Peer-Reviewed, Publishable Hypertexts: A First Look

Peter J. Nürnberg and David L. Hicks

Department of Computer Science, Aalborg Univeristy Esbjerg
Niels Bohrs Vej 8, DK-6700 Esbjerg, Denmark
{pnuern,hicks}@cs.aue.auc.dk
http://cs.aue.auc.dk/

Abstract. Despite the fact the hypermedia has long been discussed as a suitable format for academic research publications, the academic community has made little progress toward realizing hypermedia publishing. In this paper, we provide an initial framework for considering the problems facing any system designed for publishing academic hypertexts, and propose some first steps our community can take toward realizing this goal.

1 Introduction

Hypertext and hypermedia are now part of the everyday landscape. Generally, computer-literate people are familiar and even comfortable with hypermedia documents. Hypermedia is a mainstream technology. Furthermore, it is recognized by many people as the most appropriate medium for delivery of certain types of information. From help files to API documentation, many people have become accustomed to seeing certain highly interconnected types of information delivered as hypertext. Other forms of delivery for such information, such as printed manuals or online "flat" text files may be seen as inappropriate and impoverished.

Despite this, the academic community has not adopted a mechanism by which original research contributions may be submitted and published in hypertext form. Certainly, this is not due to the nature of the information. A hallmark of quality research is its positioning within the fields to which it contributes. This requires connections to be built among pieces of existing work and the research in question. Research contributions without such connections is at best difficult for others to reuse and at worst inaccurate and misleading. (When we speak of connections, we do not mean only references to other work, although this is certainly a very common and valuable type of connection to build. We also mean to include higher-level semantic connections – observations that tie together several existing research contributions for the purposes of comparison, contrast, support, refutation, extension, or other exegesis.)

Currently, the methods we use to mark connections among works in our research contributions (such as references or extended quotations) are often clumsy,

prone to error, and difficult to use. At first glance, hypertext seems to be an ideal technology for the representation of these connections. Even the paper widely acknowledged to be the first in the hypertext research area, Bush's famous As we may think article of 1945 [2], envisioned the use of hypertext in academia as a means to convey accurately connections among existing work. Bush went so far as to describe an entire profession devoted to the generation of structure among already given data (trailblazers). However, we have yet to adopt any conventions for actually realizing these types of contributions.

In this paper, we take a first look at a mechanism by which hypertexts might be compiled, peer-reviewed, and published. We begin by listing some requirements we see as critical for any such mechanism, and discuss some of the technological issues and non-technological barriers in realizing a system that meets the requirements we generated. We close with an agenda that will help use move toward the goal of realizing peer-reviewed, publishable academic hypertexts.

2 Requirements

In this section, we examine some of the requirements for a hypertextual academic research publishing system (HARPS). Before we begin, we offer this initial working definition: a HARPS is a set of technologies and practices that enable the production and dissemination of academic research in hypertext form. That is, we are not only speaking of particular software (or hardware) systems, but also of guidelines for how people will interact with such computer systems, with one another, and with the generated artifacts. Clearly, a HARPS must reside within a larger system that enables production and dissemination of academic research in other, more traditional, forms, such as printed, non-hypertextual proceedings or journals.

We also point out that this discussion is meant only to seed further work. The set of requirements we present is by no means complete. However, we believe it represents a useful starting point for future work.

Broadly speaking, we can divide the requirements for a HARPS into two categories: infrastructure and frontend. Infrastructure requirements are largely technological. We offer them as axiomatic, but based within the framework of previous, related work. Frontend requirements are largely motivated by social and political factors. We divide our discussion of frontend requirements by addressing the motivations of various parties to the academic publishing process.

2.1 Infrastructure requirements

From a systems perspective, the infrastructure requirements for a HARPS would seem to induce another OHSWG (Open Hypermedia Systems Working Group) scenario [7]. Traditionally, the OHSWG would begin work in this application domain by generating one or more scenarios of system use, which could then be analyzed for structural abstraction requirements. Before we begin this exercise, however, it may be worth asking whether we can reuse one or more existing

OHSWG scenarios. Have we discussed the problem of publishing hypertexts before?

In fact, although there is no existing OHSWG scenario that exactly describes a system used for the purpose of publishing hypertextual academic research, there have been a number of closely related discussions. As pointed out above, Bush [2] describes the production of academic research in hypertext form. Engelbart [4] describes hypertext support for generic knowledge workers. Nrnberg et al. [5] describe the production of taxonomies by botanical taxonomists. This discussion relies on the "navigational" and "taxonomic" structural abstractions identified by the OHSWG. In this field, such taxonomies comprise a significant portion of the academic product. Nrnberg et al. [6] also describe a scenario in which linguists generate various intermediate and final hypertextual products modeling language relationships. In addition to the "navigational" and "taxonomic" structural domains, this scenario also uses "spatial" and "issue-based" structural abstractions.

In fact, at a broad level, the potential types of structural abstractions used by academics to represent their work products comprise an unbounded set. This may be a rather bleak outcome for this inquiry. Supporting such a wide variety of structural abstractions may simply be impractical.

Although some might advocate proposing a generic structural computing or component-based open hypermedia system (CB-OHS) approach to handling this open and unbounded abstraction situation (approaches specifically designed to handle arbitrary co-existing structures), we would propose that we look for a simpler, if strictly less powerful, solution. Consider the current situation in academic publishing – despite the fact that work products come in many different forms (botanical taxonomies, linguistic relationships, etc.), the forms in which research is peer-reviewed, published, and disseminated are relatively constrained. Is this a result of a poor fit between how results are generated and how they are disseminated, driven by an inadequate technology (paper)?

Perhaps, but leaving aside how we arrived at our current model, there are some powerful reasons for "normalizing" dissemination products across fields. Such uncoupling of work product format from dissemination format allows academics trained in virtually any field at least to be able to read through the results in other fields. One need not be familiar with all the intricacies of botanical taxonomies to be able to read a journal paper on the subject (although such familiarity is obviously helpful) – it is sufficient (for certain limited purposes) simply to be trained to read and understand journal papers.

What can we take away from this analysis? We may be best served by choosing structural abstractions (a hypertextual dissemination format) for research results that is uncoupled from the structures used in the construction of those results. As a starting point, we propose using the abstractions identified by the OHSWG as "navigational" hypertext abstractions (see, e.g., [7] [8]). Although discussed at great length elsewhere, briefly, this structural domains allows for the creation, manipulation, and deletion of first-class, n-sided, optionally directional associations (links) between (whole or parts of) node content. Nodes

wrap arbitrary (possibly third-party managed) content. Being first-class objects, associations themselves may associated, may be tagged with arbitrary attribute-value set pairs, may be subject to access, concurrency, and versioning control, etc.

Some may argue that a number of useful structural abstractions, such as transclusions or composites are missing from the base navigational model. Clearly, more abstractions lead to more powerful systems. However, they also lead to more complex systems. We feel that the base navigational model represents a good starting point. These abstractions support at least the final delivery format described in the aforementioned scenarios. Also, since navigational hypertext systems are by far the most common in use today, such systems presumably face a smaller acceptance threshold by users.

Another potential issue is how much "complex" functionality should be made available. Using the OHSWG "1+4" paradigm[1], how many (and to what degree) of the "4" additional services should be included. We see a clear need for versioning, since existing academic research products are also versioned, and we do not want a HARPS that is strictly less powerful than existing systems of publishing and dissemination. Access control is also clearly necessary, since copyright holders (and/or authors, if these are different) should be able to modify or delete structures, even if generic readers cannot. Concurrency and notification control, while useful in authoring, seem less so in dissemination systems, as long as systems support a many-read semantic on unlocked entities. Since versioning and access control in hypermedia are still open research areas, it is difficult to formulate the exact form of these additional services.

Finally, it should be clear that only an open systems solution can be considered here. There can be no justification for imposing a fixed (set of) interaction style(s) on all authors and readers anymore.

Several existing, freely distributable and licensable (CB-)OHS's currently under development (e.g., Callimachus [3] or Construct [9]) could be used to support basic navigational structures over an open set of applications. Thus, the basic front-end technological concerns may be reasonably expected to be handled by current technology. Continuing work by the OHSWG on standardizing structure interchange formats will fill in the remaining technological issues on this front. With respect to advanced functionality (versioning and access control), while these exist in both of the aforementioned systems, much work still lies ahead for the OHSWG on the standardization front before this area can be considered fully addressed.

To summarize, a HARPS must support:

[1] The "1+4" explanation of services was coined at OHS 3.5 in Aarhus, Denmark. The "1" refers to a base service, such as the "navigational" service. The "4" refers to versioning, access, concurrency, and notification control services that can be "layered" on top of the base service. The "1+4" formulation is meant to imply that the four "advanced" services are optional, since they may not be needed in many cases

Req. 1. an open set of applications (clients);
Req. 2. the structural abstractions of the basic OHSWG navigational model;
Req. 3. some (as yet to be fully specified) level of versioning support; and,
Req. 4. some (as yet to be fully specified) level of access control.

We explicitly avoid more expressive possibilities, such as systems that support open sets of structural abstractions, due to their inherent complexity, and the fact that we see advantages in "normalized" representations for published academic research.

2.2 Frontend requirements

There are a number of frontend requirements for any HARPS. Above all, there must be motivation and reward for using a HARPS over (or in conjunction with) traditional dissemination systems for the readers, the authors, and the reviewers. We treat each of these cases separately.

Reader motivated requirements. What advantages does a reader of hypertextual research enjoy over a reader of research published with traditional media only? Clearly, however a particular HARPS is implemented, it must provide as many of the favorable aspects of traditional publishing systems as possible. A HARPS that lacks desirable features of traditional publishing systems certainly faces a much greater adoption challenge. Where possible, features of traditional systems should be augmented. Finally, there may be features of a HARPS not able to be easily reproduced in traditional systems.

One way in which to secure many positive features of traditional systems within a HARPS is to provide a method for hypertextual research to be rendered on paper. This requirement, though easy enough to state, is in fact quite challenging to meet. Hypertextual models such as the OHSWG navigational model admit arbitrarily complicated structures – links may be n-sided, may link to both nodes and other structural abstractions (and thus be nested to arbitrary levels), and may have associated traversal semantics that take the form of arbitrary computations (that, in turn, may generate traversal destinations dynamically, interact with the reader, check traversal state, etc.) How can such structures and behaviors be rendered on paper?

There are three basic types of approaches to this problem. The first type essentially attempts to define paper renderings for all types of structures and behaviors. We believe such types of approaches are doomed to failure, because even moderately complicated structures are likely to have such complex generic renderings as to make them unreadable. The second type of approach essentially disallows certain types of structures and behaviors, arbitrarily limiting authors to a simple, "renderable" set of abstractions. While this addresses the potential unreadability problem engendered by the first type of approach, we see this as unnecessarily limiting to authors. We do not yet know what kinds of structure and behaviors authors will find useful – deciding on limitations now is premature.

The third type of approach places the rendering requirement on the author, by forcing the author to define "paper rendering semantics" for all structures and behaviors used in a given hypertextual work (as is possible in KMS [1], which provides ways in which to affect the behavior of the "linear" program, which produces a paper renderable document from a KMS hypertext.) While this may be seen as a great burden to place on authors, we feel that the advantages of doing so are great enough to merit its requirement. Over time, we expect that communities would develop shorthand ways of describing common renderings, thus reducing the onus on authors.

What features of traditional publishing systems can a HARPS augment or improve; and, what totally new features can a HARPS offer? Certainly, HARPS software can render structures and behaviors in an active way, making structures traversable, computations executable, etc. This may take very simple forms (e.g., allowing references to be retrieved by a mouse click) or very complicated forms (e.g., allowing traversal between a set of data and multiple analyses carried out dynamically). Limited depth-first searches of hypertextual research products could be used as a way of building high-level overviews or "digest versions" of work. A HARPS could also offer annotation capabilities. Annotations could be linked into hypertextual documents, and themselves become a further research product that could be distributed, shared, and even published as further work. This simple scaling from informal annotation support to full peer-reviewed publishing could be a powerful benefit for readers who themselves become authors. Finally, a HARPS should offer readers the ability to select which structure/behaviors they wish to see/execute. The OHSWG model abstraction of context might be useful here.

To summarize, a HARPS must support:

Req. 5. a method by which authors specify a "printable rendering" of their products;

Req. 6. the ability to render structures and behaviors actively in software; and,

Req. 7. tailorable and personalizable reading experiences.

Author motivated requirements. Why do authors of academic research publish? Certainly, the desire to disseminate results is one important factor. Inasmuch as this is true, advantages for readers of a HARPS are also by extension advantages for authors – if a HARPS offers benefits to readers of research work, this helps the author disseminate this work. More interesting are other motivations for authors. How can a HARPS appeal to these motivations?

An author of hypertextual research work must be able to use this work to help secure promotion, tenure, or other perquisites on the job. Traditional publishing is an important aspect of acquiring these job benefits. Work published within a HARPS must also confer these benefits to authors. Informally, we have noticed that any type of electronic publication is less likely to be weighted as highly by academic administration as traditional publications. To the degree that this is true, we must recognize that hypertextual publications will also tend

to be discounted if delivered only in electronic form. We feel, however, that the requirement on authors of generating "paper rendering" directions may mitigate the problems faced by electronic-only publications. In the foreseeable future, an important component of any HARPS will be the traditional publication of these paper rendering, even if these rendering are not the primary means of consumption by other academics.

Perhaps the most difficult hurdle faced by a HARPS will be how hypertextual work is evaluated by other academics. What kinds of structures represent significant research contributions? What types of tradeoffs are there between quantity and quality of structures? How complex do structures (and the insights behind them) need to be to constitute a workshop position statement, a conference full paper, or a journal article? All of our experience points to the fact that building hypertextual works of any sort is difficult. Consider hypertext fiction, which by sheer word count is often much shorter than comparable conventional writing. We feel that academic communities will need to develop an appropriate weighting for the construction of complex hypertextual structures (when these structures are informative and insightful, of course). Stated another way, authors will need to be justly compensated (recognized) for their hypertextual work.

To summarize, a HARPS must support:

Req. 8. a major focus on the publication and distribution of author specified "printable renderings"; and,

Req. 9. a set of conventions for the judging the complexity and value of hypertextual structures and behaviors.

Reviewer motivated requirements. In many ways, reviewer motivations may be the easiest to dismiss if one superficially considers the problems facing a HARPS. However, we feel reviewers must be carefully considered if widespread adoption of hypertextual work is to become a reality. Reviewers most often agree to review work partly out of motivation for the "public good". The academic publishing system only works if there are reviewers who are willing to assess the appropriateness, originality, and overall value of contributions. Recognizing this, many academic feel duty-bound to review, at least occasionally. This motivation seems independent of the actual publication format, but it may not be independent of the submission format (which also takes hypertextual form in a HARPS). Therefore, any HARPS must support the needs of reviewers. Submissions must be able to be read on a wide-variety of platforms, etc. This, however, would seem to be addressed by the open systems requirement (Req. 1) described above. More importantly, there must be a way for reviewers to annotate the structures and behaviors of a submission, much in the same way that reviewers can markup a paper submission with a pen.

Some electronic journals provide a method by which reviewers can "co-publish" their reviews along with an article. A HARPS should of course accommodate this possibility. We feel, however, that it is worth pointing out that a HARPS would be especially able to support this practice quite easily. by linking reviews off of the reviewed article. Reviews might be modeled as separate

contexts that are initially only viewable by a program committee, and then are made optionally available to wider audiences. More broadly, one can imagine an "open reviewer" model in which comments on articles form a second-level literature all their own. The line between "original contribution" and "review" may begin to blur.

To summarize, a HARPS must support:

Req. 10. annotations on content, structures, and behaviors; and,
Req. 11. a convenient way in which these annotations (reviews) can be linked from works, and made optionally available.

3 Agenda

An agenda to move forward toward the goal of realizing peer-reviewed, publishable academic hypertexts must consider both the infrastructure and frontend challenges faced by a HARPS. Addressing the category of infrastructure challenges will require further research focused on the lower (backend and middleware) levels within an open hypermedia environment, in that these challenges correspond to functional requirements that are most appropriately addressed at these levels. The frontend challenges are generated by the necessity to support specific user capabilities and work practices within a HARPS in order for it to succeed. Addressing this category of issues will require further work at higher levels within an open hypermedia environment, to ensure that applications provide the required functionalities for users, and that policies and guidelines are in place to ensure the required work practices are followed.

3.1 Infrastructure Issues

As mentioned earlier, several component based open hypermedia systems currently under development support basic navigational structures over an open set of applications. These systems are largely capable of handling the more basic technological issues raised earlier (Req. 1 and Req. 2). It is important, however, that along with the continued development of these systems that the standardization process of the OHSWG also progress so that the interchange of structures between systems is possible.

While some open hypermedia systems do provide support to address certain aspects of the more advance technological issues of versioning and access control (Req. 3 and Req. 4), more work is needed in these areas. A reasonable place to begin is with a detailed study and analysis of the versioning and access control requirements of a HARPS. First, an examination of the current procedures used for submitting, evaluating, and publishing academic research can be used to elucidate the requirements that are inherent in the process. In addition, existing (OHSWG) scenarios can be extended, and new ones can be created, that deal more directly with the peer review and publishing process to uncover ways in which versioning and access control functionality can be used in a HARPS to augment existing procedures. Once a more specific list of requirements has been

generated, it can be used to analyze the facilities of existing systems. This can serve as a guide for researchers in extending their systems to accommodate the infrastructure support required by a HARPS.

An additional benefit of starting with an analysis of the infrastructure requirements of a HARPS is the targeted approach it promotes for coping with the issues that are involved. Versioning and access control are challenging capabilities to provide, especially in the general open hypermedia environment. Identifying and focusing on the specific subset of functionality from each of these areas required for a HARPS will enable solutions to be developed faster than attempting to address each issue comprehensively. At the same time, the targeted solutions developed to address these issues for a HARPS will be informative for researchers considering them within the broader general open hypermedia environment.

3.2 Frontend Issues

Critical to the success of a HARPS is the provision of the required capabilities at the user level as described earlier. A number of requirements from Section 2, especially the reader motivated and reviewer motivated ones (Reqs. 5, 6, 7, 10, and 11) identified specific functionality necessary in a HARPS environment. These all must be supported in order for a HARPS to succeed. Some of these capabilities, such as creating printable rendering of works and defining annotations, exist in various forms in current applications. Others can be built utilizing OHS infrastructure support, including the expanded access control and versioning facilities described above. Providing these capabilities will most likely involve a combination of modifying existing OHS applications and middleware components and perhaps creating new ones. Regardless of the strategy utilized, it will be important for standardization to take place, so that the solution will be an open systems one.

A number of the requirements from Section 2, especially the author motivated ones (Req. 8 and Req. 9) pertain to the work practices followed by those involved in the publication of academic works. Addressing these issues requires studying current practices and policies and developing from these a set of guidelines for HARPS users.

3.3 Call to Action

Can the OHSWG workshops (the Workshop on Open Hypermedia Systems and the Structural Computing workshops) be a testbed for hypermedia publishing? This is a tantalizing idea. In this paper, we've pointed to various areas that will require more study, but how best to carry out these studies? Perhaps the question should be, "how better than to try ourselves?" We propose that within the next 2 years, both workshops adopt a HARPS track which would allow authors to experiment with compiling and submitting hypertextual work, and would allow the community to evaluate the technologies and practices used to review and publish such work. This should be done in parallel to the calls for further study

made above. One might argue that if this community cannot or will not adopt such a track, who else will?

References

1. Robert M. Akscyn, D. McCracken, and E. Yoder. KMS: A distributed hypermedia system for managing knowledge in organization. Communications of the ACM 31, 7 (Jul 1988), 820–835
2. Vannevar Bush. As we may think. Atlantic Monthly 176, 1 (Jul 1945). 101-108.
3. Dimitris Christodoulakis, Michael Vaitis, Athanasios Papadopoulos and Manolis Tzagarakis. The Callimachus approach to distributed hypermedia. Proceedings of ACM Hypertext 1999 (Darmstadt, Germany, Feb).
4. Douglas C. Engelbart. Augmenting human intellect: A conecptual framework. Stanford Research Institute Technical Report AFOSR-3223 Contract AF 49(638)-1024 (Oct 1962) Palo Alto, CA.
5. Peter J. Nürnberg, John J. Leggett, and Erich R. Schneider. Designing digital libraries for the post-literate age. Journal of Universal Computer Science 2, 9 (Sep 1996).
6. Peter J. Nürnberg, Uffe K. Wiil, and John J. Leggett. Structuring facilities in digital libraries. Proceedings of ECDL 1998 (Crete, Greece, Sep).
7. OHSWG. 1997. Open Hypermedia Systems Working Group Home Page. Available at `<http://www.ohswg.org/>`.
8. Sigfried Reich, Uffe K. Wiil, Peter J. Nürnberg, Hugh C. Davis, Kaj Grønbæk, Kenneth M. Anderson, David E. Millard and Jörg M. Haake. Addressing interoperability in open hypermedia: the design of the Open Hypermedia Protocol. New Review of Hypermedia and Multimedia 5, 1 (1999).
9. Uffe K. Wiil and Peter J. Nrnberg. Evolving hypermedia middleware services: Lessons and observations. Proceedings of ACM SAC 1999 (San Antonio, TX, Mar).