

TagStar: ein interaktives Indexierungs- und Analysewerkzeug

Mirko de Almeida Madeira Clemente¹, Mandy Keck¹, Rainer Groh¹

Technische Universität Dresden, Fakultät Informatik, Professur für Mediengestaltung¹

Zusammenfassung

Dieser Beitrag beschreibt das interaktive Visualisierungskonzept *TagStar*, welches der computergestützten und kollaborativen Erschließung von Visualisierungen und der Analyse des daraus resultierenden Datenbestandes dient. Die Visualisierungen können durch Hinzufügen von Schlagwörtern eines zugrundeliegenden Klassifikationsschemas klassifiziert und semantisch beschrieben werden. Unterstützung bieten Schlagwortempfehlungen und eine Icon-basierte Darstellungstechnik.

1 Einleitung

Im Bereich der Informationsvisualisierung werden laufend neue Darstellungen generiert, die den Betrachtern erkenntnisreiche Einblicke in komplexe Daten bieten können. In Abhängigkeit der zu visualisierenden Daten finden die verwendeten Darstellungstechniken ständig neue Variationen, die durch den Einsatz geeigneter Interaktionsmechanismen zu effizienten Werkzeugen der Datenanalyse und -exploration werden. Das Such- und Analysewerkzeug *DelViz (Deep exploration and lookup of Visualizations)*¹ bietet die Möglichkeit einen Datenbestand diverser Visualisierungen aus unterschiedlichen Blickwinkeln zu betrachten (Keck et al. 2011). Ermöglicht wird dies durch ein multidimensionales Klassifikationsschema, das eine Sammlung von vordefinierten Schlagwörtern durch mehrere Dimensionen strukturiert. Die Schlagwörter dienen der Indexierung und können zur Beschreibung der Visualisierungen verwendet werden. Um verschiedenartige Meinungsbilder zu erfassen, kann eine Visualisierung – im Folgenden „Ressource“ genannt – auch von mehreren Nutzern indexiert werden. Das Schema vereint somit Prinzipien des Social Tagging mit denen einer Expertenklassifikation, wodurch ein Spannungsfeld zwischen gegenseitigen Polen wie *frei* und *kontrolliert* oder *Laie* und *Experte* entsteht. Aus Sicht des Experten ist die Anwendung des Klassifikationsschemas durch die Nutzer von großer Bedeutung, um eine Ressource mit den bereits im

¹ Die Anwendung kann unter <http://delviz.com> aufgerufen werden.

System enthaltenen vergleichen zu können. Aus Sicht des Nutzers, der beim Social Tagging in der Regel keinen Konventionen unterliegt, stehen die für das Tagging charakteristischen Merkmale wie Schnelligkeit und minimaler Aufwand an oberster Stelle. Eine erfolgreiche Anwendung des Klassifikationsschemas, aus der sowohl die Experten als auch die Nutzer profitieren, stellt somit eine Herausforderung dar.

In diesem Beitrag wird das interaktive Visualisierungskonzept *TagStar* vorgestellt, welches dieses Problem adressiert und sowohl Nutzer als auch Experten bei der Erweiterung, Analyse und Pflege des Datenbestandes unterstützt. Es ist von klassischen Systemen zur Indexierung oder Systemen zur kollaborativen Klassifikation (*Social Tagging Systems*) wie *FaceTag* (Quintarelli et al. 2008) oder *TagMe!* (Abel et al. 2009) insofern abzuheben, dass die entwickelte Darstellungstechnik den Beschreibungsumfang und die -varianz der indexierten Ressourcen visualisiert und die Interaktion unmittelbar auf der grafischen Darstellung stattfindet.

2 Visualisierungskonzept

Für jede in der Datenbank befindliche Ressource können Detailinformationen wie Titel, Vorschaubild und ein weiterführender Link aufgerufen werden. Im Fokus des Konzepts stehen allerdings die Schlagwörter, die eine Ressource semantisch beschreiben. Durch die 55 Schlagwörter des Klassifikationsschemas werden ebenso viele beschreibende Dimensionen aufgespannt, welche den Wertebereich der positiven, ganzen Zahlen abdecken. Ordnet beispielsweise ein Nutzer einer Ressource das Schlagwort „Netzwerk“ zu und ein zweiter Nutzer derselben Ressource die Schlagwörter „Netzwerk“ und „Hierarchie“, so ist der Wert der ersten Eigenschaft gleich 2 und der letzten gleich 1.

Als Grundlage zur Visualisierung dieses Datenbestandes wird auf die *Star Plots* zurückgegriffen, da sich die Icon-basierte Visualisierungstechnik besonders für multidimensionale, diskrete Daten eignet (vgl. Ware 2004, S. 184). Auf diese Weise werden alle Attribute auf konzentrisch angeordnete Linien, deren Länge den jeweiligen Attributwerten entsprechen, abgebildet und alle Linienenden miteinander verbunden. Durch die gemeinsame Anordnung vieler Glyphen entstehen Texturmuster, welche die Eigenschaften und Charakteristiken der Daten widerspiegeln und einen effizienten Vergleich von Attributen und Datensätzen ermöglichen. Aufbauend auf der Visualisierungstechnik der *Star Plots* wurden drei verschiedene Visualisierungskonzepte entworfen, die im Folgenden näher beschrieben werden:

Die Nutzeransicht stellt eine Ressource im Detail dar und bietet dem Nutzer die Möglichkeit, eine hinzugefügte Ressource erstmals zu indexieren oder sein Meinungsbild auf eine bereits vorhandene Ressource zu übertragen. Unmittelbar nach dem Bereitstellen einer Ressource verfügt diese noch über keine Schlagwörter. Das Interface visualisiert in dieser Ausgangssituation lediglich die vom System ermittelten Schlagwortempfehlungen, die durch das im Zentrum des Interfaces (Glyph) befindliche „Empfehlungspolygon“ abgebildet werden. Andernfalls wird die Glyph auf Basis der zur Beschreibung vergebenen Schlagwörter konstruiert (vgl. Abbildung 1, links). Alle weiteren Schlagwörter des Klassifikationsschemas werden zunächst ausgeblendet und erst durch Interaktion des Nutzers (RollOver) eingeblen-

det. Zur Verdeutlichung der Datenstrukturen werden neben den Schlagwörtern auch die Dimensionen des Klassifikationsschemas durch grafische Elemente, den verschiedenfarbigen Kreissegmenten, repräsentiert. Um der Anforderung gerecht zu werden, den Datenbestand möglichst umfangreich zu indexieren, wird der Nutzer durch eine Konturlinie unterstützt. Diese verbindet benachbarte Kreissegmente miteinander, sobald aus diesen mindestens je ein Schlagwort zur Beschreibung der Ressource ausgewählt wurde. Mit jeder erneuten Schlagwortauswahl ähnelt die Konturlinie zunehmend einem geschlossenen Kreis. Dadurch soll der Nutzer angehalten werden, mindestens ein Schlagwort je Dimension zu vergeben und zudem ein Gefühl der Vollständigkeit suggeriert werden.

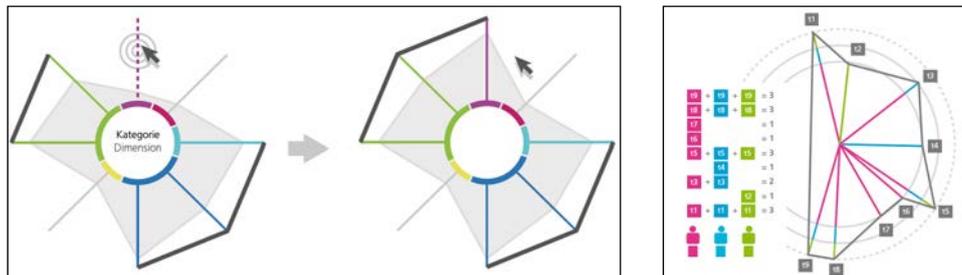


Abbildung 1: Konzept der Nutzeransicht: Glyphе als Grundlage des Indexierungs-Interfaces, das Empfehlungspolygon in der Mitte unterstützt den Nutzer bei der Auswahl eines Schlagwortes (links). Konzept der Experten-Detailansicht: die äußere Form der Glyphе spiegelt die Homogenität der ausgewählten Schlagwörter wider (rechts).

Die Experten-Detailansicht dient der genaueren Untersuchung von Ursachen, die für abweichende Merkmalsausprägungen verantwortlich sind. Der Unterschied zur Nutzeransicht besteht darin, dass der Radius der Glyphе um die Anzahl der an dem Indexierungsprozess beteiligten Nutzer erweitert wird (vgl. Abbildung 1, rechts). Haben beispielsweise drei Nutzer das gleiche Schlagwort an eine Ressource vergeben, so vergrößert sich der Wert der jeweiligen Komponente innerhalb des zugrundeliegenden Merkmalsvektors von 1 auf 3. Durch die Abbildung der Vektoren auf die Länge der konzentrisch angeordneten Linien, folgt die Kontur keinem einheitlichen Radius mehr. Anhand der äußeren Form lässt sich nun effizient erschließen, ob eine Ressource von allen Nutzern mit denselben Schlagwörtern beschrieben wurde oder ob es Abweichungen gibt.

Die Experten-Übersicht dient als ein Analysewerkzeug für die Experten. Es visualisiert die Schlagwörter einer großen Anzahl an Ressourcen gleichzeitig, wobei jede durch eine Glyphе repräsentiert wird (vgl. Abbildung 2, rechts). Durch Aneinanderreihung der Glyphen entsteht eine Textur, welche die Beurteilung des Beschreibungsumfangs und Feststellung von abweichenden Nutzermeinungen (Varianz) ermöglicht. Je kleiner die Glyphen dargestellt werden, desto stärker verschmelzen die Schlagwörter einer Dimension. Anstatt der Entdeckung einzelner Schlagwörter nachzugehen, hebt diese Unschärfe die Arbeit des Experten auf ein höheres Abstraktionsniveau. Ressourcen, die bisher nicht indexiert wurden, werden durch Punkte repräsentiert und treten durch ihr abweichendes Erscheinungsbild deutlich hervor. Dies gilt ebenso für Ressourcen mit nur sehr wenigen Schlagwörtern. Diese charakteristischen Erscheinungsbilder können dafür verwendet werden, die Beschreibung des Datenbestandes systematisch zu verbessern und Problemstellen aufzudecken. Da eine geringe Anzahl

an Schlagwörtern nicht mit der unvollständigen Indexierung einer Ressource gleichzusetzen ist, bietet dieses Vorgehen gegenüber dem Zählen von Schlagwörtern einen wesentlichen Vorteil. Die Entscheidung darüber, ob eine Ressource ausreichend indexiert wurde, wird nicht vom System sondern von den Experten getroffen. Deren Interpretation beruht auf Erfahrungswerten, die sich aus dem Umgang mit der Darstellungstechnik ergeben. Außerdem können sie auf typische Merkmale sensibilisiert und trainiert werden.

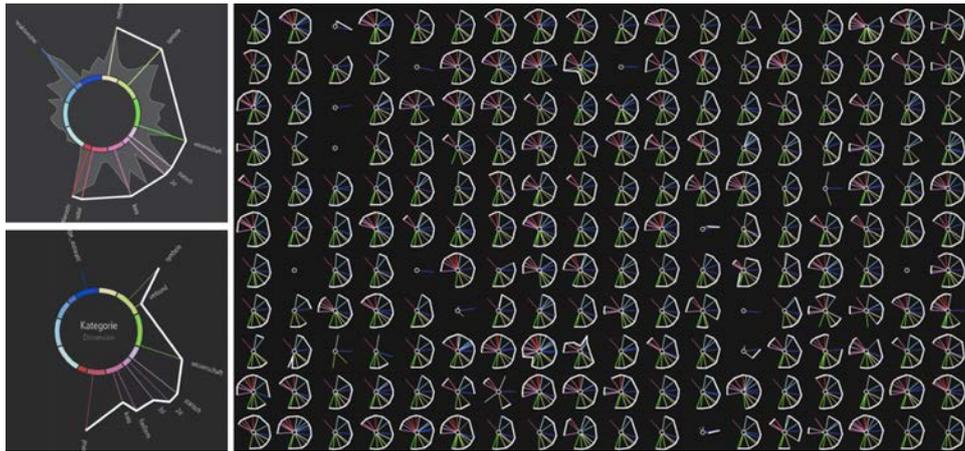


Abbildung 2: Umsetzung der Nutzeransicht (l. o.), der Experten-Detailansicht (l. u.) und der Experten-Übersicht (r.)

3 Zusammenfassung und Ausblick

Mit dem vorgestellten Visualisierungskonzept kann die Erweiterung, Pflege und Analyse des Datenbestandes in unterschiedlichen Nutzermodi vorgenommen werden. Die Sichten können der Abbildung 2 entnommen werden. Die Erweiterung des Klassifikationsschemas durch frei gewählte Schlagwörter stellt einen Aspekt für zukünftige Arbeiten dar.

Literaturverzeichnis

- Abel F., Kawase R., Krause, D. & Siehdnel, P. (2009). *Multi-faceted Tagging in TagMe!*. In 8th International Semantic Web Conference (ISWC '09).
- Keck, M., Kammer, D., Iwan, R., Taranko, S. & Groh, R. (2011). DelViz: Exploration of tagged information visualizations. *Informatik 2011 - Interaktion und Visualisierung im Daten-Web*. Berlin.
- Quintarelli, E., Resmini, A. & Rosati, L. (2008). *Facetag: Integrating Bottom-up and Top-down Classification in Social Tagging Systems*.
- Ware, C. (2004). *Information visualization: perception for design*. San Francisco: Morgan Kaufmann.

Kontaktinformationen

Technische Universität Dresden, Fakultät Informatik, Institut für Software- und Multimedia-technik, Professur für Mediengestaltung, 01062 Dresden. mirko.clemente@tu-dresden.de