

Semantische Suche zur Unterstützung des Internet-basierten Wissenstransfers

Karsten Krutz, Matthias Eck, Christian Mayerl, Matthias Riechmann,
Sebastian Abeck

Cooperation & Management, Institut für Telematik, Universität Karlsruhe
Zirkel 2, D-76128 Karlsruhe
Kontakt: krutz@cm-tm.uka.de

Abstract. Das Internet und die darüber verbundenen neuen Technologien bieten zahlreiche interessante Möglichkeiten, den Transfer-Prozess zwischen Wissensproduzenten (Autoren, Dozenten, Ausbilder) und Wissenskonsumenten (Studierende, Lernende, Auszubildende) gezielt zu unterstützen. Die Forschungsgruppe Cooperation & Management hat in den vergangenen Jahren ein Internet-basiertes Wissenstransfer-System entwickelt, das sie erfolgreich im täglichen Ausbildungsbetrieb einsetzt. Anhand dieser Lösung und den damit gesammelten Erfahrungen wird die Nutzung semantischer Netze zur Unterstützung der Suche nach Schulungsmaterial aufgezeigt. Hierbei wird nicht nur aufgezeigt, wie die strukturierte Datenbasis des Wissenstransfersystems durchsucht werden kann, sondern auch, wie sich diese Suche auf das gesamte Internet ausdehnen lässt.

1. Einführung

Wirtschaft und Politik haben erkannt, dass der Erfolg der heutigen Informations- und morgigen Wissensgesellschaft in einem nicht zu vernachlässigendem Maße davon abhängen wird, wie die Menschen die über das Internet vernetzte Informationstechnologie und die damit verbundenen neuen Möglichkeiten annehmen und in ihr tägliches Leben einbeziehen werden. Immer mehr Prozesse und Tätigkeiten, die heute noch traditionell ohne Einsatz der Informationstechnologie ablaufen, werden in Zukunft „elektronisch“ und Internet-basiert bewältigt werden. Der elektronische Handel (*e-commerce*) und das damit verbundene elektronische Bestellen und Bezahlen von Waren gehört zu den Bereichen, bei denen dieser Wandel bereits heute deutlich und für jeden erkennbar ist [1].

Ein weiterer Bereich, in dem Internet-Technologien ebenfalls intensiv erprobt werden, ist die Aus- und Weiterbildung. Selbst von nicht-kommerzieller und als neutral zu bezeichnender Seite wird der Nutzung des Internet im Bereich des Ausbildens und Lernens ein erhebliches Potential prophezeit und eine nicht zu unterschätzende gesellschaftliche Bedeutung beigemessen. Verwiesen sei hier beispielsweise auf einen eindrucksvollen Bericht der Web-Based Education Commission, der den Titel trägt „The Power of the Internet for Learning“ [2].

Mit dem steigenden Bedarf ist auch die Anzahl der Schulungsangebote geradezu explodiert. Diverse Anbieter – sowohl aus dem akademischen Sektor als auch aus der Industrie – konkurrieren um die Gunst der Lernenden. Durch den Einsatz der Internet-technologien spielt die Lokation dieser Anbieter eine nur untergeordnete Rolle und die Lernenden können auf beliebige Angebote sehr einfach orts- und zeitunabhängig zugreifen.

Somit sehen sich die Lernenden einer Vielzahl von potentiell für sie interessanten Angeboten von verschiedenen Anbietern gegenüber und benötigen Werkzeuge, die sie bei der Auswahl geeigneter Materialien unterstützen.

Ein Lösungsansatz für dieses Problem ist die Einführung einer Ablagestruktur, die neben den Inhalten standardisierte semantische Beschreibungen der Materialien enthält. Auf Basis dieser Beschreibungen kann eine semantische Suche angeboten werden. Diese Suche kann von den Lernenden in Anspruch genommen werden und ermöglicht den Autoren zudem, effizient Materialien anderer Autoren wiederzuverwenden.

Anhand des in der Forschungsgruppe Cooperation & Management entstandenen ed.tec-Systems (educational technologies) wird die im Rahmen dieses Systems konzipierte und implementierte semantische Suchunterstützung vorgestellt. Sie lässt sich sowohl für die Datenablage dieses Wissenstransfersystems als auch für die Suche im Internet vorteilhaft einsetzen.

2. Integrierte Werkzeugumgebung ed.tec

Das ed.tec-System ist in den vergangenen Jahren in der Forschungsgruppe C&M als ein modular aufgebautes System zur Unterstützung des Internet-basierten Wissenstransfers entstanden und bildet heute die technische Basis für die von der Forschungsgruppe ihren Studierenden bereitgestellten Ausbildungsdienste. Es unterstützt alle am Wissenstransfer-Prozess beteiligten Rollen durch Bereitstellung der folgenden Funktionen:

- Dem **Autor** werden Werkzeuge zur Aufbereitung von Wissensmaterialien zur Verfügung gestellt. Hiermit kann er seine vorhandenen, für die Lehrveranstaltung vorgesehenen Inhalte effizient in das Wissenstransfer-System überführen und dadurch die zahlreichen Funktionen des Systems nutzen. Dabei wird der Autor u.a. auch bei der Erstellung einer semantischen Beschreibung der Wissensmaterialien unterstützt.
- Der **Dozent** kann mittels einer automatisch generierten graphischen Oberfläche flexibel und gezielt auf die verschiedenen multimedialen Inhalte während der Lehrveranstaltung zugreifen. Außerdem werden die Aufzeichnung der Lehrveranstaltung in Form von *Screen-Capture*-Videos sowie eine Online-Übertragung in das Internet unterstützt.
- Für den **Lernenden** werden die Inhalte in einer über das Web zugreifbaren Lernumgebung strukturiert und übersichtlich angeboten. Wie bereits die Dozentenum-

gebung wird auch die Lernumgebung automatisch erzeugt, was zu einer Entlastung der Ausbildungsinstitution führt.

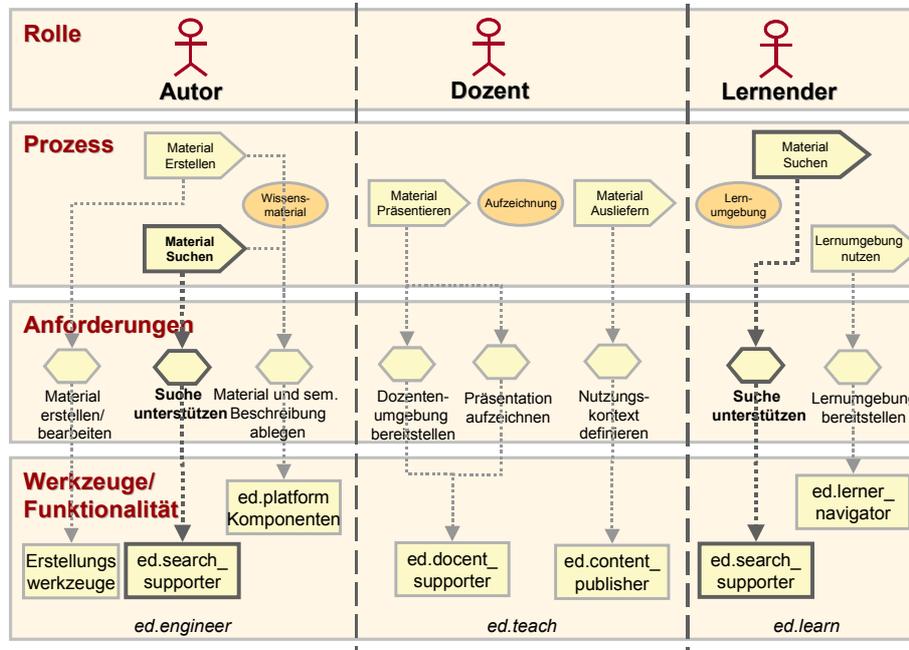


Abbildung 1: Rollen, Prozesse und Werkzeuge

Die Rollen und zentralen Prozesse sowie einige ausgewählte Werkzeuge des ed.tec-Systems sind in Abbildung 1 dargestellt.

Das wichtigste Werkzeug im Rahmen dieses Beitrags stellt der `ed.search_supporter` dar. Seine primäre Funktion ist die Unterstützung des Anwenders bei der Suche nach geeignetem Wissensmaterial. Ob der Anwender dabei ein Autor ist, der Wissensmaterial zur Wiederverwendung sucht, oder ein Lernender, der weiterführendes Material benötigt, ist hierbei nicht von Bedeutung. Durch die Nutzung semantischer Suchmöglichkeiten sowie den Einsatz eines Wissensnetzes kann in beiden Fällen die Suche sehr zielgerichtet erfolgen.

Neben der Suchfunktionalität bietet der `ed.search_supporter` dem Autor die Möglichkeit, die zur effizienten Suche benötigten semantischen Beschreibungen zu Wissensmaterialien zu erstellen und über die `ed.platform`-Komponenten strukturiert abzulagern. Weiter bietet er Funktionalität zur Unterstützung des Autors bei der Erstellung eines Wissensnetzes zu einem vorgegebenen Thema. Da ein solches Wissensnetz die wichtigsten Begriffe eines Themenbereichs und deren Beziehungen enthält, stellt es, neben seiner Funktion zur Unterstützung der semantischen Suche, ein sehr wertvolles Wissensmaterial für Lernende dar.

Um diese Such- und Navigationsfunktionalität den Lernenden effizient bereitstellen zu können, werden in der nächsten Version der Lernumgebung, dem `ed.learner_navigator`, neben der Darstellung der Materialien und der Ablagehierarchie auch Such- und Navigationskomponenten aus dem `ed.search_supporter` integriert (s. Ausblick).

Der Fokus liegt im Folgenden auf dem Werkzeug `ed.search_supporter` und den innerhalb dieses Werkzeugs realisierten Konzepten zur semantischen Suche. Zunächst soll der Teil des Stands der Technik aufgezeigt werden, der dieser Lösung zugrunde liegt.

3. Stand der Technik im Bereich der rechnerbasierten Suche

Zur Bewertung verschiedener grundlegender Suchansätze lässt sich laut [3] die Effizienz unterscheiden. In der Theorie zur Effizienz von Suchansätzen unterscheidet man zwischen zwei grundsätzlichen Arten von Wissen, dem so genannten *tacit knowledge* und dem so genannten *explicit knowledge*.

- *Tacit knowledge* (*tacit*: engl. stillschweigend) ist das Wissen, das eine Person hat und das aus ihren Informationen, Einsichten und Ideen besteht, das aber anderen nicht zugänglich ist.
- *Explicit knowledge* ist im Gegensatz dazu die gewissermaßen formalisierte Information, kodiert auf einen Informationsträger, d. h. beispielsweise in Textdokumenten.

Unter der Externalisierung (*externalisation*) versteht man die Vorgänge, die durchzuführen sind, um *tacit knowledge* in *explicit knowledge* zu verwandeln.

Zum einen wird hierzu die Aufzeichnung (*creation*) dieses Wissens in eine durch andere lesbare Form gezählt, zum anderen aber auch die Klassifizierung dieser Dokumente, um sie anderen über Suchverfahren zugänglich zu machen.

Die Internalisierung (*internalisation*) ist schließlich der gegengesetzte Prozess, der *explicit knowledge* findet und durch Verstehen dieser *explicit knowledge* in *tacit knowledge* der jeweiligen Person überführt.

Die Externalisierungseffizienz stellt bei der Suche somit den Aufwand zur Erstellung der Suchdatenstruktur dar und die Internalisierungseffizienz die Nützlichkeit der Suchdatenstruktur.

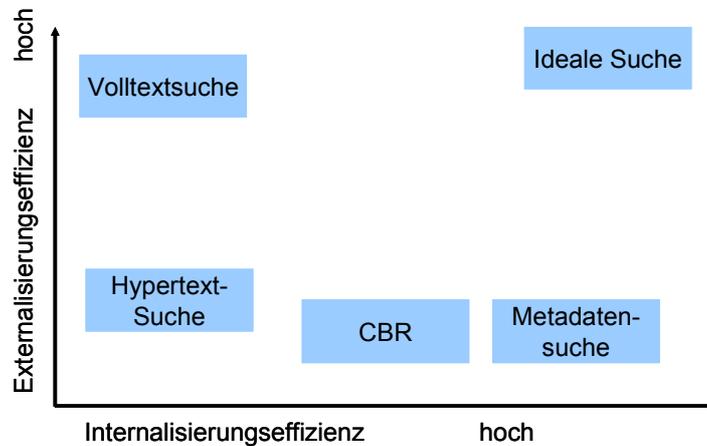


Abbildung 3: Effizienz verschiedener Suchansätze

In der Abbildung 3 ist die Effizienz von vier Suchansätzen graphisch dargestellt. Der ideale Suchansatz würde sowohl hohe Externalisierungs- als auch Internalisierungseffizienz bieten. Leider gibt es keinen Ansatz, der dieses optimale Ziel erreicht.

Die Suchdatenstruktur bei der Volltextsuche, wie sie bei vielen Suchmaschinen im Internet zu finden ist (z.B. [4]), lässt sich automatisch erzeugen. Die Aussagekraft dieser Suchdatenstruktur ist jedoch beschränkt. Daher besitzt die Volltextsuche eine hohe Externalisierungseffizienz aber eine nur geringe Internalisierungseffizienz.

Hypertext ist als Suchdatenstruktur nur bedingt geeignet, da der Aufwand zur Erstellung sehr hoch und der Nutzen zur Suche recht gering ist. Suche über eine Hypertextstruktur lässt sich nur durch die Navigation über die Verknüpfungen realisieren und damit ist eine zielgerichtete Suche schwierig durchführbar.

Durch Fallbasiertes Schließen (engl. *Case Based Reasoning, CBR*, siehe [5]) kann vorhandenes Erfahrungswissen, das in Form von sogenannten Fällen in einer Fallbasis gespeichert wird, zur Lösung neuer Probleme wieder verwendet werden. Der Aufwand zur Erstellung dieser Fallbasis ist sehr hoch und die Internalisierungseffizienz stark vom Umfang und der Qualität dieser Fallbasis abhängig.

Der letzte betrachtete Suchansatz ist die metadatenbasierte Suche. Unter Metadaten versteht man nach ISO/IEC Standard 11179 [6]:

“The information and documentation which makes datasets understandable and shareable for users.”

Eine Definition nach Tim Berners-Lee, dem Erfinder des World Wide Web und Direktor des W3C lautet:

„Metadaten sind maschinenlesbare Informationen über elektronische Ressourcen oder andere Dinge.“ (aus [7]).

Metadaten enthalten also Informationen über ein Dokument, zum einen Zusatzinformationen, die nicht mit dem eigentlichen Inhalt zu tun haben, wie beispielsweise

Autor, Veröffentlichungsdatum und Schwierigkeitsgrad, aber auch inhaltliche Informationen, wie beispielsweise Schlüsselwörter.

Da die semantische Bedeutung jedes einzelnen Metadatums feststeht, erlauben Metadaten eine semantische Suche. Dies wird auch als Grundmodell der metadatenbasierten Suche bezeichnet.

Durch die semantische Suchfunktionalität bietet die metadatenbasierte Suche eine hohe Internalisierungseffizienz. Da die Metadaten üblicherweise manuell vergeben werden müssen, ist die Externalisierungseffizienz jedoch hoch.

Da aus Sicht des Lernenden die Internalisierungseffizienz die größere Relevanz besitzt, ist die metadatenbasierte Suche für den Lernenden die beste Wahl. Aus Sicht der Autoren ist hingegen die Volltextsuche vorzuziehen, da hierbei der Aufwand zur Erstellung der Suchdatenbasis am geringsten ausfällt. Daher werden diese beiden Ansätze bei ed.tec verwendet.

4. Suchanfragenoptimierung durch Wissensnetze

Um die Qualität des Ergebnisses einer Suchanfrage zu bewerten, werden die Größen *Recall* (Vollständigkeit) und *Precision* (Genauigkeit) verwendet. Für die Berechnung sind folgende Untermengen der Gesamtmenge der Dokumente zu betrachten.

- Hits: Menge relevanter und angezeigter Dokumente
- Loss: Menge relevanter aber nicht angezeigter Dokumente
- Noise: Menge nicht relevanter aber angezeigter Dokumente
- Rejection: Menge nicht relevanter und nicht angezeigter Dokumente

Damit lassen sich Recall und Precision berechnen durch:

$$Recall = \frac{\# Hits}{\# \text{Relevante Dokumente}}$$
$$Precision = \frac{\# Hits}{\# \text{Angezeigte Dokumente}}$$

Der *Recall* beschreibt also die Fähigkeit des Systems, alle relevanten Dokumente zu bestimmen, die *Precision* dagegen die Fähigkeit, nur relevante Dokumente zu finden. Bei großen Datenbeständen wird oft auf die Berechnung des *Recall* verzichtet, da es schwierig ist, alle relevanten Dokumente anzugeben.

Eine Optimierung der Suchergebnisse auf *Recall* bzw. *Precision* kann durch eine Modifikation der Suchanfrage erreicht werden. Das dazu benötigte Wissen muss dazu in einer wissensrepräsentierenden Datenstruktur vorliegen.

Wissensrepräsentierende Datenstrukturen speichern Informationen in maschinenlesbarer aber insbesondere auch in maschinenverstehbarer Form und stellen somit die Basis für rechnerunterstütztes Wissensmanagement dar.

5. Implementierte Lösungsansätze zur semantischen Suche

Im Folgenden wird der implementierte Lösungsansatz im Rahmen des `ed.search_supporters` anhand eines konkreten Beispiels demonstriert. Wie bereits erwähnt, stellt dieses Werkzeug dem Anwender nicht nur die Suchfunktionalität zur Verfügung, sondern unterstützt ihn auch bei der Erstellung der benötigten semantischen Beschreibungen und Metadatenstrukturen. Hierbei ist insbesondere die Erstellung eines Wissensnetzes hervorzuheben.

Ein solches Wissensnetz setzt einzelne Begriffe, also beispielsweise „.NET“ oder „dotnet“ in festgelegte Relationen zueinander. Sie enthalten damit Aussagen, wie „.NET ist Synonym zu dotnet“.

Neben den Metadatendateien stellen sie eine wichtige maschinenlesbare und maschinenverstehbare Quelle für semantische Informationen über Wissensmaterialien bzw. deren Inhalt dar.

In ausführlichen Überlegungen wurde in [9] dargelegt, dass sich Wissensnetze für eine Verbesserung der Suchfunktionalität nutzen lassen. Der Ansatz wird im Folgenden an einem Beispiel demonstriert.

Um die Suchfunktionalität nutzen zu können, müssen zunächst die semantischen Beschreibungen zu den Wissensmaterialien erstellt werden. Als konkretes Beispielszenario werden im Folgenden als Wissensmaterialien einige Kapitel aus einer bestehenden Vorlesung „Entwicklung von Internet-Systemen und Web-Applikationen“ verwendet.

In dieser Vorlesung wird u.a. das Thema .NET – also die neue Middleware- und Web-Service-Technologie von Microsoft – behandelt. Die Wissensmaterialien müssen zunächst mit einer semantischen Beschreibung versehen werden. Diese Zusatzinformationen wie Autor, Sprache oder Schlüsselwörter werden in Form von Metadateien gespeichert. Im Kontext der Suche sowie der Erstellung eines Wissensnetzes sind vor allem die verwendeten Schlüsselwörter interessant. Der Autor eines Kapitels, das die Entwicklung von Webservice mit .NET behandelt, könnte z.B. folgende Schlüsselwörter vergeben:

- .NET
- dotnet
- .NET Framework
- Web Service
- ...

Ein Wissensnetz zu diesem Teil der Vorlesung würde beispielsweise folgende semantische Beziehungen zu „.NET“ enthalten.

- .NET ist eine spezielle Lösung einer Middleware
Durch die Verallgemeinerungsrelation formuliert ergibt sich damit: „Middleware-Lösung ist Verallgemeinerung von .NET“

- Eine andere Schreibweise von .NET ist dotnet
Dies lässt sich durch eine Synonym-Beziehung „.NET ist Synonym zu dotnet“ ausdrücken.
- Das .NET Framework und der Web Service (synonyme Bezeichnung Web-Service) sind Bestandteile von .NET
Daraus lassen sich die folgenden Beziehungen ableiten:
„.NET besteht aus .NET Framework“ und
„.NET besteht aus Web Service“

Durch das Werkzeug ed.search_supporter wird der Autor dabei unterstützt, diese Zusammenhänge in einem Wissensnetz formal auszudrücken.

Abbildung 5 zeigt den ed.search_supporter mit dem aufgebauten Wissensnetz. Die Darstellung der Relationen wird durch verschiedenfarbige Pfeile, analog der Darstellung in der links oben angezeigten Legende realisiert. Synonyme werden als direkt untereinander angeordnete Begriffe dargestellt.

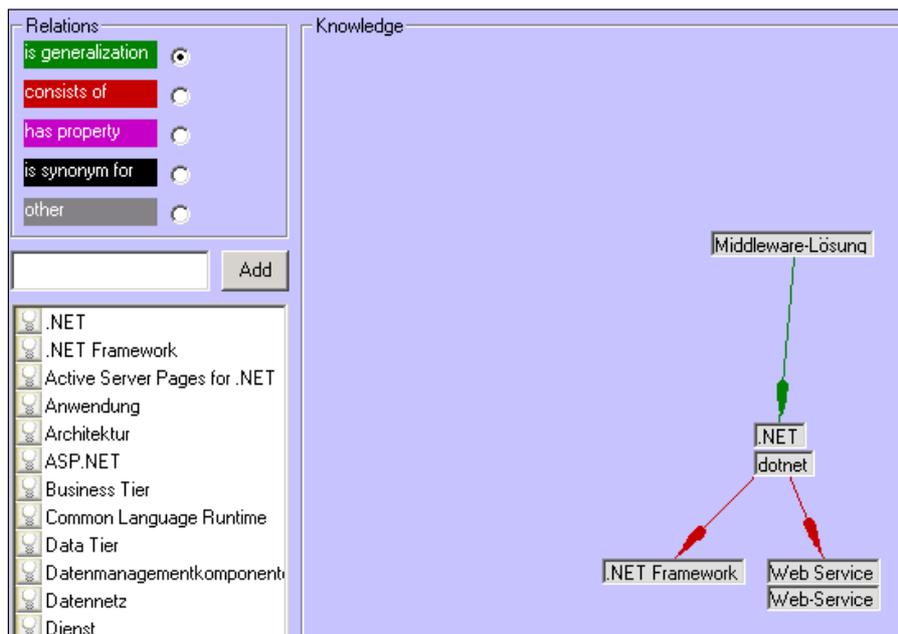


Abbildung 5: Wissensnetz zum Begriff .NET im ed.search_supporter

Durch ein derartiges Wissensnetz wird eine neue Qualität von Informationen geliefert, die für alle mit dem Wissensmaterial arbeitenden Personen wertvoll ist. Insbesondere der Lernende erhält die Möglichkeit, einen Begriff unter Nutzung des Wissensnetzes richtig einzuordnen. Zudem kann ein Anwender über das Wissensnetz durch Navigation Schlüsselworte identifizieren, nach denen er suchen kann. Dies kann ihm dabei helfen, zielgerichtetere Anfragen zu erstellen.

Weiter kann der Anwender Schlüsselworte zur Suche angeben, aus denen das Werkzeug `ed.search_supporter` verschiedene Suchanfragen generiert. Das Wissensnetz ermöglicht hierbei, die Suchanfragen so zu gestalten, dass die Anzahl der Ergebnisse der Suche erhöht oder die Qualität der Ergebnisse optimiert wird.

Die Synonym-Beziehungen zwischen „NET“ und „dotnet“ im Wissensnetz kann auf eine ODER-Verknüpfung in der Suchanfrage abgebildet werden. Damit kann also bei der Suche nach Materialien die sich mit .NET beschäftigen auch Materialien gefunden werden, die unter dem Begriff dotnet abgelegt sind. Anstelle von der Suche nach .NET wird also bei diesem Beispiel die Suchanfrage „NET OR dotnet“ verwendet.

Um die Anfragen zielgerichteter zu gestalten und um damit die Qualität des Ergebnisses zu erhöhen, wird die Vererbungs-Beziehung zwischen „Architektur“ und „NET“ in eine UND-Verknüpfung abgebildet. Damit wird erreicht, dass eine Suche nach dem Begriff „Architektur“ Ergebnisse liefert, in denen der Begriff eher im gewünschten Kontext verwendet wird. Bei der Verwendung der Suchanfrage „Architektur AND .NET“, sollte die *Precision* der Ergebnisse deutlich höher sein als bei der Suche nach „Architektur“.

Die beiden vorgestellten Ansätze lassen sich auch kombinieren, um sowohl die Qualität als auch die Quantität zu erhöhen.

Der `ed.search_supporter` generiert aus einem gegebenen Schlüsselwort eine Menge von Suchanfragen für die verschiedenen Optimierungskriterien, aus denen eine geeignete ausgewählt und durch Metadatenuche auf der `ed.tec`-Datenhaltung ausgewertet werden kann. Zudem kann diese Suchanfrage direkt an eine WWW-Suchmaschine (z. B. Google) abgesetzt werden.

Unter Nutzung des obigen Wissensnetzes lassen sich zur Suche nach Materialien zum Thema „Web-Services“ beispielsweise folgende Suchanfragen erzeugen.

- Web Service OR Web-Service
- (Web Service OR Web Service) AND .NET
- (Web Service OR Web-Service) AND dotnet
- ...

Über diese Suchanfragen kann der `ed.search_supporter` auf der `ed.tec`-Datenhaltung leicht das betrachtete Kapitel zu Webservices aus der Vorlesung „Entwicklung von Internet-Systemen und Web-Applikationen“ sowie ähnliches Material finden. Zudem kann er die Suchanfragen zur Suche nach weiterführendem Material im WWW nutzen.

Ein Autor, der diese Suchfunktionalität nutzt, kann die gefundenen und als geeignet angesehenen Wissensmaterialien durch entsprechende von `ed.tec` bereitgestellte Werkzeuge mühelos in die Schulungsmaterial-Ablage übernehmen. Dadurch stehen diese Informationen unmittelbar Dozenten und Lernenden im Rahmen der Dozenten- bzw. Lernumgebung als weiterführendes Material zur Verfügung.

6. Ausblick

In diesem Beitrag wurde eine semantische Suchunterstützung im Rahmen des Internet-basierten Wissenstransfers vorgestellt und am Beispiel des ed.tec-Systems verdeutlicht. Hierbei wurden die Vorzüge gegenüber einer Volltextsuche ohne Unterstützung der Suchanfragengenerierung durch Wissensnetze aufgezeigt.

Die vorgestellte Lösung ist bereits einsatzbereit und wird in der Lehre der Forschungsgruppe verwendet, lässt sich aber noch weiter ausbauen. In einem weiterführenden Schritt wird die Funktionalität des Werkzeugs Web-basiert angeboten und eine engere Integration der Such- und Navigationsfunktionalität in die Lernumgebung realisiert.

Dabei werden die Lernenden in der Lage sein, über die Lernumgebung eigene Schlüsselworte zu vergeben und eigene Wissensnetze (ggf. in Zusammenarbeit mit anderen Lernenden) zu erstellen. Dieses Material kann dann über die Lernumgebung der Allgemeinheit verfügbar gemacht werden.

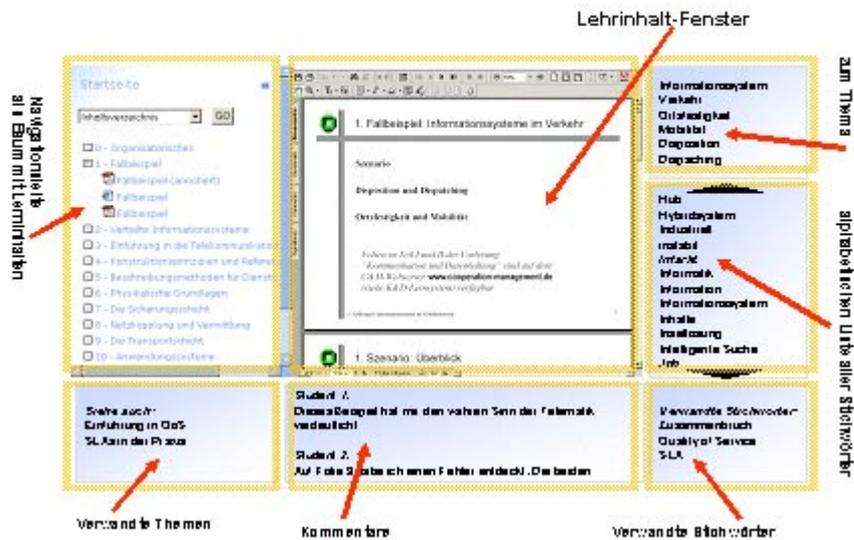


Abbildung 6: Die Lernumgebung ed.learner_navigator

Die Integration der Suchfunktionalität in der als ed.learner_navigator bezeichneten Web-basierten Lernumgebung ist in der Abbildung 6 dargestellt. Das Wissensnetz liefert über diese Lernumgebung dem Lernenden Hinweise auf andere Inhalte, die ihm Details erklären können oder verwandte und weiterführende Themen behandeln.

7. Literaturverzeichnis

- [1] P. Cunningham and F. Fröschl, *Electronic Business Revolution: Opportunities and Challenges in the 21st Century*: Springer-Verlag, 1999.
- [2] B. Kerrey and J. Isakson, "The Power of the Internet for Learning - Moving from Promise to Practice", *Web-based Education Commission*, www.hpcnet.org/webcommission, Washington, DC 2000.
- [3] K. Stanoevska, A. Hombrecher, S. Handschuh, and B. Schmid, "Efficient Information Retrieval: Tools for Knowledge Management", *Institute for Media and Communications Management, University of St. Gallen*, St. Gallen 1998.
- [4] V. C. Schöch, "Die Suchmaschine GOOGLE" in *Institut für Informatik*. Berlin: Freie Universität Berlin, 2001.
- [5] A. Schuster, R. Sterritt, K. Adamson, M. Shapcott, and E. P. Curran, "Case-based Reasoning for Complex Telecommunication Systems" in *School of Information and Software Engineering, Faculty of Informatics: University of Ulster*, 1999.
- [6] T. Meyer, "Metadaten". Oldenburg: Universität Oldenburg, 2000, pp. 25.
- [7] K. Zäpke, <http://www2.sub.uni-goettingen.de/intrometa.html>, "Einführung in Metadaten", 2001
- [8] U. Reimer, *Einführung in die Wissensrepräsentation - netzartige und schema-basierte Repräsentationsformate*. Stuttgart: Teubner, 1991.
- [9] M. Eck, *Wissensnetze zur Navigation und Semantischen Suche in der Internet-basierten Aus- und Weiterbildung: Diplomarbeit an der Universität Karlsruhe in der Forschungsgruppe "Cooperation & Management" (Prof. Abeck)*, 2001.