

Dr. Florian Hartling, Universität Halle
Bachelor-Studiengang „Fernsehproduktion“, FAM Leipzig

Skript

C Rechercheworkshop / Suchalgorithmen

Stand: Februar 2011

Achtung!

Skript ist nur zur Internen Verwendung im Seminar gedacht.

nach: Blittkowsky 2002, S. 191-201 (Im Buch noch viel ausführlichere Hinweise und Hilfen zur Online-Recherche, die aus Zeitgründen nicht näher behandelt werden können.

Strukturplan: Ablauf einer Recherche

Eine Online-Recherche kann sich nur auf Material richten, was auch im Netz vorhanden *und zugänglich* ist. Zum einen bedeutet das den Umgang mit einer riesigen Flut von Informationen, die nur als äußerst heterogene Rohmasse betrachtet werden kann und die es zu selektieren und prüfen gilt. Der Online-Rechercheur muss also Fertigkeiten entwickeln, sich in einem prinzipiell unendlichen Datenraum zurechtzufinden. Zum anderen bedeutet das den Einsatz von (kostenpflichtigen) Archiv- und Datenbankrecherchen sowie den klassischen Einsatz von Befragungen (in elektronischer Form). Erst wenn das verfügbare Material gesichtet und Wichtiges von Unwichtigem getrennt worden ist, kann die thematische Auswertung beginnen, die idealerweise zu journalistischen Beiträgen führt.

Der Ablauf einer Online-Recherche kann nach Blittkowsky in 8 Phasen geteilt werden:

Übersicht: Ablauf einer Online-Recherche kann nach Blittkowsky 2002

Vorab vorhanden: Material, Content, Rohstoff für Informationen, redundante Informationsmasse	Themenfelder wie etwa: Wirtschaft, Politik, Finanzen, IT/Computer, Internet und WWW, Kultur, Naturwissenschaft, Industrie, Verlagswelt, Bürokratie
Phase 1	Themenstellung Konkretisierung einzelner Themenaspekte
Phase 2	Formulieren von Recherchezielen
Phase 3	Herauslösen von Suchbegriffen Formulieren von Recherchefragen
Phase 4	Welche Suchdienste werden genutzt? Allgemeine SuMa (Google, Alltheweb, Altavista...) Spezielle SuMa Meta SuMa Allgemeine Kataloge (Yahoo, open directory, ...) Spezielle Kataloge
Phase 5	Einzelauswertung der Resultatlisten
Phase 6	Synoptische Prüfung der Suchergebnisse und Sondieren nützlichen Materials: informative Websites, URLs, Personennamen, E-Mail-Adressen, Fachbegriffe, PR-Material, Dokumentationen und Aufsätze, Multimedia-Daten usw.
Phase 7	Zusammenfassung der Suchergebnisse und ihre Vertiefung: Wiederholung von Suchmanövern für einzelne Begriffe, Speichernde Dokumentation von Rechercheergebnissen, direkte Bezugnahme auf Rechercheergebnisse, Auswertung, Lektüre, Analyse, Interpretation, Kontaktsuche etc.
Phase 8	Thematische Auswertung und Bezugnahme auf virulente Rechercheergebnisse

Phase 1: Themenstellung und Konkretisierung

Phase 1 bedeutet mit der Themenstellung die Basis der Recherche. Das Thema hängt ab von bereits bekannten Informationen oder eigenen Arbeitsschwerpunkten und stellt eigentlich eine Art „Vorverständnis“ der Recherche dar. Aus einem hermeneutisch strukturierten Geflecht von Begriffen, Gesprächen, Informationen, Meldungen usw. entsteht ein Thema. Dies ist auch von der zur Verfügung stehenden Zeit abhängig.

Phase 2: Formulierung von Recherchezielen

Aus dem Vorverständnis werden Rechercheziele abgeleitet. Abstraktionsvorgang: Überlegungen zum Thema werden „versprachlicht“. Am besten ist das Formulieren von Recherchefragen, wichtig ist die kritische Hinterfragung des Recherchethemas (neue Impulse und Gesichtspunkte.)

Die Rechercheziele versuchen, das Thema unter verschiedenen Aspekten zu fixieren:

- Was weiß ich bereits über das Thema?
- Was möchte ich (noch) über das Thema herausbekommen?
- In wie viele Teilaspekte kann das Thema untergliedert werden?
- Welche(r) Teilaspekt(e) ist/sind besonders hervorzuheben?
- Wie viel Zeit kann veranschlagt werden?

Hilfreich ist zudem die Erstellung eines Recherchemodells, was Themenstellung und Rechercheziele definiert. Dazu dient als Hilfsinstrument das Mindmapping.

Phase 3: Herauslösen von Suchbegriffen

Suchmaschinen (SuMas) und Webkataloge verstehen nur Suchbegriffe: Zumeist sind das Substantive, aber auch passende Adjektive und Verben („Phrasen“).

Das Ziel von Phase 3 besteht darin, elementare Suchbegriffe festzulegen: Synonyme! Die ermittelten Suchbegriffe werden dann nach Ober- und Unterbegriffen geordnet (Mindmapping!), sodass sich eine flexible Rangfolge ergibt und die Suche immer weiter präzisiert wird. Vor den Abfragen können dann einzelne Begriffe aus dem Modell ausgewählt und verschieden kombiniert werden.

Die Suche muss abgegrenzt werden gegen falsche Themen: Suchbegriffe können Mehrfachbedeutungen haben (Fußball ↔ festlicher Ball). Dies geht durch das Formulieren von ausschließenden Begriffen (NICHT Fest) oder Phrasen („Fußball WM“). (Boolsche Operatoren)

Wichtig ist zudem, dass sich während der Online-Recherche weitere, nicht vorformulierte Suchbegriffe ergeben können. Deshalb muss das Modell der Suchbegriffe

(auf Papier / im Computer) unbedingt für weitere Ergänzungen offen gehalten werden. Unnötige Suchbegriffe werden während Recherche gelöscht.

Wichtig ist auch noch: Auf welchen Sprachraum soll sich die Suche konzentrieren: deutsch, europäisch (welche Sprachen?), englisch-sprachig, international, usw... Entsprechend müssen Suchbegriffe evtl. in (einschlägiger!) Übersetzung verwendet werden (1 Sprache) oder mehrsprachig (globale Suche).

Phase 4: Welche Suchdienste werden genutzt?

Suchdienste sind eine Sache der Erfahrung und der Kenntnis. Ein erfahrener Online-Rechercheur bookmarkt die erfolgreichsten und nützlichsten Suchmaschinen und Kataloge. Recherchen werden aber nicht nur mit SuMas oder Katalogen durchgeführt, sondern auch anhand von Hinweisen in Emails, Newslettern, Websites.

Bei der Wahl der Suchdienste ist folgende Entscheidung zu treffen:

- Wo beginnt die Suche?
- Ist eine SuMa, ein Web-Katalog, ein Metasucher oder gar eine spezielle Datenbank / Website der geeignete Einstiegspunkt?

Der Einstiegspunkt und weitere Suchdienste zum Gegencheck sind entscheidend für den Erfolg der Recherche. Wichtig: Immer mehrere Suchdienste abfragen: Check / Gegencheck und Nachrecherche. Die Wahl der Suchdienste hängt ab von der Art und Weise, wie die gelieferte Information strukturiert sein soll:

Suchmaschine: Wird benutzt, um möglichst viele und breit gestreute Informationen zu finden. Relevante Treffen finden sich zumeist in den oberen Rängen, (allerdings ist das Weiterblättern bei speziellen Suchanfragen manchmal wichtig)! Nachteil: Flut von Treffern, oftmals von schlechter Qualität, gute Suchanfragen sind maßgeblich! Marktführer ist Google.

Web-Katalog: Wird benutzt, wenn man gezielt auf ein Thema und dessen Untergliederung in Rubriken zurückgreifen will. Der Erfolg einer Katalogrecherche hängt von der Qualität seiner redaktionell betreuten Themenrubriken ab. Geboten bekommt man pro Rubrik eine Auswahl von kommentierten Verweisen. Nachteil: Sehr lückenhaft und oftmals nicht auf dem neusten Stand.

Metsuchmaschine: Können die Recherche verkürzen, da sie mehrere Suchdienste parallel abfragen und auf die Ergebnisse Filtern anwenden. Nachteil: Auch hier Flut von schlechten Treffern möglich.

Das Resultat von Suchdienst-Abfragen sind Trefferlisten. Das Analysieren dieser Trefferlisten ist ein kreativer und analytischer Akt, der ein Gespür für URLs und den mit ihnen verbundenen Webseiten voraussetzt.

Wenn zu viele Treffer auf einen Suchbegriff folgen, dann sollte man nicht versuchen, die Ergebnisseiten zu durchstöbern. Eher bietet es sich an, alternative Suchbegriffe einzugeben, Phrasen zu verwenden und durch Boolesche Operatoren auszugrenzen. Ziel: Die Anzahl der Treffer von vornherein niedrig halten. Die Durchsicht von 10 Trefferlisten kann noch erfolversprechend sein, darüber hinaus sind es un-

nötige Mühen. Auch das Prüfen von Trefferlisten erfordert vom Rechercheur Erfahrungen im Umgang mit Inhaltsbeschreibungen von Webseiten.

Phase 4 ist geprägt von der Suche nach Inhalten, die im Sinne der Recherchefrage weiterhelfen können. Gleichwohl leitet die Einzelauswertung der Resultatlisten und Themenrubriken bereits zu Phase 5 über.

Phase 5: Einzelauswertung der Resultatlisten

Die Einzelauswertung der Resultatslisten beschäftigt sich vor allem mit Material-sichtung und Materialprüfung. Grundlage hierfür ist wiederum das Vorverständnis vom Thema. In dieser Phase wird die Themenstellung mit den Treffern bzw. mit den Inhalten der gefundenen Webseiten in Verbindung gesetzt. Resultatslisten werden gecheckt und Treffer beurteilt, ob sie etwas zum Thema beizutragen haben. Try-and-Error führt zu nützlichem Material.

Diese Prüfung kann oft nicht in einem Durchgang beendet werden – hier muss dann nachrecherchiert werden. Umso wichtiger ist es, das bereits recherchierte Material gut auf dem Rechner zu archivieren: Bookmark auf die entsprechende Website, besser noch komplette Website speichern oder als PDF sichern. Das gesicherte Material ist in der (möglicherweise digitalen) Recherchemappe zu dokumentieren und bildet damit eine Vorstufe zu einem eigenen Archiv.

In Phase 5 wird darum nur gesammelt. Dabei kann und wird mehr Material anfallen, als später auch im fertigen Produkt verwertet wird. Hier muss der Rechercheur Erfahrungen sammeln und ein Gespür dafür entwickeln, zum einen nicht neben dem Thema (und damit zuviel) zu sammeln, zum anderen aber sorgfältig und unter Einsatz aller zur Verfügung stehenden Mitteln zu arbeiten.

Die thematische Auswertung (die journ. Umsetzung des Themas) muss während der Prüfung nach und nach bedacht werden. Auch die Möglichkeit des Rechercheabbruchs kommt in Frage. (Thema ist doch nichts wert oder andere Recherchewege)

Phase 6: Synopsis und Verdichten der Ergebnisse

An die Sichtung des gesammelten Materials schließt an: die konsequente Zusammenfassung und erste Auswertung. Die Befunde werden auf das Thema hin konzentriert, überflüssiges verworfen. Diese Phase ist damit gekennzeichnet von kritischer Auswahl, strukturierender Anordnung und Verwerfung.

Essentiell wichtig ist hier die Prüfung der Qualität und Glaubwürdigkeit des Materials: Ist der Inhalt einer Webseite zitierfähig? Kann dem Inhalt vertraut werden oder muss ihm misstraut werden? Dabei ist es u.U. notwendig, die Quellen auf ihren Ausgawert zu prüfen, den Urheber (Impressum, Quelltext, Domainbesitzer) eindeutig zu identifizieren und den Wahrheitsgehalt zu prüfen.

Das gilt auch für Online-Medien mit hohem Bekanntheitsgrad (Spiegel.Online hatte berühmte Enten, Nachrichtenagenturen) – wichtig ist in jedem Fall kritisches Ge-

gen- und Nachrecherchieren. Jede Recherche sollte von der Frage geleitet sein, wie stichhaltig die mitgeteilte Information ist und welche Einzelaspekte sie bringt.

In dieser Phase findet auch eine Vorauswahl und Einsortierung in Informationsbereiche statt. Hier muss geprüft werden, wie das selektierte Material in die Journ. Arbeit einfließen kann. Es gilt außerdem wieder, Schwerpunkte zu bilden, die den Recherchezielen entsprechen. Auch hier ist eine fragende Haltung gegenüber dem Material wichtig. Auch hier muss evtl. noch ein zweiter Recherchedurchgang abgeschlossen werden.

So muss es auch immer möglich sein, die Recherche selbst zu bezweifeln (falsche Suchworte, falsche Fragstellung, unzulängliches Recherchematerial).

Phase 7: Zusammenfassung und Vertiefung

Zusammenfassung und Vertiefung der Suchergebnisse findet offline statt. In Einzelfällen kann durch weitere Online-Phasen Fehlendes ergänzt werden. Hier steht die Frage im Vordergrund, welche Suchergebnisse aussagekräftige Bausteine bilden, die in das bearbeitete Thema einbezogen werden können.

Erst in dieser Phase findet eine inhaltliche Auseinandersetzung und sachliche Diskussion des selektierten Materials statt. Dies geschieht in einer abschließenden Arbeitsphase, die geprägt ist von stringenter Lektüre der recherchierten Quelle und von deren Analyse und Interpretation.

Analyse und Interpretation sollte von gedanklicher Stringenz begleitet sein. Man beginnt die Materialauswertung mit einer These und der Suche nach Indizien für ihre Stimmigkeit oder Widersprüchlichkeit. Im Rahmen der Untersuchung können Thesen dann abgeändert, neu formuliert und überprüft werden.

Phase 8: Thematische Auswertung

Die thematische Auswertung besteht darin, dass der Journalist allein vor seinem Desktop aus der Materialfülle eine „Story“ macht. Das ist ganz äquivalent zum letzten Schritt in der klassischen, journalistischen Recherche nach Haller.

Problematisch ist hier, dass eine Online-Recherche eine ungeheure Materialfülle mit sich bringt, sodass die Erfassung bei knappen Zeitressourcen oft scheinbar nur oberflächlich möglich ist. Bei der konsequenten Selektion nach relevanten Fakten kann diese Gefahr aber vermieden werden.

Wichtig ist zudem eine konsequente Systematisierung des Materials. Die Arbeit am Computer und mit dem Internet muss nach festen Ordnungsprinzipien organisiert sein. Jederzeit muss ein Rückgriff auf das verarbeitete Material möglich sein. Das heißt analog zu Halle:

- Für jede Recherche wird eine elektronische Recherchemappe angelegt.

- Themenordner mit guter Unterstruktur (Haupt und Unterverzeichnisse), in die am besten gespeicherte, noch besser PDF-gesicherte Versionen der Webseiten kommen
- Unterstruktur nach Unterthemen anlegen
- Webseiten, PDF-Dokumente, Bilder, Multimedia-Inhalte, Ideen, Zwischenversionen der Texte in je einzelne Verzeichnisse
- dazu organisierte und sortierte Favoriten / Bookmarks
- extrem wichtige Funde ausdrucken, organisiert abheften
- Backups machen
- Desktop-Suche installieren und verwenden

Algorithmen

Mindmapping

Eine Mindmap bzw. Mind-Map ist eine grafische Darstellung, die Beziehungen zwischen verschiedenen Begriffen aufzeigt. Mindmaps sind eng verwandt mit Semantischen Netzen und Concept-Maps. Da MindMapping® ein eingetragenes Warenzeichen ist, hängt die genaue Definition einer Mindmap davon ab, ob man nur vom Markeninhaber zugelassene Mindmaps oder auch vergleichbare Konzepte als solche bezeichnet.

Mindmaps enthalten das zu bearbeitende Hauptthema möglichst knapp formuliert und/oder als Bild dargestellt in der Mitte des Blattes und nach außen verschiedene Hauptäste mit weiteren Unterästen, die die dazugehörigen Informationen ranggerecht darstellen. Bei der Erstellung sollten Farben und Bilder und stichwortartig Formulierungen benutzt werden, um der kreativen Arbeitsweise des Gehirns gerecht zu werden, und um die Mindmap schneller lesen und überblicken zu können. Mindmapping kann zum Beispiel eingesetzt werden zur einfachen Ideensammlung, dem Brainstorming, zur Erfassung und dem Strukturieren komplexer Inhalte, zur Erstellung von Referaten oder Vorträgen, zum Mitschreiben bei Vorträgen, zur Strukturierung von Wissen und zur Prüfungsvorbereitung. Einsatzmöglichkeiten sind also von der Schule über die Wirtschaft bis zur Wissenschaft zu finden.

Für die Prüfungsvorbereitung ist das Mindmapping deswegen gut geeignet, da in kreativen Schritten eine möglichst griffige Darstellung des Lernstoffes erarbeitet wird, und später weiter durch systematisch wiederholte Beschäftigung mit dieser Wiedergabe der gelernte Stoff gefestigt wird. Eine solche übersichtliche Anordnung der semantischen Struktur des Wissens fördert die stabile Erinnerbarkeit über lange Zeiträume.

Die kombinierte verbal-bildliche Mindmap ist eine sehr kompakte Dokumentation hierarchischer Gliederungen. Argumentiert wird, dass Kreativität und Lerneffizienz durch Synergie-Effekte gesteigert werden, die beim gleichzeitige Ansprechen sowohl der vorwiegend verbal orientierten als auch der vorwiegend bildlich orientierten Hirnhälfte (visueller Cortex) entstehen.

Mindmap-Werkzeuge liegen bereits vielfach als Software vor, von Freeware bis zu ausgefeilten Geschäftsanwendungen. Als besondere Vorteile des computergestützten Mindmappings (auch Business-Mapping genannt), gelten die einfache Handhabung sowie schnelle und präzise Strukturierung. Durch die einfachen Änderungsmöglichkeiten wird die Zusammenarbeit im Team gefördert und außerdem die Kreativität gesteigert. Durch unterschiedliche Formatierungen können Gedanken bes-

ser gewichtet werden. Mindmaps können aber auch einfach nur per Hand erstellt werden.

Quelle: Wikipedia, <<http://de.wikipedia.org/wiki/Mindmap>>

Software: Mindmapper (kostenpflichtig), Freemind (open source)

Synonymie

Synonymie (aus dem Griechischen *synonymos* synonymos für gleichnamig) ist eine Beziehung zwischen lexikalischen Zeichen. Ein Synonym ist ein lexikalisches Zeichen, das die gleiche Bedeutung hat wie ein anderes lexikalisches Zeichen.

1. Formen der Synonymie

Man unterscheidet strikte oder absolute Synonymie und Bedeutungsähnlichkeit. Strikte Synonymie setzt nicht nur voraus, dass zwei lexikalische Zeichen die gleiche denotative Bedeutung haben, sondern auch, dass sie in allen Kontexten austauschbar sind und in allen Kontexten die gleiche Wirkung haben.

Beispiele für strikte Synonymenpaare im Deutschen sind: *Orange - Apfelsine*; *Streichholz - Zündholz*.

Ein Beispiel für ein nicht-striktes Synonymenpaar ist *durcheinander - wirr*, wie an den folgenden Beispielen deutlich wird:

(1) *Er wirkt heute etwas durcheinander = Er wirkt heute etwas wirr.*

(2) *Seine Rede heute war etwas wirr != Seine Rede heute war etwas durcheinander.*

Folgende Aspekte der der konnotativen Bedeutung bewirken, dass zwei lexikalische Zeichen bei gleicher denotativer Bedeutung nicht den strengen Anforderungen strikter Synonymie genügen:

- Regionale Differenzierung: Metzger - Fleischer; Brötchen - Schrippe
- Natives Wort vs. Fremdwort: Fahrstuhl/Aufzug - Lift
- Fachwort vs. Laienwort: Rechner - Computer
- Lateinischer Ursprung vs. germanischer Ursprung: vertikal - senkrecht

In den meisten alltäglichen Verwendungssituation genügt das weniger strikte Kriterium der Ersetzbarkeit *salva veritate* in typischen Kontexten, um ein Paar lexikalischer Zeichen als Synonyme zu identifizieren:

(3) (a) *Er sitzt schon wieder am Computer* <==> (b) *Er sitzt schon wieder am Rechner.*

Aus dem in Satz 3a beschriebenen Sachverhalt folgt notwendigerweise der Satz in 3b beschriebene Sachverhalt und umgekehrt.

2. Synonymie zwischen sprachlichen Zeichen verschiedenen Typs

Synonymie besteht nicht nur zwischen lexikalischen Zeichen des gleichen Typs:

- Ein einwortiges lexikalisches Zeichen kann mit einem mehrwortigen lexikalischen Zeichen synonym sein (*stören - dazwischenfunken - in die Quere kommen*)
- Ein Wortbildungsmittel kann mit einem einwortigen oder mehrwortigen lexikalischen Zeichen synonym sein (*Online-- im Netz*)
- Eigennamen, vor allem Produktnamen, können sich zu generischen Termen und damit zu Synonymen für die Produktbezeichnung entwickeln, zum Beispiel Tempo für Papiertaschentuch. Derartige Synonyme nennt man auch Begriffsmonopol.

3. Synonymwörterbücher

Synonymwörterbücher gehören zu den Wörterbüchern mit eingeschränktem Informationsprogramm. Zum jeweiligen Stichwort werden die sinn- und sachverwandten Wörter angegeben.

Da absolute Synonymie selten ist, geben die meisten Wörterbücher dieses Typs eher bedeutungsähnliche Wörter an. Benutzer dieser Wörterbücher müssen deshalb über einen hohen Grad von Sprachkompetenz verfügen, um für einen bestimmten Kontext das passende Synonym auswählen zu können.

Man unterscheidet zwei Arten von Synonymwörterbüchern:

- distinktive Synonymiken spezifizieren die Lesarten von polysemen Lexemen und ordnen die bedeutungsähnlichen Wörter der jeweiligen Lesart zu (Bsp.: durcheinander (Person) - verwirrt; durcheinander (Sachen) - chaotisch, vermischt, wie Kraut und Rüben); eine distinktive Synonymik des Deutschen ist Schülerduden. Die richtige Wortwahl.
- kumulative Synonymiken ordnen die einem Lexem ähnlichen lexikalischen Zeichen ohne Unterscheidung der Lesarten diesem Lexem zu; eine kumulative Synonymik ist Duden Band 8. Die sinn- und sachverwandten Wörter.

Synonymenwörterbücher werden häufig verwendet, um in einem Text zu häufige Wiederholungen eines Wortes zu vermeiden. Sie können auch für die systematische Wortschatzarbeit im Zweitsprachunterricht herangezogen werden.

Das Gegenteil der Synonymie ist die Antonymie.

Quelle: Wikipedia, <<http://de.wikipedia.org/wiki/Synonymie>>

Boolesche Operatoren

Ein boolescher Operator (englisch Boolean operator) ist ein logischer Operator, der auf der booleschen Algebra beruht. Boolesche Operatoren beinhalten Ausdrücke wie UND, ODER, NICHT und XOR (ausschließendes ODER). [...]

Auch in Online-Katalogen werden sie verwendet, um Suchbegriffe miteinander zu kombinieren.

Quelle: Wikipedia, <<http://de.wikipedia.org/wiki/Boolescher%20Operator>>

ausschließende Begriffe, Phrasen