

Über die Nutzung von *TagPies* zur vergleichenden Analyse von Textdaten

Stefan Jänicke,¹ Thomas Efer,² Judith Blumenstein,³
Eva Wöckener-Gade,³ Charlotte Schubert³ und Gerik Scheuermann¹

¹Abteilung für Bild- und Signalverarbeitung, Universität Leipzig

²Abteilung für Automatische Sprachverarbeitung, Universität Leipzig

³Fakultät für Geschichte, Kunst- und Orientwissenschaften, Universität Leipzig

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 [6]. Abbildung 1a 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* [3] vorstellen, ein im Rahmen des Digital Humanities Projektes *eXChange* [2] entwickeltes Tag Cloud Layout. 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 1b zeigt einen TagPie für die häufigsten Schlagworte aus den oben genannten Shakespeare Werken. Weitere Beispiele in Abbildung 2 und 3 veranschaulichen die sprachunabhängige, vielseitige Anwendbarkeit von TagPies für verschiedene geisteswissenschaftliche Fragestellungen.

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

Abbildung 3: Das Editionsprojekt “DDR im Blick” stellt die geheimen Stimmungs- und Lageberichte der “Zentralen Auswertungs- und Informationsgruppe” (ZAIG) des MfS einer breiten Öffentlichkeit zur Verfügung [1]. Der TagPie zeigt die Kookkurrenzwörter für die Nachnamen ausgewählter, besonders in den Fokus der Stasi geratener kirchlicher Amtsträger (weiterführende Informationen in [4]).



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.

1. 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 4).¹ 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) (Abbildung 4a), “Kopf” (caput) und “Gehirn” (cerebri). Der rote Sektor liefert die Resultate für “gibbosus”. Sie zeigen auffällig viele Begriffe aus dem Sachfeld der christlichen “Moral” (Abbildung 4b), 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.² 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.

2. Vergleich von $\alpha\rho\epsilon\mu^*$, $\eta\sigma\chi^*$ und $\alpha\lambda\iota\eta\tau^*$

Für die datenbankbasierte Suche wurden die trunkierten Formen $\alpha\rho\epsilon\mu^*$, $\eta\sigma\chi^*$ und $\alpha\lambda\iota\eta\tau^*$ – jeweils Synonyme für “ruhig” bzw. “unbewegt” – verwendet. Weiter sollte in der vergleichenden Analyse untersucht werden, ob $\alpha\rho\epsilon\mu^*$ als medizinischer Fachterminus zu bezeichnen ist. Der resultierende TagPie (Abbildung 5) verdeutlicht aufgrund der vielen gemeinsamen Kookkurrenzen der Begriffe, gekennzeichnet durch die unterstrichenen Terme im mittleren Bereich, dass die drei Wortgruppen in der Literatur häufig synonym verwendet werden. $\eta\sigma\chi^*$ liefert mit großem Abstand die meisten Belege und ist in den verschiedensten Kontexten zu finden. Für $\alpha\lambda\iota\eta\tau^*$ gibt es viele Kookkurrenzen mit $\kappa\iota\omega\upsilon\tau$ (bewegen), zurückzuführen auf die aristotelische Philosophie, in der der Gegensatz zwischen Unbewegtheit und Bewegtheit eine große Rolle spielt (Abbildung 5a). Da verschiedene medizinische und anatomische Begriffe ausschließlich in den Kontexten von $\alpha\rho\epsilon\mu^*$ auftreten, kann man $\alpha\rho\epsilon\mu^*$ als medizinischen Fachterminus bezeichnen. Beispiele hierfür sind Formen von $\delta\iota\alpha\phi\omega\rho\eta\tau\iota\kappa\omicron\varsigma$ (“schweißtreibend” oder “streuend”, u.a. von Tumoren und Medikamenten) oder schlicht $\eta\tau\omicron\rho$ “Herz”. Die entsprechenden Belegstellen finden sich dann u. a. in den medizinischen Schriften des Galen und

¹Das Ergebnis berücksichtigt die Belegstellen aller Deklinationsformen der Suchterme.

²vgl. physisch: Cels. IV 1,5; VIII 1,23; Iuv. 6,109; 10,294; moralisch: christliche Exegeten

Oribasius (Abbildung 5b). Weitere Analysen der einzelnen Belegstellen, aber auch weitere Visualisierungen mithilfe der TagPies (z. B. mit Eingrenzung auf das Corpus Hippocraticum) sind geplant, um die qualitative Auswertung der Ergebnisse fortzuführen.

Zusammenfassung

Die *Distant Reading* Visualisierung TagPies ermöglicht den Vergleich verschiedener Kategorien von Schlagworten innerhalb einer Tag Cloud. Im Rahmen des Projektes eXChange hat sich die Verwendung von TagPies zur vergleichenden Analyse von Fachtermini in antiken Texten als wertvoll erwiesen. In einer Posterpräsentation würden wir typische Anwendungsszenarien der beteiligten Geisteswissenschaftler demonstrieren.

Literatur

- [1] DDR im Blick, 2015. <http://www.ddr-im-blick.de/> (Aufgerufen am 9. September 2015).
- [2] eXChange: Exploring Concept Change and Transfer in Antiquity, 2015. <http://exchange-projekt.de/> (Aufgerufen am 8. September 2015).
- [3] S. Jänicke, J. Blumenstein, M. Rücker, D. Zeckzer, and G. Scheuermann. Visualizing the Results of Search Queries on Ancient Text Corpora with Tag Pies. *Digital Humanities Quarterly*, 2015. To appear.
- [4] C. Kuras, T. Efer, C. Adam, and G. Heyer. 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 A. Fred and J. Filipe, editors, *Proceedings of the 6th KDIR*, pages 360–365. INSTICC, SCITEPRESS – Science and Technology Publications, October 2014.
- [5] M. A. Montemurro and D. H. Zanette. Keywords and Co-Occurrence Patterns in the Voynich Manuscript: An Information-Theoretic Analysis. *PLoS one*, 8(6):e66344, 2013.
- [6] F. Viégas, M. Wattenberg, and J. Feinberg. Participatory Visualization with Wordle. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1137–1144, Nov 2009.

