**SINCLAIR — Voyant**

<0 images>

**Voyant Tools 2.0: The New, the Neat, and the Gnarly**

Sinclair, S. and Rockwell, G.

**Context**

Voyant Tools (voyant-tools.org) is a web-based reading and analysis environment for digital texts. Users can create their own corpus of texts to study by pointing to URLs or uploading files in a variety of formats (plain text, XML, HTML, PDF, MS Word, RTF, etc.). Voyant allows users to navigate between macro views of the corpus (e.g., a word cloud visualization of the entire corpus) and micro views (e.g., a reading individual occurrences of a specific term in context). The default interface provides access to a basic set of tools for reading texts and studying word frequency and distribution. There are also more tools available in various pre-defined or user-defined ‘skins’ (a layout of tools that are coupled).

Voyant Tools is deliberately designed to be user-friendly and welcoming for text analysis. Voyant currently averages nearly 50,000 visits and about 750,000 tool invocations per month (not counting the downloadable instances of VoyantServer). This will be the seventh consecutive workshop of Voyant, with past sessions focusing on different aspects (pedagogy, multilingualism, customizability, standalone version, etc.).

The 2015 workshop will focus on the second major release of Voyant Tools (2.0), which represents an entire rewrite of the codebase to address several of the major shortcomings and irritants of the currently available version 1.0. Version 2.0 is currently available in a beta version online with a major release due in early spring 2015. In addition to performance improvements throughout, the search and filtering functionality have been vastly enhanced, and Voyant now supports proximity and n-gram operations. Voyant 2.0 also has improved corpus handling. Documents can be reordered or added to corpora on the fly, and there is a lightweight access management layer that differentiates between full access, full-text access, and expressive/consumptive access.

We have designed this workshop to be of interest both to new users of Voyant, who will get an introduction to the platform, and to existing users, who will discover all the new functionality 2.0 has to offer. As always, a crucial aspect of the workshop will be to get feedback from the community.

**Workshop Outline**

*1. Introduction to Text Analysis with Voyant (1 hour)*

We will begin with a general introduction to text analysis using Voyant aimed at those who haven’t used it before. We will provide a brief overview of Voyant’s user interface and discuss its strengths and weaknesses. We will provide initial text collections that users can use with Voyant, with a view to having participants experiment subsequently with their own text collections.

*2. Voyant 2.0: What’s New? (1 hour)*

This second part of the workshop will focus on what’s new in Voyant 2.0. Examples of changes include more powerful proximity and fuzzy searching of terms, infinite scrolling instead of paginated scrolling for tabular data, in-place modifications of corpora (adding documents or re-ordering them), and new tools (collocate networks, n-gram wordtrees, etc.). This part of the workshop will be useful both for users familiar with the old Voyant (to understand the changes and enhancements) and also to newcomers who will get a better sense of the variety of available tools.

*3. Voyant: Text Repository or Analytic Platform? (1 hour)*

One benefit of the enhanced scalability of Voyant Tools 2.0 is the ability to bridge the gap between existing text repositories (typically focused on searching for documents) and analytic platforms (for text mining). We are collaborating with several large-scale content providers (like TCP-EEBO and Érudit.org) to create custom Voyant skins that allow users to search and filter within very large text collections in order to create smaller worksets of relevant documents for analysis. Because everything is happening in Voyant, the jump from text repository to text analysis is smooth and efficient (very few text repositories allow mass downloading of worksets, but even when they do, additional steps are typically required for re-ingesting the workset into an analytic platform). This component of the workshop will demonstrate some of our existing collaborations and describe how other content providers and projects might be able to leverage this hybrid functionality.

**Workshop Leaders**

*Stéfan Sinclair*, sgsinclair@gmail.com, is an associate professor in digital humanities at McGill University. His research focuses primarily on the design, development, and theorization of tools for the digital humanities, especially for text analysis and visualization. He has led or contributed significantly to projects such as Voyant Tools, the Text Analysis Portal for Research (TAPoR), and BonPatron. Other professional activities include serving as associate editor of *Digital Humanities Quarterly*, as well as serving on the executive boards of ACH, CSDH/SCHN, ADHO, and centerNET.

*Geoffrey Rockwell*, grockwel@ualberta.ca, is a professor of philosophy and humanities computing at the University of Alberta, Canada. He has published and presented papers in the area of philosophical dialogue, textual visualization and analysis, humanities computing, instructional technology, computer games, and multimedia. He was the project leader for the CFI (Canada Foundation for Innovation)–funded project TAPoR, a Text Analysis Portal for Research (tapor.ca), which has developed a text tool portal for researchers who work with electronic texts. He is the author of *Defining Dialogue: From Socrates to the Internet* (Humanity Books).

**Target Audience**

A wide range of DH practitioners interested in text analysis, particularly for research, teaching, or technical support. Voyant Tools workshops are typically fully subscribed; we prefer to limit registration to about 25 people to allow us to help participants as needed.

**Format**

Half-day.