„To See or Not to See" – an Interactive Tool for the Visualization and Analysis of Shakespeare Plays

Thomas Wilhelm, Manuel Burghardt, and Christian Wolff

Media Informatics Group
University of Regensburg
D-93040 Regensburg
{firstname.lastname}@ur.de

Abstract

In this article we present a web-based tool for the visualization and analysis of quantitative characteristics of Shakespeare plays. We use resources from the *Folger Digital Texts Library*[1] as input data for our tool. The Folger Shakespeare texts are annotated with structural markup from the *Text Encoding Initiative* (*TEI*)[2]. Our tool interactively visualizes *which character* says *what* and *how much* at a particular *point in time*, allowing customized interpretations of Shakespeare plays on the basis of quantitative aspects, without having to care about technical hurdles such as markup or programming languages.

1   Introduction

Traditionally, humanities scholars have accessed and interpreted literary texts mainly using hermeneutic methods. With digital texts becoming increasingly available, these hermeneutic approaches may be supplemented by quantitative, computer-supported methods (cf. Moretti's *distant reading* (2007)). Although quantitative analysis of literary texts is an obvious approach with regard to the rise of digital resources and tools, many humanists wince at the available data sets and software. While Stone (1982, p. 300) noted the general belief that "it may be part of the humanistic tradition to be anti-machine", an increasing number of digital humanities projects[3]

---

[1] http://www.folgerdigitaltexts.org/; all hyperlinks in this article were last accessed on 25 March, 2013
[2] http://www.tei-c.org/Guidelines/P5/
[3] Cf. for instance the projects and publications of the European Association for Digital Humanities (ALLC): http://www.allc.org/

seems to refute this hypothesis. We believe that traditional humanists' skeptical and reluctant use of digital tools and resources can be explained by the substantial technical hurdles that come along with them.

In order to analyze digital texts beyond the level of mere character or word frequencies, the text needs to be annotated by means of a machine-readable markup language. The creation and analysis of such annotated texts is costly and time consuming (McLoughlin 2008, p. 10), as it requires profound technical knowledge of markup languages as well as programming languages, skills, that are not easily found with scholars from the humanities. Pierazzo (2011) describes the technical skills that have to be acquired in addition to the actual humanist skill set, illustrating the additional expenses that a humanities scholar has to face when trying to create an electronic edition of a text:

- Extensible Markup Language (XML)
- Text Encoding Initiative (TEI)
- Extensible Stylesheet Language Transformations (XSL/T)
- HyperText Markup Language (HTML)
- Cascading Style Sheets (CSS)
- various kinds of databases
- […]

However, there is a tendency to provide interfaces that hide technical complexity from the user and alleviate access to digital tools and resources for humanists. Among the most prominent examples for such interfaces are the *MONK Workbench*[4] and *Voyant*[5], both web-based tools that can be categorized as *rich-prospect browsers* (Ruecker et al. 2011, p. 2ff.), i.e. interfaces that initially display every item of a given collection, but also provide mechanisms for manipulating which specific items are displayed, with either more or less detailed information.

> "Rich-prospect browsers have the benefit of providing the user with a visual basis for understanding what is available in a collection. This kind of visual knowledge is particularly suitable for many collections of digital cultural objects, where a meaningful image of each item is readily available, and users may not be aware of everything that is in the collection." (Ruecker et al. 2011, p. 4)

---

[4] http://monkpublic.library.illinois.edu/monkmiddleware/public/index.html
[5] http://voyant-tools.org/

We present a user-friendly, web-based tool that makes use of freely available annotated texts (*Folger Shakespeare Library*) in the TEI format and provides a rich prospect browsing interface to the data, enabling scholars to access Shakespeare plays without further technical expertise.

2   Shakespeare in the Digital Age

With William Shakespeare being widely regarded as one of the greatest writers of all times, his texts have been transferred to the digital realm relatively early. Mullin (2003) gives a comprehensive synopsis on Shakespeare in the digital age: Shakespeare became *digital* in the 1960s, when a massive concordance of his complete works was created by a German research team by means of punch cards. This endeavor has evolved into a sophisticated long-term project, which today is known as the *Shakespeare Database* at the University of Münster[6]. While other projects focused on a hypertext approach to Shakespeare's work, multimedia pioneer Larry Friedlaender (mid-1980s) moved beyond the text level to include images and interactive features such as a virtual *Globe stage*, where users could place figures and share their setting with others via the Web.

More recent projects are dedicated to the quantitative analysis and visualization of digital Shakespeare texts. It is due to ambitious annotation projects such as the *Folger Digital Texts* initiative[7], which provides a growing collection of TEI-annotated Shakespeare plays, that we can study quantitative phenomena of Shakespeare beyond the level of mere word counts. The annotated data is freely available for analysis and can be shaped according to specific research needs by using the XML[8] transformation component XSL/T[9] or JavaScript commands that allow for the manipulation of the treelike XML/TEI document structure. Several tools try to hide such technical requirements by means of a graphical user interface (GUI): One example can be found with *MONK* (*Metadata Offer New Knowledge*)[10], a web-based environment that allows humanists to discover and analyze patterns in written texts. Among other texts, the *MONK Workbench* includes 37 plays and 5 works of poetry by William Shakespeare, which may be compared according to the frequency of certain words or features. Another digital

---

[6] http://www.uni-muenster.de/Shakespeare
[7] http://www.folgerdigitaltexts.org/fdt_documentation.pdf
[8] http://www.w3.org/TR/xml/
[9] http://www.w3.org/TR/xslt20/
[10] http://www.monkproject.org/

Shakespeare resource is the *Open Source Shakespeare* (*OSS*)[11], which offers keyword and advanced search (e.g. concordance search or character search) as well as basic statistics on word frequencies, speech and play lengths, genres and characters. Thiel (2009) builds on data provided by the *WordHoard project*[12] in order to visualize Shakespeare texts as huge posters in an exhibition context, enabling new reading experiences and interpretations of Shakespeare. Sinclair et al. (2011) present the *Mandala Browser*[13], a tool for the visualization of quantitative aspects of Shakespeare texts in typical classroom scenarios. Their tool can be used to answer simple questions as for instance "which character speaks the most" in a graphic way.

3 Building an Interactive Tool for the Visualization and Analysis of Shakespeare Texts

In contrast to the related work described above, our tool focuses on an interactive and flexible visualization of quantitative aspects of Shakespeare's plays. The interface was designed to provide access to these aspects in a user-friendly and intuitive way that enables new perspectives on Shakespeare. In order to avoid the so-called "designer as user problem" (Warwick, Terras et al. 2008), we interviewed a Shakespeare expert[14] before we started to build the actual tool. Together we have defined requirements and basic functionalities for an interactive tool that makes use of the TEI-annotated data from the *Folger library*. We decided that the tool should take account of the different characters of the plays as well as the discourse time (as opposed to narrative time), and provide corresponding statistical information about word frequencies (=length of acts, scenes and speeches). Following the concept of *rich-prospect browsing*, the interface serves as a synopsis of the whole play, which can be interactively explored in more detail along the timeline. The details of the implementation as well as a description of the basic functionality and interactive features are presented in this section.

---

[11] http://www.opensourceshakespeare.org/
[12] *WordHoard* is "An application for the close reading and scholarly analysis of deeply tagged texts" , cf. http://wordhoard.northwestern.edu/userman/index.html
[13] http://mandala.humviz.org/
[14] Our expert has a master's degree in English literature, and is currently writing a PhD dissertation on the productive reception of William Shakespeare from the 19th to the 21st century. We would like to thank Johannes Molz for his helpful advice. In addition, two of the authors of this paper hold a degree in English language and literature.

3.1 Source of Data: Folger Digital Texts

In December 2012, the *Folger Shakespeare Library* released twelve TEI-annotated plays as part of the *Folger Digital Texts* project (cf. Table 1):

(1) Hamlet	(7) Much Ado About Nothing
(2) Julius Caesar	(8) Othello
(3) King Lear	(9) Romeo and Juliet
(4) Macbeth	(10) Taming of the Shrew
(5) The Merchant of Venice	(11) The Tempest
(6) A Midsummer Night's Dream	(12) Twelfth Night

Tab. 1: List of TEI-annotated Shakespeare plays provided by Folger.

The digital texts are part of the *Folger edition*, which was created by the renowned Shakespeare scholars Barbara Mowat and Paul Werstine[15]. Folger claims to possess the world's largest Shakespeare collection, including copies of the *First Folio* from 1623, and its edition is widely used in classrooms across the United States. The annotated plays make use of the tags provided by TEI (e.g. *speech*, *dramatis personae*, *gender*, *stage directions*, etc.) and are freely available for further analysis and interpretation. Figure 1 shows the TEI-encoded source code of Hamlets famous soliloquy, taken from the Folger edition:

```
<sp xml:id="sp-1762" who="#HAMLET">
<speaker xml:id="spk-1762">
<w xml:id="w0271840">HAMLET</w>
</speaker>
<ab xml:id="ab-1762">
<lb xml:id="lb-17615"/>
<join type="line" xml:id="ftln-1762" n="3.1.64" ana="#verse" target="#w0271850
#c0271860 #w0271870 #c0271880 #w0271890 #c0271900 #w0271910 #c0271920 #w0271930
#c0271940 #w0271950 #p0271960 #w0271970 #c0271980 #w0271990 #c0272000 #w0272010
#c0272020 #w0272030 #p0272040"/>
<w xml:id="w0271850" n="3.1.64">To</w>
<c xml:id="c0271860" n="3.1.64"> </c>
<w xml:id="w0271870" n="3.1.64">be</w>
<c xml:id="c0271880" n="3.1.64"> </c>
<w xml:id="w0271890" n="3.1.64">or</w>
<c xml:id="c0271900" n="3.1.64"> </c>
<w xml:id="w0271910" n="3.1.64">not</w>
<c xml:id="c0271920" n="3.1.64"> </c>
<w xml:id="w0271930" n="3.1.64">to</w>
<c xml:id="c0271940" n="3.1.64"> </c>
<w xml:id="w0271950" n="3.1.64">be</w>
<!-- […] -->
```

Fig. 1: Excerpt from Hamlet's soliloquy (Hamlet, Act 3, Scene 1) taken from the TEI-encoded Folger Edition.

---

[15] http://www.folger.edu/Content/About-Us/Publications/Folger-Editions/

The basic tags that are interpreted by our tool describe the following characteristics of the play:

- *structural features*, such as words, lines, speeches, scenes and acts
- *situational features*, such as stage directions (enter, exit, etc.) or the death of a character
- *character information*, such as name, role and gender

These annotations allow to analyze the text at different levels of granularity, like for instance the whole play, a single act or multiple speeches of one character. Furthermore, complex references in almost any of the aforementioned entities explicitly connect different sections of the plays. These references are especially useful for automatic analysis and visualization of complex phenomena (cf. for instance the stage direction "All but Hamlet exit.").

3.2 Technical Approach and Implementation

In order to build a web application for visualizing XML-encoded Shakespeare plays we have transformed the data using XSL/T: These transformations specify *how* and *which features* of the input file (e.g. a person) should be displayed in the target format HTML. A simple example rule would be: *For each speech element in the input file, generate a corresponding HTML element containing its text and additional metadata in the output file*. The transformation rules are executed by a local Saxon XSL engine[16], which creates the specified HMTL output. As most current web browser implement an XSL 1.0 engine, it would be possible to provide the XSL/T-stylesheet and let users perform the transformation on their own computer. However, performance considerations (plays may have more than 5000 lines of code) as well as feature limitations of the browsers' XSL 1.0 engines[17] led to the use of an offline XSL transformation engine. Users are served the pre-generated XSL/T output when they access the tool. Since the input, in this case the TEI-encoded Shakespeare plays, rarely changes, this approach seems appropriate. Each play is transformed using the same set of rules defined by the XSL/T stylesheet. These rules may also be applied to plays that are released in the future, provided that Folger uses the same TEI-based tag sets. After the transformation, interaction behavior is added by means of JavaScript, while CSS is used to layout the data and place all the elements of the play relative to each other.

---

[16] Saxon Home-Edition 9.4, http://saxon.sourceforge.net/#F9.4HE
[17] Some functions needed to provide certain information are only available in XSL/T 2.0, e.g. *for-each-group*.

Figure 2 illustrates the transformation process from the static primary data to an interactive web page.
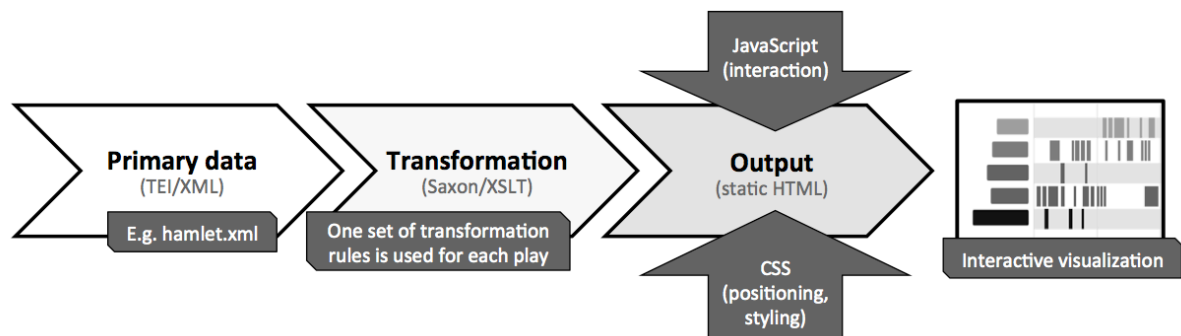

Fig. 2: The transformation process from static XML-data to an interactive web page.

All techniques used in the transformation process are based on W3C (CSS, HTML, XML, XSL/T) or ECMA (JavaScript) standards. Accordingly, these techniques are to be found at the core of most modern web applications, and are thus widely supported by current browsers.

3.3 Interface and Functionality: An Interactive and Visual Shakespeare Experience

We present a web-based tool[18] with a two-dimensional interface, where the two axes describe *characters* and *discourse time* (cf. Figure 3).
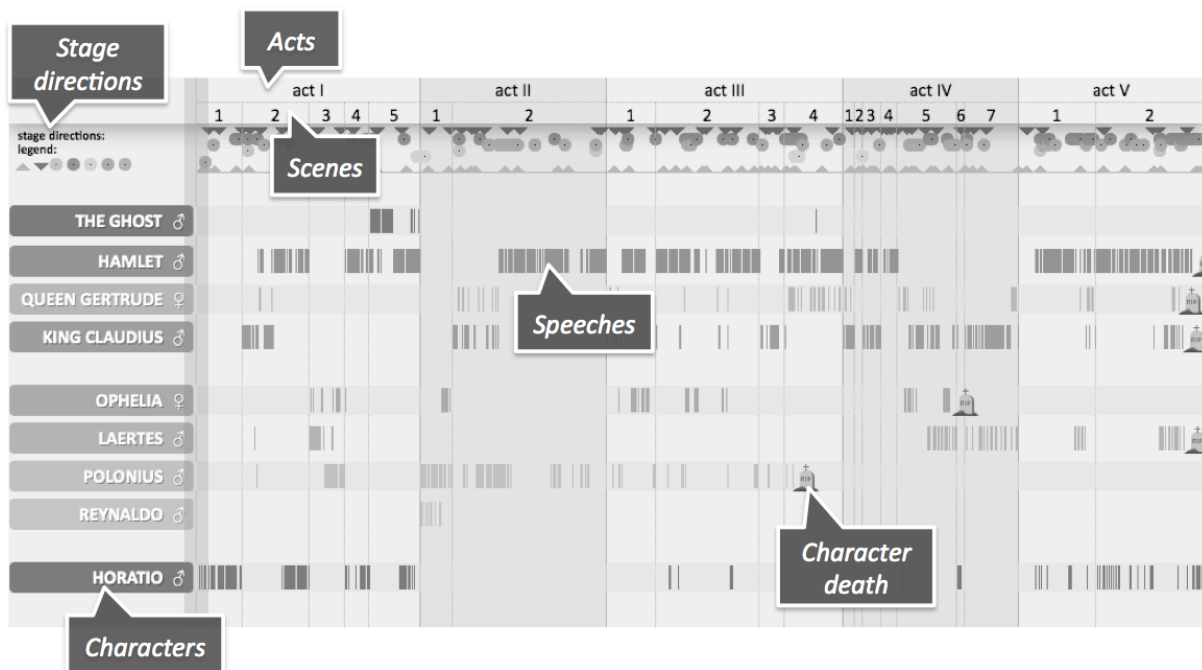

Fig. 3: Visual synopsis of Hamlet, displaying the characters' speeches and deaths as well as stage directions along a discourse timeline, which is based on text lines, scenes and acts.

---

[18] The tool is available as beta version at: www.thomaswilhelm.eu/shakespeare/

The *dramatis personae* (and respective gender and role information) are listed vertically on the left side of the screen, the timeline with different acts and scenes is displayed horizontally. The matrix gives a synopsis of the play, illustrating how much a character speaks (speeches are displayed as colored blocks) throughout the acts and scenes of the play. We use the number of text lines as a basic measure to implement a *discourse timeline*, and to position the speech elements accordingly. The same technique is used to visualize the play's structure by means of acts and scenes, to display stage directions and to visualize information about a person's death. The characters, stage directions, speeches, acts and scenes may be interactively explored to reveal more detailed information (cf. Figure 4). By selecting an *act* or a *scene*, statistics for lines, words and stage directions are displayed together with information about which characters are active in the current segment. The *speech* elements contain statistics for lines and words, information about foreign words, songs, poems or letters (if present in respective speech), and a representation of the actual text segment of the speech.
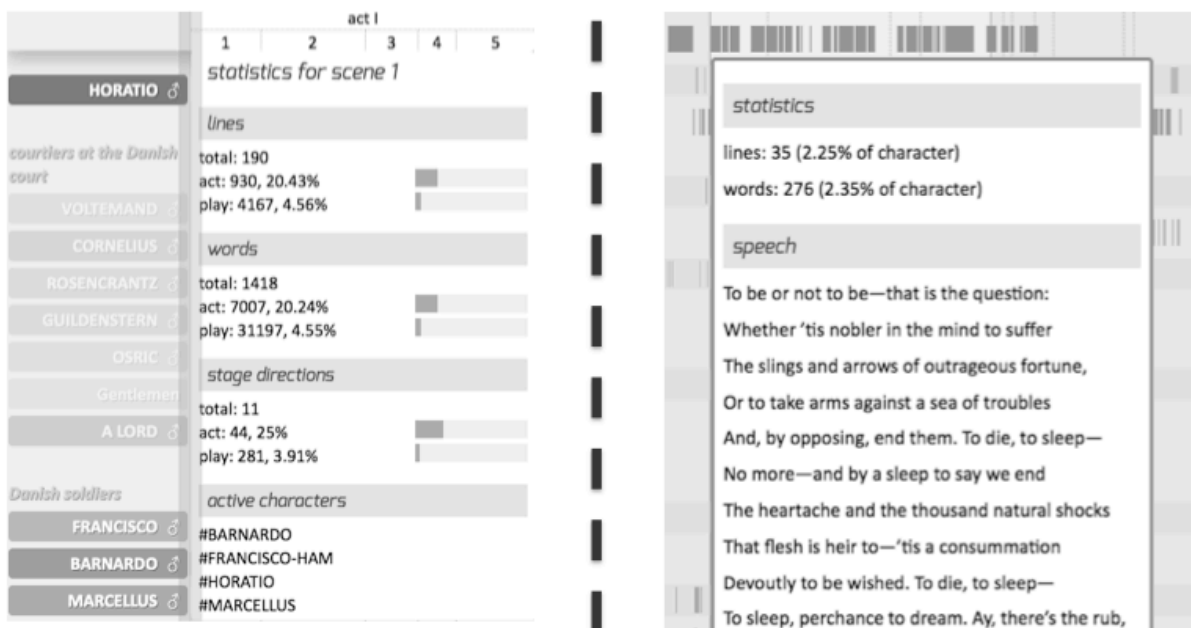

Fig. 4: Statistical details about the acts and scenes (left image) as well as the actual text of the speeches (right images) can be interactively explored by the users.

Users may also choose to use a slider component that can be dragged to any point along the timeline. The slider can also be controlled by using the arrow keys to navigate through the play speech by speech, creating a new way to explore Shakespeare texts.

3.4 Use Cases

Our tool is intended for a variety of use cases: Readers, who are new to Shakespeare, can quickly get an impression of the overall structure of a play and see which characters speak how much and in which scenes, as the interface can be interpreted as a *visual summary*. The tool could also be used in a classroom scenario, to reveal quantitative properties of a Shakespeare play, which may serve as the starting point for further interpretations. Another usage scenario might include stage directors, who are interested in the maximum number of active characters on stage at one point in time, to get a rough idea about how many different actors are needed for the production. Eventually, the tool may spark new approaches for interpretation on basis of the available quantitative aspects. Scholars can for instance easily examine relations between gender and length of speeches, or investigate how certain characters are distributed throughout the play.

4 Discussion and Outlook

Although the overall quality of the annotated Folger texts is fairly good, it is however important to note, that the data may contain small annotation errors, and might no be 100% correct. We are however confident, that the people at Folger will correct such flaws in future releases.

In its current stage of development, our Shakespeare tool is a proof of concept, that successfully uses the freely available dataset from Folger. It also provides interesting new perspectives on Shakespeare for different groups of users: While readers can experiment with an interactive way of exploring plays, scholars might find it helpful to have easy access to some basic statistics. The tool is however still under development, with a number of features yet to be implemented in future versions. Among those planned features is an export function for a selection of speeches, which may be investigated in more detail by using other tools, such as *Voyant*. Statistics will become more detailed in future versions, too. We are also thinking about alternative visualizations other than the character-timeline matrix, e.g. an interactive stage, where virtual characters present their speeches and carry out basic stage directions

Our next steps will be to evaluate the user interface for its actual usability as well as its suitability for the use cases described above. We are happy to receive any feedback about issues concerning the usability as well as the usefulness of our tool by the Shakespeare community.

Literature

McLoughlin, T. (2008). Bridging the Gap. In M. Rehbein & S. Ryder (Eds.), *Jahrbuch für Computerphilologie* (10th ed., Vol. 10, pp. 37–54).

Moretti, F. (2007). *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.

Mullin, M. (2003). Shakespeare on the Web. In L. Davis (Ed.), *Shakespeare Matters: History, Teaching, Performance* (pp. 119–137). Newark: University of Delaware Press.

Pierazzo, E. (2011). The Role of Technology in Scholarly Editing. Retrieved March 25, 2013, from http://epierazzo.blogspot.de/2011/10/role-of-technology-in-scholarly-editing.html

Ruecker, S., Radzikowska, M., & Sinclair, S. (2011). *Visual Interface Design for Digital Cultural Heritage*. Farnham et al.: Ashgate Publishing.

Sinclair, S., Ruecker, S., Gabriele, S., Patey, M., Gooding, M., Vitas, C., & Bajer, B. (2011). Meditating on a Mandala in Class: Studying Shakespeare′s Plays with a Visual Exploration Tool for XML Texts. *Media : Culture : Pedagogy*, *15*(1).

Stone, S. (1982). Humanities Scholars: Information Needs and Uses. *Journal of Documentation*, *38*(4), 292–313.

Thiel, S. (2009). Understanding Shakespeare. Towards a Visual Form for Dramatic Language and Texts. *B.A. thesis*. Retrieved March 27, 2013, from http://www.understanding-shakespeare.com/