

Westfälische Wilhelms-Universität Münster
Institut für Informatik
Arbeitsbereich Didaktik der Informatik
Themensteller: Prof. Dr. Marco Thomas
Zweitgutachter: Prof. Dr. Achim Clauzing
Sommersemester 2011
Masterarbeit
Konzeption einer kontextorientierten Unterrichtsreihe
zum Thema Suchmaschinen
Philip Rehorst

Inhaltsverzeichnis

1. Einleitung	1
1.1. Motivation	1
1.2. Fragestellung	2
1.3. Aufbau der Arbeit	2
2. Bezug zur Fachdidaktik	3
2.1. Entwicklung	3
2.2. Informatik im Kontext	5
3. Sachanalyse – Suchmaschinen	9
3.1. Auffinden von Dokumenten (Crawling)	11
3.2. Automatische Textanalyse	12
3.3. Verarbeitung der Suchanfrage und Sortierung der Ergebnisse	16
3.4. Implikationen von Suchmaschinen	20
4. Verknüpfung mit dem Lehrplan	21
4.1. Richtlinien und Lehrpläne	22
4.2. Vorgaben zum Unterricht in der Sekundarstufe II	24
4.3. Bildungsstandards für die Sekundarstufe I	25
5. Beschreibung der Unterrichtsreihe	27
5.1. Überblick	27
5.2. UE 1 – Aufbau von Suchmaschinen	31
5.3. UE 2 – Durchsuche des WWW: Crawling	33
5.4. UE 3 – Automatische Textanalyse	35
5.5. UE 4 – Binäre Suchbäume	37
5.6. UE 5 – Binäre Suchbäume: Standardoperationen	40
5.7. UE 6 – Standardoperationen implementieren	42
5.8. UE 7 – Implementierung testen	44
5.9. UE 8 – Sortieralgorithmen	46
5.10. UE 9 – Ranking von Suchergebnissen	48
5.11. UE 10 – Informationsqualität im Internet	50
6. Bewertung der Unterrichtsreihe	52
6.1. Rückmeldung von Koubek	52
6.2. Stellungnahme zur Kritik von Koubek	53
6.3. Rückmeldung von Engbring	54
6.4. Stellungnahme zur Kritik von Engbring	55

7. Überarbeitung der Unterrichtsreihe	55
7.1. Alternative Gestaltung einzelner Einheiten	56
7.2. Weitere Unterrichtseinheiten	59
8. Fazit und Ausblick	64
Literatur	67
A. Tabellarische Unterrichtsentwürfe	71
A.1. UE 1 – Aufbau von Suchmaschinen	71
A.2. UE 2 – Durchsuchen des WWW: Crawling	72
A.3. UE 3 – Automatische Textanalyse	73
A.4. UE 4 – Binäre Suchbäume	74
A.5. UE 5 – Binäre Suchbäume: Standardoperationen	76
A.6. UE 6 – Standardoperationen implementieren	77
A.7. UE 7 – Implementierung testen	78
A.8. UE 8 – Sortieralgorithmen	79
A.9. UE 9 – Ranking von Suchergebnissen	80
A.10. UE 10 – Informationsqualität im Internet	81
A.11. Überarbeitet: UE 2 – Durchsuchen des WWW: Crawling	82
A.12. Überarbeitet: UE 3 – Automatische Textanalyse	83
A.13. Umweltschutz durch effiziente Algorithmen	84
A.14. Ranking 2 – Wer bestimmt was ich finden kann?	85
B. Material	86
B.1. UE 1 – Aufbau von Suchmaschinen	86
B.2. UE 9 – Ranking von Suchergebnissen	92
B.3. UE 10 – Informationsqualität im Internet	95
C. Rückmeldungen	98
C.1. Prof. Dr. Jochen Koubek	98
C.2. Dr. Dieter Engbring	98
D. Erklärung des Studierenden	100

1. Einleitung

1.1. Motivation

„Wer suchet, der findet“ – so lautet ein bekanntes Sprichwort. Suchen ist allerdings, besonders im Zusammenhang mit den immer stärker wachsenden Datenmengen, schon lange keine triviale Aufgabe mehr. Im World Wide Web existieren Unmengen an Dokumenten mit Inhalten, die auf irgendeine Art verschlossen werden müssen, um nicht nur der Verfasserin¹ zugänglich zu sein.

Hinter der puristischen Webseite von Google steckt ein großes und vielfach auch undurchschaubares Informatiksystem. Einzelne Teile dieses Systems basieren auf grundlegenden algorithmischen Problemen, wie Suchen und Sortieren. Die gezielte Suche nach Informationen ist ein fundamentaler Bestandteil der Lebenswelt (fast) aller Menschen. Ob im Internet nach Texten zu bestimmten Themen, im sozialen Netzwerk nach Freundinnen und Bekannten oder auf der Telefonliste nach einer Klassenkameradin gesucht wird, ändert nichts daran, dass es sich um den Versuch handelt, aus einer größeren Datenmenge eine benötigte Information zu ermitteln. Nicht minder schwierig ist es, aus den angebotenen Informationen diejenigen herauszufiltern, welche die benötigten Inhalte entsprechend dem aktuellen Stand der Wissenschaft enthalten und nicht nur das Wissen der Autorin widerspiegeln. Sascha Lobo schreibt dazu in seiner Kolumne bei Spiegel Online:

Die Gesellschaft muss auf das anwachsende, aufgeblähte, verschwörungsdurchseuchte Wissensarchiv, auf die Query-Realität, auf die unbewusste und algorithmische Filterung der Informationen reagieren. Ein bisher an deutschen Schulen vernachlässigtes Fach wird zur Schlüsselqualifikation: das Wissen, wie Wissen zustande kommt, die Epistemologie. (Lobo 2011)

Aufgrund der fachwissenschaftlichen Tiefe und der lebensweltlichen Relevanz des Themenkomplexes Suchmaschinen liegt die Behandlung von Suchmaschinen im Rahmen von *Informatik im Kontext* nah. *Informatik im Kontext*² ist

¹In den Kapiteln 1, 2 und 8 findet das generische Femininum, in den Kapiteln 3, 4 und 6 das generische Maskulinum Verwendung. Kapitel 5 und 7 sollen auch für sich allein stehen können, daher werden dort überwiegend sowohl die weibliche, als auch die männliche Form verwendet. Zu jeder Zeit sind jedoch alle Geschlechter gleichermaßen angesprochen.

²Fortan: IniK.

ein Konzept zur Gestaltung von Informatikunterricht, welches sich an drei wesentlichen Punkten orientiert:

1. Orientierung an Kontexten
2. Orientierung an Standards für Informatik in der Schule
3. Methodenvielfalt

Mit einem Unterricht, der diesen Ansprüchen genügt, sollen die Schülerinnen durch den Lebensweltbezug besonders motiviert werden und dennoch den Bezug zur fachwissenschaftlichen Systematik nicht verlieren.

1.2. Fragestellung

Die vorliegende Arbeit versucht, die Frage zu beantworten, wie sehr sich das Thema *Suchmaschinen* als Kontext im Sinne von IniK eignet. Das Ziel ist, eine Reihe vorzustellen, die im Ansatz bewertet ist und somit bereitsteht, um von Lehrerinnen erweitert und angepasst zu werden, und die im Anschluss daran an Schulklassen oder Kursen getestet werden kann.

Aufgrund der vielfältigen fachwissenschaftlichen Möglichkeiten des Themenbereichs Suchmaschinen eignet sich der Kontext m. E. für die Sekundarstufe II. Bei der Konzeption einer Unterrichtsreihe für die Sekundarstufe II muss, besonders unter Beachtung der verkürzten Schulzeit durch G-8, darauf geachtet werden, dass die vermittelten Inhalte im Hinblick auf die Abiturprüfung relevant sind. Es muss demnach eine strenge Orientierung an den Lehrplänen und an den Vorgaben für das Zentralabitur gegeben sein. Eine Unterrichtsreihe, die den Ansprüchen von IniK ansonsten gerecht wird, könnte sich sonst dennoch im Unterricht der Sekundarstufe II als undurchführbar erweisen.

Die Bewertung soll durch Experten für Informatikunterricht, also Informatiklehrer und Fachdidaktiker, insbesondere durch einen der Initiatoren von IniK, erfolgen.

1.3. Aufbau der Arbeit

Im Zuge dieser Arbeit wird zuerst das fachdidaktische Konzept IniK kurz beschrieben und in den fachdidaktischen Zusammenhang eingeordnet. Daran

schließt sich eine Beschreibung des Kontextes *Suchmaschinen* an. Daraus ergeben sich neben den fachwissenschaftlich orientierten Schwerpunkten auch Themenbereiche aus anderen Fachdisziplinen sowie an der Grenze zu anderen Disziplinen liegende informative Fragestellungen. Darauf baut eine Einordnung der Inhalte in die Vorgaben des Lehrplans für Nordrhein-Westfalen für das Fach Informatik in der Sekundarstufe II unter Bezugnahme auf die Vorgaben für das Zentralabitur in Nordrhein-Westfalen auf.

Es folgt dann eine Unterrichtsreihe, mit der die einzelnen Inhalte in modularer Form vermittelt werden können. Zu dieser Reihe gehören Materialien, die einen methodisch vielfältigen Unterricht im Sinne von IniK ermöglichen sollen. Abschließend erfolgt die Bewertung der Unterrichtsreihe durch die Analyse von Expertenbefragungen und eine Bearbeitung der Reihe, welche die Bewertung berücksichtigt.

2. Bezug zur Fachdidaktik

Das folgende Kapitel gibt einen kurzen Überblick über das fachdidaktische Konzept *Informatik im Kontext*.

2.1. Entwicklung

IniK wurde als Konzept für den Informatikunterricht bei den Königsteiner Gesprächen 2008 von Jochen Koubek und Helmut Witten vorgestellt. Unterrichtskonzepte dieser Art gibt es schon seit einiger Zeit in den Naturwissenschaften.

Die Idee eines lebensweltlich orientierten Informatikunterrichts gab es ebenso schon zuvor.

2.1.1. Fachdidaktik Informatik

Als Alternative zum bis dahin stark algorithmenorientierten Informatikunterricht entstand zu Beginn der 1980er Jahre der anwendungsorientierte Ansatz. Die Algorithmen sollten hier nicht mehr den Kern des Unterrichts darstellen, sondern nur noch einen notwendigen Teil. Der lebensweltliche Bezug sollte

durch die Analyse und Lösung von tatsächlichen Problemen aus der Lebenswelt der Schülerinnen hergestellt werden.

Die Lösung des Problems erfolgt dann in einem mehrschrittigen Verfahren, welches stark an das Wasserfallmodell im Software-Engineering angelehnt ist. Dabei steht am Ende eine Benutzungsphase, in der wieder auf die ursprüngliche Problemstellung Bezug genommen werden soll.

In der praktischen Umsetzung ist der Rückbezug zum Kontext nur schwer zu realisieren gewesen. Die bearbeiteten Probleme sind entweder viel zu komplex, um in der Schule von nur einer Person und nicht von einem Team aus Expertinnen bearbeitet zu werden, oder das Problem ist sinnvoller ohne die Erstellung bzw. Bearbeitung eines Computerprogramms lösbar, sodass die Arbeit schon mit der Problemanalyse endet.³

2.1.2. Naturwissenschaften

In den Naturwissenschaften Biologie, Physik und Chemie gibt es schon seit einiger Zeit Konzepte, welche die Anwendungsbereiche der Wissenschaft mit den Kontexten im Alltag in Verbindung bringen. Diese sind entstanden, weil die fachwissenschaftlichen Inhalte und ihre Vermittlung im Aufbau der Fachsystematik nicht mit den Erwartungen und der Motivation der Schülerinnen in Einklang gebracht werden konnten. Möglicherweise erklärt sich daraus die erschreckend geringe Anzahl der Teilnehmerinnen an entsprechenden Kursen in der gymnasialen Oberstufe, was der Aufnahme eines naturwissenschaftlichen Studiums durch zahlreiche Abiturientinnen wohl nicht förderlich ist.

In den drei genannten Naturwissenschaften gibt es heute fachdidaktische Konzepte, die dieser Problematik entgegenwirken sollen. Zwar sind die Ausprägungen und das Verständnis von *Kontext* teilweise unterschiedlich, jedoch besteht Übereinstimmung darin, dass im Kern Inhalte anhand eines Kontextes vermittelt werden. Dieser Kontext stellt einen roten Faden dar, der sich durch die gesamte Unterrichtsreihe zieht und nicht nur als Aufhänger fungiert.⁴

³Vgl. Forneck 1992, S. 179–230.

⁴Vgl. Koubek et al. 2009, S. 269–270.

2.2. Informatik im Kontext

IniK soll zu einem fachlich fundierten Informatikunterricht führen, der sich weder starr an der Fachsystematik orientiert und die Inhalte mit Hilfe kleiner Puzzles oder Rätsel ähnlichen Aufgaben an die Schülerinnen vermittelt, noch eine reine Anwenderschulung ist. Die Schülerinnen sollen die Bedeutung der Informatik im Alltag erkennen, bewerten und die Abläufe verstehen können.

Um eine hohe Qualität der Unterrichtsreihen zu erreichen, folgen die Entwürfe für IniK drei Prinzipien: Der Kontext muss für die gesamte Reihe sinnstiftend sein, ein reiner Aufhänger ist nicht ausreichend.⁵ Um die inhaltliche Qualität zu sichern, sind die Standards für Informatikunterricht in der Sekundarstufe I maßgeblich. Nicht zuletzt sollen verschiedene Unterrichtsmethoden zur Anwendung kommen, damit der Unterricht vielfältig und motivierend bleibt.

Da noch keine Standards für den Informatikunterricht in der Sekundarstufe II vorliegen, wird sich in dieser Arbeit sowohl auf die Standards für die Sekundarstufe I als auch auf die Vorgaben für den Unterricht in der gymnasialen Oberstufe in Nordrhein-Westfalen bezogen.

2.2.1. Kontextorientierung

Ein Kontext darf nicht mit einem sogenannten Aufhänger oder Einstiegsproblem verwechselt werden. Der Kontext wird nicht unmittelbar nach Beginn der Reihe wieder zur Seite gelegt, sondern zieht sich, wie bereits angedeutet, erkennbar durch alle Unterrichtsstunden der Reihe. Als Kontext eignen sich viele alltägliche Situationen, in denen die Schülerinnen mit Informatiksystemen mittelbar oder unmittelbar in Berührung kommen. Bereits umgesetzte Beispiele sind unter anderem E-Mail und Verschlüsselungen, Chatbots oder der Bahnverkehr. Dabei meint Kontext hier nicht nur eine einzelne Situation, sondern durchaus mehrere zusammengehörende Situationen, die mit dem Thema in engem Zusammenhang stehen. Die Kontexte sollten sich, zumindest teilweise, im Erfahrungshorizont der Schülerinnen wiederfinden.

⁵Vgl. Koubek et al. 2009, S. 271–275.

Bei der Bearbeitung des Kontextes sollen auch Themen aufgegriffen werden, die nicht zum Kerngeschäft der Fachwissenschaft Informatik gehören. Beispielsweise sollen ethische oder moralische Komponenten von Informatiksystemen ebenso untersucht werden, wie die informatischen Inhalte, die der Funktionalität zu Grunde liegen. Gerade im Rahmen der Softwareentwicklung gehört die Beschäftigung mit dem Einsatzgebiet grundsätzlich zum Aufgabengebiet einer Informatikerin. Engbring und Pasternak (2010) schreiben dazu:

Der Bezug auf den Kontext liegt für die Informatik nahe, da sie auch in ihren [sic!] fachsystematischen Handeln (vor allem bei der Software-Entwicklung), darauf angewiesen ist, dass der Anwendungsbereich einer eingehenden Analyse unterzogen wird.

Im Rahmen der Dekontextualisierung wird dann vom konkreten, beispielhaften Kontext auf das allgemeine fachwissenschaftliche Prinzip, welches im Kontext zum Einsatz kommt, übergeleitet. Dabei darf der Kontext nicht aus den Augen verloren werden. Der Bezug zum ursprünglichen und zu anderen Kontexten – im Rahmen der Rekontextualisierung – darf nicht verloren gehen.⁶

Zur Abgrenzung von der Anwendungsorientierung schreiben Koubek et al. (2009):

Die Anwendungsorientierung geht von den fachlichen Inhalten aus, die durch Kontexte angereichert werden. Die Kontextorientierung geht im Gegensatz dazu von den Kontexten aus, die den Unterricht strukturieren.

2.2.2. Standardorientierung

Es ist eine Kernaufgabe von Schule allgemeingültiges und dauerhaft relevantes zu lehren. Die Orientierung an Kontexten soll also nicht dazu führen, dass über die Darstellung und das Erleben des Exemplarischen der Anspruch an die Weiterverwertbarkeit des Gelernten vergessen wird.

Seit 2008 gibt es die Bildungsstandards der Gesellschaft für Informatik e. V. als Empfehlung, welche Mindeststandards für die Sekundarstufe I darstellen.

⁶Vgl. Koubek et al. 2009, S. 272–274.

Unterrichtsreihen für IniK orientieren sich an diesen Standards und legen in ihrer Beschreibung jeweils dar, welche der in den Bildungsstandards genannten Kompetenzen und Inhaltsbereiche mit einer Unterrichtseinheit gefördert bzw. erlernt werden sollen.

Zum aktuellen Zeitpunkt bestehen die Bildungsstandards nur für die Sekundarstufe I. Aus diesem Grund sind auch die meisten der bis jetzt veröffentlichten Unterrichtsentwürfe für diese Stufe konzipiert. Für die Einordnung des Unterrichtsentwurfes werden daher zusätzlich zu den Bildungsstandards auch die Vorgaben für den Unterricht in der gymnasialen Oberstufe in Nordrhein-Westfalen herangezogen. Diese Vorgaben stellen die inhaltlichen Voraussetzungen für ein erfolgreiches Ablegen der Abiturprüfung im Rahmen des Zentralabiturs in Nordrhein-Westfalen dar.

Bei der Einführung des Zentralabiturs in Nordrhein-Westfalen haben die Vorgaben für den Unterricht viel Kritik erfahren. Mittlerweile sind diese aber überarbeitet worden und spiegeln einen wichtigen Teil der fachwissenschaftlichen Inhalte der Informatik wider. Der Bereich *Informatik und Gesellschaft* wird leider nach wie vor nicht berücksichtigt.

2.2.3. Methodenvielfalt

Ein weiterer Gesichtspunkt bei der Gestaltung von kontextorientierten Unterrichtsreihen ist die anzustrebende Methodenvielfalt.

Die Umgestaltung des Unterrichts zu einem stärkeren lebensweltlichen Bezug sollte aber nicht nur inhaltlich erfolgen, sondern die Möglichkeit der stärkeren Partizipation der Schüler berücksichtigen.

(Koubek et al. 2009, S. 275)

Beispielsweise ist der Einsatz von verschiedenen Medien besonders im Informatikunterricht durch eine generell höhere Aufgeschlossenheit gegenüber den sogenannten *Neuen Medien* gut realisierbar. Ebenso findet sich besonders im Informatikunterricht häufig ein sinnvolles Einsatzfeld von projektorientiertem Unterricht. Hierbei sollte allerdings berücksichtigt werden, dass Kontextorientierung keinesfalls mit Projektorientierung gleichzusetzen ist.⁷ Softwareprojekte, wie sie in der Wirtschaft umgesetzt werden, sind nicht eins zu eins in ein

⁷Vgl. Koubek et al. 2009, S. 271.

Unterrichtsprojekt zu übersetzen. Dies wird zum Beispiel am wesentlich höheren Stellenwert des Produktes in einem Softwareprojekt deutlich. Ein Unterrichtsprojekt kann sehr wohl äußerst gelungen sein, auch wenn am Ende kein nutzbares Produkt entstanden ist. Ein Softwareprojekt würde in dieser Situation spätestens von der Kundin für gescheitert erklärt werden.

Um den Kontext zu erkunden, lassen sich beispielsweise arbeitsteilig angelegte Methoden wie das Gruppenpuzzle oder die auf Kooperation zielende Methode Think-Pair-Share verwenden. Humbert (2006) zeigt eine Reihe von Beispielen für Unterrichtsmethoden, die im Informatikunterricht angewandt werden können.

Je nach Kontext sind auch Anleihen bei typischen Methoden der angrenzenden Fächer zielführend. Wenn, wie in dieser Arbeit, ein Bezug zu sozialwissenschaftlichen Themen – Stichwort *Informatik und Gesellschaft* – hergestellt wird, können Methoden wie das Rollenspiel, Debatten oder Podiumsdiskussionen zur Anwendung kommen.

2.2.4. Kritische Anmerkungen

Das Konzept IniK ist in dieser konkreten Form noch vergleichsweise jung. Engbring und Pasternak (2010) bemerken dazu, dass es sich 2010 bei IniK im Wesentlichen um eine Sammlung von Unterrichtsbeispielen gehandelt hat, denen ein umfassendes theoretisches Fundament fehlt. Im Folgenden werden einige Probleme und kritische Anmerkungen aufgeführt, die in diesem Zusammenhang zu beachten sind.

Wie oben bereits erwähnt, ist es notwendig, den Kontext im gesamten Verlauf der Reihe nicht aus den Augen zu verlieren. Diese Gefahr besteht besonders im Zusammenhang mit der Dekontextualisierung. Wenn von den konkreten Problemen, die im Kontext erkannt werden, auf die allgemeinen Probleme der Fachwissenschaft geschlossen wird, könnten bei fehlendem Bezug zum Kontext einzelne Schülerinnen zurückgelassen werden.

Das Risiko liegt darin, dass dabei jene Teile der Klasse zurück gelassen werden, die in den Prinzipien das Problem nicht mehr wieder erkennen und die Bäume vor lauter Wald nicht mehr sehen. (Koubek et al. 2009, S. 273)

Ebenso darf die Fachsystematik nicht vernachlässigt werden. Pasternak und Vahrenhold (2009) schreiben dazu:

Es reicht nicht aus, bei einem Unterricht nach dem Konzept ‚Informatik im Kontext‘ das Hauptaugenmerk auf die rein lebensweltlichen Kontexte zu legen, sondern es müssen gleich gewichtig die Fachinhalte einschließlich [sic!] der Prozesse berücksichtigt werden. Nur so lässt sich vermeiden, dass der Informatik-Unterricht zwar von der Kontextorientierung ausgeht, jedoch wieder zur Produktorientierung zurückführt.

Insbesondere eine Anwenderschulung, wie sie aus den Ideen des anwendungsorientierten Ansatzes in den 1980er Jahren entstanden ist, ist nicht im Sinne eines allgemeinbildenden Informatikunterrichts.

Trotz der Beachtung der nicht kerninformatischen Inhalte ist das Zentrum des Unterrichts in der Informatik angesiedelt. Im Informatikunterricht sollten nicht die Inhalte anderer Fächer unterrichtet werden, vor allem, wenn dabei die vorhandene Unterrichtszeit für informative Inhalte nicht mehr zur Verfügung steht. Optimal wäre an dieser Stelle ein fächerverbindender Unterricht. „Auf diese Unterrichtsform sind aber weder die Lehrpläne noch die Unterrichtsmaterialien noch die Kollegien vorbereitet.“ (Koubek et al. 2009, S. 272) Hier wird m. E. ein Widerspruch zwischen dem Anspruch, die Fachsystematik nicht aus dem Blick zu verlieren, und der Fokussierung auf den Kontext deutlich. Beide Aspekte müssen bei der Konstruktion von Unterrichtseinheiten berücksichtigt werden.

Bei der Auswahl der Kontexte dürfen die persönlichen Interessen der Lehrerin nur eine untergeordnete Rolle spielen. Im Zentrum der Überlegungen sollte ein Informatikunterricht stehen, der den Vorgaben der Standards und der Lehrpläne gerecht wird.

3. Sachanalyse – Suchmaschinen

Im folgenden Kapitel werden die grundlegenden Elemente des Informatiksystems *Suchmaschine* erläutert.

Eine wichtige Eigenschaft von Suchmaschinen sehen wir am Beispiel des Marktführers Google. Bereits direkt nach der Gründung fiel Google durch ei-

ne schlichte Seite auf. Mehr als ein Logo, das Google-Doodle⁸, ein Textfeld und zwei Knöpfe, um die Anfrage zu starten, waren nicht sichtbar. Auch wenn andere Suchmaschinen mit weniger aufgeräumten Seiten aktiv sind, so ist nach wie vor der Kern der Suchmaschine nicht unmittelbar erkennbar. Bei anderen Suchmaschinen gibt es mehr Werbeanzeigen und Zusatzangebote, doch auch hier gilt: Eine Suchmaschine verhält sich wie ein Eisberg, der größte und wichtigste Teil versteckt sich unter der Oberfläche. Innerhalb dieses nicht sichtbaren Teils gibt es eine Reihe von Teilsystemen, die ineinander greifen. Diese Systeme lassen sich grundsätzlich in drei Teile mit drei Aufgabenbereichen trennen.

Der *Crawler* hat die Aufgabe, neue oder veränderte Seiten im World Wide Web zu finden und bei Bedarf herunterzuladen. Die geladene Seite wird auf weiterführende Links untersucht. Auf diese Weise ‚krabbelt‘ der Crawler förmlich von Seite zu Seite.⁹

In einem zweiten Schritt erfolgt die eigentliche Analyse des heruntergeladenen Dokuments. Das Ziel in dieser Phase ist der Aufbau eines Index, der den Inhalt des Dokuments möglichst genau widerspiegeln soll.

Für die Benutzung schließlich müssen eingehende Suchanfragen bearbeitet und die dazu passenden Ergebnisse in eine Reihenfolge gebracht werden. Aufgrund der potentiell großen Anzahl der Suchergebnisse, ist der Sortierungsschritt besonders wichtig. Beim Ranking unterscheiden sich die Anbieter verschiedener Suchmaschinen. Googles großer Erfolg beruhte zumindest in der Anfangszeit neben der einfachen Oberfläche auf der Einführung eines neuen Algorithmus zum Ranking der Ergebnisse.¹⁰ Im Folgenden wird im Detail auf diese drei Teilsysteme einer Suchmaschine eingegangen.

⁸Das Google-Logo, welches bei verschiedenen Gelegenheiten in abgewandelter Form gezeichnet wird, wird als Doodle bezeichnet. Am heutigen 21. Juni wird beispielsweise auf der Nordhalbkugel ein Doodle zum Thema Sommersonnenwende gezeigt, auf der Südhalbkugel ein Wintersonnenwenden-Doodle.

⁹Vgl. beispielsweise Glöggler 2003, S. 25ff.

¹⁰S. Abschnitt 3.3.3, Seite 18.

3.1. Auffinden von Dokumenten (Crawling)

In der Struktur des WWW und in der Anforderung, ohne nähere Anhaltspunkte im Netz Dokumente finden zu können, liegt die Begründung für die Notwendigkeit eines Crawlers¹¹. Es gibt kein zentrales Register, in dem eine neue oder veränderte Webseite eingetragen wird. Typischerweise gibt es aber eine Reihe von Verknüpfungen, genannt Links, zwischen verschiedenen Webseiten. Links sind unidirektional angelegt, auch wenn es vermutlich häufig gegenseitige Verlinkungen zwischen zwei Seiten gibt. Das WWW kann daher als gerichteter Graph modelliert werden. Um in einem zusammenhängenden Teil dieses Graphen Webseiten zu finden kann eine Breiten- oder Tiefensuche durchgeführt werden. Üblicherweise wird auf die Breitensuche zurückgegriffen, da diese schneller relevante Webseiten findet.¹² Da von den gängigen Suchmaschinen nur ein geringer Anteil der existierenden Seiten durchsucht wird, ist es unabdingbar, dass diejenigen Seiten, die berücksichtigt werden, von höherer Relevanz sind, als diejenigen, die nicht beachtet werden. Aufgrund des hohen Aufwands lassen sich nicht alle erreichbaren Seiten durchsuchen. In der Praxis existieren eine Reihe von Aufgaben die bei der Konstruktion eines Crawlers beachtet werden müssen, auf die ich an dieser Stelle aber nicht weiter eingehe.¹³

Eine erste Untersuchung der Dokumente findet bereits beim Crawling statt. Beispielsweise müssen die Links aus dem Dokument extrahiert werden. Auch das Ermitteln von Duplikaten kann bereits an dieser Stelle sinnvoll sein. Besonderer Aufmerksamkeit bedürfen hier Dokumente mit Adressen, die offensichtlich dynamische Inhalte vermuten lassen. Wenn Daten per HTTP-GET an den Webserver übergeben werden, liegt ein dynamischer Inhalt nahe. Falls das Dokument bereits zuvor besucht wurde, muss zusätzlich festgestellt werden, ob eine Änderung vorliegt.

Eine weitere Herausforderung für einen Crawler, gerade bei dynamisch erstellten Webseiten, ist die Nichterreichbarkeit mancher Dokumente. Dies betrifft sowohl die angesprochenen dynamischen Seiten, als auch besonders ge-

¹¹ Andere gebräuchliche Begriffe sind Spider oder (Web-)Robot.

¹² Vgl. Najork und Wiener 2001.

¹³ Vgl. Olston 2010, S. 178–179.

schützte Bereiche oder Webseiten, die keine eingehenden Verweise besitzen. Diese Teile des WWW werden auch als Deep-Web oder Invisible Web bezeichnet.¹⁴

Glöggler (2003) fasst die Funktion eines Crawlers wie folgt zusammen: „Ein Webrobot ist diejenige Systemkomponente der Suchmaschine, die für die Erfassung von neuen und veränderten Ressourcen im Internet verantwortlich ist.“

3.2. Automatische Textanalyse

Die nun vorliegenden Dokumente müssen, damit die Menge aller Dokumente effizient durchsucht werden kann, aufbereitet werden. Zu dieser Analyse bzw. Aufbereitung gehören verschiedene Schritte, die im Folgenden erläutert werden. Es gibt selbstverständlich eine Vielzahl von Dokumenten, welche mehr oder weniger gut automatisch analysiert werden können. Für ein Verständnis der grundlegenden Funktionalität einer Suchmaschine ist allerdings die Be trachtung von Texten meines Erachtens ausreichend. Daher befasst sich dieser Abschnitt nur mit Textdokumenten.

Wenn im Folgenden von Wörtern gesprochen wird, so geschieht dies unter der Annahme, dass der zu analysierende Text zuvor in einzelne Wörter aufgeteilt wurde.

3.2.1. Text bereinigen

Eine Aufgabe im Zuge der Analyse ist das Entfernen von Teilen des Dokuments, die kein Text sind und die auch keine weitere Relevanz für die Analyse besitzen. Hierzu gehören bei Webseiten zum Beispiel Scripte oder HTML-Tags, die keine Aussage über die inhaltliche Bedeutung eines Wortes machen. Unsichtbare Elemente und Bilder oder andere eingebettete Dokumente fallen auch in diese Kategorie. Zu beachten ist dabei, dass strukturierende Elemente der Webseite nicht verloren gehen sollten, da durch diese die inhaltliche Relevanz eines Wortes ermittelt werden kann. Beispielsweise gibt ein Wort aus

¹⁴Vgl. Lewandowski 2005, S. 139.

einer Überschrift mit hoher Wahrscheinlichkeit den Inhalt eines Textes vergleichsweise gut wieder. Auch die Wahl einer größeren Schriftart oder -stärke sollte, genauso wie die Position im Text, berücksichtigt werden.¹⁵ Informationen dieser Art müssen mit den verbliebenen Wörtern gespeichert werden und können bei der Gewichtung als Indexterm und somit schließlich beim Ranking eine Rolle spielen.

3.2.2. Stoppwörter

In einem typischen Text sind etwa 40 Prozent der Wörter ohne inhaltliche Bedeutung.¹⁶ Wenn ein bestimmtes Wort in annähernd jedem Text vorkommt, so kann anhand dieses Wortes keine Unterscheidung zwischen den Texten erfolgen. Diese Wörter werden *Stoppwörter* genannt und vor der Erstellung des Index entfernt. Zur Menge dieser Wörter gehören neben anderen zum Beispiel bestimmte und unbestimmte Artikel, Konjunktionen, Präpositionen und Pronomen. Typische Stoppwörter haben primär oder ausschließlich eine syntaktische Funktion.¹⁷

3.2.3. Inhaltlich relevante Wörter

Eine andere Klasse von Wörtern sind diejenigen Wörter, die in einem Text häufig vorkommen, jedoch selten in anderen Texten zu finden sind. Diese Wörter besitzen eine hohe inhaltliche Relevanz. Zur Messung dieser beiden Faktoren werden die Termfrequenz und die inverse Dokumentenhäufigkeit berechnet. Die Idee, die Relevanz von Wörtern über statistische Eigenschaften zu bestimmen, geht auf Hans Peter Luhn zurück.

Luhn ging es zunächst um einen Ansatz für automatisch generierte Abstracts, wenn er anschließend über die ermittelten Worthäufigkeiten auf die Signifikanz von Sätzen abhebt. Seine Grundannahme, die Signifikanz von Wörtern für die Bedeutung eines Textes auf statistischem Wege ermitteln zu können, wurde in der Folgezeit jedoch in der Hauptsache als eine Ausgangsthese für eine Indexierung gewählt. Für die statistischen

¹⁵Vgl. Stock 2007, S. 323–324.

¹⁶Vgl. Nohr 2005, S. 37.

¹⁷Vgl. auch Stock 2007, S. 222–226.

Verfahren der Indexierung ist sie es bis heute geblieben. (Nohr 2005, S. 43)¹⁸

Die Termfrequenz (TF) berechnet sich wie folgt:

$$TF = \frac{(\text{Häufigkeit eines Terms im Dokument})}{(\text{Anzahl der Wörter im Dokument})}$$

Eine hohe Termfrequenz spricht für eine hohe Bedeutung des Wortes. Wenn ein Wort in jedem Dokument häufig vorkommt, so ist die inhaltliche Bedeutung gering. Andersherum beschreibt ein Wort, welches nur in wenigen Dokumenten vorkommt, den Inhalt des Dokumentes sehr gut. Aus diesem Grund wird die Inverse Dokumentenhäufigkeit (IDF) berechnet. Die zugehörige Formel lautet:

$$IDF = \log_2 \frac{(\text{Gesamtzahl der Dokumente in der Datenbank})}{(\text{Anzahl der Dokumente, in denen der Term vorkommt})}$$

Diese beiden Werte werden häufig als Produkt in Kombination für die Gewichtung herangezogen.¹⁹

3.2.4. Index erstellen

Aus den verbliebenen Wörtern wird jetzt ein Index erstellt. Ein simpler Ansatz ist die Volltextindexierung, welche jedes Wort in der Form, in der es im Text vorkommt, in einen Index aufnimmt. Dieser Ansatz ist aus Sicht eines Entwicklers sehr einfach umzusetzen. Ob die Volltextindexierung bereits als automatisches Verfahren zur Indexierung bezeichnet werden kann, ist strittig. Dennoch kommt die Volltextindexierung in der Praxis sehr häufig vor,²⁰ da sie sehr einfach zu implementieren ist. Die Problematik verschiedener Flexionsformen wird bei einer reinen Volltextindexierung auf die Seite des Nutzers geschoben, welcher die verschiedenen Formen möglichst vollständig in seiner Anfrage beachten muss. Empirische Untersuchungen belegen die Unterle-

¹⁸Vgl. Luhn 1958.

¹⁹Vgl. Stock 2007, S. 326.

²⁰Vgl. Nohr 2005, S. 38.

genheit einer reinen Volltextindexierung.²¹ In diesem Zusammenhang schreibt Nohr (2005, S. 59) zu den in Abschnitt 3.2.3 genannten Verfahren:

Die statistischen Verfahren sollten wenigstens Teillösungen der informationslinguistischen Ansätze berücksichtigen (wie bspw. eine Grundformenreduktion), da in der Folge die Termfrequenzermittlung zu besseren Ergebnissen kommen kann.

Durch eine Reduzierung auf Wortstämme lässt sich die Volltextindexierung bereits deutlich verbessern. Die deutsche Sprache ist eine Sprache mit vielen verschiedenen Flexionsformen. Unabhängig von der Wortform ist jedoch die inhaltliche Bedeutung dieselbe. Es ist also von Vorteil, wenn beispielsweise die Wörter „Wohnmobil“, „Wohnmobile“ und „Wohnmobilen“ zu einem gemeinsamen Indexterm zusammengefasst werden. Zur sinnvollen Verarbeitung ist es weiterhin notwendig, die Sprache des Textes zu kennen. Eine automatische Ermittlung der Dokumentsprache ist relativ zuverlässig möglich.²² Weiterhin sollte eine Rechtschreibkorrektur stattfinden, damit nicht aus Tipp- und Rechtschreibfehlern neue Indexterme entstehen, die so die Gewichtung beeinflussen.

Neben der oben bereits genannten Entfernung der Stopwörter und der Berücksichtigung verschiedener Flexionsformen sollten während einer computerlinguistischen Bearbeitung mit dem Ziel der Indexerstellung Mehrwortbegriffe (Phrasierungen) als zusammengehörend erkannt werden. Außerdem sollten zusammengesetzte Wörter (Komposita) in ihre Bestandteile zerlegt werden.²³

Am Ende der Verarbeitung steht der Aufbau eines invertierten Index. Im Gegensatz zum Index, welcher auf das Vorkommen innerhalb eines Dokuments verweist, wird ein invertierter Index über mehrere Dokumente gebildet und verweist auf das Dokument, in dem der Indexterm vorkommt. Innerhalb dieses Index muss effizient gesucht werden können. Da der Index regelmäßig erweitert wird, ist eine Datenstruktur notwendig, die sowohl eine effiziente Suche zulässt, als auch Operationen wie Einfügen neuer Elemente und Änderungen kostengünstig erlaubt. Ein naheliegender Ansatz sind an dieser Stelle Baumstrukturen, die sicherstellen, dass der Baum für eine effiziente Suche ausreichend balanciert bleibt, beispielsweise B-Bäume oder AVL-Bäume. Die

²¹Vgl. Nohr 2005, S. 37–38.

²²Vgl. beispielsweise Sibun und Jeffrey 1996 und Martins und Silva 2005.

²³Vgl. Nohr 2005, S. 60.

Variante der B*-Bäume wird z. B. regelmäßig in bekannten relationalen Datenbanksystemen (z. B. MySQL und PostgreSQL) eingesetzt.

3.3. Verarbeitung der Suchanfrage und Sortierung der Ergebnisse

Der dritte Teil einer Suchmaschine ist die Verarbeitung der Suchanfragen und die Sortierung der Ergebnisse. Bevor der Index durchsucht werden kann, muss auch die Anfrage auf ähnlich Weise wie in Abschnitt 3.2 durchsucht werden. Außerdem müssen Schlüsselworte für die Suche erkannt werden (vgl. Abschnitt 3.3.1). Besonders der Sortierung der Ergebnisse, dem sogenannten Ranking, kommt eine sehr hohe Bedeutung zu, da die Menge an Webseiten so groß ist, dass ein Benutzer nicht die geringste Chance hätte, alle Seiten zu betrachten, die möglicherweise zu seiner Anfrage passen. Es ist daher notwendig, diejenigen Dokumente zu ermitteln, die am Besten passen, und diese zuerst zu präsentieren.

Einzelne Ranking-Methoden werden nicht allein, sondern in Kombination mit mehreren Verfahren angewandt. Dadurch wird sowohl ein besseres Ergebnis erzielt, als auch den Betreibern von Webseiten die Beeinflussung der Reihenfolge der Suchergebnisse ohne Berücksichtigung der eigentlichen Qualität ihrer Seite erschwert.

3.3.1. Boolesches Retrieval

Ein erster Ansatz ist die Kombination mehrerer Begriffe mit Hilfe der booleschen Algebra. Die Begriffe werden mit einem der drei Operatoren AND, OR und NOT verknüpft. Dabei entstehen allerdings nur genau zwei Mengen von Dokumenten, die ohne Weiteres nicht in eine Ordnung zu bringen sind. Die Menge aller Dokumente, welche die Bedingung erfüllen, und diejenigen, die dies nicht tun. Bei Katalogen, z. B. in Bibliotheken, kann diese Vorgehensweise, in Kombination mit einer Sortierung nach Jahr, Autor oder Titel, durchaus ausreichend sein. In Bezug auf die große Anzahl von Dokumenten im WWW ist die Ergebnismenge allerdings so groß, dass sie einer weiteren Bewertung bedarf.

3.3.2. Vektorraummodell

Eine verbreitete Methode zum Ranking der Suchergebnisse greift auf eine Methode zur Bestimmung der Ähnlichkeit zwischen Texten zurück, das Vektorraummodell.²⁴ Die zu untersuchenden Texte werden dabei auf Vektoren abgebildet. Für jedes mögliche Wort existiert eine Dimension im Vektor. Die Elemente des Vektors werden nach der Häufigkeit des Vorkommens der Wörter im Text belegt. Über die Bestimmung der Ähnlichkeit der Vektoren wird die Ähnlichkeit zwischen den Texten, denen die Vektoren zugehören, ermittelt. Anstelle eines zweiten Textes wird bei Suchanfragen jedoch die Ähnlichkeit der Anfrage mit den vorhandenen Texten ermittelt.

Für ein kurzes Beispiel nehmen wir an, dass unsere Dokumentensammlung nur zwei Dokumente umfasst. In beiden Dokumenten steht jeweils nur ein Satz: „Nicht für die Schule, sondern für das Leben lernen wir!“ und „Schule“ ist ein deutscher Film aus dem Jahr 2000.“ Unser Index besteht unter Anwendung der Regeln zur Indexerstellung aus Abschnitt 3.2.4 aus folgenden Wörtern: 2000, deutsch, Film, Jahr, Leben, lernen, Schule und sondern. Daraus ergeben sich für die beiden Sätze die Vektoren $\vec{a} = (0, 0, 0, 0, 1, 1, 1, 1)$ und $\vec{b} = (1, 1, 1, 1, 0, 0, 1, 0)$. In längeren Texten kämen natürlich einige Indexterme mehr als einmal vor, so dass auch größere Zahlen im Vektor vorkommen würden. Jemand der sich für den Film interessiert, sucht möglicherweise mit der Suchanfrage „Schule Film“, aus der der Vektor $\vec{q} = (0, 0, 1, 0, 0, 0, 1, 0)$ entsteht. Da $\vec{q} \cdot \vec{a} < \vec{q} \cdot \vec{b}$ gilt, ist der Vektor \vec{b} dem Vektor der Suchanfrage ähnlicher und damit der zweite Satz der Suchanfrage ähnlicher als der erste Satz.

Die Ermittlung der Ähnlichkeit allein mit dem Skalarprodukt führt dazu, dass lange Texte, welche Vektoren mit großen Beträgen erzeugen, gegenüber kurzen Texten bevorzugt werden. Um dieses Problem zu umgehen wird auf das sogenannte Cosinus-Maß zurückgegriffen.²⁵ Dabei wird das Skalarprodukt durch das Produkt der Beträge beider Vektoren geteilt. Damit wird der Cosinus des inneren Winkels der Vektoren ermittelt. Die Ähnlichkeit der Vektoren bzw. der Texte wird jetzt nur noch in Abhängigkeit von der Richtung der Vektoren bewertet.

²⁴Vgl. Stock 2007, S. 104 und Glöggler 2003, S. 73–75.

²⁵Vgl. Stock 2007, S. 335f.

3.3.3. PageRank

Eine bedeutende Veränderung bei der Sortierung von Suchergebnissen wurde durch die Entwicklung von PageRank ausgelöst.

Ganz im Sinne von Larry Page: Wir möchten Ihnen ‚genau das liefern, was Sie suchen‘. Eine der Hauptinnovationen bei der Gründung von Google war der PageRank, eine Technologie, mit der die ‚Wichtigkeit‘ einer Webseite anhand der Links von anderen Seiten sowie weiteren Faktoren bestimmt wurde. (Google inc. 2011)

Page et al. (1999) konnten erfolgreich relevante Webseiten allein durch die Position der Seite im Graph des durchsuchten Webs ermitteln.

Die Idee von PageRank basiert darauf, dass davon ausgegangen wird, dass ein Dokument nur dann von einem anderen Dokument verlinkt wird, wenn es vom jeweiligen Autor für relevant befunden wird. Ein Link ist also als Empfehlung, eine Seite zu besuchen, zu verstehen. Eine Seite mit vielen Empfehlungen, also vielen eingehenden Links, ist vermutlich besser als eine Seite mit wenigen eingehenden Links.

Eine Seite gibt den eigenen PageRank-Wert an die Seiten weiter, auf die ein Link gesetzt ist. Damit nicht der komplette Wert weitergegeben wird, gibt es einen sogenannten Dämpfungsfaktor d . Der PageRank einer Seite berechnet sich aus einem Startwert $1 - d$ und der Summe der Anteile des PageRank-Wertes der Seiten, die auf die Seite verweisen. Der Startwert ist notwendig, um Senken auszugleichen, die durch Kreise im Webgraphen, in die nur Links hinein, aber keine hinausführen, entstehen. Außerdem werden so auch einzelnen Seiten, auf die kein Verweis existiert, mit einem PageRank-Wert ausgestattet. Der Dämpfungsfaktor spiegelt die Wahrscheinlichkeit wider, mit der ein beliebiger Surfer (Random-Surfer-Modell) auf einen beliebigen Link klickt.²⁶

Sei B_u die Menge aller eingehenden Links der Seite u und N_u die Anzahl der ausgehenden Links der Seite u . Für die Berechnung des PageRank $R(u)$ einer Seite u , welche iterativ über mehrere Durchläufe erfolgt, gilt folgende Formel.

$$R(u) = (1 - d) + \sum_{v \in B_u} \frac{R(v)}{N_v}$$

²⁶Vgl. Page et al. 1999, S. 5.

(Page et al. 1999, S. 3–5) Auf diese Weise wird der Startwert zur Summe der eingehenden PageRank-Werte addiert. Weitergegeben wird der PageRank einer Seite geteilt durch die Anzahl der ausgehenden Links.

3.3.4. Weitere Verfahren

Neben den bereits angesprochenen Verfahren gibt es eine Reihe weiterer Methoden, die von Google und anderen Suchmaschinen verwendet werden.

Heute verwenden wir neben dem PageRank-Algorithmus mehr als 200 Faktoren, um Websites einzuordnen, und wir aktualisieren diese Algorithmen jede Woche. So bieten wir beispielsweise personalisierte Suchergebnisse an, die auf Ihrem Webprotokoll und Ihrem Standort basieren. (Google inc. 2011)

Genauso verwendet Yahoo eine Reihe von verschiedenen Kriterien.

Die Yahoo! Suche bestimmt die Reihenfolge der Webseiten nach deren Relevanz für eine bestimmte Suchanfrage und analysiert dafür deren Eigenschaften einschließlich der Texte, des Titels und der Beschreibung, Quelle, Verlinkungen und weiteren Charakteristika des Dokumentes. (Yahoo! 2011)

Auch die in Abschnitt 3.2.3 genannten Verfahren tragen zum Ranking der Dokumente bei. Weiterhin gibt es Verfahren, die der Idee von PageRank ähnlich sind, bei denen jedoch, ausgehend von als vertrauenswürdig angesehenen Seiten, Reputation weitergegeben wird. Bei diesen Verfahren werden bestimmten Webseiten höhere Werte zugewiesen, weil sie Expertenwissen auf ihrem Gebiet bieten oder besonders vertrauenswürdig sind, wie z. B. Internetauftritte von Universitäten. Da davon ausgegangen wird, dass solche Seiten nicht auf schlechte Seiten verlinken, wird durch diese Seiten ein höherer Wert weitergegeben.²⁷

Eine Methode, die nicht nur auf automatischer Analyse beruht, ist die Messung der Click-Popularity. Ausgehend davon, wie häufig eine Seite innerhalb der Ergebnisliste ausgewählt wird, wird dabei die Relevanz in Bezug auf die verwendeten Suchbegriffe bestimmt.²⁸

²⁷Vgl. Nascimento et al. 2004.

²⁸Vgl. Erlhofer 2011, S. 252–254.

Mit der Anzeige der Ergebnisse, ist der technische Anteil dieser Sachanalyse im wesentlichen beendet. Im nächsten Abschnitt werden die Wechselwirkungen zwischen dem Informatiksystem Suchmaschine und der Gesellschaft betrachtet.

3.4. Implikationen von Suchmaschinen

Da viele Menschen ihre Recherchen im WWW mit der Benutzung einer Suchmaschine beginnen, besteht für die Betreiber der Suchmaschinen die Möglichkeit, große Mengen an Daten über die Benutzer zu sammeln. Wenn zum Beispiel verschiedene Suchanfragen einzelnen Personen zugeordnet werden, lassen sich Profile der Nutzer anlegen. Datenschützer betrachten diese Möglichkeiten mit Sorgen. Die Betreiber der Suchmaschine argumentieren mit einer Personalisierung der Suchergebnisse und der präsentierten Werbung.

Ebenso können die finanziellen Interessen Probleme hervorrufen. Nach eigenen Angaben besteht die Haupteinnahmequelle von Google aus Werbung. Zum aktuellen Zeitpunkt werden Anzeigen innerhalb der Suchergebnisse gekennzeichnet. Eine Garantie für die Zukunft stellt dies aber nicht dar, sodass möglicherweise die Reihenfolge der angezeigten Ergebnisse nicht mehr vom Inhalt und damit der Relevanz, sondern von einer „Gebühr“ o. ä. abhänge.

Die Betreiber von Suchmaschinen bestimmen, welche Inhalte eine Suchmaschine findet, aber auch welche Inhalte, zumindest mit dieser Suchmaschine, nicht gefunden werden können. Dadurch wird Zensur möglich. Jüngste Beispiele sind die Anforderungen der chinesischen Regierung an Google. Google wollte die geforderte Zensur der Suchergebnisse nicht mehr weiter durchführen. Dies hat letztendlich dazu geführt, dass das Unternehmen China verlassen hat und Anfragen an google.cn an die Seite google.com.hk weiterleitet. In der Sonderverwaltungszone Hong Kong gelten andere, weniger strenge Zensurvorschriften. Auch in Deutschland werden teilweise Suchergebnisse nicht ausgegeben. Beispielsweise werden Suchanfragen mit Begriffen aus dem Bereich Erotik, aber auch potentiell volksverhetzende Anfragen nur eingeschränkt ausgeführt. Die Beweggründe für die Zensur sind nicht immer klar ersichtlich. Die Zensur von Seiten, die die nationalsozialistische Vergangenheit Deutsch-

lands verherrlichen, lässt sich mit dem Paragraphen 130 des Strafgesetzbuches rechtfertigen.²⁹ Bei der Zensur anderer Inhalte sind aber auch wirtschaftliche Interessen zumindest zu vermuten.

Nach diesem Überblick über die grundlegenden Mechanismen einer Suchmaschine folgt im nächsten Kapitel die Verbindung der Inhalte mit den Vorgaben aus Lehrplänen und Bildungsstandards.

4. Verknüpfung der Inhalte mit dem Lehrplan in NRW, den Vorgaben für den Unterricht in der Sek. II und den Bildungsstandards für die Sek. I

In diesem Kapitel soll eine erste Verbindung zwischen den fachlichen Inhalten und den aktuell gültigen Richtlinien hergestellt werden. Im hierauf folgenden Kapitel werden, bezugnehmend auf die einzeln vorgestellten Unterrichtseinheiten, erneut Verknüpfungen zu den hier eingeführten Richtlinien und Standards hergestellt. Es wird auf den Lehrplan für das Land Nordrhein-Westfalen sowie auf die Vorgaben für den Unterricht in der gymnasialen Oberstufe in Nordrhein-Westfalen Bezug genommen. Zusätzlich werden auch die Bildungsstandards – vorgestellt von der Gesellschaft für Informatik e. V. – betrachtet. Die Bildungsstandards liegen derzeit nur für die Sekundarstufe I vor. Bildungsstandards für die gymnasiale Oberstufe sind noch in Vorbereitung.

Ein Grundsatz für den Entwurf von Unterrichtseinheiten im Konzept IniK ist die Standardorientierung (siehe Abschnitt 2.2.2, Seite 6). Die fachlichen Inhalte des Kontextes *Suchmaschinen* sind m. E. so vielschichtig, dass es sich anbietet, den Kontext für den Unterricht in der Sekundarstufe II zu nutzen.

Da zum jetzigen Zeitpunkt noch keine Bildungsstandards vorliegen, greife ich auf zwei andere Quellen zur Sicherung der inhaltlichen Qualität zurück. Grundlagen für den Unterricht sind auf der einen Seite der Lehrplan für das Fach Informatik, welcher 1999 veröffentlicht wurde, aber auch die „Vorgaben zu den unterrichtlichen Voraussetzungen für die schriftlichen Prüfungen im Abitur in der gymnasialen Oberstufe“, welche die Mindestvoraussetzungen

²⁹Vgl. Schulz und Held 2007.

für das erfolgreiche Ablegen der Abiturprüfung im Fach Informatik darstellen. Ein unreflektiertes „teaching-to-the-test“, bei dem nur die geforderten Fähigkeiten trainiert werden, ist aus pädagogischer Perspektive abzulehnen. Jedoch darf die schulische Realität bei dem Entwurf einer Unterrichtsreihe nicht aus den Augen verloren werden. Wenn nicht ersichtlich ist, dass die genutzte Unterrichtszeit auch eine sinnvolle Prüfungsvorbereitung darstellt, ist die zu erwartende Akzeptanz bei den Lehrern gering. Die Vorgaben für den Unterricht repräsentieren zentrale Elemente der Informatik zumindest teilweise. Es sind beispielsweise Übereinstimmungen mit den Fundamentalen Ideen nach Schwill zu erkennen.³⁰ Selbstverständlich müssen die Vorgaben immer wieder kritisch geprüft werden. Die zu den Vorgaben für das Jahr 2007 geäußerten Kritikpunkte sind offensichtlich zumindest teilweise in die aktuellen Vorgaben eingearbeitet worden.³¹

4.1. Richtlinien und Lehrpläne

Die Lehrpläne sind an vielen Stellen recht allgemein gehalten. Die im vorhergehenden Kapitel angesprochenen Inhalte finden sich im Bereich „Fachliche Inhalte“ wieder. In diesen Bereich ist der Aspekt „Modellieren und Konstruieren“ eingeordnet.

Unter der Überschrift „Daten und Algorithmen abstrahieren“ werden folgende Inhalte aufgezählt:

- allgemeine Strategien und Standardlösungen kennen lernen und anwenden
- Formen des Strukturierens einsetzen
- problembezogene Objekte und ihre Wechselwirkungen spezifizieren
- ein Lösungskonzept als Denkschema entwickeln

(MfSuWWF NRW 1999, S. 12)

³⁰Vgl. MSW NRW 2008 und Schubert und Schwill 2011, S. 65–78.

³¹Vgl. Gesellschaft für Informatik Fachgruppe Informatische Bildung 2005.

Die konzeptionelle Bearbeitung der verwendeten Datenstrukturen, die innerhalb der Suchmaschine vorkommen, lässt sich hier einordnen. Konkrete Beispiele sind Arrays, Lineare Listen und Bäume, auf denen verschiedene Algorithmen ausgeführt werden. Für die im Rahmen der Stoppwortbehandlung notwendigen Vergleiche von Wörtern ist beispielsweise eine binäre Suche verwendbar. In diesem Zusammenhang sind auch Sortieralgorithmen notwendig, da für die Anwendung der binären Suche ein sortiertes Feld vorausgesetzt wird. Weitere konkrete Beispiele sind Standardoperationen auf binären Suchbäumen, welche modelliert werden können.

Weiter heißt es unter der Überschrift „Lösungen nach einem Programmierkonzept realisieren, überprüfen und weiterentwickeln“:

- Lösungskonzepte implementieren und testen
- Lösungen dokumentieren
- Lösungen nach vorgegebenen Kriterien bewerten
- Problemlösungen optimieren und weiterentwickeln

(MfSuWWF NRW 1999, S. 13)

Den Schülern sollte die Möglichkeit gegeben werden, über die Modellierung hinaus die vorgestellten Datenstrukturen und Algorithmen kennen zu lernen. Dazu ist es durchaus sinnvoll, Teile der Datenstrukturen und Algorithmen auch zu implementieren, so wie es in den angegebenen Richtlinien vorgesehen ist. Bei der praktischen Umsetzung sollte darauf geachtet werden, dass die Unterrichtszeit nicht für das Finden von Syntaxfehlern oder die Erstellung von Programmteilen, welche zwar für die Lauffähigkeit notwendig sind, aber im gegebenen Zusammenhang nicht Unterrichtsgegenstand sind, verschwendet werden muss. Mit der Entwicklungsumgebung BlueJ lassen sich beispielsweise kleine Projekte erstellen, die ohne Umfangreiche Main-Methoden oder grafische Benutzeroberflächen auskommen, aber dennoch lauffähig sind und so ein Testen der eigenen Umsetzung zulassen.

Weiterhin ist eine Bewertung der kennengelernte Inhalte notwendig. Die Datenmengen, mit denen im Kontext von Suchmaschinen umgegangen wird,

sind riesig. Bei der Verarbeitung ist es unerlässlich, die Effizienz der verwendeten Techniken im Auge zu behalten. Dies gilt sowohl bei der Indexierung, als auch bei der Verarbeitung der Suchanfragen. Die Richtlinien fordern eine Analyse der eingesetzten Datenstrukturen und Algorithmen im Bereich „Analysieren und Bewerten“. Unter der Überschrift „Algorithmen, Sprachkonzepte und Automatenmodelle beurteilen“ finden sich folgende Vorgaben:

- Grenzen von Verfahren und Methoden abschätzen
- formale Sprachen und Grammatiken untersuchen
- Syntaxregeln und Beschreibungssysteme beurteilen
- Automatenmodelle und akzeptierte Sprachen analysieren und beurteilen
- Effizienzuntersuchungen durchführen

(MfSuWWF NRW 1999, S. 14)

Auch der Bereich der kontextfreien Sprachen wird im Kontext Suchmaschinen benötigt, zum Beispiel bei der Suche nach weiterführenden Links während des Crawling.

Über den Bereich der Kerninformatik hinausgehend habe ich in Abschnitt 3.4 gesellschaftliche Implikationen von Suchmaschinen beschrieben. Die Notwendigkeit, diese auch im Unterricht aufzunehmen, findet sich in den Richtlinien im Abschnitt „Lernen im Kontext der Anwendung“ auf Seite 16ff.

4.2. Vorgaben zum Unterricht in der Sekundarstufe II

Mit der Einführung des Zentralabiturs wurden konkretere Vorgaben für die im Unterricht zu behandelnden Inhalte notwendig. Viele der Inhalte des Kontextes Suchmaschine finden sich auch in den sogenannten unterrichtlichen Voraussetzungen wieder.

In diesen Voraussetzungen finden sich sowohl die Datenstruktur Baum mit der Anwendung der Standardoperationen wieder, als auch der im Zusammenhang mit dem Crawling beschriebene Graph.³² Auf der Internetpräsenz des

³²Vgl. MSW NRW 2008, MSW NRW 2009 und MSW NRW 2010.

Ministeriums für Schule und Weiterbildung Nordrhein-Westfalen sind die Prüfungsaufgaben der vergangenen Jahre, aber auch Beispielimplementierungen für binäre Suchbäume und andere Elemente der unterrichtlichen Voraussetzungen abrufbar, welche den Schluss erlauben, dass beispielsweise Kenntnisse über eine konkrete Implementierung erwartet werden.³³

Leider finden Inhalte des Themenbereichs *Informatik und Gesellschaft* nach wie vor keine angemessene Repräsentation in den Vorgaben.

4.3. Bildungsstandards für die Sekundarstufe I

Die 2008 vorgestellten Bildungsstandards stellen Mindeststandards für die Sekundarstufe I dar. Auch wenn die in dieser Arbeit beschriebene Unterrichtsreihe für die Sekundarstufe II vorgesehen ist, finden sich in den Bildungsstandards viele Anknüpfungspunkte, aus denen die Relevanz des gewählten Kontextes hervorgeht.

Die in den Bildungsstandards beschriebenen Kompetenzen werden in Inhalts- und Prozessbereiche aufgeteilt. Innerhalb der Inhaltsbereiche werden folgende Hauptpunkte genannt:

- Information und Daten
- Algorithmen
- Sprachen und Automaten
- Informatiksysteme
- Informatik, Mensch und Gesellschaft

Siehe (GI e. V. 2008, S. 12–13).

In jedem dieser Inhaltsbereiche sind Anknüpfungspunkte zum Kontext Suchmaschine zu finden. Einige Beispiele werde ich an dieser Stelle nennen, für weitere Verbindungen verweise ich auf das nächste Kapitel.

³³Vgl. <http://www.standardsicherung.schulministerium.nrw.de/abitur-gost/fach.php?fach=15> – Abgerufen am 24. Juni 2011.

Die vernetzte Struktur der Dokumente im WWW kann zum Beispiel dem Punkt „Information und Daten“ zugeordnet werden. In den Bildungsstandards heißt es im Zusammenhang mit der Erstellung von eigenen Webseiten:

Unter dem Gesichtspunkt von Daten und Information kommt [...] die Vernetzung hinzu, die durch Graphen dargestellt wird, deren Knoten die Dokumente und deren gerichtete Kanten die Verweise zwischen Dokumenten sind. [...] Das Augenmerk liegt darauf, wie die insgesamt darzustellende Information in einzelne Dokumente gegliedert und durch Verweise miteinander in Beziehung gesetzt wird, sodass sich die Leserinnen und Leser der Dokumente gut orientieren können. (GI e. V. 2008, S. 27)

Dies gilt m. E. nicht nur bei der Erstellung eigener Seiten, sondern auch bei der Analyse der bereits vorhandenen Webseiten.

Beim Inhaltsbereich *Algorithmen* wird eine Beschäftigung mit einer algorithmischen Denkweise und mit der Darstellung und Interpretation von gegebenen Algorithmen erwartet. Im Kontext Suchmaschinen lassen sich verschiedene Algorithmen wiederfinden. Darunter auch Standardlösungen zum Suchen oder Sortieren.

Die Bildungsstandards legen auf eine Verknüpfung der fachwissenschaftlichen Inhalte mit den fachübergreifenden Inhalten aus dem Bereich *Informatik, Mensch und Gesellschaft* großen Wert; bei den Voraussetzungen für den Unterricht in der gymnasialen Oberstufe fehlen diese. Die in Abschnitt 3.4 genannten Wechselwirkungen passen in diesen Inhaltsbereich.³⁴

Neben den Inhaltsbereichen gibt es eine Reihe von Prozessbereichen, welche grundlegend für die Informatik sind. Sowohl Inhalts-, als auch Prozessbereiche sind untereinander und wechselseitig nicht völlig losgelöst zu betrachten.

Die Prozesskompetenzen werden an der Arbeit mit den Inhalten erworben, ohne die Inhalte wären viele von ihnen nicht spezifisch für die Informatik. Umgekehrt stünden die Inhalte ohne Prozesse in der Gefahr, zu einer Wissenssammlung für Quizshows zu verkommen. (GI e. V. 2008, S. 45)

Konkret werden folgende Prozessbereiche genannt:

- Modellieren und Implementieren

³⁴Vgl. GI e. V. 2008, S. 41–44.

- Begründen und Bewerten
- Strukturieren und Vernetzen
- Kommunizieren und Kooperieren
- Darstellen und Interpretieren

Siehe (GI e. V. 2008, S. 13–14)

Die Einbindung der Prozessbereiche in die Unterrichtsreihe wird m. E. im Rahmen der konkreten Einheiten leichter deutlich, als es in einer allgemeinen Form gezeigt werden kann. Offensichtlich ist jedoch, dass die Modellierung und Implementierung der genannten Datenstrukturen und Algorithmen einen nicht unerheblichen Platz in der Gestaltung eines Unterrichts zum Thema Suchmaschinen einnehmen kann.

Konkrete Bezüge zu den Inhalts- und Prozessbereichen, aber auch zu den Richtlinien, Lehrplänen und unterrichtlichen Voraussetzungen stelle ich im nächsten Kapitel her.

5. Beschreibung der Unterrichtsreihe

Die Unterrichtsreihe zum Thema Suchmaschinen besteht aus zehn einzelnen Unterrichtseinheiten, die etwa zur Hälfte aus Doppelstunden bestehen. In den folgenden Abschnitten wird zuerst ein kurzer Überblick über den Inhalt der Unterrichtsreihe gegeben, danach werden die einzelnen Einheiten vorgestellt.

Die im Anhang, ab Seite 71 vorgestellten tabellarischen Unterrichtsverläufe, insbesondere die dargestellten zeitlichen Angaben, sind als Vorschläge zu verstehen. Es bedarf selbstverständlich einer Anpassung an die konkrete Lerngruppe und andere Rahmenbedingungen.

Im Anhang B (Seite 86) finden sich Materialvorschläge für einzelne Unterrichtseinheiten.

5.1. Überblick

Suchmaschinen haben eine Portalfunktion. Sie bieten für viele Nutzerinnen und Nutzer einen Einstieg in das World Wide Web. Davon sind auch Schü-

lerinnen und Schüler nicht ausgenommen. Immerhin 14% (Jungen) bzw. 15% (Mädchen) der Internetnutzung dienen der Informationssuche.³⁵ Wenn Jugendliche im Internet nach Informationen suchen, so ist eine Suchmaschine das Mittel der Wahl, noch vor dem Nachschlagen in der Wikipedia.³⁶ Die Benutzung einer Suchmaschine ist den Schülerinnen und Schülern demnach nicht fremd. Es erscheint mir wichtig, den Schülerinnen und Schülern ein tieferes Verständnis von der Funktionsweise einer Suchmaschine zu vermitteln. Die in dieser Arbeit vorgestellte Unterrichtsreihe soll diesem Zweck dienen.

Zu Beginn wird eine Übersicht über die Funktionsweise des hier thematisierten Informatiksystems *Suchmaschine* gegeben. Dazu wird es in mehrere Teile aufgebrochen. Daran anschließend erfolgt ein kurzer Überblick über die Funktionsweise der einzelnen Teile. Einen großen Anteil der Unterrichtszeit wird von der Modellierung und Implementierung der Datenstruktur *binärer Suchbaum* und des Sortierverfahrens *Sortieren durch Einfügen*³⁷ eingenommen.

Baumstrukturen bieten eine effiziente Möglichkeit wachsende Datenmengen dauerhaft schnell durchsuchen zu können. Sie eignen sich also für die Erstellung eines erweiterungsfähigen Indexes. Auch wenn der einfache binäre Suchbaum nicht davor geschützt ist aus der Balance zu geraten und damit seine Eigenschaften bzgl. der effizienten Durchsuchbarkeit zu verlieren, stellt er m. E. einen geeigneten Ausgangspunkt für eine spätere Betrachtung anderer Baumstrukturen dar. Auf die Erkenntnisse, die bei der Betrachtung des binären Suchbaums erlangt wurden, kann im Sinne eines Spiralcurriculums zurückgegriffen werden.

Der weitere Schwerpunkt, das Sortieren, begründet sich aus dem Kontext durch die Notwendigkeit für das Funktionieren einer binären Suche, welche im Rahmen der Erkennung von Stoppwörtern, eingesetzt werden kann. Die Wahl ist hier auf das Sortieren durch Einfügen gefallen, da dieser Algorithmus einfach nachzuvollziehen ist und weil er auf kleinen Feldern sowie auf schon größtenteils sortierten Feldern effizient arbeitet. Viele Implementierungen von Quicksort greifen beispielsweise bei kleinen Feldern auf Sortieren durch Einfü-

³⁵Vgl. Medienpädagogischer Forschungsverbund Südwest 2010, S. 29.

³⁶Vgl. ebd., S. 32.

³⁷Die Bezeichnungen *Sortieren durch Einfügen* und *Insertion Sort* werden synonym gebraucht.

gen zurück.³⁸ Natürlich kann an dieser Stelle auch ein anderer oder ein weiterer Algorithmus besprochen werden.

Den Abschluss können zwei Unterrichtseinheiten bilden, welche sich mit der Relevanz und Qualität von Dokumenten im WWW beschäftigen. Automatische Messungen dieser Größen sind für das Ranking der Suchergebnisse einer Suchmaschine notwendig. Diese beiden Einheiten können als Exkurse aber auch an anderen Stellen der Unterrichtsreihe verwendet werden.

Für die Unterrichtsreihe wird keine besondere Software oder außergewöhnliches Material benötigt. Material welches für die Durchführung der einzelnen Unterrichtseinheiten benötigt wird, ist in den entsprechenden Abschnitten aufgeführt. Für die Visualisierung der Algorithmen und Datenstrukturen können verschiedene Programme hilfreich sein.

5.1.1. Lernziele

Die Schülerinnen und Schüler sollen in der Reihe die Komplexität des Informatiksystems Suchmaschine erkennen. Dafür sollen sie die einzelnen Teile des Systems kennenlernen und ihre Funktionsweise erläutern. Weiterhin sollen die Schülerinnen und Schüler Probleme erkennen, die mit dem Sammeln großer Datenmengen, wie es Betreibern von Suchmaschinen möglich ist, entstehen können. Sie sollen für die Auswirkungen einer Informationskontrolle durch Suchmaschinen als Tor ins WWW – Stichwort Zensur – sensibilisiert werden.

5.1.2. Voraussetzungen

Für die erfolgreiche Durchführung der Unterrichtsreihe müssen auf Seiten der Schülerinnen und Schüler einige Voraussetzungen gegeben sein oder es muss entsprechend mehr Zeit für die einzelnen Einheiten eingeplant werden um die fehlenden Voraussetzungen auszugleichen. Ich werde im Folgenden inhaltliche und methodische Voraussetzungen nennen, welche m. E. erfüllt sein sollten.

Inhaltliche Voraussetzungen Die Schülerinnen und Schüler sollten über einen adäquaten Algorithmenbegriff verfügen. Wenn im Zuge der Reihe ver-

³⁸Vgl. Sedgewick 2002, S. 155.

schiedene Algorithmen besprochen und ansatzweise analysiert werden sollen, muss ein allgemeines Verständnis für Algorithmen als präzise, automatisch ausführbare Handlungsanweisungen vorhanden sein. Notwendig erscheint mir auch die Kenntnis von Kontrollstrukturen, Variablen und Funktionen. Sinnvoll ist es darüber hinaus auch, wenn das Konzept der Rekursion bereits eingeführt ist. Zum einen sind die iterativen Implementierungen verschiedener Algorithmen deutlich unübersichtlicher, zum anderen ist auch die Referenzimplementierung der Klasse *BinarySearchTree*³⁹ rekursiv gestaltet.

Weiterhin sollte das Konzept der Objektorientierung bekannt sein, um die Modellierung des Binärbaums erfolgreich durchführen zu können.

Methodische Voraussetzungen Den Schülerinnen und Schülern sollten die im Folgenden dargestellten Unterrichtsmethoden bekannt sein. Alternativ können diese Methoden auch innerhalb der Unterrichtsstunde vermittelt werden, dafür muss aber entsprechende Zeit eingeplant werden. Viele Methoden werden auch in anderen Fächern eingesetzt und können daher aus diesen bereits bekannt sein.

Grundsätzlich ist sollten die Schülerinnen und Schüler kooperativ in Gruppen zusammen arbeiten können. Diese Fähigkeit wird seit einiger Zeit bereits in der Primarstufe und der Sekundarstufe I eingeübt, so dass grundlegende Fähigkeiten vorhanden seien dürften. Eine erfolgreiche Gruppenarbeit bedarf einer effektiven Zusammenarbeit im Team, die allen Mitgliedern gleichstarke Beteiligung erlaubt und nicht einzelne Schülerinnen und Schüler benachteiligt. An dieser Stelle muss die Lehrkraft Hilfestellungen leisten.

Um bei arbeitsteiligen Methoden, die erarbeiteten Ergebnisse den Mitschülern vorstellen zu können und damit diese auch alle Schülerinnen und Schüler erreichen, müssen geeignete Präsentationstechniken angewendet werden. Es ist von Vorteil, wenn die Grundlagen dieser Fähigkeiten schon vor Eintritt in die Oberstufe erlernt werden.

Konkret wird unter anderem das *Gruppenpuzzle*, eine arbeitsteilige, Methode verwendet. Bei dieser Methode werden zuerst Inhalte in sogenannten Ex-

³⁹Vgl. <http://www.standardsicherung.schulministerium.nrw.de/abitur-gost/fach.php?fach=15>

pertengruppen erarbeitet, um dann in den Stammgruppen, durch die jeweiligen Experten, weitergegeben zu werden. Bei dieser Methode ist es wichtig, dass alle Schülerinnen und Schüler die Vermittlung durchführen können.

In ähnlicher Weise verläuft der *Museumsgang*, bei dem stärker auf die Präsentation Wert gelegt wird. Die Schülerinnen und Schüler besuchen gruppenweise die einzelnen „Exponate“ des Museums und einer der Schülerinnen und Schüler übernimmt die Vorstellung bzw. Erklärung.

Bei der Methode *Think-Pair-Share* erarbeiten sich die Schülerinnen und Schüler zuerst eine Thematik in Einzelarbeit, um sich danach mit einem Partner oder einer Partnerin, auszutauschen. Daran schließt sich eine Präsentationsphase an, in der die Inhalte im Plenum besprochen werden. Es ist auch möglich eine weitere Phase einzuschieben, in der erst in einer Vierergruppe die Präsentation geübt wird. Diese Methode ist teilweise auch als *Think-Pair-Square* bekannt.

Im Exkurs zum Thema *Qualität von Informationen im WWW* wird eine Podiumsdiskussion, kombiniert mit einem Rollenspiel, als Unterrichtsmethode eingesetzt. Die Schülerinnen und Schüler sollen sich dabei in unterschiedliche Rollen hineinversetzen um ein kontroverses Thema zu durchdringen. Die Diskussion dient der Darstellung der verschiedenen Positionen.

5.2. UE 1 – Aufbau von Suchmaschinen

In der Einführungsstunde wird ein erster Überblick über den Kontext gegeben. Die Schülerinnen und Schüler sollen die Black Box öffnen und neugierig auf die Funktionsweise einer Suchmaschine werden.

5.2.1. Lernziele

Die Schülerinnen und Schüler sollen erkennen, dass es sich bei einer Suchmaschine um ein komplexes Informatiksystem handelt, für dessen Funktion eine Reihe von einzelnen Schritten ineinander greifen müssen. Sie sollen verinnerlichen, dass ein großer Teil der Arbeit einer Suchmaschine vor dem Eingang eines konkreten Suchauftrages erfolgt und somit keine Suche in Echtzeit stattfindet.

Nach der Stunde können die Schülerinnen und Schüler das System in drei bzw. vier Teile trennen und die Teile sowie deren Aufgabe benennen.

5.2.2. Material

Folgendes Material wird benötigt:

- Materialkiste zum durchsuchen mit verschiedenen Dokumenten, Videos u. ä.
- Präsentationsmaterial, z. B. Plakate, Fasermaler, Klebeband
- (mehrere) Computer mit Internetverbindung
- Arbeitsblätter (siehe Anhang B.1 auf Seite 86)

5.2.3. Ablauf

Der tabellarische Unterrichtsentwurf findet sich auf Seite 71. Für diese Unterrichtseinheit sind etwa 90 Minuten eingeplant. Die Unterrichtsreihe beginnt mit einem Impuls, für den die Lehrkraft eine Kiste mit Dokumente – verschiedene Medien sind durchaus wünschenswert – mitbringt und dem Kurs einen Suchauftrag erteilt. Als Beispiel wurde die Anfrage „schriftliche Division“ genannt, andere Beispiele sind ebenso möglich. Die Schülerinnen und Schüler finden passende Dokumente, wie Schulbücher oder Anleitungen, Lehrvideos oder ähnliches zwischen vielen nicht passenden Materialien wie Romanen, Zeitungen, Zeitschriften, Werbeprospekten u. v. m. Bei der Auswahl der Dokumente für die Kiste sind der Phantasie keine Grenzen gesetzt, wichtig ist nur, dass die Schülerinnen und Schüler in der Lage sind, die Kiste in der gegebenen Zeit zu durchsuchen und das passende Material zu finden.

Nach einer Vorstellung des Themas der Unterrichtsreihe und einem Überblick über den zeitlichen Ablauf der Reihe durch die Lehrkraft, erarbeiten die Schülerinnen und Schüler in arbeitsteiliger Gruppenarbeit jeweils eine Präsentation in Plakatform zu einem der drei grundlegenden Teile einer Suchmaschine.

Die Plakate werden für einen Museumsgang im Raum verteilt aufgehängt. Die Gruppen werden neu gemischt, so dass aus jeder vorherigen Gruppe mindestens eine Schülerin oder ein Schüler in den neuen Gruppen ist. Diese Gruppen gehen dann zu den Plakaten und stellen sich die Inhalte gegenseitig vor.

Eine Sicherung kann zum Beispiel über (Digital-)Fotos der Plakate, welche dann an alle Schülerinnen und Schüler verschickt werden, erreicht werden.

5.2.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Aufdecken der Funktionsweise bekannter Werkzeuge“ *Methoden und Formen selbstständigen Arbeitens – fachspezifische Vorgehensweise*
- „Selbstständige und projektorientierte Arbeitsformen“ *Methoden und Formen selbstständigen Arbeitens*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „verstehen die Grundlagen des Aufbaus von Informatiksystemen und deren Funktionsweise“ *Informatiksysteme*
- ... „kommunizieren fachgerecht über informative Sachverhalte“ *Kommunizieren und Kooperieren*

5.3. UE 2 – Durchsuche des WWW: Crawling

Die zweite Unterrichtseinheit beschäftigt sich mit der Beschaffung neuer und veränderter Dokumente. Der Vorgang des Crawling entspricht, in vereinfachter Form, dem Versuch alle Knoten des WWW-Graphen, also die Webseiten und andere Dokumente, zu besuchen.

5.3.1. Lernziele

Die Schülerinnen und Schüler erfahren in dieser Stunde, dass sich das World-Wide-Web als gerichteter Graph modellieren lässt. Sie erhalten darüber ein implizites Verständnis für Graphen. Eine Definition könnte eingeführt werden.

Die Lerngruppe lernt weiterhin zwei mögliche Algorithmen kennen, mit denen die Knoten des Graphen der Reihe nach besucht werden können. Sie kennen die Unterschiede zwischen einer Breiten- und Tiefensuche und können Vor- und Nachteile, im Bezug auf Speicherplatzbedarf und das Erreichen potentiell hochwertiger Seiten durch Crawler von Suchmaschinen, einschätzen.

5.3.2. Material

Folgendes Material wird benötigt:

- Arbeitsblatt mit Informationen über gerichtete Graphen und Platz zum abbilden einer Beispieleseite
- untereinander verlinkte Beispieleseiten, z. B. die Schulhomepage
- Texte zur Einführung von Breiten- und Tiefensuche

5.3.3. Ablauf

Der tabellarische Unterrichtsentwurf findet sich auf Seite 72. Für diese Unterrichtseinheit sind etwa 45 Minuten eingeplant.

Diese Einheit ist zweigeteilt. Im ersten Teil erstellen die Schülerinnen und Schüler aus einer Menge von Beispieleseiten ein grafisches Modell der Seite. Für diesen Zweck eignet sich möglicherweise die Schulhomepage. Es ist anzustreben, dass es sich um eine reale Seite handelt, jedoch besteht bei vielen aktuellen Seiten oder auch bei der Arbeit mit manchen Content-Management-Systemen die Problematik, dass die Seiten sehr unübersichtlich sind.

Im zweiten Teil beschäftigen sich die Schülerinnen und Schüler mit zwei Algorithmen um alle Knoten des zusammenhängenden Graphen zu besuchen. Dabei werden die jeweiligen Algorithmen anhand der Ergebnisse des vorherigen Teils nachvollzogen.

Im Anschluss daran werden die Methoden im Plenum vorgestellt. Dabei sammelt die Lehrkraft die verschiedenen Eigenschaften, z. B. tabellarisch an der Tafel.

5.3.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Ein Informatikmodell gewinnen: Probleme eingrenzen und spezifizieren, reduzierte Systeme definieren“ *Fachliche Inhalte – Modellieren und Konstruieren*
- „Selbstständige und projektorientierte Arbeitsformen“ *Methoden und Formen selbstständigen Arbeitens*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „erstellen informative Modelle zu gegebenen Sachverhalten“ *Modellieren und Implementieren*
- ... „stellen Informationen in unterschiedlicher Form dar“ *Information und Daten*
- ... „interpretieren Daten im Kontext der repräsentierten Information“ *Information und Daten*

5.4. UE 3 – Automatische Textanalyse

Die nächste Unterrichtsstunde beschäftigt sich mit der Untersuchung der zuvor beim Crawling gefundenen Dokumente. Dabei wird an dieser Stelle nur die Untersuchung von Texten betrachtet.

5.4.1. Lernziele

Die Schülerinnen und Schüler sollen einen Überblick über die Möglichkeiten der automatischen Indexierung erhalten. Die Lerngruppe erfährt, dass statistische Eigenschaften von Wörtern in natürlichsprachlichen Texten mit der Bedeutung für den Inhalt des Textes korrelieren und verinnerlichen die Bedeutung einer linguistischen Bearbeitung der Wörter, bevor sie in einen Index

übernommen werden. Sie setzen sich weiterhin mit den Möglichkeiten zur Strukturierung und Textauszeichnung auseinander.

Nach der Unterrichtseinheit können die Schülerinnen und Schüler die Vorgänge bei der automatischen Erstellung eines Index wiedergeben. Sie kennen die Begriffe Stoppwort, Termfrequenz, inverse Dokumentenhäufigkeit und invertierter Index und können diese erklären. Sie können die Regeln, welche bei der automatischen Indexierung notwendig sind, anwenden.

5.4.2. Material

Folgendes Material wird benötigt:

- kurzer Text und zugehöriger Volltextindex
- Text mit Informationen über die automatische Textanalyse
- Stopwortliste
- Arbeitsblatt mit Vorgaben zur Erstellung von Indexterminen

5.4.3. Ablauf

Der tabellarische Unterrichtsentwurf findet sich auf Seite 73. Für diese Unterrichtseinheit sind etwa 45 Minuten eingeplant.

Die Stunde beginnt mit einem Impuls für den die Lehrkraft einen kurzen Text und einen daraus generierten Volltextindex vorstellt. Der Index sollte mit Hinweisen auf die Häufigkeit des Vorkommens der einzelnen Indexterme versehen sein. Der Index ist kaum kürzer als der ursprüngliche Text und eignet sich somit wenig für die Bestimmung des Inhalts. Daraufhin erhalten die Schülerinnen den Auftrag selbstständig einen Index für den gleichen Text zu erstellen, der ihrer Meinung nach sinnvoller ist. Die Überlegungen werden mit einem Partner besprochen und danach im Sinne des Think-Pair-Share im Plenum vorgestellt. Dabei sollen allgemeingültige Regeln erstellt werden, nach denen ein sinnvoller Index erzeugt werden kann. Die Lehrkraft sammelt die Ideen z. B. in Form einer Mindmap.

Im nächsten Schritt bearbeiten die Schülerinnen und Schüler in Einzelarbeit einen Text zum Thema und besprechen die Erkenntnisse zuerst mit einem Partner um sie danach wieder im Plenum vorzustellen. Die Lehrkraft ergänzt dabei die Aufzeichnungen aus der vorangegangenen Vorstellung.

Zur Vertiefung und Sicherung erstellen die Schülerinnen und Schüler zum Abschluss erneut einen Index unter Berücksichtigung der erlernten Regeln.

5.4.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Fachübergreifende und fächerverbindende Sichtweisen“ *Methoden und Formen selbstständigen Arbeitens*
- „Typische Einsatzbereiche, Möglichkeiten, Grenzen, Chancen und Risiken der Informations- und Kommunikationssysteme untersuchen und einschätzen“ *Fachliche Inhalte – Analysieren und Bewerten*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „nutzen formale Sprachen zur Interaktion mit Informatiksystemen und zum Problemlösen“ *Sprachen und Automaten*
- ... „kommunizieren fachgerecht über informative Sachverhalte“ *Kommunizieren und Kooperieren*
- ... „kooperieren bei der Lösung informatischer Probleme“ *Kommunizieren und Kooperieren*

5.5. UE 4 – Binäre Suchbäume

Die folgende Unterrichtseinheit befasst sich vertiefend mit den informatischen Inhalten, die bei der Indexerstellung von Bedeutung sind. Der direkte Zusammenhang mit Suchmaschinen besteht in der Notwendigkeit zum einen effizient die Wörter im Dokument mit einer Stoppwortliste abgleichen zu können

und zum anderen einen Index zu verwalten, der effizient durchsucht und ständig erweitert werden kann. Als Methoden zur Umsetzung werden eine sequentielle Suche auf einer linearen Liste, eine binäre Suche auf einem sortierten Feld und ein binärer Suchbaum eingeführt.

5.5.1. Lernziele

Die Schülerinnen und Schüler lernen in dieser Stunde grundlegende Eigenschaften der Datenstrukturen lineare, verkettete Liste, Feld und binärer Suchbaum in Kombination mit den oben genannten Algorithmen kennen. Am Ende der Einheit können die Schülerinnen und Schüler Vor- und Nachteile der verschiedenen Datenstrukturen und Algorithmen erläutern und deren Anwendungsmöglichkeiten in Abhängigkeit vom Einsatzgebiet abwägen.

5.5.2. Material

Folgendes Material wird benötigt:

- Arbeitsblätter mit Informationen zu den einzelnen Datenstrukturen und Algorithmen
- evtl. Material um die einzelnen Strukturen in Form von kleinen Rollenspielen nachzuspielen oder sie auf dem Tisch zu modellieren
- Präsentationsmaterial (Plakate, Karteikarten, Pinnwände, Pinnwandnadeln)

5.5.3. Ablauf

Der tabellarische Unterrichtsentwurf findet sich auf Seite 74. Für diese Unterrichtseinheit sind etwa 90 Minuten eingeplant.

Diese Unterrichtsstunde beginnt mit der Anbindung an den Kontext. Die Lehrkraft erläutert die Problemstellung – die Verwaltung der Stoppwortliste und des Index. Als Methode soll an dieser Stelle ein Gruppenpuzzle eingesetzt werden, bei dem sich drei Gruppen jeweils mit einer Möglichkeit zur Verwaltung einer Menge von Stoppwörtern auseinandersetzen. Während dieser Erarbeitungsphase sollen die Schülerinnen und Schüler die Algorithmen selbst

„ausprobieren“, in dem sie einzelne Felder bzw. Objekte mit Karteikarten darstellen oder selbst in Form eines kleinen Rollenspiels umsetzen.⁴⁰ Diese Methode kann auch für die Präsentation in den Stammgruppen, zusätzlich oder anstelle klassischer Präsentationsmaterialien, genutzt werden.

Nachdem die Schülerinnen und Schüler in den Stammgruppen einen Überblick über die Vor- und Nachteile der einzelnen Datenstrukturen und Algorithmen erhalten haben, stellen sie diese gegenüber und kommen zu einer Empfehlung zur Nutzung einer Methode bei der Suche nach Wörtern in der Stopwortliste. Diese Ergebnisse werden im Plenum präsentiert und von der Lehrkraft gesammelt.

In einem abschließenden Schritt wird über die Aufgabe der Speicherung des Index beraten. Den Schülerinnen und Schülern sollte an dieser Stelle klar sein, dass es für diesen Einsatzzweck erforderlich ist, dass die Datenstruktur problemlos und kostengünstig erweitert werden kann. Aus der Präsentation der Ergebnisse der zweiten Beratung soll sich eine Einigung auf binäre Suchbäume ergeben, welche das Thema für die nächsten Stunden darstellen. Zur Vorbereitung sollen die Schülerinnen und Schüler als Hausaufgabe die Umsetzung der Operationen Löschen und Hinzufügen von Knoten vorbereiten.

5.5.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Daten und Algorithmen abstrahieren“ *Fachliche Inhalte – Modellieren und Konstruieren*
- „Algorithmen, Sprachkonzepte und Automatenmodelle beurteilen“ *Fachliche Inhalte – Analysieren und Bewerten*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „kommunizieren fachgerecht über informative Sachverhalte“ *Kommunizieren und Kooperieren*

⁴⁰Vgl. Dissmann 2003.

- ... „kennen und verwenden grundlegende Operationen zum Zugriff auf die Bestandteile strukturierter Daten“ *Information und Daten*
- ... „kennen Algorithmen zum Lösen von Aufgaben und Problemen aus verschiedenen Anwendungsgebieten und lesen und interpretieren gegebene Algorithmen“ *Algorithmen*
- ... „erstellen informatische Modelle zu gegebenen Sachverhalten“ *Modellieren und Implementieren*

5.6. UE 5 – Binäre Suchbäume: Standardoperationen

Die folgende Unterrichtseinheit beschäftigt sich mit Standardoperationen auf binären Suchbäumen. Der Suchbaum kann nur dann funktionieren, wenn diese Operationen implementiert sind. Die Stunde vertieft damit die Erkenntnisse aus der vorangegangenen Stunde und bereitet die konkrete Implementierung eines binären Suchbaums vor.

5.6.1. Lernziele

Die Schülerinnen und Schüler lernen in dieser Unterrichtseinheit die Funktionsweise der Operationen *Suchen*, *Hinzufügen* und *Löschen* auf einem binären Suchbaum kennen. Außerdem lernen sie die Möglichkeit der In-Order-Traversierung kennen, die bei der Anwendung auf einem binären Suchbaum eine sortierte Liste erzeugt.

Die Schülerinnen und Schüler können am Ende der Stunde die einzelnen Abläufe der Operationen nachvollziehen und in Grundzügen erläutern.

5.6.2. Material

In dieser Unterrichtseinheit wird, wie in der letzten Stunde, Material zu Modellierung eines Binärbaums benötigt.

5.6.3. Ablauf

Der tabellarische Unterrichtsentwurf findet sich auf Seite 76. Für diese Unterrichtseinheit sind etwa 45 Minuten eingeplant. Möglicherweise ist diese Zeit sehr knapp bemessen, so dass zusätzliche Zeit eingeplant werden sollte.

Die Stunde beginnt mit einem Überblick über die, teilweise schon implizit eingeführten, Operationen. Danach erarbeiten die Schülerinnen und Schüler die Funktionsweise der einzelnen Operationen in Einzelarbeit. Wenn sie daran anschließend Struktogramme für mindestens zwei der Operationen erstellen, sollen sie dafür bekannte Konzepte wie das der Rekursion anwenden.

Die erstellten Struktogramme werden von den Schülerinnen und Schülern gegenseitig auf Korrektheit geprüft, so dass sich die Gruppe untereinander Hilfestellungen geben kann. Für offene Fragen oder Probleme steht die Lehrkraft zur Verfügung. Mit der Hausaufgabe, die Operationen zu verbalisieren und in Pseudocode zu implementieren wird die nächste Stunde vorbereitet.

5.6.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Daten und Algorithmen abstrahieren“ *Fachliche Inhalte – Modellieren und Konstruieren*
- „Ein Informatikmodell gewinnen: Probleme eingrenzen und spezifizieren, reduzierte Systeme definieren“ *Fachliche Inhalte – Modellieren und Konstruieren*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „erstellen informative Modelle zu gegebenen Sachverhalten“ *Modellieren und Implementieren*
- ... „implementieren Modelle mit geeigneten Werkzeugen“ *Modellieren und Implementieren*

- ... „entwerfen und realisieren Algorithmen mit den algorithmischen Grundbausteinen und stellen diese geeignet dar“ *Algorithmen*

5.7. UE 6 – Standardoperationen implementieren

Diese Stunde setzt die Arbeiten der letzten Stunde fort. Der zuvor modellierte binäre Suchbaum soll nun konkret in einer Programmiersprache, hier Java, umgesetzt werden.

5.7.1. Lernziele

Die Schülerinnen und Schüler vertiefen in dieser Stunde ihr Verständnis für die Funktionsweise der Standardoperationen des binären Suchbaumes. Sie festigen ihre Kenntnisse der Modellierung mit der UML. Die Schülerinnen und Schüler können anschließend eine vorgegebene Methode, auf der Grundlage von Modellen, kodieren.

5.7.2. Material

Folgendes Material wird benötigt:

- Arbeitsblätter mit Informationen zu den einzelnen Operationen
- Vorlagen für ein Klassendiagramm
- evtl. Software zur Erstellung von Klassendiagrammen⁴¹
- Klassendokumentation der Klasse „BinaryTree“

5.7.3. Ablauf

Der tabellarische Unterrichtsentwurf findet sich auf Seite 77. Für diese Unterrichtseinheit sind etwa 90 Minuten eingeplant.

Zu Beginn der Stunde stellen die Schülerinnen und Schüler ihre, in Pseudocode erstellten, Umsetzungen der Operationen vor. Dabei sollte am Ende eine

⁴¹Beispielsweise das umfangreiche Opensourceprojekt ArgoUML (<http://argouml.tigris.org/>).

konkrete Lösung für alle bekannt sein, so dass es eine belastbare Grundlage für die spätere Implementierung gibt.

Im Anschluss erstellen bzw. ergänzen die Schülerinnen und Schüler in Zweiergruppen ein Klassendiagramm mit den benötigten Klassen *BinarySearchTree*, *Item* und *Node*. Bei Item handelt es sich um eine abstrakte Klasse, die Realisierung erfolgt durch die Klasse Node. Die Klasse Node stellt ein Objekt im späteren Index dar.

Die erstellten Klassendiagramme werden dann im Plenum vorgestellt und zu einer gemeinsamen Basis zusammengeführt. Darauf folgt die eigentlich Implementierung. Hierfür kann auch auf eine modifizierte Version der Klasse *BinarySearchTree* aus der Materialsammlung für das Zentralabitur (NRW) zurückgegriffen werden. Um die Gruppe nicht mit einer unnötig komplexen Entwicklungsumgebung zu überfordern bietet sich die Verwendung einer reduzierten Entwicklungsumgebung, beispielsweise BlueJ an. Dabei besteht auch der Vorteil, dass auf die Erstellung aufwendiger Methoden für eine grafische Benutzeroberfläche verzichtet werden kann. BlueJ bietet Möglichkeiten die Methoden ohne weiteren Aufwand auszuprobieren.

Als Abschluss der Stunde werden über Probleme im Zusammenhang mit der Implementierung gesprochen. Möglicherweise zeigt sich, dass die Schülerinnen und Schüler mehr Zeit für die Implementierung benötigen. Darauf sollte, nach Möglichkeit, Rücksicht genommen werden.

5.7.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Daten und Algorithmen abstrahieren“ *Fachliche Inhalte – Modellieren und Konstruieren*
- „Lösungskonzepte nach einem Programmierkonzept realisieren, überprüfen und weiterentwickeln“ *Fachliche Inhalte – Modellieren und Konstruieren*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „erstellen informatische Modelle zu gegebenen Sachverhalten“ *Modellieren und Implementieren*
- ... „implementieren Modelle mit geeigneten Werkzeugen“ *Modellieren und Implementieren*
- ... „entwerfen und realisieren Algorithmen mit den algorithmischen Grundbausteinen und stellen diese geeignet dar“ *Algorithmen*
- ... „kooperieren bei der Lösung informatischer Probleme“ *Kommunizieren und Kooperieren*

5.8. UE 7 – Implementierung testen

In der nächsten Stunde sollen die Implementierungen auf Funktionsfähigkeit getestet werden. Dabei geht es nicht um einen vollständigen Klassentest im engeren Sinne. Im Verlauf der Stunde sollen Testdaten einzeln in den Suchbaum eingefügt werden, so dass dieser zu einer Liste, mit den bereits bekannten Nachteilen, entartet.

5.8.1. Lernziele

Die Schülerinnen und Schüler sollen in dieser Stunde zum einen erkennen, welche Bedeutung der Test eines selbst erstellten Programms hat. Weiterhin sollen sie erkennen, dass ein binärer Suchbaum, bei einer sortierten Eingabe, zu einer Liste entartet.

5.8.2. Material

Folgendes Material wird benötigt:

- Arbeitsblatt „Binärbaum testen“
- Vorbereitete Testumgebung

5.8.3. Ablauf

Der tabellarische Unterrichtsentwurf findet sich auf Seite 78. Für diese Unterrichtseinheit sind etwa 45 Minuten eingeplant.

Am Anfang dieser Unterrichtseinheit testen und korrigieren die Schülerinnen und Schüler die erstellten Implementierungen. Dabei sollen sie sich gegenseitig unterstützen. Bei größeren Problemen soll durch die Lehrkraft eingegriffen werden. Im nächsten Schritt werden durch die Schülerinnen und Schüler sortierte Daten in den Baum eingefügt, so dass dieser zu einer Liste entartet. Die Schülerinnen und Schüler protokollieren ihre Beobachtungen, die im anschließenden Unterrichtsgespräch diskutiert werden. Als Abschluss erfolgt in Form eines Lehrervortrags ein Hinweis auf andere Baumtypen, welche dieses Problem der Balancierung lösen,. Es besteht hier auch die Möglichkeit für einen längeren Exkurs zu anderen Baumtypen.

5.8.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Lösungskonzepte nach einem Programmierkonzept realisieren, überprüfen und weiterentwickeln“ *Fachliche Inhalte – Modellieren und Konstruieren*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „implementieren Modelle mit geeigneten Werkzeugen“ *Modellieren und Implementieren*
- ... „reflektieren Modelle und deren Implementierungen“ *Modellieren und Implementieren*
- ... „entwerfen und realisieren Algorithmen mit den algorithmischen Grundbausteinen und stellen diese geeignet dar“ *Algorithmen*

5.9. UE 8 – Sortieralgorithmen

In der achten Unterrichtseinheit wird die binäre Suche aus der vierten Unterrichtseinheit (Abschnitt 5.5, S. 37) wieder aufgegriffen. Dort wurde die Liste der Stoppwörter als sortiert angenommen. In der Realität liegen diese aber nicht unbedingt sortiert vor, so dass eine Sortierung erst stattfinden muss.

5.9.1. Lernziele

In dieser Stunde lernen die Schülerinnen und Schüler verschiedene Sortieralgorithmen kennen. Sie können nach der Stunde die grundsätzliche Idee der drei bearbeiteten Algorithmen, BubbleSort, InsertionSort und Quicksort wiedergeben, sowie Aussagen zur Effizienz der Algorithmen geben. Sie können weiterhin den Ablauf von InsertionSort beschreiben.

5.9.2. Material

Folgendes Material wird benötigt:

- Video „BubbleSort“ von AlgoRhythms
- Hilfen für die Recherche zu den Algorithmen
- Plakate und weiteres Material für einen Museumsgang
- Vorbereitete Testumgebung

5.9.3. Ablauf

Nach einer kurzen Einführung, in der auf die Motivation für Sortieralgorithmen Bezug genommen wird, wird der Gruppe ein Video vorgeführt, in der eine Gruppe von Tänzern BubbleSort ausführt. Die Schülerinnen und Schüler erhalten dabei den Auftrag, im Ablauf nach einer Verarbeitungsvorschrift zu suchen. Sie stellen danach ihre Beobachtungen vor. Dabei sollte eine grobe Handlungsvorschrift für BubbleSort erkennbar werden.

Daraufhin erhalten die Schülerinnen und Schüler den Auftrag, arbeitsteilig in drei Gruppen, eine Internetrecherche zu Informationen über die Sortieralgorithmen BubbleSort, InsertionsSort und QuickSort durchzuführen. Dabei

sollte auf die unterschiedliche Komplexität der Algorithmen Rücksicht genommen werden. Eine Möglichkeit zum Umgang mit diesem Problem wäre es beispielsweise, der QuickSort-Gruppe bereits eine Linkliste an die Hand zu geben, um ihnen so einen Vorteil zu verschaffen. Auch sollten die Informationen bzgl. der Eigenschaften von InsertionSort und BubbleSort tiefer gehen wohingegen bei QuickSort der Fokus auf den Ablauf gelegt werden kann. Die Gruppen bereiten ihre Ergebnisse für eine Präsentation in Plakatform vor.

Im Anschluss an die Recherche findet ein Museumsgang statt, bei dem gemischte Gruppen zu den einzelnen Plakaten gehen und jeweils die zugehörigen Mitglieder die Algorithmen vorstellen.

Die beiden folgenden Unterrichtsstunden haben wieder allgemeinere Themen zum Inhalt und können entweder als Abschluss der Reihe dienen oder auch bereits zwischendurch als Exkurs benutzt werden. Daher wird nun entweder zu diesen Stunden übergeleitet oder die Reihe beendet. Die Thematik der Sortieralgorithmen kann natürlich an dieser Stelle auch weitergeführt bzw. vertieft werden.

5.9.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Daten und Algorithmen abstrahieren“ *Fachliche Inhalte – Modellieren und Konstruieren*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „entwerfen und realisieren Algorithmen mit den algorithmischen Grundbausteinen und stellen diese geeignet dar“ *Algorithmen*
- ... „veranschaulichen informatische Sachverhalte“ *Darstellen und Interpretieren*

5.10. UE 9 – Ranking von Suchergebnissen

Diese Unterrichtseinheit beschäftigt sich mit der Aufbereitung von Suchergebnissen, dem Ranking. Die Anzahl der Dokumente die einen Suchbegriff enthalten ist häufig so groß, das es unmöglich ist auch nur einen kleinen Teil manuell zu begutachten. Diese Stunde soll einen Eindruck von den Möglichkeiten zur automatischen Untersuchung auf Relevanz vermitteln.

5.10.1. Lernziele

Die Schülerinnen und Schüler lernen in dieser Stunde verschiedene Methoden zum Ranking von Suchergebnissen kennen. Sie können am Ende der Stunde folgende Methoden benennen und die grundlegenden Konzepte bzw. Ideen erläutern:

- boolesches Retrieval
- Vektorraummodell
- PageRank

Außerdem werden die, aus der Stunde zur automatischen Textanalyse bekannten, Begriffe Termfrequenz und inverse Dokumentenhäufigkeit wieder aufgegriffen.

5.10.2. Material

Folgendes Material wird benötigt:

- Beispieldokumente, welche alle einen Begriff enthalten, aber teilweise nur am Rande mit diesem zu tun haben
- Informationen zum Thema Ranking (Siehe Anhang B.2 auf Seite 92)

5.10.3. Ablauf

Der tabellarische Unterrichtsentwurf findet sich auf Seite 80. Für diese Unterrichtseinheit sind etwa 45 Minuten eingeplant.

Als Impuls werden den Schülerinnen und Schülern drei verschiedene Texte vorgelegt, welche sich mit unterschiedlichen Themen befassen, aber einen gemeinsamen Begriff enthalten. Beispielsweise ein Text über Wirbelstürme, eine Rezension des Kinofilms Twister und ein Text über den US-Bundesstaat Nebraska. In allen drei Texten kommt der Begriff Tornado vor, thematisch gehört aber nur der erste zu einer Suchanfrage nach dem Begriff Tornado. Die Schülerinnen und Schüler können diese Sortierung selbst schnell vornehmen, da sie nur drei Dokumente zur Auswahl haben. Da die Anzahl der Dokumente im WWW deutlich größer ist, sind hier automatische Methoden notwendig. Die Lerngruppe erhält Texte zum Thema *Ranking* und bearbeitet diese im Sinne der Methode „Think-Pair-Share“. In dieser Stunde findet der Teil „Share“ allerdings nicht im Plenum sondern, je nach Kursstärke, in zwei bis drei Gruppen statt. Dort wird eine Präsentation der Methoden vorbereitet. Eine zufällig ausgewählte Gruppe präsentiert vor dem Plenum und wird von den anderen Schülerinnen und Schülern bei Bedarf ergänzt.

5.10.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Technische, funktionale und organisatorische Prinzipien von Hard- und Software kennen lernen und einordnen“ *Fachliche Inhalte – Analysieren und Bewerten*
- „Systematiken und Theorien zur Lösung spezifischer Anwendungssituationen“ *Lernen im Kontext der Anwendung*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- „kommunizieren fachgerecht über informative Sachverhalte“ *Kommunizieren und Kooperieren*
- „erkennen und nutzen Verbindungen innerhalb und außerhalb der Informatik“ *Strukturieren und Vernetzen*

- ... „begründen Entscheidungen bei der Nutzung von Informatiksystemen“ *Begründen und Bewerten*

5.11. UE 10 – Informationsqualität im Internet

Die als letzte vorgestellte Stunde hat die Problematik der unklaren Qualität von Internetquellen zum Thema. Es werden verschiedene Positionen innerhalb der Debatte vorgestellt.

5.11.1. Lernziele

In dieser Stunde sollen die Schülerinnen und Schüler erkennen, dass sie nicht jeder Informationsquelle vertrauen können. Dabei sollen sie verinnerlichen, dass es nicht im wesentlichen auf den Übertragungsweg von Informationen ankommt, sondern auf die Vertrauenswürdigkeit der Quellen. Die Schülerinnen und Schüler achten nach dieser Stunde bei Recherchen im Internet, aber auch an anderen Stellen, darauf, wer die Informationen veröffentlicht hat, welche eigenen Interessen möglicherweise verfolgt werden und welche Quellen durch den Autor verwendet bzw. nicht verwendet wurden.

5.11.2. Material

Folgende Materialien werden benötigt:

- Texte zum Thema Informationsqualität: „Informationsqualität im Internet“ – Staatskanzlei NRW 2009, „Lügen im Netz“ – Ude 2008 (Diese und weitere Texte finden sich als Referenz im Anhang B.3.1 auf Seite 95 und elektronisch auf dem beiliegenden Datenträger.)
- Rollenkarten für die verschiedenen Rollen während der Diskussion (Anhang B.3.2 auf Seite 96)

5.11.3. Ablauf

Der tabellarische Unterrichtsentwurf findet sich auf Seite 81. Für diese Unterrichtseinheit sind etwa 90 Minuten eingeplant.

Zum Einstieg in diese Stunde wird auf die Stunde zum Thema Ranking Bezug genommen. Es steht die Frage im Raum, ob Dokumente mit einem hohen Ranking auch automatisch Dokumente mit hoher Qualität sind.

Zuerst verschaffen sich die Schülerinnen und Schüler einen ersten Überblick über diese Problematik mit der Bearbeitung von zwei Texten. Dafür steht ein von der Staatskanzlei des Landes NRW finanziert Text und ein Artikel aus der Wochenzeitung *Die Zeit* vom 31. Januar 2008 zur Verfügung. Im Verlauf vertiefen die Schülerinnen und Schüler die Informationen auf der Grundlage eigener Recherchen und weiterer Texte (Siehe Anhang B.3.1) und tauschen sich mit ihren Mitschülern im Think-Pair-Share-Verfahren aus.

Für den nächsten Abschnitt der Stunde bereiten die Schülerinnen und Schüler in vier Gruppen ein kleines Rollenspiel vor. Anhand der Materialien und einer Rollenkarte werden die Positionen von vier fiktiven Personen vorbereitet (Siehe Anhang B.3.2). Als Rollen werden ein Blogger, ein Wissenschaftler, ein Zeitungsverleger und ein Wikipedia-Autor vorgeschlagen. Im Anschluss daran wird eine kurze Podiumsdiskussion durchgeführt.

Nachdem die Schülerinnen und Schüler aus ihren Rollen entlassen wurden und sie über ihre Erfahrungen in der Rolle berichtet haben, wird in einem Unterrichtsgespräch ein Fazit gezogen. Dabei sollte zumindest herausgearbeitet werden, dass eine Informationsrecherche immer auch eine Bewertung der Informationen erfordert.

5.11.4. Bezug auf Lehrplan und Standards

Lehrplan Bezug auf den Lehrplan Informatik:

- „Erzeugung eines Gesamtbildes von Informatik ausgehend von der Anwendungsvielfalt“ *Lernen im Kontext der Anwendung*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „benennen Wechselwirkungen zwischen Informatiksystemen und ihrer gesellschaftlichen Einbettung“ *Informatik, Mensch und Gesellschaft*

- ... „reagieren angemessen auf Risiken bei der Nutzung von Informatiksystemen“ *Informatik, Mensch und Gesellschaft*
- ... „stellen Fragen und äußern Vermutungen über informative Sachverhalte“ *Begründen und Bewerten*
- ... „reagieren angemessen auf Risiken bei der Nutzung von Informatiksystemen“ *Informatik, Mensch und Gesellschaft*
- ... „begründen Entscheidungen bei der Nutzung von Informatiksystemen“ *Begründen und Bewerten*

6. Bewertung der Unterrichtsreihe

Für eine erste Bewertung der entworfenen Unterrichtsreihe, wurde Kapitel 5 einschließlich der Anhänge A und B, mit der Bitte um eine Rückmeldung im Bezug auf die Einhaltung der IniK-Kriterien und der generellen Unterrichtbarkeit der Reihe, an Prof. Dr. Jochen Koubek (Universität Bayreuth) und Dr. Dieter Engbring (Universität Paderborn) geschickt. In diesem Kapitel werden die erhaltenen Rückmeldungen zusammengefasst. Daran anschließend folgt jeweils eine Stellungnahme zur erhaltenen Rückmeldung. In einem weiteren Kapitel werden Vorschläge für Veränderungen und Ergänzungen der Reihe gemacht. Dabei wird auf die erhaltenen Rückmeldungen Bezug genommen.

6.1. Rückmeldung von Koubek

Die Rückmeldung liegt als Anhang auf Seite 98 vor. Koubek beschreibt den Kontext *Suchmaschinen* als generell gut geeignet für eine IniK-Unterrichtsreihe. Er bemängelt allerdings, dass die Möglichkeiten des Kontextes nicht genutzt werden. Suchmaschinen werden nur als Aufhänger zu Beginn der Reihe benutzt und verschwinden dann bis zur zehnten Unterrichtseinheit aus dem Fokus. Dazwischen wird aus Koubeks Sicht nur auf die technische Dimension eingegangen.

Die Inhalte werden zwar durchaus als wichtig erachtet, sie dürfen aber den Kontext nicht vollständig verdrängen. Der Kontext muss während der gesam-

ten Reihe greifbar bleiben und darf nicht nur als Ausgangspunkt dienen. Es muss dabei nicht in jeder Stunde der Reihe gleichstark auf den Kontext eingegangen werden, aber die Verbindung muss eindeutig erkennbar sein. Die zugehörigen Kompetenzen der Bildungsstandards sind im Allgemeinen die im Inhaltsbereich *Informatik, Mensch und Gesellschaft* genannten. Diese Inhalte finden sich laut Koubek aber erst am Ende der Reihe wieder. Diese Konstruktion birgt dann ganz besonders die Gefahr, dass die wichtigen, kontextbezogenen Kenntnisse und Kompetenzen nicht mehr erworben werden, da die zugehörigen Stunden anderen Terminen zum Opfer fallen.

Die vorgelegte Reihe orientiert sich durchaus an den Bildungsstandards und dem Anspruch eines methodisch vielfältigen Unterrichts, was positiv zu bewerten ist. Jedoch ist für IniK der Bezug zum Kontext und zur Lebenswelt der Schüler ein entscheidendes Kriterium, welches nicht ausreichend Berücksichtigung gefunden hat.

6.2. Stellungnahme zur Kritik von Koubek

Koubek kritisiert einen zu großen Anteil der technischen Aspekte von Suchmaschinen. Dies ist nachvollziehbar. Bei Betrachtung der Überschriften der Unterrichtseinheiten fällt ein Übergewicht an informationstechnischen Inhalten auf. Viele Einheiten bieten m. E. das Potenzial für tiefere Gesellschaftsbezüge. Einige wurden in Kapitel 3 genannt. Diese Bezüge werden in der Beschreibung der Reihe aber bestenfalls implizit deutlich. Eine Konkretisierung wird in Kapitel 7 vorgestellt.

In der Beschreibung wird nicht explizit genannt, dass die vorliegende Reihe für die Sekundarstufe II konzipiert ist. Besonders bei Kursen, deren Teilnehmer die schriftliche Abiturprüfung ablegen, ist es im Bezug auf die schulische Realität notwendig, sich stark an den entsprechenden Vorgaben zu orientieren. Leider fehlt der Bezug zur Kontextorientierung in den Vorgaben für NRW (siehe Abschnitt 4.2). Der Konflikt zwischen einer Orientierung an der Fachsystematik und einer Orientierung an Kontexten wurde in Abschnitt 2.2.4 bereits genannt. All dies darf aber natürlich kein Argument sein, in einer Reihe, die dem Konzept IniK folgen will, den Kontext aus den Augen zu verlieren.

6.3. Rückmeldung von Engbring

Die Rückmeldung liegt als Anhang auf Seite 98 vor. Engbring beschreibt die vorgelegte Reihe als sehr anspruchsvoll. Bezugnehmend auf die Vorgaben durch Lehrpläne und Voraussetzungen für das Zentralabitur wird deutlich, dass die Reihe vom inhaltlichen Anspruch für einen Leistungskurs oder einen sehr leistungsstarken Grundkurs konzipiert ist.

Er weist darauf hin, dass IniK häufig als spezifisch für die Sekundarstufe I angesehen wird (vgl. dazu auch Abschnitt 2.2). Engbring sieht in IniK jedoch auch eine gute Möglichkeit, Inhalte der Sekundarstufe II zu vermitteln. Dadurch kann den Schülern der Anwendungsbezug der theoretischen Konzepte verdeutlicht werden. Er stellt im Bezug auf die Reihe positiv heraus, dass die Informatikinhalte stark betont werden, auch wenn dadurch der Anspruch sehr hoch ist.

Wie Koubek bewertet auch Engbring den Einsatz der verschiedenen Unterrichtsmethoden positiv. Die Methoden sorgen für einen schülerorientierten Unterricht, setzen aber unbedingt voraus, dass sie vom Lehrenden beherrscht werden. Auch das für IniK vorgeschlagene Schema mit einer Aufteilung in verschiedene Phasen erkennt Engbring im vorgelegten Entwurf.

Die Tatsache, dass einige Unterrichtseinheiten darauf ausgelegt sind, ohne den Einsatz eines Computers auszukommen, wird von Engbring als nützlich angesehen. Er geht davon aus, dass besonders im Grundkursbereich Schüler dadurch weniger leicht abgehängt werden.

Bezugnehmend auf die Implementierungsaufgaben stellt er fest, dass der Aufwand hier sehr groß ist. Besonders leistungsschwächere Schüler erfahren dabei Nachteile.

Weiterhin merkt Engbring an, dass es seiner Ansicht nach sinnvoll ist, den Bezug zu Bildungsstandards und Lehrplänen bzw. Voraussetzungen nicht nur zu benennen, sondern konkreter zu erläutern. Damit würde auch eine Präzisierung der Lernziele sowie des Unterrichtsablaufs erreicht.

6.4. Stellungnahme zur Kritik von Engbring

Engbring wurde in seiner Eigenschaft als aktiver Lehrer gebeten, sich bei seiner Rückmeldung im Wesentlichen auf die Frage nach einer generellen Unterrichtbarkeit zu beziehen. Die Frage nach der Einhaltung der IniK-Kriterien sollte nur an nachgeordneter Stelle betrachtet werden.

Engbring sieht in der Reihe einen Weg, Inhalte, die in der gymnasialen Oberstufe verortet sind, praxisbezogen an Schüler zu vermitteln. Die entwickelte Reihe hat genau dies zum Ziel. Die vom Lehrplan geforderten Inhalte sollen in einem sinnstiftenden Zusammenhang vermittelt werden. Gleichzeitig soll ein Blick über den Tellerrand der Kerninformatik ermöglicht werden.

Die Befürchtungen in Bezug auf eine mögliche Überforderung der Schülerrinnen und Schüler durch zu anspruchsvolle Unterrichtseinheiten ist nachvollziehbar. Besonders während der Phasen der Implementierung können ungeahnte Probleme auftreten. Es ist zu überdenken, ob für diese Phasen mehr Zeit eingeräumt werden sollte, oder ob es sich als sinnvoller erweist, den Anspruch an dieser Stelle zu verringern.

Letztendlich ist aber eine abschließende Bewertung nur mit einer konkreten Durchführung im Unterricht möglich. Dabei muss dann, wie beschrieben, auf die konkrete Lerngruppe eingegangen werden.

7. Überarbeitung der Unterrichtsreihe

Im folgenden Kapitel werden einzelne Unterrichtseinheiten ergänzt und es werden zwei zusätzliche Unterrichtseinheiten vorgestellt. In Kapitel 6 wurde von Koubek angeregt, den Kontext in der Unterrichtsreihe stärker zu berücksichtigen. Ein weiterer Ansatz um dieses Ziel zu erreichen ist es, Dazu könnten u.a. die mehr informationstechnisch ausgerichteten Unterrichtseinheiten in Exkurse ausgelagert werden. Eine weitere Option für die Lehrkraft ist es, die eher theoretischen Inhalte nur anzureißen und, im Sinne eines Spiralcurriculums, zu einem späteren Zeitpunkt wieder aufzugreifen.

In diesem Zusammenhang stellt sich m. E. die grundsätzliche Frage, ob das Ziel verfolgt werden soll, den gesamten Unterricht an den Kriterien von IniK

auszurichten. Zunächst stellt sich die grundsätzliche Frage, wie IniK am sinnvollsten im Schulalltag eingesetzt wird. Unterricht ausschließlich konsequent am Kontext zu orientieren, ist in der Oberstufe m. E. schwer umsetzbar: Es besteht schon aus praktischen Gesichtspunkten, unter Beachtung der Obligatorik, die Notwendigkeit, die informationstechnischen Inhalte stark zu betonen. Dabei ist dann auch zu beachten, dass die Richtlinien für den Unterricht in der gymnasialen Oberstufe eine stärker wissenschaftspropädeutische Ausrichtung des Unterrichts fordern.⁴² Es ist also nicht möglich, nur Unterricht durchzuführen, der sich ausschließlich an IniK orientiert. Alternativ könnte man eine IniK-Reihe nur gelegentlich zwischen den ansonsten klassischen informationstechnischen Unterricht legen. Dann tritt jedoch m. E. das von Koubek genannte Problem der in die Unendlichkeit verschobenen kontextorientierten Elemente für die gesamte Reihe auf.⁴³ Ein guter Kompromiss ist es, nur an IniK orientierte Unterrichtsreihen durchzuführen, sich aber bei der Umsetzung Abweichungen zu erlauben. Wünschenswert ist auch eine stärkere Berücksichtigung der Kontexte in der Obligatorik.

7.1. Alternative Gestaltung einzelner Einheiten

Im folgenden Abschnitt werden exemplarisch einige der bereits vorliegenden Unterrichtseinheiten ergänzt bzw. abgewandelt.

7.1.1. Überarbeitet: UE 1 – Aufbau von Suchmaschinen

In dieser Einheit sollen die Schülerinnen und Schüler das Informatiksystem Suchmaschine nicht mehr nur als BlackBox begreifen. Sie lernen dabei sowohl die technische Teilung des Systems kennen, als auch die Auswirkungen der Funktionsweise für ihren Umgang mit dem System. Der Unterschied zwischen einer Suche in Echtzeit und einer Suche in einem zuvor erstellten Index muss stärker betont werden. Eine neue Internetseite wird nicht sofort in einer Suchmaschine sichtbar. Evtl. fällt hier auf, dass manche Seiten aktueller im Index

⁴²Vgl. MfSuWWF NRW 1999, S. XI.

⁴³Vgl. Anhang C.1

vertreten sind, als andere. Beispielhaft können hier der Microbloggingdienst Twitter, Internetauftritte von Zeitungen oder die Wikipedia genannt werden.

Bei der Betrachtung der Analogie zwischen dem Suchen in der Kiste und dem Suchen im WWW fällt möglicherweise auch ein Bruch auf, wenn klar wird, dass die Verlinkungen, welche für das Crawling notwendig sind, in der Materialkiste wenig bis gar nicht vorhanden sind.

Durch diese Verschiebung der Perspektive soll der Bezug zur Lebenswelt der Schülerinnen und Schüler deutlicher werden. Damit wird auch der Bezug zu den Bildungsstandards im Bereich *Informatik-Mensch-Gesellschaft* ergänzt. Konkret findet sich folgender Punkt hier wieder: „Schülerinnen und Schüler [...] kommentieren automatisierte Vorgänge und beurteilen deren Umsetzung“ (GI e. V. 2008, S. 42)

7.1.2. Überarbeitet: UE 2 – Durchsuchen des WWW: Crawling

Lernziele Das Resultat des Crawling ist ein Abbild eines Teils des WWW. Den Schülerinnen und Schülern soll, neben den technischen Aspekten einer Breiten- und Tiefensuche, deutlich werden, dass das genannte Abbild nicht den aktuellen Stand darstellen kann. Wenn sie eine Suchanfrage starten, werden sie also niemals sicher sein, ob sie wirklich die aktuellsten Ergebnisse finden oder ob die angezeigte Liste vollständig ist. Auch die Macht der Betreiber von Suchmaschinen sollte an dieser Stelle schon Erwähnung finden. Wenn bestimmte Seiten bewusst nicht vom Crawler besucht werden, verschwinden sie für viele Nutzer des WWW von der Bildfläche.

Wenn Schülerinnen und Schüler eigene Inhalte im WWW veröffentlichen wollen, so ist ihnen im Allgemeinen auch daran gelegen, dass diese Inhalte gefunden werden können. Nach dieser Unterrichtsstunde ist ihnen deutlich, dass dafür die Verlinkung von anderen Seiten im Netz zwingend notwendig ist.

Ablauf Ein entsprechend angepasster tabellarischer Unterrichtsentwurf findet sich in Anhang A.11 auf Seite 82.

Ein guter Einstieg in die Stunde ist es, die fehlende Aktualität von Suchmaschinenindexen zu zeigen. Dafür ist es notwendig, dass eine neu erstellte Internetseite bekannt ist, die aber noch nicht unmittelbar im Index einer

Suchmaschine zu finden ist. Eine Möglichkeit ist es, die Liste der neuen Artikel der Wikipedia zu nutzen. Wenn dann mit dem Titel eines neuen Artikels als Suchanfrage gesucht wird, fällt auf, dass dieser Artikel noch nicht von der Suchmaschine gefunden wurde. Problematisch ist bei diesem Vorgehen, dass neue Artikel der Wikipedia sehr schnell in den Index aufgenommen werden. Zielführende Ergebnisse konnten mit Yahoo! und der Einstellung „Seiten auf deutsch“ erreicht werden.

Ein zufällig ausgewählter Artikel wird hingegen problemlos gefunden. Die Tatsache, dass neue Artikel binnen weniger Minuten zumindest von Google gefunden werden, unterstreicht die in der ersten Einheit genannten Unterschiede in der Aktualität des Index im Bezug auf verschiedene Webseiten.

7.1.3. Überarbeitet: UE 3 – Automatische Textanalyse

Lernziele Die Schülerinnen und Schüler erfahren in dieser Stunde, dass sie selbst zwar einfach einen Text in wenigen Wörtern zusammenfassen können, ein Computer diese Fähigkeit aber nicht hat. Die Maschine besitzt keine abrufbaren Vorkenntnisse, um zu erkennen, welche Wörter eines Textes wichtig sind und welche nicht. Die Schülerinnen und Schüler erkennen, dass der vorgestellte Volltextindex allein keine sinnvolle Methode zur Lösung des Problems ist. Sie lernen dann die Möglichkeiten kennen, dieses Problem dennoch erfolgreich zu lösen. Dabei liegt der Fokus dann auf den Unterschieden zwischen menschlicher und künstlicher Intelligenz.

Ablauf Ein entsprechend angepasster tabellarischer Unterrichtsentwurf findet sich in Anhang A.12 auf Seite 83.

Die Unterrichtsstunde beginnt jetzt mit einem Arbeitsauftrag. Die Schülerinnen und Schüler erhalten einen kurzen Text und sollen die fünf wichtigsten Wörter des Textes herausschreiben. Dafür bietet sich zum Beispiel ein kurzer Ausschnitt aus einem aktuellen Zeitungsartikel an. Die Schülerinnen und Schüler überprüfen, ob sie die gleichen Wörter ausgewählt haben. Eine hohe Übereinstimmung lässt auf eine treffsichere Auswahl schließen. Im Anschluss daran wird als Frage in den Raum gestellt, ob ein Computer einen ähnlichen Index erstellen kann. Die intuitive Antwort wird vermutlich negativ ausfallen.

Daraufhin erhalten die Schülerinnen und Schüler einen Volltextindex, welcher sich auch maschinell leicht erstellen lässt.

Der weitere Ablauf gleicht dem bereits vorgestellten. Im Rahmen der Sicherung sollte noch einmal auf die oben genannten Lernziele eingegangen werden.

Auch hier liegt wieder ein Bezug zu folgendem Bereich der Bildungsstandards vor: „Schülerinnen und Schüler [...] kommentieren automatisierte Vorgänge und beurteilen deren Umsetzung“ (GI e. V. 2008, S. 42)

7.1.4. Überarbeitet: UE 9 – Ranking

Zusätzlich zu den bereits genannten Zielen liegt es nahe, den Schülerinnen und Schülern in dieser Unterrichtsstunde zu vermitteln, dass die enorme Anzahl der potenziellen Suchergebnisse zu einer Art Abhängigkeit vom Rankingalgorithmus führt. Niemand wird alle Suchergebnisse untersuchen, da deren Anzahl häufig sechsstellig oder noch größer ist. Dadurch muss man sich aber als Nutzer darauf verlassen, dass die zuerst genannten Ergebnisse wirklich auch diejenigen mit der größten Relevanz sind. Damit wird bereits zu einer der zusätzlichen Unterrichtseinheiten übergeleitet, welche die Macht von Suchmaschinenbetreibern zum Thema hat.

Um dieses Ziel zu erreichen, sollte die verwendete Textgrundlage (siehe Anhang B.2) angepasst werden.

7.2. Weitere Unterrichtseinheiten

Im Folgenden werden zwei Unterrichtseinheiten vorgestellt, die den Kontextbezug der Reihe stärker herausstellen sollen.

7.2.1. Umweltschutz durch effiziente Algorithmen

In dieser Stunde geht es um den Zusammenhang von effizienten Datenstrukturen und Algorithmen mit „handfesten“ Zielen wie dem Sparen von Energie und Kosten. Diese Stunde kann zwischen den verschiedenen eher technischen Einheiten eingeschoben werden.

Vor einigen Jahren stand der Energieverbrauch der IT-Industrie einige Tage im Fokus der Medien. Schlagzeilen wie „Einmal googlen entspricht einer

Stunde Licht“ waren in vielen Zeitungen und Magazinen zu finden. Ob die grundeliegenden Statistiken belastbar sind und ob der direkte Zusammenhang zwischen einer einzelnen Suchanfrage und dem angegebenen Energieverbrauch so eindeutig ist, wie er dargestellt wird, darf in Frage gestellt werden. Dennoch ist klar, dass auch der Betrieb einer Suchmaschine nicht ohne die Nutzung von Energie möglich ist. Der Energieverbrauch steigt mit dem Aufwand der zum Erreichen der Ergebnisse nötig ist. An dieser Stelle kommen effiziente Datenstrukturen und Algorithmen ins Spiel. Diese sind notwendig, um nicht mehr Energie als nötig zu verbrauchen.⁴⁴

Lernziele Die Schülerinnen und Schüler lernen in dieser Einheit den Zusammenhang zwischen theoretischen Modellen und realen Erfordernissen kennen. Dabei finden sowohl ökonomische als auch ökologische Ziele ihren Platz. Die Gruppe kann nach dieser Stunde argumentativ darlegen, aus welchen Gründen die Anwendung möglichst sparsamer Algorithmen und Datenstrukturen erstrebenswert ist.

Material Folgendes Material wird benötigt:

- Artikel „Einmal googlen entspricht einer Stunde Licht“ – Schultz, In: Spiegel Online 2007⁴⁵

Ablauf Der tabellarische Unterrichtsentwurf findet sich auf Seite 84. Für diese Unterrichtseinheit sind etwa 45 Minuten eingeplant.

Die Schülerinnen und Schülern erhalten als Einstieg in die Unterrichtsstunde den Artikel „Einmal googlen entspricht einer Stunde Licht“. Der Artikel wird in Einzelarbeit bearbeitet. Im anschließenden Unterrichtsgespräch werden spontane Schüleräußerungen gesammelt. Daran anschließend wird die Diskussion auf den Zusammenhang zwischen verwendeter Software und dem Energiebedarf gelenkt. Bei Bedarf kann auf unterschiedlich starke Lüfteraktivität bei der Nutzung verschiedener Anwendungen verwiesen werden. Im Verlauf

⁴⁴Auf theoretischer Ebene beschäftigt sich der Sonderforschungsbereich 876 – *Verfügbarkeit von Information durch Analyse unter Resourcenbeschränkung* an der TU Dortmund, Teilprojekt A2, auch mit dieser Thematik.

⁴⁵<http://www.spiegel.de/netzwelt/tech/0,1518,492078,00.html>.

des Gesprächs sollte zur Sprache kommen, dass die Benutzung von mehr Hardware zur Kompensation ineffizienter Algorithmen, selbst wenn sie funktioniert, auch ökonomische wie ökologische Nachteile birgt. Die Gruppe erhält dann den Auftrag, einem fiktiven, jüngeren Schüler zu erläutern, warum es im Allgemeinen sinnvoll ist, Software effizient zu gestalten. Die Lehrkraft übernimmt dann die Rolle des jüngeren Schülers und lässt sich die Argumentation von einem/r zufällig ausgewählten SchülerIn erläutern. Die Argumentation kann im Anschluss im Plenum diskutiert werden.

Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Fachübergreifende und fächerverbindende Sichtweisen“ *Methoden und Formen selbstständigen Arbeitens*
- „Selbstständige und projektorientierte Arbeitsformen“ *Methoden und Formen selbstständigen Arbeitens*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- ... „stützen ihre Argumente auf erworbenes Fachwissen“ *Begründen und Bewerten*
- ... „untersuchen an Beispielen die Probleme der Produktion, Nutzung und Entsorgung elektronischer Geräte“ *Informatik, Mensch und Gesellschaft*

7.2.2. Ranking 2 – Wer bestimmt was ich finden kann?

Wie in Unterrichtseinheit 9 (siehe Abschnitt 5.10 und 7.1.4) bereits angesprochen, hat eine Suchmaschine und damit auch das verwendete Ranking einen starken Einfluss darauf, welche Inhalte für einen Nutzer tatsächlich erreichbar sind.

An dieser Stelle gibt es verschiedenen Interessenten, die versuchen (können), Einfluss zu nehmen. Anbieter, welche aus gewerblichem Interesse eine Webseite betreiben, sind im Allgemeinen daran interessiert, dass ihr Webangebot bei möglichst vielen Anfragen weit oben erscheint. Dabei wird nicht unbedingt nur darauf zurückgegriffen, relevante Inhalte anzubieten und eine gute Vernetzung zu erreichen. Manipulationsversuche hat es in der Vergangenheit immer wieder gegeben. Auch Manipulationsversuche bei denen bestimmte negative Begriffe mit einem (politischen) Gegner verbunden wurden, hat es bereits mehrfach gegeben. Diese Manipulationen sind unter dem Begriff „Google-Bombe“ bekannt geworden. Betreiber von Suchmaschinen versuchen diesen Bestrebungen durch Verbesserungen ihrer Rankingverfahren entgegenzuwirken. Allerdings könnten auch die wirtschaftlichen Interessen der Betreiber von Suchmaschinen dazu führen, dass beispielsweise bezahlte Anzeigen nicht länger deutlich als solche gekennzeichnet werden. Ebenso können staatliche Stellen versuchen auf die Betreiber der Suchmaschinen einzuwirken, um eine Zensur zu erreichen. Ein weiteres aktuelles Beispiel für die Macht von Suchmaschinen ist der jüngst aufgetretene Konflikt zwischen belgischen Zeitungsverlegern und Google. Dort hatte Google aufgrund eines verlorenen Prozesses die betroffenen Zeitungen kurzerhand komplett aus dem Index gestrichen.⁴⁶

Lernziele Die Schülerinnen und Schüler erfahren in dieser Stunde, dass sie keinesfalls sicher sein können, mit Hilfe einer Suchmaschine immer alle relevanten Informationen zu einem Inhaltsbereich finden zu können. Sie erkennen, dass die Reihenfolge von Suchergebnissen auf Manipulationen zurückgehen kann. Sie können nach der Unterrichtsstunde die verschiedenen Interessenlagen beschreiben und Beispiele für die Auswirkungen dieser Interessen nennen.

Material Folgendes Material wird benötigt:

- Liste mit funktionierenden „Google-Bomben“ für den Impuls

⁴⁶Vgl. beispielsweise <http://heise.de/-1280764> – geprüft am 23. Juli 2011.

- evtl. alternative Texte, falls die Recherche nicht stattfinden kann oder zu keinen Ergebnissen führen sollte

Ablauf Der tabellarische Unterrichtsentwurf findet sich auf Seite 85. Für diese Unterrichtseinheit sind etwa 45 Minuten eingeplant.

Die Gruppe erhält als Impuls den Auftrag, nach mehreren Begriffen zu suchen und die erhaltenen Ergebnislisten kritisch zu betrachten. Da die einzelnen „Google-Bomben“ nicht mehr zwingend bei jeder Suchmaschine funktionieren, müssen die Suchanfragen ausführlich getestet werden und den Schülerinnen und Schüler für jede Anfrage die zu benutzende Suchmaschine vorgegeben werden. Im Juli 2011 wurden mit Google die Anfragen „verfassungsfeindliche Partei“ (CDU auf Platz 1) und „creator of the universe“ („Church of the Flying Spaghetti Monster“, auf Platz 3) und mit Yahoo! die Anfragen „Verräterpartei“ (SPD auf Platz 1) „miserable failure“ (George W. Bush auf Platz 8) erfolgreich getestet. Im anschließenden Unterrichtsgespräch wird herausgearbeitet, dass diese Ergebnisse offensichtlich nicht aufgrund des Inhalts der Webseiten entstanden sind.

Die Schülerinnen und Schüler werden dann in zwei Gruppen aufgeteilt und erhalten den Auftrag weitere Manipulationen von Suchergebnissen zu recherchieren. Dabei legt eine Gruppe den Schwerpunkt auf politische motivierte Manipulationen, die andere Gruppe bearbeitet wirtschaftlich motivierte Manipulationen.

Die gesammelten Informationen werden während eines weiteren Unterrichtsgesprächs in Form einer Mindmap zusammengestellt und so gesichert.

Bezug auf Lehrplan und Standards

Lehrplan Bezug zum Lehrplan Informatik:

- „Aufdecken der Funktionsweise bekannter Werkzeuge“ *Methoden und Formen selbstständigen Arbeitens*
- „Fachübergreifende und fächerverbindende Sichtweisen“ *Methoden und Formen selbstständigen Arbeitens*

- „Selbstständige und projektorientierte Arbeitsformen“ *Methoden und Formen selbstständigen Arbeitens*

Bildungsstandards Bezug zu den Bildungsstandards:

Schülerinnen und Schüler ...

- … „benennen Wechselwirkungen zwischen Informatiksystemen und ihrer gesellschaftlichen Einbettung“ *Informatik, Mensch und Gesellschaft*
- … „wenden Kriterien an, um Seriosität und Authentizität von Informationen aus dem Internet zu beurteilen“ *Informatik, Mensch und Gesellschaft*

Mit den genannten Vorschlägen wird m. E. sowohl der Kontextbezug einzelner Einheiten, als auch der der gesamten Reihe gestärkt.

8. Fazit und Ausblick

In der vorliegenden Arbeit wurde eine Unterrichtsreihe entworfen, die zum Ziel hat, kontextorientiert verschiedene Inhalte der Informatik zu vermitteln. Dabei wurden informationstechnische Inhalte berücksichtigt, aber auch Inhalte, welche eher an den Schnittstellen zu anderen Disziplinen – in dieser Reihe regelmäßig die Sozialwissenschaften – zu finden sind.

Suchmaschinen eignen sich grundsätzlich gut als Kontext für eine Unterrichtsreihe, da sie Schülerinnen allgemein bekannt sind, aber eine überraschende theoretische Tiefe besitzen. Die schwierigen theoretischen Hintergründe haben in der vorliegenden Arbeit allerdings dazu geführt, dass die Unterrichtsreihe sehr anspruchsvoll wurde. Es ist durchaus möglich, dass an einigen Stellen mehr Zeit als veranschlagt benötigt wird. Während der Auswertung der Reihe sind weitere Probleme mit der Gestaltung einer Reihe zum Thema Suchmaschinen aufgefallen. Es hat sich als schwierig erwiesen, dem Kontext als auch der Fachsystematik gleichermaßen gerecht zu werden. Wenn die schwierigen theoretischen Inhalte im Unterricht ausführlich behandelt werden, ist es schwer, die Anbindung an den Kontext nicht zu verlieren. Gleichzeitig soll

die Reihe unbedingt nicht nur einem fachdidaktischen Konzept folgen, sondern auch in der Unterrichtspraxis nutzbar sein. Damit unterliegt sie den Sachzwängen des Schulalltags, welche sich an verschiedenen Konfliktlinien manifestieren. Zu nennen sind hier die Ansprüche der Schüler an eine konkrete Prüfungsvorbereitung, die der Fachkonferenz an ein von allen Lehrerinnen unterrichtbares internes Curriculum, die Vorgaben des zuständigen Ministeriums und der Anspruch der Lehrerin, einen guten Unterricht in *allen* von ihr unterrichteten Fächern zu gestalten.

Die genannten Schwierigkeiten sind kein Argument, die Bemühungen einzustellen. Der generelle Widerspruch zwischen Fachsystematik und einer durch die Anwendung vorgegebenen Systematik ist in allen Fächern erkennbar. Mit diesem Widerspruch muss umgegangen werden. Sowohl die Realität in Schulen, als auch die Obligatorik unterliegen stetigen Veränderungen. Die Weiterentwicklung der Schulinformatik voranzutreiben ist eine der Aufgaben der Fachdidaktik. Eine Entschärfung der Problematik würde m. E. eintreten, wenn die Forderung nach einer stärkeren Berücksichtigung des Themenbereichs *Informatik und Gesellschaft* innerhalb der Lehrpläne und der Prüfungsaufgaben im Zentralabitur erfüllt wird.

Der Fokussierung auf die Fachsystematik liegt sicher auch die wissenschaftspropädeutische (weil studienvorbereitende) Ausrichtung der gymnasialen Oberstufe zu Grunde. Es lässt sich beobachten, dass viele Schulabgängerinnen trotz der Berechtigung ein Studium aufzunehmen, einen Ausbildungsberuf ergreifen. In wie weit dadurch die Legitimation der Oberstufe in der aktuellen Form zumindest teilweise verloren geht, ist eine weiterführende Fragestellung, welche in dieser Arbeit keine Beachtung gefunden hat.

Aufgrund der Hinweise von Koubek und Engbring wurde die Reihe überarbeitet. Dabei sind sowohl zwei Unterrichtseinheiten hinzugekommen, als auch einige Einheiten abgewandelt worden. Der Kontextbezug ist m. E. jetzt deutlicher zu erkennen.

Im Hinblick auf die zu Beginn der Arbeit geäußerte Fragestellung lässt sich feststellen, dass der Kontext Suchmaschinen grundsätzlich für eine kontextorientierte Unterrichtsreihe geeignet ist. Es bieten sich viele verschiedene Inhalte an, von denen in dieser Reihe nur ein Teil genutzt wurde. Dabei lässt sich

sowohl der informationstechnische Anteil als auch der Bezug zu anderen Disziplinen noch erweitern. Inwieweit die Kombination des Konzepts IniK mit dem Kontext Suchmaschinen und dem Unterricht in der gymnasialen Oberstufe vollständig kompatibel ist, wurde nicht abschließend geklärt. In jedem Fall gibt es Schwierigkeiten, welche Beachtung finden müssen. Für eine umfassendere Beurteilung ist es auch unbedingt notwendig, die Unterrichtsreihe in der Anwendung zu evaluieren. In diesem Fall bietet sich ein viel zitiert Spruch vom früher in Dortmund wie in Münster aktiven Fußballer Alfred „Adi“ Preißler an, auf den mich auch Herr Engbring in diesem Zusammenhang bereits hingewiesen hat: „Grau ist alle Theorie – entscheidend is' auf'm Platz.“

Für die Weiterentwicklung der Reihe ist neben den bereits angesprochenen Punkten auch ein Ausbau der Materialsammlung erstrebenswert. Neben weiteren Textgrundlagen bietet sich m. E. auch eine Software für die Simulation der beschriebenen Funktionen einer Suchmaschine an. Auf diese Weise könnten auch der Einfluss der einzelnen Faktoren veranschaulicht werden.

Literatur

- Dissmann, Stefan:** Handlungsorientiertes Erlernen von Programmkostruktion anhand von Rollenspielen, In: **Hubwieser, Peter (Hrsg.),** Informatische Fachkonzepte im Unterricht, Band 32 Bonn: Ges. für Informatik, 2003
- Engbring, Dieter und Pasternak, Arno:** Einige Anmerkungen zum Begriff IniK, In: **Diethelm, Ira (Hrsg.),** Didaktik der Informatik, Band 168 Bonn: GI, 2010, <URL: http://ls11-www.cs.uni-dortmund.de/people/paster/forschung/inik/inik_kurz.pdf>
- Erlhofer, Sebastian:** Suchmaschinen-Optimierung: Das umfassende Handbuch, Galileo Computing , 5. Auflage Bonn: Galileo Press, 2011
- Forneck, Hermann-Josef:** Bildung im informationstechnischen Zeitalter: Untersuchung der fachdidaktischen Entwicklung der informationstechnischen Bildung, 1. Auflage Aarau: Sauerländer, 1992
- Gesellschaft für Informatik e. V.:** Grundsätze und Standards für die Informatik in der Schule 2008, <URL: http://www.gi-ev.de/fileadmin/gliederungen/fb-iad/fa-ibs/Empfehlungen/bildungsstandards_2008.pdf>
- Gesellschaft für Informatik Fachgruppe Informatische Bildung:** Stellungnahme zu den Vorgaben für die Abiturprüfung 2007 mit zentral gestellten schriftlichen Aufgaben (im Fach Informatik) 2005, <URL: <http://www.nw.schule.de/gi/zentralabitur/StellungnahmeGI-FachgruppeInformatischeBildungNRW.pdf>> – Zugriff am 28. April 2011
- Glöggler, Michael:** Suchmaschinen im Internet: Funktionsweisen, Ranking-Methoden, Top-Positionen ; mit 25 Tabellen, Xpert.press Berlin: Springer, 2003
- Google inc.:** Technologieüberblick 2011 <URL: <http://www.google.de/intl/de/corporate/tech.html>> – Zugriff am 10. April 2011
- Humbert, Ludger:** Didaktik der Informatik: Mit praxiserprobtem Unterrichtsmaterial, Leitfäden der Informatik , 2. Auflage Wiesbaden: Teubner, 2006

Koubek, Jochen und Schulte, Carsten und Schulze, Peter und Witten,

Helmut: Informatik im Kontext (IniK): Ein integratives Unterrichtskonzept für den Informatikunterricht, In: **Peters, Ingo-Rüdiger (Hrsg.),** Informatische Bildung in Theorie und Praxis Berlin: LOG-IN-Verl, 2009

Lewandowski, Dirk: Web searching, search engines and Information

Retrieval, In: Information Services & Use, 25 Nr. 3/4 2005, (URL: <http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=19476138&site=ehost-live>), 137–147

Lobo, Sascha: Desinformation: Im Netz der Besserwisser, In: Spiegel Online, 2011, (URL: <http://www.spiegel.de/netzwelt/netzpolitik/0,1518,760463,00.html>) – Zugriff am 04. Mai 2011

Luhn, Hans Peter: The Automatic Creation of Literature Abstracts, In: IBM Journal of Research and Development, 2 Nr. 2 1958, 159–165

Martins, Bruno und Silva, Mário J.: Language identification in web pages, In: Proceedings of the 2005 ACM symposium on Applied computing, SAC '05 New York, NY, USA: ACM, 2005, (URL: <http://doi.acm.org/10.1145/1066677.1066852>), 764–768

Medienpädagogischer Forschungsverbund Südwest (Hrsg.): JIM 2010 – Jugend, Information, (Multi-)Media: Basisstudie zum Medienumgang 12- bis 19-Jähriger in Deutschland Stuttgart, 2010, (URL: <http://www.mpfs.de>) – Zugriff am 6. Juli 2011

Ministerium für Schule und Weiterbildung des Landes NRW: Vorgaben zu den unterrichtlichen Voraussetzungen für die schriftlichen Prüfungen im Abitur in der gymnasialen Oberstufe im Jahr 2011: Vorgaben für das Fach Informatik 2008, (URL: <http://www.standardsicherung.schulministerium.nrw.de/abitur-gost/getfile.php?file=1176>)

Ministerium für Schule und Weiterbildung des Landes NRW: Vorgaben zu den unterrichtlichen Voraussetzungen für die schriftlichen Prüfungen im Abitur in der gymnasialen Oberstufe im Jahr 2012: Vorgaben für das Fach Informatik 2009, (URL: <http://www.standardsicherung.schulministerium.nrw.de/abitur-gost/getfile.php?file=2033>)

Ministerium für Schule und Weiterbildung des Landes NRW: Vorgaben zu den unterrichtlichen Voraussetzungen für die schriftlichen Prüfungen im Abitur in der gymnasialen Oberstufe im Jahr 2013: Vorgaben für das Fach Informatik 2010, {URL: <http://www.standardsicherung.schulministerium.nrw.de/abitur-gost/getfile.php?file=2710>}

Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes NRW (Hrsg.): Richtlinien und Lehrpläne für die Sekundarstufe II - GymnasiumGesamtschule in Nordrhein-Westfalen: Informatik, 1. Auflage Frechen: Ritterbach Verlag, 1999

Najork, Marc und Wiener, Janet L.: Breadth-first crawling yields high-quality pages, In: Proceedings of the 10th international conference on World Wide Web, WWW '01 New York, NY, USA: ACM, 2001, {URL: <http://doi.acm.org/10.1145/371920.371965>}, 114–118

Nascimento, Mario A. und Özsu, M. Tamer und Kossmann, Donald und Miller, Renée J. und Blakeley, José A. und Schiefer, K. Bernhard (Hrsg.): (e)Proceedings of the Thirtieth International Conference on Very Large Data Bases, Toronto, Canada, August 31 - September 3 2004 Morgan Kaufmann, 2004

Nohr, Holger: Grundlagen der automatischen Indexierung: Ein Lehrbuch, 3. Auflage Berlin: Logos-Verl, 2005

Olston, Christopher: Web Crawling, In: Foundations and Trends® in Information Retrieval, 4 Nr. 3 2010, 175–246

Page, Lawrence und Brin, Sergey und Motwani, Rajeev und Winograd, Terry: The PageRank Citation Ranking: Bringing Order to the Web 1999, {URL: <http://ilpubs.stanford.edu:8090/422/>}

Pasternak, Arno und Vahrenhold, Jan: Rote Fäden und Kontextorientierung im Informatik-Unterricht, In: **Peters, Ingo-Rüdiger (Hrsg.):** Informatische Bildung in Theorie und Praxis Berlin: LOG-IN-Verl, 2009, 45–56

Schubert, Sigrid und Schwill, Andreas: Didaktik der Informatik, 2. Auflage Heidelberg: Spektrum Akad. Verl, 2011

- Schulz, Wolfgang und Held, Thorsten:** Der Index auf dem Index?
Selbstzensur bei Suchmaschinen, In: **Machill, Marcel und Beiler, Markus (Hrsg.)**, Die Macht der Suchmaschinen Köln: von Halem, 2007
- Sedgewick, Robert:** Algorithmen, 2. Auflage München: Addison-Wesley, 2002
- Sibun, Penelope und Jeffrey, Reynar C.:** Language Identification:
Examining the Issues, In: Proceedings of the 5th Symposium on Document Analysis and Information Retrieval 1996, 125–135
- Stock, Wolfgang G.:** Information-Retrieval: Informationen suchen und finden, Band 1 München: Oldenbourg, 2007
- Yahoo!:** Wie funktioniert das Ranking der Web-Sites? 2011 {URL: <http://help.yahoo.com/1/de/yahoo/search/ranking/ranking-01.html>} – Zugriff am 27. Juli 2011

A. Tabellarische Unterrichtsentwürfe

A.1. UE 1 – Aufbau von Suchmaschinen

Verlaufsplan Einführungsstunde „Informatiksystem Suchmaschine“ – ca. zwei Schulstunden

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
10	Impuls/ Einführung	Suchen in mitgebrachten Materialien und Stellen die Ergebnisse zusammen; Es kann alles genutzt werden was in der Kiste gefunden wird	Gibt Suchauftrag für die Recherche in der Kiste mit den mitgebrachten Materialien; z. B. „schriftliche Division“	Materialkiste mit (Schul-)Büchern, Werbeprospekten, Zeitungen, CDs, DVDs, Zeitschriften u. ä.
5	Einführung	Suchen bei einer Suchmaschine im WWW und ermitteln die Dauer der Recherche sowie die Qualität der Suchergebnisse		PC mit Internetverbindung, Browser
5	Einleitung		Stellt Thema der Reihe vor	Lehrvortrag
30	Erarbeitung	Erarbeiten die einzelnen Themenbereiche (Crawling, Indexing, Ranking) in Gruppenarbeit und stellen die Ergebnisse für die spätere Präsentation zusammen	Gibt Hilfestellungen	Gruppenarbeit; Plakate; verschiedenfarbige Stifte
30	Präsentation / Sicherung	Stellen Thema ihrer Gruppe während des Museumsgangs vor	Gibt Hilfestellungen falls Unklarheiten auftreten	Museumsgang; Material zum Aufhängen der Plakate
5	Sicherung	Plakate auf eigenen USB-Stick kopieren o. ä.	(zuvor fotografierte) Plakate verteilen/ verschicken	
5	Reserve			

A.2. UE 2 – Durchsuchen des WWW: Crawling

Verlaufsplan Vertiefung Crawling – ca. eine Schulstunden

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Einführung			
10	Erarbeitung	Stellen die Verknüpfungen der Beispieleite graphisch dar	Stellt Thema der Stunde „Modell des World-Wide-Web“ vor	Arbeitsblatt „Crawling“, untereinander verlinkte Beispieleiten, z.B. Schulhomepage
20	Erarbeitung	Erarbeiten in zwei Gruppen Breiten- und Tiefensuche und führen ihren Algorithmus beispielhaft mit dem zuvor erstellen Graphen durch		Gruppenarbeit, Einführungstexte Breiten- und Tiefensuche
10	Sicherung	Stellen die Ergebnisse einschließlich Vor- und Nachteile der jeweiligen Algorithmen vor	Sammelt Vor- und Nachteile an der Tafel	Unterrichtsgespräch

A.3. UE 3 – Automatische Textanalyse

Verlaufsplan Automatische Textanalyse – ca. zwei Schulstunden

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Einführung		Erläutert Thema der Stunde und stellt kurzen Text einschließlich des zugehörigen Volltextindex vor	Lehervortrag
10	Erarbeitung	Erstellen „sinnvollen“ Index für den gleichen Text und besprechen diesen mit einem Partner		Think-Pair
10	Sicherung	Erstellte Indizes und dabei angewendete Regeln werden vorgestellt	Sammelt Vorgehensweise in einer Mindmap	Share; Unterrichtsgespräch
30	Erarbeitung	Bearbeiten Texte zur automatischen Textanalyse		Think
10	Erarbeitung	Besprechen die Inhalte der Texte mit einem Partner		Pair
10	Erarbeitung (Sicherung)	SchülerInnen stellen Inhalte und offene Fragen vor	Klärt offene Fragen und fügt neue und bekannte Vorgehensweisen zusammen	Share; Unterrichtsgespräch
10	Sicherung	Erstellen weiteren Index unter Beachtung aller Regeln	Gibt bei Bedarf Helfestellungen bzw. klärt Nachfragen	Stopwortliste; Vorgaben für Indexterme
5	Reserve			

A.4. UE 4 – Binäre Suchbäume

Verlaufsplan Einführung Binärbaum – ca. zwei Schulstunden

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Einführung		Gibt Ausblick auf den Verlauf der Stunde und erläutert Problemstellung: „Elemente, zum Beispiel Stoppwörter, müssen gespeichert werden.“	Lehrvortrag
5	Vorbereitung	Vorbereitung des Raumes und Einteilung in Expertengruppen für ein Gruppenpuzzle	Leitet Gruppenbildung an (bei Bedarf)	
25	Erarbeitung	Erarbeiten in Expertengruppen die Eigenschaften der Datenstrukturen und Algorithmen: Binärer Suchbaum, Lineare Liste und sortiertes Array mit binärer Suche;	Klärt Fragen und gibt Hilfestellungen	Gruppenarbeit; Arbeitsblätter für jede Gruppe; Präsentationsmaterial, Fäden als Kanten zwischen den Knoten, Karteikarten, starke Pappe, Pinwandnadeln, Material für kleine Objektspiele;
30	Erarbeitung	Stellen die Eigenschaften in den Stammguppen vor	Stellt eine korrekte Präsentation in den Gruppen sicher	Gruppenarbeit; s.o.
10	Erarbeitung	Stellen Vor- und Nachteile der einzelnen Möglichkeiten im Bezug auf die Stopwortliste gegenüber und kommen zu einer Empfehlung		Gruppenarbeit; s.o.
5	Sicherung	Präsentieren Ergebnisse der Beratung	Sammelt dabei Vor- und Nachteile an der Tafel; stellt neue Fragestellung: „Welche Möglichkeit eignet sich für die Verwaltung des Index?“	Unterrichtsgespräch

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Erarbeitung	Beraten kurz über die neue Fragestellung		Gruppenarbeit; s.o.
5	Sicherung	Präsentieren die Ergebnisse der letzten Beratung	Moderiert die Präsentation und stellt Hausaufgabe: Stichwortliste zur Umsetzung der Operationen Löschen und Hinzufügen im binären Suchbaum.	Unterrichtsgespräch

A.5. UE 5 – Binäre Suchbäume: Standardoperationen

Verlaufsplan Standardoperationen Binärbaum – ca. eine Schulstunde

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Einführung		Gibt Rückblick auf letzte Stunde und stellt Ziel der heutigen Stunde vor: Knoten finden, Einfügen und löschen; Baum der Reihe nach durchlaufen → In-Order-Traversierung	Lehrvortrag
15	Erarbeitung	Machen sich mit den genannten Operationen, im Hinblick auf Eingabe, Ausgabe und Voraussetzungen, vertraut		Arbeitsblatt mit Informationen über die Operationen
15	Erarbeitung	Setzen mind. zwei der Operationen in Struktogrammen um	Klärt Fragen	Evtl. Werkzeug zum Erstellen der Struktogramme
5	Sicherung	Tauschen Struktogramme mit Partnern und überprüfen die erhaltenen Struktogramme	Hilft bei der Durchführung des Tauschs, falls notwendig	
5	Sicherung		Stellt Hausaufgabe: Standardoperationen in Pseudocode implementieren	Lehrvortrag

A.6. UE 6 – Standardoperationen implementieren

Verlaufsplan Standardoperationen für Binäräbäume implementieren – ca. zwei Schulstunden

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
15	Einführung	Stellen Lösungen in Pseudocode vor (HA letzter Stunde)	Gibt Rückblick auf letzte Stunde und stellt Ziel der heutigen Stunde vor: Knoten finden, Einfügen und löschen; Baum der Reihe nach durchlaufen → In-Order-Traversierung	Unterrichtsgespräch
5		Verteilung in Zweiergruppen und Zuweisung der Operationen	Leitet Gruppenbildung an (bei Bedarf)	Arbeitsblätter für die einzelnen Operationen
25	Erarbeitung	Entwickeln bzw. ergänzen Klassendiagramme für die Klassen BinarySearchTree, Item und Node		Evtl. Vorlagen auf Basis der Materialien für das Zentralabitur
10	Sicherung	Stellen die erstellten Klassendiagramme vor	Erstellt ein gemeinsames Diagramm als Vorlage für die Implementierung	Tafel/Beamer
30	Erarbeitung	SchülerInnen beginnen die Umsetzung der Operationen in Java (Bluel)	Unterstützt bei Problemen	Einzel- oder Partnerarbeit; Vorlage mit entfernten Methoden; Klassendokumentation „BinaryTree“
5	Sicherung	Schwierigkeiten bei der Umsetzung werden besprochen	Gibt Hinweise zum Umgang mit den Schwierigkeiten	

A.7. UE 7 – Implementierung testen

Verlaufsplan Testen der Implementierung – ca. eine Schulstunde

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
25	Erarbeitung / Sicherung Ergebnisse letzter Stunde	Testen die eigenen Lösungen aus der letzten Stunde und nehmen, falls notwendig Korrekturen vor; Unterstützen sich gegenseitig	Gibt Hilfestellungen bei Problemen mit der Testumgebung und der Implementierung	Testumgebung mit Bluel; Arbeitsblatt „Binärbaum testen“; Einzel- und Partnerarbeit
10	Erarbeitung	Flügen geordnete Elemente der Reihe nach in einen leeren Binärbaum ein und Protokollieren Beobachtungen		Testumgebung mit Bluel; Arbeitsblatt „Binärbaum testen“, Einzelarbeit
10	Sicherung	Stellen Beobachtungen des Tests vor	Sammelt und strukturiert die Beobachtungen; weist in kurzem Lehrvortrag auf weitere Baumtypen, mit „automatischer“ Balancierung hin	Unterrichtsgespräch/Lehrvortrag

A.8. UE 8 – Sortieralgorithmen

Verlaufsplan Sortieralgorithmen Übersicht – ca. zwei Schultunden

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Einführung		Erläutert Thema der Stunde „Elemente der Listen und Felder sind nicht automatisch sortiert.“	Lehrvortrag
10	Impuls	Beobachten Vorgänge im Video und versuchen die Vorschriften zu erkennen		Video AlgoRhythms „BubbleSort“ ¹
10	Erarbeitung	Stellen Beobachtungen vor	Notiert Beobachtungen an der Tafel	Unterrichtsgespräch; Tafel o. ä.
35	Erarbeitung	Recherchieren in drei Arbeitsgruppen Material zu BubbleSort, InsertionSort, Quicksort und bereiten Plakate vor	Gibt Helfestellungen	arbeitsteilige Gruppenarbeit; Material zur Unterstützung der Recherche; Präsentationsmaterial
25	Sicherung	Museumsgang: SuS gehen in Gruppen zu den einzelnen Plakaten/Präsentationen und erläutern den Ablauf der jeweiligen Algorithmen und ihre Vor- und Nachteile	Museumsgang	
5	Überleitung		Abschluss der Reihe oder Überleitung zum Thema Ranking bzw. Informationsqualität	Lehrvortrag

¹ <http://youtu.be/lYZQPjUT5B4>

A.9. UE 9 – Ranking von Suchergebnissen

Verlaufsplan Methoden zum Ranking von Suchergebnissen – ca. eine Schulstunde

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Impuls + Einführung		Zeigt drei Texte die den Begriff „Tornado“ enthalten und bittet SuS diese danach zu bewerten welcher Text am besten auf die Suchanfrage „Tornado“ passt	Unterrichtsgespräch
10	Erarbeitung	Bearbeiten Text zum Thema „Ranking“		Think-Pair-Share: Think
10	Erarbeitung	Tauschen sich mit einem Partner über die Möglichkeiten des Ranking aus		Think-Pair-Share: Pair
10	Erarbeitung / Sicherung	Bereiten in zwei bis drei Gruppen eine Präsentation der Inhalte vor.	Steht für Nachfragen zur Verfügung	Think-Pair-Share: Share; Präsentationsmaterial oder entsprechende Software
10	Sicherung	Eine Gruppe präsentiert die Ergebnisse, die anderen Gruppen ergänzen		

A.10. UE 10 – Informationsqualität im Internet

Verlaufsplan Exkurs Qualität von Internetquellen – ca. zwei Schulstunden

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Einführung		Erläutert Thema der Stunde, motiviert durch die Frage, ob ein passendes Dokument (Rückbezug zum Ranking) auch ein qualitativ hochwertiges Dokument ist	Lehrvortrag
15	Erarbeitung	Schaffen sich einen ersten <u>Überblick</u> über das Thema „Informationsqualität im Internet“		Einzelarbeit; Texte: „Informationsqualität im Internet“, „Lügen im Netz“, Weitere Recherche im Internet ist explizit erlaubt
20	Erarbeitung	Vertiefen die Informationen mit weiteren Quellen		Think-Pair-Share; Pair; weitere Texte: „Im Netz der Besserwissen“, „Wie die Print-Lobby Kinder indoctriniert“, „Kritik an Wikipedia“
15	Erarbeitung	Bereiten in vier Gruppen die Rollen laut zugeloser Rollenkarte vor	Lost Rollenkarten zu	Podiumsdiskussion (Rollenspiel); Rollenkarten
25	Sicherung	zufällig ausgewählte SchülerInnen stellen die Teilnehmer der Diskussion; Zuschauer stellen bei Bedarf Zwischenfragen	Moderiert die Diskussion	Podiumsdiskussion (Rollenspiel)
10	Sicherung	Teilnehmer der Diskussion nehmen Abstand zu ihren Rollen ein und reflektieren die Diskussion	Entlässt Teilnehmer aus ihren Rollen und bittet um eine abschließende Reflektion der Diskussion	Unterrichtsgespräch

Informatik im Kontext Suchmaschinen – Exkurs Informationsqualität

A.11. Überarbeitet: UE 2 – Durchsuchen des WWW: Crawling

Verlaufsplan Vertiefung Crawling (erweitert) – ca. eine Schulstunden

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Einführung		Stellt Thema der Stunde „Modell des World-Wide-Web“ vor; erteilt den Auftrag den Titel neuer (möglichst nicht älter als 60 Minuten) und zufällig Wikipedia-Einträge in verschiedenen Suchmaschinen zu suchen.	Lehrvortrag
5	Erarbeitung	führen Suche durch und erstellen Notizen		Einzel- oder Partnerarbeit
5	Sicherung + Überleitung	stellen Ergebnisse der Suche vor	moderiert das Unterrichtsgespräch; stellt Beispieleite vor und erteilt Arbeitsauftrag	Unterrichtsgespräch
5	Erarbeitung	Stellen die Verknüpfungen der Beispieleite graphisch dar		Arbeitsblatt „Crawling“, untereinander verlinkte Beispieleseiten, z. B. Schulhomepage
15	Erarbeitung	Erarbeiten in zwei Gruppen Breiten- und Tiefensuche und führen ihren Algorithmus beispielhaft mit dem zuvor erstellten Graphen durch		Gruppenarbeit, Einführungstexte Breiten- und Tiefensuche
10	Sicherung	Stellen die Ergebnisse einschließlich Vor- und Nachteile der jeweiligen Algorithmen vor	Sammelt Vor- und Nachteile an der Tafel	Unterrichtsgespräch

A.12. Überarbeitet: UE 3 – Automatische Textanalyse

Verlaufsplan Automatische Textanalyse – ca. zwei Schulstunden

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
10	Impuls	Suchen die fünf aussagekräftigsten Wörter aus einem kurzen Text	Erteilt den Arbeitsauftrag und stellt kurzen Text zur Verfügung	Lehrervortrag, dann Einzelarbeit
5	Erarbeitung	Prüfen die Ähnlichkeit der erstellten Listen	Fragt die Anzahl der Unterschiede innerhalb der Paare ab; Stellt dann die Leitfrage und das Thema der Stunde vor: „Kann ein Computer das auch so gut und einfach?“	Partnerarbeit dann Unterrichtsgespräch
10	Erarbeitung	Sammeln und formalisieren Regeln um gute Indexterme zu finden	Stellt Volltextindex als einfach zu erstellen aber wenig aussagekräftigen Index vor; erteilt Arbeitsauftrag	Lehrervortrag, dann Think-Pair
5	Sicherung	Erarbeitete Regeln werden vorgestellt	Sammelt Vorgehensweise in einer Mindmap	Share; (Unterrichtsgespräch)
30	Erarbeitung	Bearbeiten Texte zur automatischen Textanalyse		Think
10	Erarbeitung	Besprechen die Inhalte der Texte mit einem Partner		Pair
10	Erarbeitung (/Sicherung)	SchülerInnen stellen Inhalte und offene Fragen vor	Klärt offene Fragen und fügt neue und bekannte Vorgehensweisen zusammen	Share; Unterrichtsgespräch
10	Sicherung	Erstellen weiteren Index unter Beachtung aller Regeln	Gibt bei Bedarf Hilfestellungen bzw. klärt Nachfragen	Stoppwortliste; Vorgaben für Indexterme

A.13. Umweltschutz durch effiziente Algorithmen

Verlaufsplan ökologische und ökonomische Vorteile effizienter Programmierung – ca. eine Schulstunde

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
10	Einstieg	Bearbeiten den Text	Verteilt Texte und erteilt Arbeitsauftrag	Lehrvortrag, dann Einzelarbeit; Text „Einmal googeln entspricht einer Stunde Licht“
10	Erarbeitung	Diskutieren den Zusammenhang zwischen eingesetzten Algorithmen und Datenstrukturen und den entstehenden Kosten bzw. Umwelteinflüssen	Moderiert das Gespräch	Unterrichtsgespräch
15	Erarbeitung	Sammeln Argumente um einem hypothetischen jüngeren Schüler die Notwendigkeit effizienter Programmierung zu erläutern		Partnerarbeit
5	Sicherung	Ein Schüler erläutert seine Argumentation, die anderen notieren Ergänzungen zu den eigenen Aufzeichnungen	Stellt den interessierten jüngeren Schüler dar	Rollenspiel
10	Sicherung	Ergänzen weitere Argumente	Moderiert das Gespräch und sammelt Argumentation an der Tafel	Unterrichtsgespräch

A.14. Ranking 2 – Wer bestimmt was ich finden kann?

Verlaufsplan Ranking 2 – Wer bestimmt was ich finden kann? – ca. eine Schulstunde

Zeit	Phase	Aktivität SuS	Aktivität L	Sozialform/ Methode/ Material
5	Impuls	Führen Suchaufträge aus	Erteilt Arbeitsauftrag vorgegebene Suchanfragen mit der angegebenen Suchmaschine durchzuführen und die Ergebnisliste kritisch zu betrachten	Lehrvortrag, dann Einzelarbeit; Liste mit funktionierenden „Google-Bomber“
10	Erarbeitung	Diskutieren die Gründe für die teilweise merkwürdigen Ergebnisse	Moderiert das Gespräch	Unterrichtsgespräch
20	Erarbeitung	Recherchieren in zwei Gruppen (politische Motive & wirtschaftliche Motive) verschiedene Möglichkeiten und Gründe Suchergebnisse zu manipulieren	Teilt Gruppen ein; gibt bei Bedarf Hilfestellungen	Gruppenarbeit
10	Sicherung	Stellen Ergebnisse vor	Sammelt und strukturiert die Ergebnisse	Unterrichtsgespräch

Informatik im Kontext Suchmaschinen – Ranking 2 – Wer bestimmt was ich finden kann?

B. Material

B.1. UE 1 – Aufbau von Suchmaschinen

Informatiksystem Suchmaschine – Dokumente finden

Hinter einer Suchmaschine steckt mehr als ein Eingabefeld und ein Knopf zum Abschicken der Anfrage. Suchmaschinen sind komplexe Informatiksystem in denen viele Teile ineinander greifen müssen, damit sowohl die hohe Geschwindigkeit, als auch die hohe Präzision der Suchergebnisse gewährleistet wird.

- 5 Eine Suchmaschine kann grob in drei Abschnitte aufgeteilt werden. Zuerst müssen Dokumente im World-Wide-Web gefunden werden. Diese Aufgabe übernimmt ein Crawler (auch Spider oder (Web-)Robot genannt). Im nächsten Schritt werden die Seiten verarbeitet und ein sogenannter Inverser Index wird erstellt. Solch ein Index ist mit dem Index in einem Lexikon oder einem Fachbuch vergleichbar. Zuletzt müssen die Suchanfragen, welche über das sichtbare Eingabefeld eingetragen werden, verarbeitet werden und die Ergebnisse müssen in eine Reihenfolge gebracht werden, so dass die besten Ergebnisse möglichst weit oben erscheinen.
- 10

15 Im folgenden Text wird der Prozess des Crawling näher beschrieben.

15 **Crawling**

Da es keine zentrale Stelle gibt in der alle Webseiten registriert werden kann ist es erforderlich, dass Suchmaschinen neue Webseiten selbstständig finden können. Diese Aufgabe übernimmt ein Teil der Suchmaschine der Crawler genannte wird.

- Ein Crawler benutzt die gleichen Links wie jemand der einfach durch das WWW surft.
- 20 Das Programm folgt den Links auf einer Seite. Ausgehend von einer Startseite sucht der Crawler alle Links auf einer Seite und speichert diese in einer Liste ab. Dann kann er diese Links der Reihe nach besuchen um die Dokumente, die sich hinter den Links verbergen herunter zu laden, zu speichern und wiederum auf neue Links zu untersuchen. Auf diese Weise arbeitet sich der Crawler durch die Dokumente im WWW und „krabbelt“ von einer
- 25 Seite zur nächste.

- Es gibt allerdings nicht nur einfach zu analysierende HTML-Dokumente im WWW sondern auch Bilder, Videos und anderen Medien, die sich nicht so einfach durchsuchen lassen. Die Betreiber von Suchmaschinen versuchen ständig ihre Techniken diese Dokumente zu durchsuchen zu verbessern.
- 30 Andere Dokumente können von einem Crawler garnicht erst erreicht werden. Manche liegen möglicherweise in einem geschützten Bereich des Webservers. Andere Dokumente werden dynamisch vom Server erstellt und der Crawler kann die notwendigen Daten nicht an den Server senden um ein relevantes Dokument zu erzeugen. Diese Dokumente werden auch als Teile des Deep Webs bezeichnet, da sie für eine einfache Suchmaschine zu tief verborgen liegen.
- 35

Natürlich gibt es auch Seiten, die von einer Suchmaschine nicht in den Index eingetragen werden sollen. Eine Möglichkeit den Crawler zu stoppen ist die Benutzung einer Textdatei mit dem Namen „robots.txt“ welche im Hauptverzeichnis der Webseite abgelegt wird.

40 Diese Technik funktioniert natürlich nur dann, wenn die Software der Suchmaschine sich an die dort aufgeführten Regeln hält.

Leider gibt es auch andere, die sich die Technik des Crawling zu Nutze machen. Programme, die aus einer Webseite nur die Links und E-Mail-Adressen filtern werden benutzt um große Mengen an E-Mail-Adressen zu erhalten an die dann Werbung, häufig Spam genannt, geschickt werden kann.

Quelle:

Erlhofer, Sebastian (2011): Suchmaschinen-Optimierung. Das umfassende Handbuch. 5., aktualisierte und erw. Bonn: Galileo Press (Galileo Computing)

Glöggler, Michael (2003): Suchmaschinen im Internet. Funktionsweisen, Ranking-Methoden, Top-Positionen. Berlin: Springer (Xpert.press)

Greifeneder, Horst (Hg.) (2010): Erfolgreiches Suchmaschinen-Marketing. Wie Sie bei Google, Yahoo, MSN & Co. ganz nach oben kommen. 2. Auflage. Wiesbaden: Gabler Verlag / GWV Fachverlage GmbH Wiesbaden

Informatiksystem Suchmaschine – Gefundene Dokumente verarbeiten

Hinter einer Suchmaschine steckt mehr als ein Eingabefeld und ein Knopf zum Abschicken der Anfrage. Suchmaschinen sind komplexe Informatiksystem in denen viele Teile ineinander greifen müssen, damit sowohl die hohe Geschwindigkeit, als auch die hohe Präzision der Suchergebnisse gewährleistet wird.

Eine Suchmaschine kann grob in drei Abschnitte aufgeteilt werden. Zuerst müssen Dokumente im World-Wide-Web gefunden werden. Diese Aufgabe übernimmt ein Crawler (auch Spider oder (Web-)Robot genannt). Im nächsten Schritt werden die Seiten verarbeitet und ein sogenannter Inverser Index wird erstellt. Solch ein Index ist mit dem Index in einem Lexikon oder einem Fachbuch vergleichbar. Zuletzt müssen die Suchanfragen, welche über das sichtbare Eingabefeld eingetragen werden, verarbeitet werden und die Ergebnisse müssen in eine Reihenfolge gebracht werden, so dass die besten Ergebnisse möglichst weit oben erscheinen.

15 Im folgenden Text wird der Prozess der Erstellung eines Index vorgestellt.

Indexerstellung

Die Anzahl der Dokumente im WWW ist enorm. Diese Seiten bei jedem eingehenden Suchauftrag komplett zu durchsuchen ist unmöglich. Es würde schlicht zu lange dauern. Es muss also eine Möglichkeit geben große Teile des Aufwands einer Suchmaschine nur einmal entstehen zu lassen.

Viele wissenschaftliche Bücher und auch Lexika enthalten auf den hinteren Seiten einen Index. Dieser Index ist eine Liste von Stichworten. Wenn man Informationen zu einem der Stichwörter sucht, muss man nur das Stichwort im Index suchen und kann dort eine Seitenzahl oder mehrere Seitenzahlen ablesen. Auf den zugehörigen Seiten sind dann Informationen bezüglich des Stichworts zu finden. Wenn man solch einen Index nicht nur für ein Dokument erstellt und auf die Seitenzahlen verweist, sondern einen Index für viele Dokumente erstellt und dort auf das einzelne Dokument verweist spricht man von einem invertierten Index.

Da ein invertierter Index immer noch enorm groß ist, muss sichergestellt werden, dass die Suche nach den Stichworten in sehr kurzer Zeit erfolgen kann. Es gibt verschiedene Möglichkeiten wie man dafür sorgen kann, dass auch auf großen Datenmengen eine effiziente Suche stattfinden kann. Sehr häufig werden Baumstrukturen eingesetzt. Diese bieten sowohl die Möglichkeit einzelne Daten schnell wieder zu finden, als auch flexibel erweitert zu werden.

35 Aus dem oben genannten Vergleich mit Büchern ergibt sich sofort, dass ein Index natürlich nicht aus allen Wörtern eines Textes entstehen kann. Es ist also auch wichtig das nur die Wörter für den Index ausgewählt werden, die eine hohe inhaltliche Bedeutung für das Dokument haben. Eine weitere Schwierigkeit, besonders in der deutschen Sprache, ist die Tatsache, dass es viele verschiedene Formen des gleichen Wortes geben kann. Die Wörter 40 sind dann zwar unterschiedlich, aber die inhaltliche Bedeutung bleibt identisch. Bei der

Erstellung des Index muss also sichergestellt werden, dass die Wörter in den Index aufgenommen werden, nach denen hinterher auch gesucht wird und nicht eine andere Form, die ein wiederfinden verhindert.

Quelle:

Erlhofer, Sebastian (2011): Suchmaschinen-Optimierung. Das umfassende Handbuch. 5., aktualisierte und erw. Bonn: Galileo Press (Galileo Computing)

Glöggler, Michael (2003): Suchmaschinen im Internet. Funktionsweisen, Ranking-Methoden, Top-Positionen. Berlin: Springer (Xpert.press)

Greifeneder, Horst (Hg.) (2010): Erfolgreiches Suchmaschinen-Marketing. Wie Sie bei Google, Yahoo, MSN & Co. ganz nach oben kommen. 2. Auflage. Wiesbaden: Gabler Verlag / GWV Fachverlage GmbH Wiesbaden

Informatiksystem Suchmaschine – Gefundene Dokumente verarbeiten

Hinter einer Suchmaschine steckt mehr als ein Eingabefeld und ein Knopf zum Abschicken der Anfrage. Suchmaschinen sind komplexe Informatiksystem in denen viele Teile in-

- 5 einandergreifen müssen, damit sowohl die hohe Geschwindigkeit, als auch die hohe Präzision der Suchergebnisse gewährleistet wird.

Eine Suchmaschine kann grob in drei Abschnitte aufgeteilt werden. Zuerst müssen Dokumente im World-Wide-Web gefunden werden. Diese Aufgabe übernimmt ein Crawler (auch Spider oder (Web-)Robot genannt). Im nächsten Schritt werden die Seiten verarbeitet und ein sogenannter Inverser Index wird erstellt. Solch ein Index ist mit dem Index in

- 10 einem Lexikon oder einem Fachbuch vergleichbar. Zuletzt müssen die Suchanfragen, welche über das sichtbare Eingabefeld eingetragen werden, verarbeitet werden und die Ergebnisse müssen in eine Reihenfolge gebracht werden, so dass die besten Ergebnisse möglichst weit oben erscheinen.

- 15 Im folgenden Text wird die Verarbeitung der Suchanfragen vorgestellt.

Verarbeitung der Suchanfragen

Für ein gutes Suchergebnis ist es wichtig, dass möglichst viele Dokumente, die Informationen zu dem gesuchten Thema enthalten ausgegeben werden. Die Zahl der Dokumente, die

- 20 Informationen zum Thema enthalten, aber nicht gefunden werden soll möglichst gering sein. Genauso sollen möglichst nur Dokumente im Suchergebnis zu finden sein, die wirklich Informationen zum gesuchten Thema enthalten, es sollen also keine unnötigen bzw. unpassenden Dokumente ausgegeben werden.

Die Suchanfrage die eine Benutzerin in das Suchfeld einträgt, muss genau wie ein Dokument, welches vom Crawler gefunden wird, zuerst verarbeitet werden. Die Verarbeitung

- 25 der Suchanfrage verläuft bei verschiedenen Suchmaschinen unterschiedlich. Typische Schritte sind folgende: Verschiedene Formen eines inhaltlich gleichen Wortes werden auf eine Form zurückgeführt. Aus „den Häusern“ würde zum Beispiel „Haus“. Wenig oder garnicht aussagekräftige Begriffe werden aussortiert. Bestimmte und unbestimmte Artikel kommen in (fast) jedem Text sehr häufig vor und eignen sich demnach nicht als Suchwort.

- 30 Es ist auch sinnvoll die Suchanfrage einer Rechtschreibkorrektur zu unterziehen, damit nicht wegen eines Tippfehlers keine oder zu wenige Ergebnisse angezeigt werden. Häufig werden auch Vorschläge zu anderen Schreibungen oder anderen Begriffen ausgegeben. Auch andere Begriffe mit der gleichen oder einer ähnlichen Bedeutung werden der Benutzerin häufig vorgeschlagen, vor allem wenn die Ergebnismenge erstaunlich klein ist und

- 35 für die anderen Begriffe größer wäre.

Wenn eine Reihe von Dokumenten gefunden wurde, ist die nächste, sehr wichtige Aufgabe, diese Dokumente in eine Reihenfolge zu bringen. Das Ziel ist es, das passendste Dokument zuerst anzuzeigen. Für diese Sortierung werden viele verschiedene Verfahren eingesetzt. Einige beruhen auf den Eigenschaften der einzelnen Dokumente, zum Beispiel auf

- 40 der Anzahl und der Lage der Suchwörter im Dokument. Ein Kriterium kann auch die

Sprache des Dokuments und der Benutzerin sein. Dokumente in der Sprache der Benutzerin werden häufig besser bewertet als Dokumente in einer Fremdsprache, auch wenn die Suchbegriffe in einer anderen Sprache eingegeben wurden.

Andere Verfahren untersuchen, die Links die auf ein Dokument verweisen und ermitteln so ob das Dokument von anderen Webautoren für wichtig erachtet wird.⁴⁵

Auch die Auswahl der Ergebnisse bei vorangegangen Suchen wird teilweise untersucht. Falls es möglich ist wird, zum Beispiel von Google, versucht die Ergebnisse auf den Standort der Benutzerin abzustimmen. Dafür werden viele Daten über die Benutzer einer Suchmaschine gesammelt. Daher sehen Datenschützer diese Techniken sehr kritisch.

Quelle:

Erlhofer, Sebastian (2011): Suchmaschinen-Optimierung. Das umfassende Handbuch. 5., aktualisierte und erw. Bonn: Galileo Press (Galileo Computing)

Glöggler, Michael (2003): Suchmaschinen im Internet. Funktionsweisen, Ranking-Methoden, Top-Positionen. Berlin: Springer (Xpert.press)

Greifeneder, Horst (Hg.) (2010): Erfolgreiches Suchmaschinen-Marketing. Wie Sie bei Google, Yahoo, MSN & Co. ganz nach oben kommen. 2. Auflage. Wiesbaden: Gabler Verlag / GWV Fachverlage GmbH Wiesbaden

B.2. UE 9 – Ranking von Suchergebnissen

Relevanz von Suchergebnissen

Aufgrund der vielen verschiedenen Möglichkeiten, Dokumente im Internet zu veröffentlichen, ist es unmöglich, genau passende Dokumente von solchen zu unterscheiden, die vielleicht nur am Rande mit den gesuchten Begriffen zu tun haben. Daher ist eine wichtige

- 5 Aufgabe einer Suchmaschine, die Ergebnisse zu sortieren, so dass die passendsten Ergebnisse zu erst angezeigt werden. Diese Aufgabe ist für einen Menschen vergleichsweise einfach zu erledigen, die automatische Sortierung ist allerdings nicht trivial.

In der Praxis werden viele verschiedene Verfahren verwendet, um die Relevanz eines Dokuments zu ermitteln. Die Gewichtung und das genaue Vorgehen innerhalb der Verfahren

- 10 ist letztendlich das entscheidende Qualitätsmerkmal einer guten Suchmaschine. Daher werden die Details nicht gerne veröffentlicht.

Im folgenden Text werden einige Verfahren und Ansätze beschrieben, mit denen die Relevanz von Dokumenten automatisch bestimmt wird.

Boolesches Retrieval

- 15 Ein erster Ansatz ist die Kombination mehrerer Begriffe mit Hilfe der booleschen Algebra. Die Begriffe werden mit einem der drei Operatoren AND, OR, NOT verknüpft. Die Wirkung der Operatoren leitet sich aus der Bedeutung dieser englischen Begriffe ab. Eine Suchanfrage nach „(Tornado OR Sturm) NOT Film“ würde nur Dokumente als Ergebnis zulassen, in dem der Begriff Tornado oder der Begriff Sturm vorkommt, solange der Begriff Film nicht vorkommt. Im Allgemeinen ist die Ergebnismenge aber immer noch sehr groß¹. Eine Sortierung lässt sich mit dieser Methode allein nicht erreichen.
- 20

Vektorraummodell

- Eine weit verbreitete Möglichkeit zur Ermittlung der Ähnlichkeit zwischen zwei Texten oder einem Text und einer Suchanfrage ist das Vektorraummodell. Dazu werden sowohl 25 die Suchanfrage als auch die Dokumente in Vektoren übersetzt. Je höher das Skalarprodukt eines Dokumentenvektors mit einem Suchanfragevektor ist, desto relevanter ist das Dokument. Zu Erzeugung der Vektoren werden die Indexterme benutzt.

Beispiel: Vektorraummodell

- Nehmen wir einmal an, unsere Dokumentensammlung umfasst nur zwei Dokumente mit 30 jeweils einem Satz: „Nicht für die Schule, sondern für das Leben lernen wir!“ und „Schule“ ist ein deutscher Film aus dem Jahr 2000.“ Unser Index besteht aus folgenden Wörtern: 2000, deutsch, Film, Jahr, Leben, lernen, Schule und sondern; Daraus ergeben sich für die beiden Sätze die Vektoren $\alpha = (0,0,0,0,1,1,1,1)$ und $\beta = (1,1,1,1,0,0,1,0)$. In längeren Texten würden natürlich einige Indexterme mehr als einmal vorkommen, so dass auch 35 größere Zahlen im Vektor vorkommen würden. Jemand der sich für den Film interessiert, sucht möglicherweise mit der Suchanfrage „Schule Film“ aus der der Vektor

1 Am 26. Mai 2011 findet z. B. Google 120 Millionen Ergebnisse zur Anfrage „Tornado OR Sturm -Film“. Der Operator NOT wird hier durch ein Minus ersetzt.

$\hat{v}=(0,0,1,0,0,0,1,0)$ entsteht. Da gilt $\hat{v} \ll \hat{u}$ ist der Vektor \hat{v} dem Vektor der Suchanfrage ähnlicher und damit ist der zweite Satz passender als der erste Satz.

- Bei dieser Methode werden allerdings Dokumente mit vielen Worten stärker berücksichtigt. 40 Die Länge eines Textes schlägt sich im Betrag, also der Länge, des Vektors nieder. Daher berechnet man mit dem sogenannten Cosinus-Maß den Cosinus des Winkels mit dem die beiden Vektoren zueinander stehen. Dafür wird das Skalarprodukt durch das Produkt der Beträge beider Vektoren geteilt.

Termfrequenz und Inverse Dokumentenhäufigkeit

- 45 Die Häufigkeit, mit der ein Wort in einem Dokument vorkommt, gibt eine belastbare Auskunft über die inhaltliche Bedeutung des Wortes. Mit anderen Worten: Wenn in einem Text ein bestimmtes Wort sehr oft vorkommt, so geht es in dem Text vermutlich um etwas was mit diesem Begriff zu tun hat. Durch diese Erkenntnis kann man die Bedeutung eines Wortes für ein Dokument gewichten. Damit diese Gewichtung von der Länge des Textes 50 unabhängig wird, wird die Termfrequenz ermittelt, bei der durch die Anzahl der Wörter im Dokument geteilt wird. Die Termfrequenz TF berechnet sich wie folgt:
$$TF = \frac{\text{(Häufigkeit eines Terms im Dokument)}}{\text{(Anzahl der Wörter im Dokument)}}$$
 Eine hohe Termfrequenz spricht für eine hohe Bedeutung des Wortes.

- 55 Wenn allerdings ein Wort in jedem Dokument häufig vorkommt, so ist die inhaltliche Bedeutung gering. Andersherum beschreibt ein Wort welches nur in wenigen Dokumenten vorkommt den Inhalt des Dokumentes sehr gut. Aus diesem Grund wird die Inverse Dokumentenhäufigkeit (IDF) berechnet. Die zugehörige Formel lautet:
$$IDF = \log_2 \frac{\text{(Gesamtzahl der Dokumente in der Datenbank)}}{\text{(Anzahl der Dokumente, in denen der Term vorkommt)}}$$
 Diese beiden Werte werden häufig als Produkt in Kombination für die Gewichtung herangezogen.

60 **Lage und Auszeichnung**

- Bisher sind wir davon ausgegangen, dass es egal ist an welcher Stelle ein Wort in einem Text zu finden ist. Ein Wort in einer Überschrift oder auch ein besonders hervorgehobenes Wort hat aber im Allgemeinen eine höhere Bedeutung als ein Wort welches an beliebige Stelle im Text steht. Viele Texte geben in den ersten Zeilen einen kurzen Überblick über 65 den gesamten Inhalt. Die dortigen Terme sollten also stärker gewichtet werden. Eine Möglichkeit diesen Faktoren gerecht zu werden ist es die Berechnung von IF und IDF für diese Bereiche oder speziell ausgezeichnete Worte einzeln durchzuführen und beim Zusammenführen entsprechend zu gewichten.

PageRank

- 70 Die bisherigen Methoden berücksichtigen nicht, dass die Dokumente im World-Wide-Web miteinander vernetzt sind. Der Name PageRank geht auf den Autor des Algorithmus zurück welcher zugleich auch Mitgründer der Suchmaschine Google ist.

Die Idee von PageRank basiert darauf, dass davon ausgegangen wird, dass ein Dokument nur dann von einem anderen Dokument verlinkt wird, wenn es vom jeweiligen Autor für relevant befunden wird. Ein Link ist also als Empfehlung eine Seite zu besuchen zu verstehen. Eine Seite mit vielen Empfehlungen, also vielen eingehenden Links ist vermutlich besser als eine Seite mit wenigen eingehenden Links.

Eine Seite gibt den eigenen PageRank-Wert an die Seiten weiter, auf die ein Link gesetzt ist. Damit nicht der komplette Wert weitergegeben wird gibt es einen sogenannten Dämpfungs faktor. Der PageRank einer Seite berechnet sich aus einem Startwert ($1 - \text{Dämpfungs faktor}$) und der Summe der Anteile des PageRank-Wertes der Seiten die auf die Seite verweisen.

Hilltop und TrustRank

Eine Erweiterung dieses Verfahrens sind das Hilltop-Prinzip und TrustRank. Bei diesen Verfahren werden bestimmten Webseiten höhere Werte zugewiesen, weil sie Experten auf ihrem Gebiet oder besonders Vertrauenswürdig sind, wie z. B. Internetauftritte von Universitäten. Da davon ausgegangen wird, dass solche Seiten nicht auf schlechte Seiten verlinken, wird durch diese Seiten ein höherer Wert weitergegeben.

ClickPopularity

Eine andere Methode, die die Meinung der Surfer berücksichtigt, die Relevanz von Webseiten zu bestimmen ist es die Anzahl mit der ein Suchergebnis aus einer Reihe von Ergebnissen nach einer Suche ausgewählt wird. Es wird also gezählt wie häufig die Seite, nach Eingabe eines bestimmten Suchwortes, angeklickt wird.

Quelle:

Erlhofer, Sebastian (2011): Suchmaschinen-Optimierung. Das umfassende Handbuch. 5., aktualisierte und

erw. Bonn: Galileo Press (Galileo Computing)

Glöggler, Michael (2003): Suchmaschinen im Internet. Funktionsweisen, Ranking-Methoden, Top-Positionen. Berlin: Springer (Xpert.press)

B.3. UE 10 – Informationsqualität im Internet

B.3.1. Texte zum Thema Informationsqualität

Literatur

Europäisches Zentrum für Medienkompetenz: Informationsqualität im Internet 2009

Lobo, Sascha: Desinformation: Im Netz der Besserwisser, In: Spiegel Online, 2011, ⟨URL: <http://www.spiegel.de/netzwelt/netzpolitik/0,1518,760463,00.html>⟩ – Zugriff am 04. Mai 2011

Niggemeier, Stefan: Wie die Print-Lobby Kinder indoktriniert 2011 ⟨URL: <http://www.stefan-niggemeier.de/blog/wie-die-print-lobby-kinder-indoktriniert/>⟩ – Zugriff am 7. Juli 2011

Ude, Albrecht: Lügen im Netz: Kann man Informationen aus dem Internet trauen? Acht Tipps, wie man vermeidet, falschen Online-Informationen aufzusitzen, In: Die Zeit, 6 31. Januar 2008, ⟨URL: <http://www.zeit.de/2008/06/C-Internetwahrheit>⟩ – Zugriff am 07. Juli 2011

Wikipedia: Kritik an Wikipedia – Wikipedia, Die freie Enzyklopädie 2011 ⟨URL: http://de.wikipedia.org/w/index.php?title=Kritik_an_Wikipedia&oldid=90947226⟩ – Zugriff am 07. Juli 2011

B.3.2. Rollenkarten

Blogger

Du bist ein aktiver Blogger. Du schreibst regelmäßig Artikel, dein Blog wird von vielen Menschen gelesen und deine Einträge werden regelmäßig, durchaus auch kontrovers, kommentiert. Außerdem benutzt du den Microblogging-Dienst Twitter.

Bei der Erstellung deiner Blogbeiträge recherchierst du soweit es dir möglich ist. In deinem Fachgebiet kennst du dich gut aus und schreibst daher hin und wieder Blogbeiträge zu fachlichen Themen.

Der Großteil deiner Beiträge sind Kommentare zu Tagespolitischen Themen. Auch wenn du nicht immer explizit schreibst, dass es sich nur um deine persönliche Meinung handelt, gehst du davon aus, dass dem Leser diese Tatsache im Allgemeinen bewusst ist.

Wenn dir Fehler in deinen Beiträgen auffallen oder du auf solche aufmerksam gemacht wirst korrigierst du die Beiträge oder ergänzt und weißt am Ende des Beitrages darauf hin.

Du siehst in der Bloggingszene keinen Ersatz für professionellen Journalismus, aber eine sinnvolle Ergänzung. Deiner Meinung nach sind allerdings auch Zeitungsartikel nicht immer gut recherchiert.

Verleger

Du bist Verleger von zwei Regionalen Tageszeitungen.

Du stehst der Bloggingszene sehr kritisch gegenüber. Zum einen siehst du ein großes Problem darin, dass sich zunehmend Menschen über Nachrichten über ungeprüfte Kanäle wie Blogs und Microblogs informieren. Du befürchtest, dass dort häufig falsche Informationen verbreitet werden und so möglicherweise auch Schäden angerichtet werden können. Die etwas höhere Geschwindigkeit ist für dich kein Argument, schließlich sind Informationen zu aktuellen Themen auch nach sehr kurzer Zeit auf den Internetseiten deiner Zeitungen und auf anderen Webseiten klassischer Medien sowie im Rundfunk zu finden.

Natürlich siehst du auch ein wirtschaftliches Problem. Wenn immer weniger Leute deine (und natürlich auch andere) Zeitungen kaufen wirst du irgendwann nicht mehr in der Lage sein deine MitarbeiterInnen zu finanzieren. Gut recherchierte Informationen können nur von qualifizierten und erfahrenen Journalisten gesammelt werden.

Du hoffst, dass die Menschen bald erkennen, dass eine Zeitung eine Qualitätsgarantie beinhaltet, auch wenn sie nicht kostenlos zu haben ist.

Wissenschaftler

Du bist ein angesehener Wissenschaftler und bist teilweise besorgt um die Qualität der Arbeit deiner Studierenden. Häufig werden Hausarbeiten primär auf der Grundlage von Internetquellen erstellt. Dabei wird viel zu selten auf die Qualität der Quelle geachtet. Selbstverständlich veröffentlicht auch du deine Fachartikel auf der Seite deines Instituts. Auch du benutzt für deine Recherchen gerne das Internet. Um „mal schnell etwas nachzulesen“ darf es auch mal die Wikipedia sein. Allerdings bist du durchaus stolz darauf auch das „richtige“ recherchieren in Bibliotheken noch gelernt zu haben.

Wichtig ist dir, dass deine Studierenden und auch die Teilnehmer und Zuschauer dieser Diskussionsrunde verstehen, dass es nicht ausreicht eine Quelle zu finden, sondern dass sie auch bewertet werden muss. Diese Tätigkeit erfordert Erfahrung und Übung und ist, deiner Meinung nach, von Computern nicht zu leisten.

Computer, Datenbanken, Internet, das alles sind nützliche Hilfsmittel, aber ohne Anstrengung ist auch mit diesen Hilfsmitteln keine Qualität zu erreichen.

Wikipedia-Autor

Du beteiligst dich in deiner Freizeit an der Erstellung und Pflege von Artikeln der Wikipedia. Auch wenn du dich nicht in allen Bereichen gut auskennst, glaubst du trotzdem einen guten Beitrag leisten zu können. Zum Beispiel überprüfst du bei Fachfremden Artikeln ob die Regeln bzgl. Quellenangaben und anderer Qualitätsanforderungen eingehalten werden.

Du vertraust darauf, dass die große Zahl an Autoren, die für die Wikipedia schreibt, dafür sorgt, dass die Inhalte im Allgemeinen fachlich korrekt sind.

Falls doch einmal falsche Informationen in einem Artikel stehen, führst du dies darauf zurück, dass die Regeln nicht eingehalten wurden.

Außerdem merkst du an, dass die Quellenangaben den Zweck haben, dass die Informationen überprüft werden können. Wenn jemand die Wikipedia als Quelle benutzt, muss er diese Überprüfung natürlich auch vornehmen. Dies sei natürlich auch Teil der Aufgaben eines Journalisten während seiner Recherche.

C. Rückmeldungen

C.1. Prof. Dr. Jochen Koubek

14. Juli 2011:

grundsätzlich eignet sich das von Ihnen bearbeitete Thema »Suchmaschinen« gut als Kontext. Allerdings verspielen Sie seine Möglichkeiten, indem Sie nach einem kurzen Aufhänger praktisch ausschließlich auf die technischen Dimension abzielen. Erst in der 10. UE kommen Sie auf den Kontext zurück, indem kritisch die Qualität von Suchergebnissen hinterfragt wird.

Binäre Bäume, Suchen und Sortieren, Implementieren und Testen sind ja schön und gut aber Informatik im Kontext geht nicht nur von der Lebenswelt aus, vergisst sie dann zwei Monate und kehrt am Ende zurück, sondern stärkt kontextbezogene Kompetenzen (regelmäßig die Informatik-Mensch-Gesellschaft-Standards) im Verlauf der gesamten Reihe. Nicht notwendig in jeder Stunde im gleichen Maße aber auch nicht nur flankierend. Ansonsten passiert da Übliche: nach einem motivierenden Einstieg wird informationstechnischer Unterricht gemacht und die letzte Stunde muss dann wegen Ferien / Zeugnissen / Wandertagen leider entfallen, was aber nicht schlimm ist, weil da eh nur so weiches Zeug im Plan steht (das ist sicherlich nicht Ihre Meinung aber gängige Praxis im Informatikunterricht).

Sie orientieren sich an den Bildungsstandards und bemühen sich um methodische Vielfalt. Das ist sehr lobenswert, aber das tun andere Ansätze auch. Das Alleinstellungsmerkmal von IniK ist der konsequente und möglichst durchgängige Bezug an der Lebenswelt und nicht nur an der Technik. Das ist Ihr Ansatz nicht oder kaum. Das macht ihn nicht schlecht, nur halt nicht kontextorientiert im Sinne von IniK.

C.2. Dr. Dieter Engbring

18. Juli 2011:

Es ist eine (sehr) anspruchsvolle Serie von UE entstanden, die auf jeden Fall, so wie sie hier dargestellt ist, nur in der Sek. II stattfinden kann. Schaut man von der Obligatorik (z.B. für das Zentralabitur darauf) wird man feststellen, dass man sich in einem LK oder sehr starken Grundkurs befinden sollte.

Informatik im Kontext wird von vielen nur als Konzept für die SI gesehen; ich sehe es wie sie auch als einen guten Zugang zu Inhalten aus

Lehrplan und Vorgaben, wodurch für die Schüler eine Sinnstiftung stattfindet. Es ist gut, dass sie die Informatikinhalte dabei nicht nur nicht aus den Augen verlieren sondern sehr stark betonen. Das macht das Ganze jedoch sehr anspruchsvoll.

Die methodischen Zugänge, die sie wählen, sind gut, vielfältig sowie schülerorientiert. Sie orientieren sich in ihrem Vorgehen durchaus an dem Schema, dass auf der Webseite zu Inik vorgeschlagen wird (Begegnungsphase ...)

By the way beschreiben sie Unterrichtsmethoden, die vor allem von den Lehrenden beherrscht werden müssen, soll der Unterricht nicht im Chaos ändern und die Zielsetzungen erreicht werden. Schüler erhalten z.B. durch Gruppenpuzzle oder Think-pair-share ein höheres Maß an Aktivität, die lernförderlich ist, da andere Formen des Unterrichts diese Aktivität unterdrücken, die eigentlich bei den Schülern vorhanden ist. Von daher wird so etwas auch gut angenommen, wenn es nicht eingeübt ist. Methoden und methodische Kompetenzen von Schülern sind eine Baustelle, die sie nicht mehr bearbeiten müssen.

Es sind – und das ist gut so – Elemente vorhanden, Inhalte auch ohne Rechnereinsatz zu erläutern oder zu veranschaulichen. Wichtiges Mittel im Informatikunterricht, da man damit gerade in GKs Schüler, die sonst abgehängt werden, wieder eingefangen werden.

Alle Implementierungsaufgaben erfordern idR sehr viel mehr Aufwand als Schüler für andere Fächer investieren müssen. Dies scheint ein wesentlicher Grund zu sein, weswegen die Leistungsunterschiede im IU oft so groß sind.

Sie sollten diese Vorschläge als Obermenge eines möglichen kontextorientierten Zugangs zu Inhalten des Oberstufeninformatik verkaufen, aber auch andeuten, wie ein Zugang im Rahmen einer Allgemeinbildung aussehen kann. Dies kann z.B. im Ausblick mit Bezug auf die tabellarischen Darstellungen der UEs geschehen.

Beim Bezug auf Vorgaben wie z.B. zum Zentralabitur oder den Bildungsstandards ist mir immer wichtig, dass diese nicht nur tabellarisch genannt werden sondern ausführlicher erörtert werden. Dies führt in der Regel zu Präzisierungen in Bezug auf die zu vermittelnden Ziele oder Kompetenzen und des unterrichtlichen Vorgehens. Dieses Diskutieren statt nur Benennen ist aber auch ein persönliches Steckenpferd von mir, so dass auf diesen Vorschlag noch weniger eingehen müssen als auf die Vorhergehenden. Auch meine Einschätzung müssen sie nicht teilen, da es ihre Arbeit ist und bleibt.

D. Erklärung des Studierenden

Hiermit versichere ich, dass ich die vorliegende Arbeit mit dem Titel
KONZEPTION EINER KONTEXTORIENTIERTEN UNTERRICHTSREIHE ZUM
THEMA SUCHMASCHINEN

selbständig verfasst habe, und dass ich keine anderen Quellen und Hilfsmittel
als die angegebenen benutzt habe und dass die Stellen der Arbeit, die anderen
Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnom-
men wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich
gemacht worden sind.

Münster, 28. Juli 2011