Processing Link Structures and Linkbases on the Web [*]

François Bry
Institute for Informatics
University of Munich
http://www.pms.ifi.lmu.de/
francois.bry@ifi.lmu.de

Michael Eckert
Institute for Informatics
University of Munich
http://www.pms.ifi.lmu.de/
michael.eckert@stud.ifi.lmu.de

ABSTRACT

Hyperlinks are an essential feature of the World Wide Web, highly responsible for its success. XLink improves on HTML's linking capabilities in several ways. In particular, links after XLink can be "out-of-line" (i.e., not defined at a link source) and collected in (possibly several) linkbases, which considerably ease building complex link structures.

Modeling of link structures and processing of linkbases under the Web's "open world linking" are aspects neglected by XLink. Adding a notion of "interface" to XLink, as suggested in this work, considerably improves modeling of link structures. When a link structure is traversed, the relevant linkbase(s) might become ambiguous. We suggest three linkbase management modes governing the binding of a linkbase to a document to resolve this ambiguity.

Categories and Subject Descriptors

H.5.4 [**Information Systems**]: Hypertext/Hypermedia—*Navigation*; I.7.4 [**Document and Text Processing**]: Electronic Publishing

Keywords

Hyperlink, XLink, Link Modeling and Processing, Linkbase

1. INTRODUCTION

Hypertext links are an essential, if not the most important, feature of the World Wide Web. Undoubtedly, much of its success is due to the capability to link from some content to *any other* content on the Web using HTML links. One might say, the Web supports "open world linking". It makes the Web significantly different from traditional hypermedia systems, including Open Hypermedia Systems (OHS). In general, these assume that all resources (documents or document parts) an application might refer to are known in advance and addressable through unique identifiers. One might say, they assume "closed world linking."[1]

[*]This research has been funded by the European Commission and by the Swiss Federal Office for Education and Science within the 6th Framework Programme project REWERSE number 506779 (cf. `http://rewerse.net`).

[1] "open" in OHS refers to interlinking documents from arbitrary applications with minimal or no hyperlink support.

WWW 2005, May 10–14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

XML links [3] improve on HTML's links in several ways. In particular, a link after XLink can be "out-of-line", i.e., specified in another document than its source. Out-of-line links make linkbases, i.e., specifications of links outside the data (Web pages or content) the links refer to, possible. In turn, linkbases considerably ease building complex link structures as encountered in context-adaptive systems (e.g. teaching systems) or in versatile information systems (e.g. catalogs). For this reason, most OHS offer out-of-line links and support linkbases. For a simple processing of linkbases they exploit the fact that all resources the links of an application might refer to are known in advance and have unique identifiers. As a consequence, the processing of out-of-line links and linkbases of OHS is not applicable on the Web.

This work is devoted to processing out-of-line links and linkbases under the open world linking assumption of the Web. This issue has been quite neglected: While advanced linking constructs have been proposed with XLink in a proposal that has reached the status of a recommendation almost four years ago, the processing of these constructs under the conditions of the Web, i.e., its open world linking, has not received much attention.

2. LINKING ON THE WEB WITH XLINK

A peculiarity in XLink terminology is worth stressing: In XLink, an *extended link* is a collection of *arcs*. Hyperlink navigation as known from HTML's hyperlinks is expressed with arcs; extended links are only a "container" for them.

Arcs put two resources (anything addressable, e.g., whole XML documents or fragments thereof) into a directed relationship, designating one source (starting point for link traversal) and the other destination. Arcs can be *typed* with an *arcrole*, which is a URI (similar to properties in RDF).

Resources can be *local* to an arc (i.e., in the same document) or *remote*. In difference to HTML's (*inline*) links, not only the destination, but also the source of an XLink arc can be remote. In this case we speak of *out-of-line* links (or arcs) that reside in a *linkbase*, an XML document separate from the link sources.

Out-of-line links allow a separation of content and link structure that significantly eases building and maintaining sets of documents with complex link structures.

3. INTERFACES FOR LINK STRUCTURES

In modeling a collection of semantically related documents, a hyperlink rarely comes alone. Generally, a number of links together impose some higher structure, such as

- *guided tours*, sequencing a number of independent documents into a reading order, e.g. for eLearning,
- *link catalogs*, supplying a document, e.g. a table of contents, with many outbound links,
- *glossaries*, linking occurrences of terms in a set of documents to their definitions in a glossary-like document,
- *cross-references*, ranging from simple cross-references to sections within an article or a book to complex cross-references, e.g. in code, API-documentations, or meta-data annotations in Semantic Web applications.

Such structures benefit highly from a separation of content (documents) and hypertext structure (linkbases) as possible with XLink. By applying different linkbases, a single document can be viewed in different contexts, e.g., stand-alone, as part of a guided tour, with links into a glossary, or enriched with annotations, depending on a reader's needs.

Hypertext structures on the Web easily suffer from dependence on the URIs of the link endpoints. When building a structure, URIs have to be known in advance by the author, and the structure has to be modified in case a URI changes. Also, often the need to combine structures into composite structures arises, e.g., to sequence two or more guided tours into a composed guided tour.

We propose to overcome these difficulties by introducing *interface* resources in XLink. Abstracting hypertext space into a directed graph with resources as nodes and links (i.e., "arcs" in XLink terminology) as directed edges, interfaces are nodes just like normal resources, but are transparent to the user and will not be displayed. When a user traverses an arc inbound to an interface, instead of displaying it, an outbound arc will be traversed automatically.

Our notion of interfaces, discussed in detail in [1], increases flexibility and allows to decouple structure from a document's "physical" location (URI) by providing an indirection. It aids authoring, esp. collaborative authoring, and maintainability. By using them to signal well-defined junctures, interfaces also facilitate composition and reuse.

4. LINKBASE MANAGEMENT

The separation of content and link structure poses a new challenge: how can the linkbases relevant for document be found in an open world scenario, and which of the found links should actually be displayed? The procedures defining this are referred to as *linkbase management* here.

XLink uses so-called linkbase-arcs (discernible by a predefined arcrole) pointing from a document to its linkbase(s) to locate linkbases. Sensibly, this introduces no new constructs to XLink. However, it also introduces a circularity, as the linkbase-arcs underly linkbase management themselves.

Unfortunately, linkbase management with XLink's linkbase-arcs is underspecified: Consider the situation in Figure 1. A linkbase B contains a link l_1 from document D_1 to D_2 and a link l_2 from D_2 to D_3. When a user views D_1, a linkbase-arc l_b (defined inside D_1) can make the user's browser aware of the links in B and link l_1 is shown. Let the user now traverse l_1 to arrive at D_2. Note that D_2 does *not* contain a linkbase-arc to B. From the XLink recommendation it is unclear whether information about the existence of l_2 is retained (and it is displayed when viewing D_2) or not.

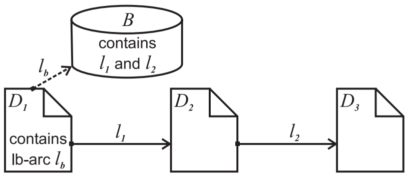

Figure 1: Ambiguity of linkbase-arcs in XLink

Neither approach is satisfactory: *Inflationary* collecting all possible links poses challenges to both the users and processing efficiency. The converse, *conservatively* considering only linkbases referred to by a document, requires the document to know all its linkbases and prohibits imposing different link structures onto a document. See also [2].

In [1], we suggest replacing XLink's linkbase-arcs with three complementary types of linkbase-arcs indicating a "binding mode" for a linkbase. A linkbase can be bound to only the current document (`transient`-arc), all documents (`permanent`-arc), and a limited range of documents (`temporary`-arc), achieving nicely the variety of behaviors required by many Web applications.

Details on the semantics and a processing algorithm for the three linkbase-arcs can be found in [1]. There, we also demonstrate the potential of our linkbase management mechanism in several usage scenarios.

5. CONCLUSION AND PERSPECTIVES

Our work is motivated by allowing reuse and composability of structures under an open world linking assumption. A document can be seen in different contexts, i.e., as part of different link structures. It should be possible for this context to be supplied not only by the author but also users that have no write-access to the original document.

Open world linking on the Web significantly differs from closed world linking in before-web systems and requires new solutions. This work has introduced the concept of interfaces for link structures and a linkbase management mechanism based on three binding modes. XLink can be extended conservatively, i.e., only additions to the language are made, to accommodate both ideas.

While our ideas have been presented as extensions to XLink in a Web of hypertext, we suggest in [1] that the raised issues and the presented solutions are also relevant to the Semantic Web and its languages. We believe that the interconnection of data and the connection of data and meta-data spread over different Web resources is still an underexplored and maybe neglected issue on the Semantic Web deserving further investigation.

6. REFERENCES

[1] F. Bry and M. Eckert. Processing link structures and linkbases and its relevance to the semantic web. Technical report, Institute for Informatics, University of Munich, 2004. `http://www.pms.ifi.lmu.de/publikationen/#PMS-FB-2004-25`.

[2] F. Bry and M. Kraus. Advanced modeling and browsing of technical documents. In *Proc. 17th ACM Symp. on Applied Computing*, 2001.

[3] S. DeRose, E. Maler, and D. Orchard. XML Linking Language (XLink) Version 1.0. W3C recommendation, World Wide Web Consortium (W3C), June 2001.