

Issues of Transclusions

Harald Krottmaier

Institute for Information Processing and
Computer Supported new Media (IICM)
Graz University of Technology, Austria
hkrott@iicm.edu

Denis Helic

Institute for Information Processing and
Computer Supported new Media (IICM)
Graz University of Technology, Austria
dhelic@iicm.edu

Abstract: Transclusions were first mentioned in the early 1960 by Ted Nelson. Generally speaking transclusions are a method of reusing parts of a document without duplicating these parts. In the proposed implementation authors of documents will be informed (on request) if parts of these documents are reused elsewhere. Authors using transclusions are informed about changes in the reused document. Reused document fragments are marked in both documents to help readers while exploring documents. The concept is old but implementations are still not available to authors and readers. After a short introduction this paper shows some applications of transclusions. Thereafter ideas on how to implement transclusions using a second generation hypermedia information system are presented.

Introduction and Problems

The idea of transclusions was born in 1960, when Ted Nelson invented Xanadu, a revolutionary information- and document management system. “Reuse without duplicating document fragments with the original context available” ([Nelson, 1995]) is the key issue of the transclusion concept. Newly assembled documents are also called *compound documents*, on the other hand, documents reused in one or more compound documents are called *transcluded documents*.

As there is no duplication of document fragments when reusing content many advantages over ordinary cut-and-pasting of data arise (see e.g. [Krottmaier and Maurer, 2001]). In the following we are going to numerate some of the advantages.

Intellectual Property: Parts of a document are not ‘stolen’ but simply ‘reused in another context’. It must be visible to the reader of a document that parts of the document are no original contributions, but references or inclusions of some other work. Dependent on the electronic format of the compound document this is easily implemented by adding hyperlinks to the original work before and after the inclusion of the reused material. Please note that in HTML documents it is possible to reuse whole images in any context stored on any server system. The reader of the document will not notice this fact without reading the source-code of the HTML document.

Disk Space: Since parts of a document are not copied into a new document there is no need to save the reused part more than once. Surrogates of the reused parts must be stored in a predefined format. There will be no waste of disk space, if the surrogate definition requires less disk space than the referenced part. It is obvious that transcluded documents must be available to the system when assembling the document. Without coping the reused part and without sophisticated caching mechanisms this fact may limit the applications of transclusions to intranet or even intra-server applications.

Update: There is simply no need of manually updating referenced parts because the content is always requested from the *original source*. Since this is an advantage in many applications (e.g. a collection of different course material which is automatically up-to-date) there are drawbacks in some other applications where many people are responsible for the content and access control is not an issue. Some

action must be taken if reused parts are changed. At least the author of the compound document must be informed (e.g. via email) by the system if there are any changes in the reused part. While there is no need to update a document it might be a time-intensive task to assemble a compound document. Therefore reusing a document which reuses parts of another document (which itself reuses parts of other documents etc.) must be limited to a certain level. Nevertheless, improvements in network infrastructure and processing speed will reduce this problem.

Two Way Reading: When displaying the transcluded document the system must add additional navigation facilities around the quotes in the compound document. This action is necessary to indicate the reuse of content (see 'Intellectual Property'). Links to the original source of the transclusion are provided. This makes it possible for the reader of a document to take a look at the original context of a quote. The reader of the transcluded document on the other hand has the ability to explore which parts of a document are no original contributions. If a fragment of a document is reused it is very likely that this fragment is important!

These are some of the most important advantages and drawbacks when using transclusions. Authors of compound documents must be aware of the risks when using this technology. The information system where these documents are stored must support authors in many ways (i.e. creation of transclusions, notification of changes, types of changes [Francisco-Revilla et al., 2001], etc.).

Applications of Transclusions

With the knowledge of the concept many existing applications can be improved. This section lists three possible applications. Please note that many other applications will profit from the idea of transclusions.

Electronic Publishing: Scholar papers often quote paragraphs or sentences of other already published papers. When using transclusions readers of the new paper are able to jump directly to the context of the quote (i.e. to the original paper). It is therefore possible for the reader of the compound document to explore the reused quote in the original context. This issue is especially useful for novices in a topic. The author of the compound document may create a special 'view' of a topic. Readers who are interested in the details may read also the original contributions to the topic. Readers of the transcluded document on the other hand are able to see which parts of a document are reused in other documents. This is a very important information! If a part of a document is reused in some other document, it is very likely that this part is very important. This additional 'meta-data' ("This part is also used in some other document(s)...") is a kind of implicit rating of a fragment of a document. In addition to attributes in link-objects (whether the part is used as a positive or negative example) useful information is added to the transcluded document!

Discussion Forum: In electronic written communication (especially in electronic discussion forums like bulletin boards or news groups etc.) many new postings refer to previous postings. These postings are (usually) immutable. In electronic journals or libraries this kind of communication is also very common (e.g. 'letters to the editor' often refer to parts of previous published articles). Referring to parts of an article should be possible by simply mark the fragment to be included in the new document or posting. Readers of the new article can be sure that the referenced part is not modified by the author of the new article. Also readers of the 'old' (i.e. transcluded) document are able to see which parts are discussed later in some other posting. On the other hand it is very likely, if the system is used in the right way and uses are used to it, that questions will be asked not more than once. A much more efficient communication will be the result when using transclusions in this kind of application. Time will also be saved because of additional structure in discussion threads.

Course Material: Very often existing course material is simply a collection of (unidirectional) links to related material stored on different server systems. Unfortunately it is not possible with current web technology to link to parts of existing material and to create links back to the collection. A very simple example will clarify this situation: imagine a single very important paragraph in an article of 100 pages. The author of the course material may add a link exactly to the paragraph, but can't include a link in the (remote stored) material *back to* the collection! Therefore the author of the collection has to cut-and-paste paragraphs of existing material into new documents. The original contexts of these paragraphs are lost.

These three applications showed the need for transclusions. Systems supporting transclusions will have a great impact in many areas. Proper use of transclusions and interoperability between these systems may help authors and readers of content stored on the web.

Requirements and Implementation

In scientific publishing the usage of quotes and references are very common. Readers of those publications have to do a lot of work if they want to explore a reference and want to find the original context of a quote used in another document. Systems like ResearchIndex ([NEC Research Institute, 2002]) are able to answer questions like “Where (i.e. on which page) is this reference used?” but it is not possible to explore the original context of the reused quote! Therefore we are motivated to implement this feature in the Journal of Universal Computer Science ([J.UCS, 2002]). In the publishing system used for serving J.UCS content is offered in several electronic formats to the readers. At the time of writing these formats are HTML, PostScript and also PDF. We consider the implementation of XML formatted documents and also eBook-formatted content.

In a first implementation we are going to support HTML documents (for both, source and destination document format) with additional hyperlinking information. The source and destination document(s) will be stored in a Hyperwave Information Server (HIS, see e.g. [Hyperwave, 2001] or [Maurer, 1996]), where it is possible to add link objects *without* being the owner of a document. Please note that inserting hyperlinks to HTML documents implies editing of the document itself. Note also that a link in HTML is unidirectional! A successor of the implementation will also consider other document formats (like PDF) and other source systems (like ordinary HTTP-server systems). Let us now sketch ideas related to the involved participants and systems:

Author of the original document: If parts of the document are reused in other documents this participant may want to be notified by the system. If the author wants to change a transcluded part of the document, the related parts should be marked. Note that in the original design of Xanadu (see e.g. [Nelson, 1995]), a “write once – read many” system was considered. Therefore it was prohibited to remove documents. Editing of documents was allowed *only* if the old version was still available and accessible by the system.

Author of a document using transclusions: It must be easy for the author who is using transclusions to determine the fragment of a document to be transcluded. Therefore easy to use tools must be implemented supporting the user in the creation of a new document. In a first implementation a simple JavaScript-based implementation will be used. While there are many drawbacks of this approach the user must not install a separate software and a prototype implementation will be available very soon. The selected part will be represented by links following the XPointer-specification (e.g. [DeRose et al., 2001b], [DeRose et al., 2001a]), therefore the selected fragment may be defined in many ways (e.g. use the 5th paragraph, or use the first paragraph after the 3rd heading etc.). The links will be stored using objects on the server system.

System, where the transcluded document is stored: The system must be able to serve parts of a document. If documents are mutable, a version control system must be integrated to make old versions of the document accessible. Additionally, the system must provide the reader of the document (depending on personal preferences) with navigation facilities which indicates transcluded parts. If the author of a transcluded document edit some part of the document, the system is responsible for informing other authors using parts of the document about the change. Please note that bidirectional links are an absolutely necessary requirement for the system to perform this task. In the first implementation we consider a HIS as server system because of the many advantages over ordinary web-systems without link management and bidirectional link concept.

System, where compound documents are stored: This system is responsible for assembling the compound document. Navigation facilities must be included by the system to support the reader of the document. If the transcluded part has changed also the reader may be informed about this change. The author of the document has to define whether the user should be informed or not. In our first implementation also this server system will be a HIS.

Reader of assembled documents: It must be easy for the user to notice the reused parts of a

document and to jump immediately to the original context of the reused parts.

Since there are many issues to consider, the first implementation will be very limited in terms of supported document formats and server systems. Nevertheless, future extensions are considered in the design of the software.

Conclusion and Future Work

This paper has shown that transclusions have many advantages for the web community and that there is a need for an easy to use system. Readers of a compound document have the possibility to *explore the original context* of the used quote. Authors of compound documents are able to easily *reuse existing material* and do not have to update reused document fragments. Finally, authors of transcluded documents know *which parts are used* by other authors due to the bidirectional behavior of the used link mechanism. Transclusions have many advantages over conventional cut-and-paste reuse of content.

Future work is addressed to a first implementation in the Journal of Universal Computer Science. Thereafter usability tests and user responses will be evaluated. After successful tests authors of the journal will be able to use transclusions of articles published in J.UCS.

References

- [DeRose et al., 2001a] DeRose, S., Maler, E., and Daniel Jr., R. (2001a). XML Pointer Language (XPointer) Version 1.0. available online <http://www.w3.org/TR/2001/CR-xptr-20010911> .
- [DeRose et al., 2001b] DeRose, S., Maler, E., and Orchard, D. (2001b). XML Linking Language (xlink) Version 1.0. available online <http://www.w3.org/TR/xlink> .
- [Francisco-Revilla et al., 2001] Francisco-Revilla, L., Shipman, F., Furuta, R., Karadkar, U., and Arora, A. (2001). Managing change on the web. In *Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries*, pages 67–76. ACM Press.
- [Hyperwave, 2001] Hyperwave (2001). Hyperwave Informatino Server. <http://www.hyperwave.com>.
- [J.UCS, 2002] J.UCS (2002). Journal of Universal Computer Science. <http://www.jucs.org> .
- [Krottmaier and Maurer, 2001] Krottmaier, H. and Maurer, H. (2001). Transclusions in the 21st Century. *Journal of Universal Computer Science*, 7(12):1125–1136. http://www.jucs.org/jucs_7_12/transclusions_in_the_21st.
- [Maurer, 1996] Maurer, H. (1996). now Hyperwave — The Next Generation Web Solution.
- [NEC Research Institute, 2002] NEC Research Institute (2002). Researchindex. <http://www.researchindex.com> .
- [Nelson, 1995] Nelson, T. (1995). The Heart of Connection: Hypermedia Unified Transclusion. *Communications of the ACM*, 38:31–33.