

Inhaltsbasierte Tag-Vorschläge in Word-Dokumenten

Björn Joop, Oliver Hetzel, Jürgen Ziegler

Universität Duisburg-Essen

Zusammenfassung

Tagging-Systeme versehen ganze Dokumente und Ressourcen mit Tags. Diese Tags können für das Wiederauffinden von Dokumenten oder für die Empfehlung von Tags verwendet werden. Tag-Empfehlungssysteme setzen jedoch häufig voraus, dass entweder andere Benutzer das zu taggende Dokument bereits mit Tags versehen haben, Tags vorgegebenen Themen und Kategorien zugeordnet werden können oder aber extrahierte Schlüsselwörter als Tags repräsentativ für ein ganzes Dokument sind. Dieser Beitrag präsentiert ein System, welches uns ermöglicht inhaltsbezogene Tags zu generieren und an beliebigen Stellen in Word-Dokumenten anzubringen. In einer Studie haben wir das Tagverhalten der Probanden und die getaggt Textstellen untersucht. Zusätzlich wurde die Suche von Informationen mithilfe Tag-Suche betrachtet.

1 Einleitung

Die Organisation und Verwaltung von Informationen wird durch immer größere Datenmengen zunehmend schwerer. Abhilfe schaffen hier komplexe Indexierungsmethoden die häufig nur von Experten zu bedienen sind. Dem entgegen stehen Tagging-Systeme – insbesondere bei Web 2.0 Anwendungen – in denen eine große Anzahl von Personen gemeinsame oder persönliche Ressourcen mit frei wählbaren Begriffen (sogenannten: Tags) versehen. Das Prinzip des Taggings wird darüber hinaus auch von Desktopanwendungen, die Inhalte über die Zuordnung von Schlagwörtern organisieren, verwendet, um so die Suche nach Informationen zu erleichtern. Getaggt werden jedoch häufig nur vollständige Objekte wie Bilder oder Textdokumente.

Insbesondere große Textdokumente können durch ein inhaltsbezogenes Tagging innerhalb des Dokuments das Auffinden von Informationen erleichtern. Jedoch entwickeln Benutzer von Tagging-Systemen über die Zeit ein stark personalisiertes Vokabular (vgl. Wetzker et al. 2010), welches den Austausch von und die Suche in Dokumenten erschwert.

In diesem Beitrag stellen wir ein Tagging-System für Word-Dokumente vor, das es den Benutzern ermöglicht, sowohl ganze Dokumente wie auch Dokumentteile beliebig zu

taggen. Das System stellt eine Tag-Empfehlungsfunktion zur Verfügung, die auf Basis von selektierten Textstellen eine Schlüsselwortextraktion und Synonymsuche durchführt, um Begriffe als mögliche Tag-Kandidaten zu identifizieren und dem Benutzer zu präsentieren. Zusätzlich besteht die Möglichkeit, auf Basis von vergebenen Tags Informationen in einem oder mehreren Dokumenten gezielt zu suchen. Wir verbinden in dem vorgestellten System automatisierte Tagging-Verfahren mit Tag-Empfehlungsprozessen und explorieren die Möglichkeit feingranular auf beliebigen Dokumentstrukturen Tags zu vergeben.

Hierfür untersuchen wir in einer ersten Studie das Verhalten der Benutzer bei der Verwendung des Systems. Wir achten insbesondere auf die Begriffe, die für das Tagging verwendet werden und vergleichen diese mit den vom System generierten Vorschlägen. Zusätzlich untersuchen wir, welche Dokumentteile die Probanden bei der Verwendung des Systems präferiert taggen. Zuletzt vergleichen wir in dieser Studie die Tag-basierte Suche mit einer Textsuche für das Auffinden von Informationen in einem oder mehreren Dokumenten.

2 Stand der Forschung

Tag-Empfehlungssysteme versuchen neue Ressourcen oder neue Benutzer in der Wahl der Tags zu unterstützen. Hierfür orientieren sie sich an Tags anderer Benutzer, die gleiche oder ähnliche Ressourcen getaggt haben. Ein Beispiel für so ein System ist das Social Bookmarking System *Bibsonomy* (Jäschke et al. 2007, Jäschke et al. 2009), in dem neue Tags entweder auf Basis von bereits getaggtten Objekten mithilfe von Collaborative Filtering und Graph-basierten Techniken, oder in Form von besonders populären Tags empfohlen werden. *AutoTag* (Mishne 2006) empfiehlt Tags für Blog-Einträge auf Basis von ähnlichen Einträgen. Sood et al. (2007) verfahren mit ihrem System *TagAssist* ähnlich, vermeiden jedoch Ambiguitäten von Tags durch die Reduzierung auf ihre Stammform und den Abgleich der Begriffe mit Wörterbüchern und Thesauri. *SparTag.us* (Hong et. al. 2008) hingegen ermöglicht es Benutzern Webdokumente an Paragraphen oder Satzteilen zu taggen.

Einen anderen Ansatz verfolgen (Medelyan et al. 2009). Sie kombinieren eine Schlüsselwortextraktion mit semantischen Informationen, um Tags für Dokumente vorzuschlagen. (Song et. al. 2008, Song et. al. 2011) erlenen Assoziationen zwischen Tags und Inhalten (Wörter) von Dokumenten um den Benutzern passende Tags zu empfehlen. (Heymann et al. 2008) extrahieren Schlüsselwörter aus Webdokumenten und verwenden ein Entropie-basiertes Ranking um Tag-Kandidaten zu identifizieren. Das System *TagHats* (Nishida & Fujimura 2010) verwendet Dokumente aus Question & Answers Communities, um ihnen automatisch Tags zuzuordnen. Mithilfe einer Schlüsselwortextraktion werden Begriffe identifiziert und auf Kategorien, Themen oder Schlüsselwörter abgebildet. Alle Ansätze (außer SparTag.us) haben gemein, dass sie sich auf das Tagging von ganzen Dokumenten konzentrieren.

Tag-basierte Empfehlungssysteme versuchen hingegen auf Basis von vergebenen Tags Objekte und Inhalte zu empfehlen. (Shepitsen 2008) und (Guo & Joshi 2010) clustern Tags und Objekte auf der Suche nach Themengruppen umso mithilfe eines Benutzerprofils Ressourcen

zu empfehlen. (Tso-Sutter et al. 2008) und (Wetzker et al. 2009) integrieren Tags in Item- und User-basierte Collaborative Filtering Mechanismen um zusätzliche Informationen für eine höhere Präzision des Empfehlungsprozess verwenden zu können. (Cantador et al. 2010) untersuchen in ihrer Arbeit die Qualität von inhaltsbezogenen Empfehlungen unter Verwendung von Tagprofilen für Benutzer und Objekte. Die Suche von Informationen kann mithilfe von Tags ebenfalls verbessert werden (vgl. Bischoff et al. 2008). (Carmel et al. 2010) gewichten Tags von Dokumenten für eine Suchoptimierung. (Zanardi & Capra 2008) hingegen erstellen ein soziales Ranking der betroffenen Ressourcen auf Basis der Beziehung von Tags und Ressourcen unter- und zueinander.

3 Systemaufbau

Unser System ist ein Addon für Microsoft Office Word und ermöglicht es Tags an beliebigen Textstellen in Dokumenten anzubringen. Das Addon besteht im Wesentlichen aus zwei Kernkomponenten: Tag-Empfehlung und einer Tag-basierten Suche. Die Tags werden im *Office Open XML*¹ (.docx-Format) als Custom-XML in Form einer Ontologie gespeichert.

3.1 Tag-Empfehlung

Die Tag-Empfehlung basiert auf einer Schlüsselwort-Extraktion. In einer Vorverarbeitungsphase wird ein selektierte Text in einzelne Wörter zerlegt und anschließend von Sonderzeichen und Stopwörtern befreit, da diese vorwiegend grammatikalische oder syntaktische Funktionen haben. Durch einen Wörterbuchvergleich werden Begriffe, die nicht eindeutig zuzuordnen sind, identifiziert und mithilfe eines Porter-Stemmers auf die Stammform reduziert. Als Anhaltspunkt für die Relevanz der Wörter in Bezug auf den selektierten Abschnitt wird die relative Häufigkeit ihrer Vorkommen berechnet.

Für die Spezifität der Terme des ausgewählten Dokumentteils wird bei der Generierung der Tag-Empfehlungen ein modifiziertes TF-IDF-Gewicht eingesetzt. Anhand der Vorkommen der Tag-Kandidaten im Rest des Dokuments kann das Gewicht:

$tfidf_t = \frac{\log(\frac{w}{t_D})}{\log(w)} * \frac{sel_t}{\max_{s \in C}(sel_s)}$ für ein Term t berechnet werden. Der erste Faktor berechnet den

IDF-Wert der nicht-selektierten Wörter w im Rest des Dokuments und des Vorkommen t_D des Terms t im nicht-selektierten Teil des Dokuments. Der zweite Faktor beschreibt die relative Termhäufigkeit der zu indexierenden Textpassage sel_t . Als Bezugsgröße wird die maximale Termhäufigkeit aller Tag-Kandidaten C aus dem Gesamtdokument herangezogen.

Wenn einer der Tag-Kandidaten bereits als Schlagwort im Dokument vorkommt, dann werden der Tag und zusätzlich alle Tag-Kandidaten der betroffenen Textstellen als potentiell relevant angesehen. Für jeden Tag-Kandidaten wird die Menge der bedeutungsgleichen Wör-

¹ <http://www.microsoft.com/germany/interop/openxml/> (Stand: 28.03.2011)

ter ermittelt und mit den bisher verwandten Tags im Dokument abgeglichen: Verwendet der Benutzer bereits Synonyme im aktuellen Dokument, so werden diese zusätzlich in die Liste der Tag-Kandidaten aufgenommen. Die vorgeschlagenen Tags werden in Form einer Tagwolke visualisiert.

3.2 Tag-basierte Suche

Der direkte Zugriff auf die Tags eines Dokuments erfolgt über eine Liste im Aufgabenbereich der Nutzerschnittstelle. Über die Auswahl von Tags lassen sich so entweder die als Tags verwandten Begriffe im Dokument markieren oder die mit den Tags annotierten Textpassagen hervorheben. Optional steht eine Ausblendung der nicht-selektierten Tagstellen zur Verfügung. Für eine gezielte Suche nach Informationen in getaggtten Dokumenten stehen zwei Funktionen zur Verfügung: es können sowohl Tags innerhalb eines Dokuments oder in mehreren getaggtten Dokumenten gesucht werden.

Die Suche in Einzeldokumenten verwendet Texteingaben des Benutzers als Suchbegriffe. Bei der Suche in mehreren Dokumenten werden relevante Such-Tags mithilfe einer Tagwolke und einer Filterfunktion ausgewählt. Eine Einschränkung auf Dokumentinhalte wie Text, Tabellen, Aufzählungen oder Bilder ist bei beiden Suchen möglich.

Die Suchergebnisse werden in Form einer Liste mit Angabe über die Art der getaggtten Dokumentstelle (Bild, Text, Tabelle oder Liste) aufgelistet. Die Suche in mehreren getaggtten Dokumenten erhält zusätzliche Informationen in der Ergebnisliste: Ein Indikator zeigt die Ausprägung der Überschneidung zwischen der Passage und der Anfrage anhand der Anzahl der enthaltenen Such-Tags relativ zur Gesamtheit aller ausgewählten Tags in der gesammelten Tagmenge der Dokumente dar. Zusätzlich wird bei Textstellen der Beginn des getaggtten Texts angezeigt.

4 Evaluation

Um die Akzeptanz der vorgeschlagenen Tags und den Nutzen der Tagfunktion in Word-Dokumenten zu untersuchen wurde eine experimentelle Studie durchgeführt. An der Studie nahmen insgesamt 21 Probanden im Alter von 23 bis 35 Jahren teil (15 männlich, 6 weiblich). Die Probanden gaben an, keine Taggingssysteme zu verwenden. Jedoch war bis auf zwei Teilnehmern das Prinzip des Taggens bekannt. Eine Kontrollgruppe umfasste 9 Personen (8 männlich, 1 weiblich) im Alter von 23-33 Jahren.

Folgende Fragen sollten mithilfe der Studie geklärt werden:

1. Verwenden Benutzer mehr Tags wenn ihnen im Dokument inhaltliche Tags vorgeschlagen werden?
2. Welche Begriffe werden für das Tagging in Dokumenten verwendet?
3. Auf welcher Dokumentenebene werden Tags vergeben?
4. Wie gezielt lassen sich Informationen in getaggtten Dokumenten wiederfinden?

4.1 Aufgaben

Um die Fragestellungen untersuchen zu können wurden folgende Aufgaben definiert:

4.1.1 Aufgabe 1: Dokument Tagging (1)

Den Probanden wurde ein einseitiges Dokument mit 435 Wörtern (insgesamt 8 Absätze) aus der deutschsprachigen Online-Enzyklopädie Wikipedia² zum Thema „*Sonne*“ vorgelegt. Nach der Lektüre des Inhalts taggten die Probanden das Dokument unter Verwendung des Test-Systems. Ihnen stand für diese Aufgabe die Tag-Empfehlung zur Verfügung.

4.1.2 Aufgabe 2: Dokument Tagging (2)

Die Probanden taggten das gleiche Dokument wie in Aufgabe 1. Hierbei stand ihnen jedoch keine Tagempfehlung zur Verfügung.

4.1.3 Aufgabe 3: Informationssuche in einem Textdokument

Den Probanden wurde ein 22-seitiger Auszug mit 9536 Wörtern aus der deutschsprachigen Online-Enzyklopädie Wikipedia zum Thema „*Geschichte Deutschlands*“ vorgelegt. Die Paragraphen des Dokuments wurden zuvor unter Verwendung der Tag-Empfehlung von der Versuchsleitung getaggt. Die Probanden erhielten 11 inhaltsbezogene Fragen, die sie nach der Lektüre des Fragebogens mündlich mithilfe des Dokuments beantworten sollten.

Diese Aufgabe wurde entweder unter Verwendung der Tag-Suche oder unter Verzicht einer Suchfunktion bearbeitet.

4.1.4 Aufgabe 4: Informationssuche in mehreren Textdokumenten

Die Probanden erhielten vier Dokumente mit Auszügen aus der deutschsprachigen Online-Enzyklopädie Wikipedia zu den Themen „*Demographie*“, „*Religion*“, „*Wirtschaft*“ und „*Geographie*“ Deutschlands mit einer Länge von 1500 bis 2000 Wörtern. Analog zu Aufgabe 3 wurden die Paragraphen aller Dokumente vorab getaggt. Die Probanden erhielten 10 inhaltsbezogene Fragen die sie nach der Lektüre des Fragebogens mündlich mithilfe der Dokumente zu beantworten hatten.

Diese Aufgabe wurde entweder unter Verwendung der Dokument-übergreifenden Tag-Suche oder unter Verzicht jeglicher Suchfunktionen bearbeitet.

4.2 Gruppenzuordnung

Nach einer mündlichen und schriftlichen Einführung in die Aufgabenstellung und in das Test-System bearbeiteten die Probanden nacheinander die vier definierten Aufgaben. Für jede Aufgabe standen maximal 5 Minuten zur Verfügung. Da die Aufgabenstellungen von Aufgabe 1 und 2 (Anbringen von Tags) unabhängig von Aufgabe 3 und 4 (Informationssu-

² <http://www.wikipedia.de> (Stand: 28.3.2011)

che mithilfe von Tags) sind, wurden die Teilnehmer gleichmäßig auf vier Gruppen mit unterschiedlicher Bearbeitungsreihenfolge eingeteilt um Reihenfolgeeffekte zu vermeiden. Ein ' kennzeichnet hier bei der jeweiligen Aufgabe die Verwendung von ungetaggtten Dokumenten.

Gruppe 1: 1 2 3' 4

Gruppe 2: 2 1 3 4'

Gruppe 3: 1 2 4' 3

Gruppe 4: 2 1 4 3'

Zusätzlich hat die Kontrollgruppe Aufgabe 3 und 4 mithilfe einer Textsuche in alternierender Reihenfolge bearbeitet.

4.3 Auswertung

Für Aufgabe 1 wurde untersucht, wie häufig die Probanden die Tag-Empfehlung nutzten und wie viele vorgeschlagene und eigene Begriffe sie für die Tags verwandten. Die Probanden befragten durchschnittlich 7,47 Mal die Empfehlungsfunktion und verwandten im Testdokument durchschnittlich 17 vorgeschlagene und 6,14 eigene Begriffe für das Tagging (insgesamt durchschnittlich 23,14 Tags pro Person). Nur einer der betroffenen Probanden hat die Empfehlungsfunktion nicht verwendet. Auffällig war jedoch, dass 3 Probanden für 66,7% des Tagaufkommens für eigene Begriffe verantwortlich waren. Der Vergleich der verwendeten Begriffe mit denen, die an den betroffenen Textstellen von der Empfehlungsfunktion generiert werden, zeigt, dass 66,27% der verwandten Begriffe auch durch die Empfehlungsfunktion generiert werden können. Von diesen drei Probanden wurden trotzdem insgesamt 75,61% der Begriffe verwendet, die die Empfehlungsfunktion generiert hätte. Von allen in Aufgabe 1 vergebenen Tags lassen sich insgesamt 88,68% von der Empfehlungsfunktion generieren. Vergleicht man dieses Ergebnis mit Aufgabe 2, so lässt sich feststellen dass die Probanden nur durchschnittlich 18,81 Tags pro Dokument verwand haben. Von diesen Tags entsprechen 73,92% (durchschnittlich 13,9 Tags pro Proband) den Begriffen, die auch von der Empfehlungsfunktion an den betroffenen Textstellen generiert werden.

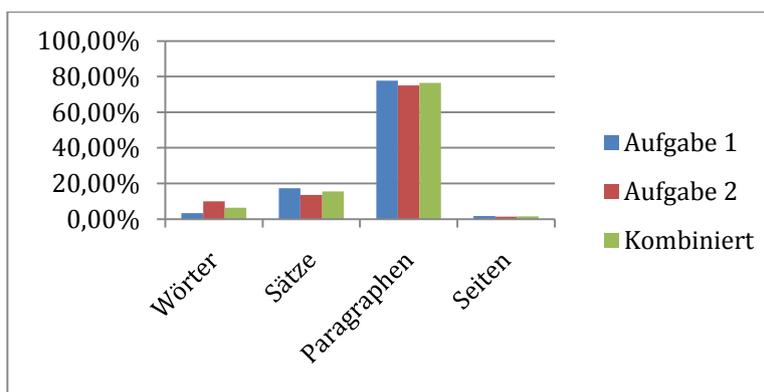


Abbildung 1: Verteilung von vergebenen Tags in Aufgabe 1 und 2 auf Dokumentteile.

Eine Analyse der in Aufgabe 1 und 2 getaggten Dokumente zeigt zudem, dass die Probanden nur selten Seiten taggten (3 bzw. nur 1 Proband). In beiden Aufgaben waren Paragraphen die am häufigsten getaggten Dokumentteile (insgesamt 76,53% aller Tags) gefolgt von Sätzen (15,58%) und Wörtern (6,32%). Nur 1,57% aller Tags wurden für Seiten verwendet.

Bei der Auswertung von Aufgabe 3 konnten wir feststellen, dass die Probanden inhaltliche Fragen in getaggten Dokumenten besser beantworten konnten als unter Verzicht auf Suchfunktionen (durchschnittlich 7,7 beantwortete Fragen mit Tag-Suche gegenüber 2,91 Antworten ohne Suchunterstützung). Die Ergebnis von Aufgabe 4 fällt ebenso eindeutig aus: unter Verzicht jeglicher Suchfunktionen konnten nur durchschnittlich 3,6 Fragen pro Proband beantwortet werden im Vergleich zu 6,36 Fragen unter Verwendung einer Dokumentübergreifenden Tag-Suche. Die Kontrollgruppe kam im direkten Vergleich auf ein nur geringfügiges schlechteres Ergebnis als die Tag-Suche von durchschnittlich 5,67 beantworteten Fragen für Aufgabe 3 und 6,33 Fragen pro Proband für Aufgabe 4 (vgl. Abbildung 2).

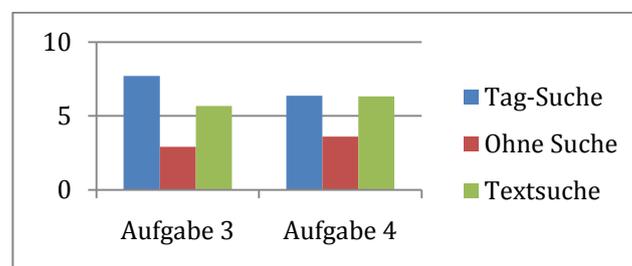


Abbildung 2: Durchschnittlich beantwortete Antworten bei Aufgabe 3 und 4 in unterschiedlichen Suchszenarien.

5 Diskussion

Die Studie zeigt eine eindeutige Tendenz bei der Frage, ob Benutzer mehr Tags vergeben, wenn sie eine Empfehlungsfunktion für Tags verwenden: Die durchschnittliche Anzahl der vergebenen Tags stieg von durchschnittlich 18,81 Tags pro Person und Dokument auf 23,14 Tags. Insbesondere die Wahl der Begriffe für Tags zeigt deutlich, dass Benutzer auf die im Dokument enthaltenen Begriffe zurückgreifen. Mehr als 73,92% der Begriffe aus Aufgabe 1 und 2 können durch die Tag-Empfehlungsfunktion, die sich auf eine Schlüsselwortextraktion und Synonymsuche beschränkt, vorgeschlagen werden. Die verbleibenden nicht-empfehlbaren Begriffe beschreiben unbekannte Synonyme und Kategorien (beispielsweise: „Definition“ oder „Daten“), oder beziehen sich auf die Dokumentstruktur („Einführung“, „Zusammenfassung“, usw.). Insgesamt bestätigt sich die Vermutung, dass beim Tagging von Dokumenten überwiegend auf Begriffe des Dokuments zurückgegriffen wird.

Die Verteilung der Tags auf Dokumentteile zeigt deutlich, dass die Probanden häufiger Tags auf Paragraphen (76,53% der vergebenen Tags) als auf Sätzen (15,58%) vergeben. Wörter und Seiten werden hingegen nur selten getaggt (6,32% bzw. 1,57% der vergebenen Tags). Obwohl eine deutliche Tendenz aufgezeigt wird, kann das Ergebnis durch den Versuchsauf-

bau begründet sein: Der Zeitdruck und die Länge des zu taggenden Textes (insgesamt eine Seite) kann bei der Erfüllung von Aufgabe 1 und 2 dazu geführt haben, dass die Probanden sich auf deutlich ausgezeichneten Dokumentstrukturen (hier: Paragraphen) konzentriert haben. Für eine endgültige Aussage müssen hier noch weitere Studien erfolgen die einerseits verschiedene Dokumentstrukturen unterschiedlicher Länge, und andererseits unterschiedliche Bearbeitungszeiträume verwenden.

Der Unterschied bei der Anzahl der beantworteten Fragen zwischen Tag-basierter und Textsuche ist hier nicht signifikant (T-Test mit unabhängige Stichproben: Signifikanz = 0,271, $F = 1,293$, $\alpha = 0,05$), obwohl die Tag-basierte Suche im Mittelwert ein besseres Ergebnis liefert: durchschnittlich 7,7 gegenüber 5,67 beantwortete Fragen pro Proband.

Bei der Suche in mehreren Dokumenten ist der Unterschied zur Textsuche ebenfalls nicht signifikant (T-Test mit unabhängigen Stichproben: Signifikanz = 0,489, $F = 0,499$, $\alpha = 0,05$), wobei hier nur ein geringfügig besseres Ergebnis vorliegt: durchschnittlich 6,36 gegenüber 6,33 beantwortete Fragen pro Proband. Es kann sein, dass die Stichprobe zu klein ist um ein signifikant besseres Ergebnis bei der Tag-Suche nachzuweisen. Zudem kann die hohe Zahl der beantworteten Fragen in der Kontrollgruppe (Textsuche) bei Aufgabe 4 in der Art der Aufgabenstellung begründet liegen: Die Probanden gaben an, dass der größte Teil der zu beantwortenden Fragen eindeutig einem Themengebiet und damit einem der vier Dokumente zugeordnet werden konnte. Dies vereinfachte die Textsuche insofern, dass die Probanden alle Fragen in einem Dokument beantworteten bevor sie mit dem nächsten fortführen. Im Vergleich dazu bearbeiteten alle Probanden der Studie mit der Tag-Suche die Fragen in der vorgegebenen Reihenfolge. Eine bessere Vergleichbarkeit der Ergebnisse lässt sich durch eine andere Wahl von Dokumenten und Fragen erzielen. Die Dokumente dürfen sich nicht in Kategorien einordnen lassen, sondern vielmehr nur ein Thema beschreiben und dabei Unterschiede im Detail und in der Art der Informationen beinhalten.

6 Zusammenfassung

In diesem Beitrag konnte gezeigt werden, dass Tagging in Dokumenten stark durch die in Dokumenten verwandten Begriffe und Worte geprägt wird. Es lässt sich vermuten, dass dies bei den Nutzern des Systems dazu führt, dass große Teile der verwandten Tag-Begriffe identisch sind. Es wurde zudem deutlich gezeigt, dass die meisten Tags an Paragraphen angebracht wurden. Zudem bestätigen die Ergebnisse der Studie die Tendenz, dass unterstütztes Tagging in Dokumenten einerseits zu der häufigeren Verwendung von Tags führt, andererseits für die Suche nach Informationen in getaggtten Dokumenten geeignet ist. Die Probanden der Studie stimmten bei der Befragung alle überein, dass das System für sie einen Mehrwert sowohl beim Tagging-Vorgang wie auch bei der Informationssuche in Word-Dokumenten darstellt.

Allerdings ist die Stichprobengröße mit 21 Probanden zu klein, um signifikante Ergebnisse bei der Tag-Suche nachzuweisen. Daher können derzeit nur Tendenzen aufgezeigt werden. Es sind noch weitere Studien notwendig (und sind geplant), um die erzielten Ergebnisse zu

verifizieren. Die aufgezeigte Tendenz, dass Benutzer viele Begriffe aus den zu taggenden Textstellen verwenden, lässt somit vermuten, dass auch fremd-getaggte Dokumenten leicht nach Informationen durchsucht werden können. Dies muss jedoch in weiteren Folgestudien nachgewiesen werden.

Literaturverzeichnis

- Bischoff, K., Firan, C. S., Nejdil, W. & Paiu, R. (2008). *Can all tags be used for search?*. Proceeding of the 17th ACM conference on Information and knowledge management, S. 193-202.
- Cantador, I., Bellogín, A. & Vallet, D. (2010). *Content-based recommendation in social tagging systems*. Proceedings of the fourth ACM conference on Recommender systems, S. 237-240.
- Carmel, D., Roitman, H. & Yom-Tov, E. (2010). *Social bookmark weighting for search and recommendation*. The VLDB Journal 19(6), 761-775.
- Guo, Y. & Joshi, J. (2010). *Topic-based personalized recommendation for collaborative tagging system*. Proceedings of the 21st ACM conference on Hypertext and hypermedia, 61-66.
- Heymann, P., Ramage, D. & Garcia-Molina, H. (2008). *Social tag prediction*. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval.
- Hong, L., Chi, E.H., Budiu, R., Pirolli, P., Nelson, L. (2008). *SparTag.us: a low cost tagging system for foraging of web content*. AVI '08 Proceedings of the working conference on Advanced visual interfaces.
- Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thie & L. Stumme, G. (2007). *Tag Recommendations in Folksonomies*. Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases, S. 50-514.
- Jäschke, R., Eisterlehner, F., Hotho, A. & Stumme, G. (2009). *Testing and evaluating tag recommenders in a live system*. Proceedings of the third ACM conference on Recommender systems, S. 369-372.
- Lu, C., Hu, X., Park, J. & Huang, J. (2011). *Post-based collaborative filtering for personalized tag recommendation*. Proceedings of the 2011 iConference.
- Medelyan, O., Frank, E. & Witten, I. H. (2009). *Human-competitive tagging using automatic keyphrase extraction*. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Volume 3.
- Mishne, G. (2006). *AutoTag: a collaborative approach to automated tag assignment for weblog posts*. Proceedings of the 15th international conference on World Wide Web.
- Nishida, K. & Fujimura, K. (2010). *Hierarchical auto-tagging: organizing Q&A knowledge for everyone*. Proceedings of the 19th ACM international conference on Information and knowledge management, S. 1657-1660.
- Shepitsen, A., Gemmell, J., Mobasher, B. & Burke, R. (2008). *Personalized recommendation in social tagging systems using hierarchical clustering*. Proceedings of the 2008 ACM conference on Recommender systems, S. 259-266.
- Sood, S. C. & Hammond, K. J. (2007). *TagAssist: Automatic Tag Suggestion for Blog Posts*. Proceedings of the International Conference on Weblogs and Social Media.

- Song, Y., Zhuang, Z., Li, H., Zhao, Q., Li, J., Lee, W. & Giles, C. L. (2008). *Real-time automatic tag recommendation*. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval.
- Song, Y., Zhang, L. & Giles, C. L. (2001). *Automatic tag recommendation algorithms for social recommender systems*. ACM Transactions on the Web 5(1).
- Tso-Sutter, K. H. L., Marinho, L. B. & Schmidt-Thie, L. (2008). *Tag-aware recommender systems by fusion of collaborative filtering algorithms*. Proceedings of the 2008 ACM symposium on Applied computing.
- Wetzker, R., Umbrath, W. & Said, A. (2009). *A hybrid approach to item recommendation in folksonomies*. Proceedings of the WSDM '09 Workshop on Exploiting Semantic Annotations in Information Retrieval.
- Wetzker, R., Zimmermann, C., Bauckhage, C. & Albayrak, S. (2010). *I tag, you tag: translating tags for advanced user models*. Proceedings of the third ACM international conference on Web search and data mining, S. 71-80.
- Zanardi, V. & Capra, L. (2008). *Social ranking: uncovering relevant content using tag-based recommender systems*. Proceedings of the 2008 ACM Conference on Recommender Systems.