

## Kapitel 6

# Informationen in Portalen finden

### **In diesem Kapitel:**

Die Auffindbarkeit – Zwischen Suchen und Finden	229
Was und wie wird gesucht?	230
Informationsklassen	232
Warum es das perfekte Suchergebnis nicht gibt	233
Das Potenz- und das Zipfsche Gesetz	234
Der Faktor Mensch (Mooers-Gesetz)	235
Technische Limitationen/Federated Search	235
Die Bedürfnisse der Benutzer	236
Annäherung an die perfekte Suche	238
Indizierung von bestehenden Webseiten	253

Die Suche spielt in quasi allen Portalprojekten und somit auch bei SharePoint eine überaus große Rolle. Dabei geht es nicht nur um das reine Durchsuchen von Dokumenten, sondern in einem Großteil der Fälle um das Finden der richtigen Person, die zu dem gesuchten Thema Auskunft geben kann. In diesem Kapitel wird gezeigt, dass es eine perfekte Suche nicht geben kann, um sich dieser jedoch im späteren Verlauf des Kapitels durch Konfiguration von SharePoint anzunähern.

Wir halten viele Vorträge über die Suche im Zusammenhang mit Microsoft SharePoint Server 2010. Zu Beginn fragen wir jeweils, wer mit dieser Suche zufrieden sei. Was denken Sie, wie viele Teilnehmer halten wohl eine Hand nach oben? Sie werden es bereits vermuten: Niemand. Und diesen Effekt haben wir jedes Mal. Dass dies so ist, hat nicht zwingend etwas mit der eingesetzten Technologie zu tun. Meist hängt es mit dem gesamten Kontext des Portals zusammen. Genau dieser Umstand macht die Suche zu solch einem schwierigen Thema. Die Suche ist keine Funktion, die ich aktivieren kann. Die Suche ist Teil des Portals, so wie die Ausbildung der Benutzer oder der Arbeitsprozess, den es abzubilden gilt.

Der Begriff *Suche* ist bereits falsch. In unserem Portalprojekt müssen wir uns zwar Gedanken über die Suche machen, noch mehr jedoch über die *Auffindbarkeit* von Objekten. Zwischen den beiden Begriffen Auffindbarkeit und Suche versteckt sich weit mehr als nur ein sprachlicher Unterschied. Das Auffinden von Informationen auf Portalen ist eine der schwierigsten Disziplinen. Dies rührt daher, dass in der Regel jeder Benutzer seine Daten zwar strukturiert, diese Strukturen aber meist nur für ihn oder für ein paar vertraute Mitarbeiter ersichtlich sind. Wie oft haben Sie sich schon bei simplen Dateiablage-Taxonomien gefragt, wo Sie jetzt diese Datei oder jene Datei ablegen sollten, sodass alle Beteiligten sie automatisch wiederfinden. Logischerweise wird dieses Problem umso größer, je mehr Beteiligte auf dem Portal arbeiten.

David Weinberger schreibt in seinem Buch »Das Ende der Schublade – Die Macht der neuen digitalen Unordnung«, dass bei der Auffindbarkeit das »Gesetz des Dschungels« herrscht. Jedes Portal, das wir kennen, ob CMS-basiertes Intranet oder Zusammenarbeitsportal, besteht in der Regel aus einer ungeordneten Vielfalt von Informationen. Diese ungeordnete Vielfalt wiederum führt dazu, dass es zu keiner bevorzugten Ordnung bei der Benutzung des Portals kommt. Denn die Benutzer müssen eigentlich die innere Ordnung des Portals nicht kennen, sondern jeweils nur den Teil, den sie benutzen möchten. Benutzer ordnen die Informationen so, wie sie es möchten und setzen diese Information auch wieder in anderer Form zusammen. Das ist der Weg, der den Boden für Innovationen ebnet. Diese Arbeitsweise auf dem Portal entspricht den Bedürfnissen der meisten Mitarbeiter.

Allerdings hat diese ungeordnete Vielfalt nicht viel Ähnlichkeit mit unserer gewohnten Sicht auf Wissen, Information und Daten. Wir gehen davon aus, dass es nur »ein Wissen« gibt, nur eine logische Sicht auf die Informationen usw. Diese falsche Sichtweise führt dazu, dass die vorgegebenen Portaltaxonomien implizieren, dass es nur eine Sichtweise gibt, nämlich diejenige, welche ein Gremium vorgeschlagen hat. Mit dieser Taxonomie wurde bereits das erste Kontingent der Auffindbarkeit abgetrennt. Für eine Vielzahl von Fällen wissen die Benutzer nun nicht mehr, wohin genau mit den Informationen. Diese Gremien (oder noch schlimmer: diese Personen) gehen davon aus, dass Information nicht mehrdeutig sein kann, sondern immer eindeutiger Natur ist, weil die Realität schließlich auch nicht mehrdeutig ist.

Ein solches Gremium braucht in der Regel immer lange, bis es zu einem Resultat kommt, weil es genau mit dem Problem der Mehrdeutigkeit beschäftigt ist. Diese Gremien kommen nie zu dem Schluss, dass vielleicht eine Taxonomie die falsche Art ist, um die Auffindbarkeit zu gewährleisten. Sie schlussfolgern eher, dass bei Unklarheiten der Benutzer etwas nicht richtig verstanden hat oder es nicht klar genug kommuniziert wurde, also ein Problem der Schulung vorliegt. Folglich wird aus dem Problem des fehlenden Wissens – mangels nicht gefundener Information – ein Problem des Könnens. Der Benutzer ist einfach nicht intelligent genug, um die Information wiederzufinden. Dies ist jetzt ein bisschen überspitzt formuliert. Es kennzeichnet jedoch

den Mechanismus, mit dem die meisten SharePoint-Administratoren ihre Portalsuche betreiben. Das Resultat ist, dass eigentlich niemand mit den Suchresultaten zufrieden ist.

Dies bedeutet: Wir haben Systeme, die es uns ermöglichen, eine Unmenge von beliebigen Informationen zu speichern. Da sind wir bereits sehr weit fortgeschritten. Woran es noch hapert, ist die Möglichkeit, wie wir die Information wieder aus dem System herausholen. Dass dies (zumindest eine Zeit lang) so bleiben wird, hängt vor allem damit zusammen, dass nicht nur die Systeme noch nicht so weit sind, sondern auch wir selbst diesen Spagat zwischen der realen Welt und unserem Informationsspeicher noch nicht gemacht haben. Grundsätzlich gibt es nur ein paar wenige Punkte zu beachten, um den Spagat mit den heutigen Systemen zu bewältigen.

Eine der ersten Überlegungen, um unser Problem mit dem Auffinden zu lösen, ist, beim Speichern ins System keine Filterung vorzunehmen. Wenn bereits auf dem Weg ins System gefiltert wird, wie etwa durch eine unlogische Taxonomie, verringert man den Wert der Fülle und schließt damit Informationen aus. Genau diese Informationen könnten aber für einen Benutzer irgendwann von großem Nutzen sein. Wird jedoch erst auf dem Weg aus dem System heraus gefiltert, steigert das den Wert des Systems durch die Fülle von Informationen, die in ihm gespeichert sind.

Jeden Tropfen Wasser in möglichst viele Seen verteilen. Jede Information muss möglichst dort sein, wo sie von Nutzen ist. In der realen Welt kann ein spezifischer Tropfen Wasser nur in einen See getropft werden. In einem digitalen System lässt sich der Tropfen einfach duplizieren und in viele Seen verteilen. Ein Informationsobjekt ist nur von Nutzen, wenn es an vielen Orten ist. Ein Onlineshop für Computerhardware kann zum Beispiel eine externe USB-Festplatte unter den Kategorien Disks, USB, externe Peripheriegeräte und unter Schnäppchen ablegen. Im Gegensatz zu der realen Welt wird die digitale Welt dadurch nicht chaotischer, sondern lediglich besser nutzbar.

The image shows a Google search interface. At the top, there are navigation links: Web, Bilder, Videos, Maps, News, Übersetzer, Google Mail, Mehr. On the right, there is a link to 'Anmelden' and a settings gear icon. The search bar contains the text 'answer to life the universe and everything' and a magnifying glass icon. Below the search bar, it says 'Ungefähr 3710'000 Ergebnisse (0.15 Sekunden)' and 'Google.com in English Erweiterte Suche'. On the left side, there is a sidebar with navigation options: Alles, Bilder, Videos, News, Mehr, Greifensee, Standort ändern, Das Web (Seiten auf Deutsch, Seiten aus der Schweiz, Übersetzte Seiten), and Alle (Letzte Stunde, Letzte 24 Std., Letzte Woche, Letzter Monat). The main search results area shows the top result: 'answer to life the universe and everything = 42' with a small calculator icon and a link to 'Mehr Informationen...'. Below this, there are three other search results from Wikipedia: '42 (Antwort) – Wikipedia', 'Phrases from The Hitchhiker's Guide to the Galaxy - Wikipedia, the ...', and '42 (number) - Wikipedia, the free encyclopedia'. Each result includes a brief description and a link to 'Weitere Ergebnisse von wikipedia.org'.

Abbildung 6.1 Douglas Adams – Die Antwort ist 42



die einzige sein. Denn diejenige Struktur, die vorgegeben wurde, ist nicht unbedingt für alle Fälle die Beste. Diese Erkenntnis alleine hilft bereits erheblich, das Suchen und Finden innerhalb von SharePoint Server 2010 zu optimieren.

### Quick Fact – Die Auffindbarkeit

Zwischen Suchen und Finden steht die Auffindbarkeit. Sie ist nicht eine *Funktion*, die ich aktivieren kann. Die Auffindbarkeit ist Teil des Portals, so wie die Ausbildung oder der Arbeitsprozess, den es abbildet. Deswegen bedingt eine gute Auffindbarkeit nicht nur Technik, sondern meist auch einen Wechsel der bisherigen Strategie, wie das Portal angeboten und betrieben wird. Dies bezieht sich insbesondere darauf, wer die Kontrolle über die Informationen und die Taxonomie des Portals hat.

## Die Auffindbarkeit – Zwischen Suchen und Finden

Das Motto jedes modernen Portals ist meist »find, use, and share«. Damit dieses Motto in die Realität umgesetzt werden kann, denken viele, es müssen lediglich einige technische Hürden genommen werden. Gerade aber die Suche ist ein Beispiel für ein Projekt, das zu einem Prozess wird. Bei vielen Unternehmen wird die Suche einmal implementiert und danach lediglich betrieben. Firmen, deren Geschäftsmodell auf der Suche von Daten basiert, sehen das natürlich anders. Google zum Beispiel forscht jeden Tag an der Optimierung der Suchalgorithmen und des Nutzens für die Benutzer des Suchportals. Lediglich bei Firmenportalen scheint die vorherrschende Meinung zu sein, dass hier nichts mehr geändert werden müsse und dass die Suche kein Prozess sei. Denn mit einem SharePoint-Portal und dem Anbinden aller verfügbaren Intranetquellen haben wir das »use and share« weitgehend gelöst. Das Gleiche trifft auf die Suche zu.

Ungelöst bleibt die Auffindbarkeit von Informationen. Das ist deshalb so kompliziert, weil das Auffinden von Informationen eigentlich kein rein technisches Problem ist, sondern es spannt den Bogen von der Technik bis hin zur sozialen Kompetenz eines Menschen.

Was macht dann eine gute Auffindbarkeit aus? Um das zu beurteilen, muss zunächst einmal die Auffindbarkeit definiert werden. Da Auffindbarkeit kein offizieller Begriff ist, sondern von uns lediglich konzeptionell verwendet wird, liegt es an uns, zu definieren, was Auffindbarkeit ist:

- Die Qualität eines Elements in Bezug auf seine Auffindbarkeit ist für einen Benutzer das zentrale Element der Suche. Es ist dabei unerheblich, ob Informationen über eine Suchmaske oder über einen Navigationspfad gefunden wurden. Der technische Architekturansatz dieser zwei Möglichkeiten ist jedoch unterschiedlich. Während bei der Auffindbarkeit über Navigationspfade die Taxonomie eine entscheidende Rolle spielt, ist bei der reinen Suche in einem globalen Suchfeld die Erfassung des möglichen Kontexts der abgesetzten Suche die Herausforderung.
- Der Grad der Einfachheit, mit der Objekte entdeckt oder lokalisiert werden können
- Der Grad der Unterstützung eines Systems in Bezug auf Navigation und Auffindbarkeit

An dieser Definition sehen Sie, dass Sie selbst die Benutzerschnittstelle als Teil der Auffindbarkeit mit einbeziehen müssen. Was im Gesamtkontext, aus Benutzersicht, am Ende auch die Suche repräsentiert.

## Was und wie wird gesucht?

Das Thema »Suchen« erzeugt bei uns viele Beratungsanfragen. Um die Aufgabe mit Erfolg abzuschließen, ist es für uns natürlich von Bedeutung, die Grundlagendaten zu kennen. So erheben wir meist zu Beginn Basisdaten über das Suchverhalten mit der aktuell vorhandenen Lösung eines Unternehmens. Meist kommen keine erstaunlichen Erkenntnisse zutage. In der Regel gleichen sich die Resultate über alle Firmen und Branchen hinweg. Aus technischer Sicht ist sicher interessant, wie die Benutzer eigentlich suchen. Sprich, mit Stichwörtern, mit ganzen Sätzen oder Stichwörter gekoppelt mit logischen Operatoren, wie etwa »and«, »or« usw. Nachfolgend finden Sie eine Zusammenfassung der wichtigen Fragen, über alle Befragungen hinweg, die wir durchgeführt haben.

### Frage: Wie suchen Sie im Moment?

	Total in Prozent	Prozent der Antworten
Ich suche mit nur einem Stichwort	74,5 %	90 %
Ich benutze erweiterte Suchbegriffe und logische Operatoren wie »and«, »or« oder »+« oder »-«-Operatoren	23,4 %	28,3 %
Andere	2,1 %	2,5 %
	100 %	–

**Tabelle 6.1** Antworten auf die Frage: »Wie suchen Sie im Moment?«

Die Benutzer, die »Andere« ausgewählt haben, geben dann folgende konsolidierten Punkte an:

- Ich benutze vordefinierte Filter
- Ich brauche die »erweiterten Suchfunktionen«, welche die aktuelle Lösung anbietet
- Ich suche nur über die Struktur des Portals und brauche die Suche gar nicht

In diesem Zusammenhang interessiert natürlich auch, wie die »Wildcard«-Suche aussieht. Im Gegensatz zu SharePoint Server 2010 ist diese Funktion beim Vorgänger Office SharePoint Server 2007 standardmäßig in der Oberfläche nicht vorhanden, kann jedoch über freie Tools oder Zusatzsoftware von Drittherstellern eingebunden werden. Die obige Frage ist deshalb besonders interessant, um festzustellen, ob in den Bereich Schulung investiert werden muss oder ob für die Benutzer die »Wildcard«-Suche zu einer solchen Gewohnheit geworden ist, dass diese von Anfang an bereitstehen muss.

### Frage: Welche »Wildcards« brauchen Sie im Moment für Ihre Suche?

	Total in Prozent	Prozent der Antworten
Ein beliebiges Zeichen, kein- bis einmal, pro Suchabfrage	27,4 %	33,3 %
Immer ein beliebiges Zeichen einmal pro Suchabfrage	5,5 %	6,7 %
Mindestens einmal oder mehrfach ein beliebiges Zeichen pro Suchabfrage	15,8 %	19,2 %
Keine	50 %	60,8 %
Andere	1,3 %	1,7 %
	100 %	–

**Tabelle 6.2** Ergebnis der Frage: »Welche »Wildcards« brauchen Sie im Moment für Ihre Suche?«

Wenn »Wildcards« eingesetzt werden, bringt uns das natürlich automatisch zur Frage, wie hoch die Anzahl der Stichwörter ist, mit denen gesucht wird?

**Frage: Mit wie vielen Suchbegriffen versehen Sie eine Suchabfrage in der Regel?**

Anzahl Suchbegriffe	Total in Prozent
1	14,2 %
2	52,5 %
3	27,5 %
4+	5,8 %
	100 %

**Tabelle 6.3** Ergebnis der Frage: »Mit wie vielen Suchbegriffen versehen Sie eine Suchabfrage in der Regel?«

Eines der weiteren interessanten Resultate ist sicher, wie die Benutzer generell mit den Suchresultaten der bestehenden Lösung zufrieden sind.

**Frage: Finden Sie im Allgemeinen die Information, die Sie suchen?**

	Total in Prozent	Prozent der Antworten
Ja	62 %	69,3 %
Nein	27,5 %	30,7 %
Keine Antwort	10,5 %	–
	100%	–

**Tabelle 6.4** Ergebnis der Frage: »Finden Sie im Allgemeinen die Information, die Sie suchen?«

Bei dieser Frage können die Teilnehmer der Befragung noch genauer spezifizieren, warum sie zufrieden oder eben unzufrieden sind. Die Antworten, die bei »Nein« geäußert werden, beziehen sich meist darauf, dass eben das Gesuchte nicht gefunden wurde. Die Antworten klaffen dabei so weit auseinander, dass kein Muster erkennbar ist. Einige Benutzer würden gerne Hierarchieinformationen von gesuchten Personen haben, andere möchten gerne Abkürzungen von Begriffen gefunden haben und andere wiederum möchten, dass alle im Unternehmen verfügbaren Anwendungen gesucht und gefunden werden können.

Aus dem Ergebnis in Tabelle 6.4 ist erkennbar, dass die Benutzer mit der Suche durchaus zufrieden sind. Dies wiegt aber die Betreiber der Lösung in einer falschen Zufriedenheit. Denn in der Regel sind auf solchen Portalen statische und strukturierte Daten abgelegt. Sobald aber mit unstrukturierten Daten auf Zusammenarbeitsportalen aufgewartet wird, genügt weder die Suchlogik noch der Suchmechanismus. Deswegen sind die meisten Anwender mit der SharePoint-Suche unzufrieden. Der Grund hierfür liegt darin, dass die Suche zwar immer noch gleich funktioniert, aber in der Relevanz deutlich schlechter geworden ist. Denn die Auffindbarkeit bei unstrukturierten Daten ist per se nicht gegeben, so wie dies bei strukturierten Daten der Fall ist. Bei strukturierten Daten können Taxonomieinformationen, wie Speicherort oder URL zur Hilfe genommen werden, um die Relevanz oder den Kontext der Suchresultate besser abzubilden. Diese Möglichkeit fällt bei unstrukturierten Daten weg.

# Informationsklassen

Dass die Auffindbarkeit bei unstrukturierten Daten so schwierig zu erreichen ist, hängt insbesondere mit der Art der Daten zusammen. Auf einem Portal und insbesondere auf Zusammenarbeitsportalen liegen diverse Arten von Informationen (siehe Tabelle 6.5).

Klassifikation	Beschreibung	Wichtigste Suchmethode	Ort der Information
Daten	Symbole für Objekte, Elemente und Eigenschaften	Schlüsselwörter	Server, Festplatte, elektronischer Speicher
Information	Sinnvolle Daten	Schlüsselwörter Menschen	Server, Festplatte, elektronischer Speicher
Wissen	Information in der Anwendung	Schlüsselwörter Menschen	Server und Menschen
Verständnis	Erklärtes Wissen	Meistens Menschen	Server und Menschen
Erfahrung	Meinungen, Werte, und Ethik angewendet auf Verständnis und Erfahrung	Fast immer Menschen	Fast immer Menschen

**Tabelle 6.5** Informationsklassen in einem Zusammenarbeitsportal

Wenn Sie die Tabelle 6.5 mit den Informationsklassen betrachten, fällt Ihnen bestimmt auf, dass eigentlich nur die Klasse »Daten« der Vorstellung entspricht, was in einem Intranet gesucht werden kann. Die Anforderungen an ein Intranetportal gehen in der Regel aber weiter. Bei den weiteren Informationsklassen fällt auf, dass sie vielfach nur im Menschen selbst vorhanden sind und nicht mehr in einem System. Zudem fällt auf, dass wir innerhalb eines Portals fast immer nur mit Schlüsselwörtern suchen können. Suchen nach komplexeren Klassen, wie etwa Erfahrung oder Verständnis, können jedoch sehr schlecht in Schlüsselwörter gefasst werden. Da der Mensch nicht indexiert werden kann, müssen die Menschen untereinander vernetzt werden. Der Begriff »Social Network« wird in diesem Zusammenhang benutzt. Hier wird offensichtlich, dass der Mitarbeiter der wichtigste Träger von komplexer Information im Unternehmen ist. Die richtige Vernetzung der Mitarbeiter kann aber nicht durch die IT-Abteilung geschehen. Diese kann in diesem Zusammenhang lediglich die Möglichkeit bieten, die Infrastruktur für die Vernetzung bereitzustellen und zusammen mit anderen Abteilungen (wie etwa Personalabteilung oder Geschäftsleitung) die Vernetzung der Menschen auf allen Ebenen anzustoßen.

## Quick Fact – Auffindbarkeit von Informationen

Die Auffindbarkeit von Informationen ist für einen Benutzer das zentrale Element der Suche. Es ist dabei unerheblich, ob Informationen über eine Suchmaske oder über einen Navigationspfad gefunden wurden. Im Allgemeinen sind die Benutzer mit der bestehenden Suche zufrieden. Allerdings liegt der Fokus der abzulösenden Portale auf statischen oder strukturierten Daten. Während die neuen Portale meist auch die Zusammenarbeit und damit neu die unstrukturierten Daten mit einbeziehen. Diese unstrukturierten Daten führen aber dazu, dass die bisherige Technik und Suchlogik versagen muss.



## Warum es das perfekte Suchergebnis nicht gibt

Die richtigen Informationen in großen Datenmengen erfolgreich zu finden, ist eine Sache. Sie so darzustellen, dass der Nutzer sie auch optimal verwerten kann, ist eine zusätzliche Herausforderung. Denn je umfassender Informationen sind, desto schwerer behält man den Überblick. Das ist aber nur die Benutzersicht. Es gibt jedoch weitere Sichten aus dem Unternehmen heraus. In Abbildung 6.3 sind diese drei Sichten exemplarisch illustriert. Jeder Benutzer ist zwar aus IT-Sicht ein Benutzer, er hat aber immer drei Sichtweisen auf die Suchresultate: die Benutzer-, die Geschäfts- und die Datensicht. Je nach Zweck der Suche verlagert sich die Sichtweise, sprich der Kontext des gewünschten Suchresultats innerhalb dieses Dreiecks.

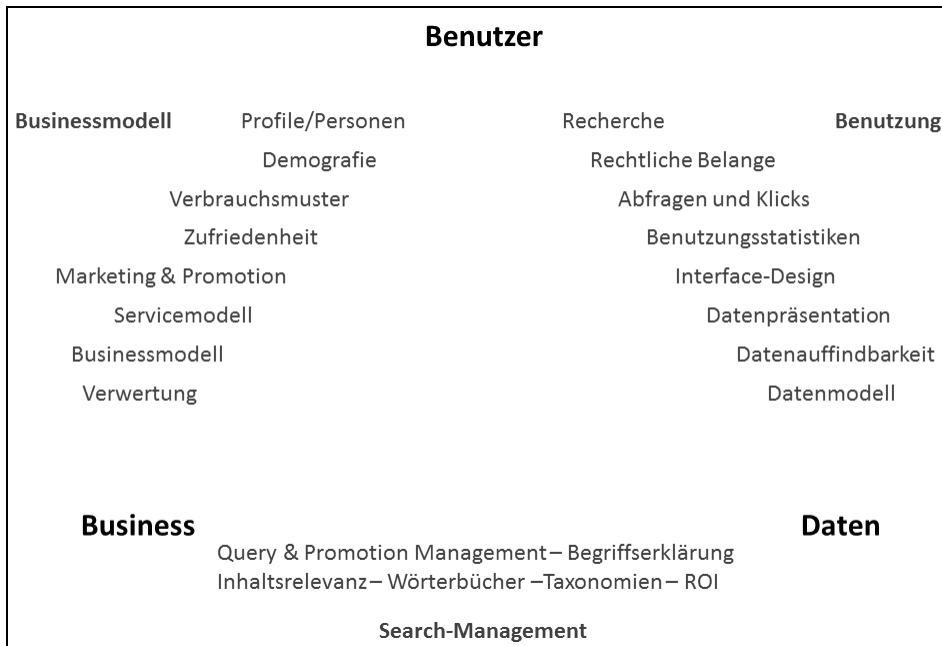


Abbildung 6.3 Das Suchdreieck

Es gibt jedoch noch andere Gründe und Mechanismen, warum es ein perfektes Suchergebnis nicht geben kann. Das »Long Tail«-Prinzip ist eine Theorie, die der US-amerikanische Journalist und Chefredakteur des Wired Magazine Chris Anderson im Jahr 2004 vorstellte, nach der ein Anbieter im Internet durch eine große Anzahl an Nischenprodukten Gewinn machen kann. Dieser Effekt trifft insbesondere für den Musik- und Bücherverkauf zu, wo selten verkaufte Titel in einem konventionellen Verkaufsgeschäft zu hohe Kosten verursachen würden, wenn sie vorrätig sein sollen. Der Name leitet sich von der Ähnlichkeit der Verkaufsgrafik mit einem langen Schwanz ab (Abbildung 6.4).

Mit dem Prinzip des »Long Tail« kann das Problem der Suche in einem Portal sehr genau illustriert werden. Das »Long Tail«-Prinzip, auf einem Portal angewendet, bedeutet Folgendes: Die Summe aller Suchabfragen nach wenig frequentierten Objekten ist größer als die Summe aller Suchen nach populären Objekten. Dies bedeutet, dass die Benutzer in der Regel Objekte suchen, die schwierig zu finden sind. Deshalb wird in diesem Zusammenhang ja gerade die Suche verwendet. Folglich wird die Suche in einem Portal verwendet, um »Long Tail«-Objekte zu finden.

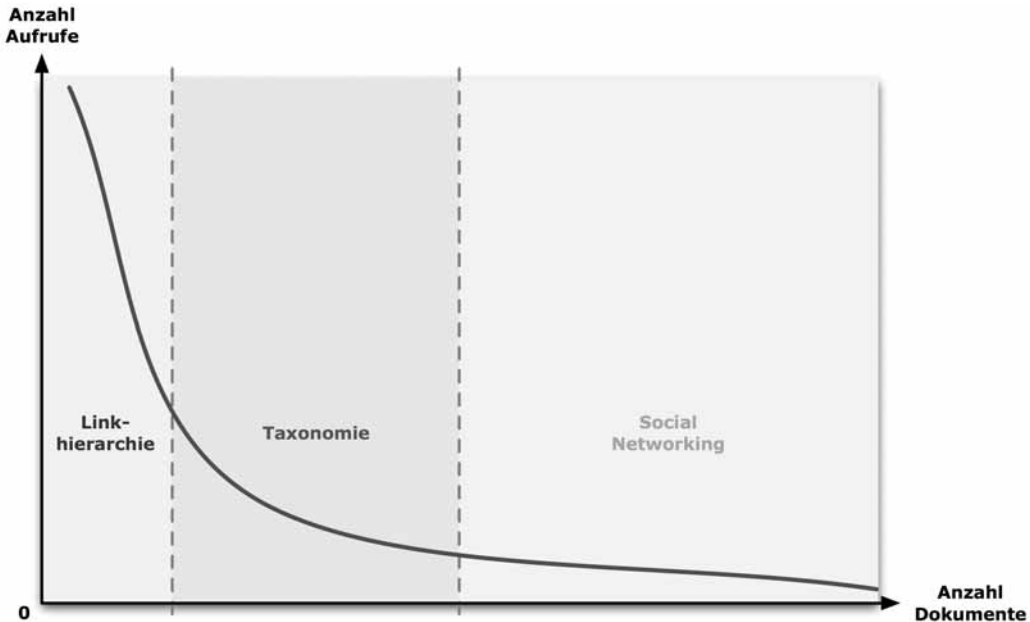


Abbildung 6.4 Das »Long Tail«-Prinzip

## Das Potenz- und das Zipfsche Gesetz

Die Potenzgesetze der Statistik gehören zu den Skalengesetzen und beschreiben die Skaleninvarianz vieler natürlicher Phänomene. In diesen Gesetzen lassen sich die Probleme der Suche innerhalb von Portalsystemen erkennen. Die Gesetze der Statistik spielen natürlich auch beim »Long Tail«-Prinzip eine Rolle. Interessant wird die Statistik aber erst, wenn Sie die Arbeiten von George Kingsley Zipf mit in die Betrachtung einbeziehen. Bekannt wurde er durch seine Arbeiten, in denen er die Häufigkeit von Wörtern in unterschiedlichen Sprachen untersuchte. Gemäß dem Zipfschen Gesetz gilt: Wenn man die Wörter einer Sprache nach ihrer Häufigkeit ordnet und dem häufigsten Wort den Rang 1 zuweist, dem zweithäufigsten den Rang 2 usw., ergibt das Produkt aus Rang und Häufigkeit eine immer annähernd gleiche Konstante. Unter den weiteren Zipfschen Gesetzen befinden sich die Erkenntnisse über den Zusammenhang zwischen Länge und Häufigkeit eines Wortes: Je häufiger ein Wort ist, desto kürzer ist es. Ferner über Alter und Häufigkeit: Je häufiger ein Wort ist, desto älter ist es auch.

Auf unser Suchproblem reduziert bedeutet dies: Je öfter ein Wort gebraucht wird, umso genereller ist seine Bedeutung. Je weniger ein Wort gebraucht wird, umso präziser ist seine Bedeutung. Sprachen aber tendieren dazu, einzelnen Wörtern mehrere Bedeutungen in unterschiedlichen Kontexten zu geben. Falls aber einzelne Wörter unterschiedliche Bedeutungen haben, wie kann dann eine Suche nach einem Wort jemals ein präzises Suchresultat ergeben?

Nehmen wir nun die Statistik und die Größe der Datenbank hinzu, ergibt sich folgender Schluss: In dem Maß, wie die Anzahl der indexierten Dokumente in einer Datenbank steigt, nimmt die Relevanz des Suchresultats bei einer Volltextsuche ab.

## Der Faktor Mensch (Mooers-Gesetz)

Calvin Northrup Mooers war ein amerikanischer Pionier der Informatik. Er ist der Begründer des Information Retrieval und entwickelte mit TRAC (Text Reckoning And Compiling) eine der ersten Programmiersprachen. Sein Gesetz besagt: »Ein Information-Retrieval-System wird tendenziell weniger benutzt, wenn es für seinen Benutzer unbequem und mühsam wäre, über Informationen zu verfügen, als sie nicht zu besitzen«. In derselben Arbeit steht auch: »Ich vermute eher, dass Menschen keine Informationen haben möchten und lieber darauf verzichten, ein System zu benutzen, gerade weil es sie informiert«.

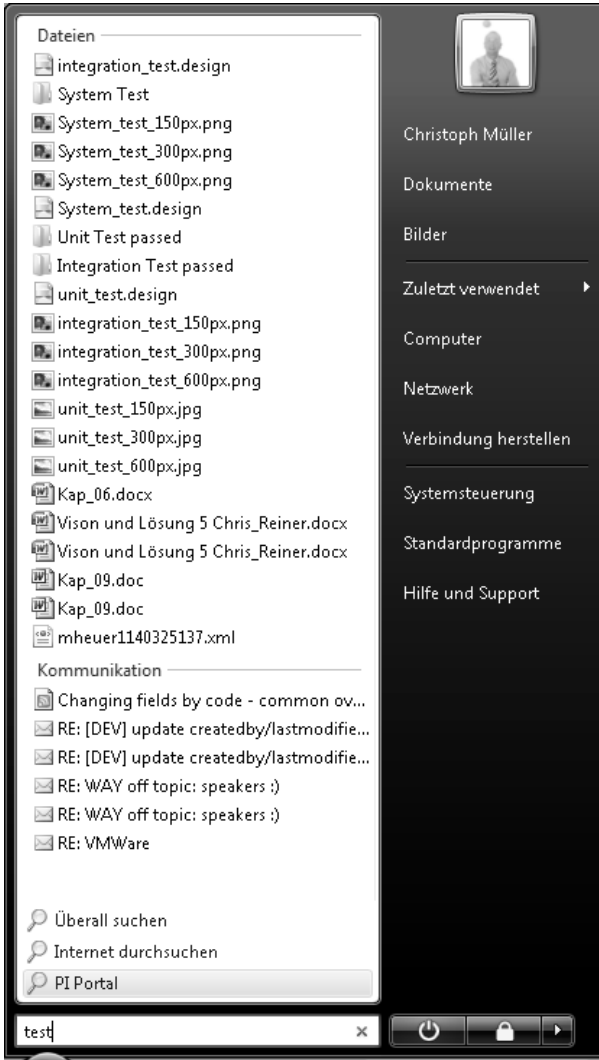
Auf ein Portal übertragen bedeutet dies, dass die Menschen Informationen nicht suchen, wenn es ihre Arbeit aufwendiger macht. Sie machen es nur, wenn sie es »unbedingt« tun müssen. Dieser Umstand trägt dem Faktum Rechnung, dass Informations- und Kommunikationstechnologien in der Regel einen niedrigen »Return on Investment« haben. Menschen tendieren dazu, ihre »Mensch-zu-Mensch-Netzwerke« für komplexere Suchen nach Information zu benutzen.

## Technische Limitationen/Federated Search

Das Problem besteht heute darin, dass zwischen dem Portalindexer und dem Clientcomputerindexer eine Kluft besteht. Diese Kluft kann im Moment nicht überbrückt werden. Dies bedeutet, dass jeweils auf dem Portal und auf dem Client separat indexiert wird. Dies wiederum bedeutet, dass für beide Quellen separate Suchen in separaten Suchmasken abgesetzt werden müssen. Die Clientindexierung geschieht unter Windows 7 über »Windows Search«. Das Pendant unter Windows XP ist »Windows Desktop Search«. Der Indexer der beiden Clientsuchmaschinen kann Dokumente, Mails und Metainformationen von binären Dateien wie etwa Bildern auslesen und speichern. Die Möglichkeit, weitere Quellen außerhalb des Dateisystems an den Indexer auf dem Client anzubinden, fehlt jedoch. Gleichzeitig fehlt auch die Möglichkeit, Clientindexdatenbanken an eine SharePoint-Suchdatenbank anzugliedern.

In der aktuellen Version 4 besteht die Möglichkeit, innerhalb der Windows-Suche eine weitere Quelle (allerdings nur eine) anzugeben. Diese Quelle erscheint im Windows-Startmenü (Abbildung 6.5). Die Suche öffnet die Resultatseite des Portals im Internet Explorer. Dies kann wohl als Workaround, nicht aber als Lösung bewertet werden.

Das Problem der Kluft zwischen Portal und Client ist Microsoft bekannt. Seit Veröffentlichung von Internet Explorer 8 kann man den Suchanbieter (Search Provider) eines SharePoint-Portals ansprechen. In SharePoint Server 2010 kann man auf der Ergebnisseite die SharePoint-Suche mit der Windows-Suche verbinden. So werden in lokalen Suchanfragen auch Inhalte aus dem SharePoint-Index abgefragt. Dies bedeutet für den Benutzer, dass er seine Suche nicht mehr auf dem Portal (Search Center) ausführen wird, sondern lokal auf dem Rechner über die gewohnte Suche.



**Abbildung 6.5** Windows 7-Startmenü mit verlinkter Portalsuche

## Die Bedürfnisse der Benutzer

Wir stellen immer wieder fest, dass es nicht einfach ist, seine Bedürfnisse in Bezug auf die Suche zu artikulieren. Klar gibt es Anforderungen an die Suche, die eindeutig und klar formuliert werden können. Es gibt aber diverse Bereiche, bei denen es nicht ganz so klar ist. Vielleicht wissen es die Benutzer noch gar nicht, dass sie bald nach etwas suchen möchten oder dass das Bedürfnis nicht in einen Kontext gesetzt werden kann. Bewährt hat sich hier, die Bedürfnisse in einen Kontext zum Inhaltstyp zu setzen (Abbildung 6.6). Dabei werden die Bedürfnisse der Benutzer nach ihrer Präzision qualifiziert. Auf der zweiten Achse werden die auf dem Portal angebotenen Inhaltstypen abgebildet. In der Abbildung 6.6 geht die Bandbreite vom redaktionellen Inhalt bis zu dem Inhalt, der aus Gründen der Vertraulichkeit nicht auf dem Portal gespeichert ist. In den so entstandenen Bereichen werden die Möglichkeiten technisch sowie organisatorisch eingefügt.

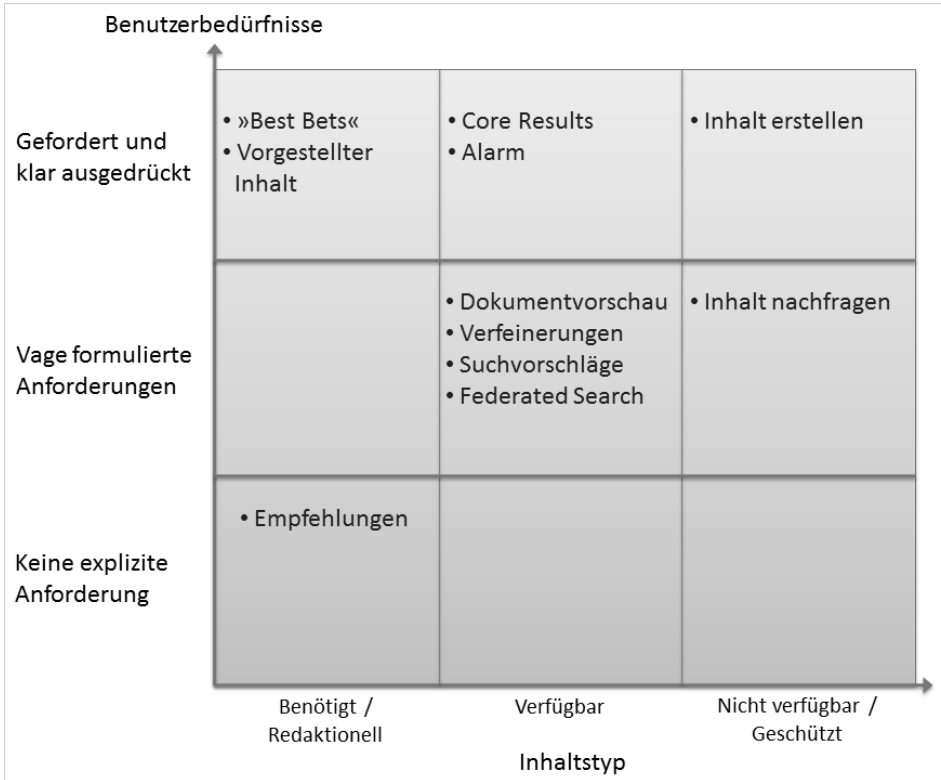


Abbildung 6.6 Inhalt und Bedürfnisse

Ist die Matrix in dieser Form einmal erstellt, lassen sich recht einfach Maßnahmen daraus ableiten. Dies sowohl für das Projektteam und deren Umsetzung in SharePoint Server 2010 als auch die organisatorischen Maßnahmen.

#### Quick Fact – Warum es das perfekte Suchergebnis nicht gibt

Das perfekte Suchresultat wird es niemals geben. Schon allein die Definition des richtigen Suchresultats macht uns Mühe. Zuviel »weiche« Faktoren, wie etwa das Geschäftsumfeld, die Erwartungen oder die sich ständig ändernden und wachsenden Datenarten auf dem Portal, machen dies zu einem unmöglichen Wunsch. Wichtig in diesem Kontext ist, dass Sie diese Erkenntnis in Maßnahmen münden lassen. Somit wird klar, dass die Suche eigentlich kein Projekt ist, sondern ein Prozess, der niemals endet, solange das Portal besteht.

## Annäherung an die perfekte Suche

Wie bereits aus den vorigen Abschnitten ersichtlich ist, gibt es die perfekte Suche nicht. Sie können nur versuchen, sich diesem Ideal anzunähern. Im Fall von SharePoint Server 2010 können Sie dies, indem Sie zum einen die Suchmaschine optimieren und zum anderen die Benutzerschnittstelle so gestalten, dass sie möglichst viel an Unterstützung für den Benutzer bietet. Beide Ansatzpunkte verlangen jedoch eine kontinuierliche Pflege und Verbesserung. Es ist daher notwendig, dass Sie die Betrachtung nicht nur auf den reinen Serverbetrieb beschränken, sondern auch die Bedürfnisse der Benutzer immer wieder einfangen und versuchen, dieses »Feedback« für die Optimierung zu nutzen. SharePoint Server 2010 bietet beispielsweise Berichte an, die das Suchverhalten der Benutzer erkennbar machen. Die Möglichkeiten des direkten Feedbacks und der Kontaktmöglichkeit zu einem Verantwortlichen für die Suche bringen ebenfalls wichtige Erkenntnisse.

## Long Tail

In einem der vorigen Abschnitte wurde das »Long Tail«-Prinzip bereits dargelegt. Es kann vor allem sehr gut herangezogen werden, um den Inhalt aus Benutzersicht zu klassifizieren. Dabei werden die Anzahl der Klicks auf ein Objekt (Anzahl der Aufrufe) und die Anzahl der vorhandenen Informationsobjekte (Anzahl der Dokumente) miteinander in Beziehung gesetzt. Ihre horizontale Position auf der so entstandenen Resultatlinie ist verantwortlich für die Gewichtung und die »Prominenz«, die ein Objekt erhält.

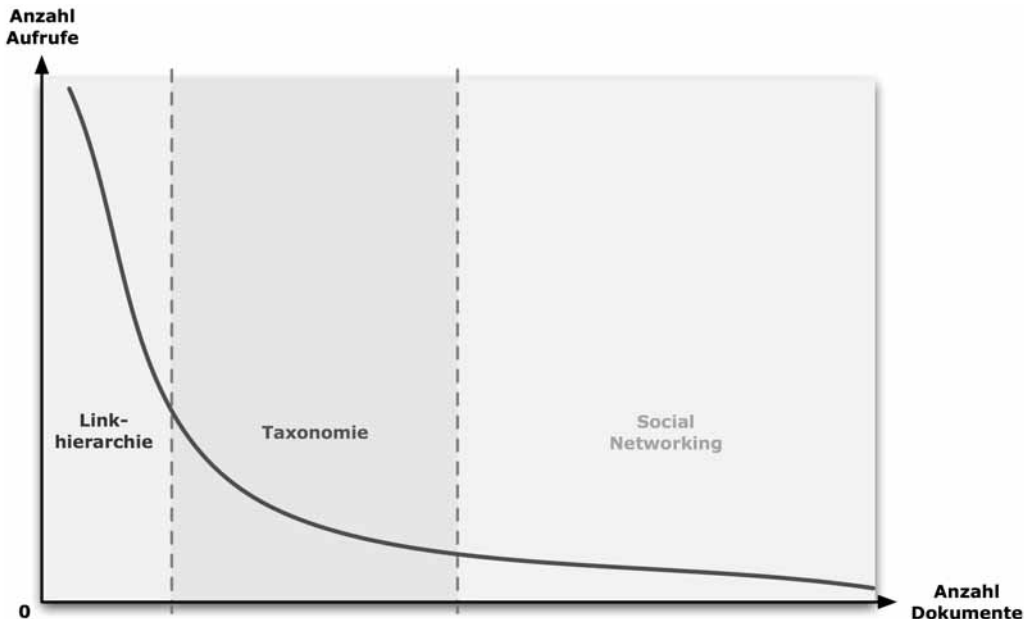


Abbildung 6.7 Abbildung der Abschnitte im »Long Tail«

Alles was in den Bereich »Short Tail« fällt, sollte prinzipiell über Links erreichbar sein, die an prominenter Stelle sichtbar sind. Diese werden sehr hoch gewichtet und werden auf den Startseiten der jeweiligen Intranetseiten platziert. Sie können jedoch auch an weiteren Orten positioniert sein:

- **Linkhierarchie** Zum Beispiel Top-Seiten, auf denen auf Unterseiten verwiesen wird. Dies kann im SharePoint Server 2010 sehr einfach anhand von Link-Webparts erreicht werden. In diesem Fall handelt es sich um redaktionell gepflegte Links.
- **Inhaltsverzeichnis und Linklisten** Ein Inhaltsverzeichnis bietet SharePoint Server 2010 bereits an. Falls die gesuchten Objekte aus dem »Short Tail« hier an oberen Positionen auftauchen, gelangt der Benutzer sehr schnell zum Ziel. Linklisten sind ebenfalls ein Standardmittel von SharePoint Server 2010, um Verzweigungen redaktionell zur Verfügung zu stellen.
- **Optimale Topnavigation und Homepage** Oft benötigte Ziele lassen sich bereits in der Hauptnavigation unterbringen und verzweigen ohne Umwege auf ein ggf. tieferliegendes Ziel, welches der Benutzer über die Navigation durch mehrere Klicks hätte finden müssen. Direkte Links oder Animationen auf der Homepage führen den Benutzer ebenfalls schnell zum Ziel. Sind diese Links zudem immer am gleichen Platz (z.B. Aktuelle Themen), findet der Benutzer sehr schnell bei aktuell aufkommenden Fragestellungen die richtige Information.
- **Aufnahme in den Zeitgeist** Der Zeitgeist ist eine Trefferliste der meistgefragten Suchen. Er wird redaktionell gepflegt und kann somit auch proaktive Themen enthalten. Mehr Details hierzu finden Sie im Abschnitt »Nutzen versus Technik« ab Seite 242.

Objekte aus dem »Mid Tail«-Bereich können nicht mehr direkt angeklickt werden. Diese Informationen findet der Benutzer jedoch durch eine entsprechend hohe Gewichtung und Positionierung auf Seiten, die unterhalb der primären Homepage liegen. Dadurch haben sie immer noch eine sehr hohe Gewichtung innerhalb des Portals. Sie werden über Navigation und Stichwortsuche gefunden. Objekte aus dem »Mid Tail« werden strukturiert durch:

- **Taxonomie** Links auf die gesuchten Objekte werden auf zweiter oder max. dritter Ebene prominent auf deren Homepage gelegt. Dadurch lassen sich die Objekte sehr schnell von Benutzern erkennen, die zu diesem Bereich gehören, aber auch Benutzer, die auf diese Bereichsebene durch die Navigation gelangt sind (z.B. durch einen Verweis auf der Portal-Homepage), ohne dass sie direkt dem Bereich angehören würden.
- **Metadaten** Objekte (z.B. Rechnung) erhalten zusätzliche Metadaten, die das Objekt eindeutig kennzeichnen. Dies ist vor allem dann sinnvoll, wenn es Merkmale gibt, die wenig an Interpretationsspielraum zulassen (z.B. eine Rechnungsnummer). Durch die Vergabe von Metadaten kann in der Suche eine direkte Eingabemöglichkeit für diese Metadaten angeboten werden. Die Suche beschränkt sich dadurch auf diese Metadaten und liefert optimierte Trefferlisten.
- **Suchbereiche** Diese ermöglichen die Unterteilung des Suchindexes. Damit hat der Benutzer die Möglichkeit, die Suche einzuschränken, ohne dass er das eingegebene Stichwort anpasst. Suchbereiche lassen sich zentral vom Administrator, als auch im jeweiligen Portal definieren und ermöglichen somit die Erstellung von eingeschränkten Bereichen, abhängig von den Anforderungen des jeweiligen Portals.
- **Beste Suchergebnisse** Der Portaladministrator kann pro Portal (Websitesammlung) definieren, welches Suchergebnis bei der Eingabe eines entsprechenden Stichworts erscheinen soll. Die besten Suchergebnisse werden an oberster Stelle in der Trefferliste angezeigt und zudem noch optisch von den anderen Treffern abgehoben. Um die »Besten Suchergebnisse« zu pflegen, muss der Administrator des Portals die Bedürfnisse der Benutzer kennen bzw. wissen, welche Stichwörter häufig angefragt wurden.
- **Separation in Websitesammlungen (Eingrenzen der Suche)** Eine automatische Einschränkung der Ergebnisse erfolgt auf der Ebene der Websitesammlungen. Wurden also beispielsweise die einzelnen Intra-

nets auf Basis einer eigenen Websitesammlung angelegt (siehe auch Kapitel 4), dann wird die Suche des Benutzers standardmäßig nur auf diesen Bereich ausgeführt.

- **Stichwortsuche** Die Stichwortsuche sucht sowohl im Volltextindex als auch in den Metadaten und nutzt dabei das Relevanzsystem von SharePoint Server 2010, um gefundene Treffer in der Trefferliste zu positionieren. Die Stichwortsuche wird von den Benutzern meistens als Erstes verwendet, um Inhalte zu finden. Durch Konfigurationen der SharePoint Server 2010-Suchmaschine kann ein Stichwort mit einem speziellen Ergebnis verknüpft werden, den sog. »Besten Suchergebnissen«. Zum erfassten Stichwort können Synonyme definiert werden, sodass Sie bei Verwendung von ähnlichen Stichwörtern zum gleichen Ergebnis gelangen.

Informationsobjekte aus dem »Long Tail«-Bereich finden sich nicht mehr von alleine. Diese werden nur noch im Kontext von Benutzern gefunden, die dieses Dokument selber erstellt oder bei der Erstellung mitgearbeitet haben. Für diese Informationen müssen Mechanismen gefunden werden, wie man die Benutzer findet, die das jeweilige Dokument kennen könnten. Dies geschieht über folgende Lösungsstrukturen:

- **Schlüsselwortsuche** Auch im Bereich des »Long Tail« kann eine Stichwortsuche durchaus zum Erfolg führen. Schließlich indiziert SharePoint Server 2010 alle Dokumente und Daten und hält diese im Index vor. Der Benutzer ist in diesem Fall allerdings etwas davon abhängig, dass er die korrekten Stichwörter wählt. Hier können ihn entsprechende Hinweise auf der Benutzerschnittstelle der Suchmaske auf die richtige Spur bringen.
- **Soziales Netzwerk** Die anderen Mitglieder einer Teamsite oder eines Diskussionsforums lassen sich ebenfalls dazu nutzen, um an weitere Informationen zu den gesuchten Themen zu gelangen. Gerade Diskussionsforen können sehr schnell zum Erfolg führen, da viele andere Benutzer potenziell auf eine Frage antworten können.
- **Persönliche Website** Sind entsprechende Informationen im Benutzerprofil der Benutzer gepflegt, kann dies beispielsweise für eine Suche nach Experten, Projektinformationen oder Ansprechpartnern genutzt werden. Oftmals werden die notwendigen Informationen dann erst auf der persönlichen Website des jeweiligen Benutzers sichtbar, wenn dieser beispielsweise Inhalte mit Tags versehen hat oder wichtige Ergebnisse aus verschiedenen Projekten auf seiner persönlichen Website vorhält.
- **Präsenz** In der Trefferliste werden Informationen zur persönlichen Nachfrage anhand von Instant Messaging angeboten.
- **Telefon und persönliche Kontakte** Der Benutzer erfragt sich die notwendigen Informationen anhand seines persönlichen Netzwerks. Die Suche liefert in diesem Fall oft nur übergeordnete Treffer, die nicht direkt das Gesuchte widerspiegeln.

#### Quick Fact – Long Tail

Die Betrachtungsweise anhand des »Long Tail«-Modells kann sehr gut genutzt werden, um den Inhalt aus Benutzersicht zu klassifizieren. Dabei sollten sich Inhalte aus dem »Short Tail« an prominenter Stelle im Portal befinden. Elemente aus dem »Mid Tail« werden über die Stichwortsuche oder durch prominente Anordnung in weiter unten liegenden Hierarchien gefunden. Für das Finden von Elementen aus dem »Long Tail« werden verstärkt soziale Netzwerke benutzt. Die Stichwortsuche kann jedoch auch noch zum Ergebnis führen.



## Folksonomie und Persönliche Website

Durch die Verfügbarkeit von Social Software und Web 2.0 passiert gerade eine Revolution im Internet. Eine, in der der Benutzer die Möglichkeit hat, seine eigene Onlineerfahrung selbst zu beeinflussen und mit anderen zusammenzuarbeiten. Benutzer sind in der Lage, Metadaten hinzuzufügen und Tags zu benutzen, um ihre digitalen Sammlungen zu organisieren, Inhalte für sich und andere zu kategorisieren und damit Klassifizierungssysteme zu erstellen. Die Klugheit der Masse und die kollektive Intelligenz tun das, was früher nur den Katalogisierungsexperten, Informationsarchitekten und Websiteautoren vorbehalten war: Sie kategorisieren und organisieren das Internet und bestimmen dadurch die Onlineerfahrung des Benutzers. Und das Erstaunliche dabei ist, dass dies auch noch funktioniert.

Diese neuen Möglichkeiten, die unter dem Begriff *Folksonomie* immer weiter Verbreitung finden, kann auch im Intranet Anwendung finden. Der große Vorteil besteht in der nicht mehr starr vorgegebenen Strukturierung und dem Vergeben von Metadaten, die meist nur für wenige Benutzer sinnvoll sind, weil sie sich in dem entsprechenden Bereich sowieso schon auskennen und die Metadaten-Taxonomie kennen. Social Bookmarking und Tagging bieten die Möglichkeit, selbst eigene Schlagwörter zu vergeben und diese mit anderen zu teilen. Die große Leistungsfähigkeit kommt dann zum Tragen, wenn möglichst viele Benutzer dieses System nutzen und dabei unterschiedliche Schlagwörter für die gleichen Objekte oder die bereits bestehenden nutzen. Damit wird die Information über mehrere Wege zugänglich gemacht und gleichzeitig bei vielfacher Nutzung von Schlagwörtern auch eine automatische Relevanz erzeugt.

Da potenzielle Suchergebnisse im »Long Tail« nur sehr schwer mit einem reinen Suchansatz zu finden sind, spielt in diesem Bereich das Thema Social Networking eine gewichtigere Rolle. Die persönliche Website im Zusammenhang mit dem Finden bekommt eine große Bedeutung. Sie zeigt zum einen Informationen aus dem Benutzerprofil an, zum anderen kann sie auch die Verbindung mit anderen Benutzern darstellen. Sie verwendet hierfür Informationen, die sich der Benutzer selbst gesetzt hat (z.B. Kollegen), aber auch Informationen, die aus Active Directory kommen oder von Microsoft Exchange Server eingebracht werden (z.B. Mitgliedschaften in Verteilergruppen). All diese Informationen bieten Zusatznutzen für andere Benutzer. Benutzer pflegen auf ihrer persönlichen Website Informationen, die ihre Erfahrungen widerspiegeln und sie damit als potenzielle Ansprechpartner leichter identifizieren lassen.

Die persönliche Website bzw. das Benutzerprofil spielt ebenfalls eine große Rolle bei der Suche nach Experten für entsprechende Themen.

Wie suchen wir nach Experten?

- Menschen suchen nach Indikatoren für Kompetenz:
  - Kompetenz wird an zwei Dinge gemessen: Was Sie sagen und was Sie erreicht haben.
  - Dem was ein anderer bereits erreicht hat, wird am meisten vertraut, da Fakten über persönliche Fähigkeiten objektiv und verifizierbar sind
- Sie vertrauen stark auf persönliche und professionelle Netzwerke:
  - Kompetenzsuchende vertrauen der Aussage von Kollegen oder Freunden
  - Gerade diese Verbindungen ermittelt SharePoint Server 2010 automatisch anhand verschiedener Informationen, die im Unternehmen vorhanden sind (z.B. Hierarchie im Verzeichnisdienst, Mitgliedschaft in Verteilergruppen usw.)
- Sie unterstellen eine schwierige Zugänglichkeit von Experten:

- Menschen unterstellen eine schlechte Verfügbarkeit, wenn sie einer Person Fragen stellen, die sehr beschäftigt und/oder ein erfahrener Mitarbeiter ist. Sie bevorzugen in diesem Fall, dass sie von einer Person vorgestellt werden, die sowohl der Suchende als auch der Experte kennt. Anhand des Netzwerks gibt es meist schon einen Anknüpfungspunkt, um mit dem Experten ins Gespräch zu kommen. SharePoint Server 2010 zeigt dieses Netzwerk an.
- Die Experten können von sich aus bereits einiges dazu beitragen, dass sich Hilfesuchende des Wissens bedienen können, indem sie beispielsweise veröffentlichtes Wissen des Experten nachlesen können (z.B. in Blogs oder veröffentlichten Dokumenten).

Aus den obigen Ausführungen wird ersichtlich, dass es wichtig ist, entsprechende Verbindungen zwischen den einzelnen Personen bzw. Mitarbeitern zu erkennen.

### Quick Fact – Folksonomie und persönliche Website

Folksonomie verwendet die Klugheit der Masse und die kollektive Intelligenz, um Inhalte zu klassifizieren. Techniken, die die Folksonomie unterstützen sind in SharePoint Server 2010 vorhanden.

Die persönliche Website, die Aktivitäten des Benutzers und das Profil kommen vor allem im »Long Tail« zum Tragen. Dort sind die Benutzer, die Informationen suchen, auf Experten angewiesen, die ihnen Auskunft geben können. Die Informationen auf der persönlichen Website und den Benutzerprofilen können dies unterstützen.

## Nutzen versus Technik

Im Hinblick auf die Nutzung der SharePoint Server 2010-Suche eröffnet sich ein weites Feld von Möglichkeiten zur Optimierung und der Erhöhung des Nutzens für den Benutzer. Dabei bedarf es oftmals keinerlei Eingriffe in die Suchtechnologie von SharePoint Server 2010.

Die Idee hinter einer Optimierung besteht aus der Trennung der Technik und dem, was der Benutzer wahrnimmt. Aus Projektsicht können wir die beiden Teile zwar nicht trennen, trotzdem müssen wir sie für die Analyse getrennt betrachten. Beide Teile optimieren, aus Benutzersicht, die Suche (Abbildung 6.8). Die Optimierung der Suchergebnisse passiert auf Basis von technischen Anpassungen und durch Modifizieren der Einstellungen. Die Optimierung des Nutzens für den Benutzer wird jedoch erreicht durch Anpassung des Inhalts oder Veränderungen an der Suchmaske. Diese beiden Ausprägungen haben zwar das gleiche Ziel, jedoch ganz andere Maßnahmen zur Folge. In den meisten Fällen ist ein Mix aus beiden Gebieten sinnvoll, um das Resultat gleichmäßig zu verbessern.

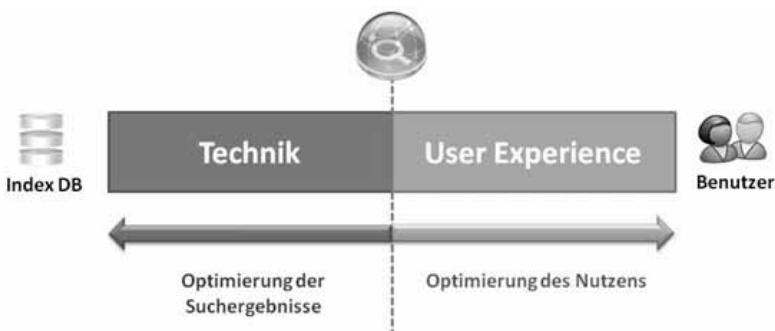
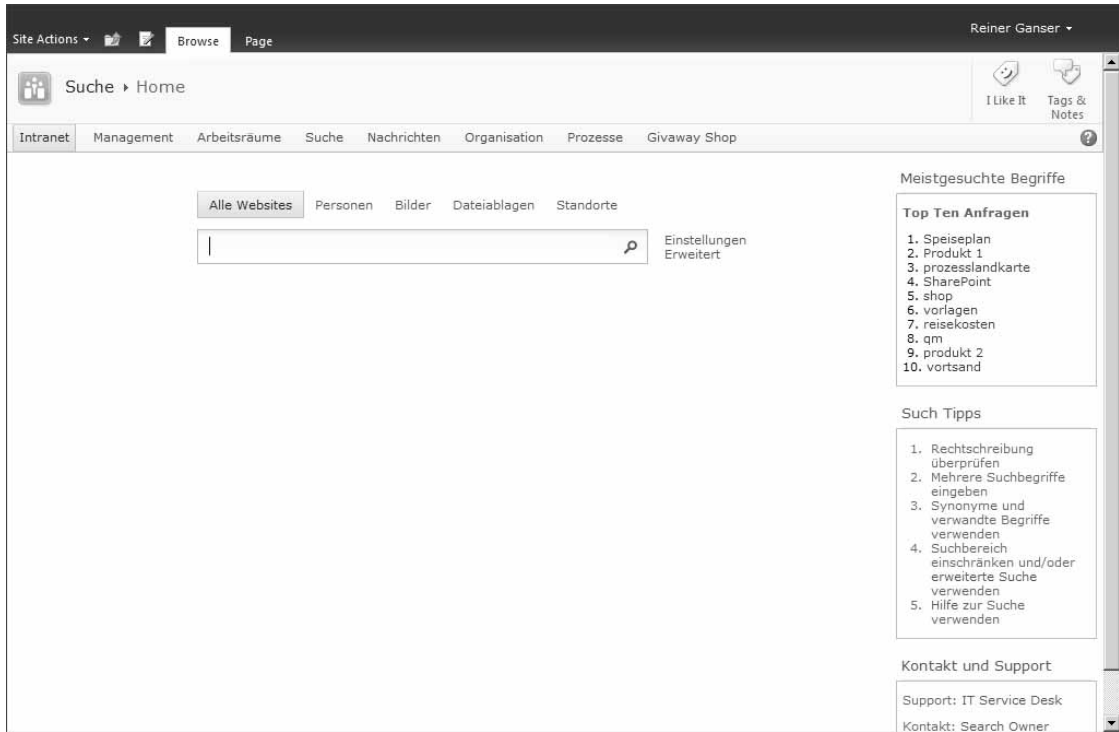


Abbildung 6.8 Benutzernutzen versus Technik

Um den Nutzen für den Benutzer zu verbessern, lässt sich die Infrastruktur von SharePoint Server 2010 sehr gut einsetzen. Es geht darum, dem Benutzer bereits beim Aufruf der Suchmaske entsprechende Hinweise anzubieten, die direkt zu einer verbesserten Suche führen. Die Abbildung 6.9 zeigt ein solches Beispiel.



**Abbildung 6.9** Einfache Suche mit Hinweisen für den Benutzer

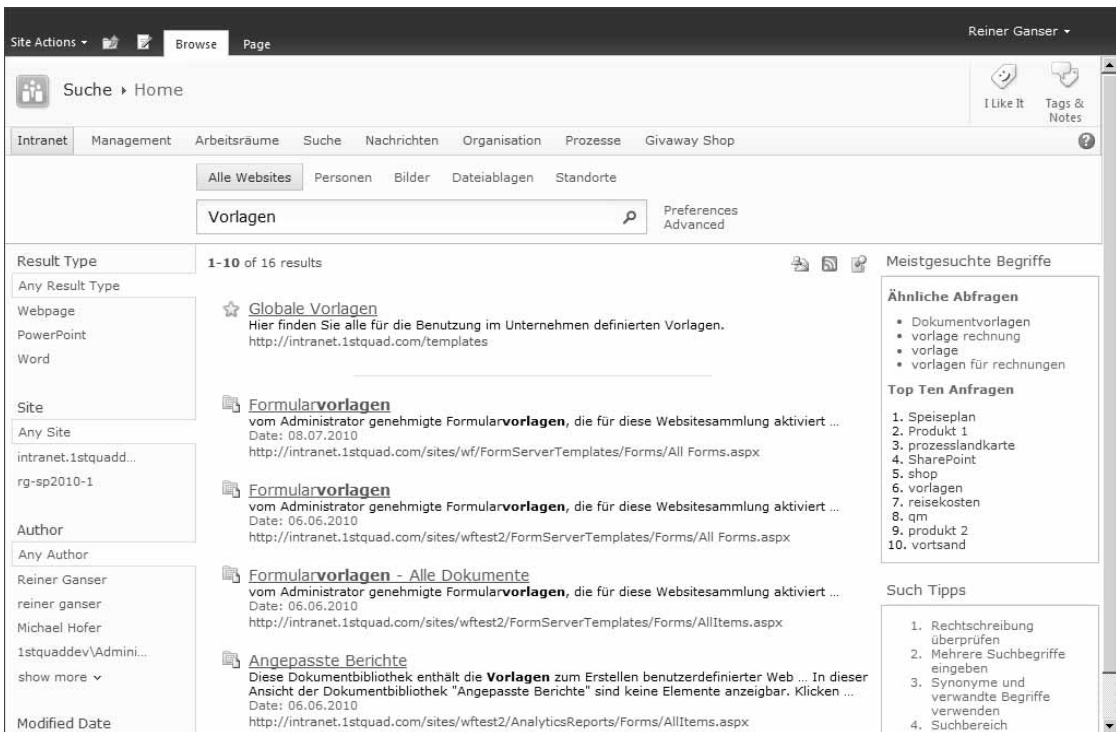
In Abbildung 6.9 sind einige Elemente enthalten, die sich zumeist mit sehr geringem Aufwand umsetzen lassen. Sie bieten jedoch dem Benutzer bereits eine sehr hilfreiche Unterstützung. Im Folgenden werden diese noch etwas näher erläutert:

- **Häufigste Suchabfragen** Der Benutzer kann bereits beim Aufruf der Suchmaske sehen, was die häufigsten Suchanfragen der anderen Benutzer waren. Sollte sich seine Suchanfrage bereits darunter befinden, kann er direkt einen Link anklicken und die passende Suche wird ausgeführt. Idealerweise steht der beste Treffer dann auch zuoberst in der Trefferliste. Dies setzt jedoch voraus, dass der »Search Business Owner« (ist verantwortlich für die Suche im Unternehmen) die besten Suchergebnisse von SharePoint Server 2010 pflegt. Gleichzeitig ist die Liste der häufigsten Suchanfragen ein Indiz für den »Search Business Owner«, welche Begriffe momentan sehr aktuell sind. Dieser kann dadurch die besten Suchergebnisse entsprechend anpassen. Die Umsetzung der häufigsten Suchabfragen kann sehr einfach erfolgen. Hierzu liefert SharePoint Server 2010 ein entsprechendes Webpart mit. Wichtig aus unserer Sicht ist, dass die häufigsten Suchabfragen bereits auf der Suchmaske aufgeführt sind und nicht erst, nachdem der Benutzer eine Suche ausgeführt hat.
- **Zeitgeist** (nicht im Screenshot von Abbildung 6.9 enthalten) Dieses Wort verwenden wir in Analogie zur Umsetzung bei Google. Hier handelt es sich ebenso um die häufigsten Suchabfragen. Diese Liste wird jedoch nicht wie beim vorigen Abschnitt automatisch erzeugt, sondern wird von den »Search Business

Ownern« redaktionell auf der jeweiligen Ebene gepflegt. Dadurch kann der Zeitgeist sowohl unternehmensweit als auch auf Abteilungs- bzw. Teamebene angeboten werden. Beim Zeitgeist handelt es sich um eine SharePoint-Liste, die manuell gepflegt und anhand der Webpartansicht auf der Suchseite angezeigt wird.

- **Suchtipps** Interessanterweise geben die meisten Benutzer in das Suchfeld nur einen Begriff ein. Die Trefferliste ist dementsprechend lang. Nur wenige kommen dann auf die Idee, mehr als ein Wort einzugeben. Solche und andere Suchtipps können das Ergebnis einer Suche erheblich beeinflussen. Es ist daher wichtig, diese dem Benutzer in knapper Form an prominenter Position darzubieten. Wichtig ist hierbei, dass die Suchtipps sehr kurz und prägnant ausfallen.
- **Support und Kontakt (Man in charge)** In vielen Fällen wird der Benutzer, trotz intensiver Bemühungen, nichts oder zu viel oder nicht das Gewünschte finden. Damit er an dieser Stelle nicht alleine gelassen wird, sollte er die Möglichkeit haben, jemanden fragen zu können. Sowohl die Suchabfrageseite als auch die Trefferliste sollte deshalb einen Ansprechpartner nennen, an den sich der Benutzer bei Fragen wenden kann.

Die aufgeführten Elemente sollten dabei auch nach Ausführung der Suche weiterhin angezeigt bleiben, damit sich der Benutzer bei Fragen beispielsweise an den Ansprechpartner wenden kann (Abbildung 6.10). Auf der Trefferliste können weitere Unterstützungen angeboten werden, die es dem Benutzer ermöglichen, entweder weiter mit geänderten oder erweiterten Suchbegriffen zu suchen oder direkt auf die vom »Search Business Owner« vorkonfigurierte Seite zu springen.



**Abbildung 6.10** Suchergebnis mit weiteren Infos für den Benutzer

In der obigen Abbildung 6.10 sind weitere Möglichkeiten zu sehen, die nachfolgend näher erläutert werden.

## Ähnliche Suchabfragen

Anhand der eingegebenen Stichwörter kann das System automatisch ermitteln, welche ähnlichen Suchbegriffe es von anderen Benutzern bereits gegeben hat, die den eingegebenen Begriff enthalten. Dies kann der Benutzer direkt dazu verwenden, um eine erneute Suche abzusetzen. Hierzu kann ein Webpart (Smart Search) von CodePlex genutzt werden (siehe Anhang C).

## Beste Suchergebnisse

Für das eingegebene Stichwort des Benutzers hat der »Search Business Owner« zuvor sog. »Beste Suchergebnisse« (Best Bet) in der Konfiguration der Stichwörter im Portal hinterlegt. Dadurch kann der »Search Business Owner« gezielt auf stark nachgefragte Begriffe reagieren und den Benutzer auf die richtige Seite lenken. Beste Suchergebnisse werden immer zuoberst in der Trefferliste angezeigt.

Die Konfiguration erfolgt zusammen mit einem Stichwort auf Ebene des Portals (Websitesammlung). Pro Stichwort können zusätzlich mehrere Synonyme hinterlegt werden (Abbildung 6.11). Diese kann der »Search Business Owner« beispielsweise anhand der häufigsten Suchabfragen oder durch Auswertung des Suchreports von SharePoint Server 2010 ermitteln.

<p><b>Stichwortinformationen</b></p> <p>Der Stichwortbegriff wird in Suchabfragen verglichen, um ein Ergebnis für die Stichwortsuche zu liefern.</p> <p>Bei Synonymen handelt es sich um Wörter, die Benutzer bei der Suche nach dem Stichwort eingeben können. Trennen Sie diese durch Semikolons.</p>	<p>Stichwortbegriff: *</p> <p>Vorlagen</p> <p>Synonyme:</p> <p>Dokumentvorlagen;dot;Offerten Vorlagen;Rechnungsvorlagen;Vorlage</p>	
<p><b>Beste Suchergebnisse</b></p> <p>Die Einstellung 'Beste Suchergebnisse' wird für dieses Stichwort empfohlen.</p> <p>Beste Suchergebnisse werden in den Ergebnissen in der aufgeführten Reihenfolge angezeigt.</p>	<p>Bestes Suchergebnis hinzufügen</p> <p>Titel</p> <p>Globale Dokumentvorlagen</p>	<p>Reihenfolge</p> <p>Entfernen 1</p>
<p><b>Stichwortdefinition</b></p> <p>Bei der Definition handelt es sich um den optionalen Text, der in den Ergebnissen der Stichwortsuche angezeigt wird.</p>	<p>Globale Vorlagen</p>	
<p><b>Kontakt</b></p> <p>Beim Kontakt handelt es sich um die Person, die benachrichtigt wird, wenn das Prüfdatum des Stichworts abgelaufen ist.</p>		<p>Kontakt:</p> <p>Michael Hofer ;</p>
<p><b>Veröffentlichen</b></p> <p>Geben Sie im Feld 'Startdatum' das Datum ein, ab dem das Stichwort in Suchergebnissen aufgeführt werden soll.</p> <p>Geben Sie im Feld 'Enddatum' das Datum ein, bis zu dem das Stichwort in Suchergebnissen aufgeführt werden soll.</p> <p>Geben Sie im Feld 'Prüfdatum' das Datum ein, an dem das Stichwort vom Kontakt überprüft werden soll.</p>		<p>Startdatum</p> <p>5/27/2011</p> <p>Enddatum (Keine Angabe, wenn nicht zutreffend)</p> <p>Prüfdatum</p>

**Abbildung 6.11** Definition eines Stichworts mit »Beste Suchergebnisse«

Zu einem Stichwort lassen sich dann ein oder mehrere »Beste Suchergebnisse« erfassen. Es handelt sich dabei mehr oder weniger um die Eintragung eines Links mit Titel und Beschreibung (Abbildung 6.12). SharePoint Server 2010 zeigt dieses dann später optisch hervorgehoben mit einem Stern und in einer anderen Farbe an. Ist ein »Bestes Suchergebnis« einmal erfasst, lässt es sich auch für mehrere Stichwörter wiederverwenden.



Bestes Suchergebnis bearbeiten

\* Bezeichnet ein Pflichtfeld.

URL: \*

http://intranet/org/templates

Titel: \*

Globale Dokumentvorlagen

Beschreibung:

Dieser Bereich enthält alle Dokumentvorlagen, die im Unternehmen verwendet werden

OK Abbrechen

**Abbildung 6.12** Erfassung eines besten Suchergebnisses

### Direkte Nutzung von bestehenden Such-/oder Informationssystemen

Um möglichst viele Möglichkeiten beim Aufruf des Suchcenters von SharePoint Server 2010 zu bieten, kann dieser auch mit einem schnellen Zugriff auf bereits bestehende Suchdienste, wie z.B. einem bereits vorhandenen Personenverzeichnis oder einer speziellen Anwendung für eine Standortsuche, erweitert werden. Der Benutzer muss dann nicht erst Favoriten im Browser anlegen, sondern weiß, dass er durch Aufruf des Suchcenters weitere Möglichkeiten hat. Eine einfache Integration besteht z.B. durch Verwendung einer zusätzlichen Webpartseite und des mitgelieferten Seiten-Viewer-Webparts. Zugegebenermaßen sind dies sehr einfache Integrationen, aber oft entscheiden die Kleinigkeiten über den Erfolg eines Systems und in diesem Fall ist der Aufwand wirklich minimal. Die Abbildung 6.13 zeigt dies am Beispiel einer bereits im Unternehmen bestehenden Standortsuche, welche über ein Seiten-Viewer-Webpart in das Suchcenter eingebunden ist.

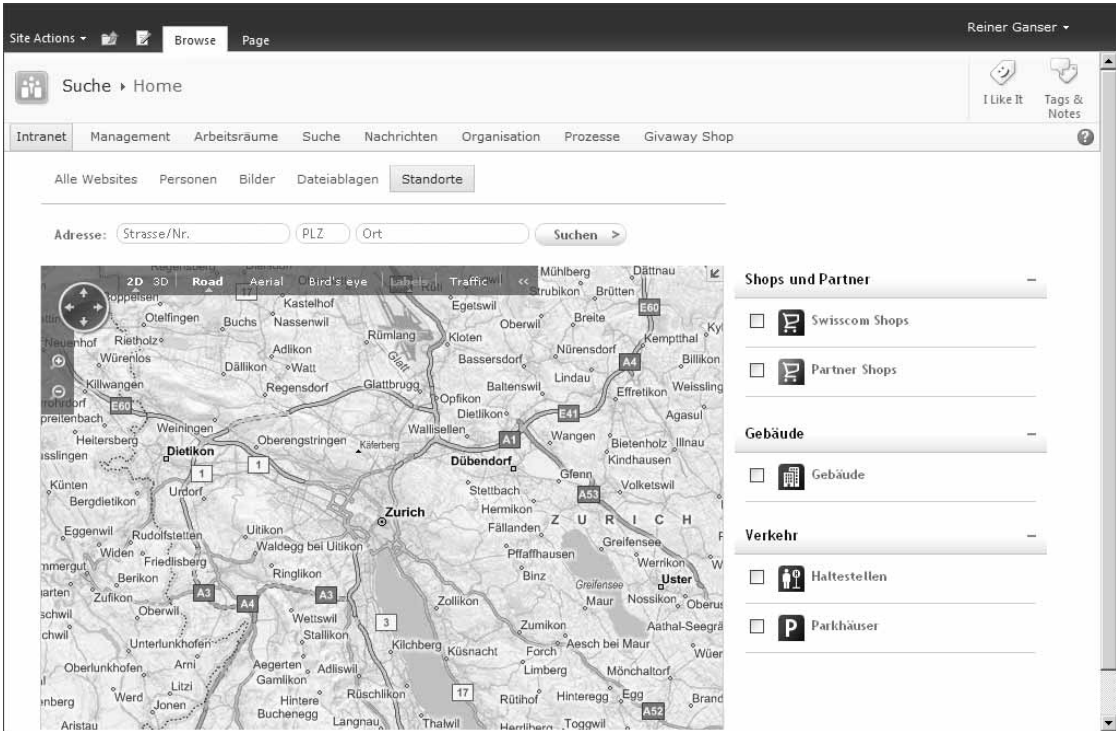


Abbildung 6.13 Nutzung von Fremdsystemen am Beispiel einer Standortsuche

### Federated Search

Um die Ergebnisse für den Benutzer weiter aufzuwerten, können auch Ergebnisse aus anderen Systemen für ihn sehr hilfreich sein. In SharePoint Server 2010 wird dies durch die sog. Federated Search (Sammelsuche) ermöglicht. Gemeint ist damit die Möglichkeit, andere Suchmaschinen oder Repositories anzufragen, die ihrerseits dann ein Ergebnis zurückliefern. D.h., dass SharePoint Server 2010 an dieser Stelle keine Indizierung des anderen Systems vornimmt, sondern dessen Suchfunktion nutzt. Das Ergebnis wird auf der gleichen Ergebnissseite dargestellt wie das der SharePoint Server 2010-Suche. Die beiden Ergebnisse werden jedoch nicht gemischt, sodass Ergebnisse aus der Federated Search nicht in die Relevanz eingehen. Eine andere Möglichkeit ist die Nutzung von Webparts, um die Ergebnisse der Federated Search anzuzeigen. Der große Unterschied zwischen der Nutzung der SharePoint Server 2010-Suche und Federated Search ist, dass in letzterem Fall der Inhalt des Fremdsystems nicht von SharePoint Server 2010 indiziert wird. Die Tabelle 6.6 zeigt die Vorteile der beiden Methoden auf.

Indizieren des Inhalts mit SharePoint Server 2010	Federated Search
Sortieren der Trefferliste nach Relevanz	Es bestehen keine zusätzlichen Speicheranforderungen für den Inhaltsindex, da diese nicht auf dem Suchserver gespeichert wird
Kontrolle über die Aktualisierungsfrequenz des Inhaltsindexes	Die Vorteile der anderen Suchmaschine können genutzt werden

Tabelle 6.6 Gegenüberstellung der Vorteile von SharePoint 2010-Suche und Federated Search

Indizieren des Inhalts mit SharePoint Server 2010	Federated Search
Definition der Metadaten, die indiziert werden sollen	Die andere Suchmaschine kann auf die speziellen Gegebenheiten des anderen Systems angepasst werden und liefert dadurch spezifischere Treffer und ist dadurch ggf. schneller
Es kann ein einziges Backup für Inhalt und Index durchgeführt werden	Auf andere Systeme kann zugegriffen werden, auf die man eigentlich keinen Zugriff hat, welche aber über die Suche zugänglich sind

**Tabelle 6.6** Gegenüberstellung der Vorteile von SharePoint 2010-Suche und Federated Search (Fortsetzung)

Die Anbindung an das andere System erfolgt über sog. Federated Locations, bei denen es sich entweder um den lokalen Suchindex von SharePoint Server 2010 oder um Lokationen, die der OpenSearch 1.0/1.1-Spezifikation entsprechen, handeln kann. OpenSearch 1.0/1.1 ist ein RSS-Format und somit leicht anzuwenden. Es existiert bereits ein vorgefertigtes Webpart zur Anzeige der Ergebnisse aus der Federated Search. Es werden bereits einige Federated Search-Connectors zum Download angeboten (siehe <http://technet.microsoft.com/en-us/enterprisearch/ff727944.aspx>). Die Umsetzung eigener Connectors ist ebenfalls möglich, sodass auf diesem Weg auch die Anbindung an ein eigenes Repository möglich ist. Die Abbildung 6.14 enthält ein Beispiel einer Ergebnisseite, die sowohl Treffer aus einer Suche in SharePoint Server 2010 anzeigt als auch den Treffer aus einer Suche nach Inhalten aus dem Internet enthält.

The screenshot shows a SharePoint 2010 search results page. The top navigation bar includes 'Websiteaktionen', 'Durchsuchen', and 'Seite'. The search bar contains the text 'SharePoint'. Below the search bar, there are tabs for 'Alle Websites', 'Personen', 'Bilder', 'Dateiablagen', and 'Standorte'. The search results are displayed in a list format, showing a mix of local intranet results and federated search results from Google News. The local results include 'SharePoint Hilfe', 'SharePoint', '1stQuad Intranet', and 'SharePoint Training Portal'. The federated results include 'Google News Results' with articles such as 'Inside SharePoint 2010, Part 2: Internet Web Sites - Paul Thurrott's SuperSite for Windows' and 'ZSL Introduces New SharePoint Add-On for Changing User Passwords - San Francisco Chronicle (press release)'. The page also features a sidebar on the left with filters for 'Ergebnistyp', 'Webseite', 'Word', 'PowerPoint', 'Website', 'Autor', and 'Weitere anzeigen'.

**Abbildung 6.14** Kombination mit Federated Search auf Internetinhalte



### Quick Fact – Nutzen versus Technik

Bei der Optimierung der Suche von SharePoint Server 2010 ist stets der Nutzen für den Benutzer im Auge zu behalten. Oft sind es Kleinigkeiten, wie z.B. ein Ansprechpartner oder Suchtipps, die zu einer besseren Akzeptanz beitragen. Die Verbesserungen betreffen dabei nicht nur die Konfiguration der Suchmaschine durch den Administrator. Ebenso muss die Oberfläche für den Benutzer so gestaltet sein, dass er weitere Unterstützung bekommt und sogar im besten Fall gar keine Suche mehr absetzen muss, weil er die notwendigen Informationen bereits auf der Suchmaske ablesen kann. Bei stark nachgefragten Begriffen kann der »Search Business Owner« das schnelle Finden der Information durch die Definition von sog. »Besten Suchergebnissen« optimieren und ermöglicht dadurch dem Benutzer das zielgerichtete Anspringen der benötigten Information, ohne sich durch mehrere Links in der Trefferliste zu klicken.

## Relevanz und Präzision

Der Aufbereitung nach Relevanz widmet sich in SharePoint eine Reihe, speziell an die Gegebenheiten in Unternehmen, angepasster Mechanismen. Die übliche Linkpopularität ist dort zumeist nicht relevant, sodass andere Algorithmen herangezogen werden müssen. SharePoint Server 2010 wertet daher automatisch die URL-Tiefe eines Dokuments aus, also wie weit unten in einem gedachten Verlinkungs- bzw. Dokumentbaum ein Dokument angesiedelt ist. Je weiter es von der Wurzel entfernt ist, als desto weniger relevant wird es eingestuft. Gleiches gilt für die sog. Klickdistanz, welche die Entfernung von als »autoritativ« vermerkten Dokumenten misst und nach zunehmender Entfernung die Relevanz weiter nach unten stuft. Höher eingestuft in der Relevanz werden die Hyperlinkwörter, also jener Text, welcher mit einem Hyperlink versehen ist. Für die Ermittlung der Relevanz werden außerdem Metadaten, wie z.B. der Titel von Dokumenten, herangezogen. Ein Kuriosum ist das sogenannte »File Type Biasing«. Dieses sorgt für eine unterschiedliche Gewichtung von Suchergebnissen je nach Typ des zugrunde liegenden Dokuments in der absteigenden Reihenfolge HTML, PowerPoint, Word, XML, Excel, Text (siehe auch Tabelle 6.7).

Relevanz	Beschreibung	Statisch/Dynamisch
URL-Übereinstimmung	Das Suchwort entspricht genau dem URL (z.B. Name der Website)	Dynamisch
Wörter im Hyperlink	Kommt das Suchwort direkt im Hyperlink vor?	Dynamisch
Gewichtung der verwalteten Eigenschaften	Einige Eigenschaften sind wichtiger als andere. Dies kann programmtechnisch angepasst werden.	Dynamisch
Titel in der Kopfzeile	Steht das Suchwort direkt in der Kopfzeile des Dokuments, erscheint dieses höher in der Rangfolge	Dynamisch
Klickdistanz	Wo ist der Link im Text platziert?	Statisch
URL-Tiefe	Objekte mit kurzer URL haben höhere Gewichtung als Objekte mit langer (tiefer) URL	Statisch

**Tabelle 6.7** Einflussfaktoren für die Relevanz

Relevanz	Beschreibung	Statisch/Dynamisch
File Type Biasing	Unterschiedliche Gewichtung anhand der Dateierweiterung. Folgende Rangfolge wird benutzt (Relevanz absteigend): HTML-Webseiten PowerPoint-Präsentationen Word-Dokumente XML-Dateien Excel-Arbeitsmappe Textdateien Listenelemente	Statisch
Spracherkennung	Beim Absetzen der Suche ermittelt SharePoint die Sprache des Benutzers anhand der Einstellungen im Internet Explorer Fremdsprachen sind weniger relevant, als die Sprache des Benutzers Die englische Sprache wird immer als relevante Sprache für den Benutzer angesehen	Statisch

**Tabelle 6.7** Einflussfaktoren für die Relevanz (Fortsetzung)

## Optimierung der Relevanz

Der Administrator hat mehrere Möglichkeiten, Anpassungen an der Relevanz vorzunehmen: Er definiert Synonyme für häufige Suchwörter, nimmt Stichwortdefinitionen vor, gibt zu häufigen Anfragen gezielt das »Beste Suchergebnis« (Best Bet) vor. Duplikate werden vom System automatisch ausgefiltert oder wahlweise ausgewiesen.

Optionen zur Verbesserung der Relevanz durch den Administrator der Suche:

- **Einstellen der Gewichtung auf Seitenebene über »Autorisierende Seiten« und »Nicht autorisierende Webseiten«** Diese Einstellung wird in der Zentraladministration von SharePoint Server 2010 durchgeführt und gilt somit über alle Portale hinweg. Seiten bzw. Webseiten, die in »Autorisierende Seiten« (konfigurierbar in der Zentraladministration) aufgeführt sind, werden weiter oben in der Treffermenge dargestellt, wogegen Links in »Nicht autorisierende Websites« herabgestuft werden.
- **Herausfiltern von Schlüsselwörtern aus den Protokolldateien** SharePoint Server 2010 protokolliert die durchgeführten Suchen und macht diese in Form eines Reports dem Administrator in der Zentraladministration verfügbar. Diese Daten können dazu dienen, um die Stichwörter und »Beste Suchergebnisse« zu definieren oder die entsprechenden Links prominent, z.B. auf der Homepage des Portals, zu platzieren. Da die Portaladministratoren, die die Stichwortdefinitionen und »Beste Suchergebnisse« konfigurieren, im Normalfall keinen Zugriff auf die zentrale Administration von SharePoint Server 2010 haben, muss hier eine aktive Zusammenarbeit zwischen dem Portal- und dem Farmadministrator erfolgen. Da sich die Reporte sehr einfach im Excel- oder PDF-Format exportieren lassen, können die Portaladministratoren oder »Search Business Owner« mit diesen Dokumenten versorgt werden. Als ebenfalls sehr praktische Vorgehensweise hat sich die Ablage der Reporte in einer speziell dafür eingerichteten Website für die »Search Business Owner« erwiesen.
- **Regelmäßiges Anpassen der Schlüsselwörter, »Beste Suchergebnisse« und der Inhaltspriorisierung** Dies wird von den »Search Business Ownern« oder den Portaladministratoren manuell anhand der Protokolldateien durchgeführt

- **Rangfolge und Rangfolge** Die Gewichtung und Berechnung der Rangfolge kann zum einen über die »Autorisierende Seiten« in der Zentraladministration oder über die Umsetzung von eigenen Rangfolge-modellen angepasst werden

Doch nicht nur der Administrator der Suchmaschine von SharePoint Server 2010 kann zur Optimierung der Relevanz von Suchergebnissen beitragen. Ebenso haben die Autoren einen entscheidenden Einfluss auf die Ergebnisse, indem sie die Dokumente entsprechend benennen und ablegen.

Optionen für die Verbesserung der Relevanz durch die Inhaltsanbieter (Content Provider) und Autoren:

- **Dokumenttitel aussagekräftig benennen** Der Indexer von SharePoint Server 2010 legt die Dokumenttitel ebenfalls im Index ab. Durch Verwendung von entsprechenden Namen, die durch Leerzeichen getrennt sind, lässt sich später danach recherchieren. Die Dokumentnamen haben eine relativ hohe Priorität für die Relevanz.
- **Leerzeichen in URL-Dateinamen einfügen** Durch das Einfügen von Leerzeichen im URL kann der SharePoint Server 2010-Indexer die Wörter einzeln für die spätere Suche verfügbar machen. Suchwörter, die im URL vorkommen, haben eine hohe Relevanz in der Trefferliste.
- **Autoren timer durchgängig und korrekt erfassen** Sucht ein Anwender nach einem Namen und findet die Suchmaschine von SharePoint Server 2010 ein Dokument, welches im Autorfeld diesen Namen aufweist, bekommt dieses Dokument eine höhere Relevanz. Deshalb sollten die Autorennamen in den Dokumenteigenschaften immer korrekt ausgefüllt werden.
- **Schlüsselwörter zu Office-, PDF- und HTML-Dokument hinzufügen** Viele Dokumentformate erlauben das zusätzliche Beschreiben von Eigenschaftswerten. Diese werden ebenfalls vom Indexer herausgelöst und in den Suchindex geschrieben. Sie stehen somit auch für die Suche zur Verfügung.
- **Wichtige Inhalte nicht zu tief in der Hierarchie der Webseite ablegen** Je weiter oben sich ein Element im Hierarchiebaum innerhalb einer Websitesammlung befindet, desto höher ist die Relevanz.
- **Bewerten von Inhalten** SharePoint Server 2010 ermöglicht die Bewertung von Dokumenten. Dokumente, welche als gut bewertet wurden, erscheinen höher in der Trefferliste als Dokumente, die zwar inhaltlich ebenfalls relevant sind, jedoch keine oder schlechtere Bewertungen haben.
- **Taggen von Inhalten** In SharePoint Server 2010 können nahezu alle Inhalte mit sog. Tags versehen werden. Dies sind Begriffe, die vom Benutzer frei vergeben oder auch gemeinschaftlich wiederverwendet werden können. Vor allem Dokumente, bei denen in seinen Tags das gesuchte Schlüsselwort vorkommt, erscheinen ebenfalls höher in der Trefferliste.

---

**HINWEIS** Die Verbesserung der Relevanz ist nicht durch einmaliges Optimieren erledigt. Vielmehr ist dies ein kontinuierlicher Prozess, für den es kein Ende gibt. Die »Search Business Owner« müssen immer wieder die angefallenen Daten (z.B. häufigste Suchbegriffe) auswerten und die Einstellungen der SharePoint Server 2010-Suche anpassen. Es lohnt sich also, für diese Aufgabe entsprechend Zeit vorzusehen. Nur so wird gewährleistet, dass die Suche auch nach mehreren Monaten oder Jahren noch Akzeptanz besitzt. Die Erfahrung von mehreren Suchprojekten hat gezeigt, dass bei Nichtbeachtung dieses Punkts die Suche relativ schnell wertlos für die Benutzer wird und einen schlechten Ruf bekommt. Dabei ist es letztlich egal, welches System eingesetzt wird.

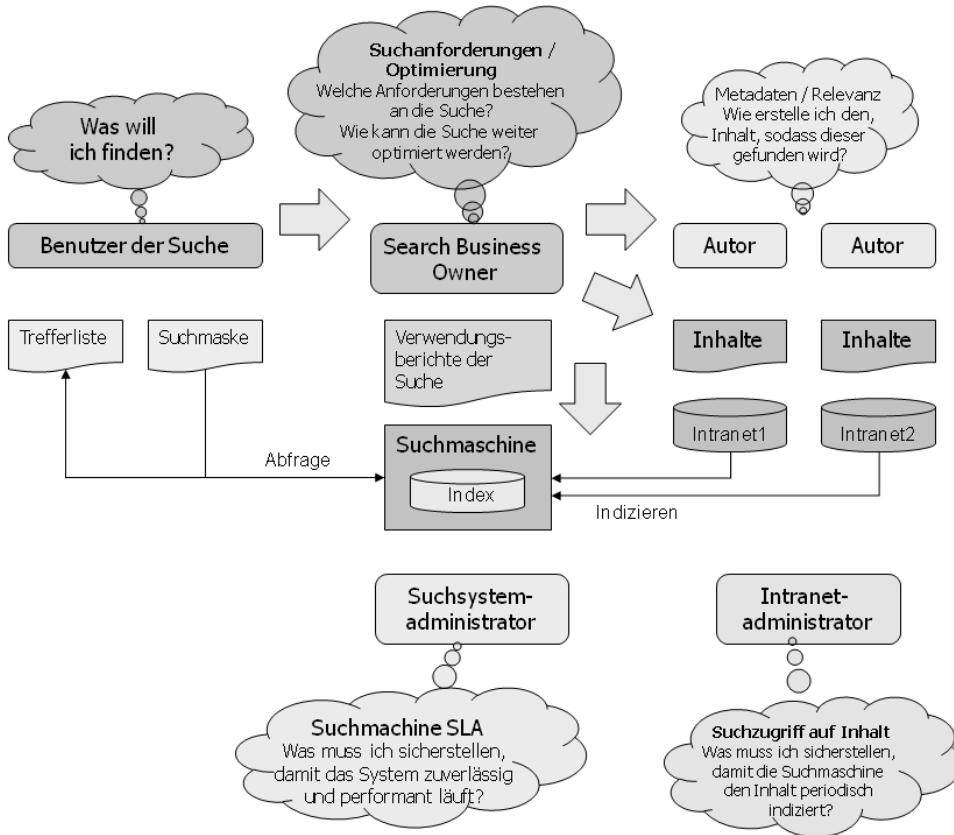
---

**Quick Fact – Relevanz und Präzision**

Wie weit oben ein Objekt in der Trefferliste von SharePoint Server 2010 steht, hängt von verschiedenen Faktoren ab. Angefangen von der Platzierung des Objekts in der Hierarchie bis zur programmatischen Anpassung der Relevanz ist ein breites Spektrum gegeben. Die Optimierung obliegt dabei nicht nur dem Administrator der Suchmaschine, sondern ist vielmehr ein Zusammenspiel von mehreren Rollen. Diese müssen an der Optimierung der Suche und Indizierung kontinuierlich gemeinschaftlich arbeiten.

**Rollen für eine kontinuierliche Qualität der Suchresultate**

Nachhaltig gute Suchergebnisse sind nur möglich durch ein Zusammenspiel von verschiedenen Elementen, welche sowohl technisch wie auch organisatorisch ganzheitlich betrachtet und verwaltet werden müssen. Dabei muss ein Zusammenspiel aus Endbenutzern, Inhaltsautoren und Technikern entstehen. Die Abbildung 6.15 zeigt dieses Zusammenspiel in einer Übersicht mit den beteiligten Personen und den für die jeweilige Rolle zu stellenden Fragen.



**Abbildung 6.15** Zusammenspiel der verschiedenen Rollen in einem Suchsystem

Die Tabelle 6.8 beschreibt die notwendigen Rollen noch etwas genauer.

Rolle	Beschreibung
Benutzer der Suche	Endbenutzer der Suche. Seine Suchanfragen werden aufgezeichnet und im Kontext verwendet, um »Short Tail«-Informationen für weitere Benutzer zu erzeugen.
Search Business Owner	Versteht den Kontext und ist damit quasi das Bindeglied zwischen Benutzer, Autoren und Technik
Autor	Der Autor des Inhalts ist daran interessiert, dass seine erzeugten Inhalte gefunden werden. Durch entsprechende Namensgebung und Platzierung in der Hierarchie kann er die Relevanz der Treffer erheblich mitgestalten.
Suchsystemadministrator	Betreibt die Plattform und optimiert die Performance (z.B. durch Einsatz weiterer Server). Er hat Zugriff auf die Zentraladministration und damit auf die Abfrage- und Ergebnisreporte der Benutzer.
Intranetadministrator	Er kann die Relevanz für sein Portal (Websitesammlung) durch die Konfiguration von Stichwörtern und »Besten Suchergebnissen« beeinflussen. Zusätzlich kann er Suchbereiche konfigurieren, um die Suche einzuschränken.

**Tabelle 6.8** Benötigte Rollen zur Erreichung einer konsistenten Qualität der Suchresultate

Aus Abbildung 6.15 und Tabelle 6.8 wird ersichtlich, dass für eine schnelle Suche mit einer guten Relevanz die Zusammenarbeit von verschiedenen Stellen im Unternehmen notwendig ist. Es ist in diesem Fall sehr wichtig, dass die Beteiligten ihre Rolle und deren Einflussmöglichkeiten kennen. Als zentrale Stelle ist sicherlich der »Search Business Owner« zu sehen. Er erkennt am schnellsten, dass gesuchte Informationen von den Benutzern nicht gefunden werden und kann dementsprechend weitere Schritte einleiten, wie z.B. die Konfiguration von weiteren »Besten Suchergebnissen« veranlassen oder selbst konfigurieren, Autorenrichtlinien erstellen, eine eventuelle Performancesteigerung bei der Technik anfragen usw.

## Indizierung von bestehenden Webseiten

Mit der Einführung von SharePoint Server 2010 werden zumeist nicht schlagartig alle bestehenden Websites im Unternehmen umgestellt. Dies sind oft erst die nächsten Schritte. In der Zwischenzeit wäre es aber wünschenswert, wenn alle Informationen in diesen Websites ebenfalls in der Suche verfügbar wären. Die Problematik dabei ist meist, dass die bestehenden Intranetauftritte auf unterschiedlichen Systemen realisiert wurden. Besonders wird dies angetroffen, wenn Abteilungen sehr autark agieren können oder wenn mehrere Geschäftsbereiche oder Firmen zusammengeführt werden. SharePoint Server 2010 mit seiner integrierten Suchfunktionalität wird dabei schnell als potenzielle Suchmaschine für die bestehenden Systeme identifiziert und soll eine komfortable Suchfunktion für die Webseiten von bestehenden Intranets anbieten. Die Gründe für diese Vorgehensweise sind unterschiedlich und reichen von »bisher keine Suchfunktion vorhanden im bestehenden CMS-System« bis »Unternehmenssuche, die alle Intranetinhalte durchsucht«.

Auf den ersten Blick scheint dies nicht sonderlich problematisch zu sein: SharePoint Server 2010 bringt bereits einen sog. Protokollhandler (Möglichkeit des Zugriffs auf das Quellsystem) für Websites mit. Trotzdem fallen die Ergebnisse einer Suche, nachdem die Quelle indiziert wurde, sehr unbefriedigend aus. Schnell wird dann erst einmal die Suchfunktionalität von SharePoint Server 2010 in Frage gestellt. Meist ist aber nicht SharePoint Server 2010 das Problem, sondern die Quellsysteme sind schlicht und einfach nicht optimiert für die Indizierung. Durch teilweise einfache Maßnahmen lässt sich jedoch die Qualität der Suchergebnisse erheblich steigern. Gerade wenn die Inhaltserzeugung selbst gesteuert werden kann, lässt sich einiges verbessern. Nachfolgend sind einige dieser Maßnahmen aufgeführt:

- **Verwendung von Indexdateien** Diese enthalten die Links auf die Seiten bzw. Dokumente, die vom SharePoint Server 2010-/Search Server 2010-Indexer indiziert werden sollen. Folgende Gründe sprechen für den Einsatz von Indexdateien:
  - Durch Indexdateien lässt sich sehr einfach steuern, welche Inhalte indiziert werden sollen und welche nicht. Lässt man den Crawler selbst die Links finden, sind oft mühselige Kleinarbeiten und Tests notwendig, bis die optimalen Einstellungen gefunden werden. Meistens sind auch mehrere Inhaltsquellen und Regeln für Crawlerauswirkungen notwendig, bis die Suche ohne Nutzung von Indexdateien die besten Ergebnisse liefert.
  - In den Webseiten wird viel JavaScript verwendet, um z.B. die Verlinkung zu erzeugen. Dies wird vor allem bei der Menüführung oft genutzt. Der SharePoint-Crawler kann allerdings mit diesen Links nichts anfangen, da diese erst auf dem Client verfügbar sind (Google geht es übrigens nicht anders).
- **Aufbau der Indexdateien** Einige Hinweise sollten hier beachtet werden:
  - Wollen Sie mehrsprachige Intranetauftritte indizieren, sollte für jede Sprache eine separate Indexdatei genutzt werden. In diesen Indexdateien sollten sich dann nur Links der jeweiligen Sprache befinden.
  - Für den Namen der Indexdatei sollten einfache URLs ohne Query-Parameter genutzt werden (z.B. *index.htm* oder *index-de.htm*, *index-en.htm* usw.)
  - Die Indexdateien sollten möglichst aktuell und mit den Indizierungszeitplänen abgestimmt sein
  - Möchten Sie nicht, dass die Indexdatei selbst im Suchergebnis auftaucht (und das ist vermutlich der Normalfall), dann sollte der folgende Meta-Tag vorhanden sein:

```
<meta name="robots" content="noindex, follow">
```

- Er bewirkt, dass zwar die Links verfolgt werden, die Indexdatei selbst jedoch nicht in den Index aufgenommen wird
- Sinnvollerweise sollte das letzte Änderungsdatum als Meta-Tag in die Indexdatei aufgenommen werden. Dadurch ist schnell zu erkennen, ob die Indexdatei aktuell ist. Dies unterstützt bei der Fehlersuche enorm, falls keine entsprechenden Treffer im Suchergebnis angezeigt werden.
- Der Head-Tag der Indexdatei sollte somit beispielsweise zumindest die folgenden Angaben enthalten:

```
<head>
  <meta name="robots" content="noindex, follow">
  <meta name="Date" content="2008-04-18T22:00:35+01:00">
</head>
```

- SharePoint indiziert Inhalte eines bestimmten URLs nur einmal. Dies ist sehr sinnvoll, bereitet aber gerade bei mehrsprachigen Auftritten oder bei der Nutzung zentraler Ressourcen in verschiedenen Auftritten Probleme. In diesen Fällen ist man nicht in der Lage, den gemeinsam genutzten Inhalt (z.B. Organigramm des Unternehmens) in mehreren Suchbereichen zu nutzen, da nur einmal indiziert wird. Umgehen lässt sich dies, indem der gemeinsame Inhalt jeweils mit eindeutigen URLs verlinkt wird. Der Indexer von SharePoint Server 2010 bzw. Microsoft Search Server 2010 indiziert diesen Inhalt dann mehrfach und stellt ihn in den entsprechenden Inhaltsquellen und somit in den Suchbereichen zur Verfügung. Zum Beispiel könnte ein zentral genutztes Organigramm der Firma unter der URL <http://<server>/central/orgchart.htm> in verschiedenen Sprachauftritten mit den folgenden URLs verlinkt werden:

- `http://<server>/central/orgchart.htm?lang=de`
- `http://<server>/central/orgchart.htm?lang=en`
- `http://<server>/central/orgchart.htm?lang=fr`
- `http://<server>/central/orgchart.htm?lang=it`
- Bei den Links innerhalb der Indexdatei sollte es sich um die endgültigen Links handeln. Links, die erst eine Weiterleitung ausführen, sollten vermieden werden.
- **Verwendung von Meta-Tags** SharePoint Server 2010/Search Server 2010 können Standard-Meta-Tags in den Webseiten verwenden. Ein Meta-Tag ist hier vor allem wichtig:
  - **Author** (z.B. `<meta name="author" content="Reiner Ganser">`): Dieses Feld wird in der Trefferliste angezeigt
  - Der Head-Tag der Webseiten sollte somit beispielsweise zumindest die folgenden Angaben enthalten:

```
<head>
  <meta name="Author" content="Reiner Ganser">
</head>
```

- **Verwendung von sprechenden Namen für den Seitennamen** Die SharePoint Server 2010-Suchmaschine stuft Seiten, bei denen das gesuchte Wort im Seitennamen vorkommt, erheblich relevanter ein. Dadurch erscheinen diese auch weiter oben in der Trefferliste. Kryptische Seitennamen (z.B. `ap2537651.htm`) haben quasi keine Bedeutung für die Relevanz in der Suche.

#### HINWEIS

In der Trefferliste der SharePoint Server 2010-Suche wird das Datum der letzten Änderung bei jedem Treffer angezeigt. Nach diesem Datum lässt sich auch sortieren. Bei der Indizierung der Inhalte von SharePoint Server 2010 stammen diese Daten von internen Systemfeldern. Woher holt sich aber SharePoint Server 2010 diese Information, wenn externe Webseiten indiziert sind? Die folgende Aufstellung bringt etwas Licht ins Dunkel:

- Wird das Datum im HTTP-Header »Last-Modified« im Format »Tue, 08 Nov 2011 14:59:42 GMT« mitgeliefert, wird dieses Datum korrekt im Suchergebnis angezeigt
- Wird ein falsches Datumsformat (z.B. 2011-01-01T09:50:08+01:00) für den HTTP-Header »Last-Modified« verwendet, wird kein Datum angezeigt
- Wird kein entsprechender HTTP-Header mitgeliefert, wird das Datum des letzten Crawls angezeigt

Laut W3C-Spezifikation kann im HTML-Header folgendes Meta-Tag angegeben werden:

```
<meta http-equiv="last-modified" content="Sun, 06 Nov 2011 14:59:42 GMT">
```

Dieses bewirkt quasi das Setzen des HTTP-Headers (siehe <http://www.w3.org/TR/html401/struct/global.html>). Leider wird dies vom SharePoint Server 2010-Indexer ignoriert. Somit ist die einzige Möglichkeit, das Datum der letzten Modifikation selbst zu bestimmen, nur durch Setzen des HTTP-Headers »Last-Modified« möglich.

Falls Sie also externe Websites indizieren und das korrekte Änderungsdatum angezeigt werden soll, dann muss das Quellsystem den HTTP-Header »Last-Modified« setzen. Nur dann lässt sich nach dem Änderungsdatum sortieren.

Ist das Setzen im Quellsystem nicht möglich, besteht die Möglichkeit, einen Proxy zwischen den SharePoint Server 2010-Crawler und dem Quellsystem zu schalten und diesen dann den HTTP-Header anhand von Meta-Tags, die im HTML-Dokument mitgeliefert werden, setzen zu lassen.

Durch die beschriebenen Maßnahmen lässt sich die Qualität der Indizierung und somit der Suche innerhalb bestehender Webseiten im Unternehmen erheblich steigern. Gerade wenn die Erzeugung dieser Webseiten selbst gesteuert werden kann, sind die beschriebenen Änderungen zumeist einfach durchzuführen bzw. einzubauen. In den meisten CMS-Systemen bedeutet es wenig Aufwand, eine Indexdatei mit den verwendeten Seiten und Links auf Dokumente automatisch zu erstellen.

## Einbinden der SharePoint Server 2010-Suche in bestehende Websites

Im vorigen Abschnitt wurden die Optimierungsmöglichkeiten für die Indizierung von bestehenden Websites beschrieben. Nach diesem Schritt liegt es nahe, die SharePoint Server 2010-Suche auch in den bestehenden Systemen als Suchmaschine zu nutzen. Die Einbindung kann dabei sehr einfach anhand von parametrisierten URL-Aufrufen erfolgen.

### Aufrufparameter für die Suche

Die SharePoint Server 2010-Suche kennt entsprechende Parameter in der URL der Suchergebnisseite. Die Einbindung in die bestehenden Intranets kann dadurch sehr einfach erfolgen, indem die Parameter programmieretechnisch beim Anklicken einer Schaltfläche oder eines Links an die URL angehängt werden. Im einfachsten Fall genügen die Parameter in Tabelle 6.9.

Parameter	Beschreibung	Beispiel
k=keyword	Suchbegriff, nach dem gesucht werden soll	<i>k=sharepoint</i>
s=scope	Suchbereich, in dem gesucht werden soll	<i>s=abteilung</i>

**Tabelle 6.9** Einfache Aufrufparameter

Die Aufrufparameter werden als Abfragezeichenfolge (Query String) hinter den Basis-URL geschrieben:

Beispiel: *http://<moss-server>/seiten/results.aspx?k=sharepoint*

Gesucht wird nach dem Schlüsselwort »sharepoint« im Standardsuchbereich.

Die einzelnen Parameter können auch kombiniert werden:

Beispiel: *http://<server>/DE/seiten/results.aspx?k=sharepoint&s=abteilunga*

Gesucht wird nach dem Schlüsselwort »sharepoint« im Suchbereich »abteilunga«.

Die obigen Parameter ermöglichen eine sehr schnelle Einbindung der SharePoint Server 2010-Suche in bestehende Websites. Es sind jedoch noch deutlich mehr Suchparameter möglich, wie in Tabelle 6.10 angedeutet.

Parameter	Beschreibung	Beispiel
v=date/relevance	Sortieren nach Datum oder Relevanz	<i>v=date</i>
start=pageno	Bestimmte Seite aus dem Suchergebnis anzeigen	<i>start=12</i>
k=propertyfilter	Nach Eigenschaften filtern	<i>k=author:rganser –site:http://intranet1</i>

**Tabelle 6.10** Weitere URL-Parameter, die für Integration in bestehende Systeme hilfreich sind



Auch hierzu einige Beispiele:

Beispiel 1: *http://<sp-server>/DE/seiten/results.aspx?k=sharepoint&v=date*

Gesucht wird nach dem Schlüsselwort »sharepoint« im Standardsuchbereich sortiert nach dem Änderungsdatum.

Beispiel 2: *http://<sp-server>/seiten/results.aspx?k=author:rganser -site:http://intranet1*

Gesucht wird nach dem Schlüsselwort »rganser« in der Eigenschaft »author« jedoch nur, wenn sich die gefundene Webseite nicht innerhalb URL *http://intranet1* befindet.

### Quick Fact – Indizierung von bestehenden Websites

SharePoint Server 2010 ist in der Lage, externe Websites (z.B. bestehende Intranetauftritte) zu indizieren. Oftmals fällt jedoch das Ergebnis eher enttäuschend aus. SharePoint Server 2010 ist aber in den meisten Fällen gar nicht der Schuldige, sondern die Quellsysteme sind schlicht und einfach nicht für die Indizierung ausgelegt. Dadurch werden Inhalte erst gar nicht erfasst, weil der sog. Crawler sie gar nicht aufspüren kann oder die indizierten Inhalte tauchen in der Trefferliste sehr weit hinten auf, weil der Indexer die Relevanz nicht korrekt eingestuft hat.

Durch meist sehr einfache Maßnahmen an den Inhalten der Quellsysteme lässt sich die Indizierung und Suche jedoch deutlich verbessern. Ist die Indizierung und die SharePoint Server 2010-Suche für externe Websites optimiert, kann diese sehr einfach in die bestehenden Systeme eingebaut werden, um eine Suchmaschine für diese Websites anzubieten.

Nachdem die Grundlagen für den Aufbau, den Betrieb und die Durchsuchbarkeit von Portalen in den letzten drei Kapiteln gelegt wurden, geht es im folgenden Kapitel um das gute Design für ein Portal.