Visualizing Varieties of Association in Orlando

Susan Brown, Universities of Guelph and Alberta
Stan Ruecker, University of Alberta
Milena Radzikowska, Mount Royal College
Matt Patey, Mount Royal College
Stéfan Sinclair, McMaster University
Jeffery Antoniuk, University of Alberta
Sharon Farnel, University of Alberta
Isobel Grundy, University of Alberta

Literary history connects writers or writing with places, publishers, periods of time. Along with links or connections, it supplies context and relationships. Awareness of connection produces contextual understanding, but the specifics of particular connections may obscure overall context and broader patterning. Literary history can now potentially shift from exclusively print-oriented "analog" forms for the recognition and representation of connections to digital modes drawing on a wide range of potential tools–both algorithmic and representational–for investigating the complex associations of the past. This paper will explore, through the literary historical textbase *Orlando: Women's Writing in the British Isles from the Beginnings to the Present*,[1] ways of presenting large, complex sets of connections without obscuring broader patterns.

The Orlando team has been carrying out a series of preliminary investigations of the dense interlinkages created by the Orlando textbase markup. This textbase provides a unique testbed for digital humanities research: an extensively encoded body of born-digital literary historical scholarship on (now 6.8 million words with 2.2 million semantic tags for everything from paragraphs to politics, plots, or relations with publishers).[2] Our work to date has involved a customized online prototype for accessing linkages where the user can adjust the search to accommodate density by selectively removing the most common sources of linkages. In a collection about British women writers, for instance, it is not useful to associate people by whether or not they have any connection to London.

The prototype (Figure 1) employs a tweaked version of the Breadth-First Search algorithm to find the shortest path between people mentioned in the textbase, not only through using other people as nodes (as in the six-degrees-of-separation concept),[3] but either individually through places, organizations, and text titles, or else through any combination of them. The six degrees prototype provides the user with a summary of the connections between two authors, showing short text extracts from the connecting documents, along with the names of the authors, organizations, titles, or so on that have been identified by the system as the basis for the connection. It returns the shortest path between two linked people. However, most of the shortest paths in the textbase are not unique but often hugely multiple: different routes, equally short. This considerably complicates the project of conceptualizing and representing the link paths.

---

[1] Brown 2006.

[2] Brown 2007.

[3] Barabási 2002.
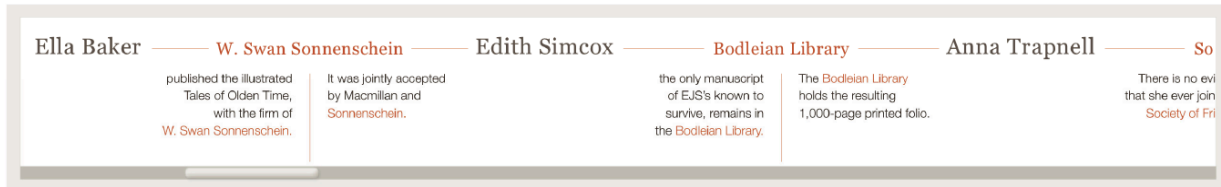
Fig 1: Current interface for linkage, here by organization names, between writers' entries in Orlando.

Even this simple prototype reveals a high degree of underlying complexity in the associations between people in the textbase, and corresponding problems in discerning, conceiving and representing patterns of association. Which linkage is primary, for instance, in Figure 1: Edith Simcox, since the Simcox entry contains mentions of both Sonnenschein and the Bodleian, or the organizations themselves? In addition, since most queries produce multiple paths, it becomes a challenge to represent efficiently multiple categories of information. Visualizations useful to literary scholars must support these complexities while representing the general as well as the specific.

We are currently working on a set of designs (e.g. Figure 2) that can summarize the multiple short paths that represent a typical connection the textbase, as well as considering the implications of different ways–perhaps even alternate views–for representing the linkages.

Fig 2. Prototype sketch summarizing multiple linkages. Here we see the summary of the six paths within Orlando materials from Ella Baker to William Henry Simcox.

Visualizing degrees of separation is one approach to studying association, but there are other strategies, especially where the XML encoding contains tags that relate to association. Another of our experiments has therefore involved using the Mandala XML browser[4] on Orlando materials. The Mandala browser and the degrees-ofseparation prototype are two very different interfaces, offering contrasting approaches to making meaning by non-verbal tools.

---

[4] Mandala Rich Prospect Browser. Dir. Stéfan Sinclair and Stan Ruecker. http://mandala.humviz.org/

As Jessop points out, seeing a visual representation that summarizes a pattern in the data is different from reading the same data as text.[5] In the former case, the viewer is presented with an object of study that is like an illustration or picture. It can be more or less complex, but if it is an effective visualization, it rewards study with insight. Similar insights might be derived by reading text, but the experience of achieving them is completely different. Reading is also a general-purpose activity that can be used for many purposes, while an interactive visual environment is intended primarily to assist in pattern-finding for subsequent deeper exploration through reading.

The Mandala browser is an attempt to create a visual environment for the iterative construction of Boolean queries and study of the patterns emerging from their results. Collection items appear as dots on the periphery. The user draws them into the interior space by creating magnets that attract matching items. Items attracted by more than one magnet are pulled into subsets between the magnets, marked by ghost magnets that are divided into pie slices, with one slice for each magnet involved.

The first Mandala visualization (Figure 3) shows part of the network connecting the Victorian novelists Elizabeth Gaskell and Charles Dickens. The number of entries where they co-occur is represented in the bottom right set of blue (Gaskell) and green (Dickens). It also allows one to see interrelationships, which are concentrated in the lower hemisphere, between all the writers in the network, and not just Gaskell and Dickens. One sees, for example, the connections between Dickens and Carlyle as well. This network of social and literary relationships is clearly denser among those writers in the lower hemisphere, that is Gaskell, Dickens, and Jane Welsh Carlyle.
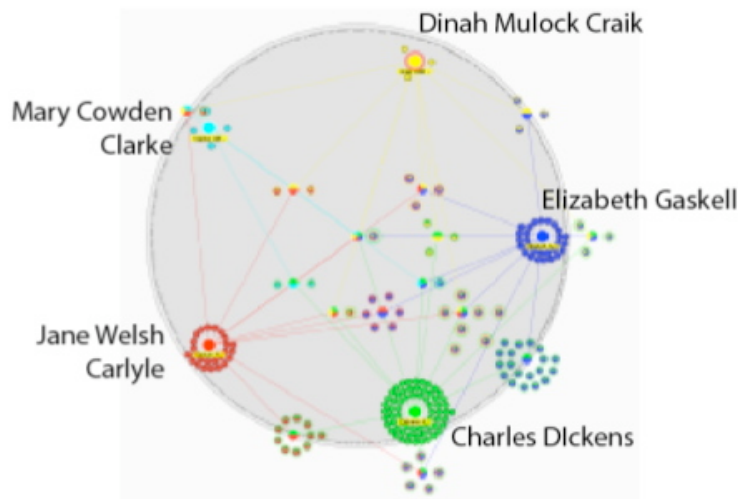
Fig 3. The Mandala browser has been used here to identify relationships between 5 Victorian writers. The dots represent the entires of other writers where the indicated names are mentioned.

---

[5] 2008.

Although the kind of visualization shown in Figure 3 only requires that a name be mentioned in an author's Orlando entry, the Mandala can also be used for examining relationships by type, rather than simply reflecting the fact of some kind of connection. The system can draw on Orlando's custom semantic XML tagset for everything from family members to politics to fictional adaptations of writers or their works, searching for names that occur within those tags to refine the context in which the mention of a name occurs and hence distinguish between different types of connection. For example, in Figure 4, different kinds of relationships to Jane Austen embedded within critical discussions can be contrasted to references to the Victorian writer Geraldine Jewsbury, whose work as a reviewer has made more impact on literary history than her own fiction.

Fig 4. The semantic tags in Orlando can be used to examine the kinds of relationships. Here there are proportionately more reception relationships for Jewsbury than for Austen, because of Jewsbury's activity as a reviewer. Austen, on the other hand, frequently appears in reception tags because there are many comparisons to her work invoked by other reviewers.

By way of contrast, Figure 5 shows a set of semantic tags that involve evaluations of intertextuality, influence, and responses to others. In this case, Austen's presence is more strongly felt and more complex than Jewsbury's.
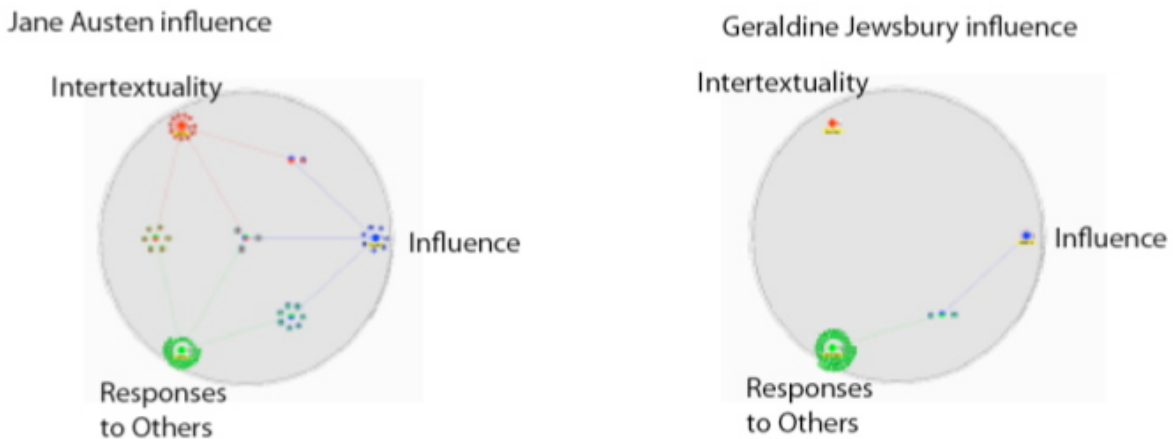
Fig 5. Austen's literary relationships with others are much more multivalent than those of Jewsbury. Whereas Orlando set out, as a literary historical project, with a conviction of the primacy of text, it now seeks to explore the visual presentation of larger patterns, in order to provide further, alternative sets of meanings and hypotheses that the current interface does not afford. These experiments relate not only to literary history but also to other areas of humanities investigation, equally rich in dense and complex interlinkages which almost defy explanation in words. Since many of these areas already employ digital documents using the Text Encoding Initiative[6] tags for person names, places, organizations and titles, they too can use new tools for the exploration of their material.

References

Barabási, Albert-Lászlo.2002. *Linked: The New Science of Networks*. Cambridge: Perseus Publishing.

Brown, Susan, et al. 2007. "Countless Links": Qualitative Query Potential in Orlando. Paper presented at *Chicago Colloquium on Digital Humanities and Computer Science* in Chicago, IL. October 2007.

Brown, Susan, Patricia Clements, and Isobel Grundy, ed. 2006. Orlando: Women's Writing in the British Isles from the Beginnings to the Present. Cambridge: Cambridge University Press Online. http://orlando.cambridge.org/.

Jessop, M. 2008. Digital Visualization as a Scholarly Activity. *Literary and Linguistic Computing* 23 (3): 281-293.

---

[6] Text Encoding Initiative. http://ww.tei-c.org/index.xml/