

## How to do ... ?

# Konzeption einer Spezialsuchmaschine für Tutorials

Sebastian Sütkler und Friederike Kerkmann

### Hintergrund

Das Department Information an der Hochschule für Angewandte Wissenschaften (HAW) Hamburg bildet in zwei Bachelor-Studiengängen und einem Master-Studiengang künftige Informationsexperten aus. Das angebotene Themenspektrum reicht von Methoden der Wissensorganisation über Informationstechnologie und Bibliotheksmanagement bis hin zu Medienforschung und Informationsökonomie. Um der Komplexität der Berufspraxis später mit geschultem Problemlösungsverhalten und Teamfähigkeit begegnen zu können, nehmen alle Studierenden im fünften Semester an einem interdisziplinären Studienprojekt ihrer Wahl teil. Durch zeitkritisches und aufgabenbezogenes Arbeiten werden so die im späteren Berufsalltag bestehenden Arbeitsbedingungen simuliert, kooperatives Verhalten und arbeitsteilige Organisationsformen erprobt und soziale wie fachliche Kompetenzen geschult. Die Studienprojekte haben einen Umfang von 12 Semesterwochenstunden und einen Workload von 540 Stunden, sodass es in diesem Rahmen möglich ist, auch verhältnismäßig umfangreiche und zeitintensive Projekte durchzuführen (Modulhandbuch, 2013).

Das beschriebene Vorhaben fand im Wintersemester 2014/15 als ein solches studiengangübergreifendes Projekt der beiden BA-Studiengänge *Bibliotheks- und Informationsmanagement* und *Medien und Information* mit 16 Studierenden sowie zwei Projektleitern statt.

Das übergeordnete Ziel des Projektes bestand darin, eine hochschuleigene Suchmaschine zu konzipieren und diese ausgehend von einer Ideenskizze über ein theoretisches Konzept hin zu einem lauffähigen Alpha-Prototyp umzusetzen, der die theoretischen Überlegungen in Ansätzen praktisch nutzbar machen sollte. Dafür waren folgende Arbeitsschritte notwendig:

0. Abgrenzung des Geltungsbereiches: Welche Art von Suchmaschine soll erarbeitet werden, für welche Domäne soll sie gelten und welche Zielgruppe ansprechen?
1. Definition von Crawler und Index: Welche Quellen sollen nach welchen Kriterien in den Index aufge-

nommen werden, wie werden diese erschlossen und was passiert mit unerwünschten Inhalten?

2. Definition von Rankingfaktoren: Nach welchen Rankingkriterien sollen die Ergebnisse in eine Rangfolge gebracht werden, wie werden die unterschiedlichen Faktoren gewichtet und welche missbräuchlichen Techniken gilt es ggf. im Blick zu behalten?
3. Definition von Suchfunktionen: Welche aktiven und passiven Suchfunktionen und Filter soll die Suchmaschine unterstützen und welche Nutzungsdaten müssen dafür ggf. erfasst werden?
4. Gestaltung der Benutzerschnittstelle: Wie soll die Benutzerschnittstelle der Suchmaschine aussehen und welche Anforderungen an Ästhetik, Usability und Barrierefreiheit gilt es für ein optimales Nutzererlebnis zu erfüllen?
5. Prototypische Umsetzung: Auf Basis welcher Suchmaschinentechnologie kann die Umsetzung erfolgen? Welche technischen Besonderheiten sind ggf. speziell für die Tutorialsuche zu beachten?

Die Abgrenzung des Anwendungsbereiches der Suchmaschine erfolgte ebenso wie die Entwicklung eines passenden Namens und Logos gemeinsam mit dem gesamten Projektteam im Rahmen eines Kreativitätsprozesses. Auf diese Weise sollte eine möglichst hohe Identifikation aller Beteiligten mit dem eigenen Produkt sichergestellt und das gegenseitige Kennenlernen untereinander angestoßen werden.

Die inhaltliche Bearbeitung der Arbeitspakete (1) Crawler/Index, (2) Ranking, (3) Suchfunktionen und (4) Benutzerschnittstelle dagegen fand in Kleingruppen mit jeweils vier Gruppenmitgliedern statt. Zugrunde lag die Erfahrung, dass kleine Teams effektiver arbeiten als große, da sie weniger Zeit mit internen Abstimmungen verbringen und Erfolge besser teilen können<sup>1</sup>. Methodisch kamen Literaturanalysen und die Recherche von Best-Practice-Beispielen zur

<sup>1</sup> Jeff Bezos, Gründer von Amazon, prägte in diesem Zusammenhang die „2-Pizzen-Regel“, die besagt, dass ein Team klein genug sein sollte, um von zwei Pizzen satt zu werden (Schmidt & Rosenberg, 2015).

Anwendung. Die Realisierung des Prototyps übernahm einer der Projektleiter, der über eine Informatik-Ausbildung und damit das erforderliche technische Wissen verfügt.

Der folgende Beitrag folgt der Struktur der Arbeitspakete und beschreibt zunächst die Definition des Geltungsbereichs der Suchmaschine, die Rahmenbedingungen von Crawler und Index, die Rankingfaktoren und Suchfunktionen ebenso wie die gestalterischen Aspekte der Benutzerschnittstelle. Im Anschluss wird die praktische Umsetzung in Form eines Prototyps erläutert und beschrieben, welche Lernerfolge die Studierenden aus diesem Projekt ziehen konnten. Die Projektbeschreibung endet mit einem Ausblick, der darstellt, welche künftigen Schritte und Entwicklungen angestrebt werden.

### Abgrenzung des Geltungsbereiches

Abhängig von Anwendungszweck und Funktionalitäten lassen sich Websuchmaschinen im Allgemeinen in vier Arten differenzieren (Lewandowski, 2015):

- (1) Universalsuchmaschinen, die alle frei zugänglichen Inhalte im Web crawlen und indexieren und allgemeine Suchfunktionen anbieten, um diese zu durchsuchen.
- (2) Archivsuchmaschinen, die der Langzeitarchivierung freier Webinhalte dienen und dafür Momentaufnahmen von Webseiten zu definierten Zeitpunkten sichern und durchsuchbar machen.
- (3) Metasuchmaschinen, die keinen eigenen Index betreiben, sondern eine Suche über die Indexe verschiedener anderer Suchmaschinen anbieten.
- (4) Spezialsuchmaschinen, die die Suche zu einer bestimmten Domäne ermöglichen und sich dabei auf einen bestimmten Ausschnitt des Web beschränken.

Angesichts der spezifischen Herausforderungen und technischen Erfordernisse von Universal-, Meta- und Archivsuchmaschine fiel die Entscheidung auf die Konzeption einer Spezialsuchmaschine. Im Vergleich zu einer Universal- oder Archivsuchmaschine wird zum Aufbau der Datenbasis einer Spezialsuchmaschine nur ein eng abgegrenzter Bereich im Web betrachtet. Technische und finanzielle Hürden, die bei Aufbau und Pflege eines universellen Index auftreten, entfallen somit. Zudem wendet sich eine Spezialsuchmaschine an eine bestimmte Zielgruppe, so dass Crawler, Index, Ranking, Suchfunktionen und Gestaltung sich an einem konkreten Bezugsrahmen orientieren können und nicht den Anspruch erfüllen müssen, jeden Nutzer gleichermaßen zu bedienen. Im Rahmen eines Brainstormings konkretisierte sich

die Idee einer Spezialsuchmaschine für Tutorials<sup>2</sup> heraus. Inhalte sollten alle Arten von zugänglichen Tutorials, Anleitungen, Anweisungen und Antworten aus Frage-Antwort-Diensten sein und verschiedene Medienformen (Text, Bild, Audio, Video sowie Mischformen) umfassen. Berücksichtigung finden sollten sowohl professionell vertriebene Dokumente (z.B. Bedienungsanleitungen von Herstellern, Kochbücher, Fitness- und Sportanleitungen, Online-Kurse an Bildungseinrichtungen) als auch User Generated Content (UGC), also Inhalte, die von (Laien-)Nutzern bereitgestellt werden. Intention war es, ein Angebot zu schaffen, das bei spezifischen Anfragen mit Tutorial-Charakter gezielt auf thematisch passende Seiten zugreift und Informationsballast, wie er bei allgemeinen Suchen in Universalsuchmaschinen wie *Google* auftreten kann, zu vermeiden, ohne gleichzeitig auf eine bestimmte Medienform reduziert zu werden, wie bspw. bei einer Suche in *Youtube*.

Eine Marktsichtung ergab, dass kaum vergleichbare Angebote existieren. So gibt es zwar kollaborativ geführte Wikis wie *wikiHow*<sup>3</sup> oder redaktionell gepflegte Portale wie *eHow*<sup>4</sup>, die allerdings keine Tutorial-Inhalte aus dem Web crawlen und auf manueller Zusammenarbeit basieren. Zum Zeitpunkt der Projektdurchführung fand die Projektgruppe lediglich ein Beispiel für eine vergleichbare Spezialsuchmaschine für Tutorials mit dem Namen *Mr. How*<sup>5</sup>, die jedoch damals schon wenig gepflegt war und inzwischen offline ist.

Nach der Entscheidung für eine Tutorial-Spezialsuchmaschine bearbeitete das Team gemeinsam die Frage nach einem geeigneten Namen und passender Farbgebung. Dies geschah bewusst zu einem frühen Zeitpunkt im Projektverlauf, um während der Arbeitsphase eine hohe Identifikation mit dem Arbeitsgegenstand zu erzielen und „dem Kind einen Namen zu geben“. In verschiedenen Denk- und Diskussionsschleifen kristallisierte sich der Name *HowTo* heraus, der übersetzt in etwa der Phrase „Wie mache ich...?“ entspricht. Die Farbgebung des Logos wurde ebenso wie das Design der Benutzerschnittstelle an das Corporate Design der HAW Hamburg angepasst, um auch eine optische Verbindung zur Hochschule herzustellen.

### Definition von Crawler und Index

Crawler und Index bilden die Basis jeder Suchma-

2 Ursprünglich bezeichnete ein Tutorial die Dokumentation eines Softwaresystems für den Endnutzer, in der im Sinne einer Einführung die wichtigsten Systemfunktionen erläutert werden (Gabler Wirtschaftslexikon, o.J.). Inzwischen umfasst der Begriff auch Anleitungen anderer Art, wie Do-it-yourself-Anleitungen für handwerkliche Tätigkeiten, Kosmetiktips oder Kochrezepte.

3 <http://www.wikihow.com>

4 <http://www.ehow.com>

5 [www.mrhow.net](http://www.mrhow.net)

schine, da diese beiden Komponenten für die Dokumentenbeschaffung und Aufbereitung der Inhalte aus dem Web verantwortlich sind. Gemeinsam bilden sie ein Abbild eines Teils des Internets, das mit Hilfe der Suchmaschine recherchierbar wird.

Der Crawler ist ein Tool, das nach festgelegten Regeln das Web durchsucht, um für die Suchmaschine relevante Inhalte aufzufinden, herunterzuladen und eventuell weitere Links auf den gesichteten Seiten zu verfolgen. Ein Crawler beginnt seine Arbeit dabei mit dem sog. Seed Set, das eine Liste von relevanten Links als Ausgangspunkt enthält. Für die Entwicklung von *HowTo* wurden für das Seed Set Webseiten zusammengestellt, die Tutorial- und Do-it-yourself-Inhalte, Anleitungen und Tutorial-Videos anbieten. Der Crawler für *HowTo* ist als Focused Crawler<sup>6</sup> mit einem Best-First-Ansatz<sup>7</sup> konzipiert, damit beim Crawling ausschließlich Webseiten mit relevanten Tutorial-Inhalten berücksichtigt werden. Zusätzlich sorgen sog. Crawling Policies dafür, dass nur indexierbare Dokumente aufgenommen werden. Solche Regeln sind auch für den Ausschluss von unerwünschten Inhalten notwendig. Dafür sollen sowohl die *robots.txt* als auch das Metatag *nofollow* berücksichtigt werden, durch die Betreiber reglementieren können, welche Inhalte ihres Angebots gecrawlt werden dürfen.

Der Crawler muss sowohl mit traditionellen statischen HTML-Seiten als auch mit Inhalten von dynamischen Webanwendungen umgehen können. Da Tutorials im Web als Text, Video und Bild vorliegen können, soll der Crawler auch zwischen diesen Medientypen differenzieren und sie dem jeweils passenden Index zuordnen können.

Um die gecrawlten Webseiten durch Suchanfragen wiederauffindbar zu machen, müssen diese anschließend mit linguistischen und textstatistischen Verfahren erschlossen und indexiert werden. Bei der Suchmaschine *HowTo* erfolgen dazu eine Volltextindexierung und ein Stemming der Terme. Da der deutschen Sprache eine sehr komplexe Grammatik zugrunde liegt und sie viele Komposita enthält, soll für die Indexierung ein Wörterbuch hinterlegt werden, das die Grundformen, Flexionen und Ableitungen enthält. Als Indexate sollen die Substantive genutzt werden, da diese elementar für die Bedeutung des Textes sind. Für die Erfassung von Multimedia-Inhalten werden die textuellen

Beschreibungen um diese Inhalte herum indexiert und Beschreibungssprachen wie MPEG7<sup>8</sup> eingesetzt.

### Definition von Rankingfaktoren

Damit Benutzer von Suchmaschinen die besten Ergebnisse für ihre Suchanfragen erhalten, nutzen Suchmaschinen Rankingfaktoren, die die Reihenfolge der Suchtreffer zu der Anfrage regeln und dafür sorgen sollen, dass die Dokumente mit der höchsten Relevanz auch auf den vorderen Platzierungen in der Trefferliste auftauchen. Eine Spezialsuchmaschine wie *HowTo* benötigt spezielle Rankingfaktoren, die sich von den Faktoren, die allgemeine Websuchmaschinen einsetzen, unterscheiden. Bei Rankingfaktoren werden sog. Onpage- und Offpage-Faktoren differenziert. Onpage-Faktoren beziehen sich dabei auf die Inhalte und Informationsarchitektur der Webseite selbst, während Offpage-Faktoren in Beziehung stehen mit der Linkpopularität der Webseite im Web, Social Signals, Brands und User Signals.

Bei der Tutorialsuche geht es vor allem um die Inhalte und deren Nützlichkeit für die Anwender. Es kommt weniger darauf an, ob ein Tutorial eine hohe Linkpopularität aufweist. Deshalb werden für *HowTo* das Vorkommen und die Formatierung von Keywords, die das Dokument repräsentieren, am höchsten beim Ranking gewichtet. Die zweitwichtigste Gruppe von Rankingfaktoren bezieht sich auf die Inhalte des Tutorials. Es wird die Multimedialität des Tutorials bewertet. Werden bei dem Tutorial neben textuellen Inhalten auch Bilder, Videos oder Tonaufnahmen angeboten, wird dieses im Ranking der Suchergebnisse bevorzugt behandelt. Die letzte hochgewichtete Gruppe bezieht sich auf die Social Signals, die die Bewertung der Nützlichkeit der Tutorials durch die Nutzer betreffen. Social Signals sind z.B. Likes auf *Facebook*, Kommentare zu den Tutorials, Erwähnungen auf *Twitter* etc., die auf die Verbreitung der Anleitungen in den sozialen Medien hinweisen und damit einen hohen Nutzen implizieren. Beim Ranking von allgemeinen Websuchmaschinen wie *Google* spielen Social Signals dagegen keine besondere Rolle.

Abbildung 1 zeigt die Einzelgewichtungen der identifizierten Rankingfaktoren für *HowTo*:

Damit Seiten, die keinen Mehrwert bieten, oder durch den Einsatz von sog. Black-Hat-Techniken<sup>9</sup> manipulierte Seiten nicht unverdient auf den vorderen Plätzen positioniert werden, sind für *HowTo* auch missbräuchliche Faktoren definiert worden, die zu einer Abstrafung im Ranking führen sollen. Diese

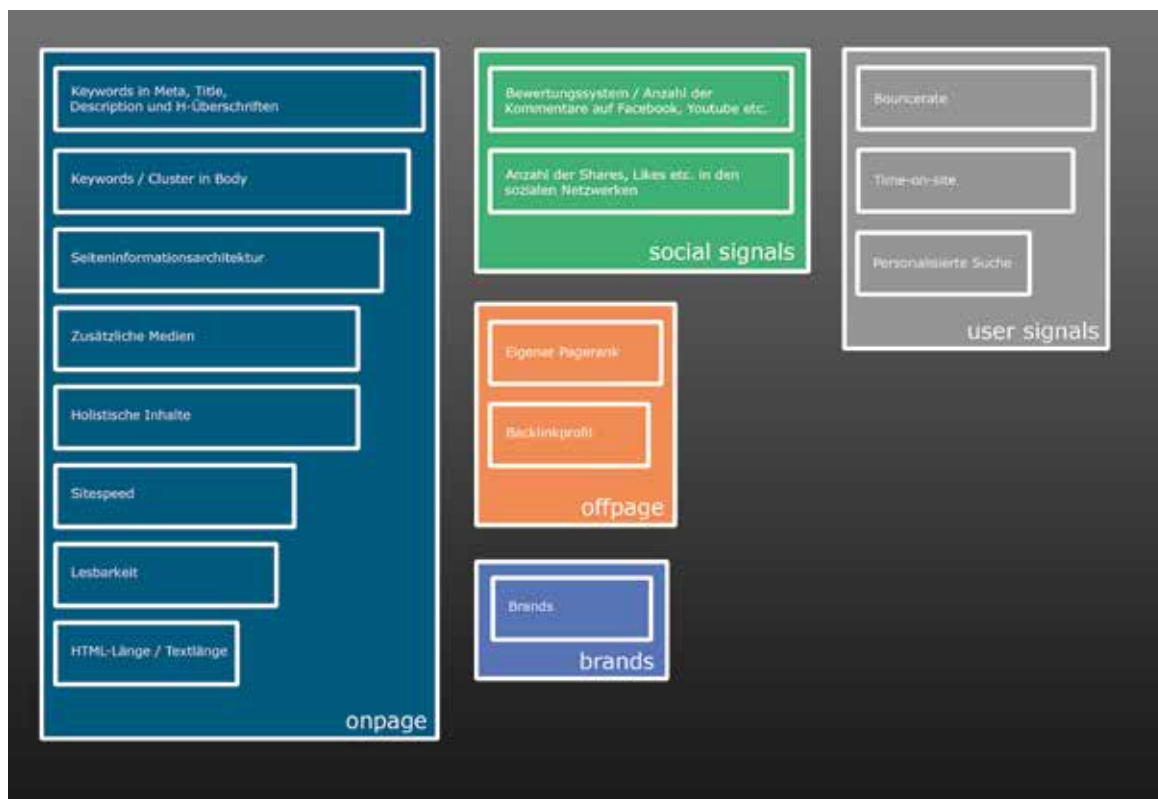
6 Focused Crawler enthalten definierte Konfigurationen, die sicherstellen sollen, dass nur Webseiten indexiert werden, die thematisch erwünscht sind. Dafür werden relevante Webseiten für das Thema durch die Nutzbarmachung der Graphenstruktur im Web selektiert (Batsakis et al., 2009).

7 Best-First-Ansätze werden dafür genutzt, die gefundenen URLs beim Crawling nach Relevanz zu sortieren, um damit die Priorisierung der Webseiten zu definieren, d. h. festzulegen, welche Webseiten vorrangig durch den Crawler abgearbeitet werden (Menczer et al., 2011).

8 <http://mpeg.chiariglione.org/standards/mpeg-7>

9 Eine Vorgehensweise der Suchmaschinenoptimierung, bei der Verstöße gegen Webmaster-Richtlinien bewusst in Kauf genommen werden.

Abb. 1: Einzelgewichtungen der Rankingfaktoren für HowTo



Rankingalgorithmen sollen prüfen, ob auf den Seiten im Index Methoden wie *Keyword-Stuffing*<sup>10</sup>, *Google Blowing*<sup>11</sup> oder der Einsatz von *Link Farms*<sup>12</sup> angewendet wurden, um dieses Vorgehen ggf. durch eine Abstufung in der Trefferliste ahnden zu können.

### Definition von Suchfunktionen und Filtern

Suchfunktionen haben die Aufgabe, den Nutzer von seiner noch nicht ausformulierten Suchanfrage im Kopf hin zum gewünschten Ergebnis zu leiten; und dies trotz vielfach sprachlich wenig elaborierten und in der Regel kurzgefassten Suchanfragen.

Will man für einen kleinen Bereich im Web ein spezielles Angebot schaffen, das eine Alternative zu den gängigen großen Suchmaschinen darstellt, so muss diese dem Nutzer spürbare Vorteile bringen und sich positiv von den größeren Konkurrenten abheben. Dabei bieten Spezialsuchmaschinen den Vorteil, dass sie sich auf eine bestimmte Domäne und einen begrenzten Anwenderkreis beziehen und dafür gezielt entsprechende Suchfunktionen anbieten können.

10 Keyword-Stuffing bezeichnet das unnatürlich hohe Vorkommen von Keywords auf einer Website.

11 Bei dieser Technik bearbeiten Mitbewerber die Seiten eines Konkurrenten mit Negativ-Ranking-Kriterien, wie z.B. dem Setzen hunderter Backlinks aus einer schlechten Nachbarschaft, so dass die Seite Links von einer Webseite erhält, die von Google negativ eingestuft wird, und damit selbst eine Abstrafung erhält.

12 Link Farms sind Webseiten, auf denen eine übermäßige Zahl von Outlinks gesetzt wird. Es gilt die ungeschriebene Regel, dass Google ab 100 Links pro Seite aufhört, diese zu verfolgen und dass die Seiten zudem abgewertet werden.

Zu den passiven Suchfunktionen zählt die Auswertung von Suchsessions. Mit Hilfe dieser Auswertung sollen zusätzlich zu den entwickelten Rankingfaktoren weitere Aspekte, wie implizites *Relevance Feedback* Berücksichtigung finden und sich dabei innerhalb der Sessions eines Nutzers positiv auf das Sucherlebnis auswirken. Für *HowTo* soll dabei sichergestellt sein, dass aus datenschutzrechtlicher Sicht nur die notwendigsten Daten erfasst und ausgewertet werden. Empfohlen wird daher für *HowTo* die Analyse auf kollektiver Ebene (d.h. die gesamte Nutzerschaft) und nicht auf Basis des einzelnen, identifizierbaren Nutzers. Bei der Interpretation von Suchanfragen soll als einziger Suchoperator die Phrasensuche unterstützt werden, da das explizite Angebot von booleschen Operatoren im Umgang mit der Websuche kaum eine Rolle spielt. Daneben sollen auch Wortstammsuche, Trunkierung, Synonymsuche und der Ausschluss von Stoppwörtern unterstützt werden.

Zu den aktiven Suchfiltern und Funktionen zählt das Angebot einer Facettensuche, die gefundene Tutorials nach Themen und Medienart klassifiziert. Mit Hilfe der Facetten kann der Nutzer dann die Suche weiter auf seine Bedürfnisse eingrenzen und so Tutorials für den gewünschten Bereich finden. Für *HowTo* ist dabei das Two-Stage-Modell angedacht, bei dem Themen in Ober- und dazugehörigen Unterkategorien eingeteilt werden. Abbildung 2 zeigt einen Ausschnitt möglicher Facetten für die Tutorialsuchmaschine.

Neben den thematischen Facetten sind darüber hin-

aus weitere Filter definiert, die sich auf Klickpopularität, Herkunft, Aktualität und Schwierigkeitsgrad der Tutorials beziehen. Zusätzlich sollen weitere Filter angeboten werden, die eine Recherche in anderen Dokumentensammlungen ermöglichen. Sie bieten dem Nutzer die Möglichkeit, vor und nach einer Suche einzugrenzen, ob er Text, Bilder oder Videos durchsuchen möchte.

### Gestaltung der Benutzerschnittstelle

Die Benutzerschnittstelle stellt den sicht- und bedienbaren Bereich der Suchmaschine für den Nutzer dar. Um eine hohe User Experience zu gewährleisten ist es wichtig, gängige Basisanforderungen an Ästhetik, Usability und Barrierefreiheit zu erfüllen. Bei der Usability für *HowTo* wurden insbesondere drei Usability-Grundsätze identifiziert, die für eine gute Nutzbarkeit sorgen sollen:

- Kohärenz und Konsistenz: Die Website von *HowTo* ist angemessen aufgebaut (Kohärenz) und alle Elemente sind stimmig gestaltet (Konsistenz).
- Schnelle Erschließbarkeit: Der Nutzer von *HowTo* erkennt auf einen Blick Zweck und Aufbau der Suchmaschine.
- Lesbarkeit: Der Kontrast zwischen Hintergrund und Schrift ist möglichst hoch. Sowohl die Schriften als auch die Symbole besitzen einen angenehmen Heligkeitskontrast.

Zur Sicherstellung der Barrierefreiheit werden eine Funktion zur Anpassung der Schriftgröße und Alternativtexte für alle grafischen Elemente angeboten, die durch Screenreader ausgelesen werden können. Zudem wird sichergestellt, dass die Suchmaschine auch mit alternativen Eingabegeräten bzw. der Tastatur bedienbar ist.

Für das Interface von *HowTo* wurde das Quasi-Standard-Layout von etablierten Websuchmaschinen wie *Google*, *Bing* oder *DuckDuckGo* adaptiert. Die Startseite der Suchmaschine zeigt wie üblich einen prominent platzierten Suchschlitz mit Button (siehe Abb. 3), um dem Nutzer einen vertrauten Sucheinstieg zu bieten. Für die Farbgebung der Seite wurde das Corporate Design der HAW Hamburg als Ausgangsbasis genutzt.

Die Suchergebnisseite orientiert sich ebenfalls an den Standards in der Ergebnispräsentation von Suchmaschinen und richtet sich nach der Leserichtung in unserem Kulturkreis. Damit eine rasche Orientierung gewährleistet ist, werden die einzelnen Suchergebnisse untereinander in der Listendarstellung gezeigt. Die Sortierung ist unmittelbar erkennbar. Der einzelne Treffer wiederum besteht aus den gängigen Ele-

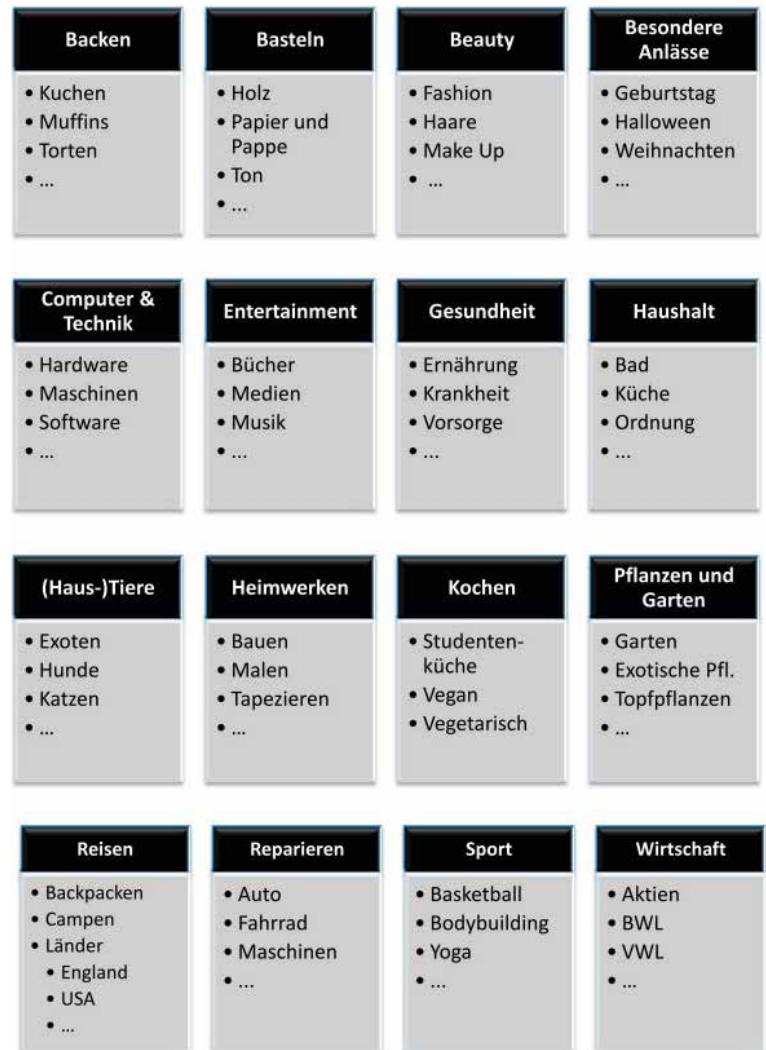


Abb. 2: Mögliche Facetten für HowTo



Abb. 3: Screenshot der Startseite von HowTo

menten Dokumenttitel, Kurzbeschreibung und URL, wobei in der Kurzbeschreibung die Suchbegriffe der Suchanfrage hervorgehoben werden.

### Prototyp

Das erarbeitete Konzept und die Layoutentwürfe bildeten die theoretische Grundlage der notwendigen Suchmaschinen-Komponenten und Ausgangspunkt der praktischen Umsetzung. Für die Realisierung des

Prototyps wurde auf die Suchmaschinentechologie *ElasticSearch*<sup>13</sup> zurückgegriffen, einer Suchmaschine auf Basis von *Lucene*, die unter der *Apache License*<sup>14</sup> vertrieben wird. Die aktuelle Version von *HowTo* ist online unter [howtosearch.de](http://howtosearch.de) verfügbar. Dabei handelt es sich um eine beispielhafte Umsetzung zur Veranschaulichung der Konzeptidee. Funktionalitäten für eine einfache Suche innerhalb der Textdokumente, Suchvorschläge während der Eingabe und eine Eingrenzung durch Facetten sind bereits nutzbar. An einer Ausweitung auf die Suche innerhalb anderer

zu erfahren und zu erproben. Denkbar sind bspw. Seminare, die sich intensiver mit den einzelnen Komponenten befassen, Veranstaltungen zur Suchmaschinenoptimierung oder Usability-Evaluationen zur Verbesserung der Gebrauchstauglichkeit.

### Lernerfolge für die Studierenden

Bis auf eigene Nutzungserfahrungen, theoretische Kenntnisse zu Information Retrieval-Systemen und zum Teil praktische Kenntnisse der Suchmaschinenoptimierung brachten die teilnehmenden Stu-

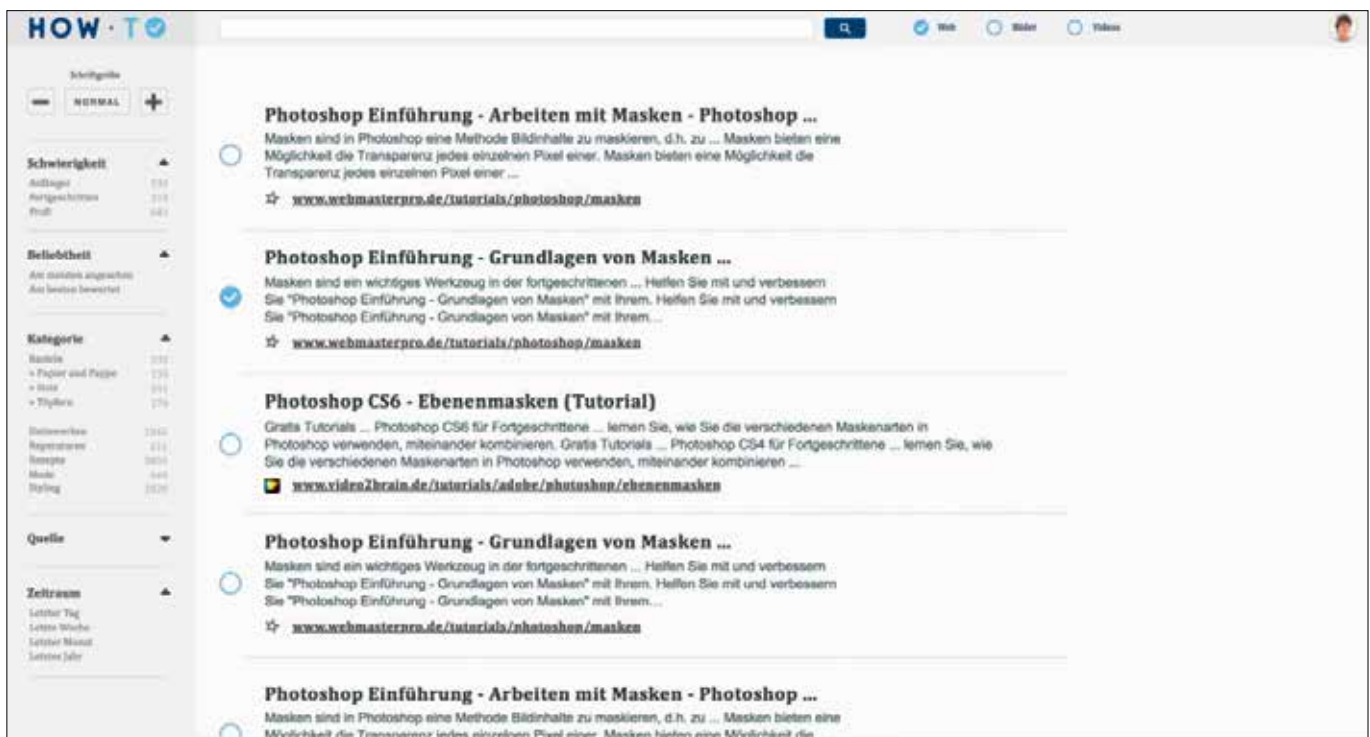


Abb. 4: Screenshot der Suchergebnisseite von HowTo

Dokumentenkollektionen wie Bilder und Videos wird derzeit gearbeitet. Im Hinblick auf die Qualität des Index und des Rankings kann *HowTo* zwar (noch) nicht mit den etablierten Suchmaschinen mithalten; dies stand jedoch auch nicht im Fokus des Projektes. Ziel war es nicht, eine ernsthafte Konkurrenz zu anderen Anbietern aufzubauen, sondern vielmehr die einzelnen beteiligten Komponenten für die Studierenden unmittelbar erfahrbar zu machen.

Unabhängig des reduzierten Funktionsumfangs und der noch wenig marktfähigen Leistung im Index und Ranking, ist es das Ziel, *HowTo* langfristig in das Lehrangebot am Department Information aufzunehmen. Studierende sollen so die Möglichkeit erhalten, den praktischen Einsatz von Suchmaschinentechologie, SEO, Retrievaleffektivität und anderen Qualitätsparametern von Suchmaschinen „am lebenden Objekt“

dierenden zu Beginn des Projektes nur wenig Vorerfahrung im Bereich Suchmaschinen mit. Vor diesem Hintergrund hat sich jede/r von ihnen auf fachlicher Ebene im Rahmen des Projektes ein solides Grundwissen über Suchmaschinentechologie, die einzelnen Komponenten sowie ihr Zusammenwirken aneignen können. Dies ist insofern berufsrelevant, da IT-(Entwicklungs-)Projekte und die Nutzung solcher Systeme in Bibliotheken und Informationseinrichtungen zum Tagesgeschäft gehören und Suchmaschinentechnologien bzw. allgemeine Information-Retrieval-Systeme in vielen Fällen die Basis ihrer Anwendungen darstellen.

Das Projekt hat den Studierenden damit einerseits Projektmanagement Erfahrung in ihrer Schnittstellenfunktion zwischen Endanwender und Programmierer ermöglicht und andererseits die für die Praxis elementaren Kenntnisse der zugrundeliegenden IT vermittelt. Durch die Gruppenstruktur mit vier Kleingruppen ergab sich zudem ein hoher Austausch- und Abstim-

<sup>13</sup> <https://www.elastic.co/products/elasticsearch>

<sup>14</sup> <http://www.apache.org/licenses/LICENSE-2.0>

mungsbedarf, der auf sozialer Ebene Kommunikationsfähigkeit und Dialogbereitschaft förderte. Keine der Kleingruppen konnte separat und unabhängig von den anderen arbeiten – jede Komponente hatte Schnittstellen zu anderen Arbeitsbereichen, jede Kleingruppe musste fortlaufend auf die (Zwischen-) Ergebnisse der anderen zurückgreifen, um selbst weiterarbeiten zu können.

## Fazit und Ausblick

Das erarbeitete Konzept definiert die relevanten Komponenten Crawler/Index, Ranking, Suchfunktionen und Benutzerschnittstelle, jeweils mit konkretem Bezug auf den Anwendungsfall Spezialsuchmaschine für Tutorials. Der entwickelte Prototyp wiederum dient der beispielhaften Veranschaulichung dieser theoretischen Grundlagen und verdeutlicht das Zusammenspiel der einzelnen Bestandteile. Der Fokus lag und liegt auf einer Anwendung in Forschung und Lehre, eine Kommerzialisierung der Suchmaschine ist nicht angestrebt.

Künftiger bzw. fortlaufender praktischer Handlungsbedarf besteht insbesondere in Ausbau und Aktualisierung des Index sowie Optimierung des Rankings und der Benutzerschnittstelle. Als nächste konkrete Maßnahme ist der Entwurf eines responsiven Designs geplant, um *HowTo* auch geräteunabhängig nutzbar zu machen. Mit Hilfe von Nutzertests sollten zudem die Maßnahmen zur gebrauchstauglichen und barrierefreien Gestaltung der Desktop- sowie der Mobilversion überprüft und ggf. korrigiert werden.

Für die Weiterentwicklung von *HowTo* müssen neben den technischen Aspekten auch juristische Fragestellungen geklärt werden, wie Aspekte des Datenschutzes, des Urheberrechts oder des Serverstandortes und sich daraus ergebende Konsequenzen, damit eine öffentliche Zugänglichmachung der Suchmaschine abgesichert ist. **I**

## Danksagung

Wir bedanken uns bei den Auslobern<sup>15</sup> des Team Award Information Professionals (TIP) 2016 für die Auszeichnung. Das Preisgeld wird voraussichtlich für die Entwicklung eines mobilen Designs für *HowTo* eingesetzt werden. Darüber hinaus bedanken wir uns bei Inga Renkhoff und Julian Preußner, die uns bei der professionellen Gestaltung des Logos und des Interface unterstützt haben.

<sup>15</sup> b.i.t.online, Schweitzer Fachinformationen, ekz.bibliotheksservice, Konferenz der bibliotheks- und informationswissenschaftlichen Hochschulen und Ausbildungseinrichtungen (KIBA), Sektion 7 des dbv sowie die Ausbildungskommission der Deutschen Gesellschaft für Information und Wissen e.V.

## Projektteam

Büusra Alici, Agneta Bock, Saskia Boos, Ramona Brase, Miriam Dachs, Falko Fesser, Mareike Franke, Sabine Lohrke, Pia Lorenzen, Wiebke Lüth, Sara Eliza Melchior, Annemarie Müller, Melis Nar, Anita Raic, Sultana Razaq, Anne Stempel

## Literaturverzeichnis

- Batsakis, S., Petraksi, E. & Milios, E. (2009). Improving the Performance of Web Crawlers. *Data & Knowledge Engineering*, 68 (10). 1001–1013.
- Gabler Wirtschaftslexikon, Springer Gabler Verlag (Hrsg.). Stichwort: Tutorial. Online verfügbar unter: <http://wirtschaftslexikon.gabler.de/Archiv/77336/tutorial-v8.html>
- Lewandowski, D. (2015). *Suchmaschinen verstehen*. Berlin: Springer Vieweg.
- Menczer, F. (2011). Web Crawling. In: Liu, B. (Hrsg.), *Web data mining: exploring hyperlinks, content and usage data*. (S. 311–362). Berlin: Springer.
- Modulhandbuch (2013). *Modulhandbuch Bibliotheks- und Informationsmanagement – Bachelor of Arts*. Online verfügbar unter: [https://www.haw-hamburg.de/fileadmin/user\\_upload/DMI-I/Studium/Studiengaenge/BIM/BIM-2013ModulHandbuch4-1.pdf](https://www.haw-hamburg.de/fileadmin/user_upload/DMI-I/Studium/Studiengaenge/BIM/BIM-2013ModulHandbuch4-1.pdf)
- Projektbericht WS 2014/15 (2015). *HowTo – Konzeption einer HAW-eigenen Spezialsuchmaschine für Tutorials*. Hrsg. von Sebastian Sünkler & Friederike Kerkmann. Auch online verfügbar unter: [http://searchstudies.org/HowTo\\_Bericht.pdf](http://searchstudies.org/HowTo_Bericht.pdf)
- Schmidt, E. & Rosenberg, J (2015). *Wie Google tickt*. Frankfurt/New York: Campus Verlag.



**Sebastian Sünkler und  
Friederike Kerkmann**

Hochschule für Angewandte  
Wissenschaften Hamburg  
Fakultät Design, Medien und  
Information



Department Information

sebastian.suenkler@haw-hamburg.de  
friederike.kerkmann@haw-hamburg.de