

# Von Tags zu semantischen Netzen - Einsatz im Ressourcen-basierten Lernen

Renato Domínguez García, Doreen Böhnstedt, Philipp Scholl, Christoph Rensing,  
Ralf Steinmetz

Multimedia Communications Lab – Technische Universität Darmstadt  
64283 Darmstadt - Germany  
Web: <http://www.kom.tu-darmstadt.de>  
Email: [vorname.nachname@kom.tu-darmstadt.de](mailto:vorname.nachname@kom.tu-darmstadt.de)

**Zusammenfassung:** Ressourcen im Internet werden zunehmend von Lernenden als Quelle für ihren individuellen Wissenserwerbsprozess genutzt. Zur Unterstützung des Managements dieser Ressourcen im Lernprozess haben wir ein System entwickelt, das mittels typisierten Taggens ein schnelles, individuelles Auszeichnen der Ressourcen erlaubt. In dieser Arbeit zeigen wir, wie das Verfahren des typisierten Taggens zur Modellierung semantischer Netze verwendet werden kann und um die Möglichkeit der semantischen Typisierung von Relationen zwischen Ressourcen und Tags erweitert werden kann.

## 1 Motivation

Die Menge an *User Generated Content* und *Open Content* oder *Open Learnware* wird aufgrund der wachsenden Anzahl und zunehmenden Verwendung von Online-Communities und Web 2.0-Applikationen täglich größer. Oft enthält dieser Content aktuelle und für den Wissenserwerb wertvolle Informationen (z. B. in Form von Wiki- oder Forenbeiträgen, Vortragsfolien oder Videos) und kann dementsprechend für das Lernen benutzt werden. Das Lernen im Web unterscheidet sich vom institutionellen Lernen in der Tatsache, dass die Inhalte nicht von Lernautoritäten ausgewählt, pädagogisch aufbereitet und erklärt werden. Das Lernen mit Web-Ressourcen erfolgt individuell. Die Ressourcen und der einzelne Lernende stehen im Vordergrund. Das Lernen findet primär durch die Interaktion des Lernenden mit den Ressourcen statt. Der Lernende muss neben dem Verstehen der Ressourcen und Aneignen von Wissen und Fähigkeiten zusätzliche sekundäre Aufgaben wahrnehmen: er muss beispielsweise Ressourcen suchen, für den Lernprozess geeignete Ressourcen selektieren und aus den Ressourcen relevante Informationen filtern [BS08a]. Er muss die Ressourcen insgesamt verwalten.

Zur Durchführung der in der Recherche nach Web-Ressourcen und deren Management anfallenden Aufgaben nutzen Lernende heute verschiedene Anwendungen - vom Browser über Suchmaschinen bis hin zu Social-Tagging-Systemen wie delicious. Taggen als Methode, Web 2.0-Ressourcen auszuzeichnen, hat sich weitestgehend durchgesetzt. Eine umfangreiche Unterstützung des Managements von Web-Ressourcen erfolgt in diesen Anwendungen aber nur unzureichend; eine lernspezifische Unterstützung fehlt vollständig. Beim klassischen Lernen mit Selbstlerneinheiten wie beispielsweise WBTs, was eine andere Form des Ressourcen-basierten Lernens ist, liegen die Lernressourcen in (verteilten) Repositories und können von den Autoren mit Metadaten beschrieben werden.

In unserem Anwendungsfall handelt es sich dagegen um allgemeine Web-Ressourcen oder Snippets davon, die zumeist nicht mit Metadaten versehen sind. Ein weiterer Unterschied zur Auszeichnung von Lernressourcen in sogenannten Lernobjekt-Repositories besteht darin, dass wir eine dem individuellen Lernprozess des Lernenden entsprechende Auszeichnung erlauben wollen, also z. B. die Zuordnung einer Ressource zu einem individuellen Lernziel.

Zur Unterstützung des Managements von Web-Ressourcen im Ressourcen-basierten Lernen haben wir in Vorarbeiten ein eigenes System, genannt ELWMS.KOM (**E-Learning WissensManagement System**), entwickelt, das nicht nur die Strukturierung und Organisation der Recherche unterstützt, sondern auch die Erfassung von semantischen Informationen über Lernressourcen wie z. B. das Thema oder Genre der Ressource oder das Lernziel, das der Lernende mit der Verwendung der Ressource verfolgt [BS08a][BS08b]. Dieses System basiert auf semantischen Netzen [Sow92] und reiht sich damit in weitere Arbeiten zur Unterstützung von Ressourcen-basiertem Lernen [Phi08] auf Basis semantischer Netze ein. Erste Evaluationen ergaben, dass Lernende ein auf persönlichen semantischen Netzen basierendes Tool zur Recherche und zum Management der zur Wissensaneignung genutzten Ressourcen sehr schätzen [BS08a].

In diesem Beitrag stellen wir zum einen dar, wie wir in unserem System die Methode des Taggens, deren Vorteile nachfolgend genauer dargestellt werden, zur Bildung von semantischen Relationen nutzen. Zum anderen beschreiben wir eine Erweiterung unseres bisherigen Ansatzes, der darin besteht, dass Relationen zwischen Ressourcen und Konzepten sowie zwischen Konzepten typisiert werden können. Diese Typisierung von Relationen stellt einen wesentlichen Vorteil semantischer Netze gegenüber einer einfachen Zuordnung von Tags dar und ermöglicht personalisierte und detaillierte Suchanfragen. Zunächst geben wir im Weiteren einen Überblick über verwandte Arbeiten und Ansätze. Danach erfolgt im Kapitel 3 eine Beschreibung der von uns vorgenommenen Analyse relevanter Relationstypen sowie der Modellierung und der Implementierung. Abschließend erfolgen eine Zusammenfassung und ein Ausblick.

## 2 Verwandte Arbeiten

In diesem Abschnitt sollen verwandte Arbeiten vorgestellt werden, die sich ebenfalls mit der Erfassung und Darstellung von semantischen Informationen beschäftigen. Insbesondere seit dem „Hype“ um das *Semantic Web* [BH01] ist die Erfassung und Erstellung von semantischen Informationen ein aktueller Forschungsgegenstand. Die Anzahl der Arbeiten, die sich mit diesem Thema beschäftigen, ist beträchtlich und es würde den Rahmen dieses Beitrags sprengen, alle Ansätze an dieser Stelle zu erwähnen. In diesem Beitrag wollen wir deshalb nur einen allgemeinen Überblick über die verschiedenen Formen der Erstellung von semantischen Informationen geben und anschließend auf *Fuzzy*<sup>1</sup>, ein sozio-semantisches Netz zur Verwaltung von Lernressourcen, eingehen.

### 2.1 Erstellung von semantischen Informationen

Die Erstellung von semantischen Informationen kann entweder automatisch, semi-automatisch oder manuell geschehen. Bei automatischen Methoden werden bestehende Informationen z. B. aus der Autorenanwendung (z. B. der Autor) oder dem Dateisystem (z. B. das Erstellungsdatum) als Information verwendet. Weiterhin werden Verfahren der semantischen Analyse genutzt, um das Thema [WB08] oder das Genre [San07] zu

---

<sup>1</sup> <http://www.fuzzy.com>

bestimmen. Andere Ansätze generieren Metadaten während der Erstellung und Nutzung der Lernressourcen [HH05, LH+07]. Bei semi-automatischen Methoden macht das System Vorschläge für semantische Informationen über Ressourcen. Dies geschieht meist durch Einsatz von Corpora, Berücksichtigung des Kontextes der Dokumente oder frühere Eingaben des Benutzers. Der Benutzer kann diese Vorschläge übernehmen oder ändern.

Die bekanntesten Methoden zur manuellen Erfassung von semantischen Informationen sind die Metadaten-Editoren, die insbesondere zur Beschreibung von Lernressourcen lange Zeit verwendet wurden [Hör05], und das Taggen, also die Verschlagwortung von Ressourcen mit Deskriptoren, wie wir es in Web-Anwendungen häufig finden. Ein umfassender Überblick über manuelle und semi-automatische Methoden zur Erfassung von semantischen Informationen wird in [UC+06] gegeben.

Taggen weist verschiedene Vorteile auf. Taggen ist aus kognitiver Sicht einfacher als das Kategorisieren von Ressourcen, das eine Entscheidung über die Zuordnung verlangt [Ras05]. Man kann einer Ressource nicht nur mehrere Tags zuordnen, sondern auch flexibel neue Begriffe als Tags anlegen und zuordnen. In Form von Tags werden auch subjektive Informationen erfasst, wie z. B. die Relevanz oder der Zweck einer Ressource bzw. individuelle Zugänge zu den Ressourcen, also z. B. ein Ereignis in dessen Zusammenhang man eine Ressource entdeckt hat.

## 2.2 Fuzzy and Sozio-semantische Netze

Fuzzy ist ein Community-Bookmark-Manager, der ursprünglich von Roy LaChica entwickelt wurde, um die kollaborative Entstehung von Semantiken im Netz zu untersuchen [LKD08]. Die Benutzerschnittstelle ist eine Webseite, auf der Benutzer nach Verweisen auf Dokumente, die fürs Lernen geeignet sind, suchen können. Da es sich um eine Community handelt, kann jeder Benutzer eigene Links oder Tags erstellen und diese werden anschließend von der Community bewertet. Das Ziel von Fuzzy ist es also, den Lernenden qualitativ hochwertige Lernressourcen anzubieten.

Fuzzy-Benutzer erzeugen Deskriptoren mittels Taggen und mittels Verlinken von Ressourcen. Die Idee dabei ist es, zwei eigentlich eher entgegengesetzte Methoden zu vereinen und davon zu profitieren: Auf der eine Seite steht die Ontologie, die ein formales Modell erfordert, und auf der anderen steht die Folksonomie, die aus der flexiblen Benutzung von Tags entsteht. Metadaten aus Folksonomien haben eher eine niedrige Qualität [LK07]. Fuzzy benutzt diese vom Benutzer erstellten Folksonomien, um eine sogenannte *Folktologie* zu erstellen. Die Folktologie ist in diesem Fall eine Ontologie für Tags und Relationen zwischen den Tags, zusammen mit einer Topic Map für Begriffe, Assoziationen und Ressourcen. Auf diese Weise entstehen sogenannte *sozio-semantische Netze*.

## 3 Modellierung und Implementierung

Wir haben das Taggen als Methode zur Erfassung semantischer Informationen und dessen Vorteile vorgestellt. In diesem Kapitel soll detailliert erklärt werden, wie in unserem System mittels Taggen semantische Informationen erfasst und damit semantische Netze modelliert werden. Dazu betrachten wir das folgende Szenario:

Ein Lernender möchte eine Seminararbeit schreiben und sucht Informationen über Weblogs und die Blogosphäre. Er findet dazu zwei interessante Webseiten, die er im Weiteren verwenden möchte. Beide Ressourcen wurden von „Max Mustermann“ geschrieben und so taggt er diese mit den Schlagworten „Max Mustermann“ und „Seminararbeit schreiben“. Das erste Schlagwort bezeichnet den Autor und das zweite das Ziel des Lernenden. Die erste Ressource ist ein Beitrag, der von „Weblogs“ handelt und in

„Madrid“ bei der Konferenz „WWW 2009“ vorgestellt wurde. Er taggt die Ressource mit den entsprechenden Schlagwörtern. Die zweite Ressource hat die „Blogosphäre“ als Thema und ist für das Ziel des Lernenden „wichtig“. Er benutzt diese Begriffe dann als Tags (siehe Abb. 1). Die zwei Ressourcen werden in Abb. 1 durch ein weißes Blatt und die Tags als Ellipsen dargestellt.

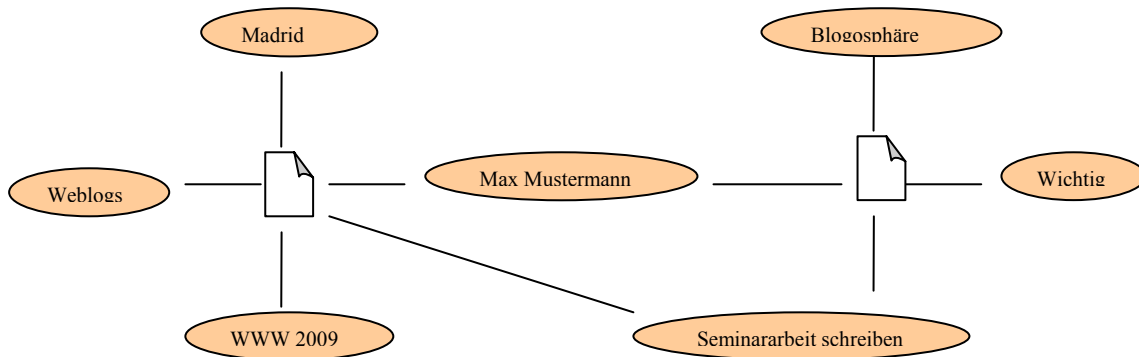


Abb. 1: Einfaches Taggen von Ressourcen

Die Verbindung eines Tags mit einer Ressource bedeutet, dass diese Ressource mit diesem Schlagwort getaggt wurde. Eine Eigenschaft des Taggens ist die Tatsache, dass alle Tags vom gleichen Typ sind: So kann z. B. nicht zwischen der Bedeutung der Tags „Madrid“ und „wichtig“ unterschieden werden: alle Tags werden als einfache Schlagwörter angesehen. Für den Benutzer ist eine individuelle Bedeutung der Tags vorhanden; für eine maschinelle Verarbeitung hingegen nicht.

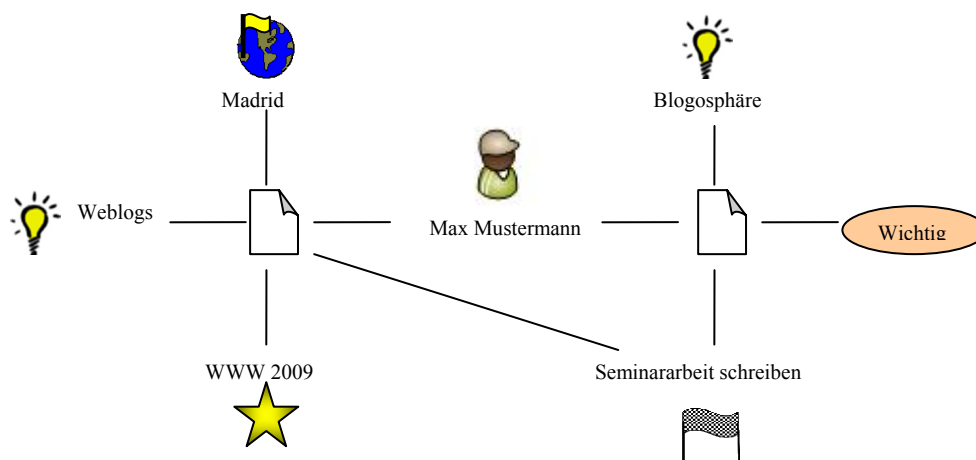


Abb. 2: Typisiertes Taggen von Ressourcen

Ein Ausweg aus dieser Situation bietet die Einführung des sogenannten *typisierten Taggens* [BS+08b]. Diese Variante des Taggens verwendet eine anwendungsspezifische Basisontologie von Tagtypen, realisiert durch Tags und einer Spezialisierung der Tags. Somit können Benutzer nicht nur Ressourcen mit freien Schlagwörtern taggen, sondern Tags auch mit eindeutigen Typen wie Personen, Ereignissen, Themen, Orten und Lernzielen versehen. Die Typisierung der Tags erlaubt es die mentale individuelle Bedeutung der Zusammenhänge zwischen den Ressourcen und der typisierten Tags abzubilden. Man reichert somit die Tags mit der semantischen Information des Typs an. In Abb. 2 wird gezeigt, wie solch eine Information im semantischen Netz modelliert wird.

Beim einfachen Taggen ist nicht ersichtlich, ob es sich bei „Madrid“ um das Thema der Ressource oder um bspw. um den Erscheinungsort handelt. Die Informationen über den Typ des Tags erlauben es den Benutzern genauere Anfragen nach Ressourcen zu erstellen.

Jetzt ist es z. B. möglich nach Dokumenten zu suchen, die „Madrid“ als Thema haben und nicht nur nach diejenigen, die mit dem Begriff „Madrid“ verbunden sind.

Typisiertes Taggen bietet aber dennoch noch keine Möglichkeit die individuelle Bedeutung der Tags für den Nutzer vollständig abzubilden. Ist eine Ressource mit dem Tag Madrid vom Typ Ort getaggt, so kann Madrid der Erscheinungsort sein oder auch der Ort, den der Benutzer individuell mit der Ressource verbindet, weil sie ihm in Madrid präsentiert wurde. Man stelle sich vor, man möchte diejenigen Artikel finden, die von der Person „Max Mustermann“ geschrieben wurden und die in Madrid präsentiert wurden. Um solche Anfragen zu unterstützen, hilft eine Typisierung der Relationen. Unser Beispiel sieht bei einer Typisierung der Relationen wie folgt aus:

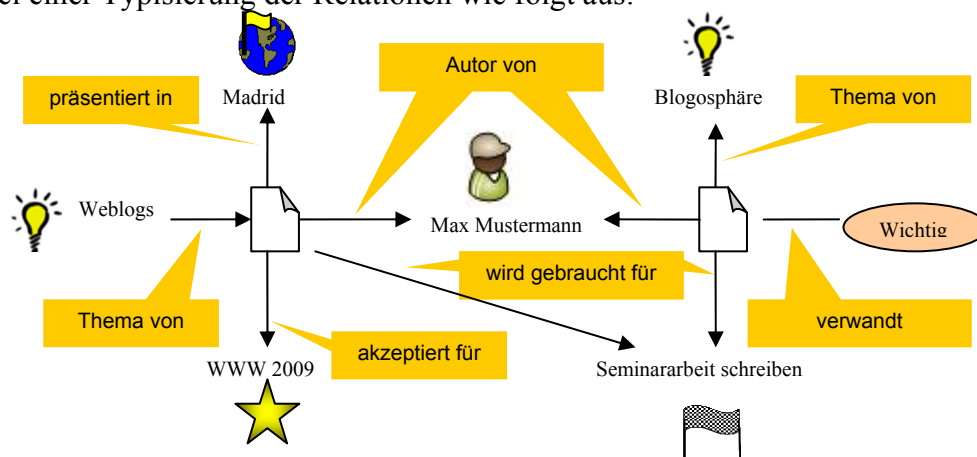


Abb. 3: Typisiertes Taggen von Relationen

Wir erweitern unser Systemkonzept und die Implementierung dementsprechend um die Möglichkeit nicht nur Tags, also Knoten des semantischen Netzes, zu typisieren, sondern auch die Relationen zwischen den Ressourcen und Tags zu typisieren. Zusätzlich ist es in unserem System nicht nur möglich Ressourcen zu taggen, sondern auch bestehende Tags wiederum zu taggen und so ein wirkliches semantisches Netz aufzubauen. Dadurch können beispielsweise einem Tag des Typs „Person“ Themen zugeordnet werden, mit denen sich diese Person beschäftigt. Auch diese Relationen sind wiederum typisierbar. Durch die Anreicherung von Relationen mit semantischen Informationen ist es nun möglich, gerichtete Relationen zu definieren.

In unserem Designprozess haben wir, ausgehend vom Anwendungsszenario des Lernens mit Web-Ressourcen, iterativ durch Aushandlung 68 Relationstypen definiert. Dabei haben wir uns, damit das System noch nutzbar bleibt, auf möglichst allgemeine Typen von Relationen beschränkt. Aus Platzgründen ist es nicht möglich, hier alle Relationen tabellarisch anzugeben, einige aussagekräftige Beispiel-Relationen sind jedoch in Tabelle 1 angegeben. Die Tabelle wird von links nach rechts gelesen: Auf der linken Seite steht der Anfangspunkt der Relation und der Endpunkt ist dann auf der oberen Zeile gekennzeichnet. Die Relation „geschrieben in“ steht als Beispiel für eine Relation zwischen „Ressource“ und „Ort“. Für einige Paare haben wir keine zusätzlichen sinnvollen Relationen finden können (z. B. von „Thema“ nach „Ort“), trotzdem existiert immer implizit eine Relation „verwandt mit“ zwischen zwei Paaren. Wir unterscheiden außerdem zwischen Personen und Benutzern. Benutzer sind die Lernenden, die in unserer Web-Anwendung registriert sind. Personen stellen Menschen und Organisationen (juristische Personen) dar, die von einem Benutzer im semantischen Netz modelliert werden. Diese beiden können aber u. U. identisch sein. Dies wäre der Fall, wenn in unserem Szenario „Max Mustermann“ auch ein Benutzer unserer Anwendung wäre.

Von/Nach	Ressource	Thema	Person	Benutzer	Ort	Ereignis	Ziel
----------	-----------	-------	--------	----------	-----	----------	------

Ressource	Referenziert	Fasst um	Geschrieben von	Gehört zum Netz von	Geschrieben in	Publiziert in	Benötigt für
<b>Thema</b>	Thema von	Synonym	Expertise von	Ist von Interesse für		Thema für	Benötigt Infos für
<b>Person</b>	Ist Erstautor	Forscht über	Arbeitet zusammen mit	Ist identisch für	Arbeitet in	Teilnehmer von	
<b>Benutzer</b>	Verwaltet	Interesse in	identisch zu		Arbeitet in	Präsentator von	Hat
<b>Ort</b>	Vortragort von		Arbeitsplatz von	Arbeitsplatz von	Ist in	Austragungsort von	
<b>Ereignis</b>	Akzeptierte	hat	Hat Teilnehmer	Hat Vortragende	Findet statt in		
<b>Ziel</b>	Resultiert aus	braucht		Ist Ziel von			Oberziel von

Tab. 1: Ausgewählte semantische Relationstypen

Der Zugriff auf die Ressourcen im semantischen Netz erfolgt für den Lernenden über ein Web-Portal. Dieses bietet auch eine grafische Oberfläche, die der Visualisierung des Wissensnetzes dient (vgl. Abb. 4).

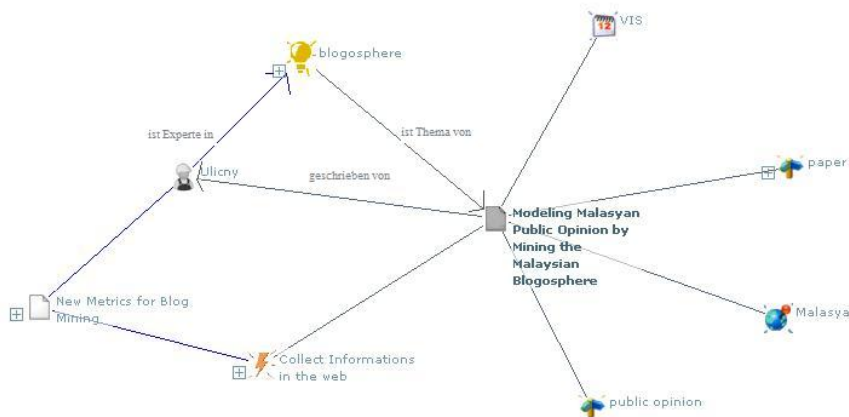


Abb. 4: Graphische Visualisierung eines Wissensnetzes

In dieser Abbildung sieht man oben links drei Relationen, die mit semantischen Informationen angereichert wurden. So erfährt man auf den ersten Blick, dass Ulicny ein Experte im Thema „Blogosphäre“ ist und dass er einen Beitrag über diesen Thema geschrieben hat. Komplexe Suchanfragen können mit Hilfe dieser Informationen gestellt werden. Ein Lernender könnte sich die Beiträge von Autoren ausgeben lassen, die Experten in bestimmten Themen sind.

Die Eingabe der semantischen Relationen geschieht mit Hilfe einer Auswahlbox (siehe Abb. 5 links). Für die Ressource (oder den Tag) werden dem Benutzer alle möglichen Knoten, d. h. Tags oder Ressourcen, die als Relationsziel in Frage kommen, angezeigt. Die Menge der möglichen Knoten kann der Benutzer über die Eingabe eines Textes einschränken. Ist das Relationsziel nicht vorhanden, kann der Benutzer dies an dieser Stelle durch Eingabe des Tags neu im semantischen Netz anlegen. Die Typisierung erfolgt durch die Auswahl des dem Relationstyp entsprechenden Formularfeldes (s. Abb. 5 rechts). Dabei werden nur diejenigen Relationstypen zur Auswahl angeboten, die den aktivierten Typ als Ausgangspunkt haben können. Diese Form der typisierten Zuordnung per Eingabefeld und Auswahlbox wurde als erste prototypische Methode zur Erfassung von semantischen Informationen implementiert.

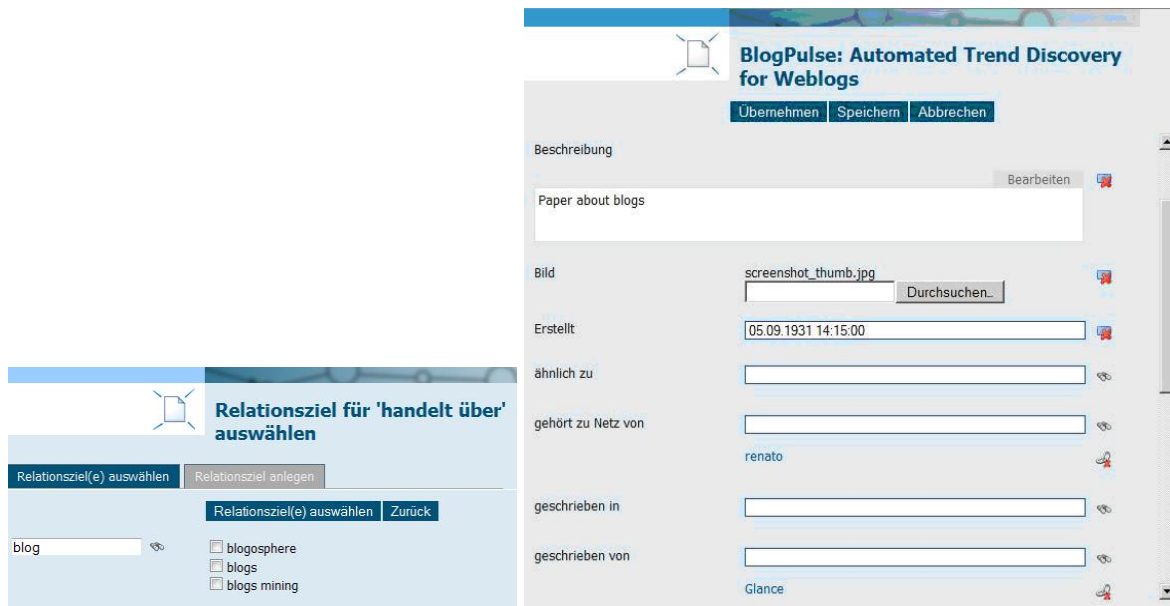


Abb. 5: Anlegen und Typisieren von Relationen in der Web-Anwendung

Eine grafische Erstellung von Relationen und deren Typisierung ist heute nur mittels einer lokalen Editoranwendung möglich (s. Abb. 6). Eine Realisierung im Web-Frontend ist geplant, wurde bisher aber noch nicht implementiert.

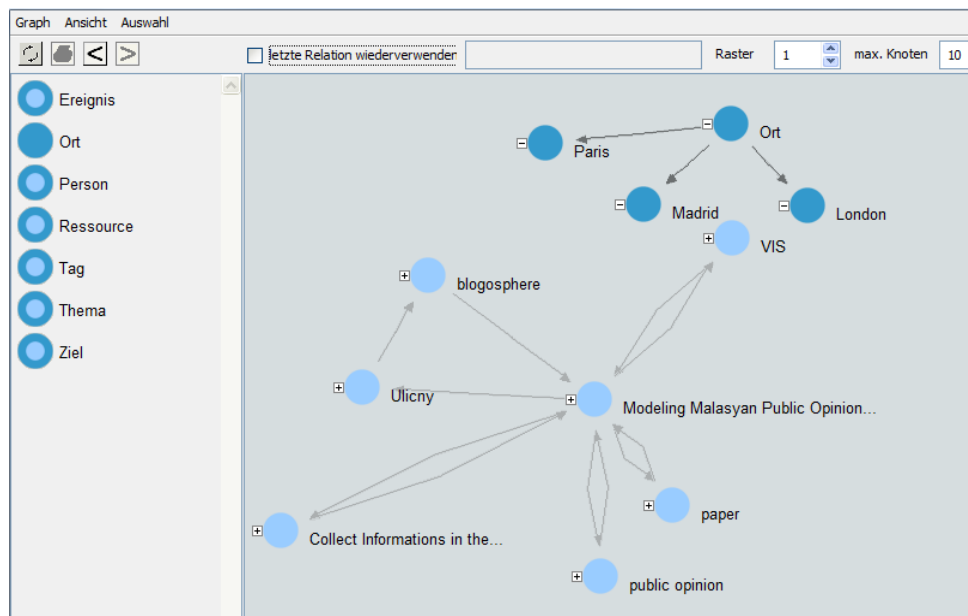


Abb. 6: Grafisches Erstellen von Relationen

## 4 Zusammenfassung und Ausblick

In dieser Arbeit haben wir eine Erweiterung unserer ELWMS.KOM-Anwendung, zur Unterstützung des Ressourcen-basierten Lernens mit Web-Ressourcen, um typisierte Relationen vorgestellt. Dies ist ein weiterer Schritt zur Entwicklung einer umfassenden persönlichen Lernumgebung, die das lernerspezifische Management von Web-Ressourcen ermöglichen soll. Die Potenziale der Typisierung von Relationen wurden beispielhaft

vorgestellt. Wir stellen damit eine Umgebung zur Verfügung, die es Lernenden erlaubt, die für ihn individuelle Bedeutung von Tags maschinenlesbar zu erfassen. Auf Basis dieser Informationen lassen sich detaillierte Anfragen stellen, und der Lerner kann einen selektiven Zugang zu bereits getaggten Ressourcen erhalten. Die nächsten Schritte bestehen aus einer Verbesserung der Benutzerschnittstellen, insbesondere die Realisierung der grafischen Erstellung von typisierten Relationen im Web-Portal, die Durchführung einer Evaluation und die Erweiterung des Wissensnetzportals um komplexe Suchanfragen. Drei unmittelbar bedeutsame Forschungsfragen sollen in Zukunft näher betrachtet werden: Werden die von uns definierten Relationstypen im Anwendungsszenario genutzt? Ist es sinnvoll, dass Benutzer eigene Relationstypen definieren und erstellen können? Welche Relationstypen - über die von uns definierten hinaus - erstellen Benutzer und wie häufig werden sie verwendet. Das im Rahmen dieser Arbeit vorgestellte System stellt die Weichen für weitere Arbeiten. Das persönliche semantische Netz des Lernenden soll im Weiteren auf Netze von Lerncommunities erweitert werden.

## Literatur

- [BH01] T. Berners-Lee, J. Handler et. al.: The Semantic Web, Scientific American, Mai, 2001.
- [BS08a] D. Böhnstedt, P. Scholl et al.: Einsatz persönlicher Wissensnetze im Ressourcen-basierten Lernen. DeLFI 2008: 2. e-Learning Fachtagung Informatik, pp. 113-124, 2008.
- [BS08b] D. Böhnstedt, P. Scholl et. al.: Collaborative Semantic Tagging of Web Resources on the Basis of Individual Knowledge Networks. In: Houben, G.-J.; McCalla, G.; Pianesi, F.; Zancanaro, M.: Proc. of First and Seventeenth Int. Conf. on User Modeling, Adaptation, and Personalization UMAP 2009, LNCS Vol. 5535, S. 379-384, Springer-Verlag Berlin Heidelberg 2009.
- [HH05] T. Hodel, R. Hamac et al.: Using Text Editing Creation Time Meta Data for Document Management. On Proc. of Conf. on Advance Information Systems Engineering, 2005.
- [Hör05] S. Hörmann: Wiederverwendung von digitalen Lernobjekten in einem auf Aggregation basierenden Autorenprozess: Dissertation TU Darmstadt, online verfügbar <http://tuprints.ulb.tu-darmstadt.de/epda/000650/>, 2005.
- [LH+07] L. Lehmann, T. Hildebrandt, et al.: Capturing, Management and Utilization of Lifecycle Information for Learning Resources. On Proc. of EC-TEL 2007, 2007.
- [LK07] R. Lachica und D. Karabeg: Metadata creation in socio-semantic tagging systems: Towards holistic knowledge creation and interchange. Saling Topic Maps. Topic Maps Research and Applications, Springer Verlag, 2007.
- [LK08] R. Lachica, D. Karabag et al.: Quality Relevance and Importance in Information Retrieval with Fuzzzy Semantic Networks. TMRA 2008, LIV series Post-Proc. paper, 2008.
- [Phi08] A. M. Philippe: Semantic Networks to support Learning. On the supplementary Proc. of the 18<sup>th</sup> Int. Conf. on Conceptual Structures in 2008.
- [Ras05] R. Sinha: A cognitive analysis of tagging (or how the lower cognitive cost of tagging makes it popular). Thoughts about Cognition, Design and Technology: Rashmi Sinha's Weblog. Aufgerufen am 08.07.2009: <http://rashmishinha.com/2005/09/27/a-cognitive-analysis-of-tagging>
- [San07] M. Santini: Automatic Identification of Genre in Web Pages. PhD Thesis, University of Brighton, 2007.
- [Sow92] J. F. Sowa: Principles of Semantic Networks, San Mateo, CA 1992 in Morgan Kaufmann Publishers.
- [UC+06] V. Uren, P. Cimiano, et al.: Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art. Journal of Web Semantics, 4(1):14-28.
- [WB08] C. Wartena and R. Brussee: Topic detection by clustering keywords. In DEXA Workshops. IEEE Computer Society, Los Alamitos, 2008.