

# Über die Nutzung von TagPies zur vergleichenden Analyse von Textdaten

**Jänicke, Stefan**

stjaenicke@informatik.uni-leipzig.de  
Leipzig University, Deutschland

**Efer, Thomas**

efer@informatik.uni-leipzig.de  
Leipzig University, Deutschland

**Blumenstein, Judith**

blumenst@rz.uni-leipzig.de  
Leipzig University, Deutschland

**Wöckener-Gade, Eva**

woeckener-gade@uni-leipzig.de  
Leipzig University, Deutschland

**Schubert, Charlotte**

schubert@uni-leipzig.de  
Leipzig University, Deutschland

**Scheuermann, Gerik**

scheuermann@informatik.uni-leipzig.de  
Leipzig University, Deutschland

## Motivation

Vor allem durch die häufige Nutzung in den sozialen Medien sind Tag Clouds heutzutage weit verbreitete Visualisierungen um den Inhalt textbasierter Daten zu veranschaulichen. Im Forschungsbereich der Informatik wurden zahlreiche Verfahren zur Berechnung von Tag Clouds entwickelt, unter anderem Wordle (Viégas et al. 2009). Abbildung 1 zeigt eine durch Wordle generierte Tag Cloud für die häufigsten Schlagworte aus den fünf Shakespeare Werken *As You Like It*, *Macbeth*, *Othello*, *Richard III* und *Romeo and Juliet*. Wie für Tag Clouds üblich, wird die Häufigkeit eines Wortes mit Schriftgröße kodiert. Leider tragen die weiteren visuellen Eigenschaften – Farbe, Position und Orientierung – der Darstellung keine Informationen. In unserer Posterpräsentation möchten wir TagPies vorstellen, ein im Rahmen des Digital Humanities Projektes eXChange entwickeltes Tag Cloud Layout. Während das Design von TagPies bereits vorgestellt wurde (Jänicke et al. 2015), sollen hier repräsentative Anwendungsszenarien illustriert werden. Ein TagPie kann als Hybrid aus Tag Cloud und

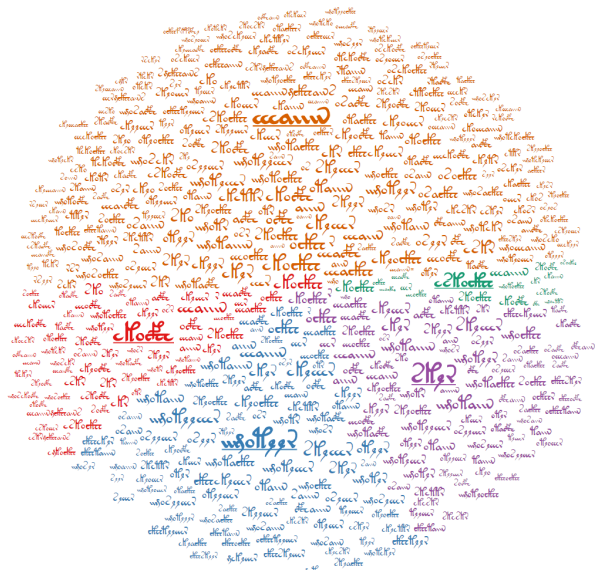
Pie Chart gesehen werden, bei dem mehrere einzelne Tag Clouds miteinander zu einer Einheit verschmelzen. Im Gegensatz zu herkömmlichen Tag Clouds verwenden TagPies neben Schriftgröße auch Farbe und Position als Informationsträger und ermöglichen damit den Vergleich von Schlagworten verschiedener Kategorien. Abbildung 2 zeigt einen TagPie für die häufigsten Schlagworte aus den oben genannten Shakespeare Werken. Wie in den folgenden Beispielen bestimmt die Größe einer Textklasse im Vergleich zu den anderen die Größe des zugehörigen Kuchenstückes im TagPie. Weitere Beispiele in den Abbildungen 3 und 4 veranschaulichen die sprachunabhängige, vielseitige Anwendbarkeit von TagPies (Jänicke 2015) – welche als Web-basierte Open Source JavaScript Bibliothek implementiert und damit projektunabhängig einsetzbar sind – für verschiedene geisteswissenschaftliche Fragestellungen.



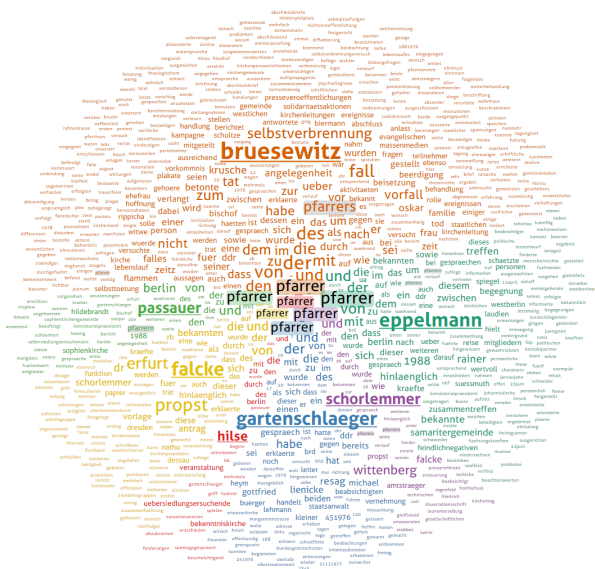
Abb. 1: Wordle Tag Cloud vereint Schlagworte aller Texte.



Abb. 2: TagPie ermöglicht den Vergleich von Schlagworten zwischen den Texten.



**Abb. 3:** Das Voynich-Manuskript ist ein mehr als einhundert Seiten umfassendes, in einer Geheimschrift verfasstes Werk (wahrscheinlich aus dem frühen 15. Jahrhundert), welches bis heute nicht entschlüsselt wurde und damit viele Wissenschaftler fasziniert. Die TagPie-Visualisierung ermöglicht die visuelle Exploration von Wort-Kontexten, welche neueren Analysen zufolge (Montemurro / Zanette 2013) interessante Muster aufweisen.



**Abb. 4:** Das Editionsprojekt “Die DDR im Blick der Stasi” (BStU 2016) stellt die geheimen Stimmungs- und Lageberichte der “Zentralen Auswertungs- und Informationsgruppe” (ZAIG) des MfS einer breiten Öffentlichkeit zur Verfügung. Der TagPie zeigt die Kookkurrenzwörter für die Nachnamen ausgewählter, besonders in den Fokus der Stasi geratener kirchlicher

Amtsträger. Weiterführende Informationen in (Kuras et al. 2014).

## Über die Nutzung von TagPies in eXChange

Das Digital Humanities Projekt eXChange untersucht den Gebrauch medizinischer und politischer Fachtermini in antiken griechischen und lateinischen Texten. Zur explorativen, vergleichenden Analyse sind TagPies in eine Rechercheplattform eingebunden, die den Geisteswissenschaftlern einen dynamischen Zugriff auf Textstellen ermöglicht, die eingegebene Suchterme enthalten. Jeder Suchterm wird innerhalb des TagPies durch einen Sektor repräsentiert, der seine häufigsten Kookkurrenzen anzeigt. Im Vergleich zu traditionellen Suchergebnislisten geben TagPies damit einen schnellen Überblick über die Kontexte, in denen die Suchterme verwendet wurden. Der TagPie ist interaktiv explorierbar und ermöglicht Close Reading durch die Selektion eines Schlagwortes via Mausclick, welches Textpassagen anzeigt, die sowohl den Suchterm als auch die gewählte Kookkurrenz enthalten. Im Folgenden sind zwei typische Anwendungsszenarien aus dem eXChange Projekt erläutert.

### Vergleich von gibbus und gibbosus

Für den Ausdruck “bucklig” findet man in lateinischen Wörterbüchern die Synonyme “gibbus” und “gibbosus”. Mit Hilfe von TagPies soll diese Synonymie überprüft werden (Abbildung 5); das Ergebnis berücksichtigt die Belegstellen aller Deklinationsformen der Suchterme. Die schwarz eingefärbte Schnittmenge im Zentrum liefert besonders auf den Körper bezogene Wörter wie “triefäugig” (lippus), “Fuß” (pede) oder “gebrochen” (fracto). Der grüne Sektor für “gibbus” zeigt weitere physische Begriffe wie “Rücken” (dorso), “Kopf” (caput) und “Gehirn” (cerebri). Der rote Sektor liefert die Resultate für “gibbosus”. Sie zeigen auffällig viele Begriffe aus dem Sachfeld der christlichen “Moral”, etwa “Begierde” (cupiditatis), “geizig” (avarum), “Mäßigung” (modestia) oder “sich rühmen” (glorietur). Der TagPie zeigt demnach eine Tendenz, dass „gibbus“ eher bei physischen Eigenschaften, „gibbosus“ eher bei moralischen Eigenschaften verwendet wurde (vgl. physisch: Cels. IV 1,5; VIII 1,23; Iuv. 6,109; 10,294; moralisch: christliche Exegeten). Ein Wörterbuch wie der Thesaurus Linguae Latinae differenziert die Semantik zwar auch in “eigentlich” und “übertragen“, aber über die Häufigkeit und Tendenz der Verwendung beider Begriffe gibt es im Vergleich zu den TagPies keine Hinweise. Damit übersteigen TagPies den “rekonstruktiven” Charakter von Wörterbüchern, indem sie auf die reale Verwendung in den Textkorpora rekurren.





*Die DDR im Blick der Stasi*. Die geheimen Berichte an die SED-Führung. <http://www.ddr-im-blick.de/> [letzter Zugriff 12. Februar 2016].

**Jänicke, Stefan** (2015): "TagPies", in: *vizcovery.org* <http://www.tagpies.vizcovery.org/> [letzter Zugriff 12. Februar 2016].

**Jänicke, Stefan / Blumenstein, Judith / Rücker, Michaela / Zeckzer, Dirk / Scheuermann, Gerik** (2015): "Visualizing the Results of Search Queries on Ancient Text Corpora with Tag Pies". To appear in *Digital Humanities Quarterly*.

**Kuras, Christoph / Efer, Thomas / Adam, Christian / Heyer, Gerhard** (2014): "The GDR Through the Eyes of the Stasi – Data Mining on the Secret Reports of the State Security Service of the former German Democratic Republic", in: Fred, Ana / Filipe, Joaquim (eds.): *Proceedings of the 6th International Conference on Knowledge Discovery and Information Retrieval (KDIR)*, Rom. Lisbon: SCITEPRESS 360–365.

**Jänicke, Stefan / Blumenstein, Judith / Rücker, Michaela / Zeckzer, Dirk / Scheuermann, Gerik** (2015): Visualizing the Results of Search Queries on Ancient Text Corpora with Tag Pies. To appear in *Digital Humanities Quarterly*, 2015.

**Montemurro, Marcelo A. / Zanette, Damián H.** (2013): "Keywords and Co-Occurrence Patterns in the Voynich Manuscript: An Information-Theoretic Analysis", in: *PloS one* 8, 6: e66344.

**Universität Leipzig Historisches Seminar** (2013): *eXChange - Exploring Concept Change and Transfer in Antiquity* <http://exchange-projekt.de/index.html> [letzter Zugriff 12. Februar 2016].

**Kuras, Christoph / Efer, Thomas / Adam, Christian / Heyer, Gerhard** (2014): "The GDR Through the Eyes of the Stasi – Data Mining on the Secret Reports of the State Security Service of the former German Democratic Republic", in: Fred, Ana / Filipe, Joaquim (eds.): *Proceedings of the 6th International Conference on Knowledge Discovery and Information Retrieval (KDIR)*, Rom. Lisbon: SCITEPRESS 360–365.

**Viégas, Fernanda B. / Wattenberg, Martin / Feinberg, Jonathan** (2009): "Participatory Visualization with Wordle", in: *IEEE Transactions on Visualization and Computer Graphics* 15, 6: 1137–1144.

**Montemurro, Marcelo A. / Zanette, Damián H.** (2013): Keywords and Co-Occurrence Patterns in the Voynich Manuscript: An Information-Theoretic Analysis. *PloS one*, 8, 6: e66344.

**Viégas, Fernanda B. / Wattenberg, Martin / Feinberg, Jonathan** (2009): "Participatory Visualization with Wordle", in: *IEEE Transactions on Visualization and Computer Graphics* 15, 6: 1137–1144.