

# Text Analysis Portal for Research, Using the Public Release

*Geoffrey Rockwell (georock@mcmaster.ca)*

*McMaster University*

*Stéfan Sinclair (sgsinclair@gmail.com)*

*McMaster University*

## Introduction

In April of 2007 the TAPoR project, after extended testing and two public beta versions, released Version 1.0 of the Text Analysis Portal for Research.<sup>1</sup> This portal gathers text analysis tools configured as web services into an environment where users can define the texts they want to work with, select the tools that best work with different text formats, and log their research. The portal is, on the one hand a broker for piping electronic texts to tools and then managing the results, and on the other hand is a study environment where projects can be managed over time. Our poster will do three things:

- Show an overview of how the portal works with annotated screen shots.
- Show the life of the project including the user testing phases.
- We will have a laptop in order to demonstrate a live version of the portal to interested.

## Demonstration

Provided we can have a live internet connection we will be prepared to demonstrate:

- How the portal can be used even without an account using the Try It feature or myLinks pages where users can publish sets of texts for analysis.
- How to get a free account from TAPoR.
- How to define texts that may be elsewhere on the internet or uploaded. How to organize these texts using tags and how to publish a set of searchable texts for students or colleagues.
- How to use the Workbench to organize a set of texts and specific tools for a project. How to run tools on texts and how to save results.

- How to log the progress of a research project and share it through TAPoR.

If a live connection is not available we will have a local version running based on the TAPoR Live CD which is a bootable version of linux and the portal that works locally.

## User Testing and the Life Cycle of a Tool Project

Version 1.0 of the portal is the result of extensive user testing which will be visually represented in the poster as will ways new users can provide us feedback as we develop the priorities for further releases.<sup>2</sup>

For those familiar with earlier versions of the portal there are a number of improvements including:

- A new type of news channel called a Research Log. Users can save results from tools to this log along with comments. They can also save the text/tool configuration to run again.
- A new feature called Analyze Text that will open a window with the text on the right and suggested tools on the left for close study of one text.
- The myLinks feature has been adapted so users can publish their public texts for others to search across and for other to analyze. This can be used, for example, for providing students with a study set of texts.
- The interface has been simplified to make it easy for users to get projects going.

## Background

The TAPoR Portal is one of the outcomes of a 5 year project funded by the Canada Foundation for Innovation (CFI) that involves 6 universities across Canada, including the University of Victoria, the University of Alberta, McMaster University, the University of Toronto, the Université de Montréal and the University of New Brunswick. Provincial funding bodies and our respective universities have also supported the TAPoR project. The portal was developed by Open Sky Solutions with a team primarily at McMaster University. The poster will recognize the supporting institutions but will focus on the portal and not present the wealth of other projects that are part of TAPoR.<sup>3</sup>

This poster will show the current version of a major text analysis tools development effort that developed a model for sustainable and modular text analysis for a broad community. This is a challenge that has concerned computing humanists since the 1980s. As Susan Hockey put it in a post to HUMANIST in 1996 that reported on a meeting she convened at CETH, "For some

time, those of us active in humanities computing have felt the need for better and/or more widely accessible text analysis software tools for the humanities. There have been informal discussions about this at a number of meetings, but so far no substantial long-term plan has emerged to clarify exactly what those needs are and to identify what could be done to ensure that humanities scholars have readily-available text analysis tools to serve their computing needs into the next century."<sup>4</sup>

The TAPoR model meets a number of the objectives described for text analysis tools in the 1990s by Hockey and others.<sup>5</sup> It is widely accessible. It provides a long-term model using web services to bring tools together where they can be used in a study environment. It allows new tools to be added and existing tools to be improved or replaced over time as new needs emerge. It allows research to be recorded and shared.

## Future Plans

**T**he TAPoR Portal also has limitations. Depending on tools provided as web services by others increases the chances of simple operations not working. Many of the tools are not multi-lingual. The tool broker model where the portal gets a remote text and passes it to a remote tool is not as efficient as an all-in-one model where texts can be preindexed. And, as with any large software project there are still inconsistencies, awkward interface panels, and bugs.

The poster will conclude by outlining future plans. TAPoR hopes to be funded for a second phase through the CFI Leading Edge Fund in order to add large-scale text datamining and aggregation tools. Currently the portal model does not pre-index large collections (except in limited cases) and provides limited aggregation or crawling tools. TAPoR 2, if funded, will continue improving the interface and will add large-scale functionality by adapting crawling, scraping and data mining tools to the study space.

4. Susan Hockey, *Humanist Discussion Group* 10.54 (23 May 1996). See <[http://lists.village.virginia.edu/lists\\_archive/Humanist/v10/0054.html](http://lists.village.virginia.edu/lists_archive/Humanist/v10/0054.html)>. There is also a note from Michael Sperberg-McQueen pointing to a trip report available at <<http://tigger.uic.edu/~cmsmcq/trips/ceth9505.html>>
5. Geoffrey Rockwell and John Bradley, "Eye-ContTact: Towards a New Design for Text-Analysis Tools", CHWP A.4. (1998). See <<http://www.chass.utoronto.ca/epc/chwp/rockwell/>>

- 
1. For more about the TAPoR project see <<http://www.tapor.ca>>
  2. TAPoR interface at <<http://www.tapor.ca/interface>> details a persona oriented investigation conducted by Audrey Carr and Joanna Dacko. Dr. Wendy Duff at the University of Toronto also conducted extensive interviews that have not been published, but which were abstracted for the portal development team. Finally, we have a layered testing process combined with the gathering of tool statistics that is being conducted on the release.
  3. *Mind Technologies* edited by Raymond Siemens and David Moorman (U of Calgary Press, 2006) includes a number of articles about the variety of projects supported by TAPoR.