

[Subscribe to List](#) [View Past Issues](#) [RSS](#)

translate

[Email not displaying correctly? View it in your browser.](#)

### Center Spotlight



Are you an alumnus of RIT or CIAS? Do you work in the imaging or graphic arts industries? Then why not reconnect with RIT and fellow alumni? Look below to see some upcoming CIAS alumni events!

**CIAS SPAS Photo Journalism Alumni Reception** (*Washington, DC*) - October 1  
[Register Now!](#)

**CIAS Graph Expo Alumni Reception** (*Chicago, IL*) - October 4  
[Register Now!](#)

**RIT Brick City Homecoming** (*Rochester, NY*) - October 15-17  
*Events include Brick City Homecoming comedian Craig Ferguson and RIT Men's Hockey vs. UMass Lowell!*  
[Register Now!](#)

## Transclusion of Dynamic Document Fragments - Part A

This month's research summary is taken from an RIT School of Print Media graduate thesis entitled *Transclusion of Document Fragments from Dynamic Text*, by Manu Choudhury. In order to provide an in-depth look at this work, we will cover the thesis in two parts. This month, we will begin with a summary of the introduction and pertinent sections of the literature review. Next month, we will finish with a look at the research objective, methodology, and research results.

### Introduction

In order to quote part of a document, people usually copy and paste the existing data into their new document (Krottmaier & Maurer, 2001). This results in a loss of connection between the source document and the destination document, and thus propagation of ideas from one document to the other is not effective (Nelson, 2007).

The reader of the new document might not be able to comprehend the quoted text and its importance in the new document unless the reader has access to the original document to understand the context in which the quoted text was originally used.

Many researchers have proposed and implemented solutions to handle the issue of lost context. The following sections will talk about some of the suggested solutions for digital/Web content, the technologies used, and some of the issues that still need to be handled to achieve re-use of content/ideas from dynamic

To find out about more alumni events, visit the [RIT Alumni Association website!](#)

---



### Print in the Mix

Print in the Mix is "a unique site demonstrating the role of print as a viable information medium in the marketing mix." This **free** resource is published by the Printing Industry Center.

#### Sample Fast Fact:

*North American consumers in the 18-34 year-old demographic prefer to learn about marketing offers via postal mail and newspapers rather than online sources such as social media platforms, according to survey research from ICOM, a division of Epsilon Targeting.*

[Read the full fast fact here.](#)

Have you visited **Print in the Mix** yet? Find out how this site can help you 'make the case' for print!

[printinthemix.rit.edu](http://printinthemix.rit.edu)

Funded by The Print Council

---



### Industry Education & Training

RIT provides training in both

documents.

### What is Transclusion?

Transclusion is a form of smart inclusion that allows an author to include other authors' work, not by duplicating the original work, but by including a reference to it (Nelson, 1995). For example:

A document (Document A) exists with content that reads, "The quick brown fox jumps over the lazy dog is a sentence that uses every character of the English alphabet."

Now, an author wishes to create another document in which they quote a portion of Document A. They want the new document to read, "The quick brown fox jumps over the lazy dog is a very interesting sentence."

In order to do so, the author can copy the required portion from Document A and then paste it in the new document and append the rest of the content to it. In this case the connection between this new document and Document A is lost, and there is a duplication of data.

The author can also do the same thing using the concept of transclusion. The author can refer to the quoted text by writing: **<transclude text from source=DocumentA, startindex=1, number\_of\_characters=43> is a very interesting sentence.**

This results in a new document with content that reads, "The quick brown fox jumps over the lazy dog is a very interesting sentence." This also ensures that the connection between the two documents is maintained and the context is not lost (Nelson, 2007).

For this to work, an editor or viewer must identify that the document contains some transcluding text and can retrieve the required text from the source document. Features, such as highlighting of the transcluding text and side-by-side comparison of the original document and the transcluded document, can also be added to enable authors to present the information comprehensively (Kolbitsch & Maurer, 2006).

Since the new document does not contain the actual text, but rather a reference to the quoted text, any changes in the source document are automatically reflected in the destination document, thereby ensuring that the new document always contains the latest information (Nelson, 2007).

technologies using world renowned instructors, comprehensive prepress and press labs, and state-of-the-art imaging facilities.

Our programs and services can help your organization make the most profitable use of new technologies, enhance productivity, boost customer satisfaction and produce a healthy bottom line.

#### **Upcoming industry education programs include:**

**September 27 - October 1**

[Orientation to the Graphic Arts](#)

**October 13 - 15**

[Color Printing Fundamentals](#)

**October 20 - 22**

[Lithographic Troubleshooting](#)

**October 26 - 28**

[Digital Printing Bootcamp](#)

For more information on these and other programs, or to register for any of these programs, visit

[\*\*printlab.rit.edu\*\*](http://printlab.rit.edu)

#### **About the eReview**

The *eReview* is a monthly publication of the Printing Industry Center at RIT for registered Affiliate companies. Articles are also published in the quarterly printed publication

## **Existing Implementations of Transclusion**

### **Using Basic HTML Tags**

The most crucial part for implementation of transclusion is a way to refer to a resource. Any page/object has a unique address associated with it, called URLs or Universal Resource Locators. Thus, any page/object can be uniquely referred to using these URLs (Berners-Lee, 1994). Using the HTML specifications, there are two tags that are used for implementation of transclusion.

**<iframe>**: By using this tag in an HTML page, the author can make reference to remote HTML documents that are shown in an inline frame. When the page is viewed, the browser fetches the source document from the URL specified in the 'src' attribute and displays the entire content of the source page (Raggett, 1999). The problem with this is that a reference to a document fragment cannot be made (Kolbitsch & Maurer, 2006).

**<object>**: Unlike the iframe tag, where only HTML pages could have been included, the object tag can be used to include objects such as images, audio, videos, Java applets, ActiveX, PDF, and Flash animations (Raggett, 1999). Again, it is not possible to include a portion of any object, only the whole object.

These tags do reflect an idea of transclusion, though very limited in nature. Moreover, these tags do not support transclusion of text.

An amendment to the HTML specification was suggested, which proposed that a new tag <text> be added. The main attributes of this tag would be the source URL, the starting index, and the length of the text to be transcluded. This would also enable transclusion of document fragments (Pam, 1997). Although the proposal seems quite promising for the implementation of transclusion on the Web, it has not been accepted, and no browsers have implemented this feature (Kolbitsch & Maurer, 2006).

### **Using Document Object Model (DOM) Nodes**

Josef Kolbitsch and Hermann Maurer designed a prototype to implement transclusion in a HTML-based environment. In their design, transclusion of document fragments was handled very well, and fragments could be of any size from a single character to the entire content of a page (Kolbitsch & Maurer, 2006).

DOM is a language-neutral interface that defines a standard way

*PrintReview.*

**[Forward this eReview.](#)**

---

## Subscriptions

You are receiving this newsletter because you registered as an Affiliate of the Printing Industry Center.

**[Update your profile.](#)**

**[Unsubscribe from this list.](#)**

---

## Contact the Center

### Director:

[Patricia Sorce](#)

### Communications Coordinator:

[Ashley Walker](#)

*(Web site, publications, general info)*

### Mailing Address:

RIT Printing Industry Center  
College of Imaging Arts & Sciences  
Rochester Institute of Technology  
55 Lomb Memorial Dr  
Rochester, NY 14623

**Ph:** 585-475-2733

**Fax:** 585-475-7279

**Web:** <http://print.rit.edu>

**Email:** [printing@rit.edu](mailto:printing@rit.edu)

**Twitter:** [RITprintcenter](#)

---

## About the Center

of accessing and updating the content, structure, and style of a document. DOM presents a document in a tree structure. ***A full discussion of the DOM structure and the tags and technologies used may be found [in the complete thesis.](#)***

The other interesting feature of their implementation was that the technologies they used included HTML, DOM, Java Script, and HTTP. These are supported by every browser on the Web; thus their implementation requires no special software or plug-ins.

This implementation takes care of document fragments very well, but when the source page has changed, it just flags a warning. Thus, this implementation does not really transclude text from dynamic pages.

### Using Asynchronous JavaScript and XML (AJAX)

AJAX is a Web development technique that is used to create interactive applications on the Web. The technique allows for any page on the client side to be updated without the need to reload it manually. (Garrett, 2005).

In AJAX, reference to a document fragment of a page is done using DOM. The start point and the end point are located and represented in an identical way, as Kolbitsch and Maurer suggested.

AJAX enables retrieval of transcluded text more interactively. The document containing the transcluded text updates itself as the source page changes without the user having to reload the document.

Although AJAX with DOM does not really help to transclude text from a dynamic page, if referencing from a dynamic page is possible, such that the reference automatically updates when the source page changes, AJAX would make the transcluded document very interactive.

### Using Extensible Markup Language (XML)

XML is a specification that defines a way to encode documents for the Web. XML provides a number of features that can be used to create documents composed of fragments from multiple sources (W3C, 2003). The granularity, i.e., the minimum amount of text that can be transcluded using this method, is the content of any tag. It is not possible to address any fragment inside any tag using this method.

Dedicated to the study of major business environment influences in the printing industry precipitated by new technologies and societal changes, the Printing Industry Center at RIT addresses the concerns of the printing industry through educational outreach and research initiatives.

Support for the Center comes from:

[Sloan Foundation](#)

[Rochester Institute of Technology](#)

[Adobe](#)

[Avery Dennison](#)

[Democrat and Chronicle](#)

[Hewlett-Packard](#)

[NewPage Corporation](#)

[NPES](#)

[Scripps Howard Foundation](#)

[VIGC](#)

[Xerox Corporation](#)

The most interesting feature of this implementation is that it can handle recursive transclusion. That is to say, if the source document itself contains some transcluded text, that transcluded text would also be fetched along with the desired contents from the source page (Di Iorio & Lumley, 2009).

Thus, XML provides very powerful tools for the implementation of transclusion. Moreover, it does not require any special plug-ins to be installed, as XML is supported by all major browsers.

Since XML refers to different nodes, even if some text is added or modified in that node, the changes are reflected in the transcluded document. Therefore, it does offer solutions to dynamic source pages, but the solution is very limited in nature since it only works if the text is modified inside the node. If another node were added to the page, then this implementation would fail to refer to the original text (Di Iorio & Lumley, 2009).

### **Purple Numbers**

Purple numbers is a way to uniquely identify an element/tag/node in a HTML/XML document. The basic idea is that a unique number is assigned to any text node, and this number is not based on the content or its hierarchical location in the DOM tree (Kim, 2003).

Purple numbers, also termed as Node Identifiers (NID), are unique to the HTML page. If the node's contents were modified, or its hierarchical position changed, the number would still be the same, and thus could be used to refer to the node (Gupta, 2009). This would enable transclusion of document fragments from a dynamic source page.

Adding these unique identifiers must be done by the author of the original page. If the source page does not contain these unique identifiers, transcluding text from those documents is not possible.

Although the concept of Purple numbers is very promising, especially for the implementation of transclusion, there are very few HTML pages on the Web that use this concept.

### **Ideal Implementation of Transclusion**

According to Ted Nelson, the major flaw in the current state of the World Wide Web is that links are unidirectional. This makes the references from the user page to the source page unidirectional as well. Thus, if the source page changes, the user document is not notified of those changes, and thus the references are not useful.

Consider a situation where all these references are bi-directional. If this were possible, the source page would have knowledge about all the other user pages that contain quoted text from the source page. It would also have knowledge of the start and the end of each selection.

With this knowledge, if the source page changed, it would automatically notify all the user pages, and provide them with a new set of transclusion parameters that would identify the selection users had initially made (Nelson, 2007).

This ideal state of Web would make it very easy for people to exchange and re-use ideas, which, in turn, would enable them to establish an effective, collaborative content- generation platform.

## Conclusion

Next month, we will finish with a look at the research objective, methodology, and research results from this thesis.

## References

- Berners-Lee, T. (1994). *Universal Resource Identifiers in WWW. A unified Syntax for the Expression of Names and Addresses of Objects on the Network as Used in the World-Wide-Web*. Retrieved October, 2009, from <http://w3.org/Addressing/rfc1630.txt>
- Di Iorio, A. & Lumley, J. (2009). From XML inclusions to XML transclusions. In *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia* (Torino, Italy, June 29 - July 01, 2009). HT '09. ACM, New York, NY.
- Kim, E. E. (2003). *Purple website*. Retrieved October, 2009 from <http://www.eekim.com/software/purple/purple.html>
- Garrett, J. (2005). *Ajax: A New Approach to Web Applications*. Retrieved October, 2009 from <http://www.adaptivepath.com/ideas/essays/archives/000385.php>
- Gupta, V. (2009). *Fine-grained addressability to support large-scale collaborative document development*. M.S. dissertation, Northeastern University.
- Kolbitsch, J. & Maurer, H. (2006, February). Transclusion in an HTML-Based Environment. *Journal of Computing and Information Technology*, 161-174.
- Krottmaier, H. & Maurer, H. (2001). Transclusions in the 21st Century. *Journal of Universal Computer Science*, 7(12), 1125-1136.
- Nelson, T. (1995, January). "The heart of connection: Hypermedia unified by transclusion". *Communications of the ACM*, 38(8), 31-33.
- Nelson, T. (2007). *Transliteration: A Humanist Format for Re-Usable Documents and Media*. Retrieved August, 2009 from <http://transliteration.org/>
- Pam, A. (1997). *Fine Grained Transclusion in the Hypertext Markup*

*Language*. Retrieved September, 2009 from <http://www.xanadu.com.au/archive/draft-pam-html-fine-trans-00.txt>

- Raggett, D. (1999). *HTML 4.01 Specification*. Retrieved October, 2009, from <http://w3.org/TR/html4/>
- W3C. (2003). *Extensible Markup Language (XML)*. Retrieved October, 2009 from <http://www.w3.org/XML/>

---

## Research Publications

To read about this research in detail, download the thesis from:  
<https://ritdml.rit.edu/handle/1850/12419>

Research publications of the Center are available at: <http://print.rit.edu/research/index>

Copyright (C) 2009 Printing Industry Center at RIT. All rights reserved.



Sent to <<Email Address>> — [why did I get this?](#)

[unsubscribe from this list](#) | [update subscription preferences](#)

Printing Industry Center at RIT · 55 Lomb Memorial Drive · CIAS Dean's Office · Rochester, NY 14623