

Relevance Assessment Tool

Ein Werkzeug zum Design von Retrievaltests sowie zur weitgehend automatisierten Erfassung, Aufbereitung und Auswertung von Daten

Dirk Lewandowski und Sebastian Sünkler

In diesem Aufsatz stellen wir ein Werkzeug vor, welches eine effiziente Durchführung von Tests zur Retrievaleffektivität von Web-Suchmaschinen ermöglicht. Das Tool besteht aus den Komponenten Suchmaschinenscraper, Verwaltung und Testdesign, Nutzerinterface für die Durchführung der Tests mit Probanden sowie Auswertung. Die Komponenten werden beschrieben, wobei insbesondere auf das Scraping, also die automatische Extraktion von Daten aus den Ergebnisseiten der Suchmaschinen, eingegangen wird. Das Tool ist in der Lage, die Ergebnisseiten der allgemeinen Web-Suchmaschinen Bing, Google und Yahoo zu erfassen. Weitere Scrapper wurden u.a. für die Produktsuche (Google Shopping, Amazon, Ebay) entwickelt.

1 Einleitung

Die Evaluierung von Web-Suchmaschinen lässt sich in vier Bereiche unterteilen [Lewandowski & Höchstötter 2007]: Qualität des Index, Qualität der Suchresultate, Qualität spezieller Suchfunktionen sowie Nutzbarkeit der Suchmaschine (also Benutzerführung und Benutzerfreundlichkeit). Besonders zur Qualität der Suchresultate, hier insbesondere zur Retrievaleffektivität der Suchmaschinen, liegen zahlreiche Studien vor (u.a. [Griesbaum 2004; Lewandowski 2008; Tawileh, Mandl & Griesbaum 2010]). Neben dem Ziel des Vergleichs ist die Kenntnis der Qualität der Treffer der verschiedenen Web-Suchmaschinen vor allem angesichts ihrer großen Bedeutung für den Wissenserwerb sowohl im Freizeitbereich als auch im professionellen Kontext wichtig. In dieser Hinsicht wäre es wünschenswert, dass solche Tests nicht nur cursorisch, sondern regelmäßig durchgeführt werden. Dem gegenüber steht jedoch der hohe Aufwand, der vor allem mit der Datenerhebung und -aufbereitung für solche Tests verbunden ist.

Um diesen Aufwand zu verringern, haben wir ein Werkzeug entwickelt, welches es einerseits ermöglicht, Tests zur Retrievaleffektivität von Web-Suchmaschinen unkompliziert zu erstellen, und andererseits, die Datenerfassung und -aufbereitung so weit wie möglich zu automatisieren. Tests, die mit dem Tool aufgesetzt werden, können dabei so frei gestaltet werden, dass verschiedene Varianten von Testszenarien abgedeckt werden. So kann beispielsweise festgelegt werden, welche Testobjekte einzeln oder in Kombination durch Probanden bewertet werden sollen. Testobjekte können dabei die Trefferbeschreibungen, die auf den Suchergebnisseiten durch die Suchmaschinen angezeigt werden als auch die Ergebnisse selbst sein.

Das entwickelte Tool besteht aus den folgenden Komponenten:

Suchmaschinenscraper

Backend zur Verwaltung der Projekte, der Projektnutzer, der Projektadministratoren und der Suchaufgaben

Frontend zur Durchführung der Tests (Relevanzbewertung durch Nutzer)

Auswertung

Die Komponenten 1 bis 3 sind bereits fertig gestellt, das Auswertungsmodul befindet sich zurzeit in Arbeit. Wir werden in diesem Aufsatz die Module kurz vorstellen und dann vor allem auf die Komponenten 1 und 2 eingehen. Komponente 1 ist technisch am anspruchsvollsten und wir schätzen unsere dort gemachten Erfahrungen als besonders hilfreich für die Entwicklung ähnlicher Anwendungen ein. Daher werden wir in einem eigenen Abschnitt auch näher auf die Probleme eingehen, die sich bei der Programmierung dieser Komponente gezeigt haben.

2 Verwandte Arbeiten

Aus den durchgeführten Studien zur Retrievaleffektivität von Web-Suchmaschinen (Übersicht in [Lewandowski 2008]) lässt sich kein klares Bild gewinnen, mit welcher technischen Unterstützung diese durchgeführt wurden. Zwar wird in manchen Studien erwähnt, dass die Relevanzbewertungen der Probanden mit Hilfe eines Browsertools erfasst wurden, in einem Großteil der Studien wird jedoch auf die Frage der technischen Unterstützung gar nicht eingegangen. Falls dies doch der Fall ist, so bleiben die Angaben vage und erlauben keine Ableitung von Empfehlungen für das Design solcher Tools. In keiner der bekannten Studien wurde erwähnt, dass die zu bewertenden Ergebnisse und/oder Trefferbeschreibungen automatisiert erhoben wurden.

Eine Marktsichtung hat gezeigt, dass ein Tool, welches die beschriebenen Aufgaben erfüllt, bisher nicht entwickelt wurde. Online-Werkzeuge, die eine automatisierte Erfassung von Suchmaschinenergebnissen (sog. *scraping*) bieten, verfolgen andere Ziele als das Relevance Assessment Tool. Beispielsweise werden diese Anwendungen im Bereich der Suchmaschinenoptimierung eingesetzt [Stokdyk 2009] oder zur Überprüfung von potenziellen Sicherheitslücken in Webseiten genutzt [Peteron 2006]. Daneben gibt es noch Suchmaschinenscraper zum Harvesting von Ergebnisseiten, die aber keine Beurteilung dieser Seiten durch Probanden vorsehen. In der Studie von [Höchstötter & Lewandowski 2009] wurde ein Suchmaschinenscraper eingesetzt, der die Ergebnisseiten der US-Versionen der gängigen Suchmaschinen auf den Einsatz unterschiedlicher Trefferquellen und -darstellungen (wie Video, Bilder, Nachrichten, usw.) hin auswertet. Da hier der Fokus auf der Auszählung der beschriebenen Elemente lag, wurden keine weitergehenden Informationen (wie beispielsweise die Trefferbeschreibungen) erfasst.

Eine Alternative zum Scraping ist die Abfrage von Suchmaschinenergebnissen mit Hilfe der von den Suchmaschinen angebotenen Application Programming Interfaces (APIs) [Tosques & Mayr 2009]. Dieser alternative Ansatz zur Gewinnung von Suchergebnissen wurde aus verschiedenen Gründen nicht gewählt. Zum einen bietet nicht jede Websuchmaschine eine API an. Zum anderen wird in der Regel bei den APIs ein anderer Suchmaschinenindex

bereitgestellt, d.h. die zurückgegebenen Ergebnisse unterscheiden sich von denjenigen, die bei einer manuellen Recherche angezeigt werden. Ein weiterer praktischer Grund liegt darin, dass die APIs sehr unterschiedlich sind und daher für den Suchmaschinen-scraping immer anders verarbeitet werden müssten, während bei einem Scraping in der Regel nur Unterschiede bei den Formatierungstags der HTML-Elemente in dem Quelltext der Trefferlisten unterschiedlicher Suchmaschinen bestehen.

3 Technische Voraussetzungen

Das Relevance Assessment Tool soll als Onlinewerkzeug einfach über jeden herkömmlichen Webbrowser aufgerufen werden können. Um diesen Zugang zu garantieren, muss es in einer Serverumgebung installiert werden, die die serverseitige Programmiersprache PHP in der Version 5.3 unterstützt und daneben noch das relationale Datenbankmanagementsystem MySQL in der Version 5.1 bereitstellt.

PHP wurde als Programmiersprache gewählt, weil sie zum einen kostenlos als Open-Source-Sprache verfügbar ist und zum anderen, weil sie einige passende Funktionen und Programm-bibliotheken für die nötigen Aufgaben eines Relevanztools mitbringt. Ein weiterer wichtiger Grund für die Wahl von PHP bestand in der Anbindung zum relationalen Datenbankmanagementsystem MySQL, das im Relevance Assessment Tool für die Speicherung der Daten genutzt wird. Als Serverumgebung für die lokale Entwicklung des Tools wurde auf das kostenlose Gesamtpaket XAMPP zurückgegriffen, das als Komplettinstallation den Webserver Apache, das Datenbank-Managementssystem MySQL, die Programmiersprache PHP und für die Datenbankverwaltung das Online-Werkzeug phpMyAdmin enthält. Für eine Nutzung auf einem anderem Webspace oder Server ist es notwendig die lokal entwickelten Komponenten und die Datenbank dorthin zu übertragen. Das Tool wurde auf dementsprechenden Speicherplatz bei einem Webhoster für die öffentliche Nutzung eingerichtet.

Alle Daten, die mit dem Relevanztool erhoben werden oder bei der Administration der Projekte anfallen, werden wie oben erwähnt in einer MySQL-Datenbank gespeichert. Insgesamt setzt sich die Datenbank aus 15 Tabellen zusammen, in denen die Projektdaten und Testdaten gespeichert werden. Ein sehr großer Vorteil bei der Nutzung einer MySQL-Datenbank für das Relevance Tool liegt darin, dass durch einfache SQL-Statements Auswertungen der Testergebnisse vorgenommen werden können, die für die spätere Entwicklung eines Auswertungsmodul nützlich sind.

4 Komponenten des Relevance Assessment Tools

Das Relevance Assessment Tool setzt sich insgesamt aus vier Komponenten mit den dazugehörigen Modulen zusammen. Dazu gehören ein Suchmaschinen-scraping, der auch als Standalone-Anwendung für andere Anwendungszwecke zur Verfügung steht, ein Backend zur Administration der Projekte, Projektbenutzer, Projektadministratoren und Suchaufgaben für die Relevanzprojekte, das Frontend zur Durchführung der Relevanztests sowie das Auswertungsmodul, welches allerdings noch nicht zur Verfügung steht, da es sich noch in der Entwicklungsphase befindet.

4.1 Der Suchmaschinenscraper

Allgemein ist ein Screenscraper (auch: Wrapper) eine Anwendung, die Informationen aus Texten extrahiert und in definierten Datenformaten speichert. Dabei werden beispielsweise bei Webseiten zwei Schritte durchgeführt. Im ersten Schritt wird die Webseite aufgerufen und im zweiten Schritt werden die Daten extrahiert, aufbereitet und gespeichert. Algorithmen zur Entwicklung von Wrappern oder Scrapern werden beispielsweise in [Kushmerick 2000] vorgestellt.

Für das Relevanztool wurde ein spezieller Suchmaschinenscraper entwickelt, der hauptsächlich auf die Funktionen aus der cURL-Library libcurl zurückgreift. cURL (Client for URLs) ist ein Programm, um Daten aus dem Internet ohne Webbrowser zu transferieren. Dabei werden alle gängigen Protokolle des Internets wie HTTP, HTTPS, FTP und FTPs unterstützt. Mit cURL kann der Aufruf eines Clients, z.B. eines Webbrowsers simuliert werden, um den Quelltext einer Webseite zu übertragen und zu speichern.

Der Suchmaschinenscraper für das Relevance Assessment Tool arbeitet in einer vorgegeben Reihenfolge. Der Projektadministrator legt Suchaufgaben inklusive der Suchanfragen an und schickt diese an den Scraper. Über cURL wird die Suchanfrage an die Suchmaschine geschickt, die die Trefferliste in HTML zurückgibt. Daraufhin wird die Suchergebnisseite geparkt und in einer DOM-Baumstruktur als XML-Dokument in einer Variablen gespeichert. In dieser Baumstruktur sind alle Tags der Suchergebnisseite hierarchisch als Knoten zugänglich. Die relevanten Daten für die Speicherung der Suchergebnisse, also beispielsweise die Treffer-URLs innerhalb der dafür vorgesehen Tags, werden über die XML-Abfragesprache XPath ausgelesen und anschließend in Arrays aufbereitet.

Jede Websuchmaschine nutzt andere Strukturen und Tags für die Gestaltung der Suchergebnisseite, auch wenn sich bei Web-Suchmaschinen Standards bei der Ergebnispräsentation herausgebildet haben [Lewandowski & Höchstötter 2009]. Beispielsweise werden die Trefferbeschreibungen auf einer Google-Suchergebnisseite durch andere HTML-Tags umschlossen als die Trefferbeschreibungen bei Yahoo oder Microsoft Bing.

Daher sind im Suchmaschinenscraper für jede eingetragene Suchmaschine Filter definiert, um die Trefferüberschrift, die Treffer-URL und die Trefferbeschreibung auszulesen. Ferner sorgen Filter auch dafür, dass zunächst nur organische Treffer erfasst und alle Elemente, die Hinweise auf die gescrapte Suchmaschine geben könnten, entfernt werden. In einem weiteren Schritt werden die aus der Trefferseite ausgelesenen URLs ebenfalls über cURL-Funktionen aufgerufen und als lokale Kopien für die Retrievaltests gespeichert. Damit wird sichergestellt, dass die Probanden die Suchergebnisse in genau der Form bewerten, die zum Zeitpunkt der gestellten Suchanfrage durch die Suchmaschinen ausgegeben wurden. Der Scraper erkennt dabei den Dateityp des Ergebnisdokuments, z.B. HTML oder PDF. Alle extrahierten Daten aus den Ergebnissen zu der Suchanfrage werden in der MySQL-Datenbank gespeichert.

Der Suchmaschinenscraper kann nicht nur für die Relevanztests eingesetzt werden, sondern auch unabhängig von den anderen Komponenten. Das Scraper-Modul kann die Ergebnisseiten diverser Suchmaschinen auswerten und in strukturierter Form exportieren. Damit können beispielsweise die Ergebnisseiten von Suchmaschinen zu mehreren Suchanfragen zusammengestellt und für andere Zwecke ausgewertet werden. Wir werden weiter unten explizit auf die Einsatzmöglichkeiten des Scrapers eingehen.

The screenshot shows the 'Relevance Assessment Tool' interface. The browser address bar displays 'http://searchstudies.org/rat/acm n/projects.php?id=62'. The page title is 'Relevance Assessment Tool'. The user is logged in as 'admin' and the project is 'Dirks Test'. The main content area shows 'Search Engine Scraper: 143- Deutsche Gesellschaft für Informationwissenschaft und-praxis'. Below this, a table titled 'Preview Scraped Results: 143- Deutsche Gesellschaft für Informationwissenschaft und-praxis' displays the following data:

Position:	URL:	Title:	Description:	Search Engine:
1	http://www.dginet.de/web/dgi/home	Home - dginet.de	LOGIN FÜR MITGLIEDER DER DGI: E-Mail Adresse Passwort Feststellasse ist gedrückt.	bing.de
2	http://www.dgd.de/	DGI-Homapage (former DGD)	Die wissenschaftliche und berufständische Fachgesellschaft der deutschen Informationsspezialisten. Informationen über Mitgliedschaft, Arbeitskreise, Tagungen und Ausbildung ...	bing.de
3	http://www.dgi-net.de/	Deutsche Gesellschaft für Infektiologie	Die DGI: stellt sich vor und bietet Informationen zu aktuellen Veranstaltungen.	bing.de
4	http://www.dgi-master.de/scripts/show.aspx?content=shop/home	DGI - Master of Science in Oral Implantology, Hannover, Prof. Dr ...	Wollen Sie Ihre berufliche Qualifikation sichern und weiterentwickeln? Wollen Sie das Profi Ihrer Praxis schärfen und das Vertrauen Ihrer Patienten festigen? Dann gehören Sie zu ...	bing.de
5	http://www.dgi-fortbildung.de/static/index.html	DGI - Fortbildung	DGI Fortbildung: Technische Innovationen und eine rasante Entwicklung prägen die Implantologie. Sie ist ein wichtiger Teil der Zahnmedizin und ihr größter Wachstumseereich.	bing.de
1	http://www.dgi-ev.de/	DGI	Mit wissenschaftlichen Empfehlungen/Stellungnahmen zu implantologischen Fragestellungen. DGI-Forschungsförderung - DGI-Mitglied werden lohnt sich! Kongressberichte	google.de
2	http://www.dgd.de/	DGI-Homapage (former DGD)	Die wissenschaftliche und berufständische Fachgesellschaft der deutschen Informationsspezialisten. Informationen über Mitgliedschaft, Arbeitskreise, ...	google.de
3	http://www.dgi-net.de/	Deutsche Gesellschaft für Infektiologie	Die DGI: stellt sich vor und bietet Informationen zu aktuellen Veranstaltungen.	google.de
4	http://www.immungenetik.de/	Deutsche Gesellschaft für Immungenetik (DGI) - Ziele	Auf den Seiten DGI - Deutsche Gesellschaft für Immungenetik finden Sie Informationen zur Förderung der Forschung und Lehre auf dem Gebiet der ...	google.de
5	http://dgi.d-grid.de/	DGI - D-Grid		google.de

At the bottom of the table, there are 'Save' and 'Cancel' buttons.

Abbildung 1: Strukturierte Erfassung der Ergebnisseiten am Beispiel von Google und Bing.

Der von uns entwickelte Scraper kann zurzeit die Ergebnisseiten von Bing, Google und Yahoo erfassen. Da das Tool zuvorderst für die Evaluierung der Websuche angelegt wurde, werden diese Scraper für die wichtigsten Suchmaschinen kontinuierlich gepflegt und stehen jederzeit für Projekte zur Verfügung. Weitere Scraper wurde für bestimmte Projekte entwickelt, beispielsweise für die Suchen von Musicload, Ebay, Amazon und Google Shopping. Der Aufwand für die Erstellung der Scraper variiert stark aufgrund der unterschiedlichen Komplexität der Ergebnispräsentation der verschiedenen Angebote.

Abbildung 1 zeigt beispielhaft die aus den Ergebnisseiten (SERPs; search engine results pages) extrahierten Informationen: Es werden die Rankingposition des Treffers, die URL, der von der Suchmaschine angezeigte Titel des Dokuments und die Trefferbeschreibung im System gespeichert. In einem zweiten Schritt werden dann die die Ergebnisdokumente selbst gespeichert, so dass in der Testsituation selbst dann nicht auf die aktuellen Versionen der Dokumente (die unter Umständen zwischenzeitlich verändert wurden), sondern auf die zum Zeitpunkt der Erfassung gültige Version zurückgegriffen wird. Mit dem Scraper ist es möglich, die Ergebnisse mehrerer Suchmaschinen gleichzeitig abzufragen; in der Abbildung sind beispielhaft jeweils fünf Ergebnisse von Google und Bing zu sehen.

Für ein weiteres Projekt wurde ein Scraper für Musicload.de entwickelt; zusätzliche Scraper, die für einzelnen Projekte benötigt werden, werden je nach Bedarf erstellt. Der Aufwand, eine neue Suchmaschine in den Scraper zu integrieren, unterscheidet sich von Suchdienst zu Suchdienst. Eine schnelle Integration von Suchdiensten ist möglich, wenn diese nur einfache Listen für ihre Suchergebnisse nutzen. Ein größerer Aufwand ist bei Trefferlisten nötig, die neben organischen Treffern noch Universal-Search-Ergebnisse, wie Bildergalerien oder Ergebnisse aus einer Nachrichtensuche, einblenden. Ferner muss der entwickelte Scraper regelmäßig gewartet werden, da sich Strukturen und Quelltext der Suchergebnisseiten oftmals verändern.

4.2 Das Backend zur Projektverwaltung

4.2.1 Testdesign

Das Backend des Tools wird über einen Webbrowser aufgerufen und zur Administration von Projekten genutzt. Nach dem Loginprozess hat der Administrator die Möglichkeit, ein neues Projekt direkt Schritt für Schritt anzulegen. Dieser Ansatz wurde gewählt, um einen schnellen und einfachen Zugang für die Gestaltung eines Testszenarios zu gewährleisten. Insgesamt werden acht Schritte durchlaufen, in denen die Daten über Formulare eingegeben werden:

Projektname

Projektbeschreibung: Diese wird den Probanden zur Orientierung vor der Durchführung des Tests angezeigt.

Projektsprache: Das Tool wurde so angelegt, dass die Benutzerführung in Sprachtemplates angelegt wird, was Tests in unterschiedlichen Sprachen ermöglicht.

Projektzugang für Nutzer: Hier kann entschieden werden, ob jeder Nutzer sich mit einem eigenen Zugangscode einloggen muss (hohe Kontrolle über die Jurorengruppe) oder ob sich alle Juroren eines Test mit der gleichen Kennung einloggen können (breite Streuung des Zugangs möglich).

Anzahl der Suchanfragen pro Juror

Zu testende Suchmaschinen: Hier können einerseits die Suchmaschinen, welche durch den Scraper unterstützt werden, ausgewählt werden. Außerdem können weitere Suchmaschinen hinzugefügt werden, deren Treffer dann allerdings manuell erfasst werden müssen und dann in strukturierter Form (.csv-Dateien) in das Tool importiert werden können.

Tasks: Hier können Suchaufgaben, bestehend aus Kurztext, Suchanfrage und Beschreibung, eingegeben oder strukturiert hochgeladen werden.

Bewertungsskalen: Es können eine oder mehrere Skalen ausgewählt oder neu erstellt werden. Das Tool unterstützt sowohl Ordinal- als auch Nominalskalen mit beliebig festzulegenden Benennungen, so dass eine große Freiheit hinsichtlich der abzufragenden Daten besteht.

Relevance Assessment Tool

Projects	Templates	User Administration	Manual	Logout
----------	-----------	---------------------	--------	--------

User: **admin** Project: **ASIST**

Project

- Open Project
- Administrare Project
- Export Results
- Create New Project

Project Users

- User Administration

Seach Tasks

- Overview
- Administrare Tasks
- Create Single Task
- Create Multiple Tasks
- Export Results

Scraper

- Scraper
- Multiple Scraper

Import Results

- Results
- Multiple Results

Administrare Project

Project Name:

Project Description:

Project End Text:

Language Template:

Search Engines:

<input type="checkbox"/> amazon.de	<input type="checkbox"/> google.de/products
<input type="checkbox"/> ask.com	<input type="checkbox"/> otto.de
<input type="checkbox"/> ask.de	<input checked="" type="checkbox"/> yahoo.com
<input checked="" type="checkbox"/> bing.com	<input type="checkbox"/> yahoo.de
<input type="checkbox"/> bing.de	
<input type="checkbox"/> ebay.de	
<input checked="" type="checkbox"/> google.com	
<input type="checkbox"/> google.de	

Search Tasks per User (old value = 1):

Mix Results:

Scale:

Scale:

Scale:

Abbildung 2: Projektübersicht in der Administratorenoberfläche (Ausschnitt).

Nachdem alle Projektinhalte definiert wurden, kann das Projekt gespeichert werden und der Administrator wird in die Projektverwaltung weitergeleitet. Hier stehen Optionen zur Verfü-

gung, die eine nachträgliche Bearbeitung von Projekten ermöglichen sowie das Interface zum Eintragen weiterer Projektadministratoren. Projekte können ebenfalls durch zusätzliche Suchaufgaben, Juroren oder Skalen erweitert werden. Eine Ansteuerung dieser Administrationsfunktionen steht auch direkt nach dem Login eines Projektadministrators zur Verfügung, indem er eines seiner Projekte über ein Dropdownmenü direkt öffnet. Weiterhin besteht die Möglichkeit, die Projektergebnisse in strukturierter Form zu exportieren, um diese beispielsweise mit Microsoft Excel auszuwerten.

Abbildung 2 zeigt die Übersicht eines Beispielprojekts in der Administratorenoberfläche. Zu sehen sind neben dem einleitenden und abschließenden Text für die Nutzeroberfläche die Auswahl der Projektsprache, die Auswahl der Suchmaschinen und die Auswahl der zu verwendeten Skalen für die Bewertung der Dokumente durch die Juroren.

4.2.2 Templates

Das System verwendet Templates, um wiederkehrende Arbeiten zu vermeiden, oder diese doch zumindest zu erleichtern. Einmal angelegte Templates können in allen weiteren Projekten wieder verwendet werden. Das Tool erlaubt die Anlage und Weiterverwendung folgender Templates:

- Sprachtemplates: Hier können sämtliche im Tool verwendete Beschriftungen, Hinweise, Fehlermeldungen, usw. in einer Sprache angelegt werden. So werden Tests in weiteren als den bereits angelegten Sprachen ohne zusätzliche Programmierung möglich.
- Skalen-Templates: Für unterschiedliche Tests werden verschiedene Skalen benötigt. Allerdings werden viele Skalen (beispielsweise die Relevanzbewertung auf einer Fünferskala) immer wieder verwendet. Einmal angelegte Skalen können in folgenden Projekten durch einfache Auswahl wieder verwendet werden.
- Suchmaschinen-Templates: Hier können Suchmaschinen, für die kein Scraper existiert, ergänzt werden. Manchmal lohnt es nicht, für ein Projekt einen Scraper für eine bestimmte Suchmaschine zu erstellen. Hier können die Ergebnisse manuell in das Tool eingespielt werden (mittels Excel-Upload).

4.3 Das Userinterface

Im Benutzerinterface (bzw. Frontend) des Testwerkzeugs werden die Daten für die Relevanzbewertungen erhoben. Das Userinterface wird über eine URL im Webbrowser aufgerufen und der Nutzer erhält Zugang zu den Suchaufgaben über einen festgelegten Zugangscode.¹ Es kann entweder ein Zugangscode für einen ganzen Test festgelegt werden oder aber ein Zugangscode pro Juror innerhalb eines Tests. Letztere Variante erlaubt eine Kontrolle über die Anzahl der durch einen Juror zu bearbeitenden Aufgaben und kann dafür genutzt werden, sicherzustellen, dass nur Juroren mit bestimmten Eigenschaften an dem Test teilnehmen können.

¹ Das Nutzerinterface kann unter <http://searchstudies.org/rat/> aufgerufen werden. Mit dem Zugangscode „RAT-Test“ kann ein Beispieldtest aufgerufen werden, in dem die Bewertung von Trefferbeschreibungen und Trefferdokumenten gezeigt wird. In diesem Test sind nur zehn Dokumente enthalten, so dass er in kurzer Zeit durchgeführt werden kann.

Vor die Relevanzbewertungen kann ein Fragebogen zur Abfrage soziodemographischer Daten gestellt werden.

Im Wesentlichen bearbeitet der Nutzer die Suchaufgaben, indem er die Ergebnisse und/oder Trefferbeschreibungen der Suchmaschinen zu den Suchanfragen mit Hilfe von Bewertungsskalen bewertet. Dabei liegen alle Ergebnisseiten und Trefferbeschreibungen in anonymisierter Form vor, so dass eine Unvoreingenommenheit der Testpersonen garantiert ist.

Das Nutzerinterface selbst ist in drei Bereiche geteilt. Im oberen Bereich sind Suchaufgabe, Suchanfrage und die Aufgabenbeschreibung sowie ein Fortschrittsbalken zu sehen (Abb. 3), während in einem Frame darunter die Ergebnisseite oder die Trefferbeschreibung angezeigt werden. Links davon finden sich die Bewertungsskalen.

The screenshot displays the 'Relevance Assessment Tool' interface. At the top, there is a browser window showing the URL 'http://searchstudies.org/rat/user_interface.php'. Below the browser, a progress bar indicates 'Progress: 0%' to '100%' with a '20%' marker. The 'Search Task' section contains the text 'tee'. On the left side, there are three evaluation scales:

- How relevant is this document to the search task shown above?**
 - 0 - not relevant
 - 1 - some elements relevant
 - 2 - partly relevant
 - 3 - mostly relevant
 - 4 - completely relevant
- Please comment on the relevancy of this document.**

Comment:
- Is this document relevant to the search task shown above?**
 - yes
 - no

Buttons for 'Next' and 'Skip' are located below these scales. The main content area shows a search result for 'tee' from the website 'kalineo'. The page title is 'Rund um den Tee finden Sie bei Kalineo köstliche Teesorten wie Sonnentor und Ronnefeldt, aber auch alles, was man für eine gemütliche Teestunde braucht: von Teekanne, Teedose und Teefilter bis zu Teespezialitäten: Weißer Tee, Grüner Tee, Schwarzer Tee'. The page features a navigation menu, a search bar, and a main content area with a large image of a teapot and a cup of tea. The text on the page describes the tea selection and provides information about the company and its products. On the right side, there is a 'Warenkorb' (shopping cart) section with a 'bestellen' (order) button and a promotional offer for 'KEINE VERSANDKOSTEN' (no shipping costs) for orders over 30€. At the bottom right, there is a section titled 'ZAHLEN SIE EINFACH & BEQUEM PER:' (pay simply and conveniently by) with a list of payment methods: RECHNUNG (invoice), LASTSCHRIFT (credit card), NACHNAHME (direct debit), and VORKASSE (advance payment).

Abbildung 3: Bewertungselemente innerhalb des Benutzerinterface.

Sind alle Testobjekte einer Suchaufgabe durch den Probanden bewertet, wird er zur nächsten Suchaufgabe geleitet, bis die vorab definierte Anzahl der Suchaufgaben für einen Juror abgearbeitet ist.

4 Beispielhafte Durchführung eines Retrievaltests

Der Nutzen des Tools soll in diesem Abschnitt beispielhaft an einem konkreten Test gezeigt werden. Als Testszenario dient hier ein typischer Suchmaschinen-Vergleichstest. Es sollen die drei bekannten Suchmaschinen Google, Yahoo und Bing verglichen werden. Verwendet werden sollen 50 Suchaufgaben, verglichen werden sollen die Ergebnisbeschreibungen sowie die Ergebnisdokumente selbst, wobei die ersten zehn angezeigten Ergebnisse berücksichtigt werden sollen. Um Reihungs- und Lerneffekte zu verhindern, sollen die Ergebnisse gemischt werden. Außerdem darf nicht ersichtlich sein, von welcher Suchmaschine ein Treffer kommt.

In einem ersten Schritt werden die Suchaufgaben erstellt. In diesem Schritt kann RAT keine direkte Unterstützung bieten; allerdings können bereits in vergangenen Tests verwendete Aufgaben wieder verwendet werden. Die Aufgaben können mit Beschreibungen und den zu verwendenden Suchanfragen als Excel-Liste in RAT hochgeladen werden, so dass das manuelle Anlegen jeder Aufgabe im Tool entfällt.

Im zweiten Schritt werden die Trefferbeschreibungen und Dokumente von den Suchmaschinen geholt. Würde dieser Schritt manuell erfolgen, müssten die Trefferbeschreibungen von den Ergebnisseiten kopiert werden, um sie später den Juroren vorzulegen. RAT extrahiert diese Beschreibungen automatisch und entfernt auch alle Elemente, die Rückschlüsse auf eine individuelle Suchmaschine zulassen. Bei einer manuellen Datenerfassung müssten zusätzlich die URLs aller angezeigten Treffer aufgerufen und die Dokumente für eine spätere Bewertung gespeichert werden. Zusätzlich wären die Suchmaschine und die jeweilige Trefferposition zu codieren.

Für die Bewertung durch die Juroren wären die Trefferbeschreibungen und die Dokumente zu mischen und dann den Juroren vorzulegen. Die erfassten Bewertungen wäre vor der Auswertung wieder zusammenzuführen, wobei die ursprünglich Reihenfolge der Treffer und die Zuordnung zu den Suchmaschinen wieder herzustellen wäre. Diese Schritte erfolgen in RAT automatisch.

Wenn man nun annimmt, dass der hier skizzierte Test 50 Suchaufgaben verwendet, so ergeben sich insgesamt 3.000 zu erfassende Relevanzurteile (3 Suchmaschinen x 50 Aufgaben x 10 Treffer x 2 Bewertungen (Beschreibung sowie Treffer)). Der Nutzen des Tools steigt mit der Komplexität bzw. des Umfangs des Tests. Mit Hilfe von RAT lassen sich auch umfangreiche Untersuchungen effizient durchführen. Grenzen bestehen vor allem hinsichtlich der Rekrutierung der Juroren und der Anzahl der durch einen Juror zu bewertenden Dokumente. Dies ist allerdings kein technisches Problem und liegt damit außerhalb der Softwareentwicklung.

5 Technische Grenzen und Probleme bei der Anwendungsentwicklung

Bei der Entwicklung des Tools traten einige Schwierigkeiten auf, die in diesem Abschnitt kurz erläutert werden sollen. Die Probleme beziehen sich ausschließlich auf den Suchmaschinenscraper. Wie zuvor erwähnt nutzen Suchmaschinen verschiedene Strukturen zur Gestaltung ihrer Ergebnisseiten. Web-Suchmaschinen wie Google und Yahoo greifen auf den Universal-Search-Ansatz zurück und reichern die Suchergebnisse durch Ergebnisse aus anderen Dokumentensammlungen an [Lewandowski 2008; Quirnbach 2009]. Beispielsweise werden neben organischen Suchergebnissen und Anzeigen (Sponsored Links) auch Module für Ergebnisse aus einer Nachrichten-, Bilder- oder Videosuche angezeigt. Der Suchmaschinenscraper soll zunächst nur organische Treffer erkennen und speichern, daher mussten für alle Suchmaschinen, die gecrapt werden sollen, Filter definiert werden, die nur die natürlichen Treffer erfassen. In der Entwicklung zeigte sich dabei, dass die Suchmaschinen nicht immer die gleichen Tags zur Formatierung ihrer Suchergebnisse nutzen. Die Formatierungstags sind aber zum Auslesen der Suchergbnisselemente entscheidend, z.B. sind die einzelnen Treffer durch Listenelemente abgegrenzt und die Treffer-URLs werden über bestimmte Tags gestaltet. Google nutzt beispielsweise für die Trefferbeschreibungen unterschiedliche Tags, auch wenn der Nutzer dies auf der Suchergbnisseite nicht erkennen kann. Es reicht also nicht, eine Struktur im Quelltext zu identifizieren und darauf den Scraper für die Suchergbnisseite zu entwickeln. Für jede Suchmaschine müssen über diverse Testanfragen alle möglichen Formatierungs-Tags identifiziert werden, um die Suchergebnisse vollständig erfassen zu können. In der Anwendungsentwicklung wurden verschiedene Ansätze zur Erkennung und Filterung getestet, darunter Stringvergleiche, reguläre Ausdrücke und Xpath. Als schnellste und zuverlässigste Variante hat sich dabei Xpath erwiesen, da sie direkte Abfragen auf die Knoten innerhalb der DOM-Baumstruktur ermöglicht. Die Aufbereitung in einer Baumstruktur hat den großen Vorteil, dass die Suchergbnisselemente bereits hierarchisch strukturiert gespeichert werden.

Neben den Problemen, die bei der Identifizierung der Strukturen der Suchergbnisseiten aufgetaucht sind, zeigten sich auch Schwierigkeiten bei der lokalen Speicherung der Ergebnisseiten. Die Dokumente werden über cURL lokal gespeichert, wobei zu berücksichtigen ist, dass von manchen Websites sog. Framebreaker genutzt werden, damit diese nicht in Frames auf externen Seiten eingebunden werden können. Dieser Prozess wird bei der lokalen Sicherung über die Entfernung von Javascript in den zu speichernden Dokumenten realisiert. Ein weiteres Problem sind interne und externe Verlinkungen auf den Webseiten. Der Benutzer soll die tatsächlich gefundene Ergebnisseite im Interface bewerten und nicht die Möglichkeit haben, innerhalb der Ergebnisseite zu navigieren. Daher wurde eine Funktion geschrieben, die alle Links auf den Ergebnisseiten unbrauchbar macht. Die Funktion sucht alle Links und überschreibt diese. Klickt ein Juror einen Link an, wird er darauf hingewiesen, dass nur der Treffer selbst, nicht aber weitergehenden Seiten zu bewerten sind.

Für die zukünftige Entwicklung des Tools ist angedacht, anstatt einer lokalen Speicherung der HTML-Seiten grafische Screenshots zu sichern. Für eine technische Umsetzung müssen aber spezielle Voraussetzungen in der Serverumgebung geschaffen werden, die in diesem Aufsatz nicht näher erläutert werden können.

Der Suchmaschinenscraper bedarf einer regelmäßigen Wartung, da sich die Suchergbnisseiten der Suchdienste häufig verändern. Dabei ist zu beachten, dass eine Erweiterung des Scrapers durch neue Suchmaschinen relativ aufwendig ist, weil jeder Suchdienst eigene

Methoden zur Aufbereitung der Ergebnisse nutzt. Die Strukturen der Suchergebnisseiten müssen daher genau analysiert werden. Dabei werden die HTML-Tags für Trefferbeschreibungen, Trefferüberschriften und Treffer-URLs im HTML-Quelltext der Suchergebnisseiten identifiziert und in den Xpath-Statements im Scraper festgehalten.

6 Fazit und Ausblick

Mit dem Relevance Assessment Tool haben wir ein Werkzeug entwickelt, das den Aufwand für Studien zur Retrievaleffektivität von Web-Suchmaschinen erheblich reduziert. Nicht nur wird den Juroren die Bewertung der Dokumente erleichtert, sondern vor allem im Bereich der Datenerfassung und Aufbereitung kann viel Aufwand eingespart werden. Mit dem Scraper lassen sich nahezu beliebig viele Ergebnisse jeder Suchmaschine leicht erfassen, innerhalb des Nutzerinterface werden die Herkunft und die Platzierung der zu bewertenden Treffer automatisch unkenntlich gemacht. Dadurch lassen sich die bisherigen Grenzen der Retrievaleffektivitätsstudien, die vor allem in der geringen Anzahl von Suchanfragen, Suchmaschinen und zu bewertender Dokumente je Suchanfrage liegen, erweitern.

Wir haben mit dem Tool bereits mehrere Studien durchgeführt. Es hat sich gezeigt, dass die Juroren das Tool ohne Probleme nutzen können, auf der Administrationsseite aber noch an der Nutzerfreundlichkeit gearbeitet werden muss. Diese ist insbesondere von Bedeutung, da das Tool auch geschaffen wurde, damit Studierende es für ihre Abschlussarbeiten einsetzen können. Diese Personen werden das Tool nur für eine einzige Untersuchung nutzen, daher soll die Einarbeitungszeit so gering wie möglich gehalten werden.

In den nächsten Schritten werden wir das Modul zur statistischen Auswertung der Ergebnisse erstellen. Damit wird sich dann der vollständige Retrievaleffektivitätstests in dem Tool abbilden lassen und der Aufwand solcher Tests weiter verringern. Um den aufgetretenen Problemen mit den gespeicherten Ergebnisseiten zu begegnen, werden wir zusätzlich unsere Serverumgebung so anpassen, dass die Erzeugung automatischer Screenshots der Ergebnisseiten möglich wird. Zuletzt werden wir die Scraper auf die von den Suchmaschinen neben den organischen Treffern präsentierten Ergebnisse (Bilder, Videos, Nachrichten, usw.) ausweiten, um eine Bewertung der vollständigen Ergebnispräsentation zu ermöglichen. Bisherige Tests beurteilen die Suchmaschinenergebnisse unzureichend, da alternative Ergebnisse mittlerweile einen nennenswerten Anteil an der gesamten Ergebnisdarstellung haben [Höchstötter & Lewandowski 2009].

Bei der Durchführung von Retrievaltests streben wir eine Zusammenarbeit mit weiteren Hochschulen und Forschungseinrichtungen an. Wir sind offen für die Wünsche und Anregungen der informationswissenschaftlichen Community.

Literatur

- Griesbaum, J. (2004). Evaluation of three German search engines: Altavista.de, Google.de and Lycos.de. *Information Research* 9(4). <http://informationr.net/ir/9-4/paper189.html> [7.11.2010].
- Höchstötter, N., & Lewandowski, D. (2009). What Users See - Structures in Search Engine Results Pages. *Information Sciences* 179(12), 1796-1812.

- Kushmerick N. (2000). Wrapper Induction. Efficiency and Expressiveness. *Artificial Intelligence* 118, 15-68.
- Lewandowski, D. (2008). The retrieval effectiveness of Web search engines: considering results descriptions. *Journal of Documentation* 64(6), 915-937.
- Lewandowski, D., & Höchstötter, N. (2007). Qualitätsmessung bei Suchmaschinen: System- und nutzerbezogene Evaluationsmaße. *Informatik Spektrum* 30(3), 159-169.
- Lewandowski, D., Höchstötter N. (2009). Standards der Ergebnispräsentation. In: Lewandowski, D. (Hrsg.): *Handbuch Internet-Suchmaschinen*. Heidelberg: Akademische Verlagsgesellschaft Aka GmbH, 2009. S. 204– 219
- Quirnbach, S. (2009). Universal Search - Kontextuelle Einbindung von unterschiedlicher Quellen und Auswirkungen auf das User Interface. In D. Lewandowski (Hrsg.): *Handbuch Internet-Suchmaschinen*. Heidelberg: Akademische Verlagsgesellschaft Aka GmbH.
- Peterson A. (2006). SEAT. Midnight Research Labs, 2006.
<http://midnightresearch.com/projects/search-engine-assessment-tool/> [7.11.2010].
- Stokdyk, D. (2009). How to Scrape a Search Engine Results Page for Your SEO Project. *Marketing OH: Internet Marketing for Web 2.0*, 2009. <http://www.marketing2oh.com/scrape-serps-for-seo-analysis/> [7.11.2010].
- Tawileh, W., Griesbaum, J., & Mandl, T. (2010). Evaluation of five web search engines in Arabic language. <http://www.kde.cs.uni-kassel.de/conf/lwa10/papers/ir1.pdf> [7.11.2010].
- Tosques, F. & Mayr, P.: Programmierschnittstellen der kommerziellen Suchmaschinen. In: Lewandowski, D.: *Handbuch Internet-Suchmaschinen*. Heidelberg: AKA, 2009, S. 116-147.



Social Media und Web Science

Das Web als Lebensraum

**2. DGI-Konferenz
64. Jahrestagung der DGI**

**Düsseldorf
22. bis 23. März 2012**

Proceedings

**herausgegeben von
Marlies Ockenfeld
Isabella Peters
Katrin Weller**



Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis