

---

## Implementing New Knowledge Environments (INKE)

Publications

---

Ubiquitous Text Analysis

Geoffrey Rockwell, Stéfan Sinclair, Stan Ruecker, & Peter Organisciak

20 December 2010

© 2010 Rockwell et al. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

<https://creativecommons.org/licenses/by-nc-sa/3.0/>

This article was originally published at:

<https://journals.tdl.org/paj/index.php/paj/article/view/13/57>

---

Citation for this paper:

Rockwell, G., Sinclair, S., Ruecker, S., & Organisciak, P. (2010). Ubiquitous text analysis. *Poetess Archive Journal*, 2(1).

## Ubiquitous Text Analysis<sup>1</sup>

© Geoffrey Rockwell 2010

© Stéfan Sinclair 2010

© Stan Ruecker 2010

© Peter Organisciak 2010

## 0. Introduction

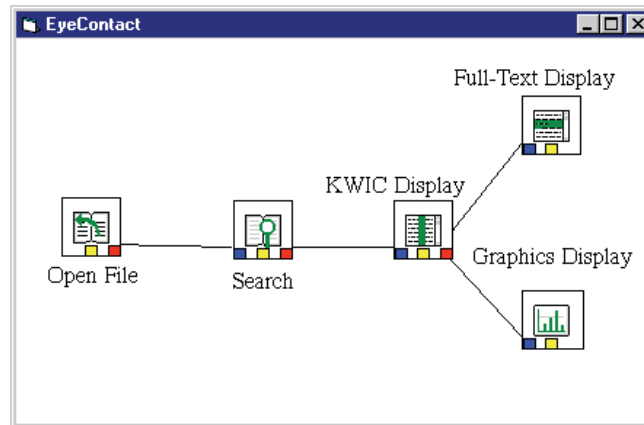
One of the problems facing e-text content publishers and text analysis tool developers is how to connect the appropriate tools with content. This problem was noted at the "Tools for Data-Driven Scholarship" (TDDS) meeting and concerns projects such as [NINES](#) that want to make analytics available.<sup>2</sup> To paraphrase the conclusions of the TDDS meeting we have three related problems,

1. Tool Connections - tools tend not to work well with each other.
2. Connecting Content and Tools - the content collections being developed don't work well with tools developed by others.
3. Visibility of Tools – tools that could work with content are hard to discover.

These conclusions confirm early usability studies of the TAPoR portal and related tools. They suggest that having users think first about tools and then about texts is the reverse the normal order of research practice.<sup>3</sup> Users do not think of tools to which they then bring texts, but instead they like to look at texts and explore what they see with tools. For this reason we have been experimenting with ubiquitous tools that can be embedded in other content sites. This paper describes the history of these experiments leading up to the Voyer Tools where results from analytical studies can be quoted in online research papers and which is being used in The Poetess Archive. To this end, the paper we will do four things:

- We will present the usability case for privileging texts over tools and presenting tools on the side, so to speak.
- This will slide into a review of various visual interface models developed by the TAPoR project and related projects for embedding tools into content interfaces.
- We will review the challenges of connecting tools reliably to content, in this case connecting with The Poetess Archive.
- We will conclude by discussing technical and open source solutions to the connection issues.

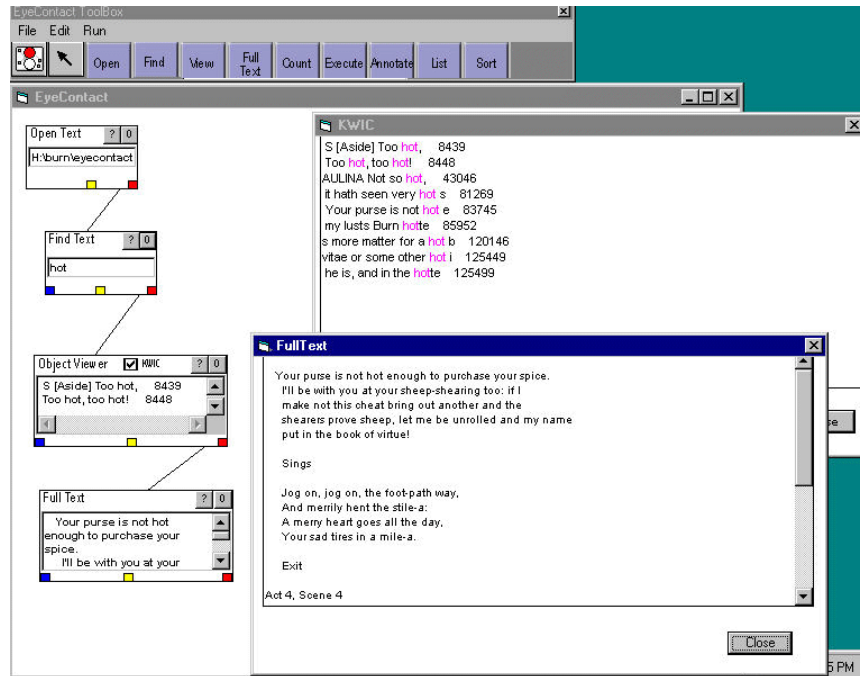
## 1. Usability Issues in Visual Analytical Interfaces



Eye-ConTact Map Display of Process

Humanists are used to looking at documents; they are not used to treating documents as tokens for processing by tools. Interfaces for text analysis like that prototyped in the Eye-ConTact project that John Bradley and Geoffrey Rockwell worked on in the late 1990s present a visual programming environment where processes are connected into a "pipe and flow" diagram, are generally too abstract for most humanists.<sup>4</sup> The Eye-ConTact interface is based on a common scientific visualization interface model where users drag out icons for processes and output and then connect the processes with "rubber-band" pipes that indicate the "flow" of data (this is similar to the Unix piping model, as well as as other visual interfaces like Yahoo! Pipes). In the process map above the Open File process opens the selected file and pipes the data to a Search process. From the Search Process only those lines that match the search criteria flow to the KWIC (Key Word In Context) Display. When the user runs the process mapped out, they can then open displays and inspect the results. The KWIC Display would show a standard KWIC of the search hits.

One immediate disadvantage of this visual programming interface model was that the user couldn't see what the settings were to the processes and they couldn't easily inspect the results. They were, in effect, distanced from the text. In a later version of Eye-ConTact we added information to the icons, turning them into miniature panels that could show key parameters like what pattern was searched for, or a sample of the results.

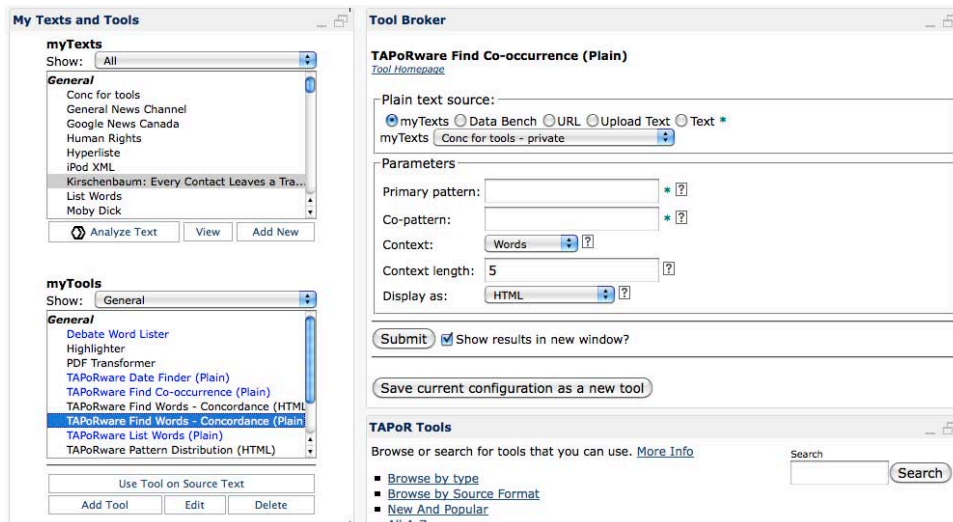


Eye-ConTact 2 Map with Mini-Panels and Expanded Displays

This interface, however still privileged the process over the text. The visualization doesn't show much of the text itself (unless you open the results windows); it shows the logic of the program being run on the text. For this reason it is likely to appeal to someone who is developing text processes, but not to someone interpreting the text. It is for this reason that SEASR, which also has a (much more sophisticated) visual programming interface, is “designed to enable digital humanities developers to rapidly design, build, and share software applications that support research and collaboration.”<sup>5</sup> The idea is that a visual programming interface is suitable for the humanist developer who can then package the data-exploration process with a different interface; it is this second interface which will offer content to users along with the data-mining or analysis that the humanist developer has already created.



TAPoR: myTexts where the user can define texts for the Library



TAPoR: Workbench where user can choose a text and then a tool to use on it

The TAPoR (Text Analysis Portal for Research) workbench model is arguably less abstract: it does not separate humanists into developers and users and so does not render quite as distant from each other algorithms and results, developers and users, methods and results.<sup>6</sup> In this model users define texts on which they want to operate in the myTexts page, define their favorite tools for the Workbench, and then in the Workbench they run tools on texts by clicking on first a text and then the tool. This launches a panel where they can set the parameters for the tool and submit the job. Results can then be saved to a Data Bench where they can be treated as a text available as input for the next job. Alas, this still effectively hides texts from the users.

Text Analysis Portal for Research  
TAPoR

Analyze Text

Texts  
Human Rights

Tools  
Compatible tools:  
TAPoRware List Words (XML)  
All tools:  
-- Select a Tool --

TAPoR Tool Broker  
TAPoRware List Words (XML)  
Tool Homepage

Parameters

Elements // ?  
List options All words except stop list below \* ?  
Term glasgow ?  
Sort criteria: By frequency \* ?  
Display as HTML \* ?

Submit  Show results in new window?

Universal Declaration of Human Rights  
Preamble

Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world,

Whereas disregard and contempt for human rights have resulted in barbarous acts which have outraged the conscience of mankind, and the advent of a world in which human beings shall enjoy freedom of speech and belief and freedom from fear and want has been proclaimed as the highest aspiration of the common people,

Whereas it is essential, if man is not to be compelled to have recourse, as a last resort, to rebellion against tyranny and oppression, that human rights should be protected by the rule of law,

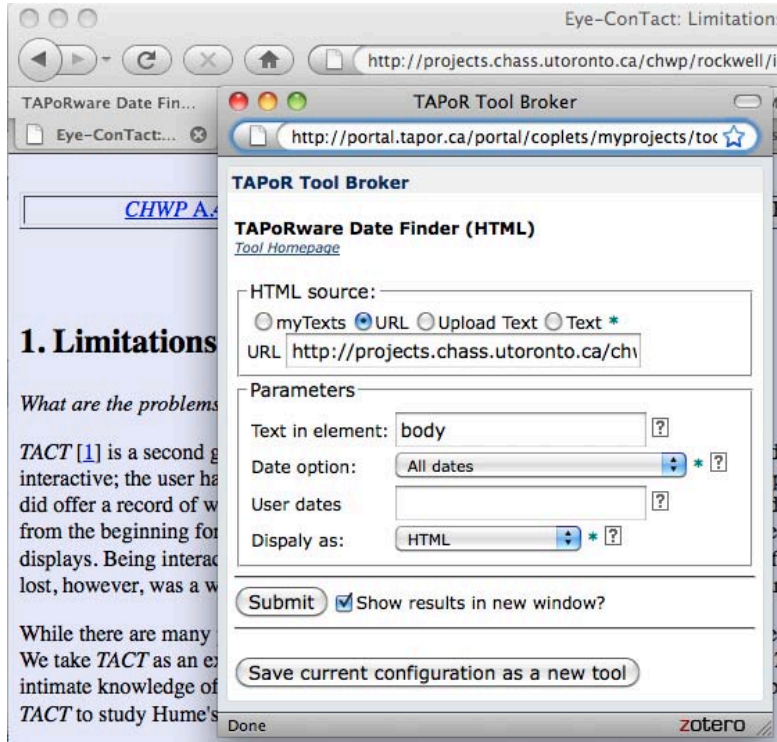
Whereas it is essential to promote the development of friendly relations between nations,

Whereas the peoples of the United Nations have in the Charter reaffirmed

TAPoR: Analyze This interface where you can see the text on the right while choosing texts

Usability interviews conducted by Wendy Duff at the University of Toronto Faculty of Information to help improve the portal interface led to speculation that a workbench was not how humanists thought of their research. Duff's interviews made clear how text-centric humanists are and led to the first of many interface experiments. We began by adding an "Analyze This" view that presents the text in one frame on the right with appropriate tools in a separate frame on the same screen. This solution, however, is only useful where a user has gone to the trouble to set up an account and define texts to study. While the TAPoR portal has the features of a bibliographic management tool it couldn't (and shouldn't) compete with specialized tools where users are likely to manage their texts, like EndNote or Zotero. We were led therefore to relinquish the idea that all users might run tools on pre-defined texts and to pursue the strategy of embedding tools into environments that already have full-text views, where there is a lot of content already published dynamically and a tool panel can be added to enhance reading. We call this ubiquitous text analysis and the rest of this paper will demonstrate a sequence of experiments.

## 2. Experiments in Ubiquitous Interfaces



TAPoR bookmarklet and the TAPoR Tool Broker window it launches

How might tools look like if they were embedded in the user's environment or the text environment? The TAPoR portal, from the beginning, was envisaged as a broker for web service tools. One of the features we were able to provide, because of the way tools are registered on the portal, was a "Detailed Info" page about each tool that provided different ways of using the tool. One way of using the tool was a "bookmarklet" that could be dragged to the Bookmarks Bar of your browser. Clicking on the bookmarklet would open a window with the tool panel already set up to analyze whatever web page you were looking at. With bookmarklets you wouldn't need to go back to the portal, you could use the tools you like in your browser. The bookmarklet approach however is still awkward. You had to know where the bookmarklets were and where to drag them and then you would end up with a bookmarklet for each tool. A better approach would be to build a Firefox or Chrome plug-in that can offer a list of tools and the parameters right in the browser.

TAPoRware Find Co-occurrence (Plain)	
<b>Tool Details</b>	
Name	TAPoRware Find Co-occurrence (Plain)
Description	Co-occurrence tool looks for two words a certain distance apart from one another. By entering a primary and secondary pattern, TAPoR will search the document for anywhere that the two patterns are within the user-specified limits of words, sentences, or lines.   <b>Note</b>: If your select context of "Words" with long context length or you select other context: Lines, Sentences or Paragraph, it is very likely that the specified pattern appears in each text of concordance more than once. If this occurs, the words of collocation will be counted more than once as well. So the counts of collocates are not accurate. For the same reason, the zScore values are not accurate. We will find way to fix this later.  
Website	http://tada.mcmaster.ca/Main/TAPoRwarePlainCoOccurrence
Type	Search
Endpoint URL	http://taporware.mcmaster.ca:9982
Service URI	http://taporware.mcmaster.ca/~taporware/webservice
Service Method	find_Cooccurrence_Plain
Soap Action	
Default Stylesheet	quickFixForEscaping.xsl
Text Compatibility	plain-text
AnalyzeThis HTML Fragment	Static URL
Fragment	<pre>&lt;form action="http://portal.tapor.ca/portal/launchTool" name="launchTool" method="get" target="_blank"&gt;   &lt;input type="hidden" name="toolName"</pre>

Detailed Information Screen for a Tool (see HTML fragment at bottom)

The Detailed Info screen also had code and HTML fragments that could be put by programmers into another web site. We generated the code that a web developer could use in their site without asking us. While this hasn't been used extensively to our knowledge it has been used by at least one journal, the Digital Humanities Quarterly.<sup>7</sup>

Summer 2010  
Volume 4 Number 1

[Summer 2010: v4 n1](#) | [XML](#) | [Print Article](#) | **Taporware Tools**

**The Landscape of Digital Humanities**

[Patrik Svensson](#) <patrik\_dot\_svensson\_at\_humlab\_dot\_umu\_dot\_se>, Humlab, Umea University

**Abstract**

The digital humanities is increasingly becoming a "buzzword", and there is more and more talk about a broadly conceived, inclusive digital humanities. The field is expanding and at the same time being negotiated, and this article explores the idea of a broadly conceived landscape of digital humanities in some depth. It is argued that awareness across this landscape is important to the future of the field. The study starts out from typologies of digital humanities, a "flythrough" of the landscape, and a discussion of what being a digital humanist entails. The second part is an exploration of four concrete encounters: ACTLab at University of Texas at Austin, the Humanities Arts Science Technology Advanced Collaboratory (HASTAC), the Humanities Computing Program at the University of Alberta, and Internet Studies. In the third part of the article, it is suggested that a model based on paradigmatic modes of engagement between the humanities and information technology can help chart and understand the digital humanities. The modes of engagement analyzed are technology as a tool, study object, expressive medium, exploratory laboratory and activist venue.

DHQ: Taporware Tools

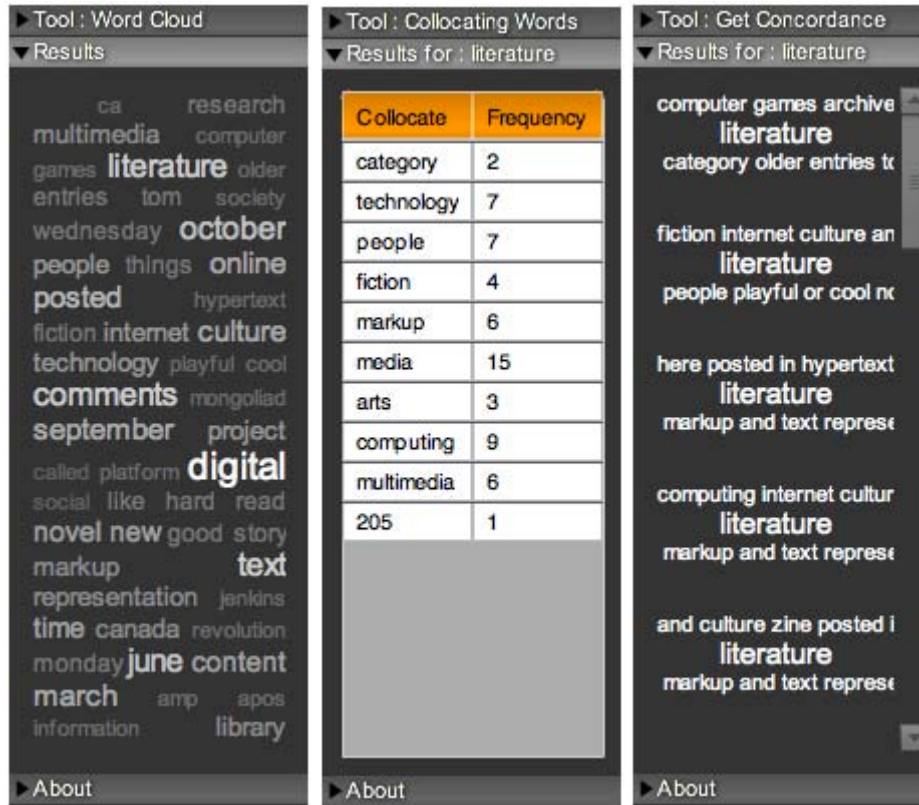
It then occurred to us that we could provide custom HTML for content providers who wanted to embed something more functional. This was inspired by the Reading Tools provided in the Open Journal System of the Public Knowledge project.<sup>8</sup> The idea was to

provide code (HTML, JavaScript and CSS) that a developer could adapt to build their own tool bar to integrate into a web site.

About the Project	
Contributors	
UBC Press	
Related Websites	
Contact Us	
<b>Analysis Tool Bar</b>	
List Words	In my analysis I am not simply these images (pro-whaling pro images are elevated to the sta seduction with which these im manipulate and subvert. One c or media is mediation (a proce simultaneously enabling and c
Concordance	
Pattern: <input type="text"/>	The photos that I analyze are
Collocation	the seduction quality (and fant disaster and futuristic hopes. )
Pattern: <input type="text"/>	nature, surge into view, viewer: catastrophe, disaster, and enc possibility. Yet in environment catastrophe is hope, a forward future. Situated at the border c
Co-occurrence	the seen and the unseen, the global awareness that not only
Pattern: <input type="text"/>	
Co-pattern: <input type="text"/>	
Summarizer	Greenpeace may have been at became known as the anti-wh: groups such as environmental Welfare Institute (AWI), World (WSPA), Whale and Dolphin C Campaign Whale, Internationa World Wildlife Fund (WWF), ar States, have followed. Scientis become the allies of these gro conscious global coalition that And in this process, they cast saviours but postulate them, t world.
Tokenizer	
Date Finder	
Principal Components Analysis	
Pattern Distribution	
Pattern: <input type="text" value="global"/>	
Visual Collocator	
Pattern: <input type="text"/>	
Weighted Centroid	
T-Bar from TAPoR	

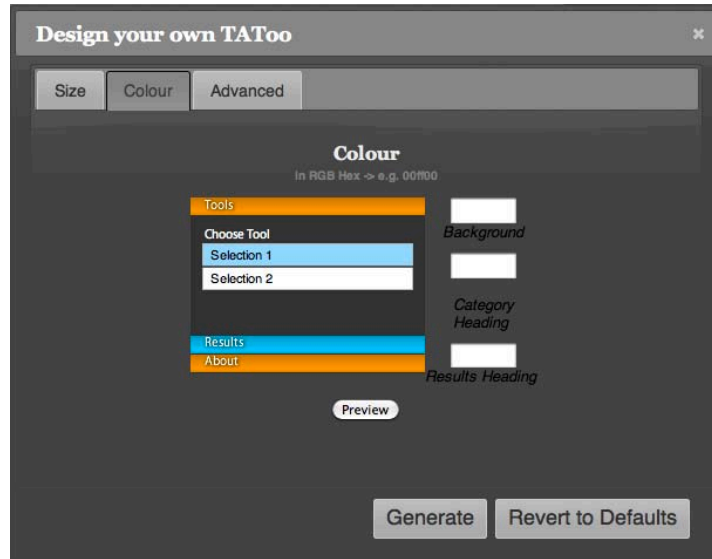
Globalization Toolbar expanded next to text on which it can operate

To test this idea we built a Toolbar for the Globalization and Autonomy Compendium, a collection of research summaries and working papers that were gathered around a SSHRC supported project.<sup>9</sup> This was our first experiment with an embedded tool bar. The code is a long span of JavaScript, CSS and HTML that is placed in the common template of a range of pages, from research summaries to position papers. The tool bar appears discretely at the bottom of the right hand navigation bar and is collapsible so as not to distract users. This is documented so others can use it, but unfortunately the code tends to conflict with other CSS and JavaScript so it has only been used on a few projects.



TAToo: Word Cloud view, Collocates view, and Concordance view

In order to avoid the problem of lots of conflicting code we then developed a YouTube-inspired Flash application originally called FlashTAT (for Flash Text Analysis Tool), but now called TAToo. TAToo, developed by Peter Organisciak, can be embedded with one <object> tag and, because the interface is handled by the Flash application, does not conflict with existing CSS and JavaScript.<sup>10</sup> This tool also has the virtue that when it loads it shows results immediately, in this case a list of high frequency words, so the user can see those results without making any choices or invoking the tool. The user can play with the results rather than having to decide to run a tool in the first place in order to see anything at all. We believe this is one of the more promising approaches to providing content providers with an easy way to embed tool interface. We have installed it in blogs and it seems generally robust.<sup>11</sup>



TAToo: Parameters editing panel

One of the features we have built into TAToo is a set of parameters that let users change the look of TAToo (size and colour), and change the instance of the Flash object to which it points so that content providers can offer up the pre-configured Flash object needed themselves. The visual parameters allow users to customize it to fit the graphic design of their blogs or web site. This tool is, however, limited to operating on the page it is embedded into. It has the advantage that the analysis runs in the client's browser which means that there is no delay as a query (and text) goes back to our server (which might be down), but the disadvantage that it can only operate on a limited amount of text.



Digital Texts 2.0 plugged into Facebook

Another approach to ubiquitous tools is to experiment with emerging social plug-in architectures. We are convinced that in the long run, especially for student- and faculty-portals (not to mention scholarly-publishing portals) we need to have social tools that users can choose from and include in their personal study space. As research content architectures stabilize it should be easier to design plug-ins for those systems. To test this hypothesis, Stéfan Sinclair and Johnny Rodgers have developed a FaceBook plug-in called Digital Texts 2.0 which gives users a social bibliography in FaceBook accounts.<sup>12</sup> Digital Texts 2.0 (dtext2.org) functions as a stand-alone web application, but it also integrates fully with Facebook, including the broadcast of related news items (added texts, comments, friends, etc.). Users can work with Digital Texts 2.0 in their Facebook space (i.e. in Facebook) or in the dedicated site if they wish more functionality.

**Now, Analyze That: Obama and Wright on Race in America**

An [experiment](#) in [extreme text analysis](#)

[Geoffrey Rockwell](#) and [Stéfan Sinclair](#)

---

Sounds like he talked a hate speech, doesn't it? Now, analyze that. (Wright, NAACP Speech)

<p><b>Introduction</b></p> <p>The <a href="#">news media</a> have Barack Obama and his spiritual father Jeremiah A. Wright Jr. locked in an oedipal drama. Obama the son tries to distance himself from his father-pastor to win the presidency while Wright struggles to continually correct the record while getting attention unlike what he is used to from the pews. Both, in different ways, are trying to tell the media what should be talked about and how. Both want the attention on more substantive issues and, in trying to redirect us, have given moving and important speeches on race and America (by which we mean the USA). Both have been trying to use the attention to redirect us to what "this time we want to talk about", or, to use Wright's blunt phrase, they challenge us directly: "now analyze that"!</p> <p>Of course the media know where the engaging human story is and it is in the age-old conflict of the son and his father, as the son comes of age as a leader.</p> <p>But, what if we took them at their word and looked away from the <a href="#">pulpit-and-pews</a> drama. What if we looked at what their speeches on race as important and tried to "analyze that" looking for the similarities and differences?</p>	<p><b>Resources</b></p> <p><a href="#">Texts Used</a> in this experiment.</p> <p><a href="#">TAPoR Portal</a> for text analysis research.</p> <p><a href="#">Experimental Notes</a>. <a href="#">May 1 &amp; 2</a> are our notes on this experiment.</p> <p><a href="#">Recipes for learning text analysis</a> including a <a href="#">comparison recipe</a>.</p>
---	---

Now Analyze That: top of page

The final model came from an experimental essay, "Now Analyze That" which presents a different way of embedding tools: interactive-tool results are woven right into the prose of the essay, allowing users to reenact the very analyses that led to claims in the essay.<sup>13</sup> Such a model connects not to content providers so much as to researchers at the time when they are writing which therefore presents new challenges to tool developers.

## This Time We Want To Talk

One of the first things we noticed was that Obama uses the word "time" far more often than Wright. (Note 2) In fact, at the climax of Obama's speech, he repeatedly uses the phrase "this time we want to talk". This table shows a [concordance](#) of all the instances of "time" in Obama:

**17 entries found.**

and should be perfected over <b>time</b> . And yet words on
and the reality of their <b>time</b> . This was one of
solve the challenges of our <b>time</b> unless we solve them together
divisive , divisive at a <b>time</b> when we need unity ;
; racially charged at a <b>time</b> when we need to come
and early sixties , a <b>time</b> when segregation was still the
prejudiced , resentment builds over <b>time</b> . Like the anger within

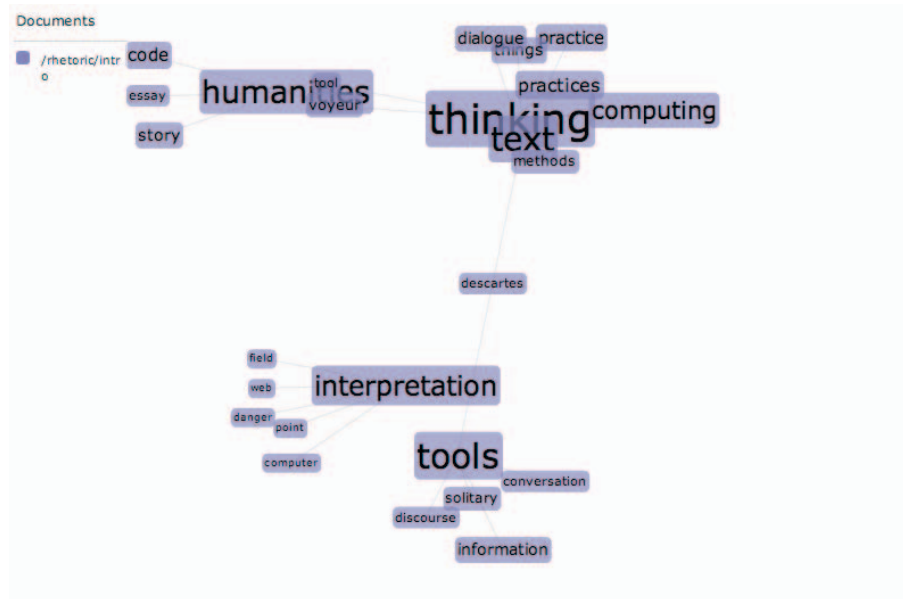
Now Analyze That: example where a concordancing tool is woven into the prose

“Now Analyze That,” the essay, takes its title from something Obama’s former minister Jeremiah Wright said in a speech about race in America. The essay compared Obama’s discourse on race before the election to Wright’s at a time when their differences were being played up in the press. The essay was itself an experiment in using text analysis to study contemporary issues like race. In the experiment Geoffrey Rockwell and Stéfan Sinclair set out to see how easy it would be to go from doing the analysis using our tools to writing up the results in a way that didn’t hide the computer-assisted textual analysis.<sup>14</sup>

To write “Now Analyze That” we used a wiki and customized the HTML code from what the TAPoR portal provides through the Detailed Info screen for each tool. The code was customized in two ways. First, we hard-coded the text to analyze so that if you ran the process you would get results for the text we were making claims about. Second we adapted the code so that we could weave the form needed to invoke a tool right into our prose as the example above shows. The sentence “This table shows a concordance of all all the instances of ‘time’ in Obama” has a field with “time” that you can edit to try other words. To involve the concordancing tool, you click the button “Obama.”

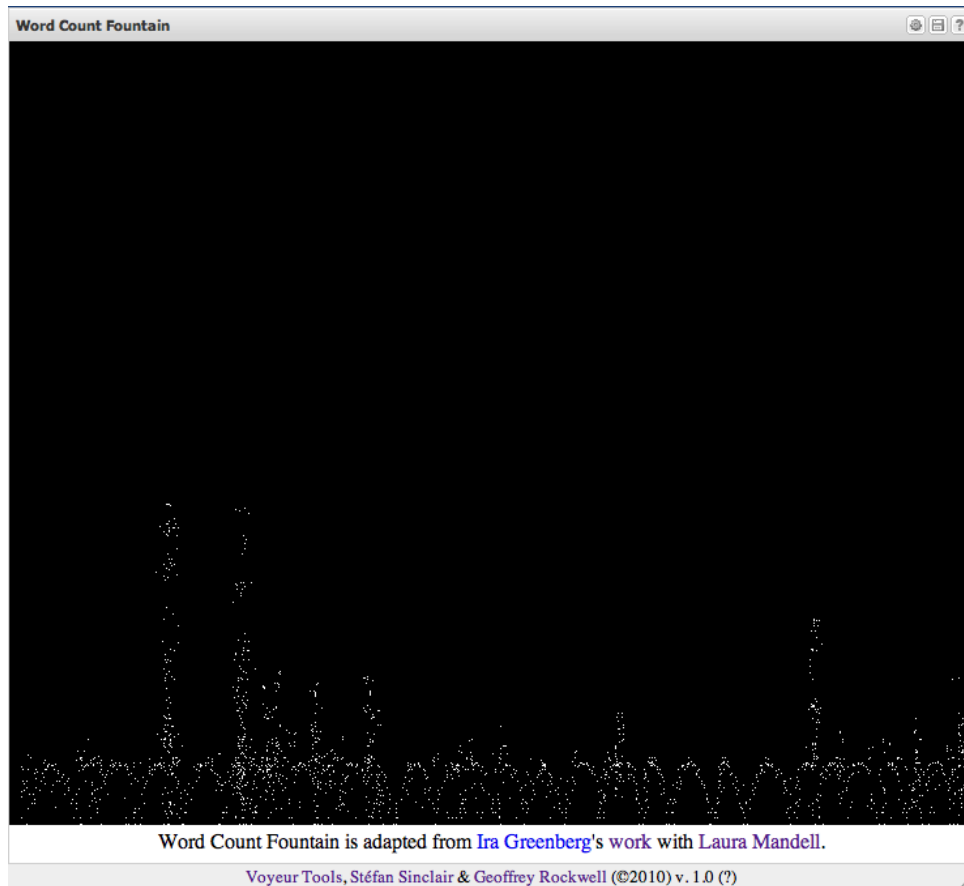
Convinced that some users would want this ability to “quote” tool results right into online papers, we adapted Voyeur Tools so that it could automatically create the code for a researcher that can be placed in their authoring environment if the finished essay is intended to go online.<sup>15</sup> The idea is to support content management systems users are already using like blogging tools or wikis. In the screen below you can see a wordle-like visualization of the high-frequency words in the DH 2010 conference abstracts that was generated by Claire Ross who then placed it in the UCL Center for Digital Humanities blog.<sup>16</sup>





Voyeur Tools: Links tool

Finally, *Voyeur Tools* is programmed so that it is fairly simple for new tools to be added. For The Poetess Archive we have incorporated two visualization tools developed by others into the Voyeur Tools framework – the Word Count Fountain tool, developed by Ira Greenberg and Laura Mandell, is an example of an existing visualization applet that was wrapped as a Voyeur tool so that it can now benefit from the broad infrastructure of Voyeur Tools (facilities for adding documents in various formats and from various locations, interacting with other tools, embedding in remote sites, etc.).



Voyeur Tools: Word Count Fountain

### 3. The Challenge of Connecting Tools to Content

We believe that such embedded tools do, in principle, offer an answer to two of the three issues identified in the Data-Driven Scholarship report:

- These tools are discoverable or visible to users, at least those reading texts within which such tools are embedded. This is obviously not the only way people should discover tools, or what the report authors really meant, but nonetheless embedded tools are becoming discoverable independently of lists of tools that can quickly become outdated. If people find tools where they are working and reading because they are ubiquitous, isn't that a form of discovery?
- Embedded tools demonstrate one way of connecting well with texts, though most of the models above like the TAPoRware tools and TAToo work only with the text on the page in which the tool is embedded. This paper documents a number of different ways we allow tools to be embedded so as to encourage the ubiquitous use of tools.
- These embedded tools do not, however, connect well with other tools unless by other tools we mean the content frameworks from WordPress to Facebook for which the tools were designed in the first place. The TAPoR Data Bench was one attempt to provide a way for data to be moved from tool to tool, but it is just too

complex for anything but a very motivated user to use. A better model is to support the interconnection of tools for the developer who then creates a process out of primitive tools. The Eye-ConTact project and SEASR are precisely such visual programming languages that visualize the flow of data through processes.

The challenge of such embedded-tool projects is magnified if the tools are placed in large content collections. Even in our smaller experiments we have had to think about reliability and scale. Some of the challenges we are currently addressing include:

- Content producers will not embed tools if they are not reliable and if they won't scale. Typically, research tool projects are not funded to run a large-scale service. One solution is to give content producers a path from experimental use, where the tool runs off our tool server, to running the tool on their own servers, giving them the code and helping them adapt the tools to guarantee reliability. One disadvantage of handing off the code is that it makes updating the tools difficult; another is that we can't centrally gather usage statistics.
- Embedded tools, especially opaque ones that use Flash, are difficult to customize to the design of the site in which they are embedded. A programmer comfortable with CSS and HTML can adapt the look of tool bars, like the one produced for the Globalization Compendium. We have provided some parameters to TAToo that allow its size and colour scheme to be customized using a special CSS file, but that undoes the advantage of a strategy where one <object> tag gets you a tool bar.
- Social plug-in models are not mature. The Facebook architecture is proprietary and Facebook is not really a content portal, though it may be where our students are most comfortable. Should Google's OpenSocial be widely adopted by providers of portal frameworks then it is possible that social tool developers could develop to one Application Programming Interface (API), thus making tools available to multiple portals and social applications.
- Differentiating content and tools can be important for scholarly work, especially for quoting results and citing resources. Although we generally want to embed tools as seamlessly as possible into content, it is also important to make clear the distinction between the two as users might want to integrate them differently into their research. The tool itself, when embedded, potentially becomes part of the content and could confuse other tools.
- This model doesn't support the interoperation of tools easily. All of these examples are optimized for embedding following established models like the YouTube object or the Facebook plug-in. We can imagine a more complex model where tools are connected in an environment like Yahoo Pipes and then exported as panels.
- The most difficult challenge ahead, however, lies in overcoming the differences between the digital library culture that mounts and maintains online text collections and the culture of text analysis tool development which is more of a research craft. We need to find venues for discussing what content providers want and connecting them with research developers in the community.

## 4. Conclusion

In conclusion, we outline a topology of models from which developers and users of content collections like The Poetess Archive can choose.

**Tool-Driven Model.** If the content you are studying doesn't have analytics available, you can always use online tools that will take uploaded content or a URL. The TAPoR portal provides access to these or you can try the TAPoRware tools directly. Voyeur Tools can also be used that way.

**Enhance Your Browser.** You can add tools to your browser for use as you surf the web. TAPoR provides bookmarklets that you can drag to your bookmark bar. These will invoke a tool on the web page being viewed.

**Embed in Your Blog.** If you have a content publishing framework like a blog, wiki or other CMS (Content Management System) you can place code in the templates for relevant pages so that analytics show up on each content page for analyzing that page. You can do this with snippets of HTML that tools like TAPoR, TAPoRware, TAToo and Voyeur Tools give you. If you are comfortable editing HTML, CSS and JavaScript then you can edit your own toolbar into your online journal, blog, wiki or other resource.

**From Research to Publication.** If you are conducting research using computer-assisted text analysis and think you might want to include results in an online venue, whether it is an online journal, your blog, or a wiki, consider using Voyeur Tools. Voyeur Tools supports going from an enhanced reading environment suitable for research to quoting results in online venues.

We are at a point of emergence when the diversity of ubiquitous tools is likely to increase. We suspect that for a while we need to reinvent wheels in different sizes, interfaces, and in different formats in order to see which ones suit emerging publishing and online research models. Rather than expecting the killer tool to magically transform our colleagues into analysts, we believe that tools in all different forms and from different developers need to be made ubiquitous so that they can be integrated into rich text contexts. Let the research stand out.

---

## 5. Links

Digital Texts 2.0: <<http://dtext2.org/>>. The project documentation is at: <[tada.mcmaster.ca/Main/DigitalTexts2](http://tada.mcmaster.ca/Main/DigitalTexts2)>.

Digital Humanities Quarterly: <<http://www.digitalhumanities.org/dhq/>>.

TAToo (previously FlashTAT): <<http://ra.tapor.ualberta.ca/~tattoo/>>. The earlier project documentation is at: <[tada.mcmaster.ca/Main/FlashTAT](http://tada.mcmaster.ca/Main/FlashTAT)>.

Globalization and Autonomy Compendium: <[www.globalautonomy.ca/](http://www.globalautonomy.ca/)>.

OpenSocial: <[code.google.com/apis/opensocial/](http://code.google.com/apis/opensocial/)>.

SEASR: <<http://seasr.org/>>.

TAPoR Portal: <[portal.tapor.ca/](http://portal.tapor.ca/)>.

TAPoRware: <[taporware.ualberta.ca](http://taporware.ualberta.ca)>.

The Poetess Archive: <<http://www.poetessarchive.com/>>.

"Tools for Data-Driven Scholarship: Past, Present, Future" Report:  
<[http://mith.umd.edu/tools/?page\\_id=60](http://mith.umd.edu/tools/?page_id=60)>.

Ubiquity: <[wiki.mozilla.org/Labs/Ubiquity](http://wiki.mozilla.org/Labs/Ubiquity)>.

Voyeur Tools: <<http://voyeurtools.org>>.

---

<sup>1</sup> This paper was originally presented at *Digital Humanities 2009*, University of Maryland, College Park, Maryland. We would like to thank the Canada Foundation for Innovation, McMaster University and the University of Alberta for support for the TAPoR project and the Social Science and Humanities Research Council of Canada for support for research projects around text analysis.

<sup>2</sup> The "Tools for Data-Driven Scholarship: Past, Present, Future" report can be found at, <[http://mith.umd.edu/tools/?page\\_id=60](http://mith.umd.edu/tools/?page_id=60)>. The Poetess Archive is at, <<http://www.poetessarchive.com/>>.

<sup>3</sup> Cherry, J., & Duff, W. "Studying the usability of TAPoR, A Text Analysis Portal for Research." Faculty of Information Studies, University of Toronto, Research Day, March 10, 2006.

<sup>4</sup> See Rockwell, Geoffrey and John Bradley. "Eye-ConTact: Towards a New Design for Text-Analysis Tools." *CHWP A.4*, publ. February 1998. <<http://www.chass.utoronto.ca/epc/chwp/rockwell/>>. Eye-ConTact was programmed by Patricia Monger at McMaster University.

<sup>5</sup> From the SEASR home page, <<http://seasr.org/>>. SEASR stands for Software Environment for the Advancement of Scholarly Research.

<sup>6</sup> The TAPoR portal is at <[portal.tapor.ca](http://portal.tapor.ca)>. To see these screens you need to get a free account. The portal was programmed by Open Sky Solutions, <<http://openskysolutions.ca/>>. James Chartrand led the programming of the portal working with Geoffrey Rockwell and Stéfan Sinclair.

<sup>7</sup> The TAPoRware tools ([taporware.ualberta.ca](http://taporware.ualberta.ca)) are distinct from the TAPoR portal. The portal is where tools are registered so that the portal can organize them. The TAPoRware tools were developed as a set of reference tools for use in the portal. The programming was led by Lian Yan at McMaster University under the supervision of Geoffrey Rockwell and the tools are now being maintained by Kamal Ranaweera. These tools are now being replaced by the Voyeur tools discussed later in the paper.

<sup>8</sup> Open Journal System, Public Knowledge Project, <<http://pkp.sfu.ca/?q=ojs>>. See also Siemens, Ray et al. "'It May Change My Understanding of the Field': Understanding Reading Tools for Scholars and Professional Readers," *Digital Humanities Quarterly*, 3:4, Fall 2009. <<http://www.digitalhumanities.org/dhq/vol/3/4/000075/000075.html>>

<sup>9</sup> See <<http://www.globalautonomy.ca/>>. To see the toolbar you need to go to an article or other full-text item. Documentation for those who want to install a toolbar is at <<http://tapor1.mcmaster.ca/~taporware/addTool.shtml>>.

<sup>10</sup> Get your TAToo at <<http://ra.tapor.ualberta.ca/~tatoo/>>.

<sup>11</sup> For example you can see it at Geoffrey Rockwell's blog, <<http://theoreti.ca>>. We also installed it by default in the Day of DH blogs created in 2010, see <[http://tapor.ualberta.ca/taporwiki/index.php/Day\\_in\\_the\\_Life\\_of\\_the\\_Digital\\_Humanities\\_2010](http://tapor.ualberta.ca/taporwiki/index.php/Day_in_the_Life_of_the_Digital_Humanities_2010)>.

<sup>12</sup> See <<http://dtext2.org/>>. A Facebook account will be required to view data from the site, including what readers are registered and what they have read.

<sup>13</sup> Rockwell, Geoffrey and Stéfan Sinclair, "Now Analyze That" <<http://hermeneuti.ca/rhetoric/now-analyze-that>>. The original version of the essay which used TAPoRware tools is at <<http://tada.mcmaster.ca/Main/NowAnalyzeThat>>. The screen shot is from the original.

<sup>14</sup> You can read about the Experiments in Text Analysis at <<http://tada.mcmaster.ca/Main/ExperimentsInTextAnalysis>>.

<sup>15</sup> *Voyeur Tools:Reveal Your Texts* is available at <<http://voyeurtools.org>>.

<sup>16</sup> See <<http://www.ucl.ac.uk/dh-blog/2010/07/07/cirrus-a-cirrus-type-wordle-of-the-dh210-conference-abstracts/>>