

Voyant Tools (voyant-tools.org)

Geoffrey Rockwell

Introduction

The construction of this new record of the human past, present, and future, as we have also suggested, is gravely affected by socio-technical inequities — not every cultural heritage or scholarly outfit around the world has access to the same infrastructure or resources. (Roopika & Gil 2022)

Data analysis and visualization tools, like much computing infrastructure, are not equitably accessible. Further, such tools have been used to discipline communities and manage information about others. While the conditions around the gathering and enrichment of data has received attention, there hasn't been the same attention to the design of globally accessible tools (D'Ignazio & Klein 2020). Perhaps it is because tools are seen as neutral (Winner 1980) compared to the messy data that is haunted by absences (Ortolja-Baird & Nyhan 2022). Visualization tools especially get interpreted as providing an objective God's eye view (Haraway 1988).

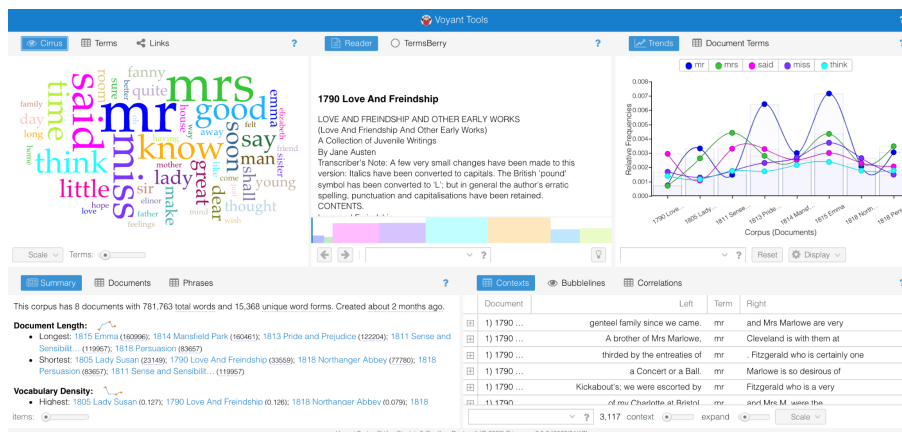


Fig. 1: Default View of Austen's Novels in Voyant

Voyant Tools is an attempt to do better. Voyant Tools is a suite of text and visualization tools freely available online through the browser (and downloadable) that make common knowledge techniques public, from data mining to topic modelling. The tools were designed by text scholars in the humanities and they are widely used around the world to study historical, cultural, literary, legal and contemporary texts and then to make analytical environments public for others. With an interface translated into 13 languages by volunteers, Voyant can be used on any accessible text in the browser without installation, which is why they have become a popular toolset for teaching, research, activism and literacy use.

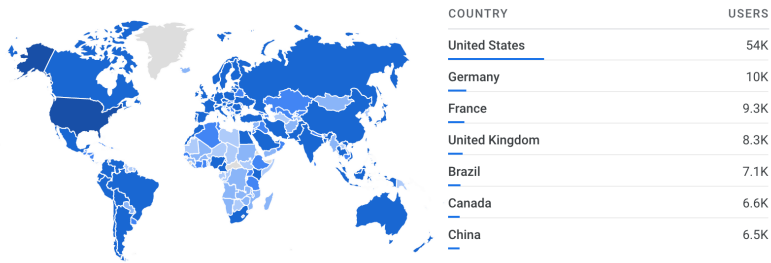


Fig. 2: Google Analytics map from 2023 showing top countries where users are coming from

The accessibility of Voyant, its ease of use, and multilingual capabilities make it popular for teaching the digital humanities. Google Analytics shows that the main server has been used by over 200,000 users in the last year. In addition, there are other servers run by other digital humanities groups not to mention the many researchers who have downloaded and run Voyant on their own. These tools are an attempt to democratize access to the tools of data mining and visualization allowing citizens to experiment and publish dashboards. They make it possible for publics to take control of their own information rather than depend on commercial tools and data scientists. They are not just accessible to anglophones in the first world; they are used in countries like Brazil, China, and India.

Announcing: The Launch of a Voyant Consortium

History

Voyant (voyant-tools.org) grew out of the HyperPo environment by Stéfán Sinclair and the later TAPoR portal project led by Geoffrey Rockwell. In 2000, Stéfán Sinclair defended his PhD thesis on “Une application d’HyperPo, un logiciel d’analyse de texte informatisée à *La Disparition de Georges Perec*” at the French Studies department at Queen’s University in Kingston, Ontario. The title translates roughly as “An application of the text analysis software HyperPo to Georges Perec’s *A Void*.”¹ In the thesis he wrote about computer-assisted text analysis including a discussion of an online in-browser analysis tool, HyperPo (short for Hypertexte Potentiel), software created as part of and for the thesis project. HyperPo pioneered many of the innovative features of Voyant. It was usable on the web in a browser. You could upload a text for analysis and visualization. It had a set of loosely concatenated panels with different views on the text, including a full text view. The problem was that it became increasingly hard to maintain and it couldn’t handle large corpora.

¹ You can Access the thesis from Theses Canada: <https://www.bac-lac.gc.ca/eng/services/theses/Pages/item.aspx?idNumber=1006925305> (Accessed August 2, 2022)

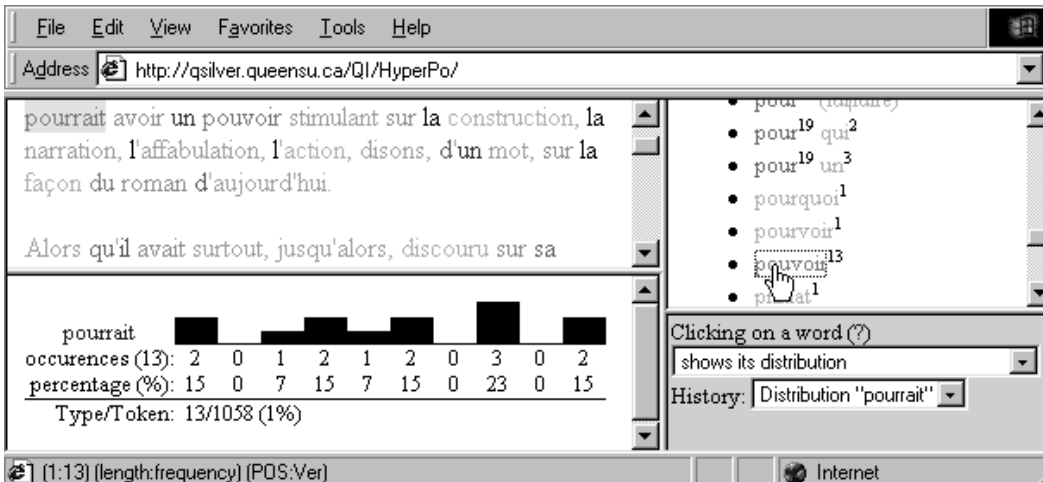


Figure 3: Screenshot from “HyperPo: The Next Generation”

The TAPoR or Text Analysis Portal for Research project (tapor.ca) brought together 6 universities across Canada to build local and shared infrastructure for text analysis. TAPoR was funded by the Canada Foundation for Innovation in 2002. The funding went to local labs, servers, special peripherals and text enrichment at the universities and shared conferences and a portal for all. The portal was designed to bring together text analysis tools, corpora, news, and online training. The development of the portal also brought Rockwell and Sinclair together at McMaster University.

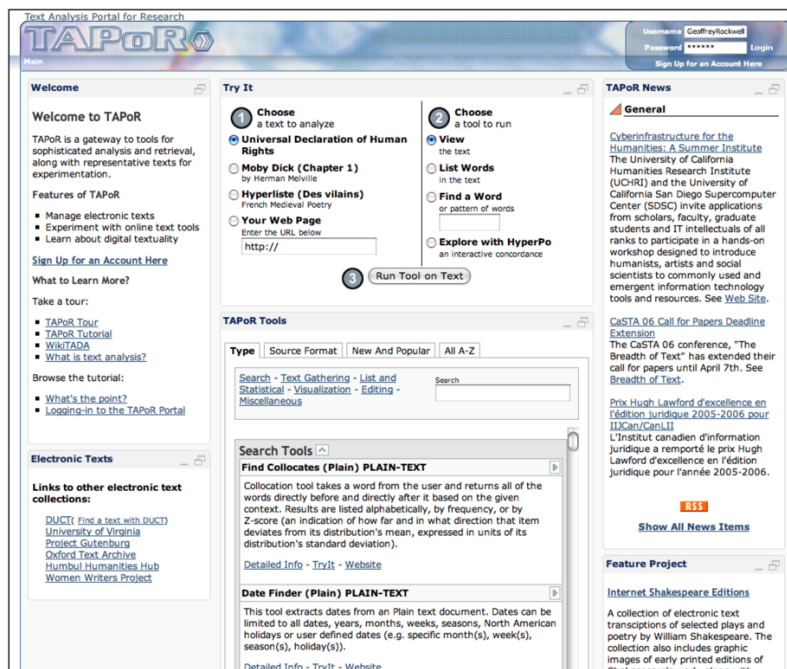


Figure 4: TAPoR Home Screen

Alas, the portal was too complicated and depended on web services at other institutions which weren't stable enough. The very idea of a portal bringing all relevant services together was too

Why notebooks for humanists? Notebook environments encourage students and researchers to document their thinking in text blocks with accompanying code blocks that do the processing and analysis. For the digital humanist, the notebook environment lets you save the interesting visualizations as interactive panels to a notebook along with your discussion of results and next steps. Spyral gives you an alternative interface to Voyant where you can call the tools you want, in the order you want, with the parameters you want, while keeping their interactivity. Notebooks are saved on the Voyant server and can then be shared or forked. As with Voyant, you don't need to install anything locally.

We believe this will provide digital humanists with a growth path. You can start by playing with texts and then move to sharing your results in an online notebook. You can try different tools and see what JavaScript code would generate that panel. Then you can play with the parameters and finally, try writing your own code. Thus DH learners can go from trying Voyant to learning to program by example.

Long Term Sustainability

To make Voyant a truly accessible and sustainable tool suite we are now developing an international Voyant Consortium that can guide its development over time. Voyant needs to be guided so that it is useful to global researchers.

We are therefore transitioning to a consortial model that has a global Community Board that can represent the needs of scholars around the world.

The Voyant Consortium with its founding Community Board will be announced at DH 2024. You will be able to register to become a member for free and learn more about the Consortium.

See voyant-tools.org for links!

Some of our reasons for shifting to becoming a Consortium include:

Co-Design and Development. A major reason for moving to a consortial model is to meaningfully involve researchers and other stakeholders from around the world. For that we need to have an organization that allows people to get engaged and have a real voice in decisions. This has to be an organization nimble enough to operate with volunteers but open enough to be inclusive globally.

Learning Support. Perhaps the greatest value of Voyant has been its ease-of-use and ease-of-learning. It is used widely in digital humanities courses as it is easy to integrate into the curriculum, thereby becoming an important part of digital humanities literacy. We hope a Consortium can help instructors use Voyant, develop learning materials, support the sharing of best practices, and support adaptation across languages and disciplinary communities.

Minimal Computing. Essential to these tools being easy-to-use and adaptable to global needs is developing them so that they can be part of minimal computing solutions that work in sites without the infrastructure we take for granted in Canada. We believe we can create a minimal “Voyant on a Stick” solution that could run off a memory key, for example. A Consortium could support such initiatives led by and for different communities.

Accessibility. We have not had the resources to make Voyant accessible. We hope a Consortium will involve and engage people with the expertise to help us make Voyant and Spyral more accessible.

Long-Term Architectural Development. Part of Voyant’s success has been its longevity. Version 2.0 has been running since 2016 and earlier versions (HyperPo and Voyeur) go back to 2000. The longevity is partly due to the long-term software architecture design developed by Dr. Stéfán Sinclair and partly due to our capacity to maintain components under the hood without taking down the service or dramatically changing it. The time has now come to review the architecture and future-proof Voyant for another decade or two. We have enlisted the support of colleagues in Computing Science at the University of Alberta for a preliminary architectural review, and the programmer Andrew MacDonald has identified priorities for upgrading. They include upgrading and, in some cases changing, the software libraries used. None of this is trivial but it needs to be done to maintain the currency of Voyant. To guide the long-term development we need a Consortium that can involve and engage colleagues with the skills to guide us collectively through.

References

- D’Ignazio, C. and L. F. Klein (2020). *Data Feminism*. Cambridge, Massachusetts, MIT Press.
- Haraway, D. (1988). “Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.” *Feminist Studies*. 14(3): 575–99.
<<https://doi.org/10.2307/3178066>>
- Knuth, D. (1984). “Literate Programming,” *The Computer Journal* 27, no. 2: 97-111.
- Ortolja-Baird, A. and J. Nyhan (2022). "Encoding the haunting of an object catalogue: on the potential of digital technologies to perpetuate or subvert the silence and bias of the early-modern archive." *Digital Scholarship in the Humanities*. 37(3): 844-867.
- Risam, R. and A. Gil (2022). "Introduction: The Questions of Minimal Computing." *Digital Humanities Quarterly*. 16(2).
<<http://www.digitalhumanities.org/dhq/vol/16/2/000646/000646.html>>
- Rockwell, G. and S. Sinclair (2016). *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Cambridge, Massachusetts, MIT Press. <<https://hermeneuti.ca>>
- Winner, L. (1980). "Do Artifacts Have Politics?" *Daedalus*. 109(1): 121-136.

In memory of Stéfán Sinclair who passed in August 2020
<https://csdh-schn.org/stefan-sinclair-in-memoriam-2/>