

Websuche und Webmonitoring

Nadine HÖCHSTÖTTER^a und Dirk LEWANDOWSKI^b

^a *Webscout GmbH, Röntgenstr. 8, 76133 Karlsruhe, Deutschland, nsh@webscout.de*

^b *Hochschule für Angewandte Wissenschaften Hamburg, Department Information, Finkenau 35, 22081 Hamburg, Deutschland, dirk.lewandowski@haw-hamburg.de*

Zusammenfassung. Die Herangehensweise an eine Websuche unterscheidet sich grundsätzlich von der ganzheitlichen Recherche von Themen im Internet, dem Webmonitoring bzw. Social Media Monitoring im Speziellen.

Bei der Suche in Websuchmaschinen geht es in den meisten Fällen darum, zu einer Suchanfrage eine oder einige passende Antworten (Webseiten) zu finden (precision-orientierte Recherche). Der Anwendungsfall Webmonitoring ist dagegen darauf ausgerichtet, zu einem Thema möglichst *alle* Veröffentlichungen im Web aufzuspüren und darzustellen (recall-orientierte Recherche). In diesem Kapitel werden die wichtigsten Qualitätsmaße vorgestellt, mit denen Suchmaschinen beurteilt werden können. Im nächsten Schritt werden die unterschiedlichen Anforderungen an Suchmaschinen gezeigt und welche Qualitätsmaße für die grundsätzlich verschiedenen Anforderungen der Ad-hoc-Recherche bzw. des Monitorings wichtig sind.

Generell zeigt sich, dass sich Suchmaschinen für die Websuche fundamental von Suchmaschinen für das Webmonitoring unterscheiden und deshalb aus Sicht des Suchenden als zwei unterschiedliche Alternativen bei der Webrecherche gesehen werden müssen. Abhängig von der Rechercheaufgabe muss der entsprechende Suchservice gewählt werden.

Schlüsselworte. Websuche, Qualität von Suchmaschinen, Webmonitoring, Social Media Monitoring

Einleitung

Seit die ersten Suchmaschinen Mitte der 1990er Jahre online gingen, hat sich viel auf dem Suchmaschinenmarkt und im Web getan. Der Suchmaschinenmarkt hat sich extrem konsolidiert: Noch vor wenigen Jahren gab es immerhin vier aktive Suchmaschinen, die eine eigene, wesentliche Teile des World Wide Web abdeckende Datenbasis pflegten, sprich ihre Suche auf einem eigenen Index bzw. Crawler aufbauten. Mittlerweile gibt es nur noch die zwei großen Player, Bing und Google. Die anderen Suchmaschinen existieren zum Teil noch „äußerlich“, d. h. als Suchinterfaces, die auf den Datenbestand einer der großen Suchmaschinen zurückgreifen, doch das Herzstück der Suche, der Index, basiert bei diesen Suchmaschinen auf Google oder Bing [1]. So greifen beispielsweise die Suchfunktionen auf den Portalen T-Online und Web.de auf den Index von Google zurück; die bei Yahoo angezeigten Ergebnisse kommen von Bing.

Doch selbst zwischen den beiden in diesem Modell “gebenden” Suchmaschinen Google und Bing besteht bei weitem kein Ausgleich in Bezug auf die Marktverhältnisse: Google beantwortet in Deutschland etwa 96 Prozent aller an die allgemeinen Suchmaschinen gestellten Anfragen [2].

Die Ergebnispräsentation der Suchmaschinen hat sich im Verlauf der Zeit ebenfalls geändert; es werden Werbung, Bilder und weiterführende Links zu tiefer gelegenen Seiten einer Website auf der ersten Ergebnisseite gezeigt (sog. Universal Search, siehe [3] und [4]). Suchmaschinen präsentieren dabei auch nicht mehr nur die Ergebnisse aus einem Index, der Seiten aus dem Web ganz allgemein gesehen enthält, sondern reichern diese klassischen sogenannten "10 blue links" mit Ergebnissen aus dem Bereich Nachrichten, Bilder, Videos und Social Media an, wofür jeweils eigene Datenbestände (Indexe) aufgebaut werden. Der Suchende bekommt so aus verschiedenen Bereichen Informationen präsentiert, um möglichst optimal die Bedürfnisse der Suche zu befriedigen.

Weitere Neuerungen sind Suchvorschläge während der Eingabe, die darauf abzielen, die Intention der Suche schon bei der Eingabe einer Suchanfrage vorwegzunehmen. Diese Vorschläge sind in den letzten Jahren immer wichtiger geworden und haben sich mittlerweile zu einem Standard in Suchanwendungen entwickelt [5, 6].

The screenshot shows the Google search results for the query "michael jackson". At the top, the Google logo and search bar are visible. Below the search bar, there are tabs for "Web", "Bilder", "News", "Videos", "Shopping", "Mehr", and "Suchoptionen". The search results are displayed in a grid-like format. On the left side, there are several text-based search results, including links to the official Michael Jackson website, Wikipedia, and various news articles. On the right side, there is a large image of Michael Jackson, followed by a grid of smaller images. Below the images, there is a section titled "Michael Jackson" with a brief biography, a list of songs, and a section for "Neueste Beiträge" (Latest Contributions) featuring a video thumbnail and a description.

Google michael jackson

Web Bilder News Videos Shopping Mehr Suchoptionen

Ungefähr 520.000.000 Ergebnisse (0,29 Sekunden)

Michael Jackson | The Official Michael Jackson Site
www.michaeljackson.com/ - Diese Seite übersetzen
Official Michael Jackson website including Michael Jackson videos, music, news, albums, pictures and more. Your number one source for all things Michael ...

Michael Jackson - Wikipedia
de.wikipedia.org/wiki/Michael_Jackson - Diese Seite übersetzen
Michael Joseph Jackson (* 29. August 1958 in Gary, Indiana; † 25. Juni 2009 in Los Angeles, Kalifornien) war ein US-amerikanischer Sänger, Komponist, ...
Villagio - Michael Jackson/Diskografie - Mark Lester - Thriller

Michael Jackson - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Michael_Jackson - Diese Seite übersetzen
Michael Joseph Jackson (August 29, 1958 – June 25, 2009) was an American singer-songwriter, dancer, businessman, and philanthropist. Often referred to by ...

Michael Jackson - Alle News & Infos zum Thema - Bild.de
www.bild.de - Alle Themenseiten
Musikalisch war er ein Genie. Doch Michael Jacksons Persönlichkeit, die seit früherster Kindheit starkem Erfolgsdruck ausgesetzt war, galt als äußerst labil. 1963 ...

Michael Jackson | Facebook
https://de-de.facebook.com/michaeljackson - Diese Seite übersetzen
Michael Jackson. 69760862 likes · 1039616 talking about this. Michael Jackson, one of the most widely beloved entertainers and profoundly influential artists of ...

Michael Jackson - YouTube
www.youtube.com/user/michaeljackson - Diese Seite übersetzen
The Official YouTube Channel of The King of Pop - Michael Jackson. For more info, visit www.michaeljackson.com.

Michael Jackson - Topic - YouTube
www.youtube.com/...UCpIOOL7QKuBhHVKLby... - Diese Seite übersetzen
Michael Jackson was unquestionably the biggest pop star of the '80s, and certainly one of the most popular recording artists of all time. In his prime, Jacks...

Michael Jackson King of Pop Forum News Community Fans ...
www.mjackson.de/ -
Website mit Infos zum Tod von Michael Jackson und großem Forum.

Michael Jackson
1.016.294 Follower auf Google+
Michael Joseph Jackson war ein US-amerikanischer Sänger, Komponist, Tänzer und Entertainer. Wegen seiner Erfolge wird er als King of Pop bezeichnet. Wikipedia
Geboren: 29. August 1958, Gary, Indiana, Vereinigte Staaten
Gestorben: 25. Juni 2009
Kinder: Paris Michael Katherine Jackson, Prince Michael Joseph Jackson, II, Prince Michael Jackson, Jr.
Filme: Moonwalker, Ghosts, The Wiz – Das zauberhafte Land, mehr

Songs

Billie Jean	1982	Thriller
Thriller	1982	Thriller
Beat It	1982	Thriller
Smooth Criminal	1987	Bad
Heal the World	1991	Dangerous

Neueste Beiträge

Michael around the world: This tribute to Michael can be found on a door in Honolulu's Chinatown District.
28. Jan. 2014

Abbildung 1. Ergebnisse zu der Suchanfrage „Michael Jackson“.

Auf der anderen Seite ist das Web aber auch extrem schnelllebig geworden und wächst rasant an. Durch Social Media Dienste wie Facebook und Twitter werden im Web laufend neue Inhalte erzeugt. Gerade wenn es um Themen geht, die stark in den Massenmedien behandelt werden, tragen die Nutzer viel bei, indem eine Vielzahl von Meinungsäußerungen und Kommentaren in den Social Media Kanälen veröffentlicht wird.

Für Suchmaschinen stellt die Integration von Social Media Inhalten eine große Herausforderung dar. Bing arbeitet dazu sowohl mit Facebook als auch mit Twitter

zusammen. Auch Google hat eine Zeitlang Inhalte von Twitter in Echtzeit in die Ergebnisseite eingebunden, ähnlich wie es beispielsweise mit Nachrichten oder Videos gemacht wird. Diese Bestrebungen wurden aber nicht fortgeführt. Ein Grund dafür mag sein, dass Google mit Google+ ein eigenes soziales Netzwerk aufgebaut hat, das zu den genannten Diensten in Konkurrenz steht.

Durch die genannten Entwicklungen, speziell im Bereich der sozialen Medien, existiert mittlerweile eine Kluft in der Suche: Auf der einen Seite befinden sich die klassischen Websucher, die eine Suchanfrage formulieren, um schnell eine spezielle Antwort zu finden. Abbildung 1 zeigt am Beispiel der Suchanfrage „Michael Jackson“, welche Elemente von der Suchmaschine Google als relevant ermittelt werden. Bei solchen Suchanfragen spielen die Artikel bei Wikipedia für die Beantwortung eine wichtige Rolle, da sie meist eine gute Zusammenfassung über Personen, Firmen, Orte oder geschichtliche Ereignisse geben.

The image shows a Google search interface for the query "dschungelcamp". The search bar is at the top left, and the results are displayed below. On the right side, there is a vertical sidebar with a search icon and a magnifying glass. The main content area is divided into several sections:

- Search Results:** A list of search results including "dschungelcamp", "dschungelcamp 2014 kandidaten", "dschungelcamp 2014", and "dschungelcamp 2013".
- News zu dschungelcamp:** A section titled "Nach dem Dschungelcamp Dschungelkönigin Melanie: Erst Stern TV, dann die Eltern" with a sub-headline "Abendzeitung München - vor 1 Stunde". It mentions that Dschungelkönigin Melanie left her mother and father waiting, and that she was first interviewed by press terms, then flew home to her parents. The photo credit is "Foto: Jan ...".
- RTL-Dschungelcamp: Das sind die Gewinner und Verlierer:** A sub-headline "Augsburger Allgemeine - vor 1 Tag".
- Dschungelcamp 2014: Teilnehmer zurück in Deutschland:** A sub-headline "T-Online - vor 22 Stunden".
- Dschungelcamp 2014 Ich bin ein Star - holt mich hier raus!** A sub-headline "www.dschungelcamp.net/". It mentions that in Germany two weeks long the Dschungel-Fieber reigned, and now it has been brought back. The Dschungelcamp has a new queen: Melanie die 1.
- Das Dschungelcamp 2014 | Ich bin ein Star - holt mich hier raus - RTL:** A sub-headline "www.rtl.de > RTL.de > Sendungen".
- Dschungelcamp 2014: Die "Ich bin ein Star - Holt mich hier raus"-Moderatoren Sonja Zietow und Daniel Hartwich begrüßen Promis im Dschungel.**
- Ich bin ein Star - Holt mich hier raus! - Wikipedia:** A sub-headline "de.wikipedia.org/wikilich_bin_ein_star_-_holt_mich_hier_raus!". It mentions that "Ich bin ein Star - Holt mich hier raus!" (often as Dschungelcamp bezeichnet) is a reality show from the United Kingdom (I'm a ...).
- Featured Snippet:** A section titled "Ich bin ein Star – Holt mich hier raus!" with a sub-headline "Fernsehserie". It mentions that "Ich bin ein Star – Holt mich hier raus!" is a reality show from the United Kingdom (I'm a ...). It also includes the premiere date "Erstaussstrahlung: 10. Januar 2004", the last episode "Letzte Folge: 6. November 2004", the sender "Sender: RTL Television", the language "Sprache: Deutsche Sprache", and the genres "Genres: Reality-TV, Action Genre, Abenteuerfilm".

Abbildung 2. Ergebnisse zu der Suchanfrage *Dschungelcamp*.

Auf der anderen Seite stehen die Suchenden, die generell wissen wollen, was überhaupt im Web zu einem Thema veröffentlicht wurde. Ein einfaches Beispiel hierzu liefert die Suche nach einem beliebigen Firmennamen oder einem aktuellen Ereignis in den Medien. Als Beispiel zeigt Abbildung 2, was Google zur Suchanfrage *Dschungelcamp* zeigt. Der Wikipedia-Artikel als eine Art "Must-have" (vgl. [7]) springt schnell ins Auge. Zudem sieht man an den Vorschlägen, die Google unterbreitet, was sonst in letzter Zeit zu diesem Thema gesucht wurde (vgl. [5]). Dass es sich hierbei um ein aktuelles Thema handelt, ist auch daran zu erkennen, dass Nachrichtentreffer eingeblendet werden.

Durch die Einblendung von Nachrichten zu dem Thema gibt es auch brandaktuelle Fundstellen, doch ist leicht erkennbar, dass die Präsentation von ca. zehn Ergebnissen nicht all das widerspiegelt, was tatsächlich im Web und in den Social Media Kanälen dazu geschrieben wurde.

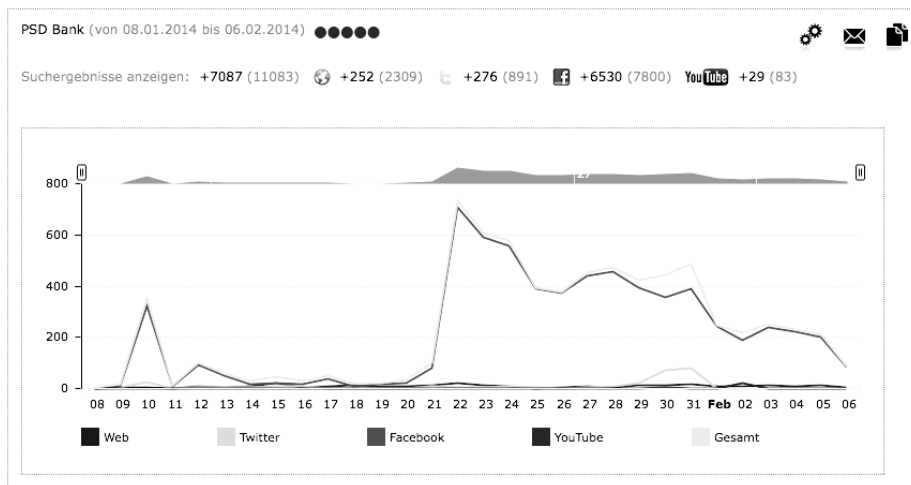


Abbildung 3. Volumina an Posts in Social Media zum Thema *PSD Bank*.

Abbildung 3 zeigt dagegen, wie viele Posts zu einem Thema auf Facebook und Twitter tatsächlich veröffentlicht werden können. Hier wird die Suchanfrage *PSD Bank* herangezogen. Man wundert sich, dass über eine Bank so viele soziale Inhalte veröffentlicht werden. Das liegt daran, dass es hier um einen Kostümpreis für den Karneval 2014 geht. In Google findet man hingegen nicht gerade aktuelle Treffer zu der Suchanfrage *PSD Bank* (Abbildung 4). Der Unterschied wird umso drastischer, da Google die Suchanfrage *PSD Bank* als Navigationssuchanfrage interpretiert [8], sprich der Suchende wird als Treffer einen Link zu einer Bankfiliale der PSD Banken erwarten. Anhand des Beispiels *PSD Bank* wird deutlich, wie unterschiedlich der Ergebnisraum ist und dem Suchenden zur Verfügung gestellt wird.

Für Firmen ist dies wichtig, um laufend über aktuelle Posts über die eigene Firma informiert zu bleiben. Im schlimmsten Fall sollten negative Meinungsäußerungen schnellstmöglich entdeckt werden, um darauf zeitnah reagieren zu können.

Vielleicht soll aber auch einfach die Reichweite von bestimmten Aktionen offline wie online gemessen werden. Social Media Monitoring ist deswegen eine der zentralen Massnahmen zur Messung der eigenen Sichtbarkeit im Web. Es kann das Marketingreporting ergänzen, das Krisenmonitoring übernehmen, Marktforschung bedeuten und unter Umständen den Kundendienst ausweiten und verbessern, wenn beispielsweise über Produkte diskutiert wird. Alle diese Aufgaben erfordern eine kontinuierliche Beobachtung von Suchergebnissen und Nennungen in Social Media Angeboten. Eine Ad-hoc-Recherche mit Suchmaschinen eignet sich für die Erfüllung dieser Aufgaben nicht, da die Datenbestände der Suchmaschinen auf der einen Seite hinsichtlich der Social Media Nennungen unvollständig sind, auf der anderen Seite der hohe zeitliche Aufwand, der für die Recherche „von Hand“ vonnöten ist, durch Monitoringsoftware erheblich reduziert werden kann. Entsprechend zeigt sich ein exponentielles Wachstum an Webmonitoringservices bis zum Jahr 2012, wobei 2013 schon ein Rückgang zu verzeichnen war¹.

¹ [http://www.goldbachinteractive.com/aktuell/fachartikel/Social Media monitoring-tool-report-2013](http://www.goldbachinteractive.com/aktuell/fachartikel/Social%20Media%20monitoring-tool-report-2013).

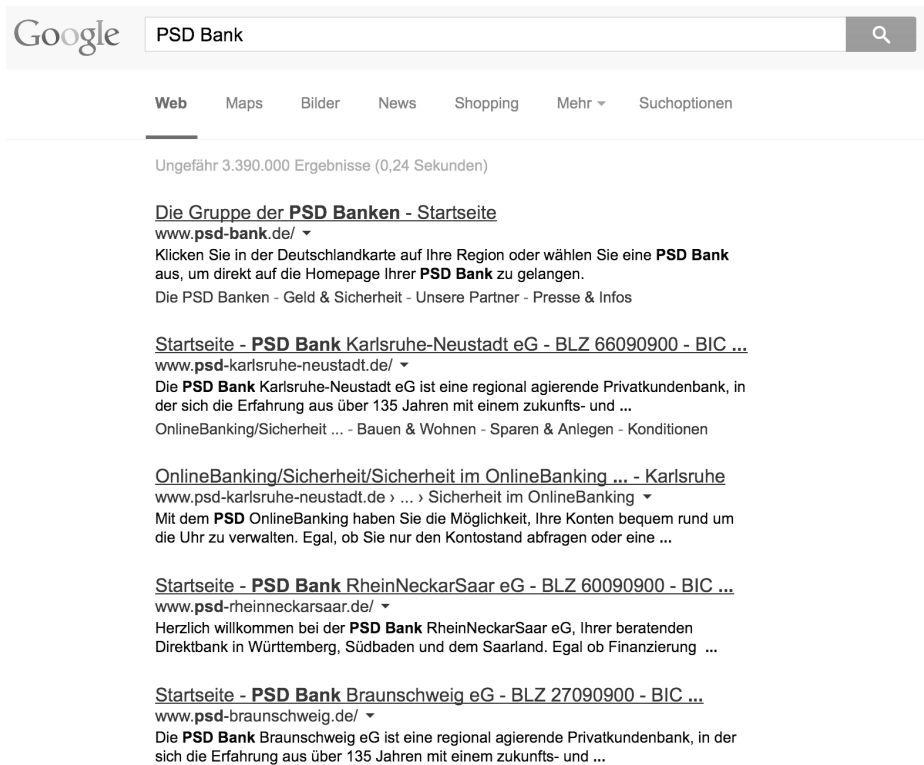


Abbildung 4. Ergebnisse bei Google zu *PSD Bank*.

In diesem Kapitel soll gezeigt werden, woran die Qualität von Suchmaschinen gemessen wird und wie Suchende an eine Webrecherche herangehen. Bei der Websuche und beim Webmonitoring gibt es aber ganz unterschiedliche Suchansätze, die dazu führen, dass in Abhängigkeit der Suche auch der passende Service gewählt werden muss. Das Beispiel der Suche nach *PSD Bank* macht das deutlich: Möchte jemand schnell die Website zu einer Filiale einer PSD Bank finden oder ist jemand daran interessiert, was gerade allgemein auf Facebook veröffentlicht wird? Je nach Anforderung der Suchaufgabe ist die richtige Wahl oder die Kombination der geeigneten Technologie wichtig und zur zufriedenstellenden Beantwortung der Suchaufgabe erforderlich.

1. Definition Websuchmaschinen und Webmonitoringsysteme

Im Folgenden werden klassische Websuchmaschinen und Webmonitoringsysteme definiert und ihre Arbeitsweise kurz dargestellt. Beide stellen Retrieval-Systeme dar; sie haben also die Aufgabe, aufgrund einer Suchanfrage in unstrukturierten Daten Informationen zu finden.

1.1. Definition Websuchmaschine

„Eine Suchmaschine (auch: Websuchmaschine; Universalsuchmaschine) ist ein Computersystem, das Inhalte aus dem World Wide Web mittels Crawling erfasst und über eine Benutzerschnittstelle durchsuchbar macht, wobei die Ergebnisse in einer nach systemseitig angenommener Relevanz geordneten Darstellung aufgeführt werden [9].“ Nach der Eingabe einer Suchanfrage wird bei Websuchmaschinen meist eine kurze Ergebnisliste geliefert (zehn Positionen sind meist üblich), die den gesamten Dokumentenraum repräsentieren und systemseitig als die relevantesten Treffer angesehen werden. Diese Ergebnisse werden in der Regel durch den Titel, den Link und einen kurzen Text (Snippet) wiedergegeben, um dem Suchenden eine Vorschau auf das Dokument zu geben bzw. eine schnelle Evaluation der Ergebnisse zu ermöglichen.

Die wesentlichen Bestandteile bzw. Aufgabenbereiche einer Websuchmaschine sind:

- Entwicklung und Pflege eines eigenen Index, d. h. der Dokumentbasis,
- Verarbeiten von Suchanfragen und das Bereitstellen von Ergebnissen,
- Aufbereitung der Ergebnisse, bspw. Erstellung einer Rangliste anhand der Relevanz.

In der Regel erfolgt die Datenbeschaffung, also das Sammeln der Texte im Web, automatisiert durch einen sogenannten Webcrawler.

Der durch Websuchmaschinen abgebildete Suchprozess ist in Abbildung 5 dargestellt. Nach der Auswahl der Suchmaschine erfolgt die Eingabe einer Suchanfrage, woraufhin eine Trefferseite (*search engine results page*; *SERP*) ausgegeben wird. Auf dieser Seite wählt der Nutzer einen Treffer aus, wobei ihm für seine Auswahlentscheidung Informationen aus den Trefferbeschreibungen (*Snippets*) zur Verfügung stehen. Das so ausgewählte Trefferdokument wird evaluiert; es folgt ggf. eine Suche oder Navigation innerhalb der Website, auf der das Dokument gefunden wurde. Die Recherche kann an jeder Stelle des Suchprozesses abgebrochen werden; weiterhin ist die Rückkehr zu einem der vorangegangenen Schritte möglich.

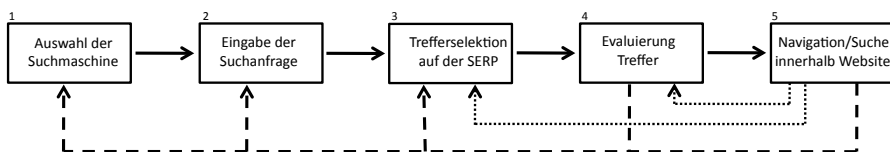


Abbildung 5. Suchprozess in der Websuche [10].

1.2. Definition Webmonitoringservice

Unter Webmonitoring wird allgemein die zeitkritische und systematische Sammlung und Auswertung von Daten im Web verstanden. Auch der Einsatz von Software zur Messung der Besucherzahlen einer Website fällt unter den Begriff Webmonitoring. Im Kontext dieses Kapitels ist unter Webmonitoring die Beobachtung und Analyse von Inhalten im Web zu verstehen.

Der Begriff Webmonitoring wird mittlerweile nochmals durch das Social Media Monitoring eingeschränkt: hier werden nur benutzergenerierte Inhalte (*User Generated Content*, *UGC*) in sozialen Netzwerken beobachtet und ausgewertet. Im Sinne einer

umfassenden Beobachtung von Inhalten im Web ist das Social Media Monitoring unter Webmonitoring zu subsumieren.

Ein Webmonitoringservice sollte ähnliche Komponenten wie eine Websuchmaschine beinhalten, jedoch sind diese durch weitere technische Komponenten erweitert:

- Die Entwicklung und Pflege eines eigenen Index, d. h. der Dokumentbasis mit starkem Fokus auf Social Media,
- das kontinuierliche Verarbeiten von Suchanfragen und das Bereitstellen von möglichst allen Ergebnissen,
- die Aufbereitung der Ergebnisse, bspw. Erstellung einer Rangliste anhand der Neuigkeit oder nach Gruppierung der Quellenart (bspw. Facebook, Twitter).

Webmonitoring kann automatisiert oder manuell erfolgen. Automatisiertes Monitoring erfolgt nur mit einer Software; die Ergebnismenge wird automatisch erstellt und dem Nutzer zur Verfügung gestellt. Diese Technik ist kostengünstig, kann aber in einer Menge unbrauchbarer Ergebnisse resultieren.

Beim halbautomatisierten Webmonitoring werden die unterschiedlichen Quellen mithilfe einer Software zur Verfügung gestellt. Im nächsten Schritt arbeiten dann ausgewählte Personen mit der Datenmenge, indem sie die Suchanfragen weiter einschränken oder die Ergebnisse säubern. Meist bedarf es einer halbautomatisierten Bearbeitung der Ergebnisse.

Eine weitere Form des halbautomatisierten Webmonitorings ist die Nutzung unterschiedlicher Softwareprogramme, die jeweils Daten aus einem spezialisierten Dienst zusammentragen. Die Daten aus den einzelnen Diensten müssen dann manuell aggregiert und aufbereitet werden. Einen Überblick über hilfreiche Werkzeuge für diese Art des Monitorings gibt [11].

Das rein manuelle Webmonitoring bedient sich keiner Software, sondern wird durch eine Person durchgeführt, die Suchanfragen oder Rechercheaufträge in verschiedenen Netzwerken durchführt, die Ergebnisse zusammenträgt und auswertet. Es ist leicht zu erkennen, dass diese Methode zeit- und kostenaufwendig ist. Vor allem hat die recherchierende Person nicht die Möglichkeit, alle Quellen zu durchsuchen, d. h. es gehen unter Umständen wichtige Beiträge verloren.

Der Suchprozess beim Webmonitoring ist im Grunde genommen aufgebaut wie in Abbildung 5, nur dass hier die Suche kontinuierlich abläuft. Es wird nicht ad hoc ein Suchbedürfnis befriedigt, sondern über einen langen Zeitraum hinweg ein Suchauftrag bedient. Ein Teil der Recherche befasst sich mit der Evaluation der Treffer.

1.2.1. Formen des Webmonitorings

Es sind zwei grundsätzliche Formen des Webmonitorings zu nennen:

- Quantitatives Webmonitoring setzt den Fokus auf die reine Messung der Häufigkeit von Texten, die einen bestimmten Begriff enthalten. Diese Begriffe können zum Beispiel Marken-, Personen- oder Unternehmensnamen sein. Die Zahl der gefundenen Beiträge bietet einen guten Überblick, wie relevant ein Thema ist und wie angeregt es in verschiedenen Kanälen diskutiert wird. Zudem kann leicht ein Überblick darüber geschaffen werden, ob ein Thema eher in Facebook oder in den Onlinemassenmedien wie Spiegel.de etc. erörtert wird.

- Qualitatives Monitoring bedeutet dagegen, dass die Daten nicht nur in ihrem Umfang erfasst, sondern auch bewertet werden. Hier ist die „Färbung“ (sogenannte *Sentiments*) der Textstelle wichtig, d. h. handelt es sich generell um eine negative oder positive Äußerung. Zusätzlich wird die Wichtigkeit der Quelle bewertet. Eine Äußerung in einem Forum mit 10.000 aktiven Lesern wird eine höhere Reichweite und Sichtbarkeit erzielen als ein Beitrag auf einer privaten Website, die nur wenige Besucher hat.

Oft wird eine Mischung aus quantitativem und qualitativem Webmonitoring durchgeführt. Das quantitative Webmonitoring bietet eine schnelle Übersicht über die Brisanz und die Publikumswirksamkeit eines Themas, das qualitative dann eine intensivere Sicht auf die Stimmen im Web.

1.3. Horizontale und vertikale Suchmaschinen

Horizontale Suchmaschinen (auch: *Universalsuchmaschinen*) haben den Anspruch, möglichst das gesamte Web zu indexieren und auf eine Suchanfrage mit den bestmöglichen Ergebnissen zu antworten, indem eine kurze repräsentative Ergebnisliste angezeigt wird. Bei den großen Suchmaschinen wie Bing und Google handelt es sich um horizontale Suchmaschinen.

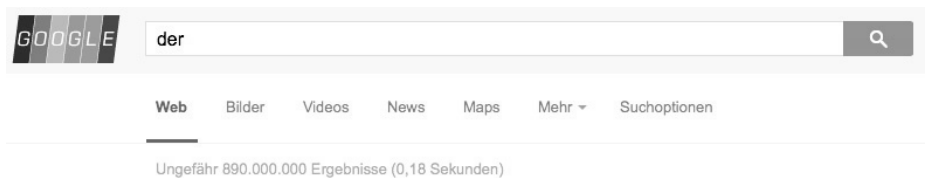


Abbildung 6. Anzahl gefundener Ergebnisse bei der Suchanfrage *der*.

Abbildung 6 zeigt, dass geschätzte 890.000.000 Ergebnisse zu der Suchanfrage *der* gefunden wurden. Klickt man sich jedoch durch alle durch Google verfügbaren Ergebnisse, dann ist bei der 42. Ergebnisseite bei Google Schluss (siehe Abbildung 7).

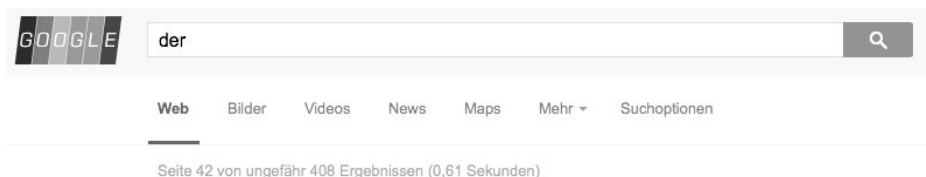


Abbildung 7. Ansicht auf der Ergebnisseite 42 zur Suchanfrage *der*.

Es wurden insgesamt nur 408 Ergebnisse angezeigt, das entspricht in diesem Beispiel 0,00000046 Prozent der geschätzten Anzahl der Ergebnisse. Selbst wenn man also in den Tiefen des Web nach einem bestimmten Dokument suchen möchte oder einfach tatsächlich alles sichten möchte, was zu einem bestimmten Themengebiet

veröffentlicht wird, ist dies schlichtweg nicht möglich. Die Suchanfrage *der* dient hier nur als drastisches Beispiel, da sehr viele Treffer generiert werden.²

Um wirklich tiefer liegende Webseiten und Informationen zu erhalten, muss die Suche stark eingegrenzt werden. Mittlerweile werden aber auch selbst zu langen Suchanfragen auf größere Domains wie amazon oder wikipedia zurückgegriffen bzw. werden weiter hinten andere Deep Links als vorne angezeigt, die aber zur gleichen Website gehören. Zudem bleibt der Ergebnisraum wie oben gezeigt auf einen kleinen Ergebnisraum beschränkt.

Vertikale Suchmaschinen (auch: Spezialsuchmaschinen [12]) beschränken ihren Index willentlich auf ein bestimmtes Gebiet bzw. auf bestimmte Inhalte im Web. Dieser Bereich soll ebenfalls möglichst vollständig indexiert sein. Beispiele sind Suchmaschinen für bestimmte Dokumenttypen (bspw. Nachrichten oder Blogeinträge) und Suchmaschinen für bestimmte Themen (etwa Verbraucherschutz).

Die Auswahl der für die richtige Information in Frage kommenden Webseiten und Datenbanken ist in diesen Fällen kleiner. Dies führt dazu, dass relevante Fundstellen in den Suchmaschinenergebnisseiten tatsächlich nach vorne bzw. in der Rangliste nach oben kommen. Tatsächlich erscheinen hier Ergebnisse auf der ersten Ergebnisseite, obwohl sie bei den horizontalen Suchmaschinen u. U. gar nicht angezeigt werden. Allerdings ist es nicht immer einfach, für eine Anfrage eine themenspezifische Suchmaschine zu finden bzw. gibt es natürlich nicht zu jedem Thema eine spezielle Suchmaschine.

Neben der thematischen Einschränkung des Index kann sich eine Spezialsuchmaschine aber auch auf ein bestimmtes Dateiformat oder einen bestimmten Medienkanal beschränken. Dazu gehören zum Beispiel Flickr (Bilder) oder YouTube (Videos).

Heutige Websuchmaschinen gehören prinzipiell zu den horizontalen Suchmaschinen, was die erste Intention betrifft, nämlich möglichst alles zu indexieren und die bestmögliche Antwortmenge zu präsentieren. Webmonitoringservices wollen ebenfalls möglichst alles erfassen, jedoch nicht eine beste Antwort zu einer Suchanfrage geben. Da sie sich oft auf Social Media Kanäle wie Facebook und Twitter fokussieren, gehören sie eher zu den vertikalen Suchmaschinen, da sie in diesen Bereichen tatsächlich bis in das Deep Web [13, 14, 15 und 16] durchdringen wollen. Zudem geben sich Webmonitoringservices nicht den Anspruch, das gesamte Web zu indexieren, sondern nur zu speziellen von Nutzern angelegten Suchanfragen die komplette Dokumentenmenge zusammenzustellen.

Da Google und Bing nicht mehr nur einen Webindex pflegen, sondern zusätzlich auch Video- und Bildersuche anbieten, gibt es hier zusätzlich einen Drift in die Richtung vertikaler Suchmaschinen. Ein deutliches Beispiel ist die Übernahme von Youtube durch Google. Der klassische Webindex konnte so durch einen reinen Videoindex und mittlerweile Filmindex erweitert werden. Das führt dazu, dass Google mittlerweile dem Suchenden die Wahl lässt, wo er suchen möchte: Web, Bilder oder Videos, etc. Damit werden die vertikalen Suchmaschinen, die Google bietet, direkt zur Verfügung gestellt. Bei der Websuche werden Treffer wie schon oben erwähnt aus den anderen Indexes bereitgestellt und durchmischte Ergebnislisten angezeigt (Universal Search).

² Es muss hier darauf hingewiesen werden, dass es sich bei den Angaben zur Anzahl der gefundenen Treffer nur um grobe Hochrechnungen handelt.

2. Gegenüberstellung Websuche und Webmonitoring

In diesem Abschnitt werden die grundsätzlichen Herangehensweisen an die Websuche bzw. an das Webmonitoring erklärt. Es wird herausgearbeitet, mit welcher Aufgabe Nutzer an die Suche oder auch an eine umfangreiche Recherche im Rahmen des Webmonitorings herangehen.

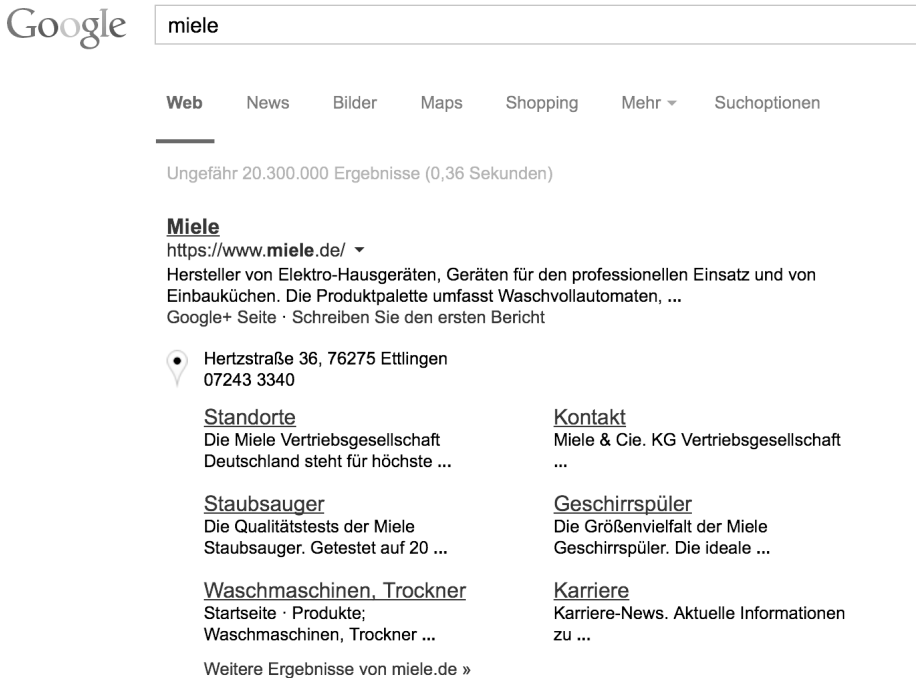


Abbildung 8. Suchanfrage *miele* und das erste Ergebnis.

2.1. Anfragetypen in der Websuche

Suchmaschinen bieten Zugriff auf eine Unmenge verschiedener Informationen und werden zur Befriedigung unterschiedlichster Informationsbedürfnisse genutzt. Grundlegend lassen sich Suchanfragen nach [17] in drei Arten einteilen: navigationsorientiert, informationsorientiert, und transaktionsorientiert. Im Lauf der Jahre wurden verschiedentlich Vorschläge zur weiteren Untergliederung bzw. einer Verfeinerung des Modells gemacht [18, 19, 20 und 21], wobei allerdings keine Einigung über eine Erweiterung oder Ergänzung des Modells erreicht werden konnte.

2.1.1. Navigationsorientiert

Der Nutzer sucht bei navigationsorientierten Anfragen gezielt nach einer Seite, die er bereits kennt, oder von der er glaubt, dass sie existiert.

Typische Navigationssuchanfragen sind beispielsweise Firmennamen oder Namen berühmter Persönlichkeiten. Das Informationsbedürfnis des Nutzers ist i. d. R. nach dem Auffinden der Seite, d. h. eines einzigen Dokuments, befriedigt. Navigations-suchanfragen machen bis zu 50 Prozent aller Anfragen aus [12].

Websuchmaschinen versuchen diese Anfragen mit der entsprechenden URL zu beantworten. Bspw. steht zur Suchanfrage *miele* die Website *miele.de* an erster Stelle. (Abbildung 8). Navigationsorientierte Suchanfragen bestehen meist aus nur einem Begriff oder sogar aus dem tatsächlichen Domainnamen wie *miele.de* oder *buecher.de*.

2.1.2. Informationsorientiert

Der Nutzer sucht bei informationsorientierten Anfragen relevante Informationen zu seinem Thema. Es geht also nicht darum, ein bestimmtes Dokument zu finden, sondern sich auf der Basis mehrerer Dokumente ein Bild von einem Thema zu machen bzw. eine Meinung zu bilden. Im Gegensatz zu den navigationsorientierten Anfragen gibt es hier kein „richtiges“ Dokument, sondern erst die Zusammenstellung mehrerer, auch unterschiedlicher Dokumente durch die Suchmaschine bietet eine gute Ergebnismenge.



Abbildung 9. Spezialfall: Recherche nach dem ersten Präsidenten von Amerika; die Suchanfrage lässt sich mit einem einzigen Dokument beantworten.

Die Anzahl der Dokumente, die zur Befriedigung des Informationsbedürfnisses nötig ist, lässt sich nicht eindeutig bestimmen: Während manche Nutzer (in manchen Fällen) einem Thema auf den Grund gehen möchten, und daher ihre Recherche erst nach der Sichtung einer Vielzahl von Dokumenten beenden, ist in vielen Fällen die Sichtung nur einiger Dokumente für die Befriedigung des Informationsbedürfnisses ausreichend. Und es gibt sogar den Sonderfall, dass ein einziges Dokument ausreicht, nämlich dann, wenn ein Nutzer sich nur cursorisch informieren möchte. Für solche Fälle zeigen Suchmaschinen häufig Wikipedia-Artikel an vorderster Position an (Abbildung 9; siehe auch [4] und [7]).

Informationsorientierte Suchanfragen sind meist länger als navigationsorientierte; sie bestehen meist aus mehreren aneinandergereihten Suchbegriffen.

Eine spezialisierte Ausprägung der informationsorientierten Suche ist das "Leisure Searching" [22], bei dem man mit einer Suchanfrage startet und durch mehrere Frage-Antwort-Zyklen bei einem ganz anderen Themengebiet landet, da auf dem Weg zur Beantwortung immer mehr interessante Fakten aufkommen. So kann eine Suche mit der Recherche nach dem ersten Präsidenten Amerikas beginnen und mit der Suche nach dem amerikanischen Bürgerkrieg enden.

2.1.3. Transaktionsorientiert

Der Nutzer sucht bei transaktionsorientierten Suchanfragen nach Internetseiten, mit denen er zu arbeiten gedenkt. Dies sind zum Beispiel Internetshops (mit direkter Kaufmöglichkeit), Chats usw. Ein prominentes Beispiel ist hier die Suche nach Produktnamen, um anschließend das günstigste Produkt online zu kaufen. Suchmaschinen reagieren auf solche Suchanfragen mit der Präsentation von kommerziellen Links (Adwords), Ergebnissen aus Preissuchmaschinen und sogenannten Shoppingergebnissen.

The screenshot shows a Google search for "miele waschmaschine". The search bar at the top shows the query and a magnifying glass icon. Below the search bar, there are tabs for "Web", "Shopping", "Bilder", "Videos", "News", "Mehr", and "Suchoptionen". The main results area displays "Google Shopping-Ergebnisse für miele waschmaschine". It features a row of five product images with their respective prices and retailers: Miele Waschmaschine (798,00 € at Miele-Shop K...), Miele Waschmaschine (1.099,00 € at OTTO), Miele Gewerbe (2.459,00 € at Miele-Shop K...), Miele W 3371 Wps Edition... (899,00 € at redcoon.de), and Miele Waschmaschine (2.972,80 € at Intermedical24). To the right of these results are several advertisements, including "Neu: Miele Waschmaschine" from preisvergleich.de, "Miele bei OTTO" from miele.otto.de, "Miele Waschmaschine" from amazon.de, and "Miele Waschmaschine" from miele-kowalschik.de. At the bottom, there are more links related to Miele washing machines, including a price comparison link and a link to the Miele website.

Abbildung 10. "Shopping"-Ergebnisse" bei der Suchanfrage *Waschmaschine Miele*.

Die Abbildung 10 zeigt eine typische Ergebnisstruktur für eine transaktionsorientierte Suchanfrage. Der Anteil an Werbung ist gegenüber dem Anteil bei Navigationssuchanfragen deutlich höher.

2.2. Informationsbedürfnisse im Webmonitoring

In den folgenden Unterpunkten werden die Haupteinsatzgebiete für Webmonitoringservices erklärt. Die diesen Einsatzgebieten zugrunde liegenden Informationsbedürfnisse unterscheiden sich dabei fundamental von denen in der Websuche.

2.2.1. Reputation

Bei der Reputationsanalyse wird meist nur nach dem eigenen Firmennamen bzw. Produktnamen gesucht. Es ist vor allem wichtig, negative Äußerungen über das eigene Unternehmen, die eigene Organisation oder Person schnellstmöglich zu entdecken und gegebenenfalls entsprechend zu reagieren.

Eine weitere Möglichkeit ist der Ausbau der Kundenbeziehung und des Kundenservice. Im Falle negativer Äußerungen kann direkt auf die Bedürfnisse des Kunden eingegangen werden und so Dienstleistungen und Produkte verbessert werden.

 Grüne Walzbachtal [x] | Beitrag vom 27.01.2014 21:23:45 Uhr     

GRÜNE WALZBACHTAL KRITISIEREN GENEHMIGUNG DES EINSATZES VON 100% ERSATZBRENNSTOFFEN (ABFÄLLEN) BEI DER ZEMENTHERSTELLUNG DER FIRMA LAFARGE WÖSSINGEN // Die Grünen Walzbachtal begrüßen die verschärften Anforderungen des Regierungspräsidiums im Genehmigungsbescheid gegenüber dem Antrag der Firma Lafarge, kritisieren jedoch insbesondere folgenden Punkt: // Obwohl ein „Zielwert“ für Quecksilber unterhalb des gültigen Grenzwerts festgelegt wird, könnte die Umweltbelastung hier durchaus weiter reduziert werden. Die kontinuierliche Eindüsung von Aktivkohle, als eine Möglichkeit hierzu, wird jedoch von der Firma Lafarge als unverhältnismäßig abgelehnt und von Seiten des Regierungspräsidiums wird mitgeteilt, dass mit den jetzt festgelegten Zielwerten das derzeit technisch und rechtlich Mögliche erreicht ist. Nach unserer Auffassung entspricht dies aber nicht der bestmöglichen Lösung für die Bevölkerung! Wir werden uns daher weiterhin für saubere Luft in Walzbachtal einsetzen. // Ortsverband der Grünen Walzbachtal Walzbachtal, 25.Januar 2014

Abbildung 11. Post der Grünen Walzbachtal gegen das Unternehmen Lafarge.

Die Abbildungen 11 und 12 zeigen Beispiele von negativen Äußerungen, wie sie häufig auf Facebook veröffentlicht werden. Im ersten Beispiel klären die Grünen auf, dass die Firma Lafarge aufgrund der Umweltbelastung nicht im Walzbachtal ansässig werden sollte.

 Marv Apmc [x] | Beitrag vom 05.02.2014 15:44:54 Uhr     

Die Wahrheit über DSDS .. // Schaut euch das mal an Freunde // Krass die Wahrheit kann schon manchmal bitter sein.. // Die Lieder suchen die Produktion aus.. // Die Kandidaten müssen sich erst vor den Produzenten beweisen um erst denen zu Kriterien entsprechen und dann erst zu der Jury weiter gelassen wird.. // Ach ja und die Geschichten und alles die von den Kandidaten stammen soll stammt auch von der Produktion..

Abbildung 12. Post über die „Wahrheit“ bei DSDS

Das zweite Beispiel zeigt, wie schnell Facebook-Nutzer auf Nachrichten in der Presse reagieren und sie auch weiterverbreiten. Solche Posts „vermehren“ sich sehr schnell, indem sie einem Virus gleich an Freunde und Bekannte weitergegeben werden.

2.2.2. Resonanzanalyse

Bei der Resonanzanalyse geht es vor allem darum zu sehen, inwieweit Veröffentlichungen und Marketingaktionen in den verschiedenen Onlinekanälen aufgegriffen wurden. Hier ist es vor allem wichtig, auf einen Blick relevante Zahlen zu bekommen, die bei Google gänzlich fehlen.

In Abbildung 13 ist zu sehen, wie häufig beispielsweise über die PSD Bank in verschiedenen Kanälen berichtet wurde. Dabei lässt sich der zeitliche Verlauf nachvollziehen und ggf. mit Ereignissen und Aktionen in Verbindung bringen.

2.2.3. Marktforschung

Mithilfe von der Recherche interessanter Themen kann bspw. untersucht werden, wie sich Trendthemen entwickeln. Als Beispiel sind hier Cocktail-Trends zu erwähnen. Gibt es neue Namen und Rezepte, die auf Facebook diskutiert werden, wenn ja, wie viele Personen diskutieren darüber und ist es ein lohnenswerter Trend, einen solchen Mode-Cocktail kommerziell zu vertreiben (Bsp.: Cocktail *Hugo* von Käfer).

2.2.4. Wettbewerb

Auch im Hinblick auf die sozialen Aktivitäten der Mitbewerber eines Unternehmens ist das Webmonitoring interessant. Wie betreibt die Konkurrenz Social Media, was für Aktionen werden durchgeführt, wie stark wächst die Zahl der Fans bei Facebook, was schreiben die Kunden und wie ist die Resonanz auf die Aktivitäten. Dies sind ein paar der interessanten Fragen, aus denen Unternehmen durch die Wettbewerbsbeobachtung mittels Webmonitoring lernen können.

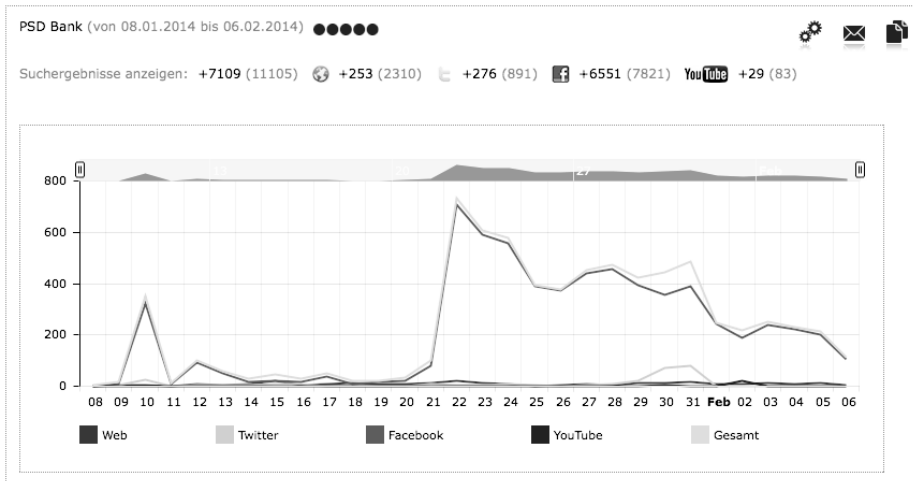


Abbildung 13. Resonanz auf Faschingwettbewerb bei der PSD Bank.

3. Messung der Qualität von Suchmaschinen und Webmonitoringsystemen

Wie gezeigt wurde, haben Suchmaschinen und Webmonitoringsysteme zwar zahlreiche Gemeinsamkeiten, aus den jeweils spezifischen Zielen ergeben sich allerdings nicht nur verschiedene Informationsbedürfnisse, sondern bei der Messung der Qualität dieser Systeme auch unterschiedliche Anforderungen und Kenngrößen. In den folgenden Abschnitten wird zuerst die Qualitätsmessung bei Websuchmaschinen, dann die Qualitätsmessung bei Webmonitoringsystemen behandelt. Schließlich werden Bereiche benannt, die für die Bestimmung der Qualität beider Systeme relevant sind.

3.1. Klassische Qualitätsmaße für Information-Retrieval-Systeme

Die beiden klassischen Maße zur Bestimmung der Güte von Treffermengen in Information-Retrieval-Systemen sind *Precision* und *Recall* [23, 24]. Zwar wurden im Lauf der Jahre zahlreiche weitere Kennzahlen entwickelt und erprobt, diese basieren aber in der Regel auf den beiden Klassikern. Während Recall und Precision auf einen impliziten Nutzermodell beruhen, werden neuere Kennzahlen mit einem expliziten Nutzermodell im Hintergrund entwickelt [25], um das tatsächliche Suchverhalten besser abbilden zu können.

Bei der Messung von Precision und Recall wird davon ausgegangen, dass ein Nutzer eine Suchanfrage eingegeben hat und dann die vom System ausgegebenen Treffer sichtet. Dabei wird ein Treffer entweder als relevant oder als nicht relevant bewertet. Weiterhin wird vorausgesetzt, dass die Anzahl der insgesamt im Datenbestand vorhandenen Treffer zu der jeweiligen Suchanfrage gemessen werden kann. Aus diesen Bewertungen lassen sich nun die Kennzahlen berechnen.

Die *Precision* misst nun, wie hoch der Anteil der relevanten Treffer an den insgesamt zur Suchanfrage ausgegebenen Treffern ist. Werden etwa 10 Treffer ausgegeben, von denen 4 relevant sind, ergibt sich eine precision von $4/10 = 0,4$. Diese Messung kann über eine Vielzahl von Suchanfragen durchgeführt werden; aus den Werten für die einzelnen Suchanfragen wird dann der Mittelwert gebildet. Da es oft nicht möglich ist, die gesamte ausgegebene Treffermenge zu bewerten (und dies unter der Voraussetzung, das reale Nutzerverhalten abbilden zu wollen, oft auch gar nicht sinnvoll ist), wird die precision oft nur bis zu einer bestimmten Trefferposition ("cut-off-Wert") gemessen.

Der *Recall* als "Gegenstück" zur Precision gibt an, inwieweit die in der Datenbank enthaltenen Treffer *vollständig* in der Treffermenge enthalten sind. Er errechnet sich als die Zahl der gefundenen relevanten Dokumente durch die Anzahl der insgesamt im Datenbestand vorhandenen relevanten Dokumente. Wurden zum Beispiel 10 relevante Dokumente gefunden, wobei im Datenbestand insgesamt 50 relevante Dokumente vorhanden sind, so ergibt sich ein Recall von $10/50 = 0,2$. Die Schwierigkeit bei der Messung des Recalls ist, dass in der Regel nicht bekannt ist, wie viele relevante Dokumente zu einem Thema insgesamt im Datenbestand vorhanden sind. Dieses Problem verstärkt sich im Vergleich zu klassischen Information-Retrieval-Systemen bei Suchmaschinen noch, da sie über riesige, kaum abgrenzbare Datenbestände verfügen [26].

Precision und *Recall* sind gegenläufig: Erhöht man – etwa durch präzisere Suchanfragen – die Precision, sinkt in der Regel der Recall und umgekehrt. Die beiden Extremfälle stellen auf der einen Seite eine Suchanfrage dar, die ein einziges relevantes Dokument ergibt (und damit eine Precision von 1 erreichen würde), auf der anderen Seite eine Suchanfrage, die schlicht alle im Datenbestand vorhandenen Dokumente abfragen würde (und damit einen Recall von 1 erreichen würde). Beide Fälle sind jedoch aus Nutzersicht nicht zufriedenstellend.

Im Kontext dieses Kapitels sind die beiden Werte Recall und Precision besonders von Bedeutung, da mit ihnen der unterschiedliche Anspruch an Websuchmaschinen und Webmonitoringsysteme erklärt werden kann. Bei der Websuche geht es meist darum, eine bestimmte Menge von relevanten Dokumenten zu einem Thema zu finden. Wie in den vorangegangenen Abschnitten zu den unterschiedlichen Anfragetypen gezeigt wurde, kann diese Zahl zwischen nur einem einzigen und einer Vielzahl von Dokumenten liegen. In den weit häufigsten Fällen werden Websuchmaschinen allerdings dazu benutzt, nur eines bis zu wenigen relevanten Dokumenten aufzufinden.

Eine recall-orientierte Recherche ist in der Websuche dagegen eher selten. „Recall-orientiert“ bedeutet, dass möglichst *alle* relevanten gefunden werden sollen. Dass dies zu Lasten der Precision geht, spielt hier oft nur eine untergeordnete Rolle: Der Nutzer ist bereit, auch eine Vielzahl von nicht relevanten Dokumenten zu sichten, wenn dies denn dazu führt, dass schließlich möglichst alle relevanten Dokumente gefunden werden.

Das klassische Beispiel für eine solche recall-orientierte Suche ist die Patentrecherche. Dort kommt es darauf an, alle relevanten Dokumente zu finden, um nicht zu

einem Zeitpunkt, wenn bereits viel Geld in eine eigene Entwicklung geflossen ist, feststellen zu müssen, dass es auf diese Entwicklung bereits einen Patentschutz gibt.

Aber auch das Webmonitoring kann als eine recall-orientierte Recherche angesehen werden: Es kommt hierbei nicht nur darauf an, die von einer Suchmaschine auf den vorderen Plätzen ausgegebenen Dokumente zu finden, sondern es geht vielmehr darum, möglichst alle relevanten Dokumente (bzw. Nennungen in verschiedenen Medienkanälen wie Facebook etc.) zu finden. Neben der oben beschriebenen Erweiterung der Quellenbasis spielt dabei auch die Erweiterung der Suche auf eine Vielzahl von Suchanfragen eine Rolle. Beim Webmonitoring handelt es zudem um eine kontinuierliche Suche, die nicht mit einer abgeschlossenen Suche befriedigt ist, sondern täglich oder in sehr kurzen Intervallen durchgeführt wird, um im Ablauf der Zeit ein Spiegelbild der Meinungen und Stimmungen im Web zu erhalten.

Ein weiterer interessanter Wert ist der sogenannte *Fallout*. Ein Webmonitoringssystem liefert vielleicht alle Treffer zu einem Begriff, aber unter Umständen ist die Suchanfrage zu unspezifisch und dadurch sind 90 Prozent der Treffer nicht themenverwandt. Oft muss hier der Suchende die Suchanfragen neu konfigurieren und schärfen. Beispielsweise liefert die Suchanfrage *miele* alles zu Miele, aber es sind eigentlich nur die Ergebnisse zu einer speziellen Miele-Waschmaschine von Interesse. Dann muss die kontinuierliche Suchanfrage *miele* eingeschränkt werden.

3.2. Messung der Qualität von Suchmaschinen

Während mit Precision und Recall sowie weiteren, auf diesen Kennzahlen basierenden Werten die Qualität der von einer Suchmaschine ausgegebenen Treffer gemessen werden kann, sind für eine umfassende Bewertung der Qualität von Suchmaschinen nach dem Modell von Lewandowski und Höchstötter [27] vier Bereiche zu berücksichtigen: Die Qualität des Index, die Qualität der Ergebnisse, die Qualität der Suchfunktionen und die Nutzerfreundlichkeit.

3.2.1. Indexqualität

Der Index einer Suchmaschine ist der ihr zugrunde liegende Datenbestand. Im Gegensatz zu den Datenbeständen in klassischen Datenbanken ist ein Suchmaschinenindex allerdings erstens nicht klar in seiner Abdeckung der Dokumentenmenge bestimmbar (welcher Anteil der im Web vorhandenen Dokumente findet sich im Index wieder?) und zweitens beständigen Veränderungen und Aktualisierungen des Web unterworfen [28]. Daraus ergibt sich, dass es für die Bestimmung der Qualität einer Suchmaschine eine Rolle spielt, inwieweit sie in der Lage ist, ein vollständiges und aktuelles Abbild des tatsächlichen World Wide Web zu bieten.

Doch auch wenn man die Aktualität aus der Betrachtung herauslässt, können sich unbefriedigende Resultate ergeben. So wurde festgestellt, dass die Inhalte aus unterschiedlichen Ländern in den Indexen der Suchmaschinen zu einem unterschiedlichen Grad abgedeckt werden [29]. Dies bedeutet, dass der Grad der *globalen* Vollständigkeit noch nichts über die Abdeckung in einem bestimmten, für einen Anwendungsfall relevanten Bereich aussagt. So würde sich der Einsatz einer Suchmaschine, die zwar einen Großteil des Web, allerdings kaum dessen deutschsprachigen Bereich abdeckt, für einen Suchenden in Deutschland nicht empfehlen.

Hinsichtlich der Aktualität der Datenbestände der bekannten Suchmaschinen konnte gezeigt werden, dass auch bekannte Webseiten nicht immer aktuell in den

Datenbeständen vorhanden sind und dass die Aktualisierungszyklen nicht unbedingt regelmäßigen Mustern folgen [30, 31].

3.2.2. Ergebnisqualität

In Abschnitt 3.1 wurde bereits auf die Messung der Qualität der Suchresultate mit Kennzahlen eingegangen. Aus der Ausrichtung der bekannten Suchmaschinen auf eine precision-orientierte Recherche ergibt sich, dass ihre Rankingfunktionen darauf ausgerichtet sind, einige hoch relevante Dokumente auf die vorderen Plätze der Trefferliste zu bringen. Dies liefert bei eher recall-orientierten Recherchen (etwa im Rahmen eines Webmonitorings „von Hand“) vordergründig erst einmal relevante Ergebnisse; in der weiteren Listung der Ergebnisse setzt sich diese Relevanz aber häufig nicht fort.

3.2.3. Qualität der Suchfunktionen

Bei einer Suchmaschine sollten spezielle Suchfunktionen angeboten werden, die den Nutzer bei der Suche unterstützen (beispielsweise Angabe der gewünschten Sprache oder eines Dokumenttyps) und ihm damit erlauben, die Qualität der Ergebnisse zu beeinflussen, indem beispielsweise ungewünschte Resultate ausgeschlossen werden.

Dem gegenüber steht allerdings das Verhalten der Suchmaschinennutzer, die meist nur wenig kognitive und zeitliche Energie in die Formulierung ihrer Suchanfragen stecken [32]. Dies wird auch in zahlreichen Studien, die tatsächliche Suchanfragen kommerzieller Suchmaschinen auswerten, bestätigt (u.a. [33] und [34]). Die Eingabe elaborierter Suchanfragen mit Operatoren oder sonstigen Kommandos ist in der Websuche selten und in vielen Fällen fehlerhaft.

Dieses Nutzerverhalten führt nun wiederum dazu, dass sich Suchmaschinen daran ausrichten. Erweiterte Suchfunktionen werden teils gar nicht mehr angeboten (so unterstützt Bing zwar erweiterte Suchen, sofern sie durch eine Qualifizierung der Sucheingabe geschehen, bietet jedoch kein erweitertes Suchformular an), teils sind sie nur eingeschränkt funktionstüchtig (so etwa das Boolesche OR bei Google, das in Kombination mit anderen Operatoren nicht funktioniert).

Beispiele für Suchfunktionen, die nur eingeschränkt funktionieren, sind die datumsbeschränkte Suche [35] und die Beschränkung nach Sprache [36].

3.2.4. Nutzerfreundlichkeit

Die Nutzerfreundlichkeit von Websuchmaschinen beruht im Wesentlichen darauf, dass die Benutzerschnittstelle einfach gehalten ist und die Ergebnispräsentation sich nicht allzu weit von erlernten Mustern entfernt.

Die Benutzerschnittstelle der gängigen Suchmaschinen besteht zunächst einmal nur aus einem einzigen Suchschlitz, in den die Suchanfrage in beliebigem Format eingetragen werden kann. So lassen sich Schlüsselwörter aneinanderreihen, ganze Sätze eingeben, Fragen stellen oder mithilfe der von der jeweiligen Suchmaschine unterstützten Befehle komplexe Suchanfragen formulieren. Dadurch bedienen Suchmaschinen mit einem einzigen Nutzerinterface nicht nur eine Vielzahl von Informationsbedürfnissen, sondern auch Nutzergruppen von Anfängern bis zu Rechercheprofis.

Die Suchergebnisseite (s. o.) folgt bekannten Mustern und baut sich zentral um die geordnete Ergebnisliste auf. Veränderungen in der Präsentation der Ergebnisse führen die Suchmaschinenbetreiber graduell durch; auch größere Veränderungen wie die

Universal Search oder Googles Knowledge Graph stellen Ergänzungen der Ergebnisse dar und verändern deren Struktur nicht grundlegend.

Neben den beiden genannten Faktoren spielen für die Benutzerfreundlichkeit einer Suchmaschine natürlich auch zahlreiche weitere Standards für die Gestaltung von Elementen der Benutzerinteraktion eine Rolle. Zusammengefasst und speziell für das Thema Suche aufbereitet finden sich diese in [6].

3.3. Messung der Qualität von Monitoringsystemen

Bei der Messung der Qualität von Monitoringsystemen kommen nicht nur die klassischen Maße zum Tragen, sondern auch spezielle Maße, um diese Systeme vergleichbar zu machen. Die wichtigsten Punkte werden im Folgenden aufgeführt.

3.3.1. Quellenabdeckung

Die Quellenabdeckung ist einer der wichtigsten Punkte. Hier spielt vor allem eine Rolle, welche Social Media Netzwerke beobachtet werden. Normalerweise sollten die bekanntesten Social Media Services wie Youtube, Facebook usw. integriert sein.

Aber auch die Abdeckung im Web ist von Interesse. Hier sollten zumindest die wichtigsten Onlinemedien und Onlinejournale, Zeitungen, Blogs und Foren regelmäßig abgerufen werden.

Ein weiterer wichtiger Punkt ist die Trainierbarkeit des Systems. So stellt sich etwa die Frage, ob der Webmonitoringservice die Datenbasis erweitert. Können die Nutzer auch Quellen, die für sie von besonderem Interesse sind, einfach vorschlagen, so dass diese in Zukunft regelmäßig vom Webmonitoringservice besucht werden?

3.3.2. Aktualität von Social Media Inhalten

Das wichtigste Qualitätsmaß beim Webmonitoring ist die Aktualität bzw. Echtzeitbereitstellung der Social Media Inhalte im Web. Die Nutzer wollen ja schließlich möglichst schnell mitbekommen, was im Web veröffentlicht wird, um schnellstmöglich darauf reagieren zu können. Bei der Aktualität ist zu unterscheiden zwischen der Aktualität des Index (also des zugrundeliegenden Datenbestands) und der Aktualität der Suchergebnisse durch das Webmonitoringsystem.

3.3.2.1. Index

Der Index sollte so aktuell wie möglich gehalten werden. Posts von Facebook und anderen zeitkritischen Diensten sollten möglichst schnell im Index zur Verfügung stehen. Meist erwarten die Nutzer die Bereitstellung in Echtzeit bzw. sofort, was aber technisch praktisch nicht zu ermöglichen sind.

3.3.2.2. Ergebnisse

Die kontinuierlichen Suchanfragen oder die sogenannten Suchprofile sollten möglichst laufend abgefragt werden, damit die Ergebnisse immer auf dem neuesten Stand bleiben. Eine wöchentliche oder auch nur tägliche Abfrage und Bereitstellung der

Ergebnisse ist gerade bei brisanten Themen zu selten. Die Frequenz muss hier möglichst hoch sein; eine sekundeschnelle Taktung ist wünschenswert.

3.3.3. Individualisierbarkeit

Webmonitoringservices müssen einen hohen Grad an Individualisierbarkeit bieten. Zum einen wird gewünscht, dass „eigene“ Quellen durchsucht werden. Aber oft soll auch die Oberfläche des Service das Design und Logo der Firma, die den Webmonitoringsservice nutzt, zeigen.

Ein weiterer wichtiger Punkt ist, wie individuell und schnell die kontinuierlichen Rechercheaufträge verändert und neu formuliert werden können. Die Nutzer brauchen dies, um schnellstmöglich auf Suchbedürfnisse reagieren zu können.

3.3.4. Reporting

Die meisten Nutzer wollen sich nicht stündlich in das System einloggen, um zu sehen, ob unter Umständen neue Ergebnisse zu ihrer Suchanfrage gefunden wurden. Interessanter ist der Fall, dass im Fall eines neuen Ergebnisses direkt eine Benachrichtigungsemail an den Nutzer gesendet wird.

In anderen Anwendungsbeispielen wird ein Report für jeden Tag, jede Woche oder jeden Monat eingerichtet, in dem die wichtigsten Ergebnisse und Kennzahlen zusammengefasst sind.

3.3.5. Aggregation der Ergebnisse

Webmonitoringsysteme können natürlich nicht einfach nur Ergebnislisten präsentieren, die die neuesten Ergebnisse zuoberst anzeigen. Meist muss auf einer übersichtlichen und übergeordneten Ergebnisseite gezeigt werden, wie viele Veröffentlichungen wie Posts, Kommentare oder Onlineartikel wo veröffentlicht wurden. Auszüge der wichtigsten Ergebnisse sind zur weiteren Navigation ebenfalls hilfreich. Der Nutzer kann dann selber anhand des ersten Eindrucks entscheiden, ob er durch alle Ergebnisse manuell durchgehen möchte, um diese einzeln zu evaluieren.

3.4. Allgemeine Herausforderungen

Es gibt nicht nur Herausforderungen, die für Websuchmaschinen oder Webmonitoringsysteme wichtig bzw. spezifisch sind, sondern auch solche, die bei beiden eine große Rolle spielen.

Manche dieser Anforderungen spielen mal bei dem einen, mal bei dem anderen Suchsystem eine wichtigere Rolle; dies liegt daran, dass die Herangehensweise und die Intention bei der Suche eine andere ist. Auch durch die Ergebnispräsentation hat hier die eine oder andere Art von Suchmaschine „kleine“ Vorteile oder aber auch Nachteile.

3.4.1. Mehrdeutigkeit

Suchanfragen sind oft unpräzise. So kann die Suchmaschine nicht selbstständig entscheiden, ob beim Begriff *Laster* nach einem Lkw oder einer schlechten Angewohnheit gesucht werden soll (semantische Korrektheit). Umgekehrt sollte die Suchmaschine

nicht zu stur auf dem eingegebenen Begriff bestehen. Sie sollte auch Synonyme einbeziehen, damit beispielsweise die Suchanfrage *Rechner Linux* auch Seiten findet, die statt *Rechner* das Wort *Computer* enthalten.

3.4.2. Grammatik

Viele mögliche Treffer gehen verloren, weil der Nutzer nach einer bestimmten grammatikalischen Form eines Suchbegriffes sucht. So findet die Suche nach dem Begriff *Auto* zwar alle im Suchindex enthaltenen Seiten, die dieses Wort in der eingegebenen Form enthalten, nicht aber jene mit dem Wort *Autos*.

Manche Suchmaschinen erlauben die Suche mittels Wildcards, mit denen sich dieses Problem teilweise umgehen lässt (z. B. berücksichtigt die Suchanfrage *Auto** auch die Begriffe *Autos* oder *Automatismus*), allerdings muss der Nutzer diese Möglichkeit auch kennen.

Weiterhin wird oft Stemming verwendet, wodurch Wörter auf ihren Grundstamm reduziert werden. So ist einerseits eine Abfrage nach ähnlichen Wortformen möglich (*schöne Blumen* findet so auch *schöne Blume*), außerdem wird die Anzahl der Begriffe im Index des Suchsystems reduziert. Die Nachteile des Stemmings können durch eine linguistische Suche ausgeglichen werden, indem alle Wortvarianten erzeugt werden. Eine weitere Möglichkeit ist der Einsatz statistischer Verfahren, mit denen die Suchmaschine die Anfrage z. B. durch das Auftauchen verschiedener verwandter Begriffe auf Webseiten danach bewertet, ob beispielsweise mit der Suche nach *Auto reparieren* auch die Suche nach *Autos reparatur* oder *Automatismus repariert* gemeint gewesen sein könnte.

3.4.3. Satzzeichen

Fachbegriffe und Produktbezeichnungen, zu deren Eigennamen ein Satzzeichen gehört (z. B. Apples Webservice, .Mac oder C/net), können bei den gängigen Suchmaschinen nicht effektiv gesucht und gefunden werden. Lediglich für ein paar sehr häufige Begriffe (z. B. .Net, C#, oder C++) wurden Ausnahmen geschaffen.

Beim Webmonitoring werden aber gerade solche Suchanfragen oft nachgefragt, da Unternehmen tatsächlich wissen wollen, wie häufig beispielsweise *H&M* in Social Media Netzwerken diskutiert wird.

3.4.4. Datenmenge

Suchmaschinen können nicht mit der gleichen Schnelligkeit, mit der im Web und vor allem in Social Media Plattformen veröffentlicht wird, die Textstellen indexieren und in der Suche verfügbar machen. Dabei ist der den Suchmaschinen unbekannte Teil – das sogenannte Deep Web [13, 14, 15] – noch gar nicht eingerechnet. Gerade hier haben Websuchmaschinen den Vorteil, dass sie sich nicht den Anspruch geben müssen, alles zu präsentieren, was in Social Media Plattformen, die ein Bestandteil des Deep Web sind, veröffentlicht wird.

Webmonitoringsysteme crawlen dafür nicht das gesamte Web, sondern nur die Teile, die für eine „gebuchte“ Suchabfrage interessant sind. Ansonsten wäre ein solcher Service technisch nicht machbar. Es werden also erst dann Daten gesammelt, wenn es dazu auch einen Suchbedarf gibt bzw. eine Suchanfrage für die kontinuierliche Suche

angelegt wird. Websuchmaschinen sammeln dagegen „alle“ Daten und bekommen dann eine mehr oder weniger bekannte Suchanfrage gestellt.

3.4.5. Aktualität von Webinhalten

Viele Webseiten werden häufig aktualisiert, was die Suchmaschinen zwingt, diese Seiten nach definierbaren Regeln immer wieder zu besuchen. Dieses ist auch notwendig, um zwischenzeitlich aus der Datenbasis entfernte Dokumente zu erkennen und nicht länger als Ergebnis anzubieten. Das regelmäßige Herunterladen der mehreren Milliarden Dokumente, die eine Suchmaschine im Index hat, stellt große Anforderungen an die Netzwerkressourcen (Traffic) des Suchmaschinenbetreibers.

3.4.6. Spam

Mittels Suchmaschinen-Spamming versuchen manche Websitebetreiber, die Ranking-Algorithmen der Suchmaschinen zu überlisten, um eine bessere Platzierung für gewisse Suchanfragen zu bekommen. Sowohl den Betreibern der Suchmaschine als auch deren Kunden schadet dies, da nun nicht mehr die relevantesten Dokumente zuerst angezeigt werden. Durch das sehr angepasste Ranking der Websuchmaschinen und durch neue Algorithmen ist Spam bei Websuchmaschinen im Gegensatz zu früher allerdings weit weniger anzutreffen [37]. Es müssen schon sehr spezielle Suchanfragen formuliert werden, um solche Treffer zu generieren.

Webmonitoringsysteme haben nicht das Problem des Spammings aufgrund von dedizierten Rankingalgorithmen, sondern durch Social Media Spammer, sprich Personen, die mehrere bspw. Twitter-Accounts gleichzeitig betreiben und gleiche Posts in Minutentaktung veröffentlichen. Dadurch kommen viele redundante Daten zustande, die zudem auch in den Ergebnissen auftauchen. Das Problem der redundanten Daten haben auch Suchmaschinen. Aber die erste Ergebnisseite wird davon kaum betroffen sein. Durch Regeln und Algorithmen kann aber auch die Ergebnisliste von solchen Treffern bereinigt werden. Bspw. kann definiert werden, dass nur alle zehn Minuten Posts von einer Quelle in den Index gelangen dürfen. Außerdem können bekannte Spammer auf einer Blacklist geführt und damit ausgeschlossen werden.

3.4.7. Technik

Suchen in sehr großen Datenmengen technisch so umzusetzen, dass die Verfügbarkeit hoch ist und die Antwortzeiten niedrig bzw. kaum spürbar, stellt große Anforderungen an den Suchmaschinenbetreiber. Die Datenmenge muss auf mehrere kleinere Indexe verteilt werden, in denen gleichzeitig gesucht werden kann.

Websuchmaschinen haben hier gegenüber Webmonitoringsystemen einen Vorteil, da ad hoc „nur“ die erste Ergebnisseite „gut aussehen“ muss. Von Webmonitoringsystemen wird hingegen erwartet, dass möglichst alle Ergebnisse mit den entsprechenden visuellen Auswertungen ad hoc bereit stehen. Das ist aber meist nicht machbar, da tatsächlich mehrere Indexe durchsucht werden müssen, um möglichst alle Ergebnisse bereitzustellen. Die Folge sind längere Antwortzeiten, die um ein Vielfaches höher liegen als die von Websuchmaschinen.

3.4.8. Recht

Suchmaschinen werden meist international betrieben und bieten somit Benutzern auch Ergebnisse von Servern, die in anderen Ländern stehen. Da die Gesetzgebungen der verschiedenen Länder unterschiedliche Auffassungen davon haben, welche Inhalte erlaubt sind, geraten Betreiber von Suchmaschinen oft unter Druck, gewisse Seiten aus ihren Ergebnissen auszuschließen.

So zeigen die marktführenden Suchmaschinen seit 2006 bei aus Deutschland stammenden Suchanfragen keine Webseiten mehr von Websites als Treffer an, die von der Bundesprüfstelle für jugendgefährdende Medien als jugendgefährdend eingestuft wurden. Diese Praxis erfolgt von Seiten der Suchmaschinen freiwillig als automatisiertes Verfahren (Filtermodul) im Rahmen des Vereins Freiwillige Selbstkontrolle Multimedia-Diensteanbieter.

3.4.9. Datenschutz

Vor allem bei einer Personensuchmaschine ist der Datenschutz ein heikles Thema. Wenn über eine Personensuchmaschine eine Suche nach einem Namen gestartet wird, betreffen die Ergebnisse der Suche nur Daten, die allgemein zugänglich sind. Diese Daten sind auch ohne die Suchmaschine für die Allgemeinheit ohne Durchführung einer Registrierung bei einem Dienst oder Ähnlichem zugänglich. Die Personensuchmaschine selbst hält keine eigenen Informationen vor, sondern vermittelt nur den Zugang zu diesen. Korrekturen oder Löschungen müssen an der jeweiligen Ursprungsquelle vorgenommen werden.

4. Fazit

Die Ad-hoc-Recherche in Suchmaschinen und das Webmonitoring erfüllen unterschiedliche Anwendungszwecke. Während eine Websuche in erster Linie dem Auffinden einer mehr oder weniger großen Zahl von relevanten Dokumenten zu einem Thema dient, wird das Webmonitoring durchgeführt, um zu einem vorgegebenen Thema möglichst alle Dokumente zu finden. Dabei ist es von Bedeutung, neu im Web auftauchende Dokumente möglichst zeitnah zu registrieren, um entsprechend reagieren zu können.

Häufig werden jedoch Ad-hoc-Recherchen durchgeführt, wenn eigentlich ein Webmonitoring angebracht wäre. Dabei wird dann oft auch übersehen, dass mit der Recherche in allgemeinen Suchmaschinen bestimmte, oft hoch relevante Inhalte nicht mit abgedeckt werden. Hier ist vor allem an die Inhalte aus Sozialen Netzwerken, die von den Suchmaschinen nicht hinreichend erfasst werden, zu denken.

Eine kontinuierliche Beobachtung von Webinhalten kann automatisch, halb-automatisch oder von Hand geschehen. Bei letzterer Möglichkeit muss allerdings beachtet werden, dass trotz einer Vielzahl von verfügbaren Werkzeugen ein hoher zeitlicher Aufwand entsteht, ohne dass damit eine Erfolgsgarantie gegeben wäre.

Während sich also mit der Hilfe von Websuchmaschinen einige Aufgaben des Monitorings mit wesentlichen Beschränkungen bewerkstelligen lassen, sind umgekehrt auch ausgereifte Webmonitoringlösungen selbst keine Suchmaschinen im klassischen Sinne. Sie sind nicht darauf ausgelegt, spontan entstandene Informationsbedürfnisse ad hoc in Sekundenschnelle zu befriedigen. Vielmehr erfordert ein erfolgreiches Web-

monitoring ein kompetent erstelltes Suchprofil und eine Auswahl von relevanten Quellen, um zu optimalen Ergebnissen zu kommen. So kann dann wiederum bei einem neuen Thema, das nicht bereits ins Webmonitoring eingespeist wurden, die Recherche in einer Websuchmaschine der richtige Startpunkt sein.

Literaturangaben

- [1] Lewandowski, D. (2013b). Suchmaschinenindices. In D. Lewandowski (Hrsg.), *Handbuch Internet-Suchmaschinen 3: Suchmaschinen zwischen Technik und Gesellschaft* (S. 143–161). Berlin: Akademische Verlagsgesellschaft AKA GmbH.
- [2] Schmidt, H. (2012, March 12). Googles Marktanteil steigt auf 96 Prozent in Deutschland. Focus Online.
- [3] Quirnbach, S. (2009). Universal Search - Kontextuelle Einbindung von unterschiedlicher Quellen und Auswirkungen auf das User Interface. In D. Lewandowski (Hrsg.), *Handbuch Internet-Suchmaschinen* (S. 220–248). Heidelberg: Akademische Verlagsgesellschaft Aka GmbH.
- [4] Höchstötter, N., & Lewandowski, D. (2009). What users see – Structures in search engine results pages. *Information Sciences*, 179(12), 1796–1812.
- [5] Lewandowski, D., & Quirnbach, S. (2013). Suchvorschläge während der Eingabe. In D. Lewandowski (Hrsg.), *Handbuch Internet-Suchmaschinen 3: Suchmaschinen zwischen Technik und Gesellschaft* (S. 273–298). Berlin: Akademische Verlagsgesellschaft AKA GmbH.
- [6] Quirnbach, S. M. (2012). *Suchmaschinen: User Experience, Usability und nutzerzentrierte Website-Gestaltung*. Heidelberg: Springer.
- [7] Lewandowski, D., & Spree, U. (2011). Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality? *Journal of the American Society for Information Science and Technology*, 62(1), 117–132.
- [8] Lewandowski, D. (2011). Query Understanding. In D. Lewandowski (Hrsg.): *Handbuch Internet-Suchmaschinen 2: Neue Entwicklungen in der Web-Suche*. Heidelberg: Akademische Verlagsgesellschaft AKA; 2011:55–75.
- [9] Lewandowski, D. (2013a). Suchmaschinen. In R. Kuhlen, W. Semar, & D. Strauch (Hrsg.), *Grundlagen der praktischen Information und Dokumentation*. (6. Ausgabe., S. 495–508). Berlin: De Gruyter.
- [10] Lewandowski, D. (2012). Informationskompetenz und das Potenzial der Internetsuchmaschinen. In W. Sühl-Strohmenger (Hrsg.), *Handbuch Informationskompetenz* (S. 101–109). Berlin: De Gruyter.
- [11] Plieninger, J. (2011). Tools für das Monitoring. In Lewa (Hrsg.), *Handbuch Internet-Suchmaschinen 2: Neue Entwicklungen in der Web-Suche*. Heidelberg: Akademische Verlagsgesellschaft AKA GmbH.
- [12] Lewandowski, D. (2009). Spezialsuchmaschinen. In D. Lewandowski (Hrsg.), *Handbuch Internet-Suchmaschinen* (S. 53–69). Heidelberg: Akademische Verlagsgesellschaft AKA GmbH.
- [13] Bergman, M. K. (2001). The deep Web: Surfacing hidden value. *Journal of Electronic Publishing*, 7(1), 1–17.
- [14] Devine, J., & Egger-Sider, F. (2009). *Going Beyond Google: The Invisible Web in Learning and Teaching* (S. 111–127). New York, London: Neal-Schuman Publishers.
- [15] Devine, J., & Egger-Sider, F. (2014). *Going beyond Google again: Strategies for using and teaching the invisible web*. London: Facet Publishing.
- [16] Sherman, C., & Price, G. (2001). *The Invisible Web: Finding Hidden Internet Resources Search Engines Can't See*. Cyberage Books.
- [17] Broder, A. (2002). A taxonomy of web search. *ACM Sigir Forum*, 36(2), 3–10.
- [18] Calderon-Benavides, L., Gonzalez-Caro, C., & Baeza-Yates, R. (2010). Towards a Deeper Understanding of the User's Query Intent. In *SIGIR 2010 Workshop on Query Representation and Understanding* (S. 21–24). New York: ACM.
- [19] Kang, I. H., & Kim, G. C. (2003). Query type classification for web document retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval* (S. 64–71). Division of Computer Science, Department of EECS, KAIST: ACM.
- [20] Rose, D. E., & Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web* (S. 13–19). ACM.
- [21] Lewandowski, D., Drechsler, J., & Mach, S. Von. (2012). Deriving Query Intents From Web Search Engine Queries. *Journal of the American Society for Information Science and Technology*, 63(9), 1773–1788.

- [22] Elswiler, D., Wilson, M. L., & Lunn, B. K. (2011). Understanding Casual-Leisure Information Behaviour. In A. Spink & J. Heinström (Hrsg.), *New Directions in Information Behaviour* (S. 211–241). Bingley: Emerald Group Publishing.
- [23] Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts And Technology Behind Search*. Harlow: Addison Wesley.
- [24] Stock, W. G. (2007). *Information Retrieval: Informationen suchen und finden*. München: Oldenbourg.
- [25] Carterette, B., Kanoulas, E., & Yilmaz, E. (2012). Evaluating Web Retrieval Effectiveness. In D. Lewandowski (Hrsg.), *Web Search Engine Research*. Bingley: Emerald Group Publishing Ltd.
- [26] Lewandowski, D. (2007). Mit welchen Kennzahlen lässt sich die Qualität von Suchmaschinen messen? *Die Macht Der Suchmaschinen / The Power of Search Engines*, 243–258.
- [27] Lewandowski, D., & Höchstötter, N. (2007). Qualitätsmessung bei Suchmaschinen – System- und nutzerbezogene Evaluationsmaße. *Informatik-Spektrum*, 30(3), 159–169.
- [28] Risvik, K. M., & Michelsen, R. (2002). Search engines and web dynamics. *Computer Networks*, 39(3), 289–302.
- [29] Vaughan, L., & Zhang, Y. (2007). Equal representation by search engines? A comparison of websites across countries and domains. *Journal of Computer-Mediated Communication*, 12(3), 888–909.
- [30] Lewandowski, D., Wahlig, H., & Meyer-Bautor, G. (2006). The freshness of web search engine databases. *Journal of Information Science*, 32(2), 131–148.
- [31] Lewandowski, D. (2008a). A three-year study on the freshness of Web search engine databases. *Journal of Information Science*, 34, 817–831.
- [32] Machill, M., Neuberger, C., Schweiger, W., & Wirth, W. (2003). *Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen*. In M. Machill & C. Welp (Hrsg.), *Wegweiser im Netz* (S. 13–490). Gütersloh: Bertelsmann Stiftung.
- [33] Höchstötter, N., & Koch, M. (2009). Standard parameters for searching behaviour in search engines and their empirical evaluation. *Journal of Information Science*, 35(1), 45–65.
- [34] Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, 42(1), 248–263.
- [35] Lewandowski, D. (2004). Date-restricted queries in web search engines. *Online Information Review*, 28(6), 420–427.
- [36] Lewandowski, D. (2008b). Problems with the use of web search engines to find results in foreign languages. *Online Information Review*, 32(5), 668–672.
- [37] Brunton, F. (2013). *Spam: A shadow history of the internet*. Cambridge, Mass., London: MIT Press.