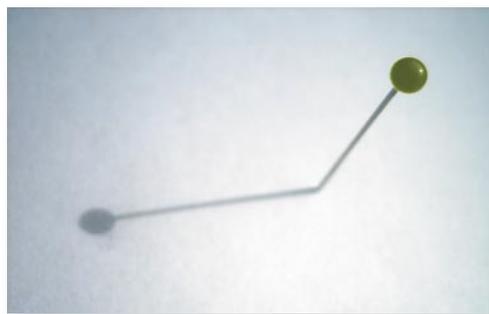


Fabian Kaiser
Sven Schimpf
Dr. Holger Schwarz
Mihály Jakob
Dr. Severin Beucker



Internetgestützte Expertenidentifikation zur Unterstützung der frühen Innovationsphasen

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Fabian Kaiser
Sven Schimpf
Dr. Holger Schwarz
Mihály Jakob
Dr. Severin Beucker

Internetgestützte Expertenidentifikation zur
Unterstützung der frühen Innovationsphasen

Impressum

Herausgeber

Fabian Kaiser
Sven Schimpf
Dr. Holger Schwarz
Mihály Jakob
Dr. Severin Beucker

Verlag

Fraunhofer IRB Verlag,
Nobelstraße 12, 70569 Stuttgart

Copyright

nova-net Konsortium, und Fraunhofer-Institut für
Arbeitswirtschaft und Organisation IAO, Stuttgart

ISBN

978-3-8167-7448-8

Erscheinungsjahr

2007

Gestaltung

Anette Grimmel, Petra Riesemann

Titelbild

www.photocase.com

Auslieferung und Vertrieb

Fraunhofer IRB Verlag
Nobelstraße 12, 70569 Stuttgart
Telefon +49(0)7 11/970-25 00
Telefax +49(0)7 11/970-25 08

www.irb.buch.de

www.publica.fhg.de

Alle Rechte vorbehalten.

Dieses Werk ist einschließlich aller seiner Teile urheberrechtlich geschützt. Jede Verwertung, die über die engen Grenzen des Urheberrechtsgesetzes hinausgeht, ist ohne schriftliche Zustimmung des Fraunhofer-Instituts für Arbeitswirtschaft und Organisation unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen sowie die Speicherung in elektronischen Systemen. Die Wiedergabe von Warenbezeichnungen und Handelsnamen in diesem Buch berechtigt nicht zu der Annahme, dass solche Bezeichnungen im Sinne der Warenzeichengesetzgebung als frei zu betrachten wären und deshalb von jedermann benutzt werden dürften.

Inhaltsverzeichnis

1	Frühe Innovationsphasen und das Internet	4
2	Zusammenfassung	5
3	Motivation für die Expertensuche	6
4	Methodischer Ansatz	7
4.1	Identifikation des Suchfeldes	8
4.2	Spezifikation des Suchfeldes	9
4.3	Identifikation von Experten	10
5	IT-Unterstützung von »EXPOSE«	13
5.1	Anforderungen an die Software	13
5.2	Technische Umsetzung	14
	Schritt 1 – Die Spezifikation des Informationsbedarfs	14
	Schritt 2 – Die Suche nach relevanten Dokumenten	17
	Schritt 3 – Die Identifikation von Experten und deren Bewertung	23
6	Praxisbeispiel	27
7	Literatur	30

1 Frühe Innovationsphasen und das Internet

Die vorliegende Broschüre ist Ergebnis des Forschungsprojektes nova-net: Innovation in der Internetökonomie.¹

Das Projekt nova-net entwickelt Theorien, Strategien und Instrumente zur Unterstützung nachhaltiger Produkt- und Serviceinnovationen in der Internetökonomie. Dabei wird davon ausgegangen, dass das Internet sowohl eine Quelle als auch ein Instrument für nachhaltige unternehmerische Innovationsprozesse darstellt.

So sind im Internet eine Vielzahl von strukturierten und semistrukturierten Informationen, wie z.B. Patent- und Handelsdaten sowie technische Spezifikationen enthalten, die in frühen unternehmerischen Innovationsphasen wichtige Entscheidungsunterstützungen liefern können. Außerdem spielt das Internet als Plattform für Kommunikationsprozesse und softwaretechnische Anwendungen eine wichtige Unterstützungsrolle im Innovationsprozess, beispielsweise für den Ideenaustausch oder auch für die Aktualisierung von Informationen in komplexen Entwicklungsprozessen.

Zentrales Ziel der Arbeiten in nova-net ist es, Unternehmen bei der Gestaltung von Produkt- und Serviceinnovationen zu unterstützen. Ein Schwerpunkt des Projektes liegt dabei auf frühen Innovationsphasen, die sich in einem zunehmend komplexen Umfeld ökonomischer, rechtlicher und gesellschaftspolitischer Anforderungen (z.B. Produkthaftung, Risikominderung und umweltpolitische Auflagen) bewegen. Gleichzeitig bestehen hier die größten Einflussmöglichkeiten auf ein neues Produkt, um Weichenstellungen bezüglich der Kostenstruktur und potenzieller Folgewirkungen zu beeinflussen.

Für die Aufgaben in den frühen Phasen des unternehmerischen Innovationsprozesses sind neue Methoden der Analyse und Strukturierung des Innovationsmanagements im Unternehmen notwendig. In nova-net wurden daher zwei Ansätze verfolgt, die folgende Teilaufgaben früher Innovationsphasen abdecken:

- Trendmonitoring im Szenariomanagement: Frühe Innovationsphasen sind durch einen hohen Bedarf an der strukturierten Darstellung maßgeblicher Einflussgrößen für die Entwicklung von Technologien und Märkten gekennzeichnet. Mit Hilfe entsprechend aufbereiteter Szenarien können Fehlentscheidungen und deren Folgekosten vermieden werden. Dieses Ziel wird im Schwerpunktthema »Trendmonitoring im Szenario-Management« mit dem entwickelten delphigestützten Szenario-Management verfolgt.
- Expertensuche und -integration: Insbesondere kleine und mittelständische Unternehmen sehen sich im Rahmen von Innovationsvorhaben vor das Problem gestellt, nicht über das spezifische Fachwissen (bspw. über eine Technologie oder einen Markt) zu verfügen, um die Realisierbarkeit einer Produktidee abschätzen zu können. Dies kann durch die Identifikation und Einbeziehung von externen Experten entsprechend kompensiert werden.

In der vorliegenden Broschüre wird der Ansatz der internetgestützten Expertenidentifikation und seine softwaretechnische Umsetzung in EXPOSE vorgestellt.

¹ nova-net ist ein Verbundforschungsvorhaben (FKZ 01AK701A), das durch das Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Programms Internetökonomie gefördert wird.

2 Zusammenfassung

In den frühen Innovationsphasen spielt die Einbindung von externen Experten insbesondere für kleine Unternehmen eine wichtige Rolle als Wissensträger auf einem Spezialgebiet und zur Bewertung von potenziell relevanten Technologiefeldern, zu denen keine oder nur unzureichende Kompetenzen im Unternehmen existieren. Die Schwierigkeit bei der Integration solcher unternehmensexterner Experten liegt jedoch oftmals bereits darin, geeignete Ansprechpartner zu identifizieren. Die zunehmende Verbreitung des Internets als Informations- und Kommunikationsmedium bietet hier ein bisher unerschöpftes Potenzial für Informationssuche und für die Identifikation externer Experten.

In der vorliegenden Broschüre wird eine strukturierte Vorgehensweise beschrieben, mit der relevante Themenfelder im und mit Hilfe des Internets identifiziert und spezifiziert, sowie, darauf aufbauend, Experten in diesen Themenfeldern gefunden werden können. Im zweiten Teil wird die im Projekt nova-net entwickelte informationstechnische Umsetzung dieser Vorgehensweise der internetgestützten Identifikation von Experten dargestellt. Am Beispiel der Firma INVIA wird dann anhand einer praktischen Problemstellung gezeigt, wie Unternehmen durch diese Vorgehensweise im Bezug auf die Planung von Technologiefeldern in den frühen Innovationsphasen unterstützt werden können.

3 Motivation für die Expertensuche

In den frühen Phasen des Innovationsprozesses spielt Expertenwissen eine maßgebliche Rolle, um das Suchfeld neuer Technologie- oder Anwendungsfelder für Innovationen einzuschränken und für die weitere Planung relevanter Themenfelder festzulegen. Experten werden in diesem Rahmen als Fachkundige gesehen, die sich das Wissen über eine bestimmte Technologie (Produkt/Prozess/Service) oder ein Themenfeld angeeignet haben. Sie finden sich sowohl innerhalb als auch außerhalb einer Organisation. Besonders für kleine und mittlere Unternehmen spielt externes Expertenwissen eine maßgebliche Rolle, da diese meist nicht in der Lage sind, alle relevanten Technologie- und Anwendungsfelder mit internen Ressourcen abzudecken. Die Problematik bei der Identifikation von Experten ist es, dass eine unstrukturierte Suche viel Zeit und Personalressourcen kosten kann und es oft kaum möglich ist, die Expertise der identifizierten Personen realistisch zu bewerten.

Die entwickelte Vorgehensweise soll es Unternehmen hierbei ermöglichen, mit über das Internet öffentlich zugänglichen Daten- und Informationsquellen relevante Technologie- und Anwendungsfelder zu definieren, externe Experten zur Unterstützung in den frühen Phasen des Innovationsprozesses zu identifizieren und mit den vorhandenen Informationen bezüglich ihrer Expertise zu bewerten.

4 Methodischer Ansatz

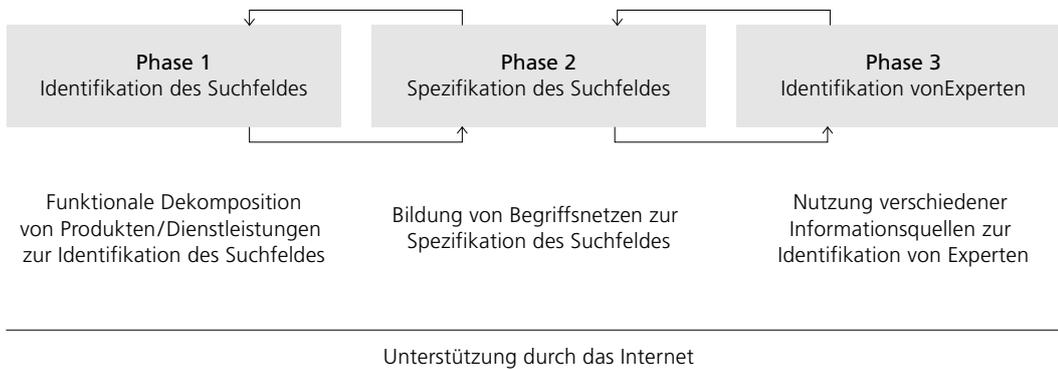


Abbildung 1: Phasen zur internetgestützten Identifikation von Experten.

Der methodische Ansatz der Vorgehensweise basiert auf den frühen Phasen der Trend- oder Technologiebeobachtung. Hierbei handelt es sich, wie in Abbildung 1 dargestellt, um die Identifikation und Spezifikation des Suchfeldes sowie die Sammlung von Informationen, in diesem Fall über die Identifikation und Einbindung externer Experten.

Die Schwerpunktsetzung im Projekt nova-net auf die Identifikation von externen Experten zur Informationssammlung und -bewertung begründet sich in der Möglichkeit, auf diesem Weg einen schnellen Überblick über ein bisher wenig bekanntes Themenfeld zu bekommen. Ein Experte zeichnet sich dadurch aus, dass er auf einem bestimmten Gebiet dauerhaft (nicht zufällig und nicht nur einzelne Male) herausragende Leistungen erbringt.² Die strukturierte Nutzung des Internets als unterstützendes Medium, stellt in diesem Zusammenhang ein erhebliches Verbesserungspotenzial im Vergleich zu den von existierenden Methoden genutzten Medien dar. Dies begründet sich vor allem in der fast unermesslich großen verfügbaren Informationsmenge, aber auch durch

den hohen Verbreitungsgrad des Internets. Andererseits ist gerade die Menge an verfügbaren Informationen die größte Herausforderung, da zur gezielten Suche ein strukturiertes und zielgerichtetes Vorgehen notwendig ist, um diese Informationsfülle auch zu bewältigen.

² Posner (1988)

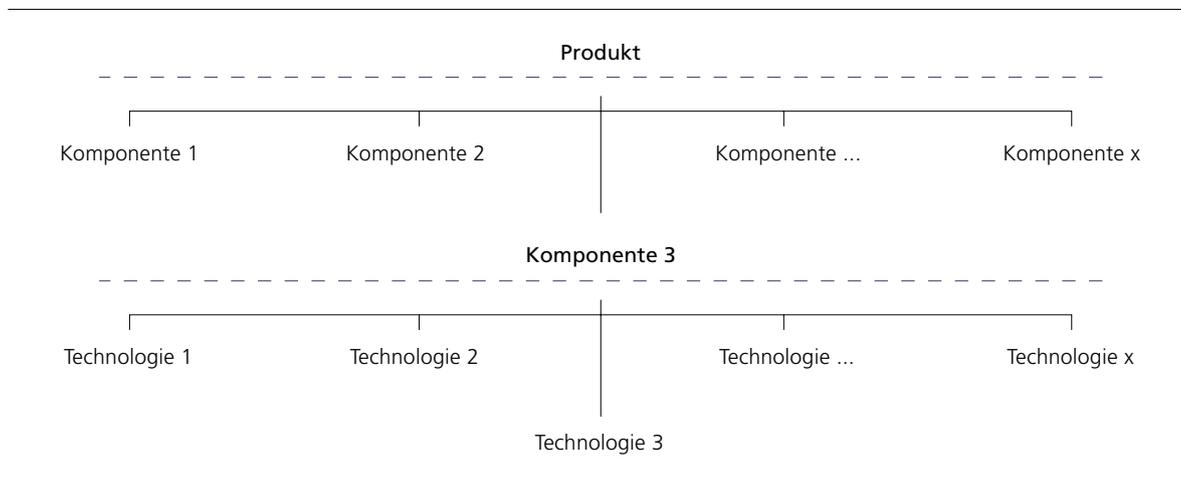


Abbildung 2: Beispiel einer technologischen Dekomposition eines Produktes.

4.1 Identifikation des Suchfeldes

Zur Definition des Suchfeldes gehört in einem ersten Schritt die Identifikation relevanter Themenfelder. Je früher bestimmt wird, welche Themenfelder betrachtet werden, umso zielgerichteter können die Ressourcen zur Sammlung und Bewertung weiterer Informationen eingesetzt werden. Dieser Schritt stellt einen maßgeblichen Wegweiser für die weitere Vorgehensweise dar.

Zur Identifikation des Suchfeldes wird auf die Methode der funktionalen und technologischen Dekomposition zurückgegriffen. Hierbei werden aktuell existierende Produkte in einem ersten Schritt nach funktionalen Aspekten zerlegt³ und diese wiederum in Technologiefelder und Technologien hierarchisch aufgegliedert. Innerhalb der Technologiefelder und Technologien werden nicht nur aktuell eingesetzte, sondern auch potenziell relevante technologische Optionen eingebunden. Die Methode der Dekomposition dient dem Umgang mit komplexen Problemstellungen.⁴ Im Falle der Identifikation des Suchfeldes können kom-

plexe Produkte hierdurch übersichtlich in Technologiefelder untergliedert und für das Produkt relevante Technologien identifiziert werden.

Diese Methode ermöglicht die schnelle und effektive Identifikation relevanter Technologien auf Basis des Produktportfolios von Unternehmen. Hierbei sollte auf eine Reduktion der Komplexität geachtet werden und die Anzahl der Ebenen entsprechend angepasst werden.

Die Rolle des Internets in dieser Phase ist es, die Identifikation von Verbindungen zwischen verwendeten Technologien oder Funktionen und potenziell relevanten Technologien zu unterstützen. Dies erfolgt durch die Analyse existierender Verlinkungen verschiedener Seiten oder Begriffe im Internet. Das Internet bietet hier gegenüber traditionellen Medien den Vorteil, dass auch

³ Specht/Behrens/Kirchhof (1999)

⁴ Ulrich/Eppinger (2004)

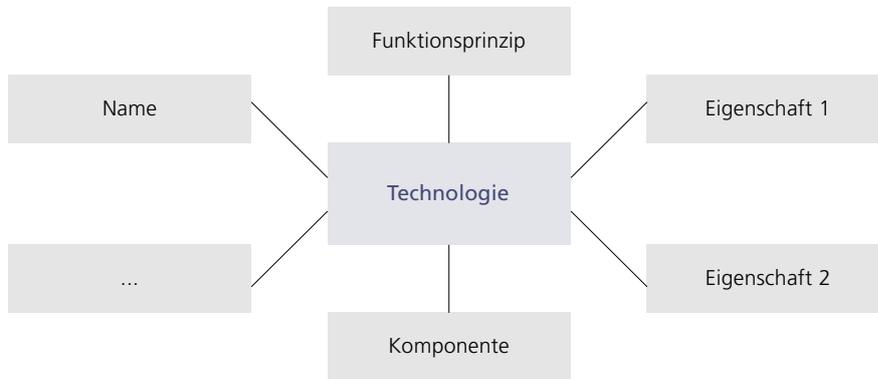


Abbildung 3: Begriffsnetz zur Spezifikation einer Technologie.

Informationen zu aktuellsten Themenbereichen direkt verfügbar sind. Geeignete Instrumente, um diese Verlinkungen im Internet zu identifizieren, stellen zum einen Enzyklopädien dar, die eine Verlinkung zwischen verschiedenen Begriffen beinhalten und zum anderen so genannte Clustersuchmaschinen, die Verbindungen zwischen verschiedenen Themenfeldern im Internet aktiv in ihrer Suche identifizieren und darstellen.

4.2 Spezifikation des Suchfeldes

Sobald potenziell relevante Themenbereiche identifiziert sind, ist der nächste Schritt, diese Themenfelder weiter zu spezifizieren. Dieser Schritt dient sowohl als Basis für die Suche nach Experten im Internet als auch zur Überprüfung von vorhandenem Vorwissen und der Relevanz der ausgewählten Themenfelder.

Zur Spezifikation der Themenfelder werden so genannte Begriffsnetze verwendet. Ein Begriffsnetz besteht aus Schlüsselbegriffen, die ein bestimmtes Themenfeld aus verschiedenen Perspektiven beschreiben. Perspektiven, die häufig Verwendung finden, sind im Falle einer bestimmten Technologie beispielsweise Name, Funktion, maßgebliche Eigenschaft, Funktionsprinzip und notwendige Komponenten dieser Technologie.

Diese Liste kann für andere Beispiele beliebig erweitert, in andere Sprachen übersetzt oder ausgedehnt werden. Für eine effiziente Identifikation von Experten im nächsten Schritt sollte zusätzlich

Instrumente	Beispiele für Webadressen
Enzyklopädien	www.wikipedia.de de.encarta.msn.com
Clustersuchmaschinen	www.clusty.de www.kartoo.com

zum Namen des Themenfeldes ein Minimum von drei Begriffen verwendet werden.

In dieser Phase dient das Internet dazu, mit möglichst geringem Aufwand weitere Informationen zu den relevanten Themenfeldern zu erhalten und auf dieser Basis geeignete Begriffe auszuwählen, um die Themenfelder weiter zu spezifizieren und die Suche zu verfeinern. Hierbei handelt es sich um einen iterativen Prozess zu dem, abhängig vom Themenfeld und dem vorhandenen Vorwissen, verschiedenste Instrumente im Internet verwendet werden können. Enzyklopädien beispielsweise eignen sich sehr gut, um mit geringem Vorwissen einen groben Überblick über ein bestimmtes Themenfeld zu erhalten. Wenn bereits Vorwissen vorhanden ist, bieten Internetcommunities, Foren oder die Webseiten von Forschungsorganisationen die Möglichkeit, das Themenfeld weiter einzugrenzen.

Instrumente	Beispiele für Webadressen
Enzyklopädien	www.wikipedia.de de.encarta.msn.com
Clustersuchmaschinen	www.clusty.de www.kartoo.com
Suchmaschinen	www.google.de www.yahoo.de
Patentdatenbanken	www.google.de/patents www.patentfamily.de
Literatursuchmaschinen	www.scholar.google.de www.dissonline.de
Internetcommunities	www.innovationsreport.de www.kompetenznetze.de
Foren	www.wissenschaft-online.de www.wissen-news.de

4.3 Identifikation von Experten

Wie bereits beschrieben, werden Experten als Fachkundige gesehen, die sich das Wissen über eine bestimmte Technologie (Produkt/Prozess/Service) oder ein Themenfeld angeeignet haben und es Unternehmen damit auf eine schnelle und unkomplizierte Weise ermöglichen, weitergehende Informationen zu einem bestimmten Themenfeld zu erhalten oder dieses zu bewerten. Bei der Identifikation von Experten dient die Spezifikation der Themenbereiche über Begriffsnetze dazu, die Gesamtmenge an möglichen Informationen zu reduzieren und die Qualität der Ergebnisse zu erhöhen.

Das Internet dient bei der Identifikation von Experten dazu, in einem ersten Schritt über verschiedene Suchstränge Personen zu identifizieren, die im Zusammenhang mit dem festgelegten Themenfeld genannt werden und diese in einem zweiten Schritt weitest möglich anhand der verfügbaren Informationen zu bewerten. Zur Bewertung werden beispielsweise Kriterien wie die Art der Informationsquelle, die Anzahl der Nennungen und die Verlinkung zwischen den Ergebnissen verwendet. Als Suchstränge werden in diesem Fall die Art der Informationsquellen wie in Abbildung 4 dargestellt bezeichnet. Diese Trennung der verschiedenen Informationsquellen begründet sich darin, dass beispielsweise Personen, die über Forschungsinstitute identifiziert werden, eher als Experten betrachtet werden können als solche, die über Suchmaschinen identifiziert werden. Damit wird über die Suchstränge bereits eine erste grobe Bewertung der Expertise identifizierter Personen ermöglicht.

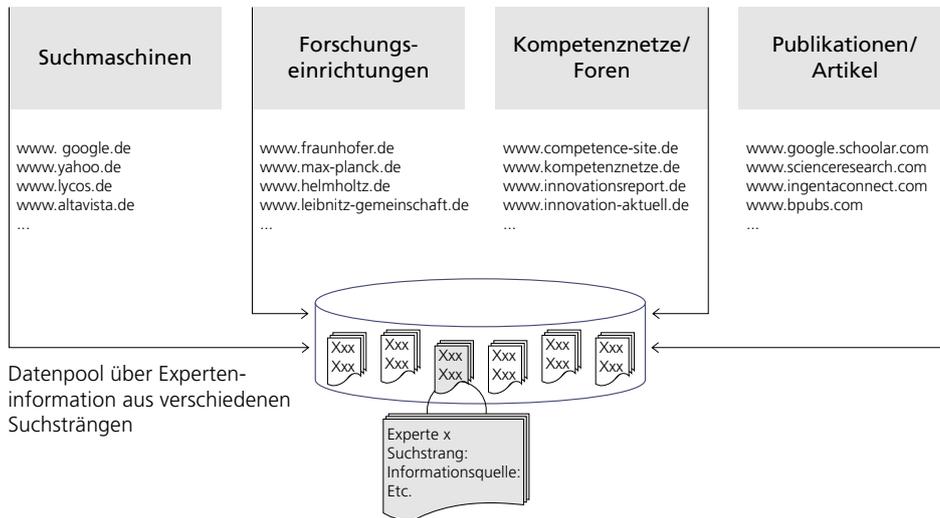


Abbildung 4: Informationsquellen zur internetgestützten Expertensuche

Aufgrund der Stärken und Schwächen der unterschiedlichen Suchstränge im Internet spielt ein ausgewogenes Portfolio an verschiedenen Informationsquellen für die Qualität der Ergebnisse eine wichtige Rolle.

Über die Verwendung verschiedener Suchstränge bei der Identifikation von Experten konnten im Rahmen des nova-net Projektes gute Ergebnisse erzielt werden, die jedoch noch mit einem recht hohen Zeitaufwand verbunden waren. Um diesen weiter zu reduzieren war das Ziel, die vorgestellte Vorgehensweise weitgehend durch die in Abschnitt 5 vorgestellte informationstechnische Lösung zu unterstützen.

Suchstrang	Stärken	Schwächen
Suchmaschinen		
www.google.de www.yahoo.de www.kartoo.com	<ul style="list-style-type: none"> – Sehr große Menge an erreichbaren Inhalten. – Hohe Aktualität. – Vorstrukturierung über Clustersuchmaschinen. 	<ul style="list-style-type: none"> – Keine Qualitätskontrolle. – Keine Thematische Zuordnung der Inhalte. – Zum Teil veraltete oder thematisch irrelevante Ergebnisse.
Forschungsinstitute		
www.fraunhofer.de www.heimholz.de www.max-planck.de	<ul style="list-style-type: none"> – Hohe Qualität der Inhalte (Inhalte werden meist vor der Veröffentlichung geprüft). – Hohe Dichte an Experten. – Thematische Schwerpunktbereiche. 	<ul style="list-style-type: none"> – Inhalte oft theoretisch. – Zum Teil nur begrenzte wirtschaftliche Relevanz der Inhalte.
Kompetenznetze / Foren		
www.kompetenznetze.de www.innovationsreport.de www.innovation-aktuell.de	<ul style="list-style-type: none"> – Verschiedene Blickwinkel (Nutzer, Industrie, Forschung etc). – Hohe Aktualität der Inhalte (auch sehr neue, unsichere Themenbereiche werden frühzeitig diskutiert). 	<ul style="list-style-type: none"> – Inhalte und Qualität stark von der Anzahl und Art der Beteiligten abhängig.
Publikationsdatenbanken		
www.google.de/scholar www.scienceresearch.com www.ingentaconnect.com	<ul style="list-style-type: none"> – Hohe Qualität (nur bereits von unabhängiger Seite bewertete Inhalte werden veröffentlicht). – Je nach Art der Veröffentlichung wissenschaftlich oder anwenderorientiert. 	<ul style="list-style-type: none"> – Veröffentlichungen oft auf einen begrenzten Autorenkreis begrenzt.

5 IT-Unterstützung durch »EXPOSE«

Die in Abschnitt 4 vorgestellte Vorgehensweise bietet einen methodischen Leitfaden für die Suche nach Experten im Internet. Die mit dieser Methode einhergehenden Arbeitsschritte sind jedoch mitnichten trivial. Insbesondere ist die vom Suchenden zu handhabende Datenmenge, bedingt durch die Fokussierung auf das Internet, äußerst umfangreich. Aus dieser Motivation heraus wurde in enger Zusammenarbeit mit Anwendern aus Industrie und Forschung eine Software entwickelt, die wesentliche Teile der vorgestellten Methode informationstechnisch umsetzt und über die Bereitstellung reiner Standard-Suchmaschinentheorie hinaus geht. Der Schwerpunkt der Arbeiten lag dabei auf den in den Abschnitten 4.2 und 4.3 genannten Techniken zur Spezifikation des Suchfeldes und zur Identifikation von Experten. Zudem wurde bei der Entwicklung besonderes Augenmerk auf die Benutzbarkeit und die Unterstützung von Personen gelegt, die keine Experten auf dem Gebiet der Websuche sind. Mit der dabei entstandenen Software EXPOSE wurde eine modulare Plattform geschaffen, die es dem Benutzer ermöglicht, das WWW effizient nach Experten zu frei definierbaren Themengebieten zu durchsuchen.

5.1 Anforderungen an die Software

Aus den oben genannten Rahmenbedingungen ergibt sich eine Reihe von Schwierigkeiten für die informationstechnische Unterstützung der Expertensuche im Internet. Die Herausforderungen für eine technische Umsetzung lassen sich auf die im Folgenden dargestellten Kernprobleme zurückführen. Sie gelten für die Suche im WWW allgemein und insbesondere für die Suche nach Experten.

- Das WWW stellt eine rasant wachsende Informationsfülle bereit, in der es sich zurecht zu finden gilt. Simple Methoden, wie die Suche nach Schlüsselwörtern, sind bei großen Suchmaschinen wie MSN, Google oder Yahoo derzeit das Mittel der Wahl. Während diese weite Teile des WWW abdecken und damit dem Benutzer zugänglich machen, unterstützen sie in erster Linie einfache Suchanfragen mit denen aber beispielsweise ein komplexes Themengebiet nur schwer zu definieren ist. Eine weitergehende Semantik, wie in diesem Fall die Identifikation von Personen mit besonderer Expertise auf dem spezifizierten Interessenfeld, wird von diesen Suchmaschinen nicht angeboten. Damit auch Benutzer, die im Umgang mit herkömmlichen Suchmaschinen wenig versiert sind, von EXPOSE profitieren, war es ein wichtiges Ziel, einen einfachen Zugang zur Suche zu schaffen. Dazu mussten Möglichkeiten gefunden werden, die Defizite herkömmlicher Suchmaschinen, insbesondere im Bereich der Suchspezifikation, zu kompensieren.
- Um die gegebenen Informationsmengen für den Benutzer handhabbar zu machen, muss durch die Suchmaschine eine sinnvolle Vorauswahl getroffen bzw. die Ergebnisliste nach vermuteter Relevanz sortiert werden. Für diesen Schritt sind Metadaten zu den betrachteten Online-Ressourcen hilfreich, wie sie von

der Semantic-Web-Bewegung vorgeschlagen werden. Mittels solcher Metadaten kann eine automatische Verarbeitung von Online-Ressourcen deutlich effizienter und effektiver erfolgen, da sie ein zumindest grundlegendes Textverständnis ermöglichen. Da derartige Metadaten jedoch nur äußerst eingeschränkt verfügbar sind, wird in EXPOSE ein anderer Ansatz gewählt, um die Bedeutung und Relevanz einer Ressource zu bestimmen. Die Berechnung und Analyse statistischer Merkmale der Ressourcen und ihrer Verknüpfungen steht dabei im Mittelpunkt.

- Da herkömmliche Suchmaschinen lediglich Dokumente liefern, die zu der vom Benutzer spezifizierten Suchanfrage passen, überlassen sie die weitere Verarbeitung des Inhalts dieser Dokumente dem Nutzer. Insbesondere bei der Suche nach Experten ist dieser der Dokumentensuche nachgelagerte Schritt äußerst aufwändig und komplex. Hier gilt es, den Benutzer entsprechend zu unterstützen, um zum einen die Ergebnismenge handhabbar zu machen und zum anderen die Identifizierung von Experten und die Bewertung ihrer Expertise zu ermöglichen. Die Visualisierung von Suchergebnissen und Bewertungen spielt in diesem Zusammenhang eine große Rolle, da grafische Zusammenhänge oft schneller und einfacher erfasst werden können als ihre Darstellung in textueller Form.

Aus diesen Anforderungen ergeben sich unmittelbar die Schwerpunkte der Software: die Unterstützung des Benutzers bei der Spezifikation des Themengebietetes, die eigentliche Suche nach themenrelevanten Dokumenten, sowie die Identifizierung von Personen und die Bewertung deren Expertise im Kontext des gesuchten Themenfeldes.

5.2 Technische Umsetzung

Die im Vorfeld genannten Anforderungen bildeten die Grundlage für die Entwicklung von EXPOSE, der nova-net Expertensuchmaschine. Abbildung 5 zeigt eine Übersicht der in EXPOSE umgesetzten Schritte und Interaktionen zwischen diesen. Einige zentrale Komponenten der Software und deren Zusammenspiel werden im Folgenden vorgestellt.

Schritt 1 – Die Spezifikation des Informationsbedarfs

In der Regel ist es für den Benutzer eine mehr oder minder große Herausforderung, seinen Informationsbedarf bzw. das Themengebiet für die Suchmaschine anhand einiger weniger Stichworte zu spezifizieren. Ist jedoch bereits diese Spezifikation unpräzise, so wird sich dies im Allgemeinen auch in dem von der Suchmaschine zurückgelieferten Ergebnis widerspiegeln. Daher wird dieser Phase des Suchprozesses von EXPOSE besondere Aufmerksamkeit geschenkt und der Benutzer entsprechend bei der Formulierung von Suchanfragen auch durch alternative Ansätze unterstützt.

Beispieldokumente als intuitive Form der Spezifikation

Deutlich einfacher als die Spezifikation mittels weniger Stichworte ist es, der Suchmaschine Beispiele für Sachverhalte oder Themen zu geben. Oftmals ist intuitiv klar, dass ein bestimmter Text relevant ist, ohne dass man als Benutzer in der Lage wäre, die einzelnen Stichworte des Textes zu bestimmen, welche ihn relevant erscheinen lassen bzw. die nötige Schärfe bieten, ihn gegen andere Texte abzugrenzen. Ein alternativer Ansatz zur Spezifikation mittels Stichworten besteht daher

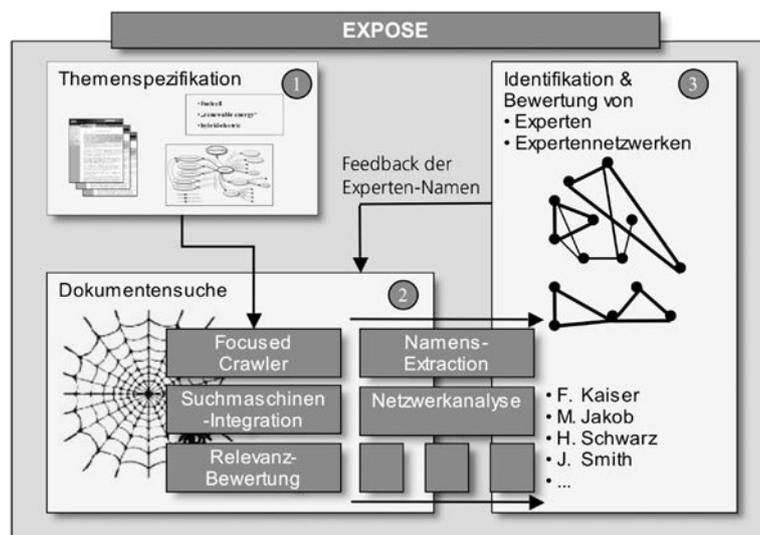


Abbildung 5: Softwaretechnische Umsetzung der Expertensuche.

darin, den Informationsbedarf anhand von Beispielen bzw. Beispieldokumenten zu spezifizieren, also eine bestimmte Textkategorie vorzugeben, in welcher man auch die Ergebnisse erwartet. Zwar ist auch dies keine exakte Methode, es ist damit jedoch möglich, eine hinreichend genaue Spezifikation für den Einstieg in die Suche zu erlangen.

Um aber solche Beispieldokumente im Rahmen einer Websuche zu nutzen, wird ein System benötigt, das ausgehend von Beispieltextrn weitere Dokumente findet, die eine möglichst große inhaltliche Nähe zu den Beispielen aufweisen. Technisch erfolgt die Bestimmung der inhaltlichen Nähe eines Dokuments zur Beispielmengr bei EXPOSE anhand der Berechnung verschiedener statistischer Merkmale der jeweiligen Dokumente. Neben der rein technischen Problematik des Ähnlichkeitsvergleichs besteht aber auch die Schwierigkeit, dem System die notwendigen Beispieldo-

kumente zu liefern. Dieses Problem stellt sich in der Praxis jedoch als das Kleinere dar, da solche Beispieldaten oftmals bereits vorhanden sind oder mittels herkömmlicher Suchmaschinen mit geringem Aufwand gefunden werden können. Dabei genügt bei der Arbeit mit EXPOSE meist eine relativ einfache Suchanfrage verbunden mit geringem Analyseaufwand, um zumindest einige Beispieldokumente zu finden, da im Fall des Einstiegs in die Suche kein gesteigerter Wert auf Exaktheit und Umfang des gesamten Suchergebnisses gelegt werden muss.

Auf den ersten Blick scheint damit nicht viel gewonnen, auch hier müssen die Ergebnisse herkömmlicher Suchmaschinen manuell durchforstet werden. In Hinblick auf das eigentliche Ziel, die Suche nach Experten, hingegen, zeigt sich der Vorteil dieses Ansatzes. Mittels einer unpräzisen Anfrage an Standardsuchmaschinen wird eine



Abbildung 7: Nach Relevanz geordnete Liste der heruntergeladenen Dokumente.

beschreiben. Allerdings ist die Semantik der Verbindungen unbekannt. Wo also bei einer vollwertigen Ontologie bspw. aus der Verbindung zwischen »Brennstoffzelle« und »Energiequelle« ersichtlich ist, dass es sich um eine »ist-ein«-Beziehung handelt, so kann in einem Begriffsnetz lediglich festgestellt werden, dass beide Begriffe in einem gewissen, wie auch immer gearteten Zusammenhang stehen. Dennoch ergibt sich daraus ein großer Vorteil für die Suche:

Über eine grafische Schnittstelle kann der Benutzer schnell den begrifflichen Kontext eines Themenfeldes erkennen und die Suchanfrage präzise formulieren bzw. modifizieren, da von der Begriffsnetz-Komponente verwandte Begriffe vorgeschlagen bzw. Abgrenzungen zu anderen Themen deutlich gemacht werden (Abbildung 6).

Schritt 2 – Die Suche nach relevanten Dokumenten

Mittels einer auf diese Art erstellten Suchanfrage kann nun die eigentliche Suchkomponente von EXPOSE angestoßen werden. Wie in Abschnitt 4.3 dargestellt, lassen sich die besten Ergebnisse erzielen, wenn verschiedene sog. Suchstränge genutzt werden. In einem ersten Schritt werden dazu mehrere Systemkomponenten angestoßen, die das WWW nach themenrelevanten Dokumenten durchsuchen. Das Ergebnis dieses Schrittes ist eine nach thematischer Relevanz geordnete Liste von Suchergebnissen (Abbildung 7). Das eigentliche Ziel, die Identifikation von Experten, ist diesem Schritt nachgelagert und wird entsprechend als Schritt 3 des Suchprozesses beschrieben.

Focused Crawling

Um die Suchanfragespezifikation mittels Beispieldokumenten sinnvoll nutzen zu können, wird zusätzlich zur Ähnlichkeitsbestimmung (vgl. Abschnitt Beispieldokumente als intuitive Form der Spezifikation) eine weitere Komponente benötigt, die für die Beschaffung der zu prüfenden Dokumente zuständig ist. Herkömmliche Suchmaschinen erwarten vom Benutzer die Eingabe von Schlüsselwörtern und liefern diejenigen Dokumente als Ergebnis zurück, die diese Schlüsselwörter enthalten, evtl. unter Berücksichtigung bestimmter Verknüpfungen zwischen den Suchbegriffen. Durch die erläuterte Änderung der Suchspezifikation, weg von Schlüsselwörtern, hin zu Beispieldokumenten, entfällt aber auch die Eingabe für Standardsuchmaschinen. Daher erfolgt die Beschaffung der zu prüfenden Dokumente in EXPOSE durch einen eigenständigen Crawler.

Zum Einsatz kommt dabei ein sog. Focused Crawler.⁵ Beim Focused Crawling werden ausgehend von einer Menge an Dokumenten, den Beispieldokumenten, die in diesen Dokumenten enthaltenen Verweise (Hyperlinks) mittels eines Crawlers verfolgt und die sich dahinter verbergenden Dokumente geladen. Daraufhin wird die inhaltliche Nähe der geladenen Dokumente zu den Beispieldokumenten verglichen (vgl. Abschnitt Beispieldokumente als intuitive Form der Spezifikation). Abhängig vom Ausgang dieses Vergleichs und unter Berücksichtigung weiterer automatisch berechneter oder benutzergesteuerter Kriterien werden die Dokumente als Ergebnis markiert und die darin enthaltenen Verweise rekursiv weiterverfolgt. Die Idee dahinter ist, dass Dokumente, die inhaltlich verwandt sind, oft über Verweise miteinander verknüpft sind, wobei solche Verweise nicht von allen Dokumenten eines Themas zu allen anderen desselben Themas gehen müssen, sondern diese auch über mehrere Zwischenschritte verbinden können. Im Unter-

schied zu Standard-Suchmaschinen werden somit lediglich sehr kleine Teile des Webs durchsucht, wobei der Fokus auf der themenrelevanten Teilmenge der verfügbaren Dokumente liegt, die stark miteinander verknüpft sind. Der Crawler folgt daher zuerst den Verweisen aus relevanten Dokumenten bevor er solchen aus irrelevanten Dokumenten nachgeht. Die Suchanfrage für den EXPOSE-Crawler kann dabei sowohl anhand bereits vorliegender Dokumente als auch durch Dokumente, die mittels herkömmlicher Suchmaschinen gefunden wurden, spezifiziert werden.

Offensichtlich ist die Nutzung eines Focused Crawlers nicht mit dem Einsatz einer Suchmaschine vergleichbar. Während eine Suchmaschine in der Regel innerhalb von Sekundenbruchteilen ein Suchergebnis präsentiert, kann ein Focused Crawler durchaus mehrere Minuten bis Stunden laufen. Ursächlich hierfür ist, dass bei einer Suchmaschine das Crawling, also das Durchlaufen des Internets sowie der Aufbau des Index, zeitlich vor der Benutzeranfrageverarbeitung erfolgt bzw. seine Aktualisierung gänzlich unabhängig von der Benutzeranfrageverarbeitung ist. Ein Focused Crawler hingegen kann erst dann mit der Suche starten, wenn er die Spezifikation für seine Fokussierung kennt, andernfalls hätte er nicht die Eigenschaft »focused«. Der Vorteil gegenüber den etablierten Suchmaschinen ist jedoch, dass eine viel detailliertere Bewertung der zurückgelieferten Ergebnisse möglich ist und neben rein syntaktischen Kriterien auch netztopologische und teilweise semantische Aspekte in die Bewertung eingebunden werden können. Mittels verschiedener Visualisierungen erleichtert EXPOSE dem Benutzer die Beurteilung der Ergebnisse indem es eben diese Zusammenhänge übersichtlich darstellt. Abbildung 8 beispielsweise zeigt die zentralen Internet-Adressen, die die für eine Suchanfrage relevantesten Dokumente bereithalten. Anhand solcher Erkenntnisse lässt sich der Focused Crawler an die in Abschnitt 4.3 angeführten Such-

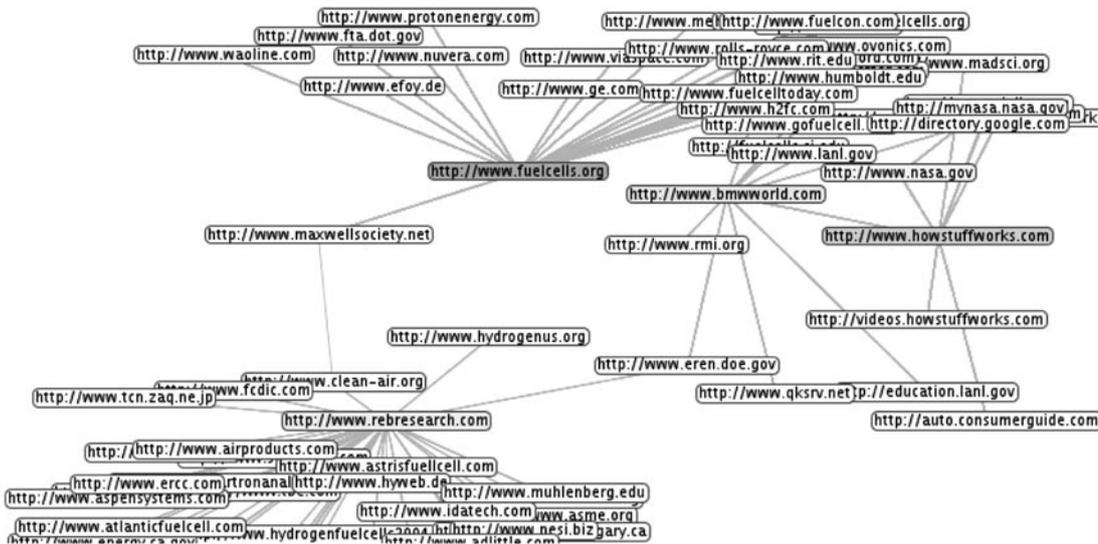


Abbildung 8: Konzentration des Focused Crawlers auf Quellen mit vielen relevanten Dokumenten.

stränge anpassen und beispielsweise auf wissenschaftliche Publikationen oder auf Dokumente einer bestimmten Quelle fokussieren, indem entsprechende Quellen mit einem stärkeren Gewicht in die Bewertung eingehen.

Alternativer Suchstrang: Ergänzung durch Standard-Suchmaschinen

Eingangs wurde auf die teilweise mangelnde Suchmaschinenkompetenz des typischen Benutzers von Suchmaschinen hingewiesen und erläutert, inwiefern dieses Problem dazu beiträgt, dass die Expertensuche mittels herkömmlicher Suchmaschinen oft nur unbefriedigend gelöst werden kann. Offensichtlich ist aber auch das »Focused Crawling« keine Lösung frei von jedweden Nachteilen. Der gravierendste hierbei ist, dass nur Quellen gefunden werden können, die tatsächlich

auf mehr oder weniger direktem Wege mittels Verweisen verknüpft sind, da der Focused Crawler bei fehlenden Verbindungen unabhängige Dokumenteninseln nicht erreichen kann. Ein Focused Crawler wird beispielsweise nur in sehr engen Grenzen Verweisen folgen, die keinen Bezug zum Thema aufweisen, da er ansonsten seine Fokussierung verlieren würde und damit nicht mehr in akzeptabler Zeit Ergebnisse liefern könnte. Das im Information Retrieval allgegenwärtige Problem des Auseinanderdriftens von Präzision & Rücklauf (Precision and Recall)⁶ zeigt sich also auch hier: Einer relativ hohen Präzision des Verfahrens steht ein geringer Rücklauf gemessen an allen theoretisch vorhandenen aber nicht gemeldeten Treffern gegenüber wenn die relevanten Quellen nicht oder nur schwach verlinkt sind.

5 Chakrabarti/van der Berg/Dom (1999)
 6 Ferber (2003)

Um einer solchen Reduktion des Rücklaufs zu begegnen sind in EXPOSE zusätzlich zum Focused Crawler auch Standard-Suchmaschinen integriert. Suchmaschinen wie Google oder Yahoo indexieren mittlerweile Milliarden von Seiten des öffentlich zugänglichen WWW und machen diesen Index über Webschnittstellen und/oder Programmschnittstellen zugänglich. Damit besitzen sie umfangreiche Informationen über Inhalt und Struktur weiter Teile des WWW. Leider sind diese Daten nicht in ihrem vollen Umfang für Außenstehende nutzbar, da die angebotenen Schnittstellen lediglich einfache, auf Seiten der Suchmaschinenbetreiber wenig rechenintensive Anfragen ermöglichen. Dennoch lässt sich die Datensammlung der Standard-Suchmaschinen auch für das Focused Crawling gewinnbringend einsetzen – weniger um die Präzision zu erhöhen, als vielmehr zur breiteren Streuung der Ergebnisse und damit zur Steigerung des Rücklaufs. Im Folgenden sollen zwei in EXPOSE zum Einsatz kommende Techniken vorgestellt werden, die sich das Wissen herkömmlicher Suchmaschinen um Inhalte des WWW zunutze machen.

Schlüsselwortextraktion aus relevanten Quellen

Betrachtet man die Worte eines Dokuments unter dem Gesichtspunkt, welchen Beitrag sie jeweils zum semantischen Inhalt liefern, so ist offensichtlich, dass einige wenige Begriffe das Themengebiet abstecken und die Restlichen zur inhaltlichen Konkretisierung und Differenzierung beitragen, bzw. lediglich Füllwörter darstellen. Letztere sind für die Suche nicht von Interesse, und werden meist von weiteren Betrachtungen ausgenommen (Beispiel: sog. Stopwörter wie »als«, »an«, »bei« etc.). Auch die Erfassung der Semantik eines Textes, hier als inhaltliche Konkretisierung und Differenzierung angesprochen, ist ein hochgradig rechenintensives Unterfangen, das heute noch nicht für allgemeine Texte ohne Einschränkungen

möglich ist. Es sei dabei wieder auf die latente Unsicherheit hingewiesen, die ohnehin den gesamten Suchprozess begleitet. Es bleibt also letztlich nur die Orientierung an den wenigen Begriffen, die das jeweilige Themengebiet skizzenhaft beschreiben. Wurden im Rahmen der Suche bereits relevante Quellen identifiziert, so kann EXPOSE die Schlüsselwörter der Texte extrahieren, um daraus eine Suchanfrage für Standard-Suchmaschinen zu generieren. Dies lässt sich insbesondere mit der Begriffsnetzkomponente zu einem nützlichen Werkzeug verbinden, indem der begriffliche Kontext der extrahierten Schlüsselwörter dargestellt wird und die Anfrage somit leicht konkretisiert werden kann.

Mit einer solchen Schlüsselwortextraktion lässt sich auch genau jenes Problem des Focused Crawlers umgehen, dass verwandte Quellen nicht zwangsläufig durch einen kurzen Verweispfad verbunden sind. Eine Standardsuchmaschine kennt im Allgemeinen deutlich mehr Quellen, welche die identifizierten Suchbegriffe enthalten, auch wenn diese nicht oder nur über relativ lange Verweispfade miteinander verbunden sind. EXPOSE kombiniert daher den Focused Crawler mit Ergebnissen anderer Suchmaschinen und stellt damit die Suche und Analyse auf eine breitere Datenbasis. Das erhöht entsprechend die Wahrscheinlichkeit, tatsächlich relevante Quellen und damit relevante Personen zu finden. Offensichtlich erhöht sich aber auch der Aufwand für die Prüfung der Suchergebnisse, da von den Suchmaschinen gemeldete Ergebnisse tendenziell mehr irrelevante Dokumente listen, als die eines Focused Crawlers. Da sich der Aufwand für diese (automatisierte) Relevanzprüfung aber in Grenzen hält, kann er meist in Kauf genommen werden, wenn dafür in der Summe der Rücklauf steigt.

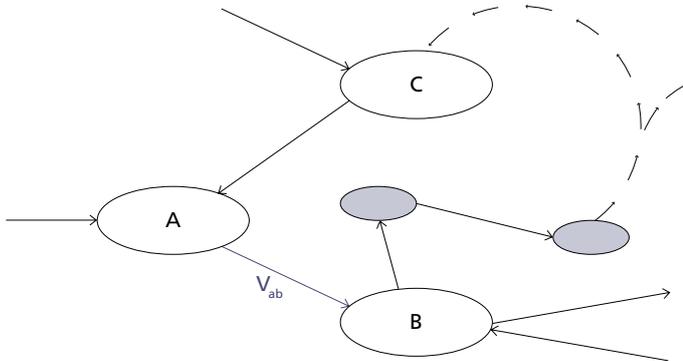


Abbildung 9: Das WWW als Linkgraph – gerichtete vs. ungerichtete Kanten.

Einen Spezialfall stellt in diesem Zusammenhang die Extraktion nicht nur beliebiger, häufig vorkommender Begriffe, sondern speziell die von Personennamen dar. So wie Suchmaschinen meist weitere Dokumente mit identifizierten Schlüsselwörtern indiziert haben, auch wenn diese nicht unmittelbar miteinander verknüpft sind, befinden sich auch mit einer gewissen Wahrscheinlichkeit weitere Dokumente in ihrem Index, die weitere Nennungen von identifizierten Personen enthalten. Unterstellt man, dass eine Person mit Kompetenzen auf bestimmten Gebieten nicht nur einmalig in diesem Kontext in Erscheinung tritt, so lassen sich auch über diesen Pfad zielgerichtet weitere Informationen gewinnen. Dieser Ansatz birgt also entscheidend mehr Semantik als die reine Fokussierung auf beliebige Begriffe. Entsprechend werden in EXPOSE auch die auf relevanten Dokumenten identifizierten Namen als Eingabe für Anfragen an Standardsuchmaschinen genutzt. In Verbindung mit den themenkennzeichnenden Schlüsselworten ist diese Art der Integration von Suchmaschinen in den Focused Crawler äußerst gewinnbringend.

Rückwärtsverweise zur Suchraumerweiterung

Das Focused Crawling basiert allgemein auf dem Prinzip, Verweisen zu folgen, die in bereits bekannten Quellen gefunden wurden. Gelangt man also von einem Dokument A über einen Verweis V_{ab} zum Dokument B, so ist dort nicht unbedingt auch einen Verweis V_{ba} zum Dokument A vorzufinden. Der Rückweg ist damit – wenn überhaupt – nicht zwangsläufig in einem Schritt möglich (Abbildung 9).

Offensichtlich will man beim Focused Crawling diesen Rückwärtsschritt auch nicht durchführen, der Inhalt des Dokuments A ist ja bereits bekannt. Wenn nun aber A noch nicht besucht sondern lediglich B über einen anderen Pfad erreicht wurde, dann ist nicht gewiss, ob A jemals erreicht wird, obwohl evtl. ein inhaltlicher Zusammenhang besteht und die Verweis-Entfernung von A nach B lediglich »eins« betragen würde. Um solche Rückwärtsverweise zu ermitteln, nutzt EXPOSE wiederum das Wissen herkömmlicher Suchmaschinen: Diese bieten oftmals die Möglichkeit, gezielt nach Seiten zu suchen, die einen Verweis auf eine bestimmte andere Seite gesetzt haben. Für den Fo-

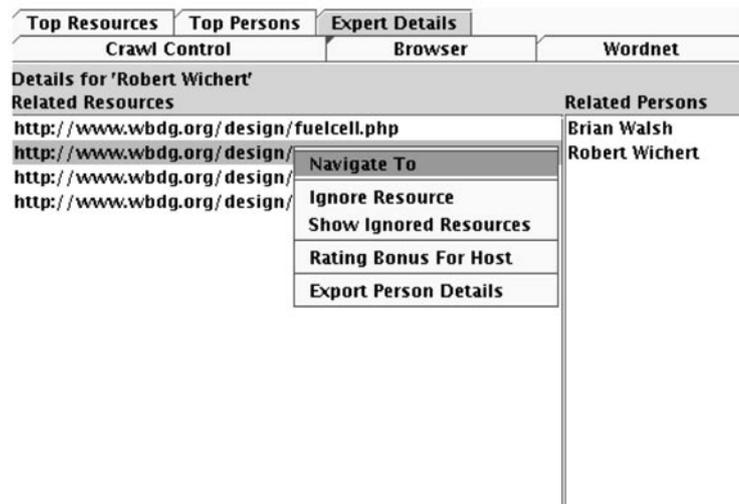


Abbildung 11: Weitere Details zu einem potenziellen Experten.

einen, die bereits gestellte Schlüsselwort-Suchanfrage zu modifizieren, zum anderen aber auch, die Liste der Beispieldokumente für den Focused Crawler anzupassen.

Schritt 3 – Die Identifikation von Experten und deren Bewertung

Mit der Identifikation relevanter Ressourcen ist, wie in Abschnitt 4.3 beschrieben, die Grundlage für Extraktion von Experteninformationen geschaffen. Um jedoch in diesen Ressourcen auch zuverlässig relevante Personen zu identifizieren, bedarf es umfangreicher syntaktischer und linguistischer Analysen, die die Semantik des Textes erfassen. Dies kann derzeit von Softwaresystemen noch nicht im nötigen Umfang geleistet werden, sofern nicht ein spezielles Training der Systeme auf Sprache, Struktur und Inhalt der zu analysierenden Texte erfolgt ist. Letzteres wiederum lässt

sich aufgrund des zeitlichen und daher auch finanziellen Aufwands im Kontext der Expertensuche nicht umsetzen. Kann jedoch eine gewisse Unsicherheit und damit einhergehend eine gewisse Fehlerhäufigkeit in Kauf genommen werden, so lässt sich der Aufwand in dieser Phase deutlich reduzieren. Für EXPOSE haben sich folgende vereinfachte Verfahren als zielführend herausgestellt.

- Suche von Personennennungen im Text anhand einer umfangreichen Namensdatenbank. Ein großer Teil aller Personennennungen kann mittels dieses einfachen Verfahrens bereits identifiziert werden, sofern einige Besonderheiten, wie beispielsweise die häufig genutzte Abkürzung von Vornamen (»K. Müller«), unterschiedliche Wortendungen (»K. Müllers Kompetenzen«) oder Änderungen der Schreibweise (»Müller, Karl«) berücksichtigt werden.

- Suche von Schlüsselbegriffen, die auf eine Personennennung in ihrer unmittelbaren Nachbarschaft schließen lassen. Beispiele hierfür sind »laut Müller«, »Müller meint/sagt/schreibt/erklärt/behauptet/...«, »Frau Müller«, und ähnliche. Auf diese Art lassen sich auch einzelne Vor- oder Nachnamen identifizieren, die andernfalls aufgrund von Mehrdeutigkeiten nicht sicher als Namen hätten identifiziert werden können. Der Begriff »Müller« existiert beispielsweise nicht nur als Name sondern auch als Berufsbezeichnung.

Alle Personen die in einem Text nach dem oben genannten Verfahren identifiziert wurden, werden von EXPOSE als »potenzielle Experten« zum betreffenden Thema registriert. Offensichtlich erlaubt jedoch allein die Nennung einer Person in einem Text noch keinen Rückschluss auf Ihre Expertise in dem vom diesem Text behandelten Themengebiet. Weitergehende Analysen sind daher notwendig, um die betreffende Person entsprechend einzuordnen. Auch hierbei unterstützt EXPOSE den Benutzer, indem verschiedene Eigenschaften des potenziellen Experten sowie Kennzahlen automatisch berechnet, ausgewertet und übersichtlich präsentiert werden (vgl. Abschnitt 4.3, sowie Abbildung 11). Zusätzlich erstellt EXPOSE ein Ranking der gefundenen Personen, wobei der Benutzer jederzeit die dieses Ranking bestimmenden Kriterien aus der Bewertung aus- und wieder einschließen bzw. in der Gewichtung anpassen kann. Im Folgenden werden beispielhaft zwei der eingesetzten Rankingkriterien beschrieben.

Häufigkeit und Kontext von Personennennungen

Auf Basis der im vorangegangenen Schritt erstellten Personenliste nimmt EXPOSE verschiedene Berechnungen und Analysen vor. Ein wichtiges Indiz dafür, dass eine identifizierte Person tatsächlich Expertise auf dem gesuchten Themenfeld besitzt ist, dass sie in diesem Zusammenhang häufiger in Erscheinung tritt. Wenn eine Person in mehreren als relevant klassifizierten Dokumenten identifiziert wird, so lässt sich daraus folgern, dass sie tatsächlich gewisse Kenntnisse auf dem betrachteten Gebiet besitzt. Ferner liegt es nahe, solchen Personen eine höhere Bewertung zuzuordnen, die in zentralen Bereichen des Textes identifiziert wurden. Zentral ist hier sowohl im Sinne der Textformatierung, als auch in semantischem Sinne zu verstehen. In Bezug auf die Textformatierung wird dem Rechnung getragen, indem die Position des Vorkommens relativ zu Textanfang und -länge bewertet wird. Die Motivation für dieses Kriterium ist, dass die Struktur der meisten Webseiten mehr oder weniger einem Schema folgt, bei dem an den Seitenrändern oftmals Navigationsleisten oder Werbebanner stehen, wohingegen sich der eigentliche Inhalt im inneren Bereich der Seite befindet. Inwiefern ein Name jedoch semantisch gesehen zentral für einen Text ist, lässt sich ohne aufwändige linguistische Analysen nicht abschließend ermitteln. Eine einfache und dennoch gute Ergebnisse liefernde Heuristik besteht nun darin, die Position der Personennennung relativ zum Auftreten relevanter Begriffe des Themenfeldes zu ermitteln und diese Nähe zu bewerten. Je dichter eine Namensnennung beispielsweise auf eine technologische Beschreibung folgt, desto größer ist die Wahrscheinlichkeit, dass diese Person Kenntnisse auf dem besagten Gebiet besitzt. Diese Daten sind daher ein wichtiger Bestandteil der Bewertung einer jeden identifizierten Person und tragen zum von EXPOSE erstellen Ranking der vorgeschlagenen Experten bei.



Abbildung 12: Verknüpfung zwischen Personen und Quellen.

Analyse von Kommunikationsstrukturen

Weitere Informationen über die vermeintliche Expertise einer Person kann über ihr Kommunikationsverhalten ermittelt werden bzw. allgemeiner über ihre Nennungen gemeinsam mit anderen Personen innerhalb eines Dokuments. Beispielsweise deutet ein reger Austausch über einen längeren Zeitraum mit bestimmten Personen auf ein gesteigertes Interesse am Themengebiet hin und damit möglicherweise auch auf eine besondere Kompetenz. Dies kann zum einen an der Häufigkeit des Austauschs festgemacht werden, zum anderen auch an den Kommunikationspartnern selbst. Sind an einer solchen Diskussion Personen beteiligt, die bereits als Experten identifiziert wurden, so deutet dies darauf hin, dass auch die anderen Beteiligten gewisse Kompetenzen in

diesem Bereich besitzen oder evtl. durch die Diskussion erlangt haben.

EXPOSE bezieht diese Kriterien in die Expertisebewertung dahingehend ein, dass Vernetzungsstrukturen von Personen über den gefundenen und nach Relevanz bewerteten Ressourcen aufgebaut werden (Abbildung 12). Innerhalb solcher Netze werden dann Kriterien wie die Netzwerkdichte um eine Person oder Ressource als weiterer Anhaltspunkt zur Bestimmung der Expertise einer Person herangezogen. Damit kann die direkte Kommunikation zwischen mehreren Personen analysiert werden, bspw. innerhalb eines Web-Forums. Es können daneben aber auch Zitate bzw. Verweise auf andere Personen berücksichtigt werden, ebenso wie die gemeinsame Nennung von zwei Personen im Kontext eines Dokuments.

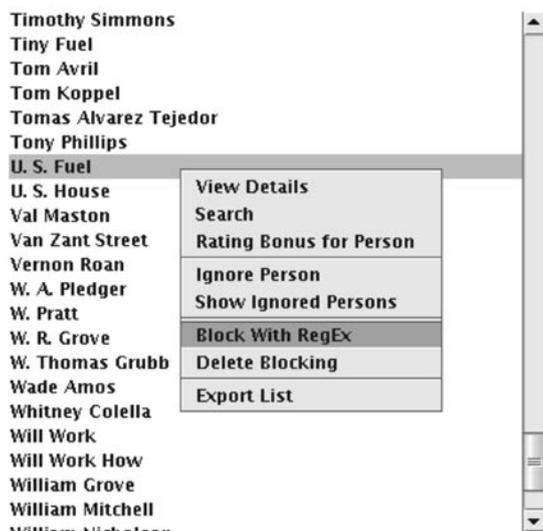


Abbildung 13: Personenliste.

Dennoch wird ein derartiges System immer eine gewisse Fehlerquote aufweisen. Diese Fehlerquote aber möglichst gering zu halten und den Benutzer bei der Auswahl seiner Experten zu unterstützen ist das Ziel von EXPOSE. Eine abschließende Bewertung jedoch muss immer durch den Benutzer erfolgen – dafür bietet EXPOSE die notwendigen und hilfreichen Werkzeuge.

Aus den oben genannten Punkten ist ersichtlich, dass eine absolute Aussage über die Expertise der identifizierten Personen nicht gemacht werden kann. Es handelt sich bei allen Ansätzen letztlich um die Arbeit mit Wahrscheinlichkeiten. Um die Wirkung der Fehlerkennungen und -bewertungen möglichst gering zu halten, gibt EXPOSE dem Benutzer jedoch verschiedene Möglichkeiten der Einflussnahme. Beispielsweise können häufig fehlerhaft erkannte Namen über eine Mustererkennung von der Bewertung ausgeschlossen werden (siehe Abbildung 13, »U. S. <xxx>« da »U.S.« in der Regel die »Vereinigten Staaten« meint und nicht eine Person mit den Initialien »U.S.«), es können Zugaben oder Abzüge in der Bewertung von bestimmten Domains oder Personen (siehe Abbildung 13, »William Grove« als (Mit-) Erfinder der Brennstoffzelle) vergeben werden und vieles mehr.

6 Praxisbeispiel

Kurzprofil INVIA GmbH, Lucido

Mit der Marke Lucido⁷ ist INVIA ein Hersteller von innovativen, qualitativ hochwertigen LED Stirn- und Handlampen. Diese sind für den Outdoor- und industriellen Einsatz konzipiert. Von insgesamt 13 Mitarbeitern sind zwei im Bereich Forschung und Entwicklung tätig.

da in diesem Bereich derzeit auf Standardbatterien zurückgegriffen wird und aus den Medien das Technologiefeld der Brennstoffzelle als potenziell relevant aufgegriffen wurde. In diesem Technologiefeld existierte jedoch keine Vorkenntnis im eigenen Unternehmen oder bei den Kooperationspartnern.

Problemstellung: »Energieversorgung für die Taschenlampe der Zukunft«

Die LED Stirn- und Handlampen der Marke Lucido von INVIA befinden sich im eher hochpreisigen Bereich und haben den Anspruch, den Kunden Technologien auf dem aktuellsten Stand zu bieten. Durch die Nutzung der Produkte in Extremsituationen im Outdoor- und Industriellen Einsatz müssen diese jedoch auch den höchsten Anforderungen in Bezug auf ihre Verlässlichkeit gerecht werden.

Durch die begrenzten Personalressourcen im Unternehmen ist INVIA bei der Neu- und Weiterentwicklung von Produkten auf die intensive Zusammenarbeit mit externen Partnern angewiesen. Diese spielen sowohl als Zulieferer, aber auch als Kooperationspartner für Entwicklungsaktivitäten eine wichtige Rolle. Bei der Dekomposition des Produktes in verschiedene funktionale Produktkomponenten stellte sich der Themenbereich der Stromquelle als besonders interessant heraus,

⁷ Zum Zeitpunkt des Abschlusses des Projektes nova-net wurde die Marke Lucido von der Firma Mammut übernommen und ist daher nicht mehr Teil von INVIA.

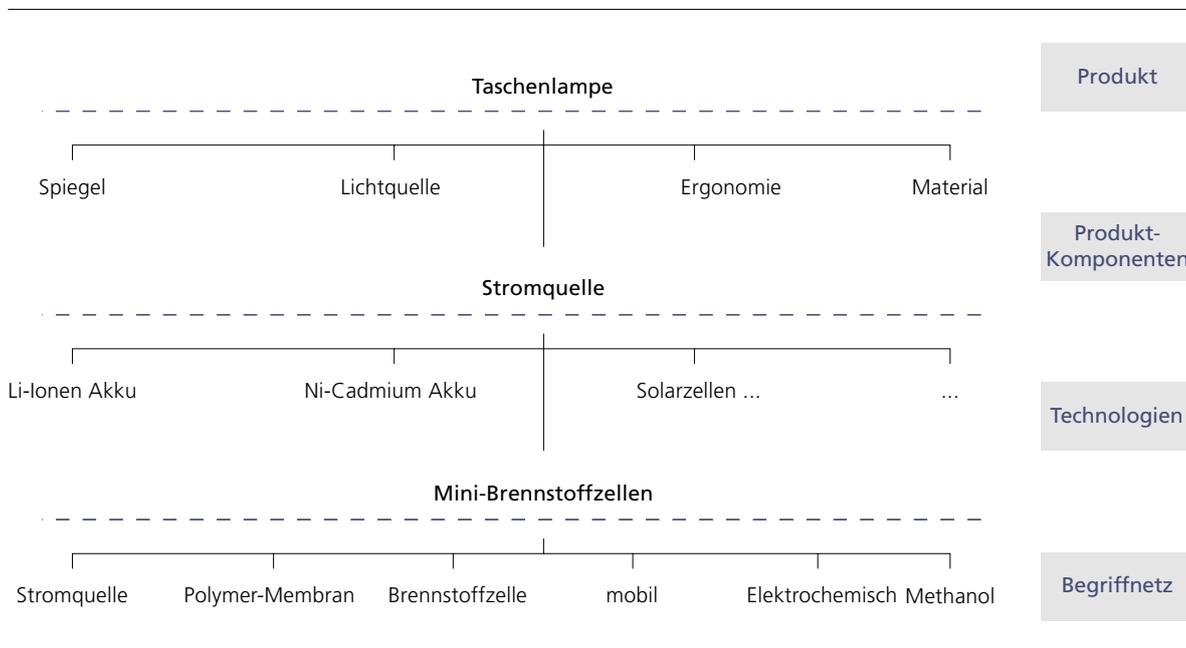


Abbildung 14: Dekomposition des Produktes und Definition eines Begriffsnetzes.

Vorgehensweise: »Von der Spezifikation der Technologie zur Identifikation von Experten«

Nach der Entscheidung, als Suchfeld das Technologiefeld der Brennstoffzelle näher zu untersuchen und es bezüglich des zukünftigen Potenzials für LED-Stirn- und Taschenlampen mittels externer Experten zu bewerten, wurde in einem zweiten Schritt das Technologiefeld der Brennstoffzelle durch ein Begriffsnetz spezifiziert. Durch die Problematik der eingeschränkten Verfügbarkeit von Wasserstoff als Brennstoff wurde der Schwerpunkt auf die Methanolbrennstoffzelle gelegt. Außerdem wurden notwendige Eigenschaften wie beispielsweise die Mobilität und die geringen Abmaße festgelegt (siehe Abbildung 14). Zur Spezifikation des Technologiefeldes konnte über die Enzyklopädie Wikipedia ein guter

Überblick über die Brennstoffzellentechnologie gewonnen werden.

Im nächsten Schritt wurde mit Hilfe des festgelegten Begriffsnetzes über verschiedene Suchstränge nach Experten gesucht. Hierbei konnten die Mitarbeiter des Fraunhofer Instituts für Solare Energiesysteme als Kompetenzträger in diesem Bereich identifiziert werden. Nach einem telefonischen Erstkontakt wurde in diesem Rahmen ein 1-Tages-Workshop durchgeführt, um die weiteren Möglichkeiten zur Anwendung der Brennstoffzellentechnologie im Bereich von LED Stirn- und Taschenlampen zu bewerten. Durch diese Maßnahme konnte die Relevanz, der Zeithorizont sowie die wichtigsten Treiber und Hemmnisse für das Technologiefeld zur Unterstützung von weiteren Entscheidungen weitgehend geklärt werden.

Fazit: »Ergebnisse und Lessons Learned«

Durch die Identifikation und weitere Spezifikation des Suchfeldes wurde anhand des Beispiels der Brennstoffzellentechnologie mit vergleichsweise geringem Ressourceneinsatz die Möglichkeit geschaffen, einen Einblick in die Technologie zu bekommen sowie das zukünftige Potenzial im Unternehmenskontext mit Hilfe von externen Experten zu bewerten. Hierbei spielten für INVIA vor allem der Zeithorizont als auch die Treiber und Hemmnisse der Brennstoffzellentechnologie eine maßgebliche Rolle. Die Sorge, dass externe Partner nicht an einer Kooperation interessiert sein könnten, erwies sich als eher unbegründet. Bei der Einbeziehung von Firmen als externen Experten stellen oft kommerzielle Interessen die Motivation für eine Zusammenarbeit dar, bei Forschungsinstituten eher der praktische Einsatz von Forschungsergebnissen und die Zusammenarbeit mit Unternehmen.

7 Literatur

Chakrabarti, S.; van der Berg, M.; Dom, B. (1999): Focused crawling: a new approach to topic-specific web resource discovery. In: Proceedings of the 8th International World-Wide Web Conference.

Ferber, R. (2003): Information Retrieval. Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web. dpunkt-Verlag.

Posner, M.I.: »What is it to be an expert?«, 1988, In Chi, M.T.H., R. Glaser & M.J. Farr (eds.): »The Nature of Expertise«. Hillsdale, N.J.: Lawrence Erlbaum.

Specht, D.; Behrens, S.; Kirchhof, R. (1999): Komplexität beim strategischen Technologiemanagement, In: ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb, Nr. 12, S. 720-724.

Ulrich, K.T.; Eppinger, S.D. (2004): Product Design and Development. Mc Graw Hill, Boston.