Informationskompetenz: ein Thema für den Informatikunterricht

Diana Jurjević, Nando Stöcklin, Werner Hartmann

Zentrum für Bildungsinformatik
Pädagogische Hochschule Bern
Muesmattstraße 29
CH-3012 Bern
diana.jurjevic@phbern.ch
nando.stoecklin@phbern.ch
werner.hartmann@phbern.ch

Abstract: Das Angebot an Online-Informationen nimmt weiter rasant zu. Die neuen Informationsmedien, speziell die Dienste im Web 2.0, erfordern neue Suchstrategien und neue Beurteilungsmethoden. Die Informationskompetenz heutiger Schülerinnen und Schüler lässt im Unterschied zu den Fertigkeiten und dem Gebrauch moderner Informations- und Kommunikationstechnologien aber zu wünschen übrig. Dem Informatikunterricht kommt hier eine zentrale Rolle in der Allgemeinbildung zu. Informationskompetenz bedingt auch ein Verständnis für die grundlegenden Konzepte von Informationsdiensten. Nach einem Überblick zu Informationsrecherche im Internet wird aufgezeigt, wie herkömmliche algorithmische Suchmaschinen und neuere Ansätze semantischer Suchdienste in Verbindung mit der geforderten Informationskompetenz im Informatikunterricht thematisiert und als authentische Lernumgebungen genutzt werden können.

1 Informationsrecherche im Internet, eine große Herausforderung

Der Umfang an Informationen im Internet ist in den letzten Jahren weiter rasant gewachsen. Besonders deutlich wird das am Web 2.0, dem Internet der Nutzer. Neben den etablierten Portalen wie Wikipedia, YouTube oder Facebook wächst ein gigantischer Ameiseisenhaufen von Microcontent, kleine und unorganisierte Informationseinheiten, wie ppt-Dokumente, Blogs oder Memos, jedoch ohne die persistenten Strukturen einer Ameisenkolonie. Diese Informationseinheiten werden nicht zentral verwaltet. Es gibt keine Stelle, welche die Inhalte zuerst sammelt, selektiert, hierarchisch anordnet und dann veröffentlicht, sondern die Inhalte liegen dezentral im Netz verteilt. Neben dem wachsenden Anteil an nutzergenerierten Inhalten im Web verändern sich auch die Formate der Inhalte. Die technischen Hürden, Fotos und Videos im Internet zu veröffentlichen, sind so gering geworden, dass diese Möglichkeit immer mehr genutzt wird. Damit hat eine Massenwanderung neuer Daten und Informationen in das Internet begonnen. 2008 betrug der Zuwachs neuer Videoclips allein auf YouTube circa 10.000 Stunden Videomaterial pro Tag. Für das Erzeugen von Bildern stehen heute weltweit mehr als

eine Milliarde Aufnahmegeräte zur Verfügung: Digitalkameras, Handys, Überwachungskameras oder Satellitenfotos sind nur einige Beispiele [Ga08].

Angesichts dieser Informationsflut hat das Selektieren von Informationen gegenüber dem Sammeln von Informationen deutlich an Bedeutung gewonnen. Gleichzeitig ist die Informationsbeschaffung auch anspruchsvoller geworden. Gerade um den Microcontent des Web 2.0 zu erschließen, müssen neue Suchstrategien entwickelt werden. Je nach Suchanfrage muss das passende Werkzeug gewählt werden: So gibt es für Blogs, Bilder und Videos spezialisierte Suchmaschinen. Die Herausforderung bei der Suche bleibt die Gleiche: Im Meer der Informationen sollen relevante Informationen von irrelevanten unterschieden werden. Bibliothekssysteme stützten sich auf Lösungsverfahren zum Auffinden von Informationen in großen Datenbeständen, welche die Informatik lange vor der Erfindung des Internets entwickelt hat. Heute verbreitet genutzte Internet-Suchmaschinen wie Google oder Yahoo sind Erben der Entdeckungen im Information Retrieval in den 1960er Jahre. Es handelt sich hier um *algorithmische* Suchmaschinen. Diesen Suchmaschinen gemeinsam ist, dass sie mittels Crawler die Inhalte des Webs durchsuchen, in einem Index erfassen und mit Hilfe statistischer Methoden zu einer Suchanfrage eine Auswahl und Rangierung der Inhalte vornehmen. Algorithmische Suchsysteme haben aber ihre Grenzen. Udi Manber, technischer Vizepräsident von Google, schildert, dass die heutigen Suchsysteme technologisch fast schon ausgereizt sind. Zur weiteren Suchmaschinenoptimierung sei heute vor allem ein besseres Verständnis der Nutzenden für die Funktionsweise von Suchmaschinen gefragt:

> „It goes two ways: The content provider should think about how users will look for their content, and the user should think about what words people use to write about their content. Very often people make the mistake of using a search engine as if they are talking to another person. They use all sorts of words that a person will understand, but are not going to be in the content they are searching for. You should think about what you expect to see in the actual page and search for that. Having said that, we're doing this, too. We will take your query and try to "understand" it and match is as best we can to the content we find on the Web." [Ma08]

Die Informationsbeschaffung wird zunehmend komplexer und mit dem Web 2.0 sind überdies neue, alternative Suchdienste entstanden, deren Funktionsweise sich deutlich abhebt von den bisher gängigen Suchmaschinen. Diese Suchdienste versuchen die bisherigen algorithmenorientierten Verfahren durch den Einbezug *semantischer* Technologien zu verbessern. So bieten zum Beispiel Social-Bookmarking-Dienste die Möglichkeit, auf den von den Benutzenden generierten Tags zu suchen, also in Inhalten, die durch die Nutzenden selbst verschlagwortet wurden. Diese Dienste können aber auch verwendet werden, um Inhalte geringer Qualität auszusortieren oder Gruppen von Nutzenden mit ähnlichen Interessen zu bilden. Eine gute Übersicht zum Status quo und den Entwicklungstendenzen bei Suchmaschinen findet sich in [LM08].

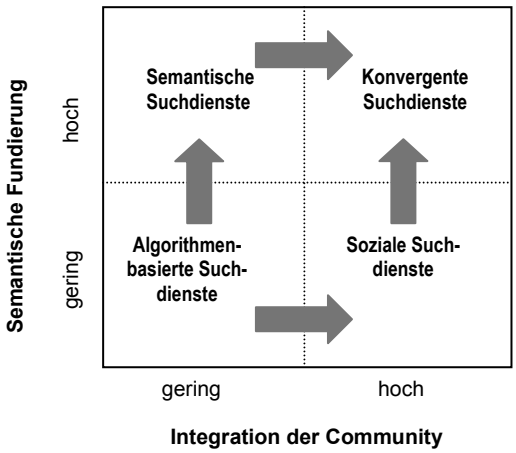

Abbildung 1: Unterschiedliche Typen von Suchdiensten, aus [LM08] S. 2

Für den Informatikunterricht bietet das Thema Suchsysteme ideale Voraussetzungen für eine authentische Lernumgebung, welche an den Alltagsproblemen der Schülerinnen und Schüler ansetzt, damit den Bezug zu ihrer Lebenswelt sicherstellt und Lernen hierdurch wahrscheinlicher macht. Der Informatikunterricht ermöglicht den Schülerinnen und Schülern, einen Blick hinter die Kulissen der kommerziellen Suchmaschinen zu werfen, die sie täglich nutzen. Grundlegende Konzepte von Informationsdiensten können aufgezeigt werden: Ausbeute und Präzision bei einer Suche, Indexierung als ein Schlüsselelement in der maschinellen Durchforstung großer Datenmengen oder Boolesche Operatoren bei Datenbankabfragen sind nur einige Beispiele.
Das Wissen über die Funktionsweise von Suchmaschinen kann den Umgang mit Informationen im Internet verbessern. Dieter Baacke [Ba97] beschreibt in seinem vierdimensionalen Medienkompetenzmodell Medienkunde als eine Dimension von Medienkompetenz. Dem Informatikunterricht kommt bei der Internetrecherche eine Schlüsselrolle zu und hat das Potenzial, die Medienkompetenz der Schülerinnen und Schüler zu steigern.

2 Informationsdienste im Wandel

Die klassischen Internet-Informationsdienste lehnen sich stark an Paradigmen aus der Bibliothekswelt an. Themenkataloge basieren auf der Einordnung von Informationen in Kategorien, algorithmische Suchmaschinen wie Google auf der Erschließung von Informationen anhand eines maschinengenerierten Stichwortindexes. Die Forschungsdisziplin Information Retrieval untersucht und evaluiert Online-Suchdienste nach definierten Kriterien. Die klassischen Retrievalmaße stammen aus der Zeit vor dem großen Internetdurchbruch in den 1990er Jahren und berücksichtigen die Veränderungen nicht, die mit den Entwicklungen im Internet einhergegangen sind. Im Vordergrund stehen statistische Maße, das Nutzungsverhalten der User wird praktisch vollständig ausgeblendet. So spielt das optimale Verhältnis von Ausbeute und Präzision eine zentrale Rolle. Auf eine Such-

anfrage soll ein Suchdienst möglichst alle zur Verfügung stehenden relevanten Informationen anzeigen (Ausbeute), aber auch genau nur diese und keine Irrelevanten (Präzision). In der Praxis sagen aber diese Maße nur wenig über die Qualität einer Suchmaschine aus. 73% aller User, die eine Suchmaschine in den USA nutzen, schauen sich nur die erste Seite der Trefferlisten an, in Europa sind es sogar 83%. Diese Zahlen ermittelten Jansen und Spink [JS05] anhand der Webprotokolle von neun Suchmaschinen. Über eine Milliarde Suchanfragen wurden im Zeitraum von sieben Jahren gesammelt und miteinander verglichen. Suchmaschinen sortieren relevante Dokumente bereits im Vorfeld aus. Durch die Beschränkung auf die erste Trefferseite fallen auch noch alle relevanten Dokumente weg, die erst weiter hinten in den Trefferlisten erscheinen. Für die Evaluation von Suchsystemen bedeutet das, dass es nicht ausreicht, wenn ein Suchsystem relevante Dokumente findet, diese müssen auch auf der ersten Seite angezeigt werden. Alle relevanten Dokumente, die nicht auf der ersten Ergebnisseite erscheinen, nimmt die große Mehrheit der Nutzenden gar nicht wahr.

Je nachdem, ob der menschliche Faktor in die Evaluation einbezogen oder ausgeklammert wird, können die Studienergebnisse unterschiedlich ausfallen. Um die Systeme nachhaltig zu verbessern, müssen sich die Evaluationen vermehrt an der Realität der Suchenden und ihren Interaktionen orientieren. Auf diese Weise könnten neue, wirksame und nachhaltige Forschungsergebnisse gefunden werden. Bereits 1995 hielt Saracevic fest:

> "The issue and challenge for any and all IR evaluations are the broadening of approaches and getting out of the isolation and blind spots of single level, narrow evaluation. How can interaction be ignored in IR evaluation at any level?" [Sa95]

Suchende nehmen nicht nur durch ihr Nutzungsverhalten Einfluss auf die Retrievalmaße, sondern sie bauen auch an neuen Arten der Informationserschließung im Internet mit. Web-2.0-Dienste stützen sich stark auf das Tagging ab. Nutzergenerierte Tags haben auf den ersten Blick den Nachteil der unkontrollierten Verschlagwortung durch Laien, der Alptraum aller Bibliotheksexperten. Bei genauerem Hinsehen stellt man aber fest, dass diese vermeintlichen Nachteile leicht durch die folgenden beiden Vorteile kompensiert werden können:

1. Von Experten vergebene Schlagworte bilden in der Regel das Wissen dieses Experten ab. Schlagworte basieren auf einem kontrollierten Thesaurus und sind deshalb oft orthogonal zur Sprachwelt von Laien. Demgegenüber sprechen die Benutzenden von Web-2.0-Diensten dieselbe Sprache wie andere Nutzende.

2. Da bei Social-Tagging-Diensten die Informationen von vielen Personen auf unterschiedliche Art verschlagwortet werden, führen Suchanfragen dort eher zum Erfolg als bei Verschlagwortung mit wenigen, in der Regel sehr spezifischen Begriffen durch Expertinnen.

Austauschplattformen wie Flickr, YouTube oder der Social-Bookmarking-Dienst Delicious sind damit eine Mischform zwischen von Experten erstellten Themenkatalogen und algorithmischen Suchmaschinen. Auf der einen Seite erschließen diese semantischen

oder sozialen Suchdienste weniger Informationen als die großen algorithmischen Suchmaschinen, die mit ihren Crawlern das Internet möglichst in der ganzen Breite und Tiefe absuchen. Rein statistisch gesehen nimmt damit die Ausbeute ab. Dieser Nachteil fällt aber kaum ins Gewicht, da die meisten User sowieso nur die ersten Treffer eines Suchdienstes inspizieren. Auf der anderen Seite bieten Suchdienste wie Delicious gewissermaßen von anderen Usern handverlesene Informationen an. Die Community dient als Filter für die Inhalte. Damit erhöht sich die Präzision bei der Suche.

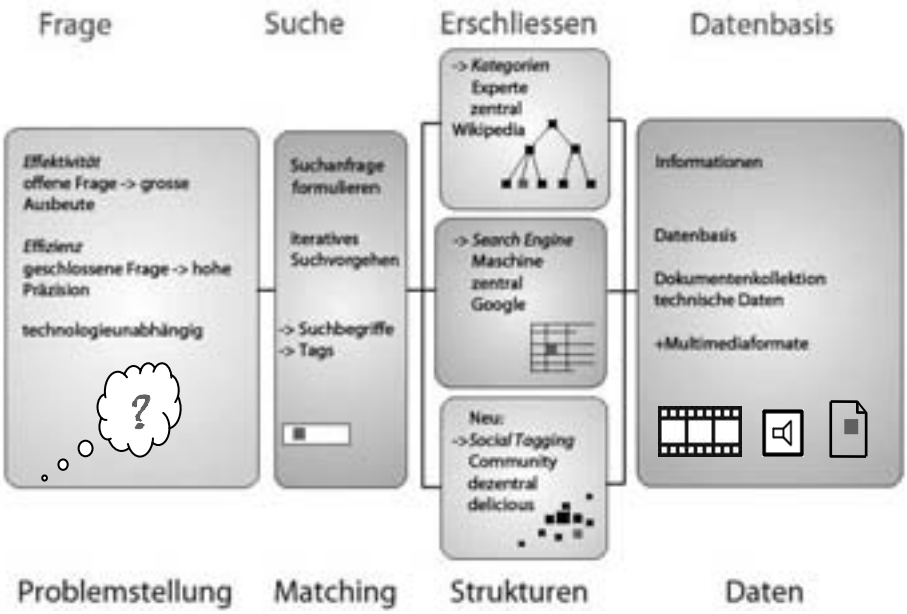

Abbildung 2: Prozessablauf der Informationssuche

Semantische und soziale Suchdienste sowie Mischformen zwischen diesen Diensten und herkömmlichen algorithmischen Suchdiensten sind in naher Zukunft aus der Landschaft der Informationsdienste nicht mehr wegzudenken. Information Retrieval ist nur ein Beispiel, wo informatische Konzepte und soziale Faktoren ein neues Ganzes bilden. Diese Entwicklung schlägt sich generell in der wachsenden Bedeutung der Web Science nieder, einem Wissenschaftszweig, der sich immer mehr neben der Computer Science etabliert und künftig im Informatikunterricht einfliessen sollte.

Bislang wurden in der Informatik die Technologien mehr oder weniger getrennt von ihren Nutzerinnen und Nutzern behandelt. Der Bereich Human Computer Interaction fristet gerade an sehr technisch orientierten Hochschulen oft ein Schattendasein. Mit dem Web 2.0 lässt sich aber das Verhalten der Online-Nutzenden bei der Entwicklung von Informatiksystemen nicht mehr ausblenden. Ben Shneiderman fordert deshalb, das Forschungsfeld der Informatik um soziale Einflussfaktoren zu erweitern. Entwicklungen von Web-Technologien müssen Nutzerbedürfnisse berücksichtigen und deren Bedarf analysieren, bis hin zur Frage: Was wollen Personen und Communities vom Web

[Sh07]? Tim Berners Lee, der Erfinder des WWW, geht noch einen Schritt weiter. Er und weitere Wissenschaftler fordern, der Forschung über das WWW mehr Gewicht zu verleihen, indem neben der Informatik (Computer Science) die neue Disziplin „Web Science“ geschaffen wird [HS08]. Der tabellarische Vergleich der Disziplin „Web Science“ mit der bedeutend älteren Disziplin „Computer Science“ manifestiert den Trend im Internet: nicht mehr die Daten und Programme stehen im Vordergrund, sondern die Menschen, die sich für diese Daten und Programme interessieren.

Computer Science	Web Science
Topics	
Computer Networks	Social networks
Packet Switching	Voice over IP, music sharing
Information	Relationships
Programming languages	Wikis, blogs, tagging
Databases, operating systems, compilers	E-commerce, e-learning, e-government, medical informatics, financial analysis
3D graphics, rendering algorithms, computational geometry, object modelling	Creating and sharing video, animation, music, photos, maps

Abbildung 3: Computer Science vs. Web Science [Sh07]

3 Informationskompetenz im Informatikunterricht

Die Entwicklungen im Bereich der Suchdienste stellen die Nutzenden vor eine große Herausforderung. Die effiziente und effektive Informationsbeschaffung ist keineswegs trivial, auch wenn das viele Suchende glauben. In einer Studie des US-amerikanischen Pew-Forschungszentrums zeigt Deborah Fallows [Fa05] auf, dass Studierende reihenweise ihre Fähigkeiten bei der Suche überschätzen. Bei erfolglosen Suchprozessen neigen viele Nutzende zur Annahme, es gäbe die gesuchte Information im Internet nicht oder sie sei nicht auffindbar. Dieser Rückschluss greift oftmals zu kurz. Wahrscheinlicher ist, dass man die falschen Suchbegriffe gewählt, am falschen Ort gesucht, nicht die richtige Anzahl Suchbegriffe eingegeben oder den richtigen Treffer übersehen hat. Solche Fehleinschätzungen werden auf Paralysen zurückgeführt: Eine große Auswahlmöglichkeit verleitet gerade dazu, falsche Entscheidungen zu treffen, weil dann die Gefahr besteht, zu einfacheren Lösungs- und Erklärungsstrategien zu greifen oder die Entscheidung ganz aufzuschieben. Die Informationsflut im Internet ist ein Paradebeispiel dafür. Barry Schwartz (2004) beschreibt das Phänomen ausführlich in The Tyranny of choice im Scientific American.

Suchende überschätzen aber nicht nur sich selbst, sondern auch die Qualität ihrer Suchergebnisse. Das Vertrauen in die Richtigkeit der Informationen im Internet ähnelt dem Vertrauen in die etablierten TV- und Printmedien. Diese Erkenntnis ist nicht neu und ein klassisches Thema in der Medienpädagogik. Informationskompetenz ist eine Schlüsselqualifikation und umfasst nach der American Library Association die Fähigkeit, Informationsbedürfnisse zu erkennen, die benötigten Informationen gezielt zu lokalisieren und evaluieren und die gefundenen Informationen für die eigenen Zwecke nutzbar zu

machen und zu speichern. Für eine ausführliche Beschreibung sei etwa auf die Standards und Indikatoren zur Informationskompetenz von Studierenden von 1998 verwiesen [Am98].

Die kompetente Nutzung moderner Informationsdienste ist ein zunehmend komplexeres Unterfangen und setzt ein Verständnis für die Grundprinzipien der Suchdienste voraus. Genau so wie eine große Bibliothek ohne ein Verständnis für Katalogisierung, Oberbegriffe und Schlagworte nicht effizient genutzt werden kann, lassen sich Suchmaschinen nicht ohne dieses Hintergrundwissen kompetent nutzen. An dieser Stelle ist die Schule gefordert, die Grundlagen für die kompetente Nutzung dieser Dienste zu vermitteln. Genau hier kommt nun dem Informatikunterricht eine wichtige Rolle zu. Bereits 2004 hat die Gesellschaft für Informatik e.V. (GI) im Memorandum „Digitale Spaltung verhindern – Schulinformatik stärken!“ gefordert [Ge04]:

> „Wenn die Informatik in der Schule stärker verankert wäre, würde dies die allgemeine IT-Kompetenz und somit letztlich die Effizienz zahlreicher Arbeitsprozesse erheblich erhöhen […]. Um die allgemeine ICT-Kompetenz zu erhöhen, reicht es nicht, dass nur das Orientierungswissen gefördert wird, also nur Grundlagen wie Modellierung, Algorithmen und Datenstrukturen oder Berechnungsmodelle vermittelt werden. Ein Schulfach Informatik muss auch unmittelbar dazu beitragen, ICT-Werkzeuge effizient im Schulalltag und später im Arbeitsprozess zu nutzen. Es sollte deshalb neben dem Orientierungswissen die Anwenderkompetenz fördern.“

Der Aufbau und die gezielte Nutzung von Informationsdiensten muss ein integraler Bestandteil des Informatikunterrichtes sein. Es reicht aber nicht, nur die informatischen Konzepte von Suchsystemen zu lehren. Um den Unterricht zeitgemäss zu gestalten, müssen Faktoren wie Nutzungsverhalten und soziale Einflussfaktoren miteinbezogen werden. Am Beispiel zweier Lernumgebungen (Soekia und Tagidex) zeigen wir auf, wie sich im Informatikunterricht das Thema Informationsrecherche und Informationskompetenz sowohl aus informatischer Perspektive als auch für alle anderen Fächer gewinnbringend vermitteln lässt. Das Thema passt in verschiedene Inhaltsbereiche der Bildungsstandards Informatik, zum Beispiel Informatiksysteme, Strukturierungsprinzipien für Dokumente erkennen und anwenden und Informatik, Mensch und Gesellschaft [Ar07].

4 Soekia – ein Blick in die Blackbox algorithmischer Suchmaschinen

In der Medienpädagogik wurden zahlreiche Methoden entwickelt, die erfolgreich zur Annäherung der Schülerinnen und Schüler an ein Medium führen. So haben sich für das Medium Film Methoden wie die Filmanalyse oder das eigenständige Erstellen von Videoproduktionen vielfach bewährt, um die Verbindung zwischen Medium und Information bewusst zu machen. Diesem pädagogischen Ansatz folgt auch Soekia [DH03].

Die Arbeitsweise von algorithmischen Suchmaschinen bleibt den Nutzenden weitgehend verborgen: Das Erfassen von Webseiten durch Webroboter (Crawling / Spidering), das Erstellen einer effizienten Datenstruktur für die Suche (Indexierung), das Finden passen-

der Dokumente zu einer Benutzeranfrage (Matching) und die Präsentation der gefundenen Dokumente in einer guten Reihenfolge (Rangierung). Die didaktische Suchmaschine Soekia ermöglicht einen Blick hinter die Kulissen von Suchmaschinen. Schülerinnen und Schüler können beispielsweise selber den Index der Suchmaschine inspizieren. Sie können die erfassten Dokumente variieren und die Auswirkungen auf die Rangliste beobachten. Soekia stellt den Index in einer lesbaren Form dar. Abbildung 4 zeigt einen Ausschnitt eines Indexes. Neben jedem Begriff steht, wie oft der Begriff in der Kollektion vorkommt und in welchen Dokumenten er wie häufig auftritt. Mit dieser Index-Darstellung lassen sich zahlreiche Fragen untersuchen. Wie verändert sich der Index beim Hinzufügen von gleichartigen Dokumenten zu einer Dokumenten-Kollektion, wie bei artfremden Dokumenten? Wie wirkt sich Wort-Normalisierung auf den Index aus? Wie unterscheiden sich die Index-Einträge von allgemeinen und spezifischen Begriffen?

Abbildung 4: Screenshots aus Soekia

Soekia ist nicht gedacht zur Nutzung als Suchmaschine. Im Unterschied zu echten Suchmaschinen steht bei Soekia weder die Geschwindigkeit noch die Menge der erfassten Informationen im Vordergrund. Ganz bewusst enthält Soekia auch Schwachstellen, die den Lernenden eigene Entdeckungen erlauben sollen. „Vor- und Nachher-Tests“ haben gezeigt, dass Soekia das Verständnis für die Funktionsweise einer algorithmischen Suchmaschine wie Google bei den Lernenden fördert. Suchanfragen werden nach einer Unterrichtssequenz mit Soekia besser und zielgerichteter formuliert und die Effizienz und Effektivität bei Recherchen steigt merklich. Soekia steht zum kostenlosen Download auf dem Bildungsserver SwissEduc [DH03] samt ausführlichen Begleitmaterialien zur Verfügung.

Die Lernumgebung wird auf verschiedenen Schulstufen ab der Sekundarstufe bis hin zur Fachhochschule eingesetzt. Eine typische Unterrichtssequenz erstreckt sich über 4-6 Lektionen und als Unterrichtsmethode bietet sich Entdeckendes Lernen besonders an. In einer kurzen Einleitung wird erläutert, wie der Crawler einer Suchmaschine Webseiten erfasst. Diese Komponente von Suchmaschinen ist in Soekia nicht enthalten. Anschliessend erstellen die Schülerinnen und Schüler eine kleine Dokumentenkollektion mit einigen wenigen kurzen Webseiten. Es hat sich bewährt, selbst mit einem einfachen Texteditor ein paar Webseiten zu einem bestimmten Thema zu erstellen. Lädt man existierende Webseiten herunter, verliert man schnell den Überblick. Selbst erstellte Seiten lassen sich leichter modifizieren, z. B. einzelne Worte mehrfach aufführen und dann den Einfluss der Häufigkeit des Vorkommens eines Suchbegriffes in einem Dokument auf die Rangierung beobachten.

Durch Einbezug verschiedener Wortformen des gleichen Wortstammes in den betrachteten Webseiten können die Auswirkungen der Wortstammnormalisierung auf den Index und damit zusammenhängend die Ausbeute einer Suche beobachtet werden. Durch die Normalisierung reduziert sich die Anzahl der Einträge im Index. Ein Suchbegriff findet damit auch andere Wortformen des Suchbegriffes, das heisst die Ausbeute der Suche nimmt zu und als unerwünschter, aber typischer Nebeneffekt die Präzision ab. In den begleitenden Unterrichtsmaterialien auf dem Bildungsserver SwissEduc ist eine ganze Reihe solcher möglicher Entdeckungen und Einsichten durch die Lernenden beschrieben.

Als interessante Herausforderung zum Abschluss einer Unterrichtssequenz mit Soekia hat sich ein Web-Spam-Wettbewerb erwiesen. Die Schülerinnen und Schüler erhalten den Auftrag, eine einfache Webseite zu einem vorgegebenen Thema (z.B. eine Autovermietung) zu erstellen. Ziel ist es, dass am Schluss die eigene Webseite bei einer typischen Suchanfrage aus dem vorgegebenen Themenkreis in der Rangliste möglichst weit oben erscheint. Neben der spielerischen Seite vermittelt ein solcher Wettbewerb auch tiefer gehende Einsichten zur Suchmaschinenoptimierung von Webseiten.

5 Tagidex

Soekia bietet die Grundlage für das Verständnis von volltextbasierten, algorithmischen Suchsystemen. Aufgrund der großen Nutzung von Soekia in unterschiedlichen Lehrkontexten arbeiten wir an einer neuen Lernumgebung, welche den neuesten Webentwicklungen Rechnung trägt. So soll die erweiterte Version den Schülerinnen und Schülern die Unterschiede zwischen einer Volltextsuche und einer tagbasierten Suche veranschaulichen. Dazu wird eine Webapplikation Tagidex entwickelt, welche es erlaubt, auf einer vergleichbar überschaubaren Kollektion von echten Websites zum Themenbereich ‚Elektrizität' mit Tagging- und Indexsuchverfahren zu suchen. Ein Prototyp von Tagidex findet sich auf [HJ09]. Die Nutzenden können selbst wählen, mit welchem Retrievalsystem die Kollektion durchsucht werden soll. Bei der tagbasierten Suche werden die dazugehörigen Tags von Social-Tagging-Diensten wie Delicious bezogen. Beispielsweise kann derselbe Suchbegriff in die Volltextsuche und in die tagbasierte Suche eingegeben werden, die Ergebnislisten können miteinander verglichen und charakteristische Merk-

malsunterschiede herausgearbeitet werden. Auf welchem Platz steht eine von Nutzenden verschlagwortete Seite im Vergleich zur Indexsuche? Wurde der erste Treffer auf eine Anfrage mit der Indexsuche auch von Nutzenden verschlagwortet? Stimmen die Schlagwörter ungefähr mit den Erwartungen überein? Tagidex verdeutlicht die unterschiedlichen Funktionsweisen der Suchsysteme. Die Schülerinnen und Schüler sollen lernen, ihre Suchanfragen zu differenzieren. Durch das Ausprobieren verschiedener Suchbegriffe und -systeme können sie erfahren, welche Suchstrategien mehr oder weniger erfolgreich sind.

Abbildung 5: Screenshot einer Suchergebnisseite aus Tagidex

Tagidex ist als didaktisches Hilfsmittel konzipiert und nicht als Lernumgebung für das Selbststudium gedacht. Bevor im Unterricht Tagidex zum Einsatz gelangt, braucht es eine Einführung in die Grundlagen der Internetrecherche, der tag- und indexbasierten Suchsysteme. Die Funktionsweise von indexbasierten Suchsystemen kann mit Soekia vermittelt werden. Bei tagbasierten Diensten wie Delicious muss zuerst die Nutzung als Suchmaschine aufgezeigt werden. Erst anschliessend können mit Tagidex Vor- und Nachteile der verschiedenen Erschliessungsmethoden erkundet werden. Es ergeben sich Fragestellungen wie z.B.: Was sind mögliche Gründe für die unterschiedliche Rangierung von Treffern bei den beiden Suchsystemen? Welche Kriterien müssen bei der Auswahl von guten Suchbegriffen berücksichtigt werden? Die Lernenden erkennen so grundlegende Unterschiede: bei indexbasierten Suchmaschinen muss man Suchbegriffe wählen, die im gewünschten Dokument enthalten sind. Bei tagbasierter Erschliessung muss man sich mehr die Benutzenden vorstellen und welche Tags diese verwendet haben könnten. Aufgrund erster Erprobungen im Unterricht kann vermutet werden, dass sich die Benutzenden mit beiden Typen von Suchsystemen zu Recht finden. Wer schon über gute Recherchekompetenzen mit herkömmlichen Suchmaschinen verfügt, nutzt zudem auch Social-Bookmarking-Dienste effizienter.

6 Suchdienste nach dem Prinzip „Do it yourself“

Mit Soekia und Tagidex stehen für den Informatikunterricht Lernumgebungen zur Verfügung, mit denen sich Experimente rund um Suchdienste durchführen lassen, bei Soekia auch auf größeren Dokumentenkollektionen. So können etwa zuerst verschiedene Websites zu einem Themenbereich heruntergeladen und anschließend mit Soekia erschlossen werden. Beispielsweise könnten Websites verschiedener Autovermietungsfirmen gesammelt werden. Schülerinnen und Schüler könnten eine eigene Website für eine Autovermietungsfirma erstellen und dann die Rangierung im Vergleich zu den heruntergeladenen Websites untersuchen. Damit kann ein immer wieder auftretendes Problem im Informatikunterricht zumindest teilweise entschärft werden: Reale Informatiksysteme sind in der Regel äußerst komplex und entziehen sich den Möglichkeiten der Schule. In der Schule können nur kleine Systeme gebaut werden. Mit Soekia und Tagidex lassen sich zumindest große und reale Datenbestände im Unterricht einbinden.

Auf fortgeschrittener Stufe im Informatikunterricht wäre es spannend, selbst einen Suchdienst zu bauen. Die einfache Indexierung bei einer Volltextsuche, die Selektion der Resultate zu einer Suchanfrage und die Rangierung der Resultate nach einfachen Relevanzkriterien sind im Rahmen eines größeren Projektes durchaus auch für die Schule zugänglich. Es besteht die Möglichkeit, nur einzelne dieser Aspekte selbst umzusetzen und für andere auf vorgegebene Open-Source-Lösungen zurückzugreifen. Soekia selbst stützt sich auf die Suchmaschine Lucene von Apache ab. Der Quellcode von Soekia ist freigegeben unter der GNU GPL und könnte als Grundlage für weitere Entwicklungen dienen.

Das Erfassen von Websites in größerem Umfang übersteigt die Mittel im Informatikunterricht. Aber auch hier bieten sich inzwischen interessante Lösungen in Form von Web Services an. So macht beispielsweise Yahoo! im Rahmen von BOSS seinen Index allgemein zugänglich. Auch die Schule kommt damit in den Genuss der Infrastruktur von Yahoo! Die BOSS API ist einfach anzuwenden, die Resultate werden als XML-Antworten zurückgeliefert und können anschließend weiter bearbeitet werden. Wir sind überzeugt, dass der Informatikunterricht ganz allgemein vermehrt die heute zur Verfügung stehenden Web Services (Wetterdienste, News-Dienste usw.) und den damit verbundenen Zugang zu realen Daten nutzen sollte.

7 Zusammenfassung und Ausblick

Suchdienste im Internet werden im Informatikunterricht oft stiefväterlich behandelt. Sie eignen sich aber aus zwei Gründen hervorragend als Thema im Informatikunterricht:

1) Der Umgang mit Suchdiensten ist keineswegs trivial, sondern wird immer komplexer. Nur wer die Konzepte hinter den Suchdiensten versteht, kann diese auch gezielt nutzen. Die Vermittlung dieses Grundlagenwissens ist eine Aufgabe, die nur der Informatikunterricht wahrnehmen kann. Mit der Vermittlung von Informationskompetenz bewegt sich der Informatikunterricht in Richtung Web Science und Medienkunde. Das trägt zur Positionierung des Informatikunterrichtes in der Schullandschaft bei.

2) Am Beispiel von Suchdiensten lassen sich viele informatische Konzepte aufzeigen: effiziente Datenstrukturen für die Speicherung und Abfrage von großen Datenmengen. Algorithmen zur Selektion und Rangierung. Einbindung von Web Services als Schnittstelle zu großen realen Informatiksystemen.

Alle Beteiligten profitieren in mehrer Hinsicht. Suchdienste bieten ein Thema welches die Schülerinnen und Schüler anspricht, weil es einen direkten Anknüpfungspunkt an ihre Lebenswelt bietet. Den Informatiklehrenden wird damit ein Anker für den Unterrichteinstieg in die Hand gegeben. Für Lehrerinnen und Lehrer anderer Schulfächer ist Recherche und Informationskompetenz im Internet ein tägliches Thema. Aber es mangelt ihnen am nötigen Grundlagenwissen um selbst das Thema Informationskompetenz angemessen zu vermitteln. Hier kommt dem Informatikunterricht eine wichtige Rolle zu. Informatisches Wissen zu Suchdiensten würde die allgemeine IT-Kompetenz erhöhen und zahlreiche Arbeitsprozesse im Schulumfeld und Berufsalltag erleichtern.

Literaturverzeichnis

[Am98] American Library Association: Information Literacy Standards for Student Learning. Chicago, 1998

[Ar07] Arbeitskreis "Bildungsstandards": Grundsätze und Standards für die Informatik in der Schule. Beilage zu LOG IN, 146(147), 2007

[Ba97] Baacke, Dieter: Medienkompetenz. Tübingen, 1997

[DH03] Dreier, M.; Hartmann, W.: www.swisseduc.ch/informatik/soekia/ (Stand 08.01.2009)

[Fa05] Fallows, D.: Search Engine Users. Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve. PEW Internet & American Life Project, Washington, 2005

[Ga08] Ganz, J.F.: The Diverse and Exploding Digital Universe. An Updated Forecast of Worldwide Information Growth Through 2011. IDC White Paper. http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf (Stand 08.01.2009)

[Ge04] Gesellschaft für Informatik (Hrsg.): Digitale Spaltung verhindern – Schulinformatik stärken! Memorandum, Ulm, 2004

[HS08] Hendler, J.; Shadbolt, N.; Hall, W.; Berners-Lee, T.; Weitzner, D.: Web Science: An Interdisciplinary Approach to Understanding the Web. Communications of the ACM, 51 (7), 2008; S. 60-69

[HJ09] Hielscher, M.; Jurjević, D.: www.r2d2.ch/tagidex/ (Stand 08.01.2009)

[JS05] Jansen, B. J.; Spink, A.: How are we searching the World Wide Web? A comparison of nine large search engine transaction logs. Information Processing and Management, 42(1), 2005; S. 248-263

[LM08] Lewandowski, D.; Maass C. (Hrsg.): Web-2.0-Dienste als Ergänzung zu algorithmischen Suchmaschinen. Berlin, 2008

[Ma08] Manber, U.: 20 (Rare) Questions for Google Search Guru Udi Manber, http://www.popularmechanics.com/blogs/technology_news/4259137.html, 16. April 2008 (Stand 08.01.2009)

[Sa95] Saracevic, T.: Evaluation of evaluation in information retrieval. Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Special issue of SIGIR Forum, 1995; S. 138-146

[Sh07] Shneiderman, B.: Web science: A Provocative Invitation to Computer Science. Communications of the ACM, 50(6), 2007; S. 25-27