**O’SULLIVAN — Nocht**

<0 figures>

**Nocht: An Open Source Tool for Text Analysis**

O’Sullivan, J. C., Hswe, P. and Long, C. P.

Computational approaches to text analysis have revolutionised the ways in which scholarly research is being conducted. A number of tools exist that help scholars, from a variety of disciplines, analyse textual data, whether literary, historical, or otherwise, using scientific methodologies. However, many of these tools are either proprietary, present a steep learning curve, or are constructed without much transparency, often leaving users with results whose means of production they do not understand. This poster will outline the development of a tool that is intuitive and completely free and open-source, so that scholars in literary studies, and indeed the broader humanities, can leverage computational methods and big data analytics in their research.

# Nocht

Nocht (trans.: to reveal, uncover), is developed, primarily, in Python, so that it is flexible, scalable, and cross-platform. It has been developed in accordance with the following principles:

• It offers users a low-barrier means of using computational approaches to text analysis.

• It is designed and developed in a humanities / arts / social sciences context.

• It is completely open-source, removing the ‘black-box’, closed-code issues.

• It brings together existing libraries and code-sets, acting as a ‘script portal’ of sorts.

At present, Nocht supports the following methodologies, though with some limitations:

• Wordcount and most frequent wordlists.

• Word / wordlist frequency plotting.

• Syntax and sentence analysis.

• Sentiment analysis.

• Topic modeling.

• Zeta analysis.

It is hoped that Nocht will further contribute to our field’s ongoing commitment to open and sustainable research tools, complementing highly regarded projects like Voyant1 and Stylo (Eder et al., 2013). Its name is an obvious tribute to the former, which has for so long been one of our field’s fundamental tools. It is hoped that Nocht will add further to the DH toolkit, as well as complement the ongoing work of Voyant’s creators in leveraging the iPython architecture.

From a technical perspective, Nocht is scalable and robust, and satisfies the needs of a wide range of scholars, many of whom wish to conduct this form of research but lack the expertise or resources to do so. In this respect, it enables scholars, both emerging and established, to engage with cutting-edge analyses across a variety of disciplines. In many cases, it draws on a series of existing libraries and proven methodologies, such as NLTK2 and matplotlib,3 and so acts as a set of original scripts as well as a portal to existing tools. A complete technical overview of the project’s features, as well as the components utilised in its modular development, will be provided at the session.

# Discussion

This poster proposes to introduce Nocht to the field, discussing possible future development directions, as well as issues to date. Some of the disciplinary particularities identified by Gibbs and Owens (2012), such as our need to enhance the usability of our tools, will be addressed. The tension between having an intuitive interface and the need for scholarly tools to produce verifiable results is particularly clear in this project. While there is a long-established requirement that such tools be user-friendly (Krug 2005), one might argue that this must be balanced with a commitment to avoiding ‘black-box’ projects; usability does not necessarily equate to understanding.

Measuring the value and success of development projects also remains problematic for scholars and practitioners working across the digital humanities. Schreibman and Hanlon’s survey (2010) finds that the majority of respondents were satisfied that their tools had been ‘successful’, enabling themselves and others to further their research. However, respondents also outlined that they had measured this success from a ‘controlled list’. As our methods continue to gain prominence beyond the core digital humanities community, we must find new metrics through which we can reliably measure the impact of our tools, not just in terms of user volumes, but in relation to the quality of research output. As a project that has sacrificed some aspects of usability and marketability in favour of broad functionality and a commitment to open principles, perhaps to its detriment, Nocht is an ideal catalyst for this debate. It is a small development with limited financial support, so it will be interesting to see if projects of this scale have a future in our discipline.

**Notes**

1. See Stéfan Sinclair and Geoffrey Rockwell, http://voyant-tools.org/.

2. Natural Language Toolkit, http://www.nltk.org/.

3. matplotlib, http://matplotlib.org/.

**References**

**Eder, M., Kestemont, M. and Rybicki, J.** (2013). Stylometry with R: A Suite of Tools. *Digital Humanities 2013: Conference Abstracts*, University of Nebraska–Lincoln, pp. 487–89.

**Gibbs, F. and Owens, T.** Building Better Digital Humanities Tools: Toward Broader Audiences and User-Centered Designs. *Digital Humanities Quarterly,* **6**(2).

**Krug, S.** (2005). *Don’t Make Me Think! A Common Sense Approach to Web Usability*. 2nd ed. New Riders Press, New York.

**Schreibman, S. and Hanlon, A. M.** (2010). Determining Value for Digital Humanities Tools: Report on a Survey of Tool Developers. *Digital Humanities Quarterly,* **4**(2).