The HyperLearning Project: Towards a Distributed and Semantically Structured e-research and e-learning Platform

Michele Barbera

Projekt HyperNietzsche, Ludwig-Maximilians-Universität München, Germany and Net7 – Internet Open Solutions–, Italy

Abstract

An increasing number of scholars think a new instrument to publish academic work is needed. In fact, the scientific journals are in a monopoly market that makes access to scientific information very expensive. The richest university libraries use 80–90% of their budgets for the purchase of scientific journals and nevertheless are able to afford only a small part of academic literature. For scholarly publications in the Humanities there is not a monopoly market—there is no market at all. Public libraries have less and less money for monographs. The Humanities are in constant crisis as far as the publication of scholarly editions is concerned. This is especially true of genetic and facsimile editions, but it is also the case for all projects where the requirements of scholarly work are in conflict with the realities of the book market. Furthermore, the access to libraries and archives holding the primary sources for scholarly work is often difficult, expensive and unsatisfactory. The HyperLearning project is an extension of the HyperNietzsche project, which tries to solve the difficulties outlined above. This short research report describes the HyperLearning project, focusing on its technological activities. In the first part we will delineate the path from HyperNietzsche to HyperLearning. The following parts are an overview of three major technical research areas of HyperLearning.

Correspondence:
Michele Barbera, Net7 – Internet Open Solutions–, via Marche 8/A, 56100 Pisa, Italy. **E-mail:** barbera@netseven.it

1 From HyperNietzsche to HyperLearning

The HyperLearning[1] project is an extension of HyperNietzsche (D'Iorio *et al.*, 2000)[2], which has been online since 2001 and received the Sofia Kovalevskaja Award of the Alexander von Humboldt Stiftung in 2001. HyperNietzsche was also financed by the French Ministry of Research, the German Ministry of Education and Research and the Deutsche Forschungsgemeinschaft.

The HyperNietzsche project team has a wide-ranging composition, with more than 20 people involved[3] (philosophers, philologists, computer scientists, jurists). During 2003, HyperLearning was submitted for approval to an Integrated Project within the 6th Framework Programme of the European Union. Although the project had been well received over the scientific evaluation led by independent experts (with 24 points ex-aequo, resulting in the second ranked project with only half a point behind the best ranked project) it was,

regrettably, not financed by the Programme Committee. Still, the project will be submitted again in the next FP6 call.

1.1 Objectives

The overall aim of the HyperLearning project is to create a framework for an advanced Research and Learning infrastructure for the Humanities. As a Research infrastructure, HyperLearning enables a decentralized community of experts, to work in a cooperative and cumulative way and, at the same time, to publish the outcome of their work on the Internet. As an e-Learning system, HyperLearning combines Research and Education, and envisages not only knowledge transfer, but also the development and enhancement of both critical skills and autonomous production of scientific contributions among graduate students and young researchers. In order to achieve this aim, the HyperLearning project pursues the following three objectives.

(1) To conclude the realization of HyperNietzsche as a framework of an IT Research infrastructure. The development of the system will go along with work on editions, transcriptions of Nietzsche's manuscripts and correspondence, as well as publication of critical essays and other scholarly contributions. Working simultaneously on these two tasks on the one hand will produce an IT infrastructure that responds effectively to the needs of actual users, and on the other hand will demonstrate how this new tool can be used effectively.

(2) To develop and implement a critical e-Learning system for higher education, which allows going beyond the e-Learning approach conceived as trained education rather than as critical learning; the former currently adopted by the present LMS and encouraged by standards such as SCORM (Sharable Content Object Reference Model) (Lambe, 2002; Welsh, 2003). This new system will be realized in the context of a new pedagogical paradigm, based on the idea of dynamic ontology, and will contribute to the improvement of current e-Learning interoperability standards.

(3) Preparatory work for the adaptation of the HyperLearning framework to other authors.

This includes writing documentation of the applied procedures, the scheduling of conferences presenting the project together with brainstorming workshops with interested scientific communities, as well as training for researchers in their first step applying this framework to other authors. A design for the IT structure of a HyperLearning Network, which would enable the interoperability of the various HyperLearning Platforms, will complete the project.

1.2 Motivations

Needless to say the software architecture of an electronic research platform has an impact on research methodologies. Think of a scholar of the pre-electronic era in a library. What does s/he do? S/he searches for a book or a review in the catalogue or in a bibliography, takes the book from the bookshelf and refers to a dictionary or a thesaurus. But what is their navigation system? When people have transposed the library, the catalogue, the thesaurus into an electronic environment, they have thought to use systems of indexing and information retrieval to link all the documents. But that is insufficient. Normally a scholar practices 20% information retrieval and 80% navigation in context, i.e. reading. One response is to encode the electronic document with a conceptual grid of metadata. But this is not enough: a scholar who reads follows twenty conceptual grids at the same time and tries simultaneously to develop his own. Scholars navigate in an ensemble of documents thanks to two criteria: the structure of the documents (for example, the order of the pages of a book or of a manuscript) and the relationships between the documents other scholars have indicated in footnotes, quotes, bibliographical references, and so on. This seems obvious, but nobody has transposed this principle into an integrated electronic environment. Nobody has used it to interlink a scholarly website. The application of computer science in the Humanities until now has been confined to computational linguistic, bibliographic or thematic databases, and digitization projects. It is time to invent models that respond better to the needs of researchers and exploit fully the possibilities of electronic media and networking.

As a practical starting point, the HyperLearning project will adopt the Research paradigm, the software infrastructure, and the legal framework already introduced by HyperNietzsche. HyperNietzsche is not merely a library of well-indexed and searchable electronic texts and studies made available online (although it will be OAI-PMH[4] compliant), nor an electronic edition available to the public as a finished product. Rather, it is a kind of electronic research matrix that creates a virtual workspace with public access to original sources for Nietzsche research, including works, manuscripts, letters, and biographical documents. Beyond these original sources, HyperNietzsche contains an ever-increasing collection of scholarly essays that have been vetted for publication by an editorial board of leading Nietzsche experts.

From a methodological point of view, this venture transposes the concepts of Open Source from computer science to the Humanities. Computer science, for its part, had merely reformulated, in its own terms, the fundamental characteristics of scientific methodology and praxis that had been in place since the 17th century. Science is, by definition, an open source social organization to the extent that it is an open, public debate over accessible objects, carried out according to ascertainable procedures. Indeed, scientific progress is hopeless in conditions of secrecy, exclusivity, or restricted access to information (Guédon, 2001). In the Humanities, working in Open Source means on the one hand being able to access the digital version of objects of study, i.e. books, manuscripts, archaeological objects, images, sounds, film sequences, etc. (Public Archives); and on the other hand it means providing free access to the results of research work (Open Publishing). And the Internet is the most suitable medium for guaranteeing public access to cultural heritage and research work.

2 HyperLearning Technologies

2.1 Dynamic contextualization and beyond

Dynamic contextualization (Barbera and Giomi, 2004) is one key innovation developed by HyperNietzsche. While the user navigates the site, moving from one page to another, this feature makes available precisely those scholarly contributions that are relevant to the one presently being viewed. For example, if the user selects a manuscript page, the system immediately makes accessible, without the need for additional complicated searches, all the facsimiles, transcriptions, and translations available for the page, as well as all the relevant text-genetic paths and critical essays that refer to the page. This material is presented in the form of an easily navigable list of hyperlinks. Likewise, if the user selects a critical essay, the contextualization mask will present a list of hyperlinks to all the Nietzsche texts and manuscripts cited by the author of that essay and all the contributions that are cited within it, or that cite the essay being viewed.

Dynamic contextualization is a first step towards a semantically structured and flexible electronic research platform. To extend the model further, the technology development and software research component of the project will work on the following.

(1) A functional unified approach to server and client side programming, based on improvements of the OCAML[5] and HASKELL[6] programming languages.
(2) Object persistent data repositories with distributed updates based on XML and XDUCE.[7]
(3) Statically typed generation and transformation of HTML and XML data.
(4) A contextualized and highly available local data repository model with an advanced caching system.
(5) A semi-decentralized peer-to-super peer (Yang and Molina, 2002) network of web services that collaborates in a virtual, distributed, and semantically structured network based on a global as view data model (Molina et al., 2004).
(6) An Open Source software that allows anyone to easily install, configure and maintain a node within the HyperLearning Network. The software will be based on *Linux* live technologies, such as *Knoppix*[8] and *Demolinux*,[9] and will adopt only Open Source technologies for the web infrastructure, such as *Apache httpd*, *php*, and *PostgreSQL*.

2.2 The HyperNetwork

The HL network is based on a p2p network similar to the one used by music sharing programs such as *Napster* and *Gnutella*. In a p2p network each node can be a client and a server at the same time; for this reason, network nodes are often called servent as a mixture of server and client.

The network services are the backbone of the whole project, which aims to build a distributed and semantically structured e-learning and e-research platform. To construct a semantically structured network we do not need to force the virtual communities aggregated around different nodes and different thematic fields of study to adopt a common ontology. Rather than super-imposing a common data scheme and access logic to all the local communities involved in HyperLearning, we will adopt a decentralized ontology based on a data approach that will preserve the existing practices, data structures, and tools of the different local sites, while ensuring a semantic interoperability of their data (Lenzerini, 2002). This will be achieved by having local communities expressing the semantics of their data by means of domain-specific ontology, and providing a suitable mapping with respect to a global ontology main-tained by a federative infrastructure (Fagin *et al.*, 2002). The ultimate goal of this part of the project is therefore obtaining a scalable, humanities-oriented, semantic data grid.

The HyperNietzsche team, in other words, has developed methods and tools that are appropriate for studying the work of Nietzsche. Roughly speaking, those methods and tools constitute what is called a 'domain specific ontology'. The HyperPuccini project meanwhile is developing methods and tools that are appropriate for the study of Puccini, and the ontology of the HyperPuccini team will therefore be different from that of the HyperNietzsche group. The HyperLearning network is designed to permit each community to retain the ontological model most appropriate to its object of study, while simulta-neously optimizing interoperability between these communities by mapping domain-specific ontology onto a global ontology. This structure is often referred to as a 'global as view' data structure.

2.3 The HyperPlatforms as network nodes

In order to have a practical and effective impact on education and research, our HyperLearning model must be so conceived as to be easily customizable for other authors or other fields of research. From a methodological point of view, this means taking particular care in order to guarantee simple installation, configuration, and interoperability between the different network nodes. Such a system would allow any researcher who has a computer, access to the Internet, and a clear knowledge of the relevant subject to install and implement a HyperLearning Platform tailored to that particular field of research. No advanced technical competence would be required. We plan to use a technology similar to *DemoLinux* or *Knoppix*, designed to enable people with no technical training to install a Linux server. In addition to the Linux kernel and basic packages, the HyperLearning server will come bundled with a certain number of other software programs, all of which will be available at no charge and freely distributed in accordance with the Open Source model. Once this software is installed, a module is launched which provides step-by-step configuration for adapting the general HyperLearning framework to the specific field of study chosen by the scholars. The real challenge of our project is to preserve the rigorous structure of a HyperLearning system while making it flexible enough to be adapted to many different objects of study and easy enough to use that it will be widely adopted by humanities researchers. It requires not only refined theoretical analysis and extensive technical competence, but also the careful study of several test cases. The various HyperLearning nodes must also be developed in such a way that they can communicate easily with one another. Imagine, for example, that Schopenhauer is cited in an essay on HyperNietzsche. The reader should be able to move from HyperNietzsche to HyperSchopenhauer with a simple click of the mouse, and so have immediate access to the original context of the passage from Schopenhauer, translations of the passage in differ-ent languages, and relevant commentaries from Schopenhauer specialists. We refer to such cases as 'crossing hypertexts'. We can also imagine 'meta

hypertexts', that is to say, HyperLearning nodes for which the object of study is a theme that is common to more than one of the author hypertexts. For example, a hypertext dedicated to certain problems in epistemology may specify an object of study that includes elements from hypertexts dedicated to Aristotle, Hume, Kant, Schopenhauer, and so on. HyperLearning platforms will be implemented to be a node of the HyperLearning network. Every HyperLearning platform instance will be able to work both as a repository and as a service provider.

2.4 Functional languages for web programming[10]

The HyperLearning platforms implemented as network nodes are more than simple static sites. They are real programs that act as a front-end to the databases, and navigation through such a site is really an incremental, interactive execution of this program, with web pages being generated as intermediate execution states. The design, development, and testing of the corresponding programs are notoriously difficult and time- and labour-intensive, due to the immense number of navigation paths through a large website, and to several characteristics of the web. We believe that the functional programming paradigm has the potential to alleviate some of these problems, and is globally a good match for the programming of complex, interactive websites. The clean semantics, reliability, and conciseness of functional languages are, of course, important factors contributing to the quality of web applications but, more specifically, we identified some points of convergence between functional languages and web applications.

By nature, interaction with a website, as well as the supporting web protocol (HTTP), is essentially stateless: when the user clicks on a link, the only information sent to the web server is the selected URL and the contents of form fields from the current page, but not a complete history of the sequences of interactions that led to this link. Modest amounts of state data can be maintained on the client side (web 'cookies' are an example of this) and on the server side (sessions), but such state data is limited in size and not completely reliable: 'cookies' can be erased by the user, and server sessions expire after some time. Web applications written in conventional imperative languages tend to rely on such state data, and consequently fail to handle correctly several user interactions: navigation through the browser history mechanism (the 'back' and 'forward' buttons); cloning of a page in two browser windows; bookmarking of an arbitrary page; etc. A better approach is to eliminate this reliance on local state data altogether, and to structure the web application in such a way that all the information needed for navigation is explicit in the URLs and fields. Besides addressing the issues above, this approach supports replication of the web application across several servers, a crucial feature for reliability and performance.

Web applications communicate with the user via on-the-fly generation of web pages in HTML format. Internally, they often use semi-structured data formats such as XML to store and access persistent data. Both HTML and instances of XML are complex data formats that must obey strict formal constraints in order to guarantee interoperability (that is, they must do so in order to display properly on all web browsers) and database integrity. These formal constraints are precisely defined in so-called DTDs or Schemas, and many tools exist to validate an HTML page or XML document against a given DTD. However, when the page is generated dynamically by a program, these validators can only establish the correctness of one run of the program. Ensuring once and for all that all runs of the program always generate valid HTML or XML documents is highly desirable, yet much harder. For this reason one of the research objects of the HyperLearning project is to adapt functional languages for use in HTML and XML type checking.

The result will be validated and tested by the software development part of the project. If this research is successful it can produce a remarkable advance in web programming techniques.

Acknowledgements

Leroy, Guido Vetere, Riccardo Giomi, Francesca Di Donato, and Hans Walter Gabler.

References

Barbera, M. and Giomi, R. (2004). The Pearl Diver Model. *Jahrbuch für Computerphilologie*, **5**: 22–30.

D'Iorio, P. et al. (2000). *HyperNietzsche: Modèle d'un Hypertexte Savant sur Internet pour la Recherche en Sciences Humaines. Questions Philosophiques, Problèmes Juridiques, Outils Informatiques*. Paris: Presses Universitaires de France (PUF).

Fagin, R., Kolaitis, P. G., Miller, R. J., and Popa, L. (2002). *Data Exchange: Semantics and Query Answering*. IBM Research Report.

Guédon, J. C. (2001). In Oldenburg's long shadow: librarians, research scientists, publishers, and the control of scientific publishing. *Proceedings of 138th ARL Membership Meeting*, Toronto, Ontario, 23–25 May 2001. http://www.arl.org/arl/proceedings/138/guedon.html (accessed 9 June 2004).

Lambe, P. (2002). *The Autism of Knowledge Management*. http://greenchameleon.com/thoughtpieces/autism.pdf (accessed 9 June 2004).

Lenzerini, M. (2002). Data integration: a theoretical perspective. *Proceedings of the ACM Symposium on Database Systems (PODS)*, pp. 233–46.

Molina, H. G. et al. (2004). *Integrating and Accessing Heterogeneous Information Sources*. http://www.db.ucsd.edu/publications/tsimmis-abstract-aaai.pdf (accessed 9 June 2004).

Welsh, E. (2003). *SCORM: Clarity or Calamity?* http://www.onlinelearningmag.com/onlinelearning/magazine/article_display.jsp?vnu_content_id=1526769 (accessed 9 June 2004).

Yang, B. and Molina, H. G. (2002). Designing a Super-peer Network. http://citeseer.nj.nec.com/fagin03data.html (accessed 9 June 2004).

Notes

1 http://www.hyperl.org (accessed 9 June 2004).
2 http://www.hypernietzsche.org (accessed 9 June 2004).
3 http://www.hypernietzsche.org/doc/committee/ (accessed 9 June 2004).
4 http://www.openarchives.org/OAI/openarchivesprotocol.html (accessed 9 June 2004).
5 http://www.caml.org (accessed 9 June 2004).
6 http://www.haskell.org (accessed 9 June 2004).
7 http://xduce.sourceforge.net (accessed 9 June 2004).
8 http://www.demolinux.org (accessed 9 June 2004).
9 http://www.knoppix.org (accessed 9 June 2004).
10 This part is based on the work of Xavier Leroy and his team done during the first submission of the project to the 6th Framework Programme of the European Union.