2010/221

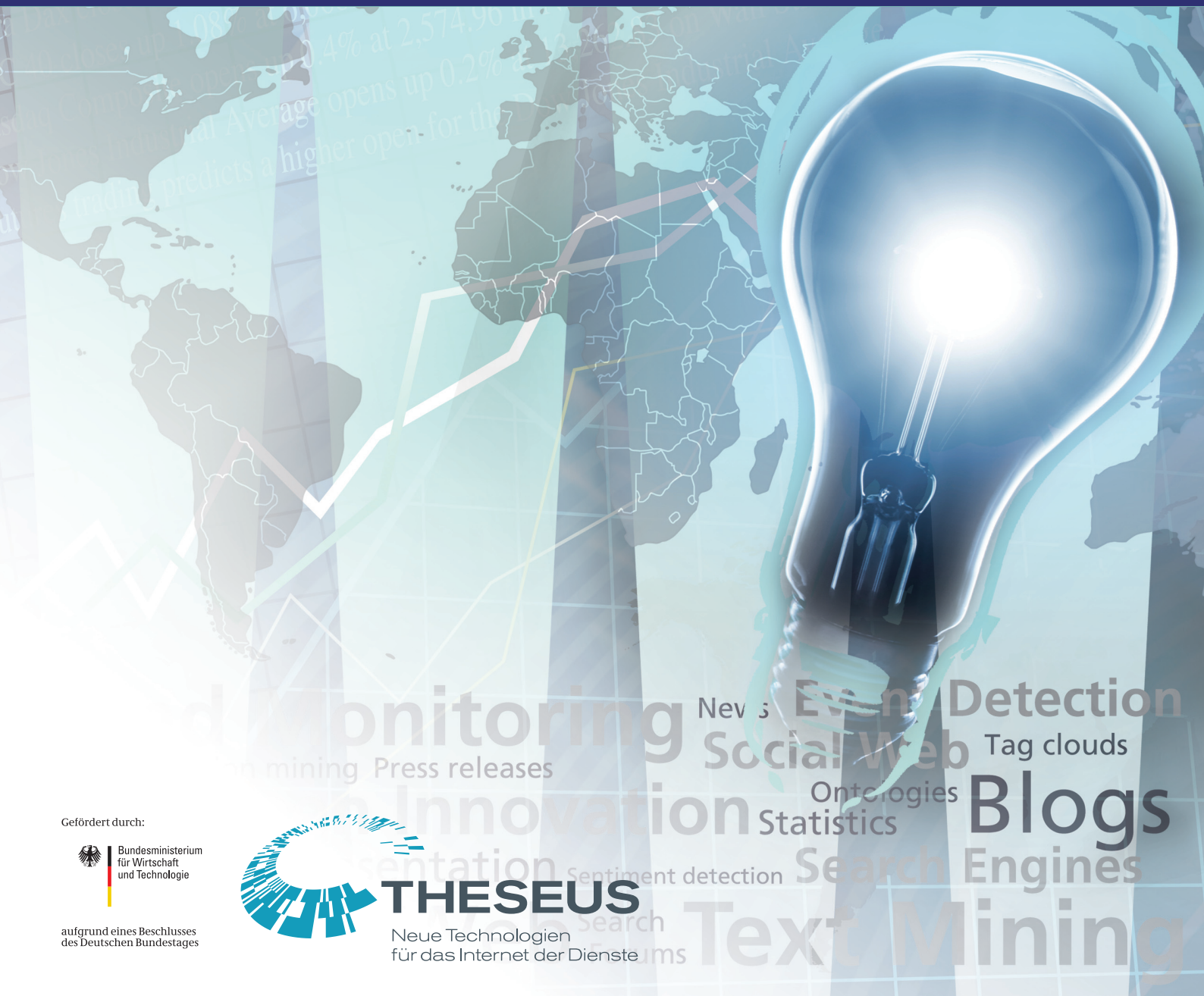
Fraunhofer IAO

FRAUNHOFER-INSTITUT FÜR ARBEITSWIRTSCHAFT UND ORGANISATION IAO

Jan Finzen, Harriet Kasper, Maximilien Kintz

INNOVATION MINING

EFFEKTIVE RECHERCHE UNTERNEHMENSSTRATEGISCH RELEVANTER INFORMATIONEN IM INTERNET



Gefördert durch:

Bundesministerium für Wirtschaft und Technologie

aufgrund eines Beschlusses des Deutschen Bundestages

THESEUS
Neue Technologien für das Internet der Dienste

Gefördert durch:

Bundesministerium
für Wirtschaft
und Technologie

aufgrund eines Beschlusses
des Deutschen Bundestages

Jan Finzen
Harriet Kasper
Maximilien Kintz

Innovation Mining

Effektive Recherche unternehmensstrategisch relevanter Informationen im Internet

Fraunhofer
IAO

THESEUS
Neue Technologien
für das Internet der Dienste

Autoren/Mitwirkende
Jan Finzen
Harriet Kasper
Maximilien Kintz
Thomas Krepp
Stefan Kobes
Jan Willems

Kontaktadresse
Fraunhofer-Institut für Arbeitswirtschaft und Organisation
Nobelstraße 12
70569 Stuttgart
Telefon: +49 (0) 711/970-51 20
Telefax: +49 (0) 711/970-51 11
E-Mail: presse@iao.fraunhofer.de
Web-Adresse: http://www.e-business.iao.fraunhofer.de

Hinweis auf das Forschungsprogramm THESEUS
Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Wirtschaft und Technologie unter dem Förderkennzeichen 01MQ07017 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.
ISBN: 978-3-8396-0139-6

Druck und Weiterverarbeitung
IRB Mediendienstleistungen
Fraunhofer-Informationszentrum Raum und Bau IRB, Stuttgart

Verlag und Druck
Fraunhofer Verlag, Fraunhofer-Informationszentrum Raum und Bau IRB
Postfach 800469, 70504 Stuttgart
Nobelstraße 12, 70569 Stuttgart
Telefon: +49 (0) 711/970-25 00
Telefax: +49 (0) 711/970-25 08
E-Mail: verlag@fraunhofer.de
Web-Adresse: http://verlag.fraunhofer.de

Für den Druck des Buches wurde chlor- und säurefreies Papier verwendet.

Copyright Fraunhofer IAO, 2010

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 Verlages 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 Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und deshalb von jedermann benutzt werden dürften. Soweit in diesem Werk direkt oder indirekt auf Gesetze, Vorschriften oder Richtlinien (z. B. DIN, VDI) Bezug genommen oder aus ihnen zitiert worden ist, kann der Verlag keine

Inhalt

Abbildungen **5**

Tabellen **8**

1 Einführung **9**
1.1 Zielgruppe 11
1.2 Aufbau der Arbeit 12
1.3 Projekthintergrund 13

2 Das Internet als Quelle unternehmensstrategischer Informationen **15**
2.1 Status Quo der internetbasierten Akquise von Informationen und deren Nutzung im Innovationsmanagement 17
2.1.1 Zusammensetzung der Befragungsteilnehmer 17
2.1.2 Rechercheaktivität 19
2.1.3 Innovationsarten 20
2.1.4 Informationsquellen 21
2.1.5 Suchwerkzeuge 23
2.1.6 Fazit 25
2.2 Web-basierte Informationen und Informationsquellen und ihrer Nutzungsmöglichkeiten 26
2.2.1 Suchmaschinen 26
2.2.2 Unternehmensnachrichten und Pressemitteilungen 31
2.2.3 Wissenschaftliche Publikationen 37
2.2.4 Patentdaten 47

3 Neue Möglichkeiten und Herausforderungen im Social Web **55**
3.1 Die Nutzer des Social Web 57
3.2 Informationsarten im Social Web 58
3.2.1 Meinungsäußerung 60
3.2.2 Soziale Netzwerke 62
3.2.3 Microblogging 64
3.2.4 Kollaboration 64
3.2.5 Weiterführende Literatur 67
3.3 Open Innovation 68
3.3.1 Web-basierte Ideenportale 69

3.3.2 Anwendungsbeispiel 78
3.3.3 Weiterführende Literatur 79
3.4 Exkurs: Onlinebasiertes Reputationsmanagement 80
3.4.1 Motivation für Unternehmen 80
3.4.2 Vorgehen im Reputationsmanagement 82
3.4.3 IT-Unterstützung für das Reputationsmanagement 84
3.4.4 Fallbeispiele Social Media 86
3.4.5 Weiterführende Literatur 90

4 Automatisierung der Informationsakquise 91
4.1 Web Intelligence- und Web Mining-Prozess 91
4.2 Softwaretechnische Unterstützung 95
4.3 Web Mining-Werkzeuge im Vergleich 104
4.4 Weiterführende Literatur 105

5 Fazit 107

Referenzen 108

Abbildungen

Abbildung 1: Kenntnisstand der Zielgruppe 12

Abbildung 2: Sprachzugehörigkeit der Internetnutzer 16

Abbildung 3: Zusammensetzung der Befragungsteilnehmer nach Branchenzugehörigkeit 18

Abbildung 4: Zusammensetzung der Befragungsteilnehmer nach Abteilungszugehörigkeit im Unternehmen 18

Abbildung 5: Durchschnittlicher wöchentlicher Aufwand für internetbasierte Informationsbeschaffung 19

Abbildung 6: Organisation der Recherche 19

Abbildung 7: Aneignung der Recherchestrategie 20

Abbildung 8: Nutzungsintensität und zukünftige Bedeutung internetbasierter Informationsbeschaffung für verschiedene Innovationsarten 21

Abbildung 9: Beurteilung des Nutzens internetbasierter Informationsquellen 22

Abbildung 10: Bewertung der Wichtigkeit von Suchwerkzeug-Features 23

Abbildung 11: Hauptprobleme beim Einsatz von Suchwerkzeugen 24

Abbildung 12: Wiederkehrende Suchroutinen 24

Abbildung 13: Aufteilung des Suchmaschinenmarktes weltweit und in Deutschland im Mai 2010 27

Abbildung 14: Architektur von Web-Suchmaschinen 27

Abbildung 15: Screenshot des neuen Google-Suchfrontends 30

Abbildung 16: Veröffentlichte Mitteilungen pro Tag auf verschiedenen Pressemitteilungsdiensten 33

Abbildung 17: Screenshots statistischer Auswertungen der Meldungsfrequenz 36

Abbildung 18: Screenshots geographischer Auswertungen der Meldungsfrequenz 36

Abbildung 19: Suchergebnisse Fraunhofer Publica 39

Abbildung 20: Suchergebnisse bei CiteSeerX 40

Abbildung 21: Suchergebnisse bei Google Scholar 41

Abbildung 22: Logo der Open Access Bewegung 42

Abbildung 23: Zitationssuche mit Scopus 45

Abbildung 24: Screenshot Google Patent Search 50

Abbildung 25: White Spot-Analyse-Prozess des Fraunhofer IAO. 53

Abbildung 26: Nutzung von Produktbewertungen anderer Nutzer 55

Abbildung 27: Auswirkungen von Nutzermeinungen 56

Abbildung 28: Social Media Prisma 58

Abbildung 29: Anwendungsbereiche von Blogs in Unternehmen. 62

Abbildung 30: Soziale Netzwerke in Deutschland 63

Abbildung 31: Open Innovation-Paradigma 68

Abbildung 32: Neugründungen von Ideenportalen (in der untersuchten Stichprobe) 70

Abbildung 33: Screenshot prominenter Ideenportale 71

Abbildung 34: Vertrauen in die verschiedenen Werbemöglichkeiten 81

Abbildung 35: Verantwortung für das Thema Social Media 81

Abbildung 36: Filtern nach Tonalität 86

Abbildung 37: Social Media Fall: Greenpeace gegen KitKat 88

Abbildung 38: Begriffsabgrenzung »Web Searching«, »Web Mining« und »Web Intelligence« 92

Abbildung 39: Typischer Web Intelligence-Prozess 93

Abbildung 40: Typische Visualisierungen von Klassifikationen 100

Abbildung 41: Typische Darstellung einer Klassifikation anhand einer Themenlandkarte 100

Abbildung 42: Beispiele von Tag Clouds. 101

Abbildung 43: Assoziationsgraphen. Firmenbeziehungen auf Basis von Pressemitteilungen 102

Abbildung 44: Visualisierung von Häufigkeitsvergleichen. 102

Abbildung 45: Visualisierung geographischer Informationen 103

Tabellen

Tabelle 1: Suchmaschinenbenutzergruppen 11

Tabelle 2: Nutzung internetbasierter Informationsquellen in verschiedenen Innovationsphasen 22

Tabelle 3: Bewertung verschiedener Pressemitteilungs-Verteilerdiensten 34

Tabelle 4: Suche nach wissenschaftlichen Veröffentlichungen im Internet 43

Tabelle 5: Beispiele üblicher Patentanalysen 47

Tabelle 6: Nutzerrollen im Social Web 57

Tabelle 7: Informationsarten im Social Web 59

Tabelle 8: Firmeneigene Ideenportale 73

Tabelle 9: »Out-of-the-box«-Lösungen für Ideenportale 74

Tabelle 10: Subjektive Bewertung einiger Ideenportale 76

Tabelle 11: Mediatoren-betriebene Ideenportale 77

Tabelle 12: Social Media Monitoring-Tools 85

Tabelle 13: Social Media-Fälle 89

Tabelle 14: Linguistische Verarbeitungsschritte 98

1 Einführung

Zwei von Fraunhofer IAO in den Jahren 2006 und 2009 durchgeführte Unternehmensbefragungen unter Innovationsakteuren und –experten kamen zu dem Ergebnis, dass das Internet sich zur wichtigsten Informationsquelle für Innovationsmanager entwickelt hat (Springer, 2006 sowie Finzen et al., 2009).

Schwerpunkte der Internetrecherche liegen neben der Sammlung aktueller technischer Informationen zum Geschäftsfeld des jeweiligen Unternehmens auf der Informationssammlung über Kunden und Anwender, Wettbewerber und Zulieferer und somit im Bereich der Marktbeobachtung und -forschung. Klassische Marktforschung, etwa durch Studien und Kundenbefragungen, ist das typische Mittel, die Kundenwünsche in der eigenen Entwicklung besser berücksichtigen zu können – das Internet und insbesondere seine derzeitige Entwicklung in den Bereichen der semantischen Inhaltserschließung (»Semantic Web«) einerseits und der Userzentrierung (als Hauptmerkmal der »Web 2.0«-Bewegung) erschließen aber viel weitergehende und mitunter weitaus mächtigere Möglichkeiten.

Kunden sind heutzutage oftmals stark vernetzt – in Diskussionsforen und anderen Communities tauschen sie sich über Erfahrungen mit Produkten und Dienstleistungen aus, die Kundenbewertung ist zu einem unverzichtbaren Bestandteil jedes Webshops geworden, Testportale erfreuen sich zunehmender Beliebtheit. Darüber hinaus gibt es Portale und Communities, an denen sich die Mitglieder der Zielgruppen von Unternehmen über neue Ideen und Erfahrungen austauschen - Informationen, die Impulse für erfolgreiche neue Produkt- oder Serviceideen liefern können.

Innovative Ideen können das Ergebnis formaler oder unstrukturierter Suchprozesse sein und aus verschiedensten Quellen stammen. Neben unternehmensinternen Quellen, wie der Forschungs- und Entwicklungs-, Marketing- und Vertriebsabteilung spielen externe Quellen wie Kunden, Universitäten und Forschungseinrichtungen, Berater, Wettbewerber, Investoren usw. eine wichtige Rolle. Das Internet bietet umfassende neue Möglichkeiten der Erschließung solcher externer Informationsquellen. Für das innovierende Unternehmen ist dieser Erschließungsprozess jedoch mitunter recht mühsam. Einerseits fehlt oftmals das Wissen, wo die relevanten Informationen zu finden sind, andererseits haben klassische Suchmaschinen eine viel zu niedrige Präzision, um die wirklich relevanten Informationen aus der Menge der Seiten herauszufiltern. Dennoch sind sie nach wie vor das am häufigsten eingesetzte Recherche-Werkzeug. Dabei stehen neben einer Fülle von fachlich

spezialisierten Suchmaschinen heutzutage auch äußerst mächtige Web-Intelligence-Werkzeuge zur Verfügung, die Unternehmen gezielt in ihren Aktivitäten der Akquisition innovationsrelevanter Informationen unterstützen.

Die Möglichkeiten dieses von uns als »Innovation Mining« bezeichneten Prozesses sind vielfältig und reichen von der Unterstützung des Suchenden durch Domänenwissen über die automatische Identifikation von Experten zu einem Themengebiet bis zum computerunterstützten Aufspüren neuer Trends, Technologien und innovativer Ideen. Die vorliegende Arbeit liefert einen Überblick über den aktuellen Stand der Technik und Forschung im Bereich Innovation Mining.

Aufgabe des strategischen Innovationsmanagements ist die »Planung, Organisation, Durchführung und Kontrolle von allen Innovationsaktivitäten, die auf das Erzielen und Bewahren von Wettbewerbsvorsprüngen ausgerichtet sind« (Goos und Hagenhoff, 2003). Innovation Mining betrachtet daher unter anderem die folgenden Anwendungsfälle internetbasierter Informationsakquise:

- Erkundung neuer Technologiefelder, z. B. durch gezielte Analyse von Patentschriften und wissenschaftlichen Veröffentlichungen
- Analyse technologischer Trends und frühzeitiges Erkennen wichtiger Ereignisse, etwa durch Überwachung einschlägiger Online-Medien
- Identifikation von Experten, Early Adoptors und Lead-Usern, z. B. durch Analyse von Diskussionsstrukturen im Social Web
- Exploration von Netzwerkstrukturen, etwa von Anwender-Communities
- Identifikation geographischer Kompetenz-Cluster, z. B. durch Analyse von Pressemitteilungen
- Überwachung von Konkurrenzaktivitäten, etwa Kampagnen, Akquisitionsplanungen etc. mittels Web Monitoring
- Frühzeitige Erkennung von Ereignissen und Diskussionen, die dem eigenen Image schaden könnten
- Erkennung von neuen Ideen und Auffinden von Verbesserungsvorschlägen für bestehende Produkte und Dienstleistungen, z. B. durch Observation offener Ideenportale

1.1 Zielgruppe

Im Kontext der vorliegenden Arbeit verstehen wir den strategischen Innovationsmanager als einen Spezialisten, der für die marktgerechte Produkt- und Dienstleistungsentwicklung seines Unternehmens verantwortlich ist und dafür insbesondere die Aktivitäten von Kunden, Partnern und Wettbewerbern überblickt. Diese Aufgabe hat Implikationen auf sein Informationsbedürfnis.

Stock und Lewandowski (2006) unterteilen die Benutzer internetbasierter Suchmaschinen in die in Tabelle 1 dargestellten Gruppen.

Tabelle 1: Suchmaschinen-benutzergruppen

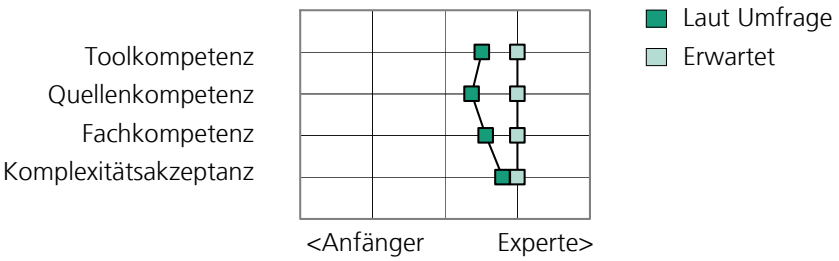
(Stock und Lewandowski, 2006)

	Beschreibung	Typische Informationsquellen
Information Professional	Experte für Recherche; professioneller Wissenshintergrund auf dem Gebiet der Informationswissenschaft; kennt relevante Datenbanken; entwickelt eigene Suchstrategien; recherchiert oft auch im Auftrag Dritter; komplexe Suchanfragen	- Web-Intelligence-Werkzeuge - Deep-Web-Datenquellen wie Patentdatenbanken oder wissenschaftliche Suchmaschinen
Professioneller Endnutzer	Fachexperte; recherchiert aufgrund fachlichen Informationsbedarfs; entwickelt kaum Recherchestrategien; Informationsbedürfnis eher ad-hoc	- Allgemeine Suchmaschinen - Einschlägige Fachdatenbanken
Laiennutzer	Recherchiert aus privaten und gelegentlich aus beruflichen Gründen; keine systematische Recherchestrategie; eher kurze Suchanfragen	- Allgemeine Suchmaschinen

Wir ordnen strategische Innovationsmanager zwischen der ersten und zweiten Gruppe ein: Sie sind Experten bzgl. der Fachdomäne(n) ihrer Firma (»Fachkompetenz«), haben aber aufgrund ihres Aufgabenprofils einen erweiterten Blickwinkel. Darum verfügen sie über ein breites Wissen darüber, wo und wie Informationen zu finden sind, die ihrem Informationsbedürfnis entsprechen (»Quellenkompetenz«). Sie recherchieren kontinuierlich und auf strategische Weise und sind im effektiven Umgang mit Suchwerkzeugen geübt (»Werkzeugkompetenz«). Um Ihre Rechercheergebnisse zu verbessern, sind sie potenziell bereit, eine höhere Werkzeugkomplexität in Kauf zu nehmen. Abbildung 1 fasst unsere Einordnung strategischer Innovationsmanager bzgl. der o. g. Aspekte zusammen. Während die helleren Markierungen unser Idealbild eines strategischen Innovationsmanagers beschreiben, zeigen die dunkleren Markierungen die Selbsteinschätzung der von uns befragten Personen[1].

[1] Weitere Erkenntnisse aus der Benutzerbefragung werden in Abschnitt 2.1 präsentiert.

Abbildung 1:
Kenntnisstand der
Zielgruppe



Die Selbsteinschätzung bzgl. Werkzeug- und Quellenkompetenz lässt Raum für Verbesserungen erkennen, und die vorliegende Arbeit soll dazu einen Beitrag leisten.

- Die für das strategische Innovationsmanagement besonders relevanten Arten von Informationsquellen werden benannt und Möglichkeiten der effizienten Nutzung vorgestellt.

- Werkzeuge und Methoden für professionelle Recherche und Analyse unternehmensstrategisch relevanter Informationen aus Web-Quellen werden untersucht.

1.2 Aufbau der Arbeit

Die Arbeit gliedert sich in drei Hauptkapitel:

- Kapitel 2 stellt zunächst das Internet als Quelle unternehmensstrategischer Informationen vor und untersucht anhand einer Unternehmensbefragung unter Innovationsmanagern das derzeitige Rechercheverhalten und die damit verbundenen Herausforderungen und Probleme. Anschließend werden »klassische« innovationsrelevante Informationsquellen wie Patentdatenbanken, Online-Verteiler von Pressemitteilungen und Suchmaschinen für wissenschaftliche Veröffentlichungen vorgestellt.

- Kapitel 3 widmet sich den Möglichkeiten und Herausforderungen des Social Webs. Dabei liegt der Fokus wiederum auf der Nutzung der Informationsquellen für die Akquisition unternehmensstrategisch relevanter Informationen. Ein wichtiges Kennzeichen der Web 2.0-Bewegung ist die Benutzerinteraktion – entsprechend geeignet sind die damit verbundenen Technologien zur Unterstützung offener Innovationsprozesse, die den Kunden als kreativen Partner in frühe Phasen der Erstellung von Produkten und Dienstleistungen einbinden. Ein Unterkapitel widmet sich darum den offenen Ideenportalen, die neben der Erschließung externen Wissens auch ein zunehmend

wichtiges Instrument der Kundenbindung darstellen, und somit auch gezielt im onlinebasierten Reputationsmanagement eingesetzt werden können – einem Thema, dem das Kapitel einen Exkurs widmet.

- Das abschließende Kapitel 4 beschäftigt sich mit weiterreichenden Möglichkeiten der Automatisierung der Informationsakquisition durch Web Intelligence-Werkzeuge. Die Möglichkeiten und Herausforderungen solcher Werkzeuge werden diskutiert und Aspekte, die bei der Auswahl entsprechender Tools berücksichtigt werden sollten, vorgestellt. Den Abschluss des Kapitels bildet eine Übersicht über derzeit am Markt verfügbare Software-Lösungen.

Um der Intention der Nutzbarkeit als Leitfaden gerecht zu werden, sind die einzelnen Abschnitte so organisiert, dass sie unabhängig voneinander gelesen werden können (wo nötig, sind entsprechende Querverweise angegeben). Zu jedem Abschnitt existiert ein Hinweis auf weiterführende Literatur.

Das vorliegende Dokument kann kostenfrei in elektronischer Form über die Webseite http://innovation-mining.net heruntergeladen werden. Dort finden sich zudem viele weitere Informationen zu den behandelten Themengebieten und insbesondere viele Querverweise zu Lösungsanbietern und Produkten.

1.3 Projekthintergrund

Die vorliegende Arbeit ist im Rahmen des TEXO-Projektes des THESEUS-Programms entstanden.

THESEUS ist ein vom Bundesministerium für Wirtschaft und Technologie (BMWi) initiiertes Forschungsprogramm mit dem Ziel, neue Technologien und Methoden für das »Internet der Dienste« zu entwickeln. Unter dem Dach von THESEUS arbeiten über 60 Forschungspartner aus Wissenschaft und Wirtschaft zusammen. Die entwickelten Technologien und Methoden werden den Zugang zu Informationen vereinfachen, Daten zu neuem Wissen vernetzen und die Grundlage für die Entwicklung neuer Dienstleistungen im Internet schaffen. Die zentrale Webseite des Forschungsprogramms ist unter http://www.theseus-programm.de zu erreichen.

Das Projekt TEXO erforscht im Rahmen des Forschungsprogramms THESEUS Dienstleistungen, die über das Internet angeboten und abgerufen werden. Betrachtet wird dabei der gesamte Lebenszyklus dieser Dienstleistungen, von der marktgerechten Gestaltung und softwaretechnischen Entwicklung der Dienste über deren Angebot und Vermittlung in Business Webs bis zu ihrer Nutzung und Weiterentwicklung. Neben der Automatisierung durch technische Dienste und der Entwicklung intuitiver Schnittstellen wird auch das Innovationspotenzial durch Feedback aus Kundenkreisen und Internet-

Communities untersucht (vgl. Stathel et al., 2008; Riedl et al., 2009). Fraunhofer IAO betrachtet in diesem Kontext die Frage, wie innovationsrelevante Informationsquellen möglichst effektiv im Rahmen eines durchgängigen Service-Innovationsprozesses eingebunden werden können. Zur Evaluation fortgeschrittener Konzepte der internetbasierten Informationsakquise wurde unter anderem das »Innovation Mining Cockpit« entwickelt – das in verschiedenen Abschnitten der vorliegenden Arbeit Erwähnung finden wird.

2 Das Internet als Quelle unternehmensstrategischer Informationen

Die Suche nach Informationen im World Wide Web unterscheidet sich in vielerlei Hinsicht vom »klassischen« Information Retrieval. Dies liegt in der Natur des Internets, das ein völlig neues Zusammenspiel von Informationsquellen ermöglicht und sich in seinem Wesen grundsätzlich von dem herkömmlicher Informationsquellen unterscheidet: Manning, Raghavan und Schütze fassen zusammen: »Das Web (…) ist beispiellos in seiner Größe, beispiellos im beinahe vollständigen Fehlen eines koordinierten Entstehungsprozesses und beispiellos in der Diversität der Hintergründe und Motivationen der teilnehmenden Benutzer« (Manning et al. 2007). Einige wichtige Daten zur aktuellen Ausprägung des Internets und World Wide Webs werden in den folgenden Abschnitten dargestellt.

Den aktuellen Statistiken des Web-Analyse-Dienstes Netcraft[2] zufolge gibt es derzeit im Web ca. 200 Millionen registrierte Hostnames und ca. 80 Millionen als »aktiv« bewertete Websites. Mit »Sites« sind nicht einzelne HTML-Seiten gemeint, sondern durch eine IP-Adresse identifizierbare Internetlokationen, die eine HTTP-Anfrage mit Antwort-Code 200 und einer Webseite beantworten. Eine Website besteht aus allen zu dieser IP-Adresse gehörenden Webseiten. Die Anzahl aller Webseiten lässt sich aufgrund der starken Größenunterschiede zwischen einzelnen Websites und insbesondere der Natur dynamischer Webseiten nur sehr schwer schätzen. Nimmt man einen von der Firma Boutell geschätzten Faktor von 273 Seiten pro Site[3] an, bestünde das Web aus beinahe 35 Milliarden Webseiten.

Nach Angaben von internetworldstats.com[4] nutzen heute knapp 1,8 Milliarden Menschen (etwa 27 Prozent der Weltbevölkerung) das Internet, dabei ist die Anzahl in den vergangenen sieben Jahren um 400 Prozent gestiegen. In Europa nutzen ca. 426 Millionen Menschen das Netz, davon gut 62 Millionen in Deutschland. Demnach nutzen in Deutschland inzwischen über 75 Prozent der Einwohner das Internet.

[2] http://news.netcraft.com

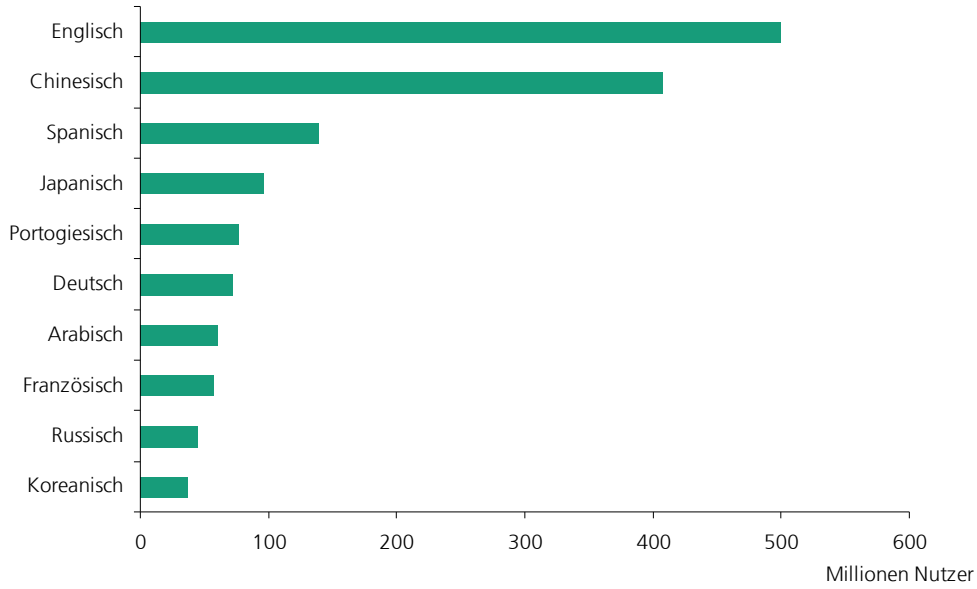
[3] Dieser Faktor entstand aus der Division der im August 2005 von Yahoo indizierten Webseiten (19,2 Milliarden) durch die damalige von Netcraft angegebene Menge der Websites (70,4 Millionen). Diese Zahlen sind mit Vorsicht zu betrachten, da sie zum einen aus unterschiedlichen Quellen stammen und zudem nicht geklärt ist, ob der für 2005 ermittelte Faktor auch heute noch gültig ist.

[4] http://www.internetworldstats.com

Die im Internet vorherrschende Sprache ist Englisch - fast ein Drittel aller Internetbenutzer spricht es, gefolgt von Chinesisch, Spanisch und Japanisch. Englisch dominiert eindeutig als »Sprache des Webs« – allerdings ist ein sehr starker Zuwachs an chinesischsprachigen Webseiten (es wird ein Anstieg von 2000 auf 2009 um fast 1.200 Prozent angenommen[5]) und chinesischsprachigen Internetbenutzern zu verzeichnen (vgl. Abbildung 2). Gleichzeitig steigen auch die Zahlen arabisch-, portugiesisch-, spanisch- und französischsprachiger Internetnutzer.

Abbildung 2: Sprachzugehörigkeit der Internetnutzer

Darstellung basierend auf Angaben von Internet World Stats (http://www.internetworldstats.com/stats7.htm)



Das Internet ist geprägt durch eine immens hohe Volatilität: Nirgendwo verbreiten sich Neuigkeiten schneller, sind Nachrichten aktueller als im Internet. Gleichzeitig findet sich auch eine immense Menge an veralteten Informationen im Internet. Nicht aktualisierte oder schlichtweg vergessene Webseiten machen einen beträchtlichen Bestandteil des Gesamtinfomationsaufkommens aus. Andererseits wird das Web zunehmend dynamisiert – die Verbreitung von Content-Management-Systemen, welche die Generierung aktueller Webseiten »on the fly« erlauben, führt dazu, dass weniger »statische« Webseiten abgelegt und nicht mehr gepflegt werden.

[5] Noch größer ist allerdings der Zuwachs der russisch- und arabischsprechenden Internetnutzergruppe mit 1.400 Prozent bzw. mit 2.300 Prozent). In Deutschland betrug der Zuwachs ca. 160 Prozent.

Längst bestehen Internet und World Wide Web nicht mehr nur aus verknüpften HTML-Seiten. Zu den verbreiteten *Datenformaten* gehören inzwischen neben den gängigen Office-Formaten (insbesondere der Microsoft-Produkte) das Portable Document Format (PDF) und diverse Multimedia-Formate: Bilder (gif, jpeg usw.), Video- und Tonformate (mp3, mp4, divx usw.) und insbesondere Streaming-Media-Formate (beispielsweise im RealPlayer oder Windows Media Player).

2.1 Status quo der internetbasierten Akquise von Informationen und deren Nutzung im Innovationsmanagement

Fraunhofer IAO führte 2009 im Rahmen der Projektarbeiten des Theseus Forschungsprogramms eine Befragung zum Einsatz internetbasierter Suchtechniken unter 1.000 »Innovation Professionals« deutscher Unternehmen durch. Der Rücklauf von 142 Fragebogen (schriftlich und online) wird als Indiz für die Aktualität und die wahrgenommene Wichtigkeit der Thematik gewertet. Im Rahmen der Befragung wurden die bei der Informationsbeschaffung im Innovationsmanagement eingesetzten internetbasierten Informationsquellen untersucht. Zudem lag ein Fokus der Befragung auf der Analyse der Zufriedenheit mit der realisierbaren Ergebnisqualität sowie den aktuellen Problemen und Grenzen internetbasierter Informationssuche.

Die Ergebnisse der Befragung sind bedingt repräsentativ für die gesamte deutsche Industrie - die Auswahl der Teilnehmer erfolgte nicht vollständig zufällig: Die Adressaten wurden teilweise gezielt selektiert, um einen möglichst hohen Anteil an »Innovation Professionals« zu erreichen: 450 Adressen stammten aus einer entsprechenden Industriekontaktliste des Instituts. 550 weitere wurden aus einer kommerziellen Adressdatenbank extrahiert. Bei letzteren wurden nur solche Adressaten ausgewählt, deren Positionsfreitext-Beschreibung auf eine innovationsbezogene Tätigkeit hinwies bzw. deren Unternehmensnamen den Begriff »innovativ« bzw. »Innovation« enthielt.

2.1.1 Zusammensetzung der Befragungsteilnehmer

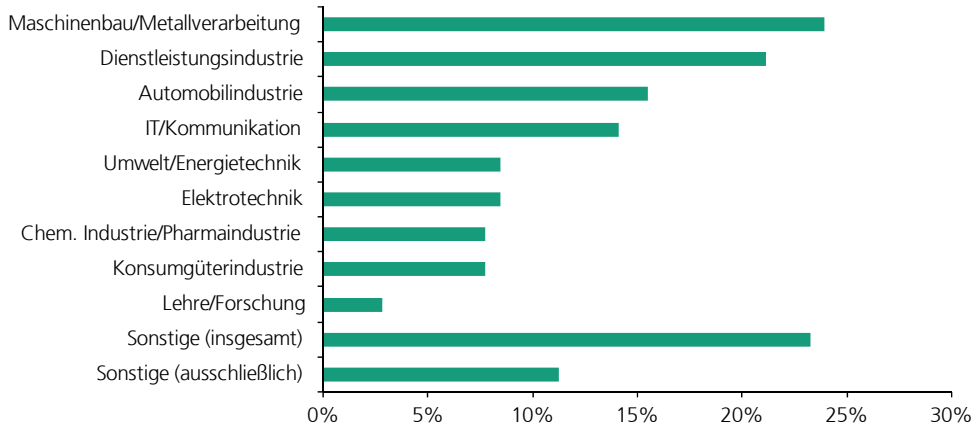
Es lagen insgesamt 142 Fragebogen für die Auswertung vor. Die Befragungsteilnehmer sind am häufigsten in den Branchen Maschinenbau/ Metallverarbeitung (24 Prozent) und der Dienstleistungsindustrie (21 Prozent) vertreten. Dann folgen Automobilindustrie (15 Prozent), IT/Kommunikation (14 Prozent) und mit jeweils 8 Prozent Umwelt/Energietechnik, Elektrotechnik und Pharma- und Konsumgüterindustrie (Abbildung 3).

Hinsichtlich der Verteilung nach Unternehmensgröße können 15 Prozent als Kleinstunternehmen (bis 9 Beschäftigte), 43 Prozent als kleine und mittelgroße

Unternehmen (unter 250 Beschäftigte) und 42 Prozent als Großunternehmen (über 250 Mitarbeiter) klassifiziert werden.

Abbildung 3: Zusammensetzung der Befragungs-teilnehmer nach Branchen-zugehörigkeit

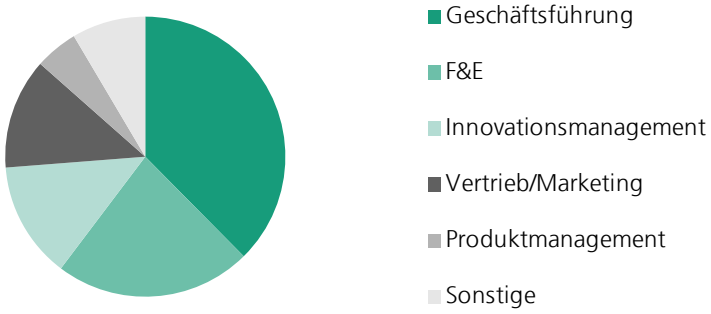
(n=142)



37 Prozent der Teilnehmer gehören der Geschäftsführung an, 23 Prozent arbeiten in der Forschung und Entwicklung und jeweils 13 Prozent in den Abteilungen Innovationsmanagement und Marketing/Vertrieb (Abbildung 4).

Abbildung 4: Zusammensetzung der Befragungs-teilnehmer nach Abteilungs-zugehörigkeit im Unternehmen

(n=142)



Um die Innovationsintensität der Unternehmen zu untersuchen, wurde der Anteil der Investition in Forschung und Entwicklung am Gesamtumsatz erfragt. Hier war auffällig, dass über 20 Prozent der Befragten mehr als 10 Prozent des Umsatzes in Forschung und Entwicklung investieren. Hinsichtlich der Unternehmensgröße waren fast ausschließlich (87 Prozent) die Unternehmen mit unter 250 Mitarbeitern in dieser Kategorie vertreten. Die Unternehmen, in denen die Befragten tätig sind, haben innerhalb der letzten drei Jahre ihre Produktpalette im Rahmen von Produktinnovationen zu 83 Prozent durch

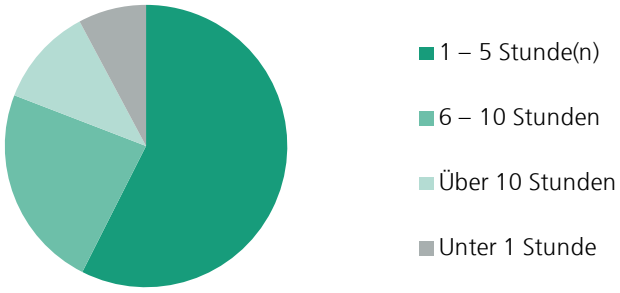
Produktverbesserungen, zu 60 Prozent durch Marktneuheiten und zu 56 Prozent durch Sortimentsneuheiten ausgebaut.

2.1.2 Rechercheaktivität

In den Abteilungen Forschung und Entwicklung, Innovationsmanagement und der Geschäftsführung ist die Mehrheit der Befragten oft an der Entwicklung neuer Ideen beteiligt. Im Gegensatz hierzu ist die Mehrheit der Befragten aus den Abteilungen Produktmanagement und Vertrieb/Marketing nur manchmal an der Ideenentwicklung beteiligt. 58 Prozent der Befragten verbringen mit internetbasierter Informationsbeschaffung wöchentlich 1 bis 5 Stunden, 23 Prozent 6 bis 10 Stunden und 11 Prozent sogar über 10 (vgl. Abbildung 5).

Abbildung 5: Durchschnittlicher wöchentlicher Aufwand für internetbasierte Informations-beschaffung

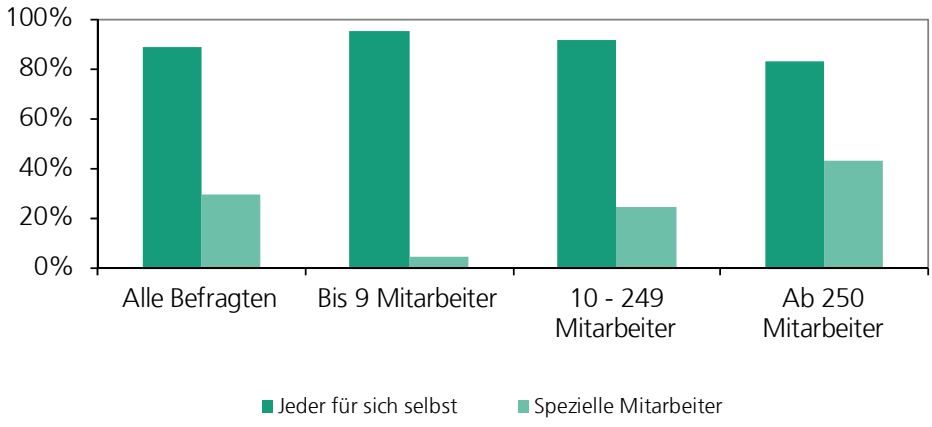
(n=142)



In 30 Prozent der befragten Unternehmen ist ein spezieller Mitarbeiter beschäftigt, der sich um die Informationsbeschaffung kümmert. Wie Abbildung 6 zeigt, steigt mit wachsender Unternehmensgröße auch die Häufigkeit spezieller Recherchemitarbeiter.

Abbildung 6: Organisation der Recherche

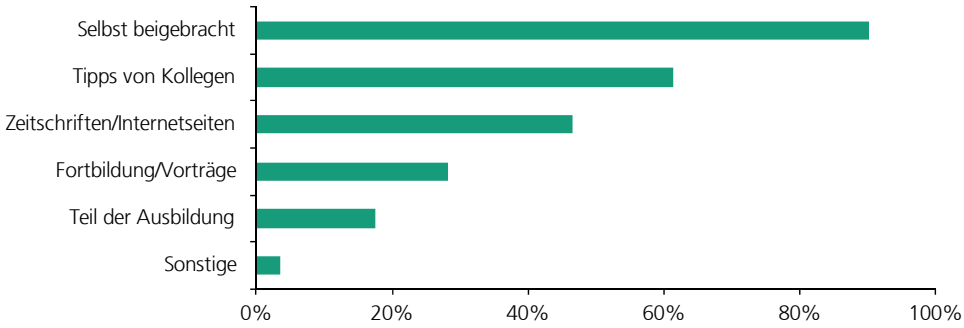
(n=142)



Bei der Aneignung der Recherchestrategie dominiert die Autodidaktik (vgl. Abbildung 7) über sämtliche Branchen und Unternehmensgrößen hinweg. Die Hälfte der Teilnehmer informiert sich in Zeitschriften oder im Internet über Recherchemethoden und –werkzeuge. Lediglich eine Minderheit nahm bisher an Weiterbildungen teil - vorrangig Mitarbeiter größerer Unternehmen - oder konnte sich entsprechendes Wissen während der eigenen Ausbildung aneignen.

Abbildung 7:
Aneignung der Recherchestrategie

(n=142)

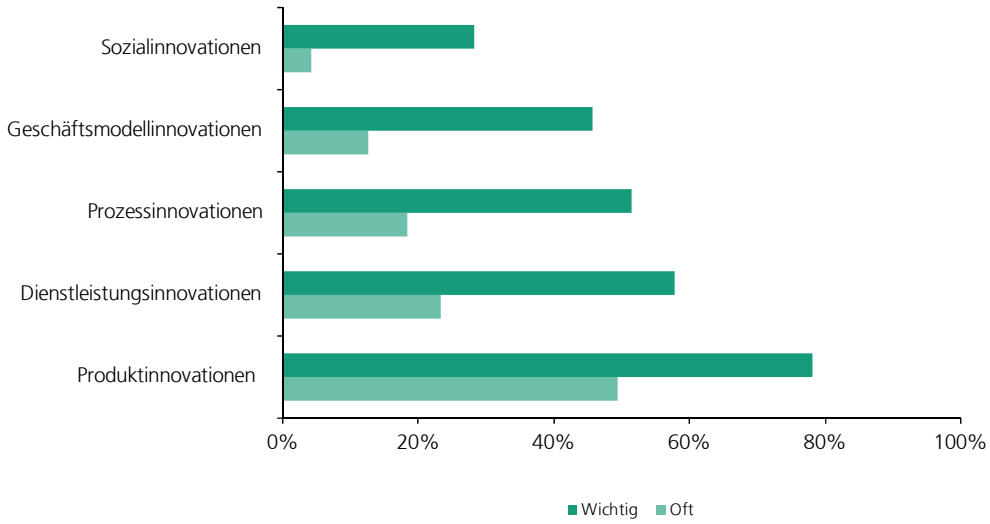


2.1.3 Innovationsarten

Internetbasierte Suchtechniken können zur Informationsbeschaffung für alle Innovationsarten eingesetzt werden. Für Produkt- und Dienstleistungsinnovationen wird internetbasierte Informationsbeschaffung aktuell am häufigsten genutzt, und auch zukünftig wird ihr die größte Bedeutung beigemessen. Für Prozess-, Geschäftsmodell- und Sozialinnovationen werden internetbasierte Informationsquellen derzeit deutlich weniger eingesetzt, doch auch hier wird zukünftige Bedeutung des Internets höher eingeschätzt. Bei der Interpretation von Abbildung 8 ist zu berücksichtigen, dass bei den Angaben zu Geschäftsmodell- und Sozialinnovationen die Non-Response-Raten vergleichsweise hoch waren.

Abbildung 8: Nutzungsintensität und zukünftige Bedeutung internetbasierter Informationsbeschaffung für verschiedene Innovationsarten

(n=142)



2.1.4 Informationsquellen

Im Rahmen der internetbasierten Informationsbeschaffung kommt eine Vielzahl verschiedener Quellen zum Einsatz. Abbildung 9 verdeutlicht die Nutzeneinschätzung der verschiedenen Quellen durch die Befragungsteilnehmer. Der größte Nutzen wird (was wenig überrascht) den Suchmaschinen und Online-Enzyklopädien zugeschrieben. Ebenfalls sehr hoher Nutzen wird den Online-Wirtschafts- und Fachzeitschriften, Forschungs- und. Technologieportalen sowie Patentdatenbanken zugeschrieben. Überraschend niedrig schneiden die typischen Web 2.0-Quellen wie Social Networks, Experten- und Verbraucher-Blogs sowie Open Innovation-Plattformen ab. Dies liegt vermutlich zumindest teilweise an dem noch geringen Verbreitungsgrad dieser recht neuen Art von Informationsquellen – insbesondere auch bezüglich einer Nutzung durch Unternehmen.

In welchen Phasen des Innovationsprozesses internetbasierten Informationsquellen bei den Befragten zum Einsatz kommen, wird in Tabelle 2 verdeutlicht. Dabei waren die Innovationsphasen wie folgt vorgegeben: 1=Innovationsanstoß, 2=Ideensammlung, 3=Ideengenerierung und 4=Ideenbewertung. Suchmaschinen werden in allen Phasen häufig eingesetzt, Forschungs- und Technologieportale sowie Onlineausgaben von Wirtschafts- und Fachzeitschriften kommen besonders in den frühen Phasen Innovationsanstoß und Ideensammlung zum Einsatz. Online-Enzyklopädien und insbesondere Patentdatenbanken werden eher für die nachfolgende Ideenbewertung genutzt.

Abbildung 9:
Beurteilung des Nutzens internetbasierter Informationsquellen

(n=142)

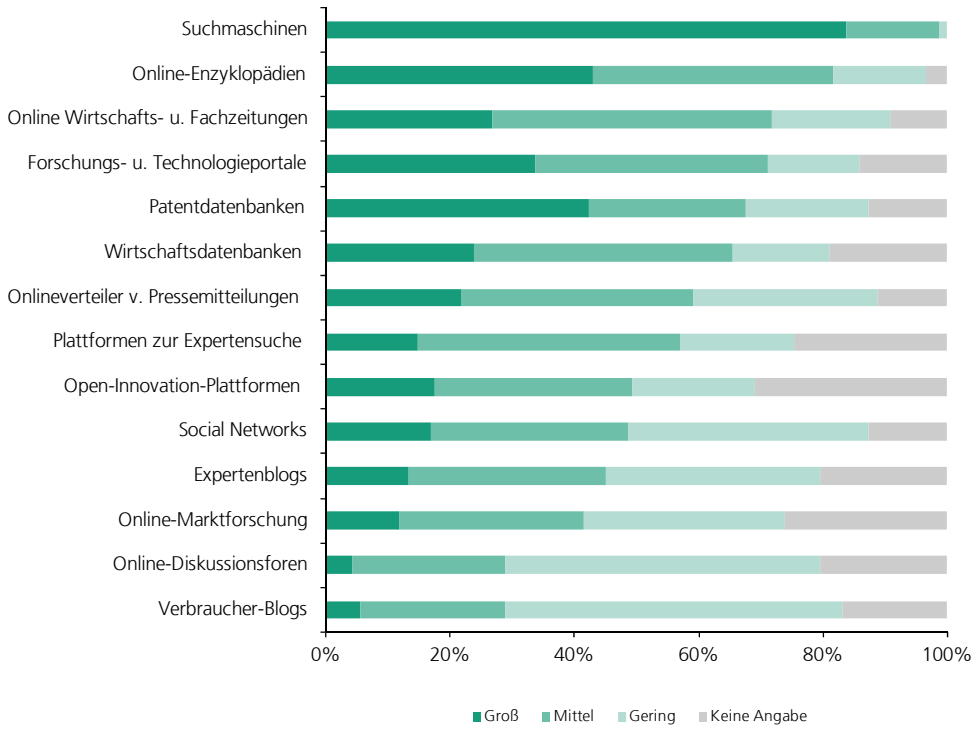


Tabelle 2:
Nutzung internetbasierter Informationsquellen in verschiedenen Innovationsphasen

(n=142)

Lesehilfe: 28 Prozent der Teilnehmer gaben an, dass Online-Verteiler von Pressemitteilungen für sie in der Innovationsphase »Innovationsanstoß« von Bedeutung sind.

	Innovationsprozessphasen				
	1	2	3	4	P_i
Suchmaschinen	36%	70%	35%	37%	16%
Online-Enzyklopädien	13%	28%	16%	32%	8%
Online Wirtschafts- u. Fachzeitungen	32%	35%	15%	18%	9%
Forschungs- u. Technologieportale	39%	32%	23%	18%	10%
Patentdatenbanken	22%	19%	14%	48%	9%
Wirtschaftsdatenbanken	11%	16%	8%	28%	6%
Online-Verteiler v. Pressemitteilungen	28%	25%	10%	9%	7%
Plattformen zur Expertensuche	8%	10%	9%	9%	3%
Open-Innovation-Plattformen	19%	12%	13%	8%	5%
Social Networks	15%	13%	9%	12%	5%
Expertenblogs	14%	14%	10%	18%	5%
Online-Marktforschung	25%	31%	14%	27%	9%
Online-Diskussionsforen	13%	13%	7%	13%	4%
Verbraucher- Blogs	11%	8%	6%	10%	4%
Prozent je Innovationsprozessphase	26%	30%	17%	26%	

Top-Ten Restliche

2.1.5 Suchwerkzeuge

Eine gute Qualität der Suchergebnisse und ein geringer eingesetzter Zeitaufwand bis zum Auffinden der gewünschten Information sind für die Befragten die wichtigsten Merkmale einer effektiven Suchmaschine, wie Abbildung 10 verdeutlicht. Entsprechend sind der hohe Zeitaufwand und nicht zufriedenstellende Ergebnisqualität derzeit die Hauptprobleme bei der Nutzung von Suchwerkzeugen (Abbildung 11). Dies ist sicherlich keine zielgruppenspezifische Einschätzung, sondern eine allgemeingültige Anforderung – jeder Benutzer wünscht sich, möglichst schnell zu guten Ergebnissen zu gelangen. Für das Innovationsmanagement spezifischer sind die in Abbildung 10 mit dem Median 4 versehenen Aspekte: Eine effektive Eingrenzung des Suchraums zur Erhöhung der Ergebnispräzision, transparente Ergebnis-Rankingverfahren, die Möglichkeit, Suchergebnisse nach Aktualität zu ordnen und die Bündelung verschiedener Suchwerkzeuge in einem einzigen Werkzeug.

Abbildung 10: Bewertung der Wichtigkeit von Suchwerkzeug-Features

(n=142)

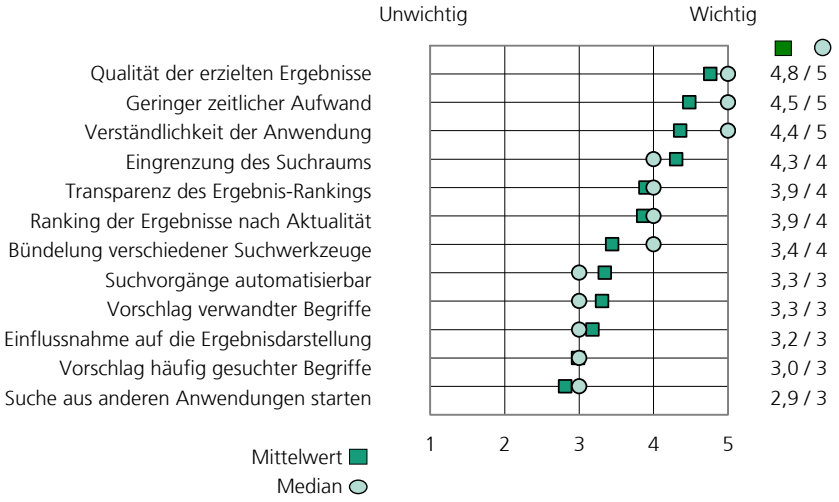
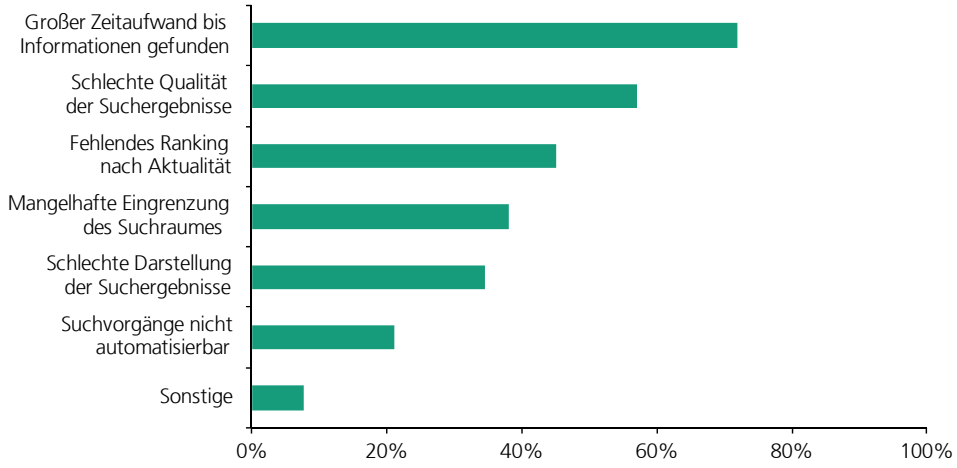


Abbildung 11:
Hauptprobleme beim Einsatz von Suchwerkzeugen

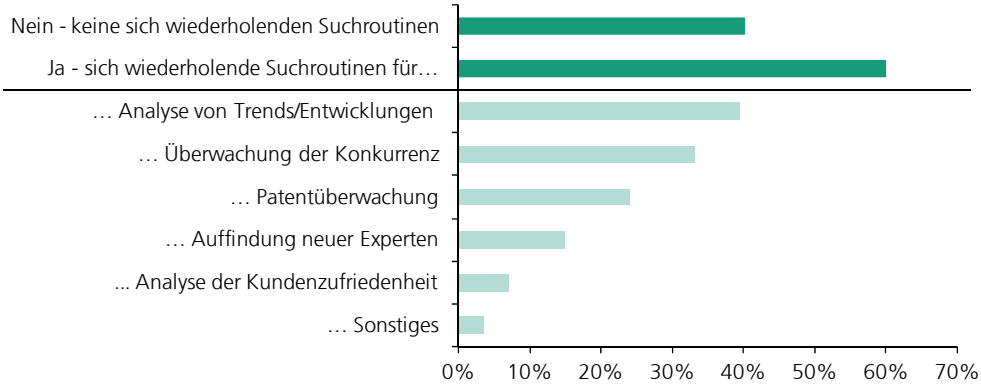
(n=142)



Über 60 Prozent der Befragten durchlaufen wiederkehrende Suchroutinen, hauptsächlich zur Analyse von Trends und Entwicklungen und zur Analyse der Konkurrenzaktivitäten (vgl. Abbildung 12). Push-Versorgung wird bereits von ungefähr der Hälfte der Befragten eingesetzt. Der Großteil der Befragten bevorzugt die Information über neue Treffer mittels E-Mail, 20 Prozent der Befragten würden sich auch per RSS-Feed informieren lassen. Bei der Eingrenzung des Suchraumes werden die Filterkriterien nach Erstellungsdatum und nach Sprache des Suchergebnisses aktuell am häufigsten genutzt und auch am wichtigsten bewertet.

Abbildung 12:
Wiederkehrende Suchroutinen

(n=142)



2.1.6 Fazit

Die hohe Rücklaufquote zur Umfrage verdeutlicht die Wichtigkeit des Themas Websuche im Internet für das Innovationsmanagement. Die Ergebnisse legen den Schluss nahe, dass durchaus eine Nachfrage nach spezialisierten Suchwerkzeugen für die Ideen- und Innovationsentwicklung existiert. Solche Suchwerkzeuge sollten mindestens

- mächtige Möglichkeiten der Suchraumbeschränkung und der Ergebnisfilterung besitzen,
- über transparente und adaptierbare Ranking-Methoden verfügen,
- die insbesondere eine Identifikation besonders aktueller Informationen zulassen und
- dokumentspezifische Eigenschaften (wie verfügbare Metadaten) einsetzen, um sowohl die Ausdrucksmächtigkeit der Suchanfragesprache zu erhöhen, als auch anwendungsfallspezifische Möglichkeiten der Ergebnisdarstellung zu unterstützen, sowie
- langfristige Informationsbedürfnisse unterstützen, etwa durch eine automatisierte Wiederholung der Suchabfrage und eine Benachrichtigung über neue Ergebnisse nach dem Push-Prinzip.

In Kapitel 4 werden entsprechende Konzepte tiefergehend diskutiert und verschiedene Werkzeuge vorgestellt, die zur Umsetzung eingesetzt werden können.

Etwas überraschend mutet die Erkenntnis an, dass typische Web 2.0-Informationsquellen wie Konsumenten-Blogs und Diskussionsforen im Innovationsmanagement (bisher) nicht als herausragend wichtig angesehen werden. Auch die Bedeutung von Open Innovation-Portale wird (bislang) eher gering eingeschätzt. Zumindest bei letzteren kann dies teilweise auf den noch geringen Bekanntheitsgrad zurückgeführt werden (vgl. Abbildung 8 auf Seite 21). Offene Ideenportale können ebenso wie Blogs und Diskussionsforen von Unternehmen als »Listening Platforms«, also als Kommunikationskanal zu Endverbrauchern eingesetzt werden und stellen somit eine prinzipiell für unternehmensstrategische Fragestellungen relevante Informationsquelle dar. In den folgenden Abschnitten werden neben »klassischen« Informationsquellen wie Patentdatenbanken und Suchmaschinen für wissenschaftliche Veröffentlichungen daher auch Blogs, Diskussionsforen und Ideenportale untersucht (Kapitel 3).

2.2 Web-basierte Informationen und Informationsquellen und ihrer Nutzungsmöglichkeiten

Die im vorigen Abschnitt dargestellten Ergebnisse einer Umfrage unter Innovationsmanagern zeigen, dass allgemeine Suchmaschinen den weithin wichtigsten Zugang zu Informationen im Internet darstellen. Es wurde jedoch auch deutlich, dass für den Zugang zu speziellen Informationen, wie etwa Patentdaten oder wissenschaftlichen Veröffentlichungen, spezielle Zugriffswerkzeuge genutzt werden sollten. In den folgenden Abschnitten werden für das Innovationsmanagement wichtige Informationsquellen beschrieben und Zugriffswerkzeuge vorgestellt, die zur Erschließung dieser Quellen genutzt werden können.

2.2.1 Suchmaschinen

Suchmaschinen gehören heute sicherlich zu den gebräuchlichsten Werkzeugen im Internet. Die eingangs skizzierte Größe des World Wide Webs hat sie zu unverzichtbaren Ausgangspunkten für Recherchen aller Art gemacht. Auch für das Innovationsmanagement stellen Sie einen wichtigen Zugangspunkt zu den im Web zu findenden Informationen dar. Hauptvorteile allgemeiner Suchmaschinen sind

- *Universalität*: Allgemeine Suchmaschinen bieten einen einheitlichen Zugang zu jeder Art von Information, da sie prinzipiell alle Webseiten indizieren, die von ihren Crawlern erreicht werden können.

- *Abdeckung*: Die Abdeckung möglichst großer Teile der im Web zu findenden Informationen spielt eine wesentliche Rolle bei Suchmaschinen.

- *Bedienbarkeit*: Bei den heute populären allgemeinen Suchmaschinen haben sich übersichtliche und minimalistisch gestaltete Benutzeroberflächen durchgesetzt.

Der Suchmaschinenmarkt wird inzwischen von einigen wenigen Firmen dominiert. Im Wesentlichen ist der Markt aufgeteilt zwischen den Anbietern Google, Yahoo und Microsoft. In den USA verfügt zudem Ask.com (früher Ask Jeeves) über geringe Marktanteile. Erwähnenswert ist noch die Suchmaschine Baidu, die durch ihre ausschließliche Präsenz im chinesischen Heimatmarkt jedoch im weltweiten Vergleich auf lediglich 0,42 Prozent Marktanteil kommt. Abbildung 13 zeigt die Marktaufteilung im Mai 2010. Google dominiert sowohl den deutschen als auch den weltweiten Markt deutlich. In Deutschland ist die Monopolstellung von Google noch deutlich ausgeprägter.

Abbildung 13: Aufteilung des Suchmaschinen-marktes weltweit und in Deutschland im Mai 2010

Darstellung basierend auf Angaben von StatCounter.com (http://gs.statcounter.com)

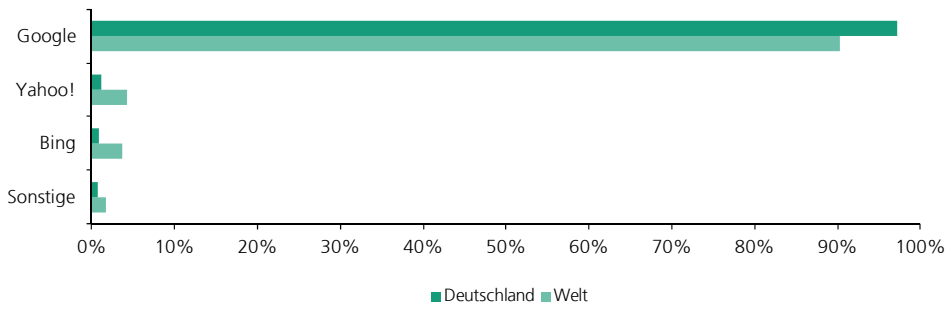
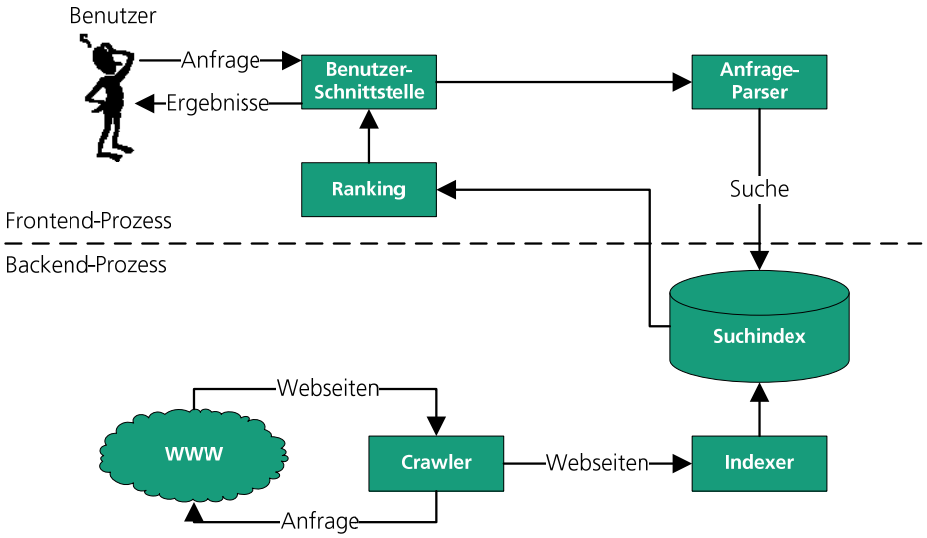


Abbildung 14 zeigt die übliche Architektur einer Web-Suchmaschine. Die Hauptbestandteile sollen im Folgenden in Kürze erläutert werden.

Abbildung 14: Architektur von Web-Suchmaschinen

Darstellung basierend auf: http://www.ibm.com/developerworks/web/library/wa-lucene



Aufgabe des *Crawlers* ist es, das Web zu durchsuchen und die gefundenen Dokumente zu indizieren. Üblicherweise beginnt der Crawling-Vorgang mit einer wohldefinierten Dokumentenmenge (»Root Set«) und folgt nach einem bestimmten Traversierungsalgorithmus den von diesen Dokumenten ausgehenden Verknüpfungen. Insbesondere aufgrund der Größe und des ungebremsten Wachstums des Internets muss der Crawlingalgorithmus hinsichtlich seiner Performanz optimiert sein.

Der *Suchindex* ist die Datenbank einer Suchmaschine. Die durch den Crawler erschlossenen Dokumente werden indiziert, die wichtigsten Begriffe und weitere für eine Suche notwendige oder hilfreiche Informationen werden gespeichert. Die Größe des Internets bedingt eine entsprechende Indexgröße, was wiederum zu entsprechenden Infrastrukturkosten auf Seite des Suchmaschinenbetreibers führt.

Der *Ranker* ist dafür verantwortlich, die gefundenen Dokumente in eine definierte Ordnung zu bringen. Das Ordnungskriterium kann dabei variieren. Leitidee ist stets, die Dokumente nach Relevanz für den Benutzer zu ordnen. Die Schwierigkeit liegt darin, dass diese Relevanz für den Nutzer ein unscharfer und in den meisten Fällen nicht einfach zu algorithmisierender Begriff ist. Faktoren, die beim Ranking berücksichtigt werden können, sind z. B. Termfrequenz, inverse Dokumentfrequenz[6], Nähe der einzelnen Suchbegriffe im Dokument, Vorkommen der Suchbegriffe an exponierter Stelle im Dokument (z. B. im Titel), Aktualität des Dokuments, Dateityp des Dokuments, Quelle des Dokuments, Anzahl der Verweise auf das Dokument, Anzahl Zugriffe auf das Dokument usw.

Das Ranking in Suchmaschinen ist von höchster Bedeutung, da aufgrund der immensen Größe des Dokumentenkorpus die Anzahl der gefundenen Treffer oftmals sehr hoch sein und der Benutzer in den meisten Fällen nicht sämtliche Treffer begutachten kann. Studien haben gezeigt, dass Benutzer nur in seltenen Fällen durch die Ergebnisseite einer Suchanfrage blättern, daher ist entscheidend, welche Dokumente im oberen Bereich der ersten Seite zurückgegeben werden – es also unter die ersten 10 bis 20 Plätze schaffen[7].

In den aktuellen generischen Suchmaschinen kommen üblicherweise Rankingverfahren zum Einsatz, die auf der Idee der Zitationsindizes basieren: PageRank (Brin und Page, 1998), HITS (Kleinberg, 1999) etc. treffen ihre Bewertung nicht, wie die ersten Suchmaschinen der 90er Jahre ausschließlich auf Basis einer Inhaltsanalyse, sondern primär auf Basis der Verlinkungsstruktur des Dokumentes. Grundannahme ist, dass auf die besten Dokumente auch am häufigsten verwiesen wird.

[6] Termfrequenz nennt man die relative Häufigkeit eines Wortes in einem Dokument. Grundannahme ist, dass wichtige Wörter im Text mehrfach vorkommen. Die inverser Dokumentfrequenz misst, wie oft ein Wort in der gesamten Dokumentenmenge vorkommt. Grundannahme ist, dass Wörter, die in besonders vielen Dokumenten vorkommen, weniger zur Relevanzbestimmung eines Dokumentes beitragen. Die Kombination von Term- und inverser Dokumentfrequenz (»TF-IDF«) wird oft als Gewichtungsmethode im Information Retrieval eingesetzt (vgl. Baeza-Yates und Ribeiro-Neto, 1999).

[7] Vgl. z. B. die Eye-Tracking-Studie des Arbeitskreises Suchmaschinen-Marketing im Bundesverband Digitale Wirtschaft (BVDW) e.V., http://www.bvdw.org/medien/bvdw-eyetracking-studie-2008?media=49.

Der Benutzer drückt seinen Informationswunsch aus, indem er eine Suchanfrage über die *Benutzerschnittstelle* eingibt. Üblicherweise besteht das Anfrage-Frontend aus einem Textfeld, in das eine Menge von Suchbegriffen eingegeben werden kann. Meist gibt es zusätzliche Möglichkeiten, das Informationsbedürfnis zu präzisieren, etwa durch eine Verknüpfung mehrerer Begriffe durch Operatoren, die Kennzeichnung von Phrasen, die Einschränkung auf bestimmte Dokumententypen oder Quellen usw. Häufig bieten Suchmaschinen ein einfaches Frontend für den »normalen« Benutzer und ein erweitertes Suchfrontend für spezifischere Suchen (»Expertensuche« oder »Profisuche«). Die Anfrage wird durch den Anfrageparser in eine von der Suchmaschine auswertbare Anfrage übersetzt. Üblicherweise werden in diesem Schritt Anfrageoptimierungen, wie etwa das Entfernen von Stoppwörtern usw. vorgenommen.

Die gefundenen und gerankten Dokumente werden dem Benutzer präsentiert. Üblicherweise geschieht dies in Form einer sortierten Liste, die neben dem Verweis auf das eigentliche Dokument einen (möglichst bzgl. der Suchanfrage relevanten) Ausschnitt oder eine Inhaltszusammenfassung sowie weitere Informationen wie z. B. Metadaten wie Autorenname, Änderungsdatum etc. enthält.

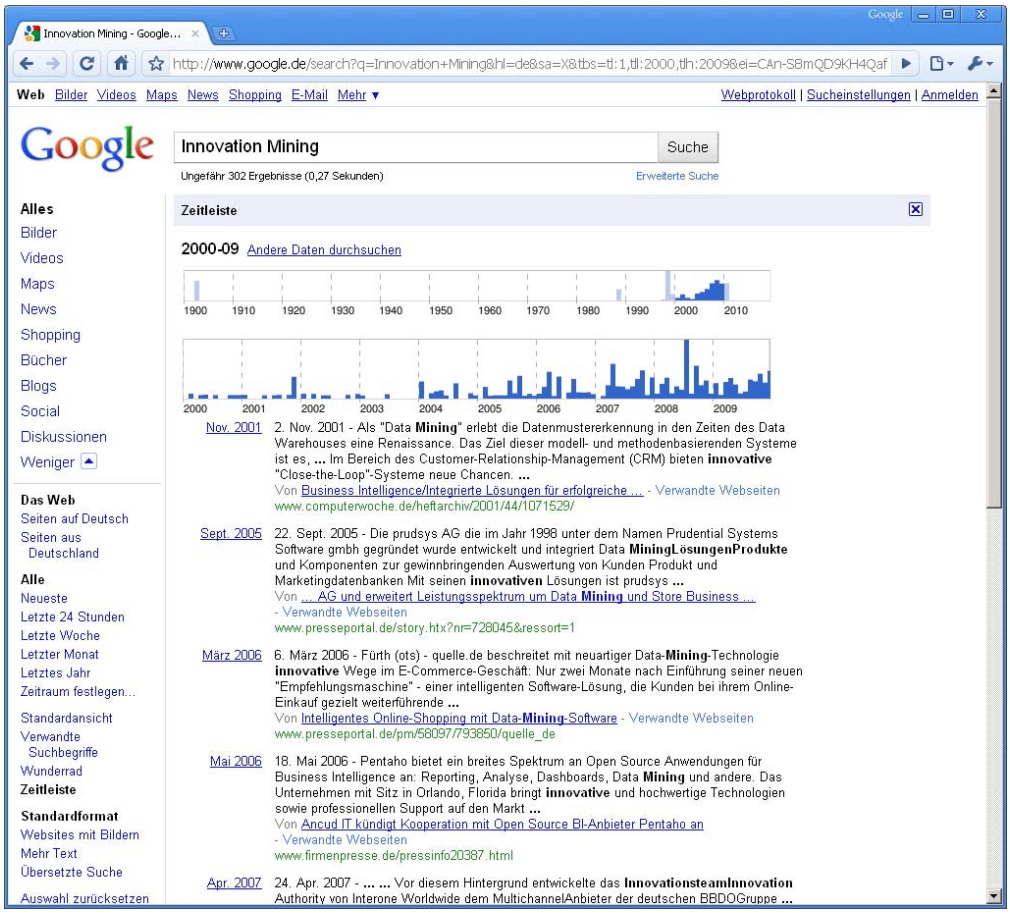
Aufgrund der für viele Fachgebiete mangelnden Präzision der generischen Suchmaschinen entsteht ein immer größerer Bedarf an spezialisierten Suchmaschinen. Dies hat u. a. folgende Gründe:

- Generische Suchmaschinen liefern oft zu viele nicht relevante Treffer.
- Generische Suchmaschinen lassen die Suchanfrage nicht in angemessener Form zu.
- Generische Suchmaschinen ranken die Ergebnisse oft nicht in angemessener Form.
- Generische Suchmaschinen stellen die Ergebnisse oft nicht in angemessener Form dar.
- Generische Suchmaschinen erlauben keine angemessene Navigation in der Dokumentenmenge.

Der Bedarf von Spezialsuchmaschinen ist auch den Betreibern der großen generischen Suchmaschinen bekannt. So bietet Google seit geraumer Zeit mit »Google Personal«, »iGoogle« und Google »Custom Search Engine« die Möglichkeit, personalisierte Suchseiten zu erstellen. Im Frühjahr 2010 hat Google auch das Frontend seiner Standard-Suchmaschine stark im

Funktionsumfang erweitert. Abbildung 15 zeigt einen Screenshot des neuen Frontends: Eines der Hauptanliegen der befragten Benutzer aus der im vorigen Abschnitt präsentierten Unternehmensbefragung wird nun adressiert – die zeitliche Einschränkung der Suchergebnisse und damit eine Fokussierung auf aktuelle Dokumente sowie eine Sortierung der Trefferliste nach Aktualität. Darüber hinaus wurden weitere Konzepte der »Facetted Search« (vgl. Hearst, 2008), also der anklickbaren Filter zur Eingrenzung der Suchergebnisse implementiert. Dadurch lässt sich die Ergebnismenge nun komfortabel auf Bilder, Videos, News, Blogs etc. einschränken, womit eine weitere wichtige Anforderung professioneller Benutzer adressiert wird.

Abbildung 15: Screenshot des neuen Google-Suchfrontends



Daneben existiert eine große Zahl weniger bekannter spezialisierter Suchmaschinen. Der »Big Search Engine Index[8]« und das Archiv von »Search

[8] http://www.search-engine-index.co.uk

Engine Land[9]« bieten umfangreiche Sammlungen von Suchmaschinen zu allen erdenklichen Themenbereichen.

2.2.1.1 Weiterführende Literatur

Eberspächer, J. und Holtel, S. (2007): Suchen und Finden im Internet. Berlin: Springer.
Sammlung von Aufsätzen, die sich mit aktuellen Entwicklungen im Suchmaschinenmarkt auseinandersetzen. Dokumentation eines Workshops des Münchner Kreises (http://www.muenchner-kreis.de).

Lewandowski, D. (2005): Web Information Retrieval: Technologien zur Informationssuche im Internet. Frankfurt am Main: DGI, 2005.
Umfassende Einführung in die Thematik der Web-Suchmaschinen.

Website Search Engine Land: http://searchengineland.com/library.
Website, die News und Informationen rund um die Themen Suchmaschinenindustrie und Suchmaschinen-Marketing zusammenstellt.

Website The Big Search Engine Index: http://www.search-engine-index.co.uk.
Umfangreiches Verzeichnis von Spezialsuchmaschinen zu jedem erdenklichen Themenfeld.

2.2.2 Unternehmensnachrichten und Pressemitteilungen

Pressemitteilungen sind ein vielgenutztes Mittel der Öffentlichkeitsarbeit in Unternehmen und können, wenn sie entsprechend ausgewertet werden, relevante Informationen zu Marktentwicklungen und Strategien von Kunden, Partnern und Konkurrenten enthalten. Der typische Rezipient einer Pressemitteilung ist ein Journalist, der einen Artikel über ein neues Thema schreiben möchte. Prinzipiell kann aber jedermann Pressemitteilungen lesen, denn die Mitteilungen können fast immer kostenlos und ohne Registrierung bezogen werden. Insbesondere werden Pressemitteilungen auch oftmals öffentlich sichtbar auf der Webseite einer Organisation platziert.

[9] http://searchengineland.com/library

2.2.2.1 Pressemitteilungen auf Unternehmenswebseiten

Unternehmen und Organisationen, die regelmäßig Pressemitteilungen veröffentlichen, präsentieren diese üblicherweise (auch) auf der eigenen Webseite. Typischerweise findet man die Meldungen in einer eigenen »Presse und Medien«-Sektion: Neuere Mitteilungen stehen zeitlich geordnet auf der Seite, ältere findet man oft im Archiv.

Es sind bereits mehrfach Fälle bekannt geworden sind, in denen Pressemitteilungen mit bewusst falschen Aussagen lanciert wurden, um kurzfristige Marktmanipulationen auszulösen[10]. Dennoch können Unternehmenswebseiten im Allgemeinen insofern als eine vertrauenswürdige und vollständige (wenn auch nicht objektive) Informationsquelle betrachtet werden, als dass eine Mitteilung, die auf der Webseite der Firma X veröffentlicht wird, tatsächlich die Meinung der Firma X darstellt und tatsächlich von der Firma selbst geschrieben wurde.

Derartig veröffentlichte Pressemitteilungen setzen natürlich voraus, dass die Leser auf die jeweilige Unternehmenswebsite gelangen, um die Pressemitteilung zu lesen – was wiederum zunächst einmal die Kenntnis der Website des Unternehmens sowie ein entsprechendes Interesse am Unternehmen seitens des Rezipienten voraussetzt. Um eine höhere Resonanz zu erreichen, werden Pressemitteilungen darum häufig alternativ oder ergänzend über Mediatoren publiziert, etwa Presseagenturen oder Presseverteiler.

2.2.2.2 Pressemitteilungsverteiler und -suchmaschinen

Allgemeine Presseagenturen wie die DPA[11] oder die Associated Press[12] sammeln Nachrichten und schicken sie an Journalisten. Oft erledigen sie eine zusätzliche Umformulierungsarbeit und sortieren Nachrichten nach Wichtigkeit. Für Pressemitteilungen von Unternehmen, die sich an andere Organisationen wenden und diese journalistische Arbeit nicht unbedingt benötigen, gibt es zusätzlich spezielle Verteiler und Webseiten, die Mitteilungen archivieren und eine Suche im Datenbestand erlauben. Typischerweise muss eine Organisation, die auf einer solchen Seite eine Pressemitteilung veröffentlichen will, einen Betrag bezahlen, während das Lesen der Meldungen kostenlos ist.

[10] Auf der Website der U.S. Securities and Exchange Commission (SEC, http://www.sec.gov) finden sich Beispiele von Kursmanipulationen mittels fingierter Pressemitteilungen, größtenteils handelt es sich dabei um Manipulationen sogenannter »Penny Stocks«, also geringwertiger Aktien, deren Kurs unter 5 US-Dollar liegt. Beispiel: http://www.sec.gov/news/press/2009/2009-117.htm.

[11] http://www.dpa.de

[12] http://www.ap.org

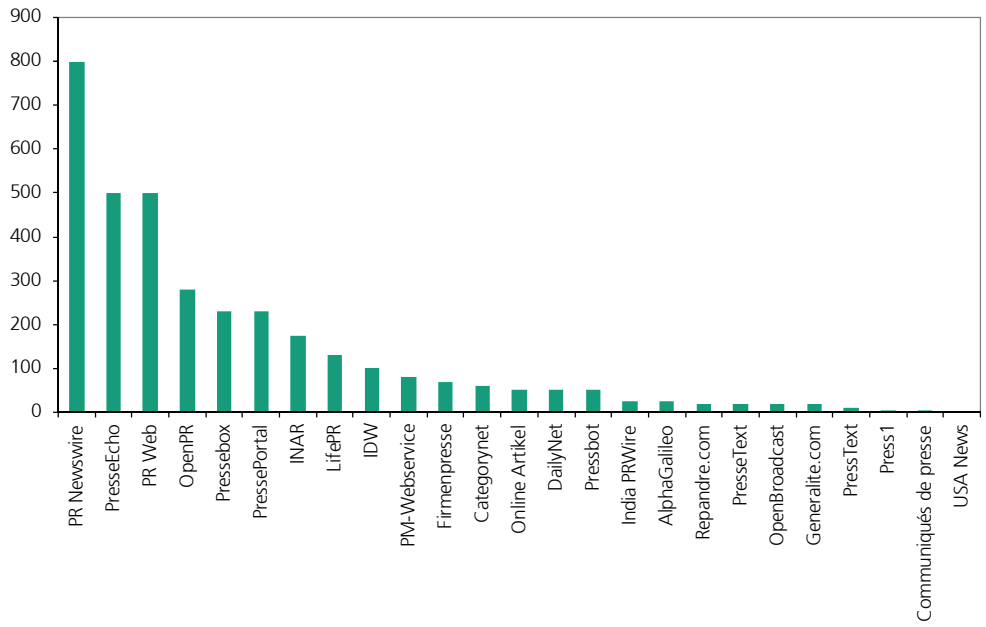
Wird eine Mitteilung auf einer solchen Seite veröffentlicht, so wird sie auf der Startseite angezeigt und über den meist von solchen Verteilern angebotenen RSS-Feed vertrieben. Oft bietet der Verteiler auch die Möglichkeit, die Mitteilung direkt an Journalisten oder ausgewählte Empfänger weiterzuleiten (meist per E-Mail-Newsletter) und sie für eine vorgegebene Zeit im Archiv zu speichern.

Unsere Marktrecherche ergab, dass in verschiedenen Ländern eigene Verteilerseiten populär sind und viele themenspezifische Verteiler existieren. Für die im Folgenden dargestellten Analysen wurden hauptsächlich deutsche und amerikanische Seiten im Bereich Wirtschaft und Wissenschaft untersucht.

Zu Beginn der Analyse wurden die wichtigsten Pressemitteilungswebsites identifiziert. Die Auswahl erfolgt aufgrund von Expertenbefragungen (so wurde etwa das Pressebüro des Fraunhofer IAO hinzugezogen) sowie auf Basis von Webrecherchen und eigenen Vergleichen (insbesondere bzgl. der veranschlagten Preise und der Anzahl der jeden Tag veröffentlichten Mitteilungen).

Alle untersuchten kostenlosen Websites nutzen zur Finanzierung Anzeigen (fast immer auf Basis von Google AdSense). Ist die Veröffentlichung einer Mitteilung kostenpflichtig, so bewegen sich die Preise zwischen 100 und 300 Euro pro Mitteilung. Der Preis variiert dabei mit den Optionen, die die Webseite anbietet (zum Beispiel Verbreitungsgrad der Veröffentlichung, Archivierung von Anhängen usw.). Diese Anzeigen stören aber oft die Lesbarkeit der Seite.

Abbildung 16: Veröffentlichte Mitteilungen pro Tag auf verschiedenen Pressemitteilungs-diensten



Wie Abbildung 16 verdeutlicht, sind nur wenige Verteiler besonders aktiv (mehr als 200 Mitteilungen pro Tag), auf vielen werden nur vergleichsweise wenig Mitteilungen pro Tag veröffentlicht (weniger als 50). Aus der »Long Tail«-artigen Verteilungskurve lassen sich zwei Konsequenzen ableiten:

1. Um den Großteil aller in einem Land veröffentlichen Pressemitteilungen zu lesen reicht es, wenige Verteiler zu betrachten.

2. Um eine vollständige Abdeckung zu erhalten, müssen jedoch auch die zahlreichen kleinen Verteiler berücksichtigt werden, denn Stichproben ergaben, dass viele Meldungen ausschließlich über die kleinen Verteiler publiziert werden.

Tabelle 3: Bewertung verschiedener Pressemitteilungs-Verteilerdiensten

Name	Sprache	Kosten (€)	Such-anfragen	Ergebnis-darstellung	RSS feed-Qualität	Gesamt-bewertung Suche	Anzahl Meldungen pro Tag
PR Newswire	EN	120	●○	●○	●●●	●●	●●●
PresseEcho	DE	-	○	●	●	●	●●●
PR Web	EN	80-230	●	○	○	○	●●●
OpenPR	DE	-	○	●	○	○	●●●
Pressebox	DE	200	●	●●	●●	●●	●●●
PressePortal	DE	315	●	●●	●●●	●●	●●●
INAR	DE	-	○	○	●●	●	●●
LifePR	DE	99	●●	●●	●●	●●	●●
PM-Webservice	DE	-	●○	●	●	●	●
Firmenpresse	DE	-	●	●	●●	●	●
Online Artikel	DE	-	●	●○	●	●	●
DailyNet	DE	-	●	●	●	●	●
Pressbot	DE	-	●	○	●	●	●
AlphaGalileo	EN-DE	50-100	●	●●	●●	●●	○
PresseText	DE	167	●●	●○	○	●	○
OpenBroadcast	DE	-	●	●	○	●	○
PressText	DE	160	●●	●●	-	●	○
Press1	DE	-	○	●	●●	●	○

Spaltenerklärung: Kosten für die Veröffentlichung einer Pressemitteilung; Bewertung des Suchanfrage-Frontends: ●●● = viele Optionen und einfache Benutzung, ○ = wenige Optionen, schwierige Benutzung; Bewertung der Suchergebnisdarstellung: ●●● = sehr klare und vollständige Darstellung, ○ = weniger Informationen, unklare Darstellung; RSS-Feed-Qualität: ●●● = Feed enthält alle Informationen, ○ = wichtige Informationen sind nicht im Feed enthalten, - = kein Feed verfügbar; Gesamtbewertung Suche: Mittelwert aus Bewertung des Suchanfrage-Frontends, der Bewertung der Suchergebnisdarstellung und der RSS-Feed-Qualität; Anzahl Meldungen pro Tag: ●●● = 200 und mehr, ●● = 100 bis 200, ● = 50 bis 100, ○ = 0 bis 50.

2.2.2.3 Anwendungsbeispiele

Im Innovationsmanagement sind Pressemitteilungen als Medium zur Marktbeobachtung von besonderer Relevanz, weil sie einem Unternehmen viel über Aktivitäten von Partnern, Zulieferern, Kunden und Konkurrenten verraten können:

- *Beobachtung der letzten Entwicklungen in einem Bereich* (z. B. einem Technologiefeld): Was sind die neuesten Erfindungen, die dem Unternehmen nützlich sein könnten?

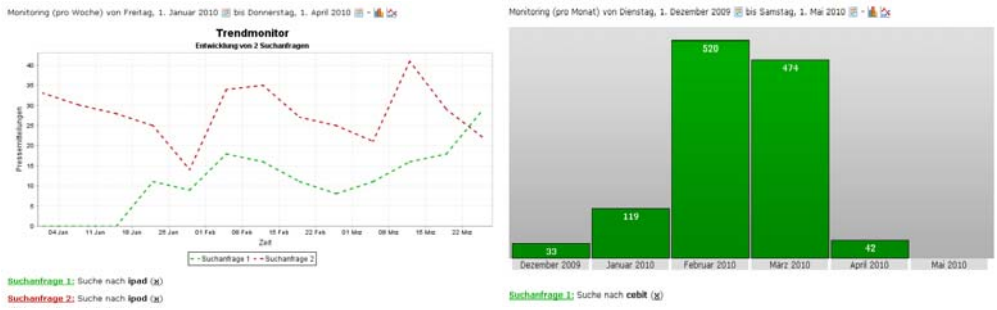
- *Wettbewerbsbeobachtung*: Welche Strategien verfolgen Wettbewerber und was bedeutet dies für die eigene Unternehmensstrategie?

- *Beobachtung geschäftlicher Kunden*: Wer sind die Kunden und welches sind ihre Bedürfnisse? Wer könnte ein neuer Kunde werden?

Pressemitteilungen, die von Partnerfirmen oder Wettbewerbern, Lieferanten oder Kunden veröffentlicht werden, stellen eine Informationsquelle dar, die bei der Beantwortung dieser Fragen helfen kann. Neben den Inhalten der neuesten Pressemitteilung kann dabei auch die Interpretation von Daten aus älteren Meldungen herangezogen werden – z. B. die Menge und Frequenz, mit der ein Unternehmen Mitteilungen zu bestimmten Themen veröffentlicht. Mit passenden Werkzeugen kann es ermöglicht werden, neues Wissen zu entdecken oder schon bekannte Informationen zu vertiefen. Abbildung 17 zeigt Screenshots einer von Fraunhofer IAO entwickelten Metasuchmaschine für Pressemitteilungen: Das linke Bild stellt die Publikationsfrequenz zu den Themen »iPod« und »iPad« gegenüber (Mitte März überstieg das Mitteilungsaufkommen zu Apples iPad erstmalig das des iPods), das rechte analysiert das Meldungsaufkommen zu einem bestimmten Thema am Beispiel »CeBIT«.

Abbildung 17: Screenshots statistischer Auswertungen der Meldungsfrequenz

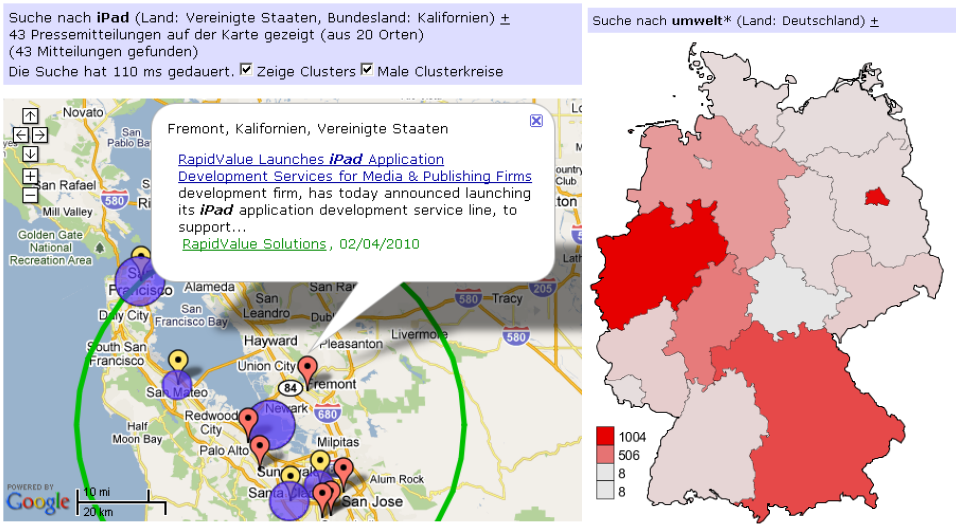
(Fraunhofer IAO Innovation Mining Cockpit)



Neben dem zeitlichen Aspekt können auf ähnliche Weise auch die in Pressemitteilungen üblicherweise enthaltenen Ortsangaben genutzt werden, um die geographische Verteilung darzustellen. Abbildung 18 verdeutlicht dies abermals anhand von Screenshots des Fraunhofer-Prototyps: Das linke Bild zeigt ein erhöhtes Aufkommen von Meldungen zum Thema »Software« im Umkreis des Silicon Valleys, das rechte Bild stellt mithilfe einer gefärbten Karte die Meldungshäufigkeit zum Thema »Umwelt« dar.

Abbildung 18: Screenshots geographischer Auswertungen der Meldungsfrequenz

(Fraunhofer IAO Innovation Mining Cockpit)



2.2.2.4 Weiterführende Literatur

Buschardt, T. und Krath, S. (2002): Pressemitteilung: Inhalt, Form, Praxis. Luchterhand, Neuwied.
Systematische Analyse guter und weniger guter Pressemitteilungen.

Falkenberg, V. (2002): Pressemitteilungen schreiben, Frankfurter Allgemeine Buch, Frankfurt am Main.
Ratgeber zum Verfassen und Managen erfolgreicher Pressemitteilungen.

Finzen, J., Kintz, M., Kett, H. und Koch, S. (2009): Strategic Innovation Management on the Basis of Searching and Mining Press Releases. In: Proceedings of the Fifth International Conference on Web Information Systems and Technologies, Lissabon, 2009: INSTICC Press, S. 347–353.
Dieses Paper diente als Grundlage des vorliegenden Abschnitts und beschäftigt sich mit der Analyse der Online-Landschaft von Pressemitteilungsverteilern sowie der Entwicklung einer Metasuchmaschine für Pressemitteilungem..

Thaeler, J. M. (2009): I Need a Killer Press Release - Now What???: A Guide to Online PR. HappyAbout.info.
Aktuelle Auseinandersetzung mit der Nutzung von Pressemitteilungen als Marketinginstrument. Schwerpunkt auf der Nutzung von Onlinemedien und begleitenden Social-Web-Aktivitäten.

2.2.3 Wissenschaftliche Publikationen

Wissenschaftler veröffentlichen ihre Erkenntnisse Konferenzbänden und Fachmagazinen. Wissenschaftliche Veröffentlichungen stellen auch für das strategische Innovationsmanagement eine wichtige Datensammlung dar, die sich mit den richtigen Tools leicht verwerten lässt. Zwei wesentliche Anforderungen an wissenschaftliche Veröffentlichungen sind Qualität und Aktualität – also Anforderungen, die auch im Innovationsmanagement an hochwertige Quellen gestellt werden:

- *Qualität:* Ein Hauptziel wissenschaftlicher Publikationen ist die Veröffentlichungen von Inhalten, die *originär* sind und sich *an der besten international üblichen Praxis* orientieren sollen (Deutsche Forschungsgemeinschaft, 1998). Die Sicherstellung einer hohen Qualität ist also bei wissenschaftlichen Veröffentlichungen ein ganz wesentlicher Aspekt – und üblicherweise ein erklärtes Ziel jedes wissenschaftlichen

Journals und jeder Konferenz. Wissenschaftliche Publikationen werden meist in einem »Peer-Review«-Verfahren evaluiert, das heißt, sie werden anonym durch externe Experten geprüft und erst nach erfolgreicher Begutachtung veröffentlicht. Die Qualitätssicherung wissenschaftlicher Beiträge variiert allerdings sehr stark. Bei Konferenzen erfolgt die Annahme von Beiträgen mitunter lediglich auf Basis kurzer Zusammenfassung, auch gibt es sehr unterschiedliche Standards im Reviewprozess. Bedingt aussagekräftig sind die Annahmequoten einer Konferenz – üblicherweise deutet eine niedrige Annahmequote (insbesondere bei einer hohen Einreichungsmenge) auf eine höherwertige Konferenz hin. Wissenschaftliche Fachmagazine haben im Allgemeinen eine striktere Qualitätssicherung, als Konferenzen – weswegen erstere üblicherweise als höherwertig angesehen werden.

- *Aktualität*: Die Aktualität einer Information ist für den Verfasser einer wissenschaftlichen Veröffentlichung von entscheidender Bedeutung. Das Interesse der Fachwelt ist ihm schließlich nur dann sicher, wenn seine Erkenntnisse dem neuesten Kenntnisstand der Wissenschaft und Technik entsprechen oder diesen (idealerweise) erweitern. Allerdings vergeht zwischen der Einreichung eines Beitrags zu einer Konferenz und der öffentlichen Verfügbarkeit des Beitrags in einem Tagungsband oftmals eine erhebliche Zeit (normalerweise mehrere Monate). Ein noch längerer Zeitraum liegt häufig zwischen der Einreichung und tatsächlichen Veröffentlichung eines Beitrags in einem Fachmagazin (mitunter mehrere Jahre). Daher gewinnt die Vorveröffentlichung auf sogenannten »Preprint«-Servern zunehmend an Bedeutung (s. u.).

Im Folgenden werden die unterschiedlichen Arten wissenschaftlicher Veröffentlichungen bzgl. ihrer Eignung für das »Innovation Mining« genauer untersucht.

2.2.3.1 Kategorisierung

Online-Datenbanken für wissenschaftliche Publikationen lassen sich in drei Kategorien unterteilen: Primäre Quellen, spezialisierte Datenbanken und spezielle Webseiten allgemeiner Suchmaschinen[13].

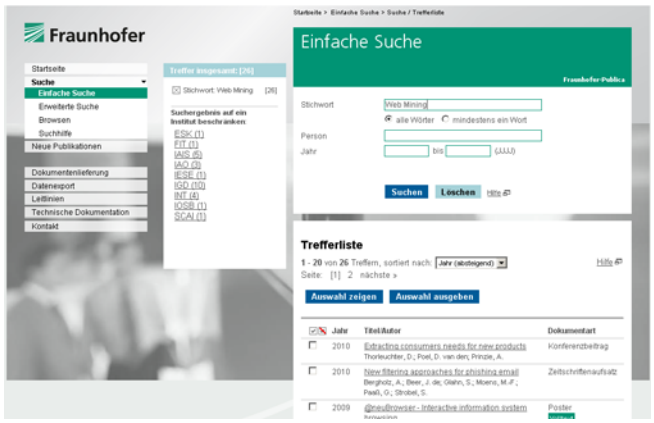
- *Primäre Quellen* sind vor allem Verlags-Angebote (wie z. B. Springer Link[14], über das ausschließlich vom Springer-Verlag veröffentlichte

[13] Anmerkung: Der Verlauf der Trennung zwischen der zweiten und dritten Gruppe ist mitunter nicht klar zu ziehen.
[14] http://www.springerlink.de

Publikationen online zur Verfügung gestellt werden), sowie Angebote von Forschungseinrichtungen (wie z. B. Fraunhofer Publica[15] für Publikationen der Fraunhofer-Gesellschaft) oder Berufsverbänden (wie z. B. IEEE Xplore[16] für Dokumente aus dem Bereich Informatik und Elektrotechnik). Diese Seiten stellen nur selten den Volltext der Dokumente kostenlos zur Verfügung und bieten Zugang nur zu einer begrenzten Anzahl an Veröffentlichungen (üblicherweise den eigenen). Die Reputation dieser Quellen hängt direkt von der Reputation des Verlags oder der Forschungsorganisation ab, wird aber oft als sehr hoch gesehen. Die Fachgebiete der Publikationen richten sich nach der fachlichen Ausrichtung des Betreibers.

Abbildung 19: Suchergebnisse Fraunhofer Publica

(http://publica.fraunhofer.de)



- *Spezialisierte Datenbanken* fassen die Daten mehrerer primärer Quellen zusammen. Sie können aus rechtlichen Gründen oft nur eine Abstract- und Zitationssuche anbieten: Um auf den Volltext zugreifen zu können, muss der Benutzer sich direkt an dem Herausgeber der Publikation wenden. Der Zugang zu den Suchfunktionen kann kostenlos oder kostenpflichtig sein. Populäre Beispiele sind CiteSeerX[17] der Pennsylvania State University, Web of Science[18] von Thomson Reuters

[15] http://publica.fraunhofer.de
[16] http://ieeexplore.ieee.org
[17] http://citeseerx.ist.psu.edu
[18] http://apps.isiknowledge.com. Anmerkung: Das »Web of Science« ist in das breiter aufgestellte »Web of Knowledge« eingebunden, das neben den Journal-Inhalten des Web of Science auch weitere wissenschaftlich relevante Inhalte (Forschungswebsites, Patentdaten etc.) enthält.

und die beiden Angebote Scopus[19] und ScienceDirect[20] der Firma Elsevier[21].

Abbildung 20: Suchergebnisse bei CiteSeerX

(http://citeseerx.ist.psu.edu)

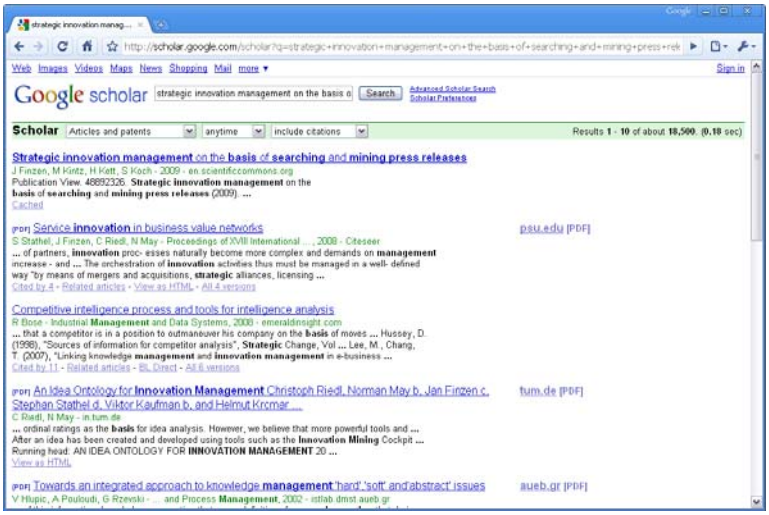


- Betreiber *allgemeiner Suchmaschinen* wie Google und Microsoft bieten heute im Allgemeinen spezielle Suchfunktion für wissenschaftliche Publikationen an (Google Scholar[22], Microsoft Academic Search[23]). Ihre Stärke ist die sehr hohe Anzahl an indizierten Dokumenten - teilweise werden auch Informationen zu Dokumenten als Ergebnis zurückgeliefert, welche gar nicht online zu finden sind, sondern lediglich aus anderen Dokumenten heraus referenziert wurden. Die Suchmaschinen zeigen mitunter neben Journal- und Konferenzbeiträgen und wissenschaftlichen Monographien auch Dissertationen, Diplom- und Studienarbeiten und Praktikumsberichte an. Diese Suche ist kostenlos, der Zugang zum Volltext ist hier ebenfalls vom Verlag abhängig.

[19] http://info.scopus.com
[20] http://www.sciencedirect.com
[21] Scopus enhält ausschließlich Abstracts und Zitationen, während ScienceDirect zwar insgesamt weniger Journale indiziert, dafür aber auch Zugriff auf Volltexte bietet.
[22] http://scholar.google.de
[23] http://academic.research.microsoft.com

Abbildung 21: Suchergebnisse bei Google Scholar

(http://scholar.google.de)



2.2.3.2 Funktionsumfang

Die Standard-Suchfunktionen von Online-Datenbanken für wissenschaftliche Publikationen erlauben normalerweise eine Suche in den Titeln und Zusammenfassungen, eine Suche nach Autor, eventuell auch nach der Organisation des Autors, nach Veröffentlichungsdatum, manchmal auch nach Konferenz oder Verlag. Seiten wie CiteSeerX erlauben eine zitationsbasierte Suche oder eine Filterung der Dokumente nach Anzahl an Zitationen. Meist beschränkt sich die Ergebnisanzeige ebenfalls auf die genannten Informationen - die Volltexte sind nur in seltenen Fällen frei verfügbar (s. u.), in den meisten Fällen findet man sie nur auf Verlagswebseiten und gegen Gebühr. Forschungseinrichtungen und –abteilungen bieten ihren Mitarbeitern oft Zugang zu den Volltexten bestimmter Quellen (etwa eine fachlich beschränkte Auswahl der von einem Verlag angebotenen Journale). Die Gebühren werden direkt von der Forschungseinrichtung bezahlt, die Datenbank erlaubt den Zugriff für einen bestimmten IP-Adress-Bereich, der Nutzer kann dann direkt im Browser ohne zusätzliche Registrierung an seinem Arbeitsplatz die Texte lesen.

Preprints

Da der Review- und Veröffentlichungsprozess bei wissenschaftlichen Fachmagazinen mitunter mehrere Jahre in Anspruch nehmen kann, veröffentlichen viele Autoren ihre Arbeit vorab auf Plattformen, die eine schnellere Verbreitung der Ergebnisse erlauben, indem sie ungeprüfte Arbeiten

veröffentlichen – sogenannte »Preprint«-Plattformen. Populäre Beispiele sind arXiv[24], Nature Precedings[25] und das Preprint-Archiv von ScienceDirect[26].

Open Access

Da viele wissenschaftliche Datenbanken die Volltexte ihrer Veröffentlichungen nur kostenpflichtig anbieten, kann der breite Zugang zur Literatur für kleine Firmen oder Organisationen schnell teuer werden. Der unfreie Zugang zu Forschungsergebnissen stellt darüber hinaus ein globales Problem dar, denn insbesondere für ärmere Länder ergibt sich hieraus ein weiterer Nachteil. Deswegen - und um die Sichtbarkeit von Artikeln im Allgemeinen zu erhöhen - hat sich in den letzten Jahren die sogenannte »Open Access«-Bewegung[27] etabliert. Diese fordert unter anderem mit dem 2003 verabschiedeten »Bethesda Statement on Open Access Publishing«[28], das von verschiedenen amerikanischen, englischen und auch deutschen Universitäten und Forschungseinrichtungen unterzeichnet wurde, den kostenlosen und zeitlich wie geographisch unbeschränkten Zugang zu Volltexten sowie die Möglichkeit und das Recht, die Inhalte zu nutzen, wiederzugeben und zu kopieren, insofern der ursprüngliche Autor benannt wird.

Für die Veröffentlichung von Texten existieren im Open Access-Ansatz zwei Methoden, die sich auch im Logo der Bewegung wiederfinden (Abbildung 22):

- Der »goldene Weg« ist die Veröffentlichungen von Volltexten in speziellen Open Access-Datenbanken.

- Der »grüne Weg« ist die Veröffentlichung von Volltexten auf privaten Homepages der Autoren, normalerweise mit einem Eintrag in einer institutionellen Open Access-Datenbank.

Abbildung 22: Logo der Open Access Bewegung



Auch die Fraunhofer-Gesellschaft hat sich 2007 verpflichtet, alle Forschungsergebnisse ihrer Mitarbeiter soweit möglich frei zugängig zu

[24] http://arxiv.org/
[25] http://precedings.nature.com
[26] http://www.sciencedirect.com/preprintarchive
[27] http://www.open-access.net
[28] http://www.earlham.edu/~peters/fos/bethesda.htm

machen und für eine kostenlose Nutzung ins Internet zu stellen. Die Publikationen findet man auf der ePrints-Seite der Fraunhofer-Gesellschaft[29].

Tabelle 4: Suche nach wissenschaftlichen Veröffentlichungen im Internet

	Name	URL	Betreiber	Spezialgebiet	Zugang	PrePrint
Primäre Quellen	Fraunhofer ePrints	http://eprints.fraunhofer.de	Fraunhofer-Gesellschaft	Nein	Open Access	Nein
	Fraunhofer Publica	http://publica.fraunhofer.de	Fraunhofer-Gesellschaft	Nein	Fraunhofer-Intern	Nein
	IEEE Xplore	http://ieeexplore.ieee.org	IEEE	EDV	Kostenpflichtig	Nein
	Nature Preceedings	http://precedings.nature.com	Nature Publishing Group	Naturwissenschaften	Kostenlos	Ja
	SpringerLink	http://www.springerlink.com	Springer	Nein	Kostenpflichtig	Nein
Spezialisierte Datenbanken	arXiv	http://arxiv.org	Cornell University	Naturwissenschaften, Wirtschaft	Nur Suche	Ja
	CiteSeerX	http://citeseerx.ist.psu.edu	Pennsylvania State University	Nein	Nur Suche	Nein
	OpenDOAR	http://www.opendoar.org	University of Nottingham	Nein	OpenAccess	Nein
	Public Library of Science	http://www.plos.org	Public Library of Science	Nein	OpenAccess	Nein
	Scopus	http://www.scopus.com	Elsevier	Nein	Kostenpflichtig	Nein
	Web of Science	http://apps.isiknowledge.com	Thomson Reuters	Naturwissenschaften, Wirtschaft	Kostenpflichtig	Nein
Such-maschinen	Google Scholar	http://scholar.google.com	Google	Nein	Nur Suche	Ja
	iSeek	http://education.iseek.com	Vantage Labs, LLC	Nein	Nur Suche	Ja
	Microsoft Academic Search	http://academic.research.microsoft.com	Microsoft	Nein	Nur Suche	Ja

2.2.3.3 Anwendungsbeispiele

Wissenschaftliche Literaturdatenbanken bieten sich für die folgenden innovationsstrategisch relevanten Analysearten an:

- *Untersuchung des wissenschaftlichen State-of-the-Art*: Mittels Analyse der aktuellen wissenschaftlichen Publikationen kann der aktuelle Forschungsstand analysiert werden. Üblicherweise beinhalten wissenschaftliche Arbeiten selbst eine Analyse des State-of-the-Art, so dass durch Lektüre einiger relevanter Texte bereits ein recht vollständiger Überblick über ein Thema gewonnen werden kann.

- *Identifikation von Forschungsbedarfsfeldern*: Die Analyse des State-of-the-Art zeigt die aktuellen Grenzen der Forschung auf. Wissenschaftliche Arbeiten enthalten zudem üblicherweise einen Ausblick auf »Future Research«.

- *Erkennung von »Hot Topics«*, an denen aktuell massiv geforscht wird. Dies kann über eine zeitliche Analyse der Anzahl von zu einem Thema

[29] http://www.verlag.fraunhofer.de/eprints/

erscheinenden Arbeiten erfolgen. Da wissenschaftliche Arbeiten im Allgemeinen ein Veröffentlichungsdatum tragen, ist diese Art Analyse einfach durchführbar.

- *Expertensuche*: Es werden Forscher oder Forschungseinrichtungen identifiziert, die an bestimmten Themenstellungen forschen.

- *Identifikation von Forschungsnetzwerken*, etwa basierend auf Zitations- oder Co-Authorship-Analysen.

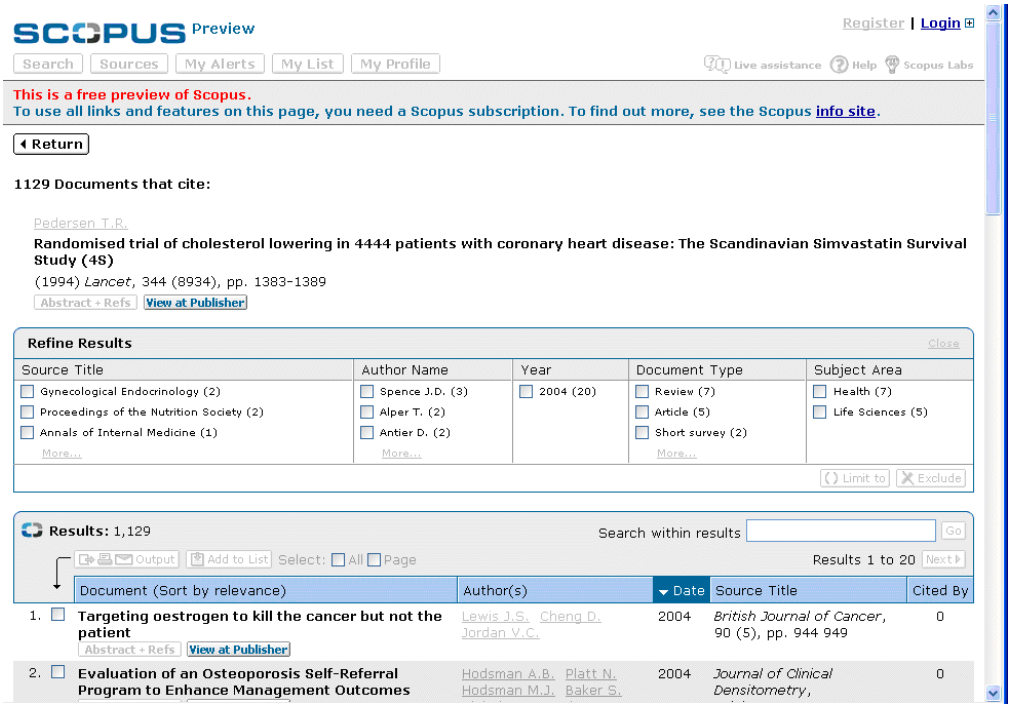
Die beiden letztgenannten Anwendungsfälle haben durch die Möglichkeiten moderner Internetdatenbanken enorm an Popularität gewonnen. Aufgrund der einigermaßen standardisierten Form wissenschaftlicher Arbeiten lassen sich mittels der relativ leicht automatisch extrahierbaren Autor- und Quellenangaben Assoziationsgraphen zwischen Autoren erstellen:

- *Co-Authorship*: Wissenschaftliche Publikationen werden oftmals von mehreren Autoren gemeinsam erstellt. Über die Frequenz des gemeinsamen Auftretens zweier Namen lassen sich Rückschlüsse auf die Intensität ihrer Kollaboration ziehen.

- *Zitation*: Wissenschaftliche Publikationen enthalten stets Verweise auf andere wissenschaftliche Arbeiten. Die Messung der Zitationshäufigkeit erlaubt nicht nur die Bestimmung eines Näherungswertes des »wissenschaftlichen Impacts« einer Arbeit oder der Expertise eines Autors, sondern auch die Analyse der »Forscher-Szene« zu einem bestimmten Thema. Abbildung 23 zeigt die Zitationssuche mit Scopus.

Abbildung 23: Zitationssuche mit Scopus

(http://info.scopus.com/documents/files/scopus-training/resourcelibrary/images/ss1_big.gif)



2.2.3.4 Weiterführende Literatur

Bosman, J., Mourik, I., Rasch, M., Sieverts, E. und Verhoeff, H. (2006): Scopus reviewed and compared: the coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar. Utrecht University Library.
In dieser 60-Seitigen Studie wird die wissenschaftliche Datenbank Scopus detailliert analysiert und mit anderen Plattformen wie Web of Science verglichen.

Deutsche UNESCO-Kommission (2007): Open Access. Chancen und Herausforderungen. Ein Handbuch Deutsche UNESCO-Kommission. Bonn 2007. Verfügbar als pdf-Dokument über Webseite der UNESCO-Kommission (http://www.unesco.de/1280.html?&L=0).
Das Handbuch definiert den Begriff »Open Access«, beschreibt die Lage in Deutschland für verschiedene wissenschaftliche Bereiche und diskutiert detailliert politische Aspekte der Open Access-Bewegung.

Eysenbach G. (1999): Challenges and changing roles for medical journals in the cyberspace age: Electronic pre-prints and e-papers. J Med Internet Res 1999;1(2).

Garfield, E. (1979): Citation Indexing - Its Theory and Application in Science, Technology, and Humanities. Wiley, New York.

Nacke, O. (1980): Zitatenanalyse und verwandte Verfahren. IDIS, Bielefeld.

Newby, G. B. und Peek, R. P. (1996): Implementing Peer Review on the Net: Scientific Quality Control. In Scholarly Electronic Journals. Scholarly publishing: the electronic frontier. Cambridge, Mass: MIT Press. S. 103–118.
In diesem 14 Jahre alten Paper werden Chancen und Risiken der Web-basierten Publikation wissenschaftlicher Forschungsergebnisse besprochen, der Fokus liegt dabei auf Aspekten der Qualitätssicherung.

2.2.4 Patentdaten

Patente stellen eine wertvolle Wissens- und Informationsquelle dar, denn in ihnen ist Wissen zu speziellen Themenbereichen strukturiert oder zumindest halbstrukturiert verfügbar. Die Recherche in Patentdatenbanken bietet die Möglichkeit, sich über den aktuellen State-of-the-Art eines Technologiefeldes zu informieren, nach neuen Produkten oder Verfahren zu suchen, potenzielle Wettbewerber oder Kooperationspartner zu finden, Rechtsverletzungen vorzubeugen und vor allem teure Doppelentwicklungen auszuschließen. Für einen Großteil der in Patentdatenbanken beschriebenen Patente ist der Patentschutz bereits abgelaufen, sodass dort beschriebene Verfahrensweisen für eine Adaption zur freien Verfügung stehen. Zahlreiche nationale Patentdatenbanken sind inzwischen über das Internet zugreifbar. Für automatisierte Textverarbeitungsverfahren wie Mining-Prozesse und Informationsextraktion eignen sich diese Datenbestände auch deshalb, weil sie (zumindest innerhalb derselben Patentdatenbank) einigermaßen homogenisiert vorliegen – und der Vorverarbeitungsaufwand sich daher in Grenzen halten lässt.

Tabelle 5:
Beispiele üblicher Patentanalysen

Darstellung übernommen aus (Siwczyk, 2009)

	Analysemethode	Beispiel	Bemerkung
Einfache statistische Methoden	Analyse der Anmelder	Wer ist Technologieführer?	Je nach Analyseziel Bildung von Kennzahlen durch die Kombination mehrerer Parameter --> Ansätze für qualitative Bewertung der Patentsituation
	Länderanalyse	In welchen Ländern werden wieviele Patente	
	Klassananalyse	IPC-Übersichten	
	Trendanalyse	In welchen Jahren wurden wieviele Patente angemeldet?	
	Zitationsanalyse	Wie oft wird ein Patent in anderen Patenten zitiert?	
Text- und Datamining-Methoden	Keyword-Analyse	Welches sind die Top-25 Keywords?	
	Freedom-to-Operate-Analyse	Welche Patente kennzeichnen den Stand der Technik und welche davon sind aktiv?	Insbesondere zur Vermeidung von Patentverletzungen in der frühen Produktentwicklungsphase
	Patentlandkarte	Welcher Wettbewerber hält welche Patente zu einem Themenfeld?	Textmining-Verfahren kombiniert mit statistischen Analysen; Visualisierung
	Whitespot-Analyse (FhG IAO)	Welche wirtschaftlich attraktiven, ungelösen Probleme gibt es in einem Technologiefeld?	Textmining-Verfahren kombiniert mit statistischen Analysen auf Basis von Patent- und Marktdaten; Visualisierung

2.2.4.1 Nationale und internationale Patentdatenbanken

Nationale und internationale Patentdatenbanken bieten die Möglichkeit, mithilfe des Internets gezielt auf ihren Datenbestand zuzugreifen:

- DEPATISnet ist der Patentserver des Deutschen Patent- und Markenamts[30]. Er enthält bibliographische Daten und Faksimiledarstellungen der in den letzten 24 Monaten erschienenen deutschen Offenlegungs- bzw. Patentschriften (das Informationsangebot ist also sehr aktuell, jedoch auf die letzten beiden Jahre eingeschränkt). Im Rahmen von esp@cenet[31], eines Kooperationsprojektes der Europäischen Patentorganisation[32] (EPO) und deren Mitgliedsstaaten, wird außerdem über DEPATISnet Zugang auf Patentinformationen der anderen Mitgliedstaaten ermöglicht.

- Die »Patent Full-Text and Full-Page Image Databases« des US-Patent- und Warenzeichenamtes[33] (United States Patent and Trademark Office, USPTO) enthalten Volltextbeschreibungen der in den USA veröffentlichten Patente seit 1976 sowie alle seit 1790 vergebenen Patente als gescannte Grafiken (das TIFF-Format erfordert ein spezielles Browser-Plug-in). Auf den Volltexten kann stichwortbasiert recherchiert werden, zudem steht eine erweiterte Suche zur Verfügung, die viele spezifisch Möglichkeiten der Suchbereichseingrenzung vorsieht (Patentnummern und –klassen, diverse Datums- und Ortsangaben, Angaben zum Patentinhaber usw.). Eine Recherche auf den TIFF-Bildern erfordert dagegen die Eingabe der Patentnummer.

Neben diesen beiden wichtigen Patentdatenbanken gibt es eine ganze Reihe weiterer vorwiegend nationaler Vertreter, eine Übersicht bietet z. B. das Patentinformationszentrums (PATON) der Technischen Universität Ilmenau[34].

2.2.4.2 Kommerzielle Patentdatenbanken

Neben den von öffentlichen Trägern betriebenen Patentdatenbanken gibt es eine Reihe von kommerziellen Anbietern von Patentdaten. Deren Mehrwert besteht insbesondere in dem vereinheitlichten Zugriff auf die Datenbestände verschiedener staatlicher Patentdatenbanken, teils mächtigere Recherchefunktionen sowie bessere Möglichkeiten der Anbindung an Inhouse-Systeme. Im Folgenden werden einige prominente Beispiele aufgeführt.

- Thompson Reuters bietet nach eigener Aussage mit Derwent World Patents Index (DWPI) die weltweit umfangreichste Datenbank für

30 http://depatisnet.dpma.de
31 http://www.espacenet.com
32 http://www.epo.org
33 http://www.uspto.gov/patft/index.html
34 http://www.paton.tu-ilmenau.de

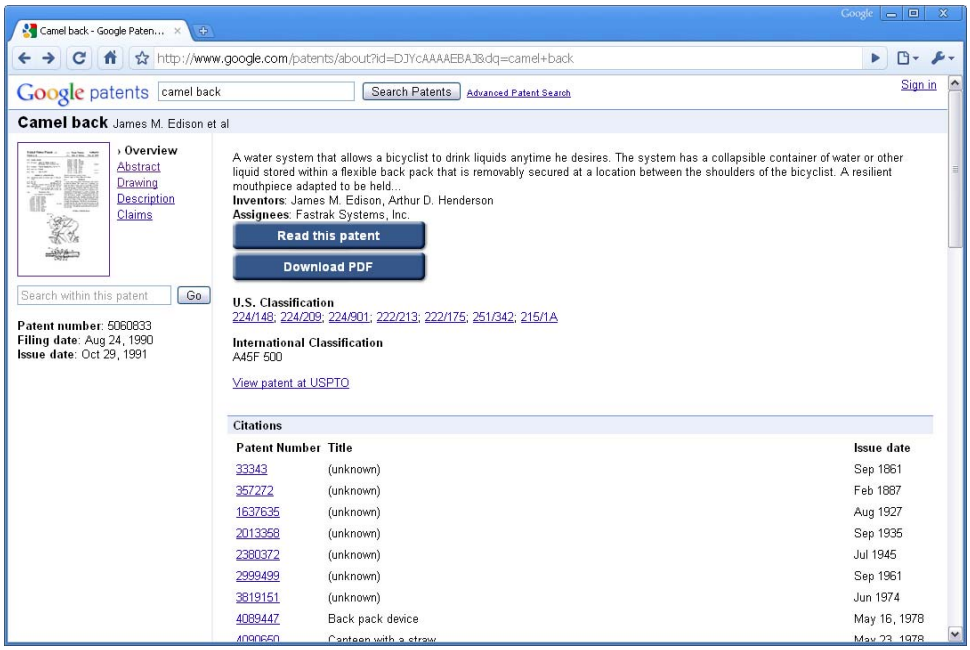
Patentschriften. Sie enthält 41 Millionen Dokumente von 41 Patentämtern. Der Zugriff auf die DWPI-Daten erfolgt über die Services Thomson Innovation[35] oder Delphion[36], darüber hinaus werden Schnittstellen zu Inhouse-Systemen und den Dialogsystemen Dialog[37], Questel.Orbit[38] und STN[39] angeboten.

- Questel.Orbit betreibt die Patentdatenbanken PlusPat[40] und FamPat[41]. Nach Firmenaussagen enthält PlusPat über 60 Millionen Patentdokumente von 92 Patentbehörden. Das Hauptkriterium bei PatFam ist die Gruppierung von Patentinformationen zu bestimmten Erfindungen (da oftmals Einzelaspekte derselben Erfindung in separaten Patenten geschützt werden). Der Zugriff erfolgt über Dialogsysteme oder die die Questel-eigenen Mehrwertdienste QPat[42] und PatentExaminer[43].

- Die Inhalte der USPTO sind inzwischen auch über die Google-Funktion Google Patent Search[44] recherchierbar. Google hat nach eigenen Angaben ca. 7 Millionen Patente aus dem USPTO-Datenbestand übernommen und mittels der für die Google-Buchsuche entwickelten Technologie in ein besser Browser-unterstütztes Format konvertiert. Abbildung 24 zeigt einen Screenshot einer Patentrecherche mit Google Patent Search.

35 http://www.thomsoninnovation.com
36 http://www.delphion.com
37 http://www.dialog.com
38 http://www.questel.orbit.com
39 http://www.stn-international.de/index.php?id=123
40 http://www.questel.com/Prodsandservices/pluspat.htm
41 http://www.questel.orbit.com/Prodsandservices/FamPat.htm
42 http://www.qpat.com
43 http://www.patexaminer.com
44 http://www.google.com/patents

Abbildung 24:
Screenshot Google
Patent Search



Die genannten Produkte bieten neben ihrer Hauptfunktion als Metasuchmaschine für die verschiedenen nationalen Patentdatenbanken weitere wichtige Mehrwertfunktionen wie etwa Analyseunterstützung und Report-Erzeugung bieten. Wanner et al. (2008) identifizieren neben der Suche nach Patenten zu einem bestimmten Thema die folgenden von solchen Werkzeugen angebotenen »primären« Patentverarbeitungsdienste[45]:

1. Zusammenfassen: Die wichtigsten Informationen aus einem Patentartikel werden zusammengefasst und in eine homogene Struktur gebracht.

2. Klassifizieren und Clustern: Große Mengen von Daten müssen einheitlich strukturiert werden. Neben der Organisation der Patente nach Standard-Klassifikationsschemata wie der »International Patent Classification« (IPC[46]), dem »European Patent Classification System«

[45] Wanner et al. (2008) benennen zudem die linguistische Verarbeitung als weiteren Dienst. Da dieser dem Benutzer jedoch nur mittelbar zur Patentanalyse dient (er stellt einen Basisdienst für die anderen Anwendungsfälle dar), soll er an dieser Stelle vernachlässigt werden.

[46] http://www.wipo.int/classifications/ipc

(ECLA[47]), der US Patent Classification (USPC[48]) oder dem Derwent World Patent Index Classification Schema (DWPIC[49]) werden mittels statistischer Verfahren »ähnliche« Patente zu Gruppen zusammengefasst.

3. Übersetzen: Patente werden in unterschiedlichen Sprachen veröffentlicht. Daher bieten einige Anbieter manuell oder automatisch erstellte Übersetzungen an.

4. Analysieren von Meta-Informationen: Wie schon im Abschnitt zu Pressemitteilungen diskutiert, lassen sich auch die in Patentdokumenten und –kollektionen enthaltenen Metadaten für interessante Analysen instrumentieren. So erlaubt etwa die Analyse der Anzahl der Patenteinreichungen die Beobachtung der Entwicklung eines Themenfeldes.

Diese Dienste werden mittels automatisierter Verfahren erbracht oder auch manuell erbracht. Die Firma Thomson beispielsweise beschäftigt eigenen Angaben zufolge 350 Experten, die sich der Aufbereitung von Patentinformationen für eine optimierte Nutzung mittels Suchwerkzeugen widmen und dabei etwa die Haupt-Claims in homogener Sprache zusammenfassen.

2.2.4.3 Anwendungsbeispiel: White-Spot-Analyse

Mithilfe von White-Spot-Analysen lassen sich Hinweise auf Marktpotenziale ableiten. Die Grundannahme besteht darin, dass diejenigen Aspekte eines Technologiefeldes, die bislang wenig durch Patente geschützt sind, höheres Potenzial für den wirtschaftlichen Erfolg zukünftiger Produkte bietet. Die White-Spot-Analyse-Methodik des Fraunhofer IAO gliedert sich dabei in zwei wesentliche Schritte (vgl. Flussdiagramm in Abbildung 25):

1. Zunächst wird die Patentabdeckung eines Technologiefeldes manuell und computerunterstützt[50] ermittelt. Dabei werden Problem-Lösungs-Zusammenhänge aus Patenttexten extrahiert und in einer Matrix zusammengefasst.

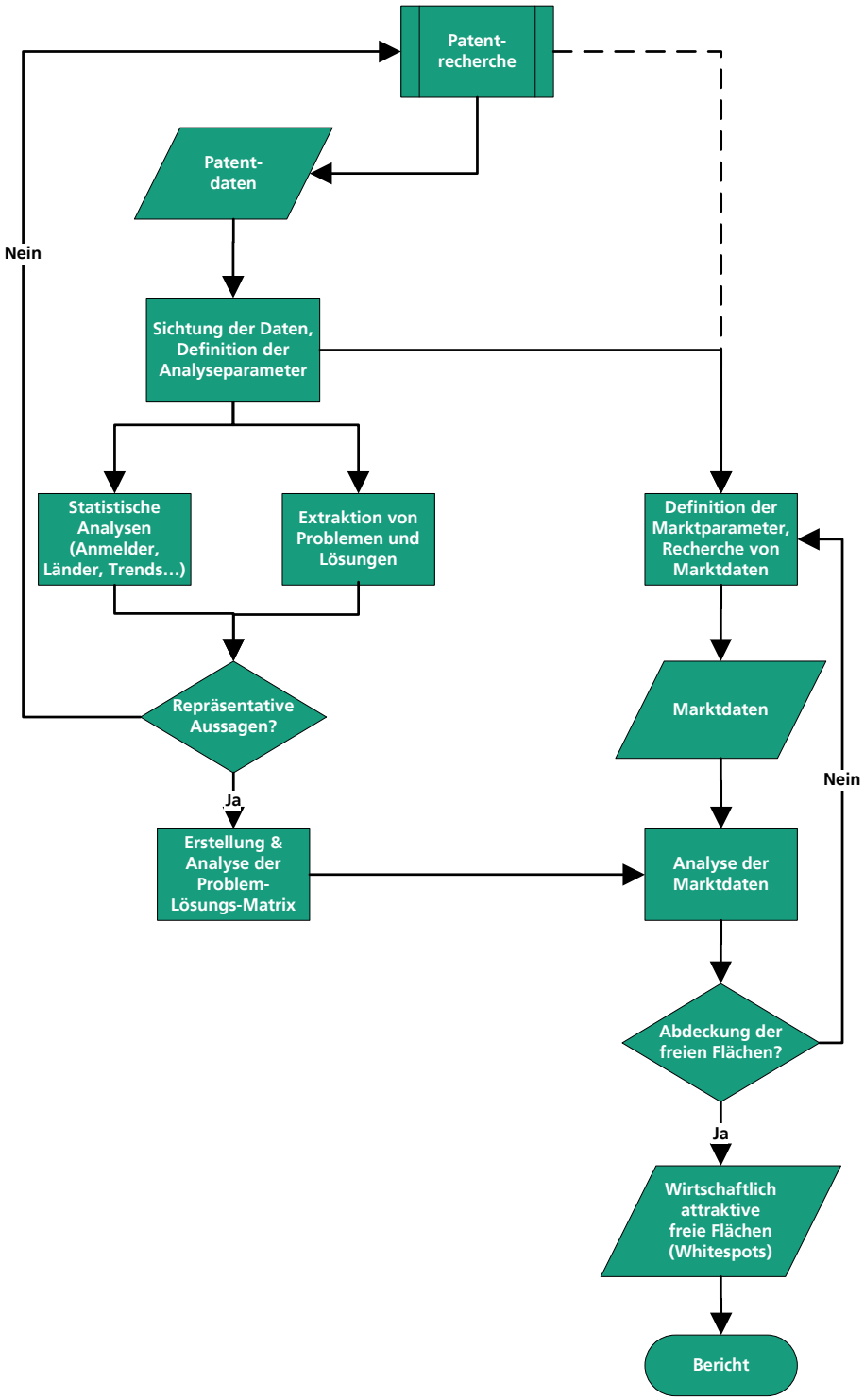
[47] http://v3.espacenet.com/eclasrch?&locale=de_ep&classification=ecla
[48] http://www.uspto.gov/web/patents/classification/
[49] http://science.thomsonreuters.com/support/patents/dwpiref/reftools/classification/
[50] Für die computergestützte Extraktion der Problem-Lösungszusammenhänge lässt sich beispielsweise die Text Mining-Software Luxid (http://www.temis.com/?id=201&selt=1) verwenden, die speziell zu diesem Zweck um ein entsprechendes Plug-in (»Skill Cartridge«) erweitert wurde.

2. In einem zweiten Schritt wird die wirtschaftliche Attraktivität der identifizierten White Spots des Technologiefeldes bewertet. Hierzu werden zum einen die jeweiligen unternehmensspezifischen Parameter berücksichtigt, zum anderen erfolgt eine Web-basierte Recherche zusätzlicher Marktdaten.

Beide Schritte werden üblicherweise mehrfach durchlaufen (vgl. die rückgerichteten Pfeile in Abbildung 25). Ein detailliertes Beispiel der Anwendung der Methodik liefert Siwczyk (2009), die White Spots im Bereich der Elektromobilität identifiziert.

Abbildung 25:
White Spot-Analyse-Prozess des Fraunhofer IAO.

Darstellung basierend auf (Siwczyk, 2009)



2.2.4.4 Weiterführende Literatur

Shih, M. J., Liu, D. R und Hsu, M. L.: Discovering competitive intelligence by mining changes in patent trends. In: Expert Systems with Applications, 37(4). 4, April 2010, S. 2882-2890.
Wissenschaftlicher Aufsatz, der eine Methodik zur Analyse von Trends in Patentdaten zum Gegenstand hat.

Siwczyk, Y. (2009): IT-gestützte White-Spot-Analyse: Potenziale von Patentinformationen am Beispiel Elektromobilität erkennen. Fraunhofer Verlag. 2009.
Veranschaulicht die Fraunhofer-Methodik der Patentdaten-White Spot-Analyse anhand des Beispiels der Elektromobilität..

Wanner, L., Baeza-Yates, R., Brügmann, S., Codina, J., Diallo, B., Escorsa, E. et al. (2008): Towards content-oriented patent document processing. In: World Patent Information, Jg. 30, H. 1, S. 21–33.
Wissenschaftlicher Aufsatz über das europäische Forschungsprojekt »PATExpert«, das sich vorrangig mit der semantischen Analyse von Patentinhalten beschäftigt.

Yang, Y., Akers, L., Klose, T. und Yang, C. B. (2008): Text mining and visualization tools - Impressions of emerging capabilities, World Patent Information, Volume 30, Issue 4, December 2008, S. 280-293.
Zusammenfassung von Ergebnissen eines Vergleichs von Textmining- und Visualisierungswerkzeugen bzgl. der Eignung für Innovationsmanagement und Wettbewerbsbeobachtung der Patentanalyse-Gruppe des Pharma-Unternehmens Bristol-Myers Squibb.

3 Neue Möglichkeiten und Herausforderungen im Social Web

Das Kommunikationsmodell im Internet hat sich verändert. Kommunikation erfolgt nicht mehr ausschließlich in eine Richtung über die offiziellen Kanäle von Unternehmen über Massenmedien hin zum Kunden, sondern im Web 2.0 vor allem auch zwischen den Kunden, sowie direkt zwischen Kunden und Unternehmen (Skibicki et al., 2008). Dabei entstehen nutzergenerierte Informationen, die für jeden im Netz auffindbar und zugänglich sind.

Der Begriff »Social Web« wird heute synonym mit dem Begriff Web 2.0 verwendet und betont die Möglichkeiten der Interaktion der Nutzer. Typische Plattformen hierfür sind Wikis, Foren, Blogs und Soziale Netzwerke. Im Abschnitt 3.2 gehen wir genauer auf diese Informationsarten ein.

Die in Abbildung 26 und Abbildung 27 dargestellten Ergebnisse einer Untersuchung von Fittkau & Maaß[51] unterstreichen, dass Verbraucherkritiken und Meinungen im Internet von Unternehmen als besonders wichtig eingeschätzt werden müssen, weil sie von mehr als 87 Prozent der Internetnutzer zumindest gelegentlich genutzt werden und die Nutzer sich dadurch in ihrer Meinung und konkreten Kaufentscheidung beeinflussen lassen.

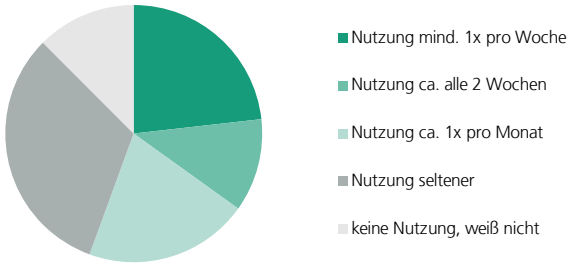


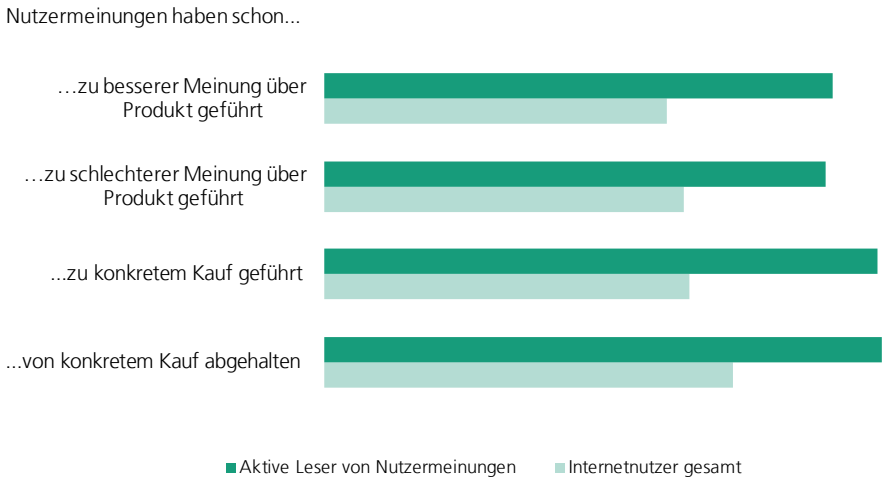
Abbildung 26: Nutzung von Produktbewertungen anderer Nutzer

Darstellung basierend auf: 29. WWW-Benutzer-Analyse W3B, Herbst 2009, Fittkau & Maaß Consulting

[51] http://www.w3b.org/trends

Abbildung 27:
Auswirkungen von Nutzermeinungen

Darstellung basierend auf 29. WWW-Benutzer-Analyse W3B, Herbst 2009, Fittkau & Maaß Consulting



Besonders herausfordernd beim Umgang mit nutzergenerierten Informationen ist, dass diese nicht den Regeln von Unternehmenskommunikation und klassischem Journalismus folgen beziehungsweise verpflichtet sind: Ausführliche Recherche, Objektivität und allgemein Korrektheit sind für Internetautoren keine Voraussetzung zur Publikation. Meinungen werden nicht immer als solche gekennzeichnet. Außerdem sind Masse und Zeit kritische Faktoren: Im Social Web entsteht eine große Menge an Inhalten, die von einer großen Anzahl an Autoren vervielfältigt und von einer noch größeren Anzahl an Lesern kommentiert werden. Das alles passiert in rasender Geschwindigkeit. Der Ruf eines Unternehmens als Ganzes oder einzelner Produkte und Dienstleistungen wird zunehmend auch durch nutzergenerierte Inhalte bestimmt. Die Informationen und deren Auswirkungen können dabei sowohl positiv als auch negativ sein. Den Aspekt des strategischen Umgangs mit kritischen Verbraucheräußerungen vertiefen wir in Abschnitt 3.4.

Die Masse der im Internet aktiven Nutzer bietet gleichzeitig auch Chancen für Unternehmen. Im Buch »The Wisdom of Crowds« (Surowiecki, 2004) wird beschrieben, wie Gruppen gemeinsame Entscheidungen treffen, die besser sind, als Lösungsansätze einzelner Teilnehmer. Eine Strategie des Outsourcings interner Entwicklungsarbeiten an eine Internet-Community wird oft als »Crowdsourcing« bezeichnet und derzeit vor allem in den frühen Phasen der Produkt- und Dienstleistungsentwicklung eingesetzt. Neben der Nutzung unternehmensexterner Kreativität können mittels Crowdsourcing auch positive Effekte auf Kundenbindung und Marketing erzielt werden. Ein typisches

Beispiel stellt der effektive Einsatz von Open Innovation-Portalen dar, auf welche wir im Abschnitt 3.3 detailliert eingehen.

3.1 Die Nutzer des Social Web

Die Beteiligung des Einzelnen im Social Web ist sehr unterschiedlich. Viele Internetnutzer sind bereits mit benutzergenerierten Inhalten konfrontiert gewesen, der Anteil derer, die Inhalte schaffen, ist jedoch vergleichsweise gering. Forrester Research unterscheidet in »Social Technographics« verschiedene Nutzerrollen aufgrund der von ihnen ausgeübten Aktivitäten und erhebt, welcher Anteil der erwachsenen Bevölkerung zu den jeweiligen Gruppen gehört. Die Nutzerrollen überlappen dabei: Beispielsweise ist jemand, der einen Blog schreibt, gleichzeitig auch regelmäßiger Leser anderer Blogs. Ein Internetnutzer zählt zu einer Nutzerrolle, wenn er mindestens einmal im Monat typische Aktivitäten dieser Rolle ausführt. Tabelle 6 stellt die Social Technographics für Deutschland und die USA einander gegenüber. Man erkennt, dass Deutsche das Social Web bisher in allen Nutzergruppen lediglich halb so oft nutzen wie US-Amerikaner, jedoch muss man analog zu den USA auch in Deutschland von einem steilen Aufwärtstrend ausgehen. Im zweiten Quartal 2007 lag der Prozentsatz der nicht im Social Web aktiven Nutzer auch in den USA noch bei 44 Prozent (Li et al., 2008).

Tabelle 6: Nutzerrollen im Social Web

Darstellung basierend auf Forrester Researchs North American Technographics® Interactive Marketing Online Survey, Q2 2009 sowie European Technographics® Benchmark Survey, Q2 2009

Nutzerrolle	% US *	% DE *	Aktivitäten
Schöpfer (Creators)	24	9	Veröffentlichen eigene Artikel in Blogs, Webseiten oder laden Multimediainhalte wie Videos und Podcasts auf entsprechende Plattformen.
Kritiker (Critics)	37	12	Bewerten Produkte und Dienstleistungen in Bewertungsportalen, kommentieren Inhalte in Blogs, beteiligen sich an Foren und Wikis.
Sammler (Collectors)	20	4	Akkumulieren Inhalte, indem sie RSS-Feeds abonnieren und Social-Bookmarking-Applikationen einsetzen.
Teilnehmer (Joiners)	59	21	Sind in mindests einem Sozialen Netzwerken angemeldet und besuchen diese Seiten.
Zuschauer (Spectators)	70	38	Konsumieren Inhalte in Blogs, Foren und auf Bewertungsseiten.
Inaktive (Inactives)	17	54	Sind nicht im Social Web aktiv.

* Prozentsatz der Erwachsenen, die mindestens einmal monatlich in diese Rolle schlüpfen

Im Januar 2010 wurde in das Social Technographics Nutzermodell zusätzlich zu den in der Tabelle definierten Rollen der »Plauderer« (Conversationalist) eingefügt[52]. 33 Prozent der erwachsenen US-Bürger gehören laut den Erhebungen von Forrester Research zu dieser Gruppe, das heißt sie befinden sich mindestens wöchentlich auf Plattformen wie Facebook und Twitter, um dort ihren Status zu ändern beziehungsweise Nachrichten zu schreiben.

3.2 Informationsarten im Social Web

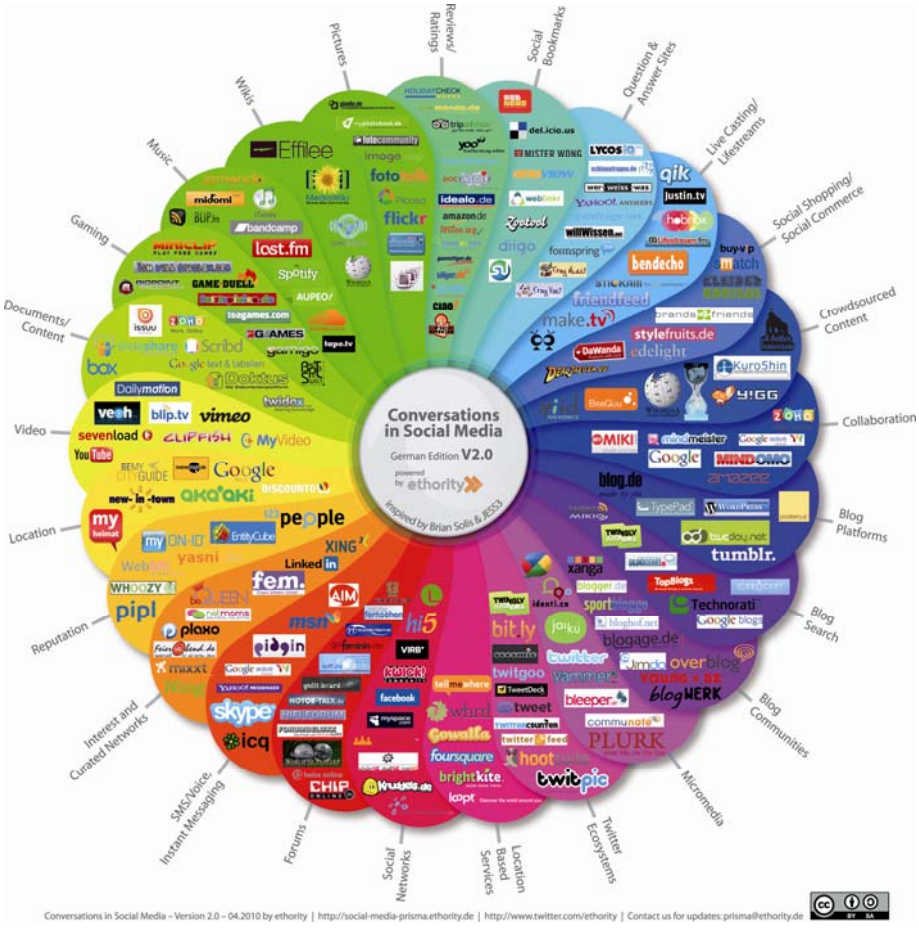


Abbildung 28: Social Media Prisma

Darstellung übernommen von ethority GmbH & Co. KG

Die Inhalte im Social Web sind sehr heterogen, und eine Klassifizierung kann anhand verschiedenster Kriterien erfolgen. Es gibt Plattformen, die sich auf bestimmte Medien-Formate wie Bilder, Video, und Musik spezialisieren, andere

[52] http://forrester.typepad.com/groundswell/2010/01/conversationalists-get-onto-the-ladder.html

wiederum definieren sich über ihre Funktionen wie Kontaktpflege, Bewertung oder Information. Meist lassen sich jedoch weder Medien noch Funktionen eindeutig zuordnen, und auch die Grenze zwischen nutzergenerierten Inhalten und redaktionellen Inhalten verwischt zunehmend.

Einen weder vollständigen noch allgemeingültigen aber trotzdem wertvollen Überblick verschafft das »Social Media Prisma«, welches ursprünglich von Brian Solis und Jess Thomas[53] entwickelt wurde. Die ethority GmbH & Co. KG hat basierend darauf die in Abbildung 28 dargestellte Grafik angefertigt und darin auch deutsche Plattformen berücksichtigt. Größe und Wichtigkeit der einzelnen Dienste können aus dieser Grafik nicht abgeschätzt werden, und auch die Zuordnung ist nicht immer eindeutig – so gehört Wikipedia sowohl zu Crowdsourcing als auch zur Kategorie Wikis, und darüber ob Apples iTunes überhaupt zu Social Media zu zählen ist, lässt sich streiten. Tabelle 7 enthält eine Kategorisierung nach Informationsarten, wie sie im Folgenden auch besprochen wird.

Tabelle 7: Informationsarten im Social Web

In manchen Kategorien gibt es unzählige Beispiele, hier ist jeweils nur ein Beispiel exemplarisch aufgeführt.

Kategorie	Unterkategorie	Beispiel	URL
Medienplattformen	Video	YouTube	http://www.youtube.com
	Bilder	flickr	http://www.flickr.com
	Dokumente	Slideshare	http://www.slideshare.net
	Musik	last.fm	http://www.lastfm.de
Meinungsäußerung	Foren	Forum Deutsches Recht	http://www.recht.de
	Blogs	netzpolitik	http://www.netzpolitik.org
	Bewertungsportale	ciao!	http://www.ciao.de
	Frage & Antwortseiten	wer-weiss-was	http://www.wer-weiss-was.de
Soziale Netzwerke	privat	Facebook	http://www.facebook.com
	beruflich	Xing	http://www.xing.de
	Social Commerce	DaWanda	http://www.dawanda.de
Kollaboration	Wikis	Wikipedia	http://www.wikipedia.de
	Social Bookmarking	Mister Wong	http://www.mister-wong.de
Microblogging		Twitter	http://www.twitter.com
Ortsbezogene Dienste		foursquare	http://www.foursquare.com
Instant Messaging		skype	http://www.skype.de
Unterhaltung	Virtuelle Welten	Second Life	http://www.secondlife.com
	Social Games	Sevengames	http://www.sevengames.de

Im Allgemeinen handelt es sich bei den in Tabelle 7 enthaltenen Anwendungen des Social Web um kostenlos verfügbare Dienste, die sich primär durch Werbung finanzieren. Kostenpflichtige Premiummitgliedschaften bieten auf manchen Plattformen zusätzliche Funktionalitäten.

[53] http://theconversationprism.com/

Charakteristisch für Informationen im Social Web sind:

- der stetige Wandel, sowie die Kombination und Vernetzung der Anwendungen untereinander
- der persönliche Bezug der Informationen
- die Geschwindigkeit, in der sich Informationen verbreiten und die insgesamt generierte Masse an Informationen

Für benutzergenerierte Multimedia-Inhalte wie Bilder und Videos gibt es separate Multimediaplattformen im Social Web. Vor allem die Videoplattform YouTube ist mit zwei Milliarden Aufrufen pro Tag[54] sehr erfolgreich und wird inzwischen auch von Unternehmen mit eigenen Channels[55] bedient. Auf Funktionsweise und Besonderheiten von Plattformen zur Meinungsäußerung, Kollaboration und auf soziale Netzwerke, wird in den nächsten Abschnitten genauer eingegangen. Eine ausführliche Diskussion zu Open Innovation-Plattformen im Allgemeinen und webbasierten Ideenportalen im Besonderen im Kontext des Innovation Minings findet sich in Abschnitt 3.3.

3.2.1 Meinungsäußerung

Die Meinungsäußerung ist ein elementarer Bestandteil des Social Webs – und spielt in vielen Ausprägungen sozialer Software eine große Rolle: Sie erfolgt in Form von Textbeiträgen in Diskussionsforen und Blogs oder in der Bewertung von Produkten, Artikeln oder Beiträgen anderer Benutzer. Kommentare sind oft subjektiv, unstrukturiert und der Einstiegspunkt in eine aktive Teilnahme am Social Web. Für Unternehmen können Kommentare und Bewertungen eine nützliche Informationsquelle darstellen, denn sie können Aufschluss über Zufriedenheit und Unmut (potenzieller) Kunden geben.

Diskussionsforen sind eine der ältesten Formen von Kooperation im Internet[56] – es handelt sich dabei um Web-Applikationen, die den hauptsächlich textbasierten Austausch von Nachrichten ermöglichen. Eine Nachricht, die in einem Forum geschrieben wird, kann von allen Internetnutzern gelesen

[54] laut eigener Aussage im YouTube-Blog aufgerufen am 17. Mai 2010 http://youtube-global.blogspot.com/2010/05/at-five-years-two-billion-views-per-day.htm

[55] Ein Channel ist die Präsenz eines einzelnen Nutzers innerhalb von YouTube. Er führt dessen Profildaten, Inhalte, Kommentare etc. zusammen und kann individuell angepasst werden.

[56] Der »Urvater« der heute populären Webforen ist das »Usenet«, dessen Wurzeln bis ins Jahr 1979 zurückreichen. Die Teilnahme an den »Newsgroups« des Usenet setzt - im Gegensatz zu den meisten Webforen – keine Registrierung voraus. Allerdings wird eine spezielle Software (»Newsreader«) benötigt.

werden. Andere (registrierte[57]) Nutzer können auf diese Nachricht antworten – es entsteht ein Nachrichtenbaum - ein sogenannter »Thread« – zu einem bestimmten Thema. Die wohl häufigste Art von Foren befasst sich mit technischen Fragestellungen, es gibt aber inzwischen Foren zu jedem denkbaren Thema.

Supportforen

Häufig werden Foren eigens von Herstellern aufgesetzt, um eine Gemeinschaft zu schaffen, in der sich Kunden gegenseitig unterstützen. Antworten in den Foren werden oft von nur wenigen Experten verfasst. Diese werden auch als »Influencer« bezeichnet, da sie einen großen Einfluss auf die Meinungen der Leser ausüben (können).

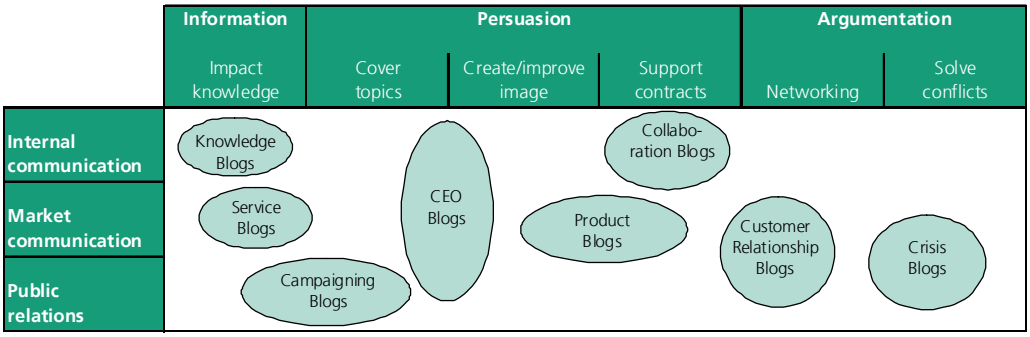
Weblogs

Weblogs (»Blogs«) sind ein wichtiger Bestandteil des Web 2.0-Paradigmas, dessen Essenz darin liegt, den Benutzer in den Mittelpunkt zu rücken. Weblogs sind sozusagen öffentliche elektronische Tagebücher, in denen einzelne Personen ihre persönliche Meinung zu unterschiedlichsten Themen kundtun. Für Innovation Mining-Aktivitäten sind Blogs genauso wie Foren deshalb interessant, weil Informationen und Meinungen hier besonders schnell und direkt veröffentlicht werden. Unternehmen können die hier veröffentlichten Texte nutzen, um Feedback zu ihren Produkten und Dienstleistungen jenseits des üblichen Feedback- und Beschwerdemanagements einzusammeln. Viele Blogs widmen sich einem bestimmten Thema oder einer bestimmten Technologie – sie sind damit ein potenzieller Ansatzpunkt für Lead-User-Identifikation und –Einbindung, zur Trendidentifikation und zur Auffindung neuer Ideen. Außerdem nutzen Unternehmen Blogs als Ergänzung ihrer Internetauftritte zur Bereitstellung aktueller Information. Weitere Anwendungsmöglichkeiten im Unternehmen sind in Abbildung 29 dargestellt.

[57] Die meisten Foren erfordern zum Schreiben bzw. Beantworten von Nachrichten eine Registrierung. Das Lesen von Nachrichten ist im Allgemeinen für alle Internetbenutzer möglich.

Abbildung 29: Anwendungsbereiche von Blogs in Unternehmen.

Darstellung basierend auf Zerfass (2005)



Bewertungsportale

In Bewertungsportalen veröffentlichen Nutzer ihre eigenen Erfahrungen mit Produkten und Dienstleistungen. Anwendung finden verschiedene Bewertungssysteme von quantitativ auswertbaren Benotungen bis hin zu ausführlichen Textbeschreibungen. Die Bewertungen und Autoren können manchmal wiederum selbst bewertet werden, was die Qualität der Informationen steigert. In Bewertungsportalen erkennt man auch die Selbstheilungseffekte in Gemeinschaften: Durch die Reaktionen anderer Benutzer werden subjektive Meinungen relativiert, und es entsteht ein objektives Bild.

3.2.2 Soziale Netzwerke

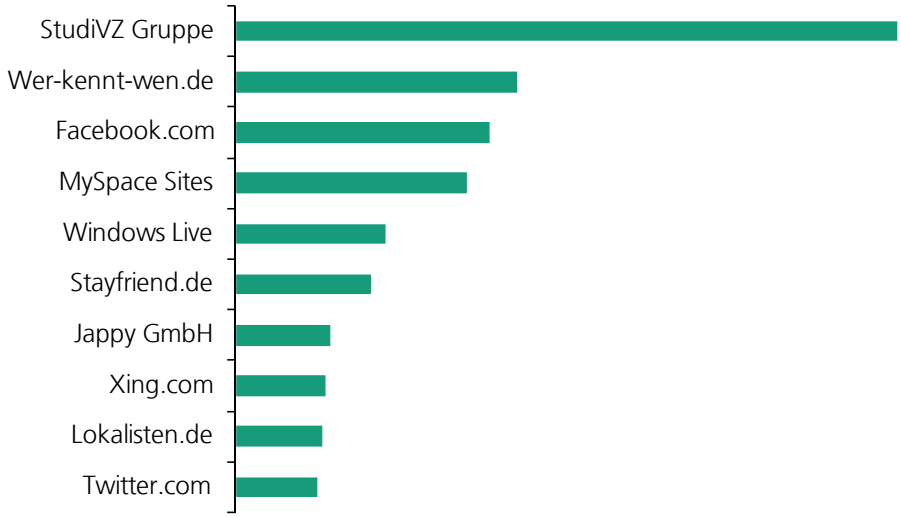
Soziale Netzwerke verzeichnen in Deutschland einen enormen Zuwachs: Im Mai 2010 gab es in Deutschland bereits fast 10 Millionen auf Facebook aktive Nutzer[58].

Abbildung 30 zeigt die im dritten Quartal 2009 erhobenen Nutzerzahlen der meistgenutzten sozialen Netzwerke in Deutschland.

[58] http://www.checkfacebook.com/

Abbildung 30:
Soziale Netzwerke in Deutschland

Darstellung basierend auf Comscore Studie für BITKOM (2009), Angaben in Millionen Nutzer.



In sozialen Netzwerken werden Nutzer durch die Erstellung eines Profils Mitglieder einer Gemeinschaft in der sie sich dann mit anderen Mitgliedern direkt verbinden. Kontakte in der realen Welt werden in der virtuellen Welt erneut geknüpft beziehungsweise bestätigt. Hinzu kommen Kontakte, die allein innerhalb der sozialen Netzwerke entstehen und bestehen. Typische Funktionen solcher Netzwerke sind die Veröffentlichung von Statusmeldungen, das Versenden von Nachrichten und das Gründen von Gruppen. Hinzu kommen immer mehr kleine Anwendungen, in denen die Benutzer miteinander interagieren können oder durch die sie ihr eigenes Profil erweitern und detaillieren können.

Ursprünglich wurden soziale Netzwerke ausschließlich im privaten Bereich genutzt, inzwischen sind jedoch auch Unternehmen mit Profilen und Fanseiten dort vertreten. Soziale Netzwerke werden zu einer digitalen Abbildung des Lebens des Nutzers. Das digitale Profil des Nutzers spielt beim Geschäftsmodell der meist kostenlosen Dienste eine wichtige Rolle. Profildaten wie Alter, Geschlecht, Wohnort, Beziehungsstatus werden von zahlenden Unternehmen genutzt, um Online-Werbung und Angebote in sozialen Netzwerken individuell auf den Kunden abzustimmen. Konsequenterweise wird daher dem Thema Datenschutz eine zunehmend wichtige Rolle beigemessen.

Eine Trennung von privaten und geschäftlichen Kontakten erfolgt in Business-Netzwerken wie LinkedIn[59] oder Xing[60]. Diese werden über alle Altersklassen hinweg genutzt (vgl. Studie von careerbuilder.de[61]). Für Unternehmen sind Business-Netzwerke besonders zur Rekrutierung von Personal und Experten interessant.

3.2.3 Microblogging

Den kostenlosen Microblogging-Dienst Twitter[62] gibt es im Social Web seit 2006. Angemeldete Benutzer können damit Textnachrichten von 140 Zeichen Länge schreiben (»twittern«). Jeder Nutzer empfängt auf seiner eigenen wiederum öffentlichen Seite Nachrichten (»Tweets«) von Nutzern und Gruppen, die er abonniert hat. Im Umfeld von Twitter nennt man dieses Abonnieren folgen (»follow«). Durch die Anzahl derer, die einen Twitter-Nutzer verfolgen und die Anzahl der Nutzer, denen er folgt, sowie die Anzahl der Nachrichten, die ein Nutzer verschickt, wird dessen Einflussbereich bestimmt. Neben privaten Inhalten werden über Twitter derzeit vor allem Links zu anderen Webseiten weitergegeben, solche Inhalte werden von anderen Nutzern oft wiederholt und weiterverbreitet (»Retweet«). Unternehmen setzen Twitter auf unterschiedliche Art und Weise ein, zum Beispiel zum Verbreiten aktueller Nachrichten, Marketing etc. Twitter selbst hat noch kein Monetarisierungskonzept implementiert.

Es gibt inzwischen zahlreiche Twitter umgebende Dienste, die zum Beispiel URLs verkürzen, das Einbinden von Bildern ermöglichen oder Twitter-Statistiken erstellen. Außerdem entstehen jenseits von Twitter weitere interessante Dienste mit Fokus auf dem Austausch von kurzen persönlichen Informationen. Beispiele dafür sind Plurk[63], FriendFeed[64] und tumblr[65].

3.2.4 Kollaboration

Das Social Web fördert nicht nur den Gedanken- und Informationsaustausch der Nutzer untereinander, sondern bietet noch weitere mächtige Möglichkeiten der Zusammenarbeit.

[59] http://www.linkedin.com/
[60] http://www.xing.com/
[61] http://facebookmarketing.de/zahlen_fakten/xing-fb
[62] http://www.twitter.com
[63] http://www.plurk.com
[64] http://friendfeed.com/
[65] http://www.tumblr.com

Wikis

Bei Wikis handelt es sich um Hypertextsysteme für Webseiten, deren Inhalte von Benutzern nicht nur gelesen, sondern auch online direkt im Browser geändert werden können. Unternehmen setzen Wikis vor allem im Intranet als Wissensmanagement-Werkzeug ein.

Die wohl bekannteste und erfolgreichste Anwendung eines Wikis ist die Online-Enzyklopädie Wikipedia[66]. Wikipedia ist auch ein Beispiel für Crowdsourcing: die Enzyklopädie wurde unentgeltlich durch Nutzer der Internetgemeinschaft erstellt und wird fortwährend erweitert. Dabei bietet sie zumeist neutrale und umfassende Inhalte zu fast jedem Thema und inzwischen in über 260 Sprachen. Statt einer Redaktion korrigieren, kontrollieren und regulieren sich die Benutzer gegenseitig. Kontrollmechanismen erlauben beispielsweise nicht, dass strittige Artikel von neuen Benutzern überarbeitet werden. Eine repräsentative Umfrage unter deutschen Studierenden[67] ergab, dass es sich um die am häufigsten genutzte Internetseite handelt und dass 52 Prozent der Studierenden die Informationen in Wikipedia für verlässlich halten. Alle anderen Online-Wissensplattformen genießen weniger Akzeptanz und Vertrauen. Auch viele Web Intelligence-Werkzeuge (siehe nächstes Kapitel) nutzen Wikipedia inzwischen als grundlegende Informationsbasis: Die Verlinkungen zwischen den einzelnen Artikeln sowie die in Wikipedia enthaltenen strukturierten Informationen (Kategorien, geographische Koordinaten etc.) können z. B. für die Erstellung von Begriffsnetzen und Klassifikationen herangezogen werden. Insbesondere im DBpedia-Projekt[68] werden die Inhalte von Wikipedia in Semantic Web-konforme Strukturen überführt und mit weiteren frei verfügbaren Datensammlungen verknüpft. Dies erlaubt dann z. B. in Verbindung mit Werkzeugen wie der »Faceted Wikipedia Search« der Firma Neophonie[69] strukturierte Suchanfragen der Form »Zeige alle Flüsse, die in den Rhein münden und deren Länge mehr als 50 km beträgt:«[70].

Social Bookmarking und Folksonomies

»Social Bookmarks« sind über das Internet nutzbare Lesezeichen, die bei einem zentralen Bookmarking-Dienst wie Delicous[71] oder Mr. Wong[72] gehostet werden. Alle im Browser aufrufbaren Quellen können vom Benutzer individuell

66 http://www.wikipedia.org
67 Kleimann, Özkilic, Göcks - HISBUS Studieren im Web 2.0 (November 2008)
68 http://wiki.dbpedia.org
69 http://www.neofonie.de
70 http://dbpedia.neofonie.de/browse/rdf-type:River/River-mouth:Rhine/Place-length~:50000~/?fc=30

71 http://delicious.com
72 http://www.mister-wong.de

mit Zusatzinformationen versehen, kategorisiert bzw. mit »Tags« annotiert und bewertet werden. Durch die Zusammenfassung der Bookmark- und Tag-Sammlungen mehrerer Benutzer entstehen sogenannte »Folksonomies«, also kollaborativ erstellte Taxonomien.

Das Setzen eines Bookmarks kann als positive Bewertung eines Seiteninhalts interpretiert werden, da es zumindest ein gewisses Interesse des Benutzers signalisiert. Außerdem finden sich in der jeweiligen Kategorie eventuell weitere interessante und ähnliche Inhalte. In Social Bookmarking-Diensten erstellen viele Nutzer gemeinsam Rangfolgen für Informationen, die nicht notwendigerweise dem Ranking der großen Suchmaschinen entsprechen.

User Innovation

Im Kontext des Innovationsmanagements sind »User Innovation Toolkits« schon lange bekannt (vgl. von Hippel, 1998). Durch die neuen Paradigmenwechsel des Internets zum Web 2.0 und Social Web haben diese stark an Relevanz gewonnen – die Beteiligung von Nutzern am Innovationsprozess ist inzwischen weit verbreitet. Im nächsten Abschnitt gehen wir darum detailliert auf das Thema Open Innovation und insbesondere die internetgestützte Ideensammlung ein.

3.2.5 Weiterführende Literatur

Richter, A. und Koch, M. (2007): Social Software – Status quo und Zukunft. Technischer Bericht Nr. 2007-01. Fakultät für Informatik, Universität der Bundeswehr München.
Wissenschaftliche Betrachtung zu Web 2.0, seine Benutzer und Plattformen.

Skibicki, K. und Mühlenbeck, F. (2008): Verbrauchermacht im Internet. 1. Auflage BrunoMedia Buchverlag, ISBN 978-3981150681.
Ein aus Verbrauchersicht geschriebenes Buch darüber, Welche Möglichkeiten es gibt, Meinungen- und Bewertungen im Internet auszutauschen.

Surowiecki, J. (2005): The wisdom of crowds. 1. ed. New York, NY: Anchor Books.
Standardwerk zum Thema Crowdsourcing. Deutscher Titel »Die Weisheit der Vielen: Warum Gruppen klüger sind als Einzelne«.

Tapscott, D und Williams, A. D. (2007): Wikinomics: How Mass Collaboration Changes Everything. How Mass Collaboration Changes Everything: Portfolio.
Beschreibung der Kollaborationsmöglichkeiten im Web und deren wirtschaftlichem Potenzial.

Weinberg, T. (2010) Social Media Marketing: Strategien für Twitter, Facebook & Co. Auflage 1, O'Reilly. ISBN 978-3897219694.
Einstiegsliteratur für Unternehmen ins Thema Social Media und dessen Möglichkeiten im Marketing. Inklusive Beschreibung der verschiedenen Social Media-Plattformen und mit vielen Anwendungsbeispielen.

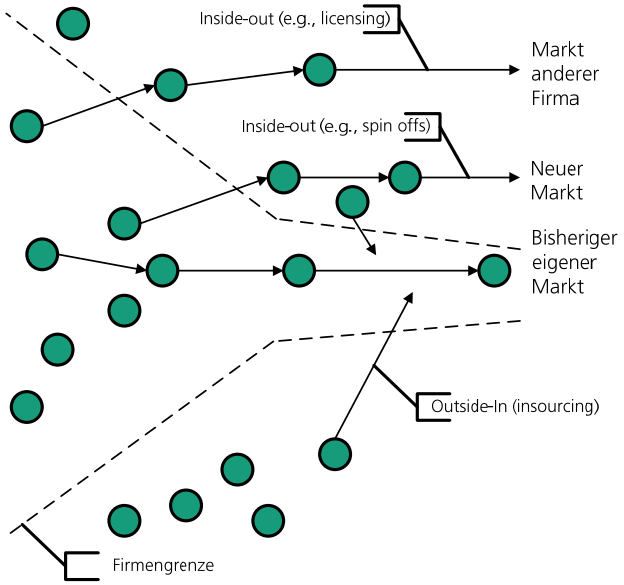
3.3 Open Innovation

Globalisierung und Fortschritt der IuK-Technologien führen seit Längerem zu einer Abflachung und Auflösung hierarchischer Unternehmensstrukturen zugunsten von Netzwerkorganisationen und einer Abwicklung auf Märkten. Diese strukturellen Änderungen wirken sich auch und insbesondere auf den Innovationsprozess von Unternehmen aus: Kunden (sowohl Endkunden als auch Mediatoren) nehmen eine zunehmend aktive Rolle bei der Gestaltung von Produkten und Dienstleistungen ein.

Der Begriff »Open Innovation« ist ein derzeit stark diskutierter Forschungsgegenstand (vgl. Chesbrough, 2006). Das Einbeziehen von Akteuren außerhalb des eigenen Unternehmens in den eigenen Innovationsprozess kann das »Floprisiko« bei der Entwicklung neuer Produkte und Dienstleistung und bei der Weiterentwicklung bestehender Angebote senken und somit viel Geld sparen.

Abbildung 31: Open Innovation-Paradigma

Darstellung basierend auf (Chesbrough, 2006)



Die Hauptansätze des Open Innovation-Paradigma liegen im »Inside-out «- und »Outside-in«-Prozess (Gassman und Enkel, 2006):

1. Durch den Inside-out-Prozess sollen Wissen und innovative Ideen, die – aus welchem Grund auch immer – von der Firma nicht genutzt werden, gewinnbringend verwendet werden (z. B. durch Gründung von Technologie-Spin-offs oder durch Lizenzierung an andere Firmen).

2. Der Outside-in-Prozess erlaubt es Firmen, die nicht über genug eigene Kapazität zur Entwicklung innovativer Konzepte verfügen, entsprechendes Know-how durch Dienstleistungen oder Lizenzen einzukaufen.

Natürlich lassen sich beide Ansätze kombinieren, was im Ergebnis dazu führt, dass Firmen ihr eigenes Wissen gewinnbringend externalisieren und gleichzeitig externes Wissen einholen. Abbildung 31 skizziert die Ansätze.

Im Innovation Mining ist insbesondere der Outside-in-Ansatz von Interesse. Die Integration von Ideen externer Partner und insbesondere von Kunden bei der Entwicklung neuer Produkte und Dienstleistungen vergrößert das eigene Innovationspotenzial, kann das »Flop-Risiko« senken und helfen, die Kundenzufriedenheit zu erhöhen.

3.3.1 Web-basierte Ideenportale

Eine im Kontext der Web-basierten Informationsakquise besonders interessante Ausprägung von Software-Werkzeugen, die Outside-in-Prozesse unterstützen, sind Web-basierte Ideenportale. Sie sind eine Ausprägung der »User Innovation Toolkits« (von Hippel, 2005) und erfreuen sich – beflügelt durch prominente Beispiele wie »My Starbucks Idea«[73] und »Dell Idea Storm«[74] wachsender Beliebtheit.

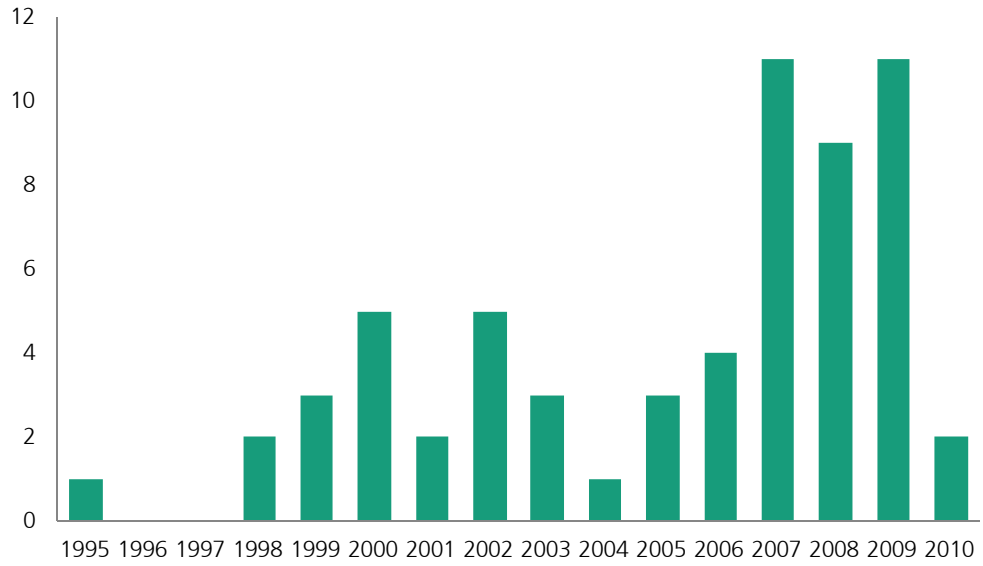
In einer Studie hat Fraunhofer IAO 77 Web-basierte Ideenplattformen untersucht. Abbildung 32 zeigt, wie viele Plattformen aus dieser Stichprobe in welchen Jahren gegründet wurden – ab 2007 erfolgten deutlich mehr Neugründungen[75]. Interessant ist auch, dass das älteste untersuchte Ideenportal, die Global Ideas Bank[76], bereits 1995 gegründet wurde.

Ein wichtiges Unterscheidungskriterium bei Web-basierten Ideenportalen ist die Offenheit des Inhalts. Viele Plattformen wählen einen geschlossenen Ansatz, indem sie Problemstellungen ausschreiben, die Ideen und Lösungsansätze der Benutzer aber nicht auf der Website veröffentlichen (z. B. InnoCentive[77]) – dies ist der übliche Weg bei Online-Ideenwettbewerben (vgl. Walcher, 2007) – allerdings gibt es Ausnahmen, so bietet etwa Atizo[78] Ideenwettbewerbe an, bei denen die Ideen offen einsehbar und diskutierbar sind. Ein großer Vorteil des geschlossenen Ansatzes liegt in der leichteren Sicherung geistigen Eigentums –

73 http://mystarbucksidea.force.com
74 http://www.ideastorm.com
75 Die Zahlen summieren sich nicht zu 77, da nicht zu allen Portalen das Jahr der Freischaltung ermittelt werden konnte.
76 http://www.globalideasbank.org
77 http://www.innocentive.com
78 https://www.atizo.com

sein größter Nachteil ist, dass er keine Diskussion in der Community zulässt. Im Kontext der internetbasierten Informationsakquisition interessanter sind somit jene Portale, die einen offenen Ansatz verfolgen und bei denen Ideen und Community-Feedback somit öffentlich sichtbar sind. Die in den folgenden Abschnitten dargestellten Ergebnisse der Untersuchungen wurden daher nur solche Portale herangezogen, die einen offenen Diskussionsansatz verfolgen.

Abbildung 32: Neugründungen von Ideenportalen (in der untersuchten Stichprobe)



Ein weiteres wichtiges Unterscheidungskriterium liegt im Betreibermodell. Hier werden im Folgenden die »firmeneigenen« von den »Mediator«-Plattformen abgegrenzt.

Firmeneigene Plattformen

Die beiden Portale von Starbucks und Dell haben viele Gemeinsamkeiten (in Abbildung 33 werden Screenshots gezeigt):

- Sie basieren beide auf Salesforce[79]-Technologie und verwenden ähnliche Workflows (Benutzer geben Ideen ein, Benutzer bewerten

[79] http://www.salesforce.com/de

Ideen, die besten Ideen werden (manchmal) umgesetzt). Dell nennt das Prinzip »View-Post-Vote-See«, während Starbucks es mit »Share-Vote-Discuss-See« bezeichnet.

- Beide Firmen bieten den Benutzern keine »quantifizierbare« Belohnung für die Teilnahme am Prozess – anscheinend reicht die Möglichkeit, dass die eigene Idee tatsächlich umgesetzt wird oder unter den am höchsten bewerteten Beiträgen zu sehen ist, als Anreiz aus.

- Beide Firmen bedienen direkt den Endkunden, und beide Firmen besitzen eine große und internetaffine Endkunden-Community.

- In beiden Fällen finden sich neben Produktideen sehr viele Beiträge, die sich mit der Verbesserung des Services befassen – die Portale dienen sozusagen gleichzeitig als Ideenplattformen und Beschwerdebriefkasten.

- In beiden Portalen gibt die betreibende Firma intensiv Feedback: So werden etwa die von der Firma aufgegriffenen Ideen auf der Hauptseite gesondert hervorgehoben (»Your ideas in action«). Auch in den Diskussions-Threads schalten sich auch immer wieder Mitarbeiter der Unternehmen ein.

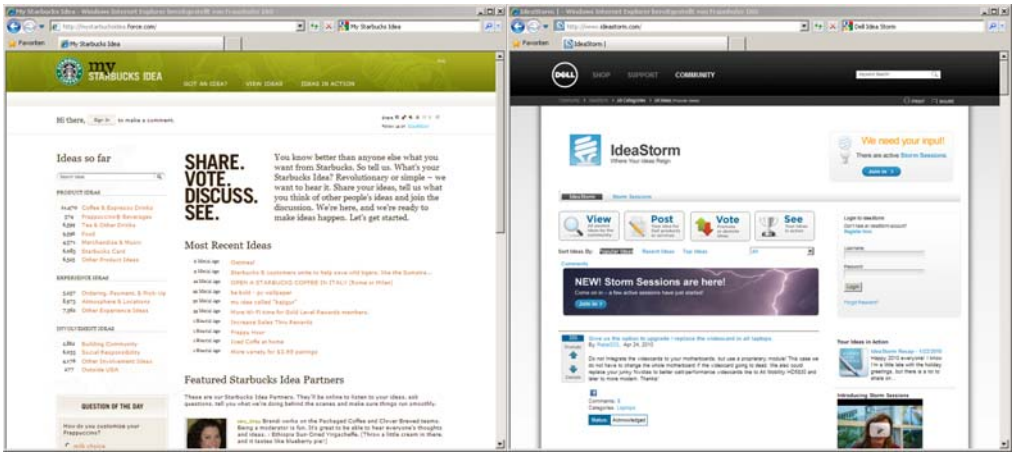


Abbildung 33: Screenshot prominenter Ideenportale

Links: My Starbucks Idea. Rechts: Dell Idea Storm

Insbesondere die letzten Punkte verdeutlichen, dass der laufende Betrieb einer solchen Plattform – und insbesondere der Aufbau und die Pflege der Mitglieder-Community - mit hohem Zeit- und (damit auch) Kosteneinsatz verbunden ist. Dennoch haben in den letzten Jahren eine ganze Reihe weiterer Firmen bzw. Organisationen ein eigenes Ideenportal als »Listening-Plattform« installiert. Tabelle 8 zeigt eine Auswahl. Unsere Untersuchungen zeigen, dass

- Firmen, die ein eigenes Ideenportal betreiben, üblicherweise im Endkundengeschäft tätig sind;

- eigene Ideenportale insbesondere bei IT-Unternehmen beliebt sind;

- derzeit die Mehrheit der Ideenportale von US-amerikanischen Organisationen betrieben werden;

- oftmals »Out-of-the-box«-Lösungen von Anbietern von Ideenportal-Software eingesetzt wird. Tabelle 9 zeigt eine Auswahl von Software-Lösungen, die bei den untersuchten Ideenportalen eingesetzt werden.

Tabelle 8:
Firmeneigene
Ideenportale

(i) Jahr der
Domainregistrierung
(ii) Erster Eintrag
(iii) Info auf Website
(iv) Aussage des
Betreibers
(v) Information aus
Pressemitteilung
oder Artikel

(a) Selbstauskunft
(b) Eigene Zählung
oder Schätzung
(c) Anzahl sichtbarer
Einträge

Name	URL	Sitz	Sprache	Online seit	Branchen-fokus	Firma	Community-größe	Anzahl Einträge	Suche	RSS	Kommen-tare	Bewertung
Acrobat Ideas	https://na5.brightidea.com	US	en	2009 (ii)	Software	Adobe	2209 (a)	610 (a)	ja	ja	ja	ja
Betavine	http://www.betavine.net/home/main/home.html	UK	en	2006 (i)	Mobilfunk	Vodafone	n/a	~40 (b)	ja	ja	ja	nein
BSA Innovation Engine	http://ideas.scouting.org	US	en	1998 (i)	Pfadfinder	BSA	2.782 (a)	443 (a)	ja	ja	ja	ja
Buglabs	http://buglabs.ideascale.com	US	en	n/a	IT	Buglabs	9.555 (a)	321 (a)	ja	ja	ja	ja
Camtasia Studio	http://camtasia.ideascale.com	CZ	en	n/a	Software	Camtasia	727 (a)	236 (a)	ja	ja	ja	ja
Cisco Idea Center	http://forum.cisco.com/eforum	US	en, de	2008 (ii)	IT	Cisco	n/a	~170 (b)	ja	ja	ja	ja
Cisco I-Prize	http://www.cisco.com/web/solutions/iprize/index.html	US	en	n/a	IT	Cisco	2.680 (a)	839 (a)	ja	nein	ja	ja
Dell IdeaStorm	http://www.ideastorm.com	US	en	2007 (ii)	IT	Dell	n/a	~14.000 (b)	ja	ja	ja	ja
Digidesign Pro Tools	http://protools.ideascale.com	US	en	n/a	Software	Digidesign	924 (a)	788 (a)	ja	ja	ja	ja
Kaneva Ideas	http://ideas.kaneva.com	US	en	2006 (i)	Software	Kaneva	4.577 (a)	2.182 (a)	ja	ja	ja	ja
My Starbucks Idea	http://mystarbucksidea.force.com	US	en	2008 (ii)	Kaffee und Service	Starbucks	n/a	~84.000 (b)	ja	ja	ja	ja
Nagios Open Innovation Network	http://ideas.nagios.org	US	en	2001 (i)	Software	Nagios	3.219 (a)	223 (a)	ja	ja	ja	ja
Netroots Nation	http://www.askthespeaker.org	US	en	2008 (i)	Öffentlicher Sektor	Netroots	3.135 (a)	357 (a)	ja	ja	ja	ja
New York State Senate	http://propertytaxideas.nysenate.gov	US	en	2009 (ii)	Öffentlicher Sektor	New York State Senate	1.407 (a)	38 (a)	ja	ja	ja	ja
Open Austin	http://openaustin.ideascale.com	US	en	n/a	Öffentlicher Sektor	Austin	n/a	92 (a)	nein	ja	ja	ja
Open Government Dialogue	http://opengov.ideascale.com	US	en	2009 (v)	Öffentlicher Sektor	National Academy of Public Administration	14.948 (a)	4.205 (a)	ja	ja	ja	ja
OpenInternet.gov	http://openinternet.ideascale.com	US	en	2009 (iii)	Öffentlicher Sektor	Federal Commu-nication Commission	3.801 (a)	344 (a)	ja	ja	ja	ja
preideas	http://www.preideas.com	US	en	n/a	IT, Mobilfunk	Palm	n/a	35 (a)	nein	ja	ja	ja
salesforce.com IdeaExchange	http://sites.force.com/ideaexchange	US	en	2006 (ii)	Software	salesforce.com	n/a	~10.000 ©	ja	ja	ja	ja
Tchibo Ideas	https://www.tchibo-ideas.de	de	de	2008 (v)	Haushalt, Freizeit	Tchibo	~7.500 (b)	~540 (b)	ja	nein	ja	ja

Tabelle 9: »Out-of-the-box«-Lösungen für Ideenportale

Name	URL	Hersteller	Referenzkunden
IdeaScale	http://ideascale.com	IdeaScale	Camtasia Studio, Buglabs
WebStorm	http://www.brightidea.com/i/webs	BrightIdea	Acrobat
IdeaNet	http://www.hyve.de/ideanet.php	Hyve AG	SAP, Ideenreich
Salesforce CRM ideas	http://www.salesforce.com	Salesforce	Starbucks, Dell
ContestSpigit	http://www.spigit.com	Spigit	Cisco

Mediatoren-Plattformen

Der große und mit hohen Kosten verbundene Aufwand der Pflege eigener Ideenportale führte in den letzten Jahren zur Gründung einer ganzen Reihe »unabhängiger« Ideenplattformen, deren Geschäftsidee darin besteht als Ideenmediator »Seeker« (also Firmen, die nach neuen Ideen oder Lösungen für bestehende Probleme suchen) und »Solver« (also Erfinder oder andere Kreative, die eine Idee anzubieten haben) zusammenzubringen. Diese Plattformen finanzieren sich weitgehend durch Gebührenzahlungen der Seeker (z. B. Atizo), Werbung (z. B. Springwise[80]) oder öffentliche Förderung (z. B. InnovationMarket[81], Ideenreich[82]). Eine spezielle Gruppe der Mediatorplattformen arbeitet gemeinnützig und versucht überhaupt nicht, Gewinn zu erzielen. Beispiele hierfür sind Portale zur Sammlung Umweltschutz-bezogener Ideen (z. B. Global Ideas Bank[83]) oder von politischen Initiativen (z. B. Open Internet[84], Open Government Dialogue[85]).

26 der 77 untersuchten Plattformen basieren auf einem offenen Mediator-Ansatz. Zu diesen Portalen konnten folgende Erkenntnisse gewonnen werden:

- *Geschäftsmodell*: Die meisten Plattformen finanzieren sich durch Werbung oder Gebühren, andere basieren auf Spenden oder öffentlichen Fördermitteln.

80 http://springwise.com
81 http://www.innovationmarket.de
82 http://ideenreich.at
83 http://www.globalideasbank.org
84 http://openinternet.ideascale.com
85 http://opengov.ideascale.com

- *Fachlicher Schwerpunkt*: Die meisten der Plattformen sind thematisch offen, einige Ausnahmen sind auf Fragestellungen des Umweltschutzes oder der IT spezialisiert.

- *Community*: Leider erhält man zu den meisten Plattformen keine oder nur unzureichende Informationen bzgl. der Größe und Zusammensetzung der Community. Die Größe der Communities, die wir bestimmen konnten, liegt zwischen 100 und 30.000 Mitgliedern.

- *Anzahl der Ideen*: Die Anzahl der auf den Plattformen einsehbaren Ideen variiert von 100 bis 40.000 und liegt meist zwischen 1.000 und 10.000.

- *Offene Ideendiskussion*: Alle untersuchten Plattformen erlauben das Kommentieren der Ideen durch die Community. Einige lassen zusätzlich das Kommentieren der Kommentare zu, bieten also ein wirkliches Diskussionsforum. Die meisten Portale erlauben zudem eine Bewertung der Ideen, die genutzten Bewertungsmechanismen variieren dabei erheblich sowohl bzgl. der verwendeten Skala als auch bzgl. der Granularität.

- *Datenmodell*: Wie wird eine Idee beschrieben, d.h., welche Eingabefelder werden befüllt? Während einige Plattformen sich auf den Titel und eine kurze textuelle Beschreibung beschränken, erlauben andere die Einstellung multimedialer Inhalte (etwa Bilder und Videos) und bieten ein komplexes Ideen-Status-Modell an (z. B. »Ongoing, Evaluated, Rejected, Idea in Market«[86]).

- *Benachrichtigungen*: Viele Plattformen bieten die Möglichkeit, den Benutzer aktiv über neue Inhalte zu informieren – per Newsletter oder RSS-Feed – eine für das Innovation Mining sehr hilfreiche Funktion. Nur wenige Plattformen erlauben allerdings das Abonnieren von Benachrichtigungen zu bestimmten Kategorien oder Stichwörtern.

Des Weiteren wurden die 26 Portale bzgl. subjektiver Kriterien untersucht:

- *Ideensuche*: Wie einfach kann der Ideen-Pool des Portals durchsucht werden? Können die neuesten und die am besten bewerteten Ideen auf einfache Weise identifiziert werden? Wie nützlich erweisen sich die angebotenen Klassifikations- oder Tagging-Mechanismen?

[86] Mögliche Ideenstatus bei Crowdspirit (http://www.crowdspirit.com).

- *Eingabe von Inhalten*: Wie einfach kann eine neue Idee eingetragen werden? Wie gut arbeitet die Kommentarfunktion? Ist der Inhalt mit Spam kontaminiert? Wie effektiv arbeiten die Bewertungsmechanismen?

- *Registrierung*: Wie einfach kompliziert ist der Registrierungsprozess? Ist die Registrierung sicher?

- *Optischer Eindruck und Usability*: Wie gut und nützlich sind Design und Layout?

Die Bewertung erfolgte anhand einer Skala von 0 (schlecht) bis 3 (sehr gut) und wurde von drei Test-Benutzern unabhängig durchgeführt. Anschließend wurde der Mittelwert gebildet. Tabelle 10 zeigt die Bewertung der insgesamt am höchsten bewerteten Portale. Tabelle 11 zeigt weitere Informationen zu den untersuchten Portalen im Überblick.

Tabelle 10: Subjektive Bewertung einiger Ideenportale

Portal	Insgesamt	Ideen suchen	Eingeben, kommentieren, bewerten	Registrieren	Design
Ideawicket	••	••	••	•	••
IdeasXchange	••	••	••	•	•
WhyNot?	••	••	••	•	○
iBridge Network	••	••	-	•	••
Lovemyidea.com	••	•	•	•	•
Halfbakery	••	•	•	○	•
Ideenreich	••	○	••	•	○

Tabelle 11: Mediatoren-betriebene Ideenportale

(i) Jahr der Domainregistrierung
(ii) Erster Eintrag
(iii) Info auf Website
(iv) Aussage des Betreibers
(v) Information aus Pressemitteilung oder Artikel

(a) Selbstauskunft
(b) Eigene Zählung oder Schätzung
(c) Anzahl sichtbarer Einträge

Name	URL	Sitz	Sprache	Online seit	Branchen-fokus	Finanzierung	Community-größe	Anzahl Einträge	Suche	RSS	Kommen-tare	Bewertung
Atizo Ideas	https://www.atizo.com	AT	de, en, fr	2008 (iv)	nein	Investoren	n/a	n/a	ja	ja	ja	ja
Brain Reactions	http://brainreactions.net/	US	en	2004 (iii)	nein	Gebühren	~2.000	~25.000 (b)	ja	ja	nein	ja
Business Idea Aggregator	http://www.busines-ideas.com.ar	AR	en	2003 (i)	E-Business	Sponsoren	n/a	~800 (b)	ja	ja	ja	ja
Cash4Idea	http://www.cash4idea.com	NL	nl, en	n/a	nein	n/a	n/a	~30 (c)	nein	nein	ja	ja
Challenge Your World	http://challengeyourworld.com	CA	en	2008 (ii)	Geschäftsideen	n/a	n/a	~170 (b)	ja	ja	ja	nein
Creativitypool	http://www.creativitypool.com/	US	en	2002 (ii)	nein	Spenden	n/a	~1.800 (b)	ja	nein	ja	ja
Erfinderprofi	http://www.erfinderprofi.de/	DE	de	2007 (ii)	nein	Werbung	n/a	~200 (b)	ja	ja	ja	ja
Global Ideas Bank	http://www.globalideasbank.org	UK	en	1995 (iii)	Umwelt	Spenden, Buchverkäufe	n/a	~6.000 (b)	ja	nein	ja	ja
Halfbakery	http://www.halfbakery.com	n/a	en	1999 (i)	nein	n/a	~30.000 (a)	~40.000 (a)	ja	ja	ja	ja
IBridgeNetwork	http://www.ibridgenetwork.org/	US	en	2007 (iii)	nein	Sponsoren	~4.000 (a)	~11.000 (b)	ja	ja	nein	nein
Idea a day	http://www.idea-a-day.com	n/a	en	2000 (iii)	nein	n/a	n/a	~3.500 (b)	nein	ja	ja	nein
Ideas Project	http://www.ideasproject.com	n/a	en	n/a	Mobilfunk	Sponsoren	n/a	~500	ja	ja	ja	nein
IdeasXchange	http://ideasxchange.com	US	en	2005 (i)	nein	Sponsoren	n/a	~450 (b)	ja	nein	ja	ja
Ideawicket	http://www.ideawicket.com	IN	en	2007 (iii)	nein	Werbung, Gebühren	n/a	~850 (b)	ja	nein	ja	nein
Ideenreich	http://www.ideenreich.at	AT	de	2005 (iv)	nein	öffentlich gefördert	~2.000 (a)	~1.500 (b)	ja	nein	ja	ja
Innovation Exchange	http://innovation-exchange.org	UK	en	2009 (ii)	nein	n/a	n/a	~1.800 (b)	ja	ja	ja	ja
Innovation Market	http://www.innovationmarket.de	DE	de	2005 (i)	nein	öffentlich gefördert	n/a	~100 (b)	ja	nein	ja	nein
InnoXchange	http://www.innoxchange.net	MY	en	2009 (ii)	nein	Sponsoren	n/a	~10 (b)	nein	nein	ja	ja
Lovemyidea.com	http://www.lovemyidea.com	UK	en	2007 (iii)	nein	Werbung	n/a	~500 (b)	ja	ja	ja	ja
Patent-net	https://www.patent-net.de	DE	de	n/a	nein	n/a	n/a	50 (b,c)	ja	ja	ja	ja
Premises Premises	http://www.premisespremises.com	US	en	2002 (ii)	nein	Privat	n/a	~150 (b)	nein	nein	ja	nein
Save Our Energy	http://www.save-our-energy.de	DE	de	2009 (ii)	Energieeffizienz	öffentlich gefördert	333 (a)	165 (a)	ja	ja	ja	ja
Should Exist	http://shouldexist.org	US	en	2000 (i)	Spenden	Spenden	n/a	~ 350 (b)	ja	ja	ja	nein
Why Not?	http://www.whynot.net	US	en	2002 (i)	nein	n/a	n/a	~5.500 (b)	ja	ja	ja	ja
Yet2Com	http://www.yet2.com	US	en	1999 (iii)	nein	Gebühren	n/a	~5.500 (b)	ja	ja	ja	ja

3.3.2 Anwendungsbeispiel

Der offensichtliche Nutzen von Ideenportalen liegt darin, Benutzerideen insbesondere der eigenen Kunden für neue Produkte und Dienstleistungen zu erhalten. Gleichzeitig eröffnen die eigenbetriebenen Portale den Unternehmen als Listening-Plattformen einen neuen Kommunikationskanal zu ihren Endkunden und dienen neben dem Ideen-Insourcing auch der proaktiven Verbesserung der Kundenbeziehungen.

Neben dem Betrieb einer eigenen Plattform oder der aktiven Nutzung einer Mediator-Plattform oder eines Ideenwettbewerbs kann aber auch bereits die Beobachtung der Ideendiskussionen auf bestehenden Plattformen (insbesondere von Wettbewerbern) nützliche wertvolle Informationen liefern: So wird es sich für den Betreiber einer kleineren Kaffeehauskette möglicherweise nicht lohnen, eine eigene Plattform zu betreiben – aber durch die Beobachtung der Starbucks-Community kann er durchaus Ideen finden, die zu den Bedürfnissen seiner eigenen Kunden passen. Unter Umständen kann er diese Idee aufgrund seiner geringeren Größe sogar einfacher und schneller umsetzen als Starbucks selbst – und sich somit (zumindest vorübergehend) einen Wettbewerbsvorsprung sichern.

3.3.3 Weiterführende Literatur

Bjelland, O. M. und Chapman Wood, R. (2008): An Inside View of IBM's Innovation Jam. In: MIT Sloan Management Review, Jg. 50, H. 1, S. 32–40.
Fasst die Erfahrungen zusammen, die die Firma IBM in ihrer 2006 veranstalteten massiv-parallelen Onlinekonferenz zur Ideenentwicklung und –diskussion mit 150.000 Teilnehmern sammelte.

Chesbrough, H. W. (2006): Open innovation. The new imperative for creating and profiting from technology. Boston, Mass.: Harvard Business School Press.
Das Standardwerk zum Thema Open Innovation.

Di Gangi, P. M. und Wasko M. (2009): Steal my idea! Organizational adoption of user innovations from a user innovation community: A case study of Dell IdeaStorm. Information product markets. In: Decision Support Systems, Jg. 48, H. 1, S. 303–312.
Schwerpunkt dieses Aufsatzes ist die Analyse der Beweggründe, die eine Firma (in diesem Fall der Computerhersteller Dell) veranlassen, bestimmte Benutzerideen aufzugreifen und andere fallen zu lassen.

Hippel, E. v. (2005): Democratizing Innovation: MIT Press Ltd.
Standardwerk zum Themenfeld der Benutzer-Einbindung in Innovationsprozesse.

Finzen, J., Kintz, M. und Kobes, S. (2010): A Comparative Study of Publicly Accessible Web-Based Idea Portals. In Proceedings of the XXI ISPIM Conference, Bilbao.
Untersuchung aktueller öffentlich einsehbarer Ideenportale, Grundlage des vorliegenden Abschnitts.

Walcher, D. (2007): Der Ideenwettbewerb als Methode der aktiven Kundenintegration: Theorie, empirische Analyse und Implikationen für den Innovationsprozess. Wiesbaden: Deutscher Universitäts-Verlag / GWV Fachverlage GmbH, Wiesbaden.
Dissertation, die sich damit auseinandersetzt, inwieweit sich internetbasierte Ideenwettbewerbe als Mittel der Kundenintegration eignen.

3.4 Exkurs: Onlinebasiertes Reputationsmanagement

Die Reputation eines Unternehmens, sein Ruf und Image, wird bestimmt durch die Informationen über das Unternehmen, seine Produkte und zum Beispiel auch seine Geschäftsführer und Mitarbeiter. Reputationsmanagement oder auch onlinebasierte Reputationsmanagement (ORM) umfasst sowohl die Beobachtung all dieser Informationen im Internet als auch die gezielte Einflussnahme darauf.

Die Informationsgewinnung über die Online-Reputation eines Unternehmens – das sogenannte Monitoring - erfolgt über Web Mining-Technologien (siehe hierzu Kapitel 4). Der folgende Abschnitt gibt einen Überblick über das Themenfeld Reputationsmanagement.

3.4.1 Motivation für Unternehmen

ORM ist ein strategisches Themenfeld, welches in vielen deutschen Unternehmen heute noch nicht etabliert ist. Zwar sind 60 Prozent der 100 größten Marken in Deutschland bereits aktiv im Social Web, jedoch ist eine umfassende Social Media-Strategie als Teil des Reputationsmanagements noch die Ausnahme. Nur 5 Prozent der Unternehmen bedienen zugleich Facebook, Twitter, YouTube und Corporate Blogs[87]. Dabei bieten diese Kanäle hervorragende Chancen, den Kunden direkt anzusprechen, um ihn gezielt positiv zu beeinflussen. Eine besondere Herausforderung besteht darin, auf negative Meinungen von Benutzern einzugehen und diese in positive Erfahrungen zu verwandeln. Abbildung 34 zeigt, dass Empfehlungen von Bekannten im Internet am meisten zählen und Konsumentenmeinungen als ebenso vertrauenswürdig wahrgenommen werden wie Markenauftritte selbst.

Die Zuständigkeit für das Thema Social Web wird in deutschen Unternehmen unterschiedlichen Abteilungen zugeschrieben (vgl. Abbildung 35). Dies ergibt sich aus der Heterogenität der Aufgaben und Ziele, die unter dem Thema zusammengefasst werden: Die Nutzung von Social Media, indem beispielsweise gezielt Kampagnen für Plattformen wie Facebook oder Twitter konzipiert werden, kann direkt messbaren Gewinn generieren und hat gleichzeitig positiven Einfluss auf die Reputation – es handelt sich dabei jedoch nicht um ein umfassendes Reputationsmanagement. Ein solches umfasst vor allem auch den Umgang mit ungewollten, unwahren und allgemein dem Unternehmensimage schadenden Informationen im Internet. Als effektives Krisenmanagement spürt Reputationsmanagement Risiken rechtzeitig auf, um Kosten zu vermeiden.

[87] Empirische Studie der Universität Oldenburg und construktiv GmbH. (2009) Wie nutzen Deutschlands größte Marken Social Media?

Abbildung 34:
Vertrauen in die verschiedenen Werbemöglichkeiten

Lesehilfe: 90 Prozent der Befragten vertrauen »vollkommen« oder »einigermaßen« Empfehlungen von Personen, die sie kennen.

Darstellung basierend auf Nielsen Global Online Consumer Survey April 2009

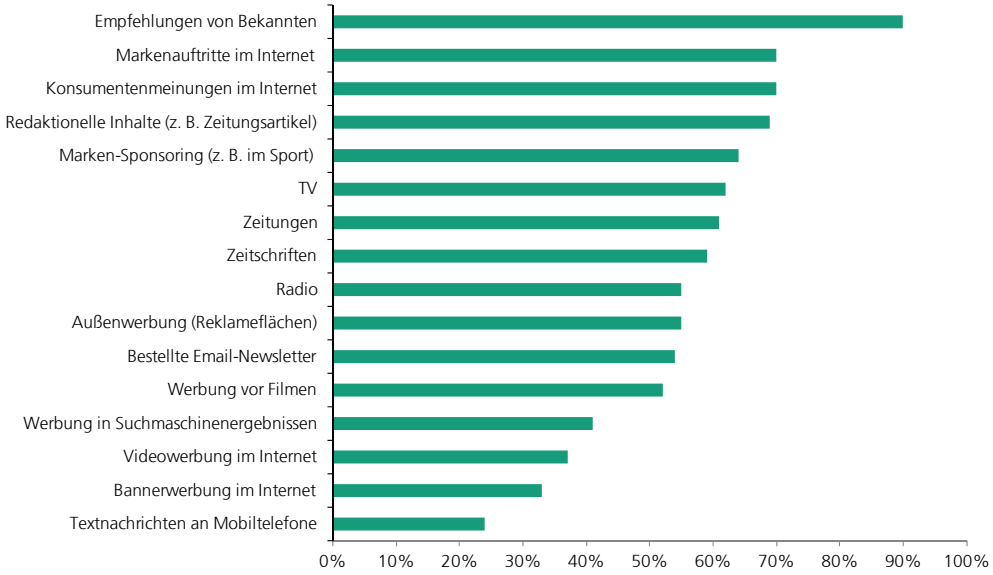
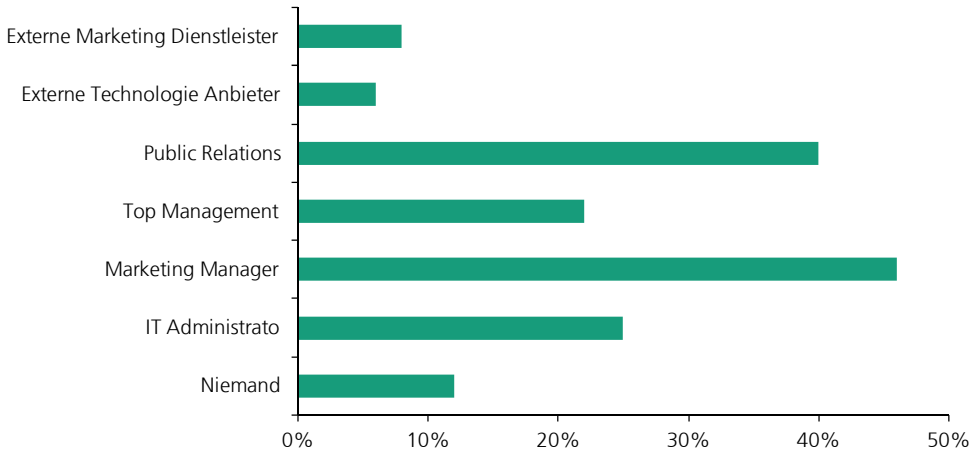


Abbildung 35:
Verantwortung für das Thema Social Media

Darstellung basierend auf Social Media Company Monitor 2009



Während Online und Social Media Marketing im Rollen- und Organisationsmodell von Unternehmen fest etabliert sind, gibt es noch keine feste Zuordnung des Reputationsmanagers. Dieser sollte aufgrund der zu treffenden strategischen Entscheidungen nahe der Unternehmensführung angesiedelt sein.

3.4.2 Vorgehen im Reputationsmanagement

Nachhaltigkeit im Umgang mit Social Media erreicht man durch die Vorab-Definition von Zielgruppen und Strategie. Über direkte Marketing-Maßnahmen hinaus muss heute in der Organisation von Unternehmen auch der Reputationsmanager einen festen Platz einnehmen. Eine gute Einführung in die Funktionsweise des Web 2.0 und Motivation und Anleitung zum Reputationsmanagement gibt das Buch »Groundswell« (Li et al. 2008) welches auf Erhebungen und Erkenntnissen von Forrester Research basiert. Die Einführung und Durchführung des Reputationsmanagements erfolgt diesem Modell zufolge in fünf aufeinander aufbauenden Stufen:

1. *Listening*: Ein Unternehmen muss sich zunächst ein Bild über die Informationslandschaft im Internet machen und, da diese sich kontinuierlich verändert, dieses Bild regelmäßig auffrischen. Angesichts der Menge an Inhalten im Netz ist die einfache Datensammlung und Auswertung über die Suche in verschiedenen Suchmaschinen oft nicht ausreichend. Social Media Monitoring-Tools sind IT-Werkzeuge, die Information anhand bestimmter Stichworte auffinden, klassifizieren und analysieren. Für kleinere Unternehmen und Privatpersonen eignen sich zum Einstieg unter Umständen auch frei verfügbare Tools, wie zum Beispiel Google Alerts[88]. Eine weitere interessante Möglichkeit ist Tattler[89], eine Open Source Lösung mit vielen Funktionen (jedoch auch einem nicht unerheblichen Installationsaufwand).
   Das Beobachten und Auswerten dessen, was über die eigene Marke, die Produkte, Dienstleistungen, Mitarbeiter und Geschäftsführer im Internet veröffentlicht wird, ist die Basis für die Definition der Strategie für Social Media und Reputationsmanagement. Krisenprävention muss dabei immer ein Fokus sein, zusätzliche Anwendungen der Datenanalyse sind oftmals möglich und sinnvoll.

2. *Talking*: Unternehmen sollten in einem zweiten Schritt nach dem Zuhören (Listening) an den Diskussionen im Web 2.0 in Foren und Blogs teilnehmen oder auch selbst Kanäle schafft, welche die Diskussion mit Kunden und Geschäftspartnern ermöglichen und fördern. Abgesehen von dem Gewinnsteigerungspotenzial solcher Maßnahmen ermöglicht die aktive Teilnahme an Diskussionen im Internet es Inhalte zu steuern, Risiken frühzeitig zu erkennen und im Krisenfall bereits in den relevanten Kanälen präsent zu sein. Beim Agieren in sozialen Netzwerken müssen Unternehmen dabei ehrlich und menschlich

[88] http://www.google.com/alerts
[89] http://tattlerapp.com

auftreten. Die Kommunikation nach Außen ist dabei am besten mit konkreten Personen verknüpft. Das Potenzial der eigenen Mitarbeiter, die zumeist schon aktiv am Social Web teilnehmen, sollte ausgenutzt werden. Social Media Richtlinien für alle Mitarbeiter unterstützen und kommunizieren die Strategie des Unternehmens und sind im aktiven Reputationsmanagement unerlässlich. Ein gutes Beispiel hierfür sind »IBMs Social Computing Guidelines«[90]. Nicht immer ist ein Vorgehen mit zahlreichen Kampagnen angezeigt - die Beschränkung auf einzelne Kanäle, in denen sich die Zielgruppe aufhält oder die Schaffung von solchen Kanälen zum Beispiel in Form von eigenen Communities kann ebenfalls zum Erfolg führen.

3. *Energizing*: Durch Beobachtung und Teilnahme am Social Web können sogenannte »lead user« oder »influencer« identifiziert werden. Das können die aktivsten Mitglieder eines Forums oder auch einflussreiche Blogger sein. Diese Personen gilt es als Fürsprecher für die eigene Marke, das eigene Produkt oder die eigene Dienstleistung zu gewinnen, zum Beispiel indem ihnen besondere Informationen oder Kontakte ermöglicht werden. Hersteller von Unterhaltungselektronik stellen ihre Produkte oft vorab bekannten Bloggern zum Test zur Verfügung. Auf dieser Stufe des »Energizing« sind direkte Kontakte entscheidend[91], hierzu zählt auch ein besonderer Fokus auf eine eigene oder bereits bestehende als besonders wichtig identifizierte Community.

4. *Supporting*: Die erfolgreiche Beteiligung am Social Web erfordert einen langen Atem und konstante Strategie-Anpassung. Die Unterstützungsphase zielt darum auf »Hilfe zur Selbsthilfe«: Unternehmen stellen Plattformen und/oder Mitarbeiter bereit, um Internetbenutzern zu ermöglichen, sich auszutauschen und gegenseitig zu unterstützen.

5. *Embracing* als letzte Stufe bezeichnet die permanente intensive Beteiligung von Unternehmen an Social Media Plattformen, zum Beispiel um Ideen für die Produktentwicklung zu gewinnen (siehe auch Abschnitt 3.3 zu Ideenportalen).

[90] IBM Social Computing Guidelines, http://www.ibm.com/blogs/zz/en/guidelines.html

[91] Für Marketingmaßnahmen vermitteln Seeding-Agenturen die Kontakte zu »lead usern« und »influencern«, im Reputationsmanagement ist es jedoch wichtig, Beziehungen zwischen Kunde und Unternehmen direkt aufzubauen.

3.4.3 IT-Unterstützung für das Reputationsmanagement

Zur zeitnahen und strukturierten Erschließung der großen Menge an nutzergenerierten, reputationsrelevanten Informationen im Internet können IT-Werkzeuge eingesetzt werden. Für diese Werkzeugklasse gibt es noch keine etablierte Bezeichnung, und auch die angebotene Funktionalität ist sehr heterogen. Durchzusetzen scheint sich der Begriff »Social Media Monitoring«, häufig wird aber auch der Begriff »Web Monitoring«[92] verwendet. Eine Auswahl solcher Werkzeuge ist in Tabelle 12 dargestellt. Um einen besseren Überblick über Eigenschaften und Möglichkeiten dieser Werkzeuge zu erhalten, plant Fraunhofer IAO im November 2010 die Veröffentlichung einer entsprechenden Marktstudie. Allgemein werden in der Funktionsweise der Werkzeuge folgende Schritte unterschieden:

- *Aggregation* von reputationsrelevanten Informationen
- *Bereinigung und Aufbereitung* der Ergebnisse, sowie die Sortierung nach Relevanz
- *Analyse* der Ergebnisse und Visualisierung

Das Sammeln der Informationen im Social Web ist der erste Schritt, den solche Werkzeuge durchführen. Es erfolgt mittels Crawling- und Scraping-Technologien (vgl. Abschnitt 4.2). Ein Werkzeug sollte die Möglichkeit bieten, die durchsuchten Quellen einzusehen und die Quellenbasis zu erweitern, da für unterschiedliche Unternehmen und Branchen spezifische Quellen relevant sind.

Die Analyse der erhobenen Daten erfolgt auf verschiedenen Ebenen. Neben quantitativer Analyse (wie viele Beiträge zum Thema, auf welchen Plattformen, wie oft wird der Suchbegriff erwähnt etc.) ist vor allem auch die qualitative Analyse der Daten wichtig. Deren Automatisierung ist nur begrenzt möglich, zum Beispiel sind zum Erkennen der Tonalität, ob also ein Beitrag positiv oder negativ zu bewerten ist, sprach- und manchmal sogar produktspezifische Kenntnisse notwendig. Viele Hersteller ergänzen ihr Analyseangebot durch menschliche Analyse, denn Ironie und Sarkasmus werden im Social Web häufig verwendet, können bisher durch Algorithmen aber nur unzureichend erkannt werden. Abbildung 36 zeigt die Filtermöglichkeit nach Tonalität (»Sentiment Detection«) bei Radian 6[93].

[92] Web Monitoring scheint als Begriff zwar korrekter als Social Media Monitoring, da die Quellen dieser IT-Werkzeuge oftmals nicht ausschließlich auf Social Media beschränkt sind, allerdings ist die Bezeichnung Web Monitoring uneindeutig, weil sie leicht mit Begriffen wie Web Analytics, wo es um die Erhebung von Klickraten und Ähnlichem geht, verwechselt wird.

[93] http;//www.radian6.com

Tabelle 12: Social Media Monitoring-Tools

Auswahl von IT-Werkzeugen, die im Bereich Social Media Monitoring eingesetzt werden können.

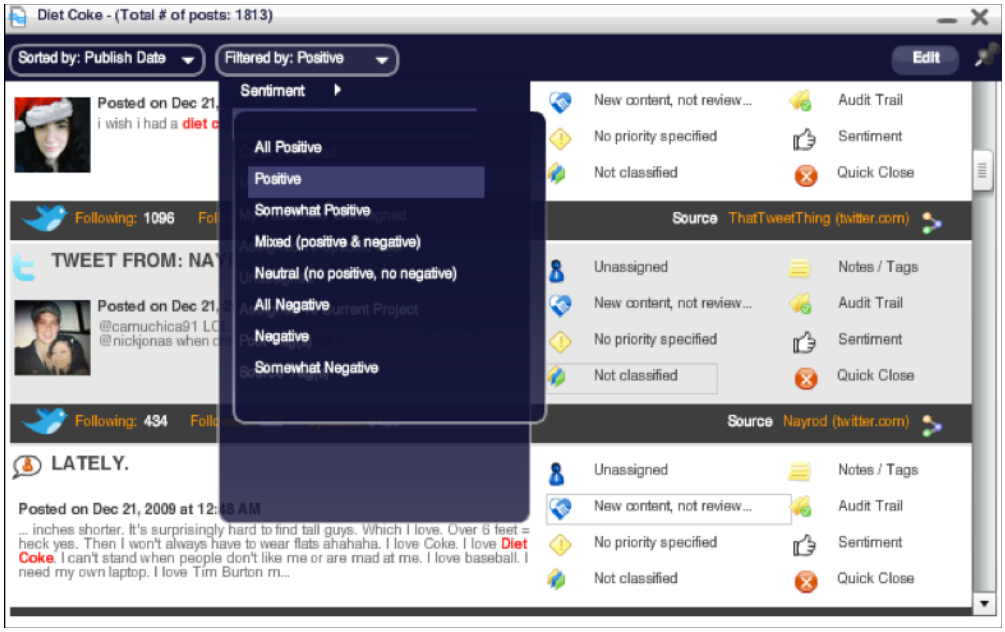
Hersteller, Produkt	Land	URL
Alterian, SM2	USA	http://socialmedia.alterian.com
Attensity360 (früher BIZ360)	USA	http://www.attensity360.com
Brands Eye	Südafrika	http://www.brandseye.com
Brandwatch	Großbritannien	http://www.brandwatch.com
Business Intelligence Group B.I.G. Screen	Deutschland	http://www.intelligence-group.com
Cogisum Web 2.0 Observer	Deutschland	http://www.cogisum.com
Complexium, Social Media MatriCX	Deutschland	http://www.complexium.de
Cymphony	USA	http://www.cymfony.com/
Digimind	Frankreich	http://www.digimind.com
Infegy Social Radar	USA	http://www.infegy.com/socialradar.php
Infospeed	Deutschland	http://www.infospeed.de
Intergrasco	Norwegen	http://www.integrasco.com
Jive Market Engagement (früher Filtrbox)	USA	http://www.filtrbox.com
Meltwater Buzz, News	Norwegen	http://www.meltwater.com
Netbreeze	Schweiz	http://www.netbreeze.ch
Nielsen BuzzMetrics	USA	http://de.nielsen.com
Nielsen BuzzMetrics	USA	http://www.nielsen.com
Radian 6	Kanada	http://www.radian6.com
Synthesio	Frankreich	http://www.synthesio.com
Sysomos MAP, Heartbeat	Kanada	http://www.sysomos.com
Trackur	USA	http://www.trackur.com
Vico Research	Deutschland	http://www.vico-research.com
Visible Technologies	USA	http://www.visibletechnologies.com
Whitevector	Finnland	http://www.whitevector.com

Frei verfügbare IT-Werkzeuge zur Beobachtung der Online-Reputation, zum Beispiel Google Alerts[94], eignen sich für Unternehmen nur bedingt, vor allem, wenn durch verschiedene Marken und Produkte ein großer Suchwortraum aufgespannt wird. Professionelle Social Media Monitoring-Tools sparen Zeit, indem sie Suchräume, Ergebnisse und Analysen zentral verfügbar machen und es ermöglichen, diese Information direkt an die zuständigen Mitarbeiter weiterzuleiten oder in Berichten zusammenzufassen. Am häufigsten werden diese Werkzeuge von Agenturen eingesetzt.

[94] http://www.google.com/alerts

Abbildung 36: Filtern nach Tonalität

Social Media Monitoring-Tools bieten vielfältige Analysemöglich-keiten, hier z. B. Sentiment Detection/ Erkennen der Tonalität in Radian 6



3.4.4 Fallbeispiele Social Media

Maßnahmen zur Verbesserung der Online-Reputation können aus erfolgreichen Social Media Marketing- und PR-Kampagnen abgeleitet werden. Einen Überblick über die Social Web-Aktivitäten der Top-100-Marken[95] gibt die Webseite ENGAGEMENTdb[96]. Am wirtschaftlich erfolgreichsten sind laut der im Juni 2009 veröffentlichten Studie[97] Unternehmen, die gleichzeitig mehrere Plattformen bedienen. Konkrete Zahlen zu dem durch seinen Twitter-Account generierten Umsatz hat der US Computerhersteller Dell im Juni 2009 in seinem Blog Direct2Dell veröffentlicht[98]: Seit 2007 wurden mit dem Twitter-Account @DellOutlet über zwei Millionen US-Dollar eingenommen. Eine weitere Erfolgsgeschichte im Social Web ist die des Wispa-Schokoriegels in Großbritannien. Ob es sich dabei um eine geplante virale Marketingaktion des Herstellers Cadbury handelt, ist strittig. Laut Cadbury[99] führte die enorme Resonanz privat initiierter Facebook-Fanseiten 2007 zunächst zu einem temporären Wiederverkauf des beliebten Schokoladenriegels, dessen Vertrieb

95 Top-100-Marken basierend auf der 2008 BusinessWeek/Interbrand Best Global Brands Umfrage: http://www.businessweek.com/magazine/toc/08_39/B4101global_brands.htm
96 http://www.engagementdb.com. Ergänzend veröffentlich die Agentur Vierpartner 2010 eine Vergleichsstudie, welche die Social Web-Aktivitäten der Top-30 der deutschen DAX-Konzerne untersucht: http://www.vierpartner.de/
97 http://www.engagementdb.com/downloads/ENGAGEMENTdb_Report_2009.pdf
98 http://en.community.dell.com/dell-blogs/b/direct2dell/archive/2009/06/11/delloutlet-surpasses-2-million-on-twitter.aspx
99 Anfang des Videos http://www.wispa.co.uk/wistory.aspx?vid=1

2003 eingestellt worden war. Aufgrund des großen Erfolges wird Wispa seit 2008 wieder regulär hergestellt, vertrieben und im Social Web beworben.

Als »Social Media-Fälle«, »Social Media Fails« oder »Social Media-Krisen« bezeichnet man Vorfälle, in denen nutzergenerierte Inhalte und deren Verbreitung und Vervielfältigung über das Web 2.0 hinaus in Online- und Printmedien der Reputation von Unternehmen massiv geschadet haben. Solche Fälle veranschaulichen die neuen Risiken und die Notwendigkeit eines Reputationsmanagements. Charakteristisch für Social Media-Fälle ist:

- Diskussionen werden schnell außerhalb von Social Media in Online-Medien, Print und TV aufgegriffen.
- Manipulation von Communities wird aufgedeckt.
- Unternehmen werden zu Hauptakteuren in Social Media–Fällen, indem sie versuchen, auf juristischem Weg Recht zu bekommen.
- Information, die von Unternehmen gelöscht wird, verbreitet und vervielfältigt sich umso mehr. Dieses Phänomen bezeichnet man als »Streisand-Effekt« – basierend auf dem Fall der Sängerin und Schauspielerin Barbara Streisand, die versucht hat, ein Foto ihres Hauses aus dem Internet zu entfernen.

Aktuelles Beispiel ist der Fall KitKat / Nestlé, welchen Greenpeace am 17. März 2010 mit der Veröffentlichung einer Website[100] und dem zugehörigen Video auf YouTube startete, siehe Abbildung 37. Angeprangert wird darin die Verwendung von Palmöl von Zulieferern, die den indonesischen Regenwald und somit den Lebensraum der Orang-Utans zerstören. Zunächst ging Nestlé juristisch gegen die Veröffentlichung des Videos auf YouTube vor, woraufhin eine alternative Videoplattform genutzt wurde. Vier Tage später waren wieder mehrere Kopien auf YouTube zu finden. Gleichzeitig startete eine Diskussion auf den Facebook-Fanseiten von KitKat und Nestlé, und nur zwei Tage nach Veröffentlichung des Videos am 19. März berichtete SkyNews auf seiner Website[101] über den Fall, es folgen Online- und Offline-Medien weltweit. Vorläufiges Ende fand das Ganze mit einer Pressemitteilung[102] von Nestlé im Mai 2010, in der der Konzern sich zum Schutz des Regenwaldes verpflichtet.

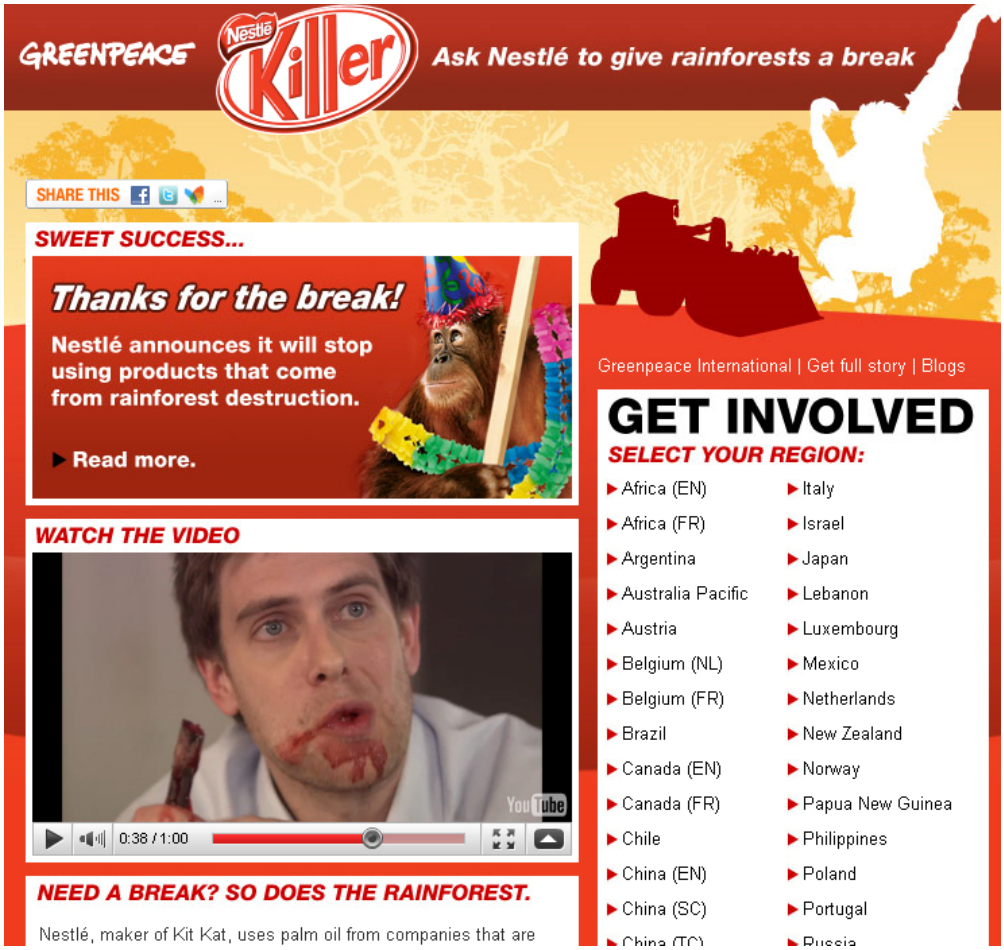
[100] http://www.greenpeace.org/international/campaigns/climate-change/kitkat/

[101] http://news.sky.com/skynews/Home/Business/Nestl-Under-Fire-For-Snarky-Attitude-To-Social-Networking/Article/201003315577468?lpos=Business_First_Buisness_Article_Teaser_Region_3&lid=ARTICLE_15577468_Nestl%3F_Under_Fire_For_Snarky_Attitude_To_Social_Networking

[102] http://presse.nestle.de/presseinfo/nestle__1444/

Abbildung 37:
Social Media Fall:
Greenpeace gegen
KitKat

Darstellung der
Kampagnenwebsite



Auswirkungen und Statistiken zu Nestlé / Greenpeace und anderen Krisen im Social Web werden tagesaktuell in Blogs diskutiert und kommentiert. Um diese Beispiele zu finden, muss man nur den Namen des Falles zusammen mit den Stichwörtern »Social Media«, »Krise«, »crisis« oder »fail« in eine Suchmaschine eingeben. Tabelle 13 beschreibt kurz einige Social Media-Fälle der letzten Zeit und nennt die Stichworte, unter denen man mehr Informationen dazu findet. Ältere und berühmte Beispiele wie »Dell's Hell« oder »Kryptonite's Blogstorm« werden als Fallbeispiele in der Literatur zum Thema Reputationsmanagement besprochen (z. B. in Beal et. al. 2008).

Tabelle 13: Social Media-Fälle

Beispiele für PR-Krisen durch Social Media.

Jahr	Stichworte + Social Media	Kurzbeschreibung
2010	BP, @BPGlobalPR	Seit 19. Mai 2010 wird die Ölkatastrophe im Golf auf dem anonymen Twitteraccount @BPGlobalPR satirisch kommentiert. Über 11.000 Follower innerhalb der ersten 5 Tage, 55.000 am 27. Mai 2010 - zum selben Zeitpunkt folgen unter 7000 dem offiziellen Twitteraccount @BP_America.
2010	Siedler 7, Ubisoft, Amazon	Schlechte Nutzerbewertung auf Amazon, als Protest gegen restriktiven Kopierschutz des Spieles.
2010	Paperchase, Eloise	Urheberrechtsverletzung der Kaufhauskette Paperchase an dem Werk der Designerin Eloise. Unternehmen reagiert erst nach Sturm im Social Web.
2010	Süddeutsche Zeitung, Trigami	Mittels der Agentur Trigami wollte die Süddeutsche Zeitung eine virale Kampagne für ihre iPhone App starten. Ergebnis war, dass Blogger gekauft werden sollten, worüber einer davon berichtete.
2009	Deutsche Bahn, Verdeckte PR	Deutsch Bahn bezahlt Positiv-Beiträge in Foren etc. Spiegel und Süddeutsche decken auf, die Intenet Community ist empört.
2009	Belkin, Produktbewertung	Auch Belkin bezahlte über Amazon Mechanical Turk für positive Bewertungen eines ihrer Produkte.
2009	Jack Wolfskin, DaWanda	Abmahnung der Social Commerce Plattform DaWanda und deren Nutzer wegen Verwendung eines Tatzensymbols welches dem Logo von Jack Wolfskin ähnelt.
2009	JAKO, Abmahnung	Abmahnung des Sportbekleidungsherstellers JAKO gegen einen Blogger, der das neue Logo des Unternehmens kritisiert hatte.
2009	Domino's Pizza, Video	Zwei Angestellte veröffentlichen auf YouTube ein Ekel-Video in dem sie sich die Zutaten der Pizza vor der Zubereitung in die Nase stecken - die Geschichte wird schnell auch in deutschen Medien aufgegriffen.
2009	Amazon, Amazonfail	Das Entfernen von Literatur mit homosexuellen Inhalten aus Suche und Rankings von Amazon löste einen Empörungssturm im Social Web aus. #Amazonfail wird auf Twitter bis heute bei jedem Serviceproblem von Amazon referenziert.
2009	Otto, MacBook, 50 Euro	Aufgrund eines Fehlers waren über otto.de Apple MacBooks zum Preis von 49,95 statt für knapp 1700 Euro zu haben. Die Nachricht verbreitete sich schnell und Otto entstanden einige hunderttausend Euro Schaden.

3.4.5 Weiterführende Literatur

Beal, A.; Strauss, J. (2008): Radically Transparent: Monitoring and Managing Reputations Online. Wiley Publishing, Inc., Indianapolis, Indiana.
Einstieg in das Thema Online-Reputation mit Beispielen und Fokus auf dem amerikanischen Markt.

Brauckmann, P. (Hg.) (2010): Web-Monitoring. Gewinnung und Analyse von Daten über das Kommunikationsverhalten im Internet. UVK Verlagsgesellschaft mbH, Konstanz.
Sammelband. Speziell hervorzuheben ist die wissenschaftliche Betrachtung von Web-Monitoring-Lösungen durch A. Plum.

ENGAGEMENTdb: The world's most valuable brands. Who's most engaged? Ranking the Top 100 Global Brands. Studie veröffentlicht Juli 2009 von wetpaint und Altimenter
http://www.engagementdb.com/downloads/ENGAGEMENTdb_Report_2009.pdf. Zuletzt geprüft am 25.05.2010.
Die Website www.engagementdb.com rankt Unternehmen nach deren Aktivitäten im Social Web und präsentiert diese im Profil. Gute Inspirationsquelle für aktuelle Entwicklungen.

Gilliatt, N.; Social Target LLC (2010): Social Media Analysis Platforms for Workgroups. Online beziehbar unter http://www.socialtarget.com/research/sma-platforms.html. Zuletzt geprüft am 25.05.2010.
Kostenpflichtige Studie über IT-Werkzeuge zum »Social Media Monitoring« in größeren Unternehmen. Focus auf amerikanischen Markt.

Li, C.; Bernoff, J. (2008): Groundswell: winning in a world transformed by social technologies. Forrester Research, Inc., Harvard Business School Publishing.
Standardwerk mit vielen Beispielen zu Social Web und den Möglichkeiten für Unternehmen, dieses zu nutzen. Basierend auf Marktforschungsdaten von Forrester Research wird der »Groundswell« strukturiert und greifbar.

Schwede, M. (2010): Die Plattformen für das Social-Media-Monitoring. Marktübersicht. Marketing & Kommunikation 3/10, S. 30-31, ISSN 1023-8913.
Fachartikel, in dem einige IT-Werkzeuge zum »Social Media Monitoring« kurz vorgestellt werden.

4 Automatisierung der Informationsakquise

In den beiden vorangegangenen Kapiteln wurden grundlegende Informationsbedürfnisse insbesondere des strategischen Innovationsmanagements betrachtet und zu den wichtigsten im Internet verfügbaren Informationsquellen in Beziehung gesetzt. Im professionellen Umfeld herrschen vielfach langfristige Informationsbedürfnisse vor, die sich nicht mit herkömmlichen, dokumentenorientierten Suchmaschinen und mitunter auch nur unzureichend durch fachlich fokussierte Web-Datenbanken befriedigen lassen. Zu den Anforderungen gehören

- der Umgang mit einer Vielzahl unterschiedlicher Informationsquellen bei gleichzeitiger Ausnutzung der Spezifika des jeweiligen Informationstyps;
- die effektive Miteinbeziehung von Benutzerwissen in den Informationsakquiseprozess, etwa indem Benutzer Websites, die ihnen wichtig sind, einbinden, und solche, die ihnen explizit nicht wichtig sind, ausschließen können;
- Unterstützung von Informationsbedürfnissen, die nicht durch das Auffinden bestimmter Dokumente befriedigt werden können, sondern durch die gezielte Analyse der Ergebnismenge. Typische Anwendungsfälle sind z. B. Trend Monitoring oder Zitationsanalysen;
- Integration und Kollaboration: Suchergebnisse oder sonstige mittels Internetrecherche gewonnene Erkenntnisse sollen mitunter in bestehende Geschäftsprozesse eingebunden werden, hierzu sind entsprechende Schnittstellen nötig.

Die genannten Punkte stehen lediglich exemplarisch für eine Vielfalt von Anforderungen, die über das Leistungsspektrum »normaler« Suchmaschinen hinausgehen und von modernen Web Mining- und Web Intelligence-Werkzeugen adressiert werden.

4.1 Web Intelligence- und Web Mining-Prozess

In der Literatur werden die Begriffe »Suche«, »Mining« und »Intelligence« oftmals vermischt und mehr oder weniger synonym verwendet, obgleich sie eigentlich unterschiedliche Konzepte beschreiben. Auf eine detaillierte Diskussion der unterschiedlichen Interpretationen dieser Begriffe soll an dieser

Stelle verzichtet werden. Im Folgenden erfolgt stattdessen lediglich eine kurze Abgrenzung der Begrifflichkeiten im Kontext dieser Arbeit. [103]

- Unter »Web Searching« verstehen wir die im Allgemeinen stichwortgestützte Suche nach Dokumenten im Internet mithilfe einer Suchmaschine.

- Im Gegensatz hierzu werden beim »Web Mining« Verfahren des Text- bzw. Data Minings angewandt, um aus der Gesamtmenge der als relevant eingestuften Dokumente weitergehende Informationen zu erschließen. Üblicherweise unterscheidet man verhaltensgetriebene Analysen (»Web Usage Mining«), also die Erkundung des Verhaltens der Websitebesucher, und inhaltsgetriebene Analysen (»Web Content Mining«), also die zielgerichtete Analyse der Inhalte (üblicherweise fremder) Websites. Die nachfolgenden Betrachtungen beziehen sich auf letzteren Aspekt.

- »Web Intelligence« kombiniert Verfahren des Web Searchings und des Web Minings mit zusätzlichen Mehrwertdiensten z. B. der Dateninterpretation, Verfahrensautomatisierung, des Reportings und des Informationsmanagements.

Dabei basieren Werkzeuge für Web Intelligence üblicherweise auf Verfahren des Web Minings, die wiederum naturgemäß auch einfache Web-Suchen einsetzen (vgl. Abbildung 38).

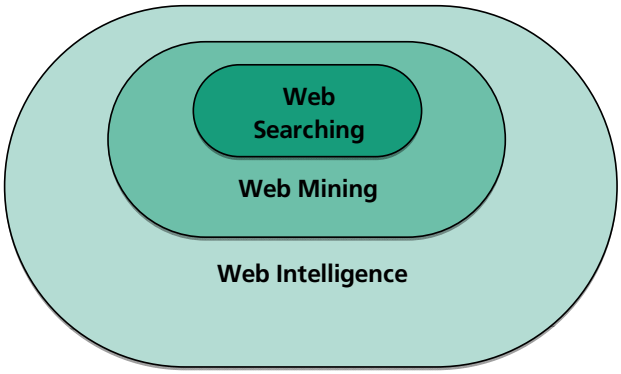


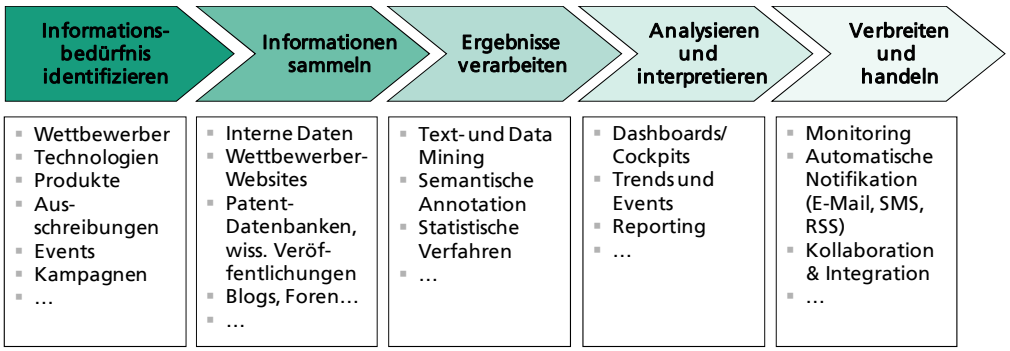
Abbildung 38: Begriffs-abgrenzung »Web Searching«, »Web Mining« und »Web Intelligence«

[103] Es sei aber ausdrücklich darauf hingewiesen, dass in der Literatur durchaus andere Definitionen der Begriffe verwendet werden. Zu einer weiterführenden Betrachtung der Terminologie sei auf Zhong, Liu und Jiming (2003) verwiesen.

Web Mining wird oftmals auch definiert als die Anwendung von Text Mining-Verfahren auf Web-basierte Dokumente (vgl. Kosala und Blockeel, 2000). Zwar greift diese Definition insofern zu kurz, als dass grundsätzlich auch die Analyse multimedialer Inhalte wie Audio-Streams, Bilder und Videos Gegenstand des Web Minings sein können. Da der Großteil unternehmensstrategisch relevanter Informationen jedoch in textueller Form vorliegt (vgl. Hippner und Rentzmann, 2006), stellen Text Mining-Module tatsächlich die wichtigsten Analysekomponenten der meisten verfügbaren Web Intelligence-Werkzeuge dar. Und auch bei der Analyse von multimedialem Content kommen oftmals Text Mining-Verfahren zur Anwendung: Aus Bild- und Videodaten werden über eine automatische Schrifterkennung Texte extrahiert, und Audio-Streams werden mittels automatischer Spracherkennung transkribiert (z. B. Autonomy Virage[104]).

Die im vorigen Kapitel beschriebenen Informationsbedürfnisse von »Innovation Professionals« lassen sich ganzheitlich nur mit fortgeschrittenen Web Intelligence-Lösungen begegnen. Abbildung 39 zeigt den typischen Ablauf eines Web Intelligence-Projektes:

Abbildung 39: Typischer Web Intelligence-Prozess



1. *Informationsbedürfnis identifizieren*: Obgleich die aus einem Mining-Prozess gewonnenen Erkenntnisse überraschend sein können (letztendlich zielt die Idee von Mining-Prozessen ja auf die Entdeckung zuvor unbekannter Tatsachen), sollte der Prozess so zielgerichtet wie möglich geplant werden. Das heißt, die übergeordneten strategischen Ziele sollten im ersten Schritt definiert werden, da diese sämtliche Folgeschritte wie Quellenauswahl und Visualisierungsspezifikation beeinflussen.

[104] http://www.virage.com

2. *Informationen sammeln*: Je nach Informationsbedürfnis werden bestimmte Informationsquellen selektiert. Falls das Informationsbedürfnis Erkenntnisse über zeitliche Entwicklungen einschließt, wie etwa Trend-Analysen zu einem bestimmten Thema, so muss der Dokumentenkorpus über einen längeren Zeitraum erstellt werden. Oftmals erfordern Web Mining-Anwendungsfälle, das Informationen so früh wie möglich gewonnen werden, also idealerweise in Echtzeit – sobald entsprechende Daten im Web veröffentlicht werden. Dies erfordert entweder einen zeitlich entsprechend eng konfigurierten Crawl- und Scrape-Ansatz oder (noch besser) eine Informationsversorgung nach dem Push-Prinzip, etwa über RSS-Feeds.

3. *Ergebnisse verarbeiten*: Nachdem der Quellenkorpus festgelegt wurde (er wird höchstwahrscheinlich im Laufe des Mining-Projektes weiter wachsen) beginnt die Ergebnisverarbeitung. Obgleich die zur Anwendung kommenden Verfahren stark variieren können, liegt die Hauptaufgabe dieses Schrittes im Allgemeinen stets darin, ein spezifiziertes Informationsbedürfnis (etwa eine Suchanfrage) mit der Dokumentenmenge des Quellenkorpus abzugleichen. In navigationalen[105] Suchprozessen endet dieser Schritt normalerweise mit der Gewichtung der Relevanz der infrage kommenden Dokumente bzgl. der Suchanfrage, um ein passendes Ranking der Dokumente im nachfolgenden Schritt zu ermöglichen. Informationsbedürfnisse, die eher informationaler Natur sind, erfordern üblicherweise weitere Schritte der Informationsverarbeitung, wie etwa Metadaten-Extraktion, um passende Analysen im nachfolgenden Schritt zu erlauben.

4. *Analysieren und interpretieren*: Die im vorherigen Schritt gesammelten Informationen werden analysiert und verdichtet, und am Ende strukturell, textuell und/oder graphisch so aufbereitet, dass das Informationsbedürfnis bestmöglich befriedigt wird. Dieser Schritt zielt darauf ab, den Benutzer möglichst gut bei der Datenanalyse zu unterstützen und erfordert u. a. die Wahl der am besten passenden Visualisierungsform, diese Visualisierung bzgl. der jeweiligen Daten zu optimieren, sowie idealerweise zusätzliche Informationen bereitzustellen, die bei der korrekten Interpretation der Daten helfen können.

5. *Verbreiten und handeln*: Wenn interessante Ergebnisse gefunden oder Erkenntnisse gewonnen werden konnten, schließen sich ggf.

[105] Zur Unterscheidung zwischen »navigationaler« und »informationaler« Suche siehe (Stock und Lewandowski, 2006)

Folgeprozesse an: Weitere Dokumente und Informationen, die zu einem späteren Zeitpunkt veröffentlicht werden, können automatisch ergänzt werden und zu einer neuen Version der Ergebnisanalyse führen. Neue Ergebnisse können dem Nutzer über einen Notification-Mechanismus zugestellt werden. Ergebnisse und Ergebnisberichte werden gesichert, zu einem späteren Zeitpunkt wieder geladen, gedruckt, zur weiteren Verarbeitung in andere Systeme exportiert oder an andere Benutzer weitergegeben.

4.2 Softwaretechnische Unterstützung

Die softwaretechnische Unterstützung von Web Intelligence-Projekten im oben genannten Sinne birgt eine ganze Reihe technischer (und organisatorischer) Herausforderungen. Diese sollen im Folgenden diskutiert werden, wobei wir uns strukturell wiederum an dem im vorigen Abschnitt besprochenen Phasenmodell orientieren.

Betriebsmodell

Eine wichtige Entscheidung bei der Softwareauswahl liegt im Betriebsmodell. Soll die Software im eigenen Haus betrieben werden oder auf ein ASP (Application Service Provider) - bzw. SaaS (Software as a Service) -Modell des Anbieters zurückgegriffen werden? Viele Hersteller unterstützen inzwischen beide Betriebsarten, bei der Entscheidung sollten folgende Aspekte berücksichtigt werden:

- Für den Eigenbetrieb sprechen insbesondere die größere Flexibilität und Aspekte der Datensicherheit.

- Für einen SaaS-Ansatz spricht der mitunter äußerst ressourcenfordernde Betrieb einer Web Intelligence-Lösung. Die Aktualisierung großer Informationsmengen bedeutet mitunter eine hohe Last für die Internetverbindung, und die Analyse großer Datenmengen erfordert eine hohe Rechenleistung.

Informationsbedürfnis identifizieren

Neben eher allgemein gehaltenen Web Intelligence-Werkzeugen, die prinzipiell jedwede Art der Analyse von Webinhalten unterstützen, haben sich viele Anbieter auf die Beantwortung bestimmter Informationsbedürfnisse

spezialisiert oder erlauben den Zuschnitt ihrer Werkzeuge durch Plug-in-Konzepte (z. B. bietet der Anbieter Temis[106] sogenannte »Skill Cartridges« für sein System Luxid an). Die Beschränkung der Einsatzmöglichkeiten der Software wird dabei üblicherweise zugunsten einer geringeren Anwendungs- und Konfigurationskomplexität in Kauf genommen. Viele Anbieter von Softwarelösungen unterstützen potenzielle Kunden im Vorfeld des Kaufs durch entsprechende Beratungsleistungen.

Informationen sammeln

Die Zusammenstellung der zu analysierenden Informationsmenge ist ein besonders wichtiger Schritt in jedem Web Intelligence-Projekt, denn die Qualität (und Quantität) der Eingangsdaten beeinflusst die Qualität der Analyseergebnisse unmittelbar. Werden wichtige Dokumente nicht erfasst, so sind die Analyseergebnisse im glücklichen Fall unvollständig, im schlimmsten Fall sind sie sogar falsch. Vor der Beschaffung einer Software sollte daher die Art der Informationssammlung bezüglich der folgenden Dimensionen bewertet werden:

- Werden die wichtigsten Quellen bereits durch den Anbieter erfasst? Viele Anbieter werben mit einer großen Menge vorerfasster Quellen oder Dokumente. Allerdings sind rein quantitative Angaben nicht leicht zu überprüfen, schwierig zu vergleichen und schwer zu bewerten. Wichtiger als die reine Menge der Quellen ist hingegen die Prüfung, ob die für das eigene Unternehmen relevanten Quellen erfasst werden.

- Lassen sich neue Quellen auf einfache Weise ergänzen? Da es unwahrscheinlich ist, dass ein Anbieter wirklich alle relevanten Quellen vorerfasst hat und zudem durch die hohe Volatilität des Webs (vgl. Kapitel 2) ständig neue Quellen hinzukommen und bestehende verschwinden, ist eine Modifizierung der Informationsquellenmenge eine wichtige Voraussetzung für den langfristig erfolgreichen Einsatz. Während viele Werkzeuge die Ergänzung neuer Quellen auf einfache Weise durch den Benutzer erlauben, erfordert es bei anderen den Eingriff des Anbieters.

- Wie aktuell ist der Informationskorpus, wie oft werden bestehende Informationsquellen auf neue oder geänderte Inhalte untersucht?

- Können Inhalte des sogenannten »Deep Webs« erfasst werden? Darunter versteht man Dokumente oder Informationen, die nicht

[106] http://www.temis.com

mithilfe generischer Suchmaschinen erfasst werden können. Beispiele sind passwortgeschützte Sektionen oder online-zugreifbare Datenbanken, die das Ausfüllen von Formulardaten vorsehen.

- Welche Dokumentformate werden verarbeitet? Neben HTML-Seiten liegen unternehmensstrategisch relevante Informationen häufig als PDF- oder Office-Dokumente vor. Auch multimediale Inhalte wie Audiopodcasts oder Videos gewinnen an Wichtigkeit.

- Werden Schnittstellen zu professionellen Datenbanken angeboten und deren Datenaustauschformate unterstützt? Beispiele sind Factiva, Thomson, LexisNexis etc.

- Können eigene (interne) Datenquellen eingebunden werden, wie etwa E-Mail- oder Dokumentenserver? In diesem Zusammenhang ist vor allem auch das geeignete Betriebsmodell (s. o.) zu berücksichtigen.

Ergebnisse verarbeiten

Da die Informationen in Texten, anders als z. B. in Datenbanken im Wesentlichen unstrukturiert vorliegen, ist eine mitunter umfangreiche Vorverarbeitung nötig, bevor Mining-Algorithmen Anwendung finden können. Zu diesen Aufgaben gehören u. a. die Erkennung von Datenformat und Zeichenkodierung, der Textsprache und ggf. Subsprache (z. B. Deutsch (Ch)) sowie struktureller Gliederungen (Titel, Überschriften usw.). Statistische Verfahren der Textanalyse operieren üblicherweise auf den einzelnen Bestandteilen eines Textes, also etwa seinen Wörter oder Phrasen. Diese werden als »Features« bezeichnet und spannen für jeden Text einen »Feature-Raum«. Wesentliche Featureklassen, die im Text Mining Anwendung finden, sind Buchstaben, Wörter, Terme und Konzepte (vgl. Feldman und Sangler, 2007). Dabei verfügen Terme und Konzepte über das höchste semantische Erschließungspotenzial und sind somit für »höherwertige« Analysen am besten geeignet. Um auf Termen und insbesondere Konzepten operieren zu können, bedarf es allerdings einer entsprechend sorgfältigen Vorverarbeitung der Texte. Diese umfasst zumeist die Interpretation der Texte auf unterschiedlichen linguistischen Abstraktionsebenen (morphologisch, syntaktisch, semantisch). Tabelle 14 fasst die wichtigsten linguistischen Verarbeitungsschritte zusammen.

Tabelle 14: Linguistische Verarbeitungsschritte

	Beschreibung	Grund
Tokenisierung	Aufteilung des Textes in Sätze, Phrasen und Wörter.	Die identifizierten Einheiten können in den weiteren Schritten separat annotiert und analysiert werden. Wichtig ist hier vor allem die korrekte Interpretation der Verwendung von Interpunktionen sowie die Erkennung von Mehrwortlexemen.
Flexionsanalyse, Stammform-reduzierung	Die konkreten Wortformen werden auf ihren jeweiligen Wortstamm abgebildet.	Komplexität der Analyse wird verringert: Unterschiedliche morphologische Ausprägungen desselben Begriffs werden unifiziert. Beispiel: Haus vs. Häuser vs. Hauses.
Stoppwort-Entfernung	Nicht inhaltsrelevante Wörter wie Artikel, Konjunktionen oder Hilfsverben werden herausgefiltert.	Die Komplexität der Analyse wird verringert.
Part-of-Speech-Tagging, syntaktisches Parsing	Analyse von Wortformen, Annotation der Wortformen und Konsituenten mit ihren Wortarten (Verb, Nomen bzw. Verbalphrasen, Nominalphrase usw.).	Die Kenntnis der grammatische Funktion der Einheiten erlaubt die Anwendung von Textextraktionsregeln.
Semantische Analyse	»Konzeptionalisierung« des Inhalts - die konkreten sprachlichen Wörter werden auf Konzepte abgebildet, zu denen ggf. Beziehungswissen im System hinterlegt ist (Ontologie).	Trägt zur Lösung von Synonomie- und Polysemieproblemen.
Eigennamen-erkennung	Identifikation von Personen- und Firmennamen, Geographischen Informationen etc.	Metadaten können für vielfältige Analysen genutzt werden, z.B. Beziehungsanalysen, Kompetenzclusteridentifikationen usw.

Wie bereits in Kapitel 2 festgestellt wurde, ist die im Internet dominierende Sprache Englisch – wobei die Menge der in anderen Sprachen (insbesondere Chinesisch und Arabisch) veröffentlichten Informationen weitaus schneller ansteigt. Bei der Auswahl eines Web Intelligence-Werkzeuges ist die Unterstützung der für das Unternehmen relevanten Sprachen zu beachten – die am Markt verfügbaren Werkzeuge unterscheiden sich in diesem Punkt erheblich. Zwar geben einige Anbieter an, auf rein statistische und somit sprachunabhängige Verfahren zurückzugreifen - ob diese Verfahren jedoch eine vergleichbare Qualität zu einer dedizierten Sprachunterstützung mittels linguistischer Regeln bieten, darf bezweifelt werden.

Auf den aus den Texten extrahierten und mit Annotationen versehenen Features können nun statistische und regelbasierte Analyseverfahren angewendet werden. Gängige Ziele dieser Analysen sind:

- *Klassifizierung und Clustering*: Die Dokumente werden den Klassen einer vorgegebenen Klassifikation oder Taxonomie zugeordnet. Ist die Klassifikation nicht vorgegeben, sondern selbst Ergebnis des Mining-Prozesses, so spricht man auch von Clustering oder Segmentierung.

- *Themenerkennung und –verfolgung (Topic Detection and Tracking)*: Nutzt Methoden des Text-Clusterings zum Entdecken wichtiger Themen in Dokumentenkollektionen. Ziel ist das Aufspüren von »Trends«, z. B. durch Extrapolation der im Verlauf einer Zeitspanne gemessenen Daten in die Zukunft. Im einfachsten Fall wird die Anzahl von Dokumenten,

die einem Thema zugeordnet werden können, über einen Zeitraum gezählt.

- *Ereigniserkennung (Event Detection)*: Wird beim Topic Tracking ein signifikanter Anstieg oder Abfall gemessen, so ist dies möglicherweise auf ein bestimmtes Ereignis zurückzuführen. Ziel der Event Detection ist, diese Ereignisse frühzeitig zu erkennen und nach Möglichkeit auch zu benennen.

- *Stimmungserkennung (Sentiment Detection)*: Die Analyse von Stimmungen in Texten hat insbesondere durch die steigende Popularität des benutzergenerierten Inhalts des Social Webs an Bedeutung gewonnen (vgl. Kapitel 3). Viele Tools bieten heute Komponenten, um Stimmungstrends zu einem Thema (z. B. einem Markennamen) zu analysieren.

- *Informationsextraktion*: Es wird versucht, bestimmte Informationen aus dem Text zu extrahieren. Ein gängiges Beispiel ist die Extraktion von Eigennamen (»Named Entity Recognition«, NER), bei der z. B. Personen- und Firmennamen identifiziert werden. Solche Metadaten können auf vielfältige Weise verwendet werden, etwa, um Experten zu identifizieren oder gezielt nach Kampagnen von Konkurrenten zu suchen.

- *Abhängigkeitsanalyse*: Es wird versucht, Relationen zwischen verschiedenen Objekten zu identifizieren.

- *Zusammenfassungen*: Es wird versucht, die wichtigsten Aussagen des Textes in kurzer Form zusammenzufassen.

Analysieren und interpretieren

Eine ganz wesentliche Herausforderung von Web Intelligence-Werkzeugen ist die Unterstützung des Anwenders bei der Analysedurchführung sowie die anwendergerechte Darstellung der Analyseergebnisse. Von der angemessenen Anwendung der Analysen und der korrekten Bewertung der Analyseergebnisse durch den Benutzer hängt letzten Endes der gesamte Wert eines Web Mining-Projektes ab. Daher sind die durch die Werkzeuge angebotenen Interaktions- und Visualisierungsansätze von entscheidender Bedeutung bei der Auswahl einer entsprechenden Software. Gängige Visualisierungen sind:

- *Kategorien*: Dokumente werden vorgegebenen oder automatisch erkannten Kategorien zugeordnet und dem Benutzer in einer übersichtlichen Struktur präsentiert. Die Darstellung erfolgt meist in

hierarchischen Strukturen, etwa Navigationsbäumen (ähnlich der Darstellung im Windows-Explorer) oder TreeMaps.

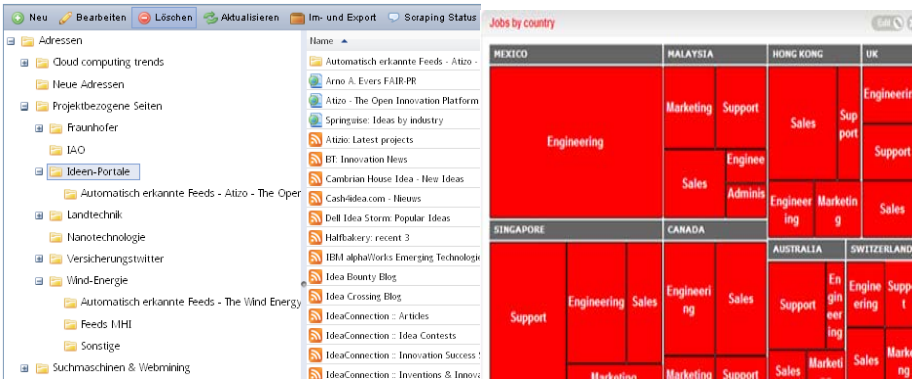


Abbildung 40: Typische Visualisierungen von Klassifikationen

Links mittels Navigationsbaum (Fraunhofer IAO Innovation Mining Cockpit). Rechts mittels TreeMap (Digimind D7)

- *Cluster*: Als »ähnlich« eingestufte Dokumente werden in einer Gruppe zusammengefasst dargestellt. Gängige Darstellungen sind Landkarten-artige Visualisierungen und hyperbolische Bäume[107].

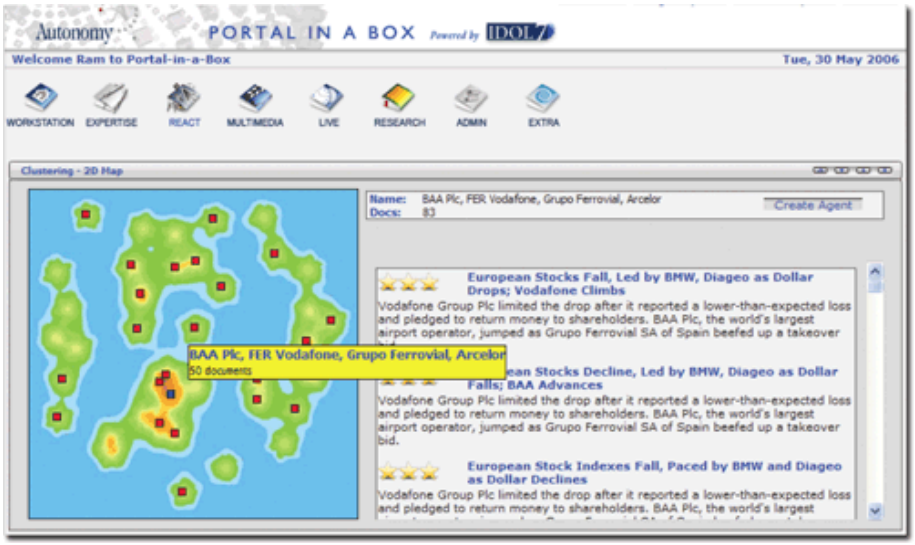


Abbildung 41: Typische Darstellung einer Klassifikation anhand einer Themenlandkarte

(Autonomy IDOL).

- *Tag Clouds*: Als relevant eingestufte Begriffe werden in Form einer Wolke dargestellt, wobei die errechnete Relevanz eines Begriffs in der

[107] Hyperbolische Darstellungen helfen dabei, große Informationsmengen explorierbar zu machen, indem die Objekte (also z. B. die Knoten eines Baums) desto kleiner dargestellt werden, je weiter sie vom Darstellungszentrum entfernt sind.

Stärke seiner Hervorhebung ausgedrückt wird – etwa durch die Schriftgröße oder Farbwahl.

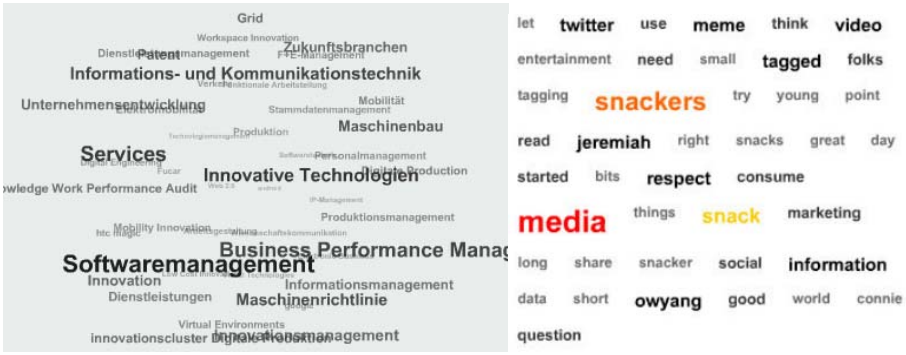


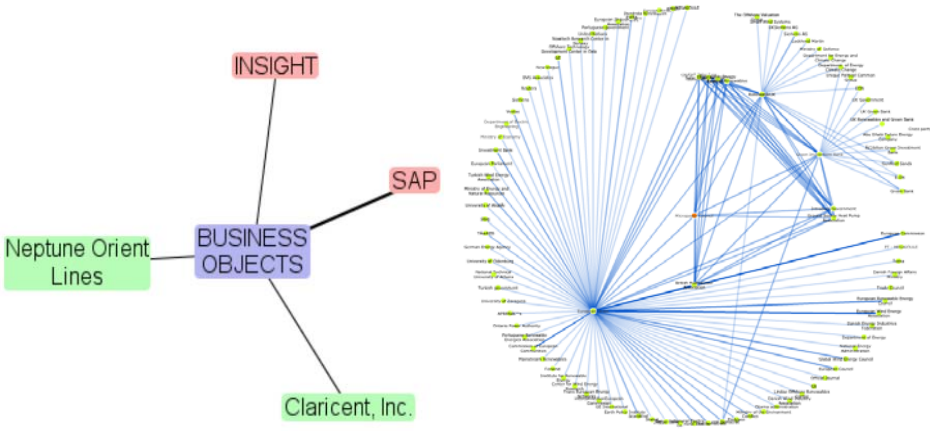
Abbildung 42: Beispiele von Tag Clouds.

Links: Animierte Tag Cloud (Fraunhofer IAO-Blog). Rechts: Farbige Hervorhebung von Tags (Radian 6)

- *Assoziationsgraphen*: Werden verwendet, um Beziehungen zwischen Objekten (z. B. Firmen oder Personen) zu visualisieren. Üblicherweise stellen dabei die Knoten des Graphen die Objekte dar, während die Kanten die Relationen zwischen den Objekten symbolisieren. Verschiedene Layoutalgorithmen helfen, Strukturinformationen des Graphen sichtbar zu machen. Beispielsweise kann der Vernetzungsgrad von Personen ein Indiz für ihre »Wichtigkeit« in einem Netzwerk darstellen. Abbildung 43 zeigt zwei Beispiele von Assoziationsgraphen, die mit dem Fraunhofer IAO Innovation Mining Cockpit erstellt wurden, und Beziehungen zwischen Firmen visualisieren: Im linken Bild wurden auf Pressemitteilungen zur Firma Business Objects analysiert – die stärke Kantenbreite der Relation zu SAP erklärt sich durch die zum Analysezeitpunkt angekündigte Übernahme von Business Objects durch SAP. Das rechte Bild veranschaulicht den unterschiedlichen Vernetzungsgrad verschiedener Firmen, die im Zusammenhang mit dem Thema »Windenergie« genannt wurden.

Abbildung 43: Assoziationsgraphen. Firmenbeziehungen auf Basis von Pressemitteilungen

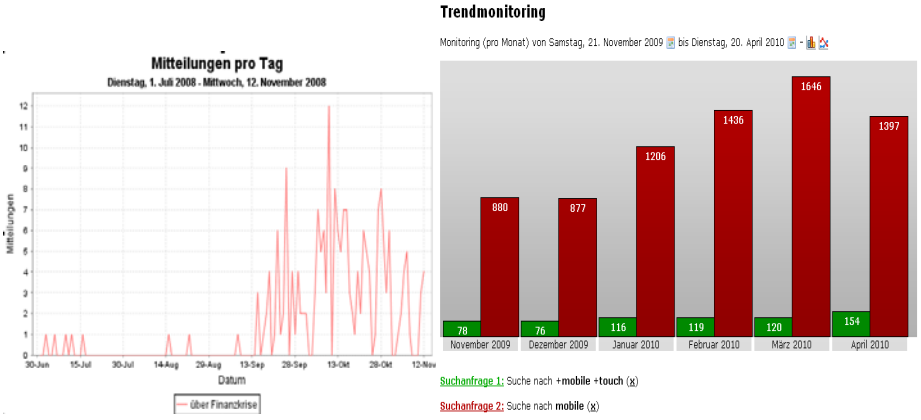
(Fraunhofer IAO Innovation Mining Cockpit)



- *Häufigkeitsvergleiche*: Trend Monitoring und Event Detection stellen wichtige Anwendungsfälle des Technologie- und Innovationsmanagements, aber auch der Wettbewerbsbeobachtung und des Reputationsmanagements dar. Bei der Visualisierung wird meist auf klassische Chart-Grafiken wie Balken- oder Liniendiagramme zurückgegriffen, mitunter kommen auch Wärmegrafiken zur Verwendung, die die Wichtigkeit bzw. Erwähnungshäufigkeit durch Farbnuancen darstellen.

Abbildung 44: Visualisierung von Häufigkeits-vergleichen.

(Fraunhofer IAO Innovation Mining Cockpit).



- *Geographische Informationen*: Häufig lassen sich aus Texten geographische Informationen relativ einfach extrahieren. Ländernamen stellen z. B. eine recht begrenzte (»geschlossene«) Wortklasse dar und erlauben eine einfache Identifikation mittels Look-Up-Listen (»Gazetteers«). Zu geographische Informationen lassen sich zudem

heutzutage leicht die Längen- und Breitengradangaben ermitteln (»Geocoding«), etwa über den entsprechenden Google-Maps-Dienst[108]. Damit lassen sich Objekte wiederum auf Landkarten darstellen – populär sind z. B. die in Abbildung 45 dargestellten Visualisierungen.

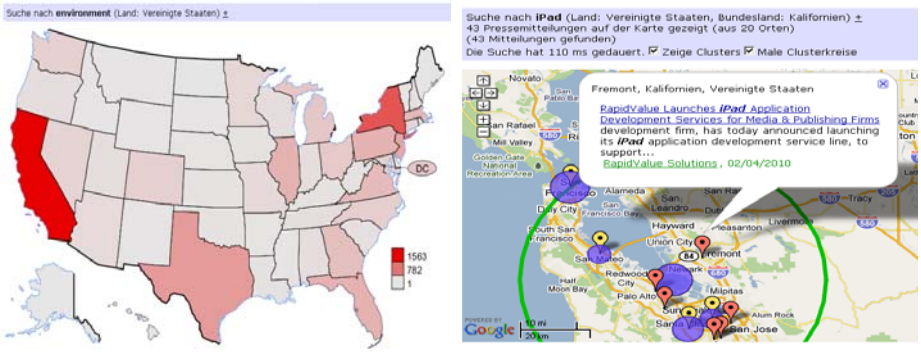


Abbildung 45: Visualisierung geographischer Informationen

(Fraunhofer IAO Innovation Mining Cockpit).

Neben der visuellen Datenaufbereitung und einer angemessenen Unterstützung bei der Datenexploration ist die Erzeugung von Analyseberichten, also das Reporting ein wichtiger Schritt in Web Intelligence-Projekten. Dies umfasst üblicherweise Dashboardsichten, die dem Benutzer die wichtigsten aktuellen Analyseergebnisse »auf einen Blick« präsentieren, so wie die Generierung von Ergebnisberichten in den gängigen Dateiformaten (Excel, PDF, HTML usw.).

Verbreiten und handeln

Web Intelligence-Projekte dienen dem Zweck, unternehmensstrategisch relevante Erkenntnisse zu gewinnen, sei es über technologische Entwicklungen, Aktivitäten der Konkurrenz oder die Zufriedenheit der eigenen Kunden. Oftmals folgt der Erkenntnis eine Handlung – etwa die Justierung der eigenen Forschungs- und Entwicklungsaktivitäten oder auch lediglich die Antwort auf einen unzufriedenen Blogeintrag. Die Wichtigkeit der Verbindung von Analyse und Aktion steigt mit der Komplexität des Handlungsumfeldes, also z. B. der eigenen Firmengröße: Gewonnene Erkenntnisse müssen an die richtige Abteilung oder den zuständigen Kollegen weitergeleitet oder an einer zentralen Stelle abgelegt werden. Dabei wird die Grenze zwischen Web Intelligence- und internem Dokumenten- bzw. Wissensmanagement zunehmend unschärfer. So wurde etwa NStein Technologies[109], ein Spezialist für Textanalysesysteme im April 2010 durch den Enterprise-Content-Management-System-Anbieter

[108] http://code.google.com/apis/maps/documentation/javascript/v2/services.html#Geocoding
[109] http://www.nstein.com

OpenText[110] übernommen[111]. Bei der Entscheidung für oder gegen ein bestimmtes Web Intelligence-Produkt sollten also folgende Aspekte berücksichtigt werden:

- Werden wichtige Kollaborations-Features unterstützt? Erlaubt das System das Teilen von Dokumenten, Reports oder auch komplexen Einstellungen mit Kollegen? Gibt es ein entsprechendes Rollen- und Berechtigungskonzept?

- Werden Schnittstellen zu eigenen Systemen angeboten, etwa Wissens- oder Dokumentenmanagementsystemen?

Da Informationsbedürfnisse im professionellen Bereich oftmals eher langfristiger Natur sind, ist die Möglichkeit der Speicherung der Such- und Analysevorgänge ein sehr wichtiges Feature. Einige Systeme verfügen zudem über Automatisierungsmöglichkeiten, die es erlauben Suchen und Analysen in bestimmten Abständen zu wiederholen und den Benutzer bei Auftreten als relevant eingestufter Ereignisse zu benachrichtigen (etwa per Hinweis am Dashboard oder per E-Mail/SMS).

4.3 Web Mining-Werkzeuge im Vergleich

In diesem Abschnitt sollte ursprünglich eine tabellarische Übersicht zu aktuell angebotenen Web Intelligence-Produkten folgen. Während der Recherchen wurde jedoch schnell klar, dass das Angebot einer sehr starken Fluktuation unterworfen ist: Auf der einen Seite werden die verfügbaren Produkte laufend um neue Funktionen ergänzt, auf der anderen Seite ändern und erweitern die Anbieter Ihre Produktpalette recht häufig. Produkte werden umbenannt oder neu geschnitten, indem etwa aus einem komplexen Webmining-Werkzeug eine eigenständige Social Media Monitoring-Lösung hervorgeht.

Daher haben wir uns entschlossen, an dieser Stelle auf eine Übersicht zu verzichten und möchten stattdessen auf unsere Website http://www.innovation-mining.net verweisen. Dort stellen wir laufend aktuelle Informationen zu Herstellern und Produkten zur Verfügung.

Für das Unterthema Social Media Monitoring möchten wir zudem auf die im Herbst 2010 erscheinende Fraunhofer-Marktstudie »Social Media Monitoring Tools« (Kasper, Dausinger und Kett, 2010) verweisen.

[110] http://www.opentext.com.
[111] http://www.opentext.com/2/global/press-release-details.html?id=2340

4.4 Weiterführende Literatur

Bügel, Ulrich; Knaf, Hagen; Laufs, Uwe: Technologieentwicklung : theoretische und praktische Unterstützung mit Data Mining. In: Knoll, Matthias (Hrsg.) ; Meier, Andreas (Hrsg.): Web & Data Mining. Heidelberg : dpunkt.-Verl., 2009, S. 21-34.
Der Aufsatz beschreibt Ansätze, Technologieentwicklungsprojekte mithilfe von Textminingverfahren zu unterstützen, die in einem Präsidialprojekt der Fraunhofer-Gesellschaft entwickelt und erprobt wurden.

Feldman, R. und Sangler, J. (2007): The Text Mining Handbook – Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press, New York, 2007.
Gut verständliches Grundlagenwerk zum Thema Textmining. Neben den gängigen Verfahren zu Klassifikation, Clustering usw. werden anschauliche Beispiele der praktischen Anwendung gegeben.

Kasper, H., Dausinger, M. und Kett, H. (2010): Marktstudie »Social Media Monitoring Tools«. Fraunhofer IRB, 2010.
Fraunhofer IAO hat im Rahmen einer unabhängigen Studie Software-Lösungen zur Sammlung und Analyse von Informationen im Social Web untersucht und rund 20 Angebote für den deutschen Markt im Profil dargestellt. Außerdem werden in der Studie Anwendungsfelder und Funktionen solcher Angebote sowie allgemein die Herausforderungen und Chancen von nutzergenerierten unternehmensrelevanten Informationen im Internet diskutiert.

Hippner, H. und Rentzmann, R. (2006): Text Mining. *Informatik-Spektrum.* Volume 29, Number 4 / August 2006. S. 287-290.
Kurze Einführung in das Themengebiet Textmining. Auch enthalten im Online-Lexikon der Gesellschaft für Informatik: http://www.gi-ev.de/no_cache/service/informatiklexikon/informatiklexikon-detailansicht/meldung/text-mining-137.html.

Kosala, R. und Blockeel, H. (2000), Web Mining Research: A Survey, SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, ACM, Volume 2, Issue 1, o.O., S. 1-10.
Dieses Paper zeichnet ein umfassendes Bild der Forschungsfragen rund um das Thema Web Mining. Obgleich bereits zehn Jahre alt, sind die viele der behandelten Fragen und Probleme auch heute noch aktuell.

Manning, C. D., Raghavan, P. und Schütze, H. (2007): An Introduction to Information Retrieval, Cambridge University Press, Cambridge, England.
Umfassende und sehr populäre Darstellung der Grundlagen des Information Retrieval.

Porter, A.L. und Cunningham, S. W. (2005): Tech mining. Exploiting new technologies for competitive advantage. Hoboken, NJ: Wiley (Wiley series in systems engineering and management).
Beschreibt Ansätze der systematischen automatisierten Analyse von Patentdaten und wissenschaftlicher Literatur zur Generierung strategisch relevanten Wissens.

Yang, Y.; Akers, L.; Klose, T.und Yang, C. B. (2008): Text mining and visualization tools - Impressions of emerging capabilities, World Patent Information, Volume 30, Issue 4, December 2008, S. 280-293.
Zusammenfassung von Ergebnissen eines Vergleichs von Textmining- und Visualisierungswerkzeugen bzgl. der Eignung für Innovationsmanagement und Wettbewerbsbeobachtung der Patentanalyse-Gruppe des Pharma-Unternehmens Bristol-Myers Squibb.

Zhong, N; Liu, J; und Yao, Y. (2003): Web Intelligence. Springer.
Umfassende Aufsatzsammlung zum Thema »Web Intelligence«, das in diesem Werk allerdings etwas breiter ausgelegt wird, als im vorliegenden Leitfaden. Neben der Wissensakquisition aus Internetquellen (»Web mining and Farming«, »Web information retrieval«) werden Themen wie »Web-based education« und »Social Network Intelligence« in eigenen Kapiteln behandelt.

Website des Web Intelligence Consortiums: http://wi-consortium.org.
Website der internationalen Non-Profit-Organisation, die sich der Forschung und industriellen Entwicklung der Web Intelligence widmet.

The Easy Bee Directory of Web 3.0 Software: http://www.theeasybee.com.
Die Website bietet ein umfangreiches Software-Verzeichnis und enthält u. a. die Kategorien »Web Content Extraction«, »Business Intelligence«, »Knowledge Management« und »Social Web Tools«.

5 Fazit

Die Nutzung generischer Suchmaschinen als alleiniges Mittel der Informationsrecherche im professionellen Umfeld erweist sich in zunehmendem Maße als unzureichend, da vermehrt unternehmensstrategisch relevante Informationen in Quellen veröffentlicht werden, die sich dem Zugriff der Suchmaschinen entziehen. Für die effiziente Informationsrecherche im professionellen Umfeld ist es daher von großer Bedeutung, über eine hohe Quellenkompetenz zu verfügen, also zu wissen, wo und wie relevante Informationen gefunden werden können. Ein Ziel der vorliegenden Arbeit war es daher, die für das strategische Innovationsmanagement besonders relevanten Arten von Informationsquellen zu benennen und Möglichkeiten der effizienten Nutzung vorzustellen.

Die Informationsbedürfnisse professioneller Benutzer unterscheiden sich darüber hinaus in vielfältiger Weise von denen privater Suchmaschinenbenutzer – insbesondere werden oftmals weniger einzelne Dokumente gesucht, sondern Informationen, die aus einer großen Menge von Dokumenten extrahiert oder interpoliert werden müssen. Auch die Ansprüche an Analysen und Mittel der Weiterverarbeitung übersteigen die Angebote heutiger Suchmaschinen bei Weitem. Zweites Anliegen der vorliegenden Arbeit war es daher, Alternativen aufzuzeigen. Neben spezialisierten Suchwerkzeugen für bestimmte Informationsarten wie Patentdaten und wissenschaftliche Veröffentlichungen kann für die professionelle Informationsakquisition auf äußerst mächtige Web Intelligence-Werkzeuge zurückgegriffen werden. Diese verknüpfen etablierte Mechanismen der Analyse unstrukturierter Informationen (»Text Mining«) mit vorgelagerten automatisierten Verfahren der Datensammlung und -aufbereitung und den Möglichkeiten eines unternehmensweiten Informationsmanagements (Reporting, Kollaboration Wissensmanagement).

Das Social Web lässt die Benutzer vom Informationskonsumenten mehr und mehr zum Informationsproduzenten werden (»Prosumer«). Es ermöglicht ein noch nie da gewesenes Maß an Interaktion und damit riesige Chancen der Einbindung von Benutzerwissen, -erfahrungen und –meinungen in unternehmensstrategische Entscheidungen. Gleichzeitig erfordert es aber auch eine klare Strategie der aktiven Teilnahme seitens der Unternehmen, denn die wachsende Macht der Kunden birgt auch hohe Risiken für die Reputation eines Betriebs. Ein drittes Anliegen dieser Arbeit war es daher, das Bewusstsein für die Chancen und Risiken zu stärken, welche die neuen Möglichkeiten des Social Webs eröffnen.

Referenzen[112]

Baeza-Yates, R und Ribeiro-Neto, B. (1999*): Modern Information Retrieval*, pp 29-30. Addison Wesley / ACM Press, New York.

Beal, A. und Strauss, J. (2008): Radically Transparent: Monitoring and Managing Reputations Online. Wiley Publishing, Inc., Indianapolis, Indiana.

Brin, S. und Page, L. (1998): The Anatomy of a Large-Scale Hypertextual Web Search Engine. In: Computer Networks and ISDN Systems 1998

Chesbrough, H. W. (2006): Open innovation. The new imperative for creating and profiting from technology. Boston, Mass.: Harvard Business School Press.

Feldman, R. und Sangler, J. (2007): The Text Mining Handbook – Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press, New York

Finzen, J. Kintz, M., Kett, H. und Koch, S. (2009): Strategic Innovation Management on the Basis of Searching and Mining Press Releases. In: Proceedings of the Fifth International Conference on Web Information Systems and Technologies, Lisbon: INSTICC Press, S. 347–353.

Finzen, J., Kintz, M. und Kobes, S. (2010): A Comparative Study of Publicly Accessible Web-Based Idea Portals. In Proceedings of the XXI ISPIM Conference, Bilbao.

Finzen, J., Krepp, T. und Heubach, D. (2009): Web Searching in Early Innovation Phases - a Survey among German Companies. In: Proceedings of the 2nd ISPIM Innovation Symposium. New York.

Gassman, O. und Enkel, E. (2006): Open Innovation: Die Öffnung des Innovationsprozesses erhöht das Innovationspotenzial. In: Führung und Organisation, H. 3, S. 132–138.

[112] An dieser Stelle sind nur diejenigen Quellen aufgeführt, die im Text referenziert werden. In den einzelnen Abschnitten der Arbeit finden sich Hinweise auf weiterführende Literatur. Ein umfassendes Literatur-Verzeichnis findet sich zudem auf der Website http://innovation-mining.net.

Goos, P., und Hagenhoff, S. (2003): Strategisches Innovationsmanagement: Eine Bestandsaufnahme. In Schumann, M. (ed.), Arbeitsbericht Nr. 11/2003 des Instituts für Wirtschaftsinformatik der Georg-August-Universität Göttingen, Göttingen.

Hearst, M. A. (2008): UIs for Faceted Navigation: Recent Advances and Remaining Open Problems. In *HCIR08 Second Workshop on Human-Computer Interaction and Information Retrieval*. Microsoft.

Hippel, E. v. (1998). Economics of product development by users: The impact of `sticky' local information. Management Science, 44(5), S. 629-644.

Hippner, H. und Rentzmann, R. (2006): Text Mining. Informatik-Spektrum. Volume 29, Number 4 / August 2006. S. 287-290

Kleinberg, J. (1999): Authoritative sources in a hyperlinked environment. In: Journal of the ACM, vol. 36 nr. 5, S. 604-632.

Kosala, R. und Blockeel, H. (2000), Web Mining Research: A Survey, SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, ACM, Volume 2, Issue 1, o.O., Seite 1-10

Lewandowski, D. (2009): Handbuch Internet-Suchmaschinen. Nutzerorientierung in Wissenschaft und Praxis. Heidelberg: Akad. Verl.-Ges. AKA.

Lewandowski, D. und Mayr, P. (2006): Exploring the academic invisible web. In: Library Hi Tech, Jg. 24, H. 4, S. 529–539.

Li, C. und Bernoff, J. (2008): Groundswell: winning in a world transformed by social technologies. Forrester Research, Inc., Harvard Business School Publishing

Manning, C. D., Raghavan, P. und Schütze, H. (2007): An Introduction to Information Retrieval, Cambridge University Press, Cambridge, England.

Riedl, C.; May, N.; Finzen, J.; Stathel, S.; Kaufman, V.; Krcmar, H. (2009): An Idea Ontology for Innovation Management (forthcoming). *International Journal On Semantic Web and Information Systems (IJSWIS)* 5(4). 1–18.

Siwczyk, Y. (2009): IT-gestützte White-Spot-Analyse: Potenziale von Patentinformationen am Beispiel Elektromobilität erkennen. Fraunhofer Verlag. 2009.

Skibicki, K. und Mühlenbeck, F. (2008): Verbrauchermacht im Internet. 1. Auflage BrunoMedia Buchverlag.

Springer, S. (2006): Nutzung von Internet und Intranet für die Entwicklung neuer Produkte und Dienstleistungen. Herausgegeben von nova-net-Konsortium. Fraunhofer IRB. Stuttgart. (nova-net Werkstattreihe). Online verfügbar unter http://www.nova-net.de.

Stathel, S., Finzen, J., Riedl, C. und May, N. (2008): Service Innovation in Business Value Networks. *Proceedings of the XVIII International RESER Conference*. (Stuttgart, Germany). 288–302.

Stock, W. G. und Lewandowski, D. (2006). Suchmaschinen und wie sie genutzt werden. WISU 35(2006)8-9, S. 1078-1083.

Surowiecki, J. (2005): The wisdom of crowds. 1. ed. New York, NY: Anchor Books.

Walcher, D. (2007): Der Ideenwettbewerb als Methode der aktiven Kundenintegration: Theorie, empirische Analyse und Implikationen für den Innovationsprozess. Wiesbaden: Deutscher Universitäts-Verlag / GWV Fachverlage GmbH, Wiesbaden.

Wanner, L., Baeza-Yates R. et al. (2008): Towards content-oriented patent document processing. In: World Patent Information, Jg. 30, H. 1, S. 21–33.

Zerfaß, A. (2005). Corporate blogs: Einsatzmöglichkeiten und Herausforderungen. BIG BlogInitiativeGermany, S. 1-9.

Zhong, N; Liu, J; und Yao, Y. (2003): Web Intelligence. Springer.

Das Internet hat sich zur wichtigsten Informationsquelle des Innovationsmanagements entwickelt. Doch mit der Menge der verfügbaren Information steigt auch die Komplexität der Recherche – es ist schwer in der Datenflut nicht unterzugehen. Der vorliegende Leitfaden erläutert daher anhand grundlegender Anwendungsfälle des strategischen Innovationsmanagements und angrenzender Disziplinen wie Wettbewerbsbeobachtung und Reputationsmanagement, wie sich internetbasierte Informationsquellen effektiv nutzen lassen, um unternehmensstrategisch relevante Analysen durchzuführen. Neben den klassischen Online-Quellen wie Patentdatenbanken und Suchmaschinen für wissenschaftliche Veröffentlichungen widmet sich ein eigenes Kapitel den Möglichkeiten und Herausforderungen des Social Web sowie dem Open Innovation-Ansatz zur Kundenintegration. Die Arbeit enthält zudem aktuelle Informationen zu Anbietern und Produkten, welche die professionelle Informationsrecherche unterstützen.

ISBN 978-3-8396-0139-6
9 783839 601396