Credit Allocation with the Social Knowledge Timeline

Aaron Mauro
mauro@psu.edu
Penn State Digital Humanities Lab
United States of America

Introduction

Humanities scholarship is becoming increasingly collaborative, participatory, and public facing. As humanists take up digital tools to conduct and share research, larger teams are needed to complete ever more complex computational tasks. When blending these heterogeneous teams--which may include faculty, librarians, staff, undergraduates, graduate students, postdocs, and community contributors--humanists have an ethical responsibility to offer a fair and transparent accounting of research activities. Tracing the evolution of research contributions is necessary for a range of issues facing digital scholarship such as authorship allocation, promotion and tenure, and reports to funders. The allocation of credit and authorship is an increasingly thorny issue for teams with a range of possible roles and a variety of research outputs and media types. There is, however, a large amount of data being generated by these teams that is capable of describing and measuring the contributions made on a variety of platforms and by multiple team member and community partners.

Despite our inheritance of social and collaborative tools, many of these systems elide the nuance and process of humanities based research. Knowledge creation is not merely a function of how much code is produced. New knowledge is often the result of a key insight made by a team of students, staff, and faculty. These insights are generated in a complex and overlapping system of mentorship, service, teaching, learning, and authorship that are deeply dependent, social, and human. With a system that is possible of visualizing the history of a digital project over the course of years, which may see the ranks of team members change over time, primary investigators and project funders will be better able to address often thorny and ethically charged issues relating to student

assessment, mentorship, authorship, promotion, and tenure. Credit, promotion, funding, and credentialing are more complex topics than ever, yet many individuals and institutions rely on simple, outdated structures to assess the value of insights made by networked teams.

Social Knowledge Creation

The Penn State Digital Humanities Lab (Penn State Behrend) in partnership with the Teaching and Learning with Technology group (Penn State University Park) has developed a prototype of an ongoing project entitled the Social Knowledge Timeline (sktimeline.net). By linking together popular collaboration tools, the SKTimeline stores, analyzes, and communicates user data in three distinct areas of social knowledge creation:

- **Collaboration Platforms:** Many scholars are turning to collaboration platforms like Slack, Yammer, and Basecamp to organize teams and foster communication within teams. These systems use an interface similar to a social media feed to pool project member input into a single narrative and eliminate the need for email. These systems help share documents and support conversations that may lead to drafting manuscripts on Google Drive and other services. These platforms offer a rich, conversational natural language data set that describes how team members mentor and support each other over time.

- **Version Control Systems:** Github and Bitbucket are two of the most common version control platforms. These tools help facilitate large programming and encoding projects by allowing multiple coders to work simultaneously. When a team member "commits" code, a commit message describes the nature of the contribution as well as the date and time. This message will offer a highly granular view of coding projects as they unfold. Similarly, by including a feed from Google Drive's own version control system, document authorship may be traced with similar precision.

- **Social Media:** Platforms like Twitter, LinkedIn, and Facebook have proven to be

fast paced and engaging areas for social and cultural exchange. Twitter has long been a particularly important site for digital humanists. The SKTimeline draws together multiple hashtags and user handles to frame preserve and contextualize this often ephemeral site of both popular and scholarly debate. Hashtags associated with digital projects, conferences, publications, and even course work can be analyzed and set in real time with other platforms.

Credit allocation in large teams is dependent on our ability to describe, quantify, and visualize our activities. By analyzing the rich natural language conversations generated by teams, the SKTimeline solves these ethical and institutional problems. The appearance of "Collaborators' Bill of Rights" for digital humanities projects in 2011 is symptom of a need for greater clarity in heterogeneous collaborative teams (Clement et al 2011). The Modern Language Association's "Guidelines for Evaluating Work in Digital Humanities and Digital Media" are similarly responding to appropriate credit allocation for researchers. There is a need for a more formalized and automated system of data collection and analysis for collaborative researchers across the university.

Machine Learning Contributor Taxonomies

The Taxonomy of Digital Research Activities in the Humanities (TaDiRAH) is used to quantify and describe user contributions. Machine learning systems like Google's Cloud Platform is used to conduct language analysis, and translation, image recognition, sentiment analysis, and keyword extraction. Custom machine learning systems has also been layered on to these services using the Tensor Flow library to learn the project specific phrasing for contributions. Additional text analysis will be conducted using standard tools like the Natural Language Toolkit (NLTK) to link to TaDiRAH's defined contributions. This project will reshape authorship and credit allocation in the humanities and beyond, but it will also be a perfect test bed for an emerging set of artificial intelligence tools that are now finding common application throughout society. In this way, the SKTimeline is representative of a broader cultural trend toward AI systems in aiding research.
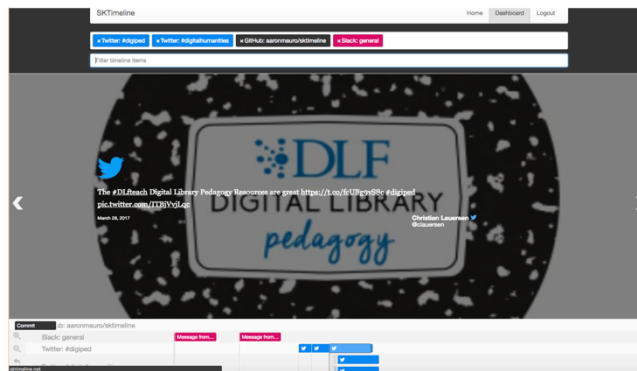

Figure 1. The Social Knowledge Timeline displaying Slack channels posts and Twitter hashtags chronologically. Images associated with posts are used for backgrounds on the timeline

Conclusion

Undergraduate course projects, ongoing faculty research with graduate researchers, digital humanities labs, and library based digital research projects are just some of the contexts this round of user testing will examine. The data collected on participating teams against interview and form based user surveys. This kind of socially oriented knowledge creation emerges from a community of practice that moves fluidly between curricular experiences and co-curricular research experiences often hosted in DH labs, libraries, and centers. The SKTimeline seeks to solve a critical problem within scholarly communication in a digital context. The SKTimeline offers a means to capture complex narratives that constitute the organic and nuanced unfolding of humanities research.

Bibliography

Alphabet. (n.d.) Google Cloud Platform. <https://cloud.google.com/products/machine-learning/>

Alphabet. (n.d.) Tensor Flow. <https://www.tensorflow.org/>.

Borek, L. (n.d.) TaDiRAH. <https://github.com/dhtaxonomy/TaDiRAH>.

Clement et al. Eds. (2011) Off the Tracks: Laying New Lines for Digital Humanities Scholars.

Media Commons Press, (2011). Web. <http://mcpress.media-commons.org/offthetracks/>. Committee on Information Technology (2012) "Guidelines for Evaluating Work in Digital Humanities and Digital Media." Modern Language Association. Web.

<https://www.mla.org/About-Us/Governance/Committees/Committee-Listings/Professional-Issues/Committee-on-Information-Technology/Guidelines-for-Evaluating-Work-in-Digital-Humanities-and-Digital-Media>.

Di Pressi et al. (2015). "A Student Collaborators' Bill of Rights." UCLA Digital Humanities. Web. <http://www.cdh.ucla.edu/news-events/a-student-collaborators-bill-of-rights/>.

Python Software Foundation. (n.d.) Python Language Reference, version 3.6. Available at <http://www.python.org>.

Ronacher, A. (n.d.) Flask. <http://flask.pocoo.org/>.