

Christoph Czupy, Arian Kuschki

# Enterprise Search – Anspruch trifft Realität – Tipps aus der Praxis

Fast keine Webseite kann es sich heutzutage leisten, auf eine eigene Suchfunktion zu verzichten. Doch die Ansprüche sind in den letzten Jahren kontinuierlich gewachsen. Fällt die Suche Ihrer Webseite zu sehr hinter die Erwartungen der anspruchsvollen Nutzer zurück? Verlassen diese oft schnell Ihre Seite und kommen – im schlimmsten Fall – nicht wieder zurück? Die folgenden Best Practices für eine bessere Sucherfahrung der Experten Christoph Czupy und Arian Kuschki zeigen Ihnen, was bei einer Webseiten-Suche im Jahr 2014 möglich ist.



„Suche“ ist eines der wichtigsten Paradigmen bei der Nutzung digitaler Inhalte, Dienste und Anwendungen geworden. Betriebssysteme wie Windows 7 oder Mac OSX nutzen vermehrt ein suchzentriertes User Interface, ebenso bekannte Webseiten wie Google (Suche, Mail und Maps), Amazon, Ebay und Airbnb. Webseiten aus vielen verschiedenen Bereichen wie arte.tv oder calgary.ca (die offizielle Webseite der kanadischen Stadt) setzen auf ein prominentes Suchfeld in bester Bildschirmlage. Und das kommt nicht von ungefähr. In den frühen Tagen des World Wide Web wurde hauptsächlich auf umfangreiche Kategorienbäume gesetzt. Erinnern Sie sich noch an das Open Web Directory? Der Aufwand, sinnvolle Strukturen manuell zu pflegen, steigt bei diesem System mit der Informationsmenge aber unverhältnismäßig. Das ist der Reiz und das Versprechen von Suchmaschinen: Gewaltige Datenmengen durch kluge Algorithmen und modernste Technik effizient durchsuchbar – und somit alles findbar zu machen. In den letzten Jahren schließt sich durch immer fortgeschrittenere Technologien zunehmend die Kluft zwischen redaktionell gepflegten Kategorisierungen und reiner Suche. Eine moderne Webseitensuche bietet heute mithilfe von Funktionen wie After-Search-Navigation, Entity Extraction und selbstlernenden Suchvorschlägen das Beste aus beiden Welten.

Folgende Typen von Website Search sind aus dem Alltag bekannt:

- 1. Intranet/Knowledge Search.** Hier geht es vor allem darum, den Angestellten eines Unternehmens effizienten Zugriff auf verschiedene Datensilos zu geben. Dateien von Netzlaufwerken, Informationen aus Wikis, Daten aus Ticketsystemen und Bugtrackern und vielen anderen Enterprise-IT-Systemen müssen hier indexiert werden, oft unter Berücksichtigung von Zugriffsrechten.
- 2. Website Search.** Hier werden die Inhalte einer öffentlichen Webseite für die Nutzer zugänglich gemacht.
- 3. E-Commerce Search.** Im Vergleich zur regulären Website Search spielt die korrekte Vertagung der Inhalte bei Online-Shops eine wichtigere Rolle, um Funktionen wie Sortierung und Filterung nach Eigenschaften zu ermöglichen.

Die eingebauten Suchlösungen populärer CMS und Online-Shops basieren oft auf einer einfachen Datenbanksuche, die in Echtzeit Ergebnisse für eine Suchanfrage zurückgibt. Dedizierte Suchlösungen, seien sie proprietär oder aus dem Open-Source-Bereich, bieten darüber hinaus jedoch ein so reichhaltiges Featureset, dass eine reine Datenbanksuche heutzutage wohl nur noch

## DIE AUTOREN



**Christoph Czupy** und **Arian Kuschki** sind Senior Web Business Consultants bei der TIE Kinetix GmbH in München. Sie lieben unkonventionelle technische Lösungen schwieriger Probleme zum geschäftlichen Wohl ihrer Kunden.

für sehr kleine Online-Shops und Webseiten sinnvoll erscheint.

Im Folgenden bekommen Sie einen Überblick, wie eine moderne Suchlösung funktioniert, und einige gute Tipps, um Ihre Webseitensuche noch etwas besser zu machen. Dieser Artikel ist bewusst produktunabhängig gehalten und betrachtet Best Practices und nützliche Features, die nicht an eine bestimmte Suchtechnologie gebunden sind.

## Wie kommt der Content in die Suchmaschine?

Egal ob Produkte für einen E-Commerce-Shop, Content in Webseiten oder PowerPoint-Dateien aus einem Firmenintranet: Bevor sie durchsucht werden können, müssen sie der Suchmaschine zur Verfügung gestellt werden. Hierfür gibt es zwei Möglichkeiten – Feeding und Crawling.

**Feeding.** Beim Feeding wird der zu indexierende Content in die Suchmaschine importiert. Je nach Technologie wird XML, JSON oder auch CSV verwendet, um Inhalte in strukturierter Form bereitzustellen. Oft passiert der Import über einen HTTP-Post-Request, da viele Suchmaschinen wie ein Webserver administrierbar sind.

**Crawling.** Hier wird ein [Webcrawler\\*](#) genutzt, der die durchsuchbaren Inhalte selbstständig und kontinuierlich findet und indexiert. Alles, was er dazu braucht, ist eine Start-URL. Es gibt HTTP-Crawler für reguläre Webseiten, aber auch für andere Datenquellen wie Netzlaufwerke und Datenbanken.

**Welche Methode sollten Sie wählen?** Crawling ist in der Regel mit weniger Aufwand verbunden. Es bietet sich besonders an, wenn ein strukturierter Datenfeed (inklusive regelmäßiger Aktualisierung) nur mit großem Aufwand erstellt werden kann. Ein Beispiel wäre der globale Webseitenverbund eines multinationalen Konzerns, der für eine Websiteseuche indiziert werden soll. Die verschiedenen Landesseiten laufen

eventuell auf unterschiedlichen Systemen und haben jeweils andere technische und redaktionelle Ansprechpartner. In diesem Fall stellt Crawling einen einfachen Weg dar, ohne großen technischen und bürokratischen Aufwand einen kompletten Index zu erstellen.

Bei einem Online-Shop hingegen liegen die Produktdaten höchstwahrscheinlich in strukturierter Form (z. B. in einem Warenwirtschaftssystem) vor und können leicht in ein Format umgewandelt werden, das die Suchmaschine akzeptiert. Hier bietet es sich an, den Feeding-Mechanismus zu verwenden, um volle Kontrolle über Metadaten und Update-Frequenz zu haben.

**TIPP:** Egal wie die Daten in die Suchmaschine kommen: Nutzen Sie auch die direkte Sicht auf den Index. Enterprise-Suchmaschinen bieten zumeist ein integriertes Suchfrontend und sind ein wertvolles Diagnosewerkzeug für Ihren Datenbestand. Ist die Crawlbarkeit Ihrer Seiten sichergestellt (Verwendung von Ajax)? Sind die URLs bzgl. „Duplicate Content“ gefährdet? Basieren automatisiert erstellte Filter auf homogen gepflegten Metadaten? Eine gute Suche baut auf strukturiert und konsequent gepflegten Inhalten auf.

## Content anreichern mit Metadaten

Volltextsuche ist gut und schön. Besser jedoch ist eine Suche, die durch relevante Metadaten angereichert wurde. Diese ermöglichen dem Nutzer, seine Suchanfrage präziser zu stellen oder sie nachträglich zu verfeinern („search refinement“; siehe nächster Abschnitt). Im Fall einer Online-Nachrichtenseite bieten sich hier z. B. Informationen zu Erscheinungsjahr, Autor oder auch Ressort eines Artikels an. Im Falle einer Produktsuche sind Farbe, Größe, Preis und Lieferzeit wichtige Datenpunkte,

die auf jeden Fall mit indexiert werden sollten. Weitere Metadaten können, je nach Business Case, dazukommen. Dies ermöglicht eine Suche wie zum Beispiel „Artikel, die den Investor Warren Buffett erwähnen, nur aus dem Ressort Wirtschaft, geschrieben zwischen 2010 und 2014“. Wenn Feeding genutzt wird, ist es in der Regel kein Problem, auch die gewünschten Metadaten mit zu übergeben, insofern die Daten gut gepflegt sind. Im Falle des Crawling muss die Suchmaschine selbst die Metadaten aus der HTML-Seite extrahieren. Standardmäßig werden HTML-Metatags als Metadaten indiziert.

**TIPP:** Analysieren Sie Ihren Business Case. Möchten Sie mithilfe der Suche Ihren Umsatz steigern, Ihre Ausgaben senken oder die Effizienz der Benutzer Ihrer Websuche erhöhen? Wer sind die Benutzer Ihrer Suche und was haben sie für Bedürfnisse? Sind es Käufer in Ihrem Online-Shop, die sehr gern in Ihrer Produktpalette stöbern oder ganz genau wissen, welche Produkte sie suchen? Oder ist es ein Verkäufer oder Ihr Einkäufer, der sich schnell einen Überblick über das Produktsortiment beschaffen möchte? Oder wollen Sie für Berater in einem Callcenter oder Ihre eigenen Mitarbeiter, die auf schnelle und relevante Informationen angewiesen sind, bei regelmäßigen Suchen wertvolle Zeit einsparen? Die richtige Auswahl von Metadaten gibt Ihrer Suche für jeden Business Case die passende Struktur.

## After-Search-Navigation zum Entdecken relevanter Inhalte

Sind alle Inhalte gut vertaggt, kann mit wenig Konfigurationsaufwand eine After-Search-Navigation aktiviert werden. Die Suchmaschine weiß nun für jedes Dokument in ihrem Index, welche Metadaten ihm zugeordnet sind, und

\*siehe Online-Glossar unter [www.websiteboosting.com](http://www.websiteboosting.com)

kann aus diesen Daten Kategorien bilden. Führt der Nutzer nun eine Suche aus, wird ihm angezeigt, welche Kategorien für die aktuellen Suchergebnisse vorhanden sind. Dies ermöglicht ihm, schnell die Ergebnismenge einzugrenzen und zu dem gewünschten Content zu gelangen. Hat der Nutzer z. B. nach „Winterschuhen“ gesucht, werden ihm Facetten wie „Farbe“, „Größe“ oder „Material“ angeboten, mit denen er seine Suche weiter eingrenzen kann. Neben jeder Facette wird auch angezeigt, für wie viele Ergebnisse sie zutrifft. Das Tolle an der After-Search-Navigation ist zudem: Wenn die Metadaten gut gepflegt sind, ist sie immer aktuell, ohne dass sie mühsam manuell angepasst werden muss. Und der Nutzer kann durch Kombination von verschiedenen Facetten beliebige Schnittmengen erstellen und den Datenbestand erkunden.

**TIPP:** Filterung auf Metadaten kann auch verwendet werden, um automatisch Übersichtsseiten zu erstellen. Wenn der Nutzer zum Beispiel bei einem Nachrichtenportal im Menü auf das Ressort „Politik“ klickt, kommt er zu einer Kategorieseite, der die Artikel nicht manuell zugewiesen wurden, sondern die von der Suchmaschine immer aktuell mit passenden Artikeln befüllt wird. So wie auf der Suchergebnisseite bei einem Klick des Nutzers auf die Facette „Ressort: Politik“ alle Artikel zu diesem Thema angezeigt werden, so kann der gleiche Mechanismus auch genutzt werden, ohne dass der Nutzer überhaupt bewusst eine Suche ausführt. Hier kann der Webseitenbetreiber viel Zeit bei der manuellen Zuweisung von Artikeln sparen und hat trotzdem immer relevante, aktuelle Übersichtsseiten.

## Entity Recognition

Der Mehrwert gut gepflegter Metadaten ist enorm. Doch was tun, wenn

der Datenbestand nicht einheitlich oder gar nicht getaggt ist? Die sogenannte Entity Recognition („Auto Tagging“) ist ein mächtiges Werkzeug, um auch in unstrukturierten Daten Muster zu entdecken und zu extrahieren. Über Wörterbücher und Regular Expressions (Regex) können Regeln definiert werden, nach denen Dokumente automatisch von der Suchmaschine getaggt werden sollen, wenn sie gewisse Kriterien erfüllen. So können Orte, Personen und Konzepte im Dokument erkannt werden. Oder es kann vorkommen, dass Artikel das Veröffentlichungsdatum nicht in einem Metatag enthalten, sondern irgendwo im Text. Über die Definition einer einfachen Regex (z. B. der Form TT.MM.JJJJ) können diese Daten leicht aus dem Text extrahiert werden. Ein anderes Beispiel sind Webseiten, bei denen die Kategorien nur aus der Seiten-URL abzulesen sind. Zum Beispiel liegen alle Artikel in einer Ordnerstruktur ‚.../kategorie/jahr/artikel.html‘. Auch hier kann Entity Extraction genutzt werden, um diese Informationen in nutzbare Metadaten umzuwandeln. Auf diese Art extrahierte Metadaten können wie reguläre Metadaten verwendet werden, um daraus Facetten, Filter oder Sortierungen der Suchergebnisse abzuleiten. Ein großer Vorteil ist hierbei, dass Entities auch von nicht technischen Nutzern angelegt werden können, denn manchmal ist es nur mit Entwicklungsaufwand und Verzögerungen möglich, Webseiten mit neuen Metatags zu versehen. Entities dagegen können in der Suchmaschine bequem z. B. von Marketingzuständigen angelegt und gepflegt werden, ohne das IT-Budget zu belasten.

**TIPP:** Lassen Sie Ihre Suchmaschine als Webseitenoptimierungs- und SEO-Tool arbeiten. Extrahieren Sie aus Ihrem Webanalyse-Tool die Suchanfragen, die ihre Nutzer in Ihre Site-search eingeben. Dadurch können Sie besser verstehen, was genau die Nutzer auf Ihrer Seite suchen, ob sie die

passenden Suchergebnisse zurückbekommen und wie zufrieden sie mit diesen Ergebnissen sind. Wichtige KPI sind hier, je nach Business Case, z. B. „visits with search“, „number of repeated searches“, „pages browsed after search“. Aus den gewonnenen Erkenntnissen lassen sich dann wiederum SEO-Maßnahmen ableiten. Gibt es zum Beispiel eine populäre Suchanfrage ohne Ergebnis, dann ist die Wahrscheinlichkeit groß, dass der Nutzer auch z. B. bei google.de für diese Anfrage kein gutes Ergebnis von Ihrer Webseite bekommen wird. Vielleicht müssen Sie hier passenden Content erstellen, vielleicht liegt aber auch ein technisches Problem vor, weswegen der Crawler Ihre Inhalte nicht indexieren kann.

## Sortierung der Suchergebnisse

Metadaten können ebenfalls genutzt werden, um die Sortierung von Ergebnissen zu ermöglichen. Im Normalfall gibt eine Suchmaschine Ergebnisse nach Relevanz geordnet zurück. Aber je nach Situation kann es sinnvoll sein, eine andere Sortierung zu wählen. Zum Beispiel gibt es in den meisten Online-Shops die Möglichkeit, die Suchergebnisse nach Preis zu sortieren. Auch Popularität und Bewertungen werden häufig angeboten. Bei einer contentlastigen Webseite bietet sich die Sortierung nach Erscheinungsdatum an, um die aktuellen Artikel zuerst anzuzeigen. Aber hier wird auch ein Problem mit dem strengen Sortieren nach einem Kriterium wie dem Datum deutlich, denn dabei werden jegliche Relevanzkriterien außer Acht gelassen. Dies ist aber oft nicht das gewünschte Resultat. Auch wenn der Suchbegriff auf einer Seite nur einmal und in einer unwichtigen Position vorkommt (also nicht im Seitentitel, der Überschrift etc.), so wird diese Seite doch auf Nummer eins der Ergebnisliste



sortiert werden, wenn sie dem Datum nach die neueste ist. Wäre es nicht toll, wenn es einen Weg gäbe, bei der Sortierung der Ergebnisse sowohl das Datum (die Frische) als auch die Relevanz mit einzubeziehen? Das geht tatsächlich, mit sogenannten Biasingregeln (Gewichtung).

**TIPP:** Erstellen Sie Ihre eigenen Metatags, nach denen Sie gewinnbringend filtern und sortieren können. So wären etwa Attribute mit Werten für „Marge“, „Preis-Leistung“, „Lagerbestand“ oder „redaktionelle Gewichtung“ sehr gut geeignet, bestimmte Business Cases wie Gewinnoptimierung, Ausverkauf von Produkten oder Filterung und Sortierung für Eilmeldungen automatisiert abzubilden. So lassen sich beispielsweise auch Upselling- oder Crossselling-Blöcke auf Produkt- oder Übersichtsseiten darstellen.

## Beeinflussung der Gewichtung von Ergebnissen

Biasing funktioniert ähnlich wie Sortierung. Sie wählen einen oder mehrere Metatags und geben ihnen je nach ihrem Inhalt eine positive oder negative Gewichtung. Im Unterschied zur absoluten Sortierung, die Relevanzkriterien ignoriert, ist hiermit eine subtile Beeinflussung des Relevanzalgorithmus der Suchmaschine möglich. So kann für Suchergebnisse eines Nachrichtenportals konfiguriert werden, dass Ergebnisse nach Relevanz sortiert werden sollen, aber mit einem Push für neuere Artikel. Einen zweiten leichten Push könnten Sie auf Basis positiver Nutzerbewertungen einstellen. In Online-Shops würde sich anbieten, Bestseller oder Auslaufmodelle/Restposten mit einem positiven Push zu versehen, je nach Geschäftsstrategie.

Spannend wird es, wenn das Biasing zur Laufzeit dynamisch angepasst wird. So kann zum Beispiel auf der Basis des

Nutzerverhaltens in einem Online-Shop die Ergebnisgewichtung angepasst werden, je nachdem, welche Aktionen der Nutzer ausführt. Sieht er sich ausschließlich die günstigeren Produkte einer Kategorie an oder hat er sogar schon einmal nach Preis sortiert oder auf eine Facette mit einer niedrigen Preisrange geklickt? In diesem Fall kann es sich lohnen, eine vordefinierte Biasingregel an alle weiteren Suchrequests dieses Nutzers anzuhängen, die günstigere Produkte in den Suchergebnissen nach oben pusht. Der Erfolg eines solchen Biasings lässt sich sehr gut anhand des Warenkorbwertes in einem A/B-Test bestimmen.

**TIPP:** Definieren Sie gern etwas mehr Gewichtungsregeln und weisen Sie diese Ihren Benutzergruppen zu. Über Cookies lassen sich Ihre wiederkehrenden User identifizieren und mit dem richtigen Set an Filtern und Gewichtungen die Suchergebnisse optimieren. Indem Sie Ihre Nutzer kennenlernen und ggf. saisonale Bedingungen einbeziehen, verschaffen Sie ihnen ein einzigartiges Sucherlebnis. So bieten Sie einem männlichen Käufer im Herbst für die einfache Suche nach „braunen Schuhen“ die richtigen Angebote für braune Winter- oder Herbstschuhe und degradieren unpassende Schuhmodelle, insbesondere jene für den Sommer, bunte oder gar für Damen, entsprechend herunter.

## Darstellung von Suchergebnissen

Nicht alle Suchergebnisse sind gleich. In einer Intranetsuche gibt es Dokumentenergebnisse, Personenergebnisse, Medienfiles und vieles mehr. In einem Online-Shop kann es Produkte geben, die in vielen Farben und Größen (Varianten) vorhanden sind. Jeden dieser Ergebnistypen können Sie mit einem

individuellen Layout versehen, um dem Nutzer möglichst weit entgegenzukommen. Dadurch werden die wichtigsten Informationen bereits in der Suchergebnisliste angeboten und ein weiterer Klick in das Dokument hinein ist eventuell nicht mehr nötig. So ermöglicht bei Officedokumenten ein Quick-Preview auf die ersten (oder alle) Seiten des Dokuments schnelle Recherchen und erspart vielleicht das Herunterladen des Dokuments (Dokument-Previews werden zum Indizierungszeitpunkt der Dokumente berechnet und sind entsprechend performant). Bei Personenergebnissen werden neben einem Foto der Person auch Jobbezeichnung, Telefonnummer/E-Mail-Adresse und weitere wichtige Informationen angezeigt und ermöglichen eine schnelle Kontaktaufnahme. Bei einem Variantenprodukt in einem Shop schließlich kann man schon in der Suchergebnisliste die verfügbaren Farben und Größen anzeigen und eine direkte Bestellung ermöglichen.

**TIPP:** Sie kennen Ihre Benutzer, Ihre Vorlieben bei Suchen und immer wiederkehrende Tätigkeiten, die mit der Suche verbunden sind. Bieten Sie Ihren Nutzern entsprechende Servicelinks an. Ein User, der nach Adressen sucht, wünscht sich eine Darstellung seiner Suchergebnisse in Maps. Bieten Sie ihm einen Link zur Navigation vom aktuellen Standort an. Ein Redakteur oder Produktmanager, der regelmäßig Suchergebnisse kontrolliert, freut sich über einen Deeplink in sein CMS oder WWS, um notwendige Anpassungen ohne Umwege vornehmen zu können. Auch Administratoren freuen sich bei Suchergebnissen über Deeplinks in die Indexdiagnose der Suchmaschine oder zum Feeder-gate. Diese Servicelinks verringern die Anzahl der zu leistenden Klicks, erhöhen die Effizienz jedes einzelnen Mitarbeiters und tragen maßgeblich zur Akzeptanz einer Suchlösung bei.



## Query Expansion

Query Expansion ist ein wichtiger Bestandteil jeder relevanzbasierten Suchmaschine. Im Wesentlichen handelt es sich um Wörterbücher, die die Äquivalenz von Worten oder Wortbestandteilen festlegen. Auf diese Weise ist eine Suchmaschine zum Beispiel in der Lage zu verstehen, dass „sprechen“, „Sprache“ und „Gespräch“ verwandte Begriffe sind und dass eine Suche nach „englische Sprache“ wohl auch Ergebnisse zu „Englisch sprechen“ zurückliefern sollte. Oder dass „Laptop“ und „Notebook“ äquivalent sind, aber auch „Kühlschrank“ und „Kuehlschrank“ (bidirektional). Auch für Akronyme wie „DNS/Desoxyribonukleinsäure“ ist diese Funktion sehr hilfreich. Neben den eingebauten Query Expansions können auch eigene Wörterbücher erstellt werden, um zum Beispiel domänenspezifische Mappings zu generieren, die von den eingebauten Wörterbüchern nicht abgedeckt werden. Synonyme werden beispielsweise in Online-Shops gern eingesetzt, um von unüblichen Farbworten auf ihre Grundfarbe zu verweisen. So kann ein Mapping „weiß » elfenbeinweiß“ erstellt werden. Das heißt, alle Suchen, die den Begriff „weiß“ enthalten, sollen auch nach „elfenbeinweiß“ suchen, während dies umgekehrt in diesem Fall nicht geschehen soll (unidirektional). Denn wenn der Kunde schon spezifisch nach „elfenbeinweiß“ sucht, will er vermutlich nicht die unspezifischeren Ergebnisse für „weiß“ erhalten.

**TIPP:** Profitieren Sie von Community-Wörterbüchern: kontextbasiert oder hinsichtlich Produktgruppen und in unterschiedlichen Sprachen. Übermitteln Sie mit der Suchanfrage des Nutzers immer auch die richtige Sprachinformation an die Suchmaschine, denn Query Expansions sind natürlich sprachabhängig. Bedenken Sie, dass Ihr Suchproxy installationsbedingt eine feste Sprache verwenden

det (in der Regel Englisch). Werten Sie also serverseitig die Sprache Ihres Benutzers aus und verwenden Sie diese bei den Suchanfragen an die Suchmaschine. So stellen Sie bei mehrsprachigen Portalen oder Shops, die über dieselbe Suchapplikation bedient werden, auch sicher, dass jeweils das korrekte Wörterbuch verwendet wird.

## Verwandte Suchvorschläge (Related Queries)

Manchmal macht es aber auch Sinn, dem Nutzer die Möglichkeit zu geben, selbst die Synonyme auszuwählen, die ihn interessieren. Related Queries sind ebenfalls Synonyme, allerdings werden diese nicht intransparent „hinter den Kulissen“ der Anfrage des Nutzers hinzugefügt, sondern als Vorschläge an den Nutzer zurückgegeben. Sie werden in der Regel manuell konfiguriert. So könnte auf die Suchanfrage nach „DNS“ von der Suchmaschine der Vorschlag zurückkommen: „Meinten Sie ‚Desoxyribonukleinsäure‘ oder ‚Domain Name System‘?“ Hierbei sind die beiden Begriffe jeweils auch auf die passende Suchergebnisseite verlinkt. Auch für manuell erstellte Rechtschreibvorschläge können Related Queries natürlich genutzt werden.

**TIPP:** Vertrauen Sie auf die hervorragende Relevanz der Suchmaschine, die schnellen Antwortzeiten und Ihre gute Contentpflege. Fügen Sie zur Frage „Meinten Sie etwa ...“ auch das Top-Suchergebnis der jeweils verwandten Suchergebnisse als Teaser an. Doppelungen bei den Webserviceanfragen und deren Suchergebnissen können Sie serverseitig einfach herausfiltern. Bieten Sie Ihren Usern den Service – womöglich ist er nur einen Klick vom gewünschten Dokument entfernt.

## Suchvorschläge

Schnelle und relevante Suchvorschläge sind für eine Webseitensuche heutzutage essenziell. Eine gute Implementierung liefert nach jedem Tastendruck Vorschläge zurück, die zu den bisher getippten Buchstaben passen. Tippt der Nutzer z. B. „Geschirr“ ein, werden nach jedem Tastendruck Vorschläge angezeigt: erst Begriffe, die mit „G“ beginnen, dann mit „Ge“, dann „Ges“ und so weiter. Folgende Merkmale muss eine gute Vorschlagsfunktion mitbringen:

- » Eine Latenz von maximal 300 ms zwischen Tastendruck und Anzeige des Vorschlags. Google und Amazon schaffen es in ca. 100 ms und das prägt natürlich die Erwartungen der Nutzer.
- » Kein „Flackern“ des Vorschlags-Popups bei schnellen Tippert. Hierzu bedarf es einer Warteperiode nach jedem Tastendruck, bevor der Vorschlagsrequest an die Suchmaschine geschickt wird.
- » Vorschläge nach Popularität geordnet.
- » Keine Vorschläge anzeigen, für die es keine Ergebnisse gibt.
- » Vorschläge jeweils zugeschnitten auf die Sprache des Nutzers und die Teilmenge des Suchindex, die abgefragt wird.
- » Ein Klick auf einen Vorschlag löst sofort die Suche aus.
- » Die Suchmaschine sollte Vorschläge entsprechend den oben genannten Kriterien selbstständig lernen, damit immer aktuelle und relevante Vorschläge für jede Suchanfrage vorhanden sind.

Die oben beschriebene (statistische) Art von Suchvorschlägen basiert auf dem Suchverhalten der Nutzer. Ein anderer, (algorithmischer) Ansatz ist es, durch Methoden des maschinellen Lernens aus dem Corpus der indizierten Dokumente Suchvorschläge zu gewinnen. Bei eher unspezifischen Suchbegriffen wie z. B. „Mitgliedschaft“ können Sie die häufigs-

ten Wortpaarungen ermitteln, die dieses Wort enthalten und im Suchindex vorkommen, etwa „Mitgliedschaft Arten“, „Mitgliedschaft Hilfe“, „Mitgliedschaft Preise“, „Mitgliedschaft Vorteile“, „Mitgliedschaft Specials“. Diese werden dem Benutzer dann als Suchvorschläge angezeigt (ähnlich den verwandten Suchanfragen).

**TIPP:** Erstellen Sie pro Contenttyp eigene Suggest-Indizes. Nutzen Sie die Stärke der Suchmaschine bei Long-Tail-Anfragen, die zunehmend an Bedeutung gewinnen. Importieren Sie für eine FAQ-Suche einfach die Fragen, für Rezepte die Titel, für eine Bilddatenbank Titel, Urheber und Datum. Lassen Sie Ihre Nutzer selbst entscheiden, welchen Empfehlungen sie folgen, oder gar neue Suchtrends setzen.

### Livesearch (Search-as-you-type)

In letzter Zeit gibt es zunehmend Websitesuchen, die nicht nur nach jedem Tastendruck des Nutzers Suchvorschläge anzeigen, sondern zusätzlich auch Suchergebnisse, teilweise mit Bildern. In manchen Fällen werden auch Kategorien zurückgegeben, im Fall eines Online-Shops zum Beispiel Marken, nach denen gesucht werden kann. Im Prinzip sind hier der Fantasie keine Grenzen gesetzt, auch Features wie verwandte Suchvorschläge oder Rechtschreibvorschläge können natürlich in Echtzeit zurückgegeben werden. Das Ziel ist, dem Nutzer schneller und mit weniger Klicks das zu geben, was er will. Allerdings gilt auch hier, dass dieses Feature nur bei sehr guter Performance (niedriger Latenz) Sinn macht. Dies hängt zum einen von der Leistungsfähigkeit der verwendeten Suchlösung ab, aber auch von der Internetverbindung zwischen Nutzer und Website. Gerade auf mobilen Geräten sind die Latenzen für einen Roundtrip

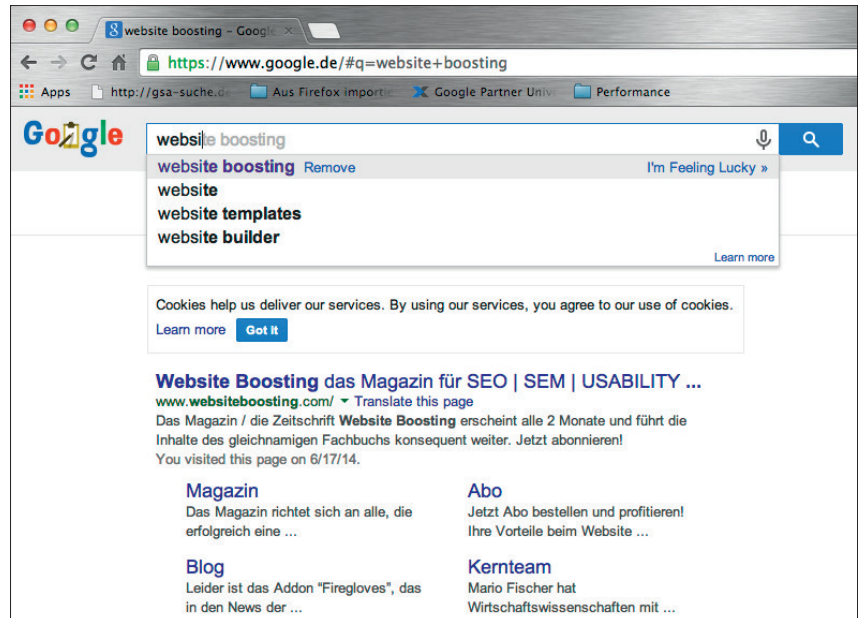


Abb. 2: Livesearch auf google.de

zwischen Browser und Webserver technisch bedingt immer noch höher als bei stationären Internetanschlüssen. Dies muss bedacht werden, wenn eine Livesearch geplant wird.

Technisch ist die Implementierung der Livesearch etwas anspruchsvoller. Denn es wird ja nicht nach ganzen Suchbegriffen gesucht, sondern nach Wortbruchteilen. Das heißt, dass hier in der Regel mit einer Substringsuche gearbeitet wird, damit schon die Eingabe von „Geschirrs“ Ergebnisse für den „Geschirrspüler“ zurückgeben kann. Hier ergibt sich jedoch das Problem der Relevanz der Liveergebnisse. Denn es ist davon auszugehen, dass keine Suchlösung für eine Substringsuche wie „Gesch“ oder gar „Ges“ besonders relevante Ergebnisse zurückgeben wird, da der Ergebnisraum einfach zu groß ist. Dies kann dazu führen, dass die Top-3-Liveergebnisse für den Nutzer überhaupt nicht relevant sind. Eine einfache Lösung für dieses Problem findet man in der Implementierung der Livesearch auf google.de, die auf clevere Art Suggest- und Suchfunktion miteinander kombiniert. Näheres hierzu in dem letzten „Tipp“.

**TIPP:** Eine einfache und sehr effektive Möglichkeit, eine Livesearch zu bauen, ist die Kombination von Suggest- und Suchfunktion. So lässt sich

das Problem der schlechten Relevanz einer Substringsuche elegant lösen, und das bei herausragender Performance. Folgendermaßen funktioniert es: Wie oben beschrieben, wird nach jedem Tastendruck des Nutzers ein Request gegen die Suchapplikation abgeschickt, um passende Suchvorschläge abzurufen. Bevor die Suchapplikation nun aber die Suchvorschläge an den Browser zurückgibt, nimmt sie den ersten Suchvorschlag (dies ist in der Regel der populärste Suchbegriff) und führt für diesen Suchbegriff eine reguläre Suchanfrage aus. Beispiel: Der Nutzer tippt „websi“ und die populärsten Suchvorschläge hierfür sind „website boosting“, „website“ und „website templates“. Es wird automatisch ein Suchrequest für „website boosting“ ausgelöst. Die Suchvorschläge werden an den Browser zurückgeschickt, während der Suchrequest noch verarbeitet wird. Sobald die Suchergebnisse verfügbar sind (typischerweise die Top 3), werden auch diese über Ajax nachgeladen und im Browser dargestellt. Diese Methode ermöglicht es, Livesearchergebnisse mit sehr niedriger Latenz darzustellen, die auf einer Mischung von Relevanz und Popularität basieren.