPhO is an organic and emergent taxonomy of the discipline of philosophy that uses artificial intelligence algorithms to "read" the *Stanford Encyclopedia of Philosophy* (SEP) in collaboration with feedback from human users. It is designed to identify philosophically-rich vocabulary insofar as it is distributed in the SEP, isolate patterns of generality and specificity, and present this to the user in the form of a topic tree. (Other visualizations are being considered as well, but they exceed the scope of the current discussion.) InPhO's tree allows users to explore the topical terrain of philosophy by expanding and collapsing parts of it as needed. An additional affordance is the ability to pass computer-configured search requests to various search services, including the SEP, Google Scholar and Noesis. A single click then supplies the user with a list of resources answering to that request.

In addition to helping with the search and browse problem discussed earlier, the affiliation between InPhO and Noesis overcomes (in part) what is sometimes called the discovery problem, the problem of making users aware of resources that they do not know exist, but that they may need to consider. If users do not know they exist, they cannot search for them in particular; hence the need for computer-aided search and browse strategies to bring them to light. Such strategies can make connections that are initially rather distant explicit and therefore available for philosophical reflection. A favored example from the affiliation is that a search for "mental representation" turns up connections with "divine illumination" in medieval philosophy that, in turn, shed light on the heritage of the concept of representation and its reception into the philosophy of mind via modern philosophy. Further research on how to use computers to find and explore these connections will, no doubt, add additional utility, and we will perhaps one day learn novel information about the profession from a variety of automated procedures.

The affiliation between Noesis and InPhO also promises to help each project overcome limitations of scope. I noted earlier in conjunction with the first Noesis, for example, a set of objections concerning what qualified as a discipline related to philosophy. If Noesis is going to be dedicated to academic philosophy online, somehow it must be sensitive to resources of a philosophical nature that do not come from within the discipline itself. Scholars from outside the formal institutional context of philosophy are often by no means philosophical amateurs and have earned a proper place in the canons of our discipline. Physicists, for instance, can and do engage in philosophy of physics, as economists sometimes engage in

philosophy of economics, medical practitioners in matters of ethics, literary critics and art historians in aesthetics, and cognitive scientists in epistemology and the philosophy of mind. It would be a mistake to define the scope of philosophy in such a way that, by definition (even by search space definition!), these others are excluded. Yet, without going back to inspecting each resource by hand for inclusion in Noesis, which in turn requires a perspective on the discipline that no human being (or small set of human beings) has, the problem is how to identify them, especially given that we do not want all of medicine, art history and cognitive science in the dataset, just those parts of them that overlap in important ways with philosophical research.

InPhO is not just a mechanism for generating a browsable topic tree, it is also a partitioned map of the semantic space of philosophy, and, as it continues to develop, it will be more so. As such, it holds part of the key to identifying philosophical documents automatically. The geographical strategy outlined above for building Noesis' fixed search spaces picks up resources that are located where professional philosophers tend to appear. If some among them work in philosophy of physics, to cite an example, then resources written by professional philosophers of physics will be picked up. However, resources in the philosophy of physics that are written by physicists may well be missed, unless they are published in an online philosophy journal or are put online by an official representative of philosophy like a professional association or academic department. Over time, InPhO will help us identify a "semantic signature" for philosophical resources that will prove useful for comparison with documents not in our search space to determine which should be added. Experiments will be necessary to determine the best way to proceed, but we do have on hand some things to try, and, in any case, if such mechanisms cannot do the trick themselves, they will no doubt go along way in pre-filtering resources for human inspection. But pursuing this course will be a last result only if our artificial intelligence strategies cannot be made to work.

Less speculatively, Noesis can be used to help overcome the limitations of scope in the InPhO project. InPhO is seeded by the SEP, and its topic tree and semantic maps are only as good as the SEP. Thankfully, the SEP is excellent and currently provides the best representation (in terms of scope and quality) of philosophy online and, possibly, in print as well. At the time of this writing, it includes 1105 articles in over 40 subject areas carefully managed and reviewed under the auspices of 115 area editors. Though it grew primarily out of the analytic tradition, it has quickly expanded beyond to include a fair representation of other areas, including ethics, feminism and continental philosophy. Its holdings in the history of philosophy are also extensive, and the credentials of its authors are hard to beat. Still, it is limited and *probably* falls short of providing the complete range of vocabulary used by the profession to the InPhO project. I say *probably* because the only way to know for sure is to compare it against the broader corpus of pro-

fessional philosophy online to see. Given that Noesis plans to map the entire space of professional philosophy online, it will prove useful as a tool to help test the adequacy of the SEP as a semantic representation of philosophy. If it is inadequate, we will be able to know this and, more importantly, know precisely where it is lacking. In conjunction with the InPhO project, Noesis then will be able to help the editors of the SEP know where additional coverage is needed.

2. Making Philosophy Transparent: Toward Real Evidence about the State of the Profession

In addition to helping users find quality and relevant philosophical resources and supplying the InPhO project with a broader pool for its semantic analysis, there are other affordances for the profession (and beyond) that make the Noesis project important. Some of these affordances are common to other initiatives and are already well documented in the literature of the open access movement. They concern matters like helping to free the direction of scholarship from the exigencies of the marketplace and bringing quality resources to developing countries that cannot afford to build print libraries of their own.

The need to free scholarship from the marketplace is critical for the future of philosophy, especially for the areas that are not part of the "mainstream" movements within it. A cursory review of market trends might easily lead one to believe that philosophy today concerns the philosophy of mind and ethics, and slightly less so the philosophy that relates to science (as with the many resources concerning evolution) and public policy. But this is an illusion created by the fact that resources on topics that will sell are more likely to get published by an industry that must be concerned with income to stay in business. Consequently, topics that are likely to attract a wider readership come to represent philosophy precisely because they are popular among readers. This might be a fair way to measure the spread of interest in a topic across the profession more broadly, but it does not do justice to what is really going on in the actual practice of the discipline, and allowing the institutions of commerce to determine what of the world's wisdom we should protect might be objectionable on other grounds as well. For instance, there is the perennial threat of a "feedback loop" that perpetuates the popularity of resources in a given area. Books on topics that sell generate interest which in turn encourages more books on the same topics and create topical "ruts" that impact the scope and utility of philosophy in potentially harmful ways. Projects like Noesis that provide focused access to open-access resources contribute to providing a more adequate representation of philosophy because they serve also to *open access* to a broader range of philosophy.

Concerning the matter of spreading resources globally, even though Noesis indexes resources primarily in English at this time, the prototype version now on-

line regularly serves pages to over fifty different countries weekly, including Argentina, Australia, Ecuador, India, Iran, Japan, Kuwait, Mexico, Rwanda, Thailand, the Russian Federation and Venezuela, to pick a representative dozen, and it is linked to pages in several different language domains, including Arabic, Danish, Dutch, French, Italian, Norwegian, Slovak, Spanish, Swiss, and, of course, English. As Noesis continues to develop, we expect a truly global presence, if we do not have that already. As a world full of people find their way to Noesis, they are simultaneously finding their way to other open access initiatives in philosophy.

So, Noesis will provide a broader representation of philosophy across a broader populace than many (any?) other avenues of professional dissemination. However, though service to the open access movement and crossing political and cultural boundaries to serve the growing academic needs of developing countries are important matters that should not be minimized, I will use the space remaining in this paper to address an issue that is not commonly a part of the literature—the literature concerning electronic resources in philosophy, at least—and that ties more centrally to the exigencies of representing philosophy itself.

2.1 Profession, Know Thyself!

The first release of Noesis was dedicated to scholarly articles and primary historical texts, much like the impressive PhilPapers project recently put online by David Bourget and David Chalmers. (See http://philpapers.org/.) In the latest release, Noesis expands beyond the scholarly output of philosophers to include information tied to the social and administrative aspects of the profession and the way that the profession taxonomizes itself in the form of professional associations, conference programs, tables of contents, bibliographies, etc. In short, we are now interested in mapping the profession of philosophy as it is practiced and not solely in terms of its scholarly output. In this way, Noesis represents parts of the profession that are not typically included in conventional representations of the discipline, but that are nonetheless a part of the way that philosophy gets practiced and determinative of its direction in history. (See Morrow & Sula, elsewhere in this volume, for hints about how social elements of the profession might be relevant here.)

The profession of philosophy is so varied that it is possible for someone to be at the center of one of its parts and virtually unknown to people working elsewhere within it, as demonstrated by a recent exchange on the Philos-L mailing list concerning the status of the intellectual accomplishments of Slavoj Zizek. (Some people were arguing that he was a non-contender and a "no name" in philosophy, while others were arguing that he was something of a shining star with maximal impact on certain sectors within the profession. It turns out, by the way, that he is

an important figure, but that his research is not disseminated in what, according to some, are the "more mainstream" sectors of philosophy.) Professional blind spots are easily created by the fact that we often take our exposure to the profession to be representative of the whole, even while we realize that we only frequent the venues or read the texts that suit our research and teaching interests. Anecdotes abound about what is going on elsewhere in the profession, and they often substitute as evidence of the true state of affairs, leading to value claims about what is important and what is not. (I am not making indictments here. Human reason is limited by perspective and this limitation, of course, affects our judgments, even when we are careful to employ reason "objectively".) Furthermore, the fact that other people share my judgment often looks like corroborative evidence in favor of its objectivity, when it may mean only that my view is popular (and among *my* interlocutors, at that!), especially if those sharing my views were indoctrinated into the profession by the same teachers, institutions, or schools of thought. Even though most of us in the profession are trained in informal fallacies and other strategies of poor reasoning, we often seem to succumb to them all the same, especially when it comes to "water cooler" talk about the profession that nonetheless has an impact on people's careers and the direction of the discipline.

In the past, appeal to anecdote was perhaps excusable, given that other kinds of evidence were not readily available. In the absence of access to facts on the profession, impressions were allowed to suffice, but this situation is changing quickly. Recent changes in communications technology over the past decade and a half make it easy for us to study neighboring and distant institutions, look up the scholarly credentials of our colleagues, see who has published what and when, and also how such publications are being received, etc. Since 1994, a new data pool on the profession has emerged, and while it is already present in piecemeal form to anyone wishing to have a look, it has yet to be collected, collated and cataloged. Here is one place, then, where a tool like Noesis can be instrumental in making us self-reflective about the profession we practice and our place within it.

Without going into the details of data mining (and admitting that we still have a way to go with what is really still an emerging technology), it is possible to identify a variety of empirical data on the profession that are already available in the Internet terrain that will be marked out by Noesis' geographical search space strategy. Here is a partial list, leaving out for the moment items that refer to philosophical content and acknowledging that a significant amount of information on content is available as well:

- which departments have the widest diversity of faculty in terms of coverage

- which departments are focused on particular topics and the relative strength of their faculty, where strength is defined according to a variety of user-configurable criteria
- which departments are the most unstable in terms of faculty departures and arrivals
- which types of courses are most frequently taught where
- which universities are most successful at placing their graduates in teaching positions
- which undergraduate institutions fare best at placing their graduates into top graduate programs and, ultimately, into the profession
- which scholars are most widely published and with whom they are most often published

The answers to such questions (and others like them) that we will be able to cull from Noesis in time will not be like the impressions and anecdotes that have powered our discipline in the past. They will be data-driven answers that reflect in real-time genuine quantitative metrics about a constantly changing profession.

In this connection, I cannot resist the need to mention here *The Philosophical Gourmet*, compiled by Brian Leiter and, hence, also known as the Leiter Report, a website that purports to provide metrics on "faculty quality and reputation" for Ph.D. granting philosophy departments in the English speaking world. (See http://www.philosophicalgourmet.com/.) The fact that the report has become the standard metric of overall departmental ranking, despite its claims to the contrary, in just a short time attests to the fact that metrics are needed. At the same time, it demonstrates how quickly the Internet can act in getting information out that can restructure our profession, since it is no longer uncommon to hear professionals speak of "Leiter rankings" in the hiring of faculty or to find graduate students acknowledge that the Leiter report weighed heavily in their selection of a graduate program.

Reports such as these may present themselves as descriptive, but their use makes them prescriptive. If philosophy has anything to offer the success of our species and the well-being of its individuals, it is imperative that we represent it adequately, optimally, fairly, and accessibly. Given that how philosophy is represented determines in no small measure what philosophy is, both in theory and in practice, and what it will become, attempting an accurate and adequate representation might be something of a moral imperative as well, and this applies no less to metrics about faculty quality and reputation than it does to other aspects of the profession. Such reports therefore should invite caution and careful scrutiny. In the case of *The Philosophical Gourmet*, for instance, 95% of the advisory board for the 2009 Leiter report work at Leiter ranked schools, and 57% of them work at schools that are ranked in the top ten for their respective countries. More impor-

tantly, however, is that *80% of the respondents surveyed to determine the rankings have connections to schools that ranked in the top ten* for their respective countries, while a full 97% of them have connections to schools somewhere in the rankings. 81% of the respondents work at Leiter ranked schools, and 93% of them went to ranked schools. Furthermore, the institutions to which respondents are connected are distributed across the top of the rankings. I just noted that 80% have connections to top ten schools. But 15% of them have connections to second-ranked (overall) Oxford, 10% to fourth-ranked Princeton, 8% to third-ranked Rutgers and another 8% to seventh-ranked Harvard. Even though respondents are not allowed to rank their home institutions or those from which they received their degree, it is still quite clear that people from a handful of institutions are handing out high marks to other institutions in that same handful to the neglect of several institutions that are not presented for assessment in the first place. Leiter explains:

> The survey presented 99 faculty lists [from institutions up for assessment], from the United States, Canada, United Kingdom, and Australia and New Zealand. Note that there are some 110 PhD-granting programs in the U.S. alone, but it would be unduly burdensome for evaluators to ask them to evaluate all these programs each year. The top programs in each region were selected for evaluation, plus a few additional programs are included each year to 'test the waters'. http://www.philosophicalgourmet.com/ reportdesc.asp

Top programs are pre-selected to determine which of the existing programs are in the top. Certainly, something smells a little fishy, and even an undergraduate in a critical thinking class would be tempted to see several fallacies operating here. Notwithstanding the possibility that Leiter may have hit on a heuristic of some sort that does in fact track faculty quality and reputation across the profession, the fact remains that his anecdotal survey approach can only leave us guessing about this possibility. We need some way either to verify or debunk the report, and here is one place where Noesis may be able to help. Furthermore, given the wealth of information about the profession that is available online and what is at stake, soon there will be simply no need to fall back on such simple anecdotal measures in any case. I suspect that we will find they are not worth much, or perhaps more so, that genuine evidence may provide us with better intuitions about the profession and reshape our anecdotes to fit better with the actual state of affairs.

Noesis' contribution to this issue of acquiring knowledge about the profession here is akin to the one mentioned above about testing the adequacy of the *Stanford Encyclopedia of Philosophy* as a representation of the semantic content of the profession. Though various data mining techniques will, no doubt, prove useful

over time, the planned custom and dynamic search space strategies described above go part of the way toward responding to the need for adequate information. Using the procedures outlined above, we will be able to partition the profession in any way we see fit, run a set of searches across the various partitions and compare the results. Thus, comparative analysis is one of the affordances that will emerge naturally from Noesis. Though human users will be able to build their own partitions for comparative purposes, more interesting exploration will, no doubt, follow from automated procedures that create, compare and then adjust partitions until evidence-based sectoring gives us a computer-mediated picture of what the profession looks like. There is no reason to think that such a picture be static or that there be a single picture either. Dynamic representations and representations adjusted to serve a variety of user-defined purposes are definite possibilities.

3. Conclusion

The really interesting work for the Noesis project is in the next phase of development. For now, we are interested in circumscribing the profession of philosophy online, building an evidence pool as it were, a search space, for further analysis down the line. Even without the vision of the profession sketched immediately above, Noesis promises to offer an unprecedented access to philosophy as it is practiced in real-time. Combined with other tools, such as the Indiana Philosophy Ontology Project, we will provide a research tool never before possible in history.

Acknowledgements

References

Comment: The papers below that discuss earlier versions of Noesis present some ideas that will never be developed. Over the past decade unanticipated changes in technology have somewhat resituated the project's direction. The current paper represents our most recent vision along with a summary of highlights from the earlier versions.

Beavers, A. (1999). "Noesis: Philosophical Research On-Line: An Experiment in Progress," *Newsletter on Philosophy and Computers*, American Philosophical Association, Volume 98.2.

Beavers, A. (1998). "Evaluating Search Engine Models for Scholarly Purposes: A Report from the Internet Applications Laboratory," *D-Lib Magazine: The Magazine of Digital Library Research, The Corporation for National Research Initiatives*, December.

Buckner, C., Allen, C., & Niepert, M. (2010). "From Encyclopedia to Ontology: Toward Dynamic Representation of the Discipline of Philosophy," *Synthese*, current volume.

Hölscher, C. & Strube, G. (2000). "Web search behavior of Internet experts and newbies," *Computer Networks, 33*, 337-346.

"The ISC Domain Survey," *Internet Systems Consortium*. https://www.isc.org/solutions/survey. Retrieved on March 31st, 2009.

Jansen, B., Spink, A., & Saracevic, T. "Real life, real users, and real needs: A study and analysis of user queries on the web," *Information Prorcessing & Management 36.2,* 207-227.

Leiter, Brian. (2009). "The 2009 Philosophical Gourmet Report: Brian Leiter's Ranking of Graduate Programs in Philosophy in the English-Speaking World." http://www.philosophical gourmet.com/. Retrieved on March 31st, 2009.

Linden, G., Jacobi, J., & Benson, E. (1998). "Collaborative recommendations using item-to-item similarity mappings." Patent number 626649. Google Patent Search. Retrieved on March 31st, 2009.

Morrow, D. & Sula, C. (2010). "Naturalized Metaphilosophy." *Synthese*, current volume.

Niepert, M., Buckner, C., & Allen, C. (2007). "A Dynamic Ontology for a Dynamic Reference Work," JCDL Proceedings, forthcoming.

Shatz, D. (2004). *Peer review: A criticial Inquiry*. Rowman & Littefield.

Simon, H. (1965). *The Shape of Automation: For Men and Management*. New York: Harper & Row.

Spink, A., Wolfram, D., Jansen, B., & Saracevic, T. "Searching the web: The public and their queries," *Journal of the American Society for Information Science and Technology, 52.3*, 226-234.

Suber, P. (2002). "Noesis: Is It a Library with Built-In Searching or a Search Engine with a Built-in Library?" *Campus Technology*.

Uzgalis, B. (2000). "Searching Phenomenology and Cyberspace: An Interview with Anthony Beavers," *Newsletter on Philosophy and Computers*, American Philosophical Association, Volume 00.1.