

Tippfehler in Bibliothekskatalogen

Möglichkeiten einer EDV-gestützten Ermittlung

Henning Klauß

Ein nicht unerheblicher Anteil von Benutzer-Recherchen am OPAC und Vorkazessionsarbeiten laufen deshalb ins Leere, weil die gesuchten Titelaufnahmen aufgrund von Tippfehlern¹ in der Datenbank nicht ermittelt werden können, weil mit zutreffender Schreibweise gesucht wird, was mit Tippfehler katalogisiert wurde.² Ob es sich bei den Tippfehlern tatsächlich um die „banalste Schadensweise“ (Kuhn 2000) defizitärer Titelaufnahmen handelt, mag dahingestellt sein. Fest steht, dass dann, wenn sich die korrekte Schreibweise nicht in einem verknüpften Feld findet, Schaden entsteht: Verzicht auf die Lektüre von etwas, dessen Lektüre dem betreffenden Forschungsprojekt nützlich gewesen wäre, vermeidbare Neubeschaffung, vermeidbare Fernleihen, vermeidbare Wege zu Bibliotheken, die das Gesuchte ebenfalls haben ... Ein Teil der hier aufgeführten Schäden ist also unmittelbar kostenverursachend; mittelbar kostenverursachend im kulturellen, wissenschaftlichen bzw. volkswirtschaftlichen Sinne sind alle hier aufgeführten Schäden.

Der durch Tippfehler entstehende Schaden überwiegt *bei weitem* den, der z.B. durch Verwechslung von Verfasser-, Urheber- und Sachtitelwerk, fehlerhafte Auflagen- oder Sprachbezeichnungen oder gar den des nicht rechtzeitigen Wechsels vom Hauptsachtitel zu dessen Zusatz usw. entsteht - womit nicht gesagt ist, dass derlei Fälle irrelevant sind. Angesichts der Relevanz der Tippfehler war und bin ich erstaunt, wie wenig diese bisher Gegenstand bibliothekarischer Diskussion bzw. Literatur gewesen sind. (Eine gewisse Ausnahme hiervon beinhaltet die Literatur über automatische Indexierung).

Von den Tippfehlern soll im Folgenden der Teil besprochen werden, der nicht mit Normdateien verknüpft, aber retrievalrelevant ist: Hauptsachtitel, Zusatz dazu usw. (in manchen Bibliothekskatalogen sind auch die Felder für z.B. Gesamttitel, Bandangaben und Fußnoten indexiert) nicht jedoch Auflagenbezeichnung, Erscheinungsort, Verlag, Kollationsangaben usw. Die Stoßrichtung

1 Genau genommen handelt es bei diesen Problemfällen um Schreibfehler. Diese können auch durch Defizite in Kopierfunktionen der verwendeten Bibliothekssoftware (insbes. Sonderzeichen) entstanden sein; es ist daher nicht jeder Schreibfehler auf Fehler der Katalogisierer zurückzuführen. Das Gros der Schreibfehler resultiert allerdings aus Tippfehlern.

2 Natürlich kann eine mit einem Tippfehler behaftete Titelaufnahme zufällig ermittelt werden, weil die Suchanfrage ebenfalls genau diesen Tippfehler beinhaltet.

ist hierbei nicht die Entwicklung einer fehlertoleranten Bibliothekssoftware (Schulz 1994) oder die Überprüfung/Unterstützung der sachgerechten Eingabe der Benutzer (Weller/Drabenstott 1996), sondern die Korrektur des zugrunde liegenden Problems, nämlich der (durch Bibliothekare) eingegebenen Daten in die Datenbank. Damit ist der weitere Versuch, eine Recherche-Unterstützung zu entwickeln, nicht erledigt, aber die Bedingung von dessen Möglichkeit geschaffen. Es handelt sich bei den hier dargestellten Überlegungen lediglich um Reparatur-Bemühungen, also die Korrektur eines Defizits. D.h. nicht, dass die Bemühungen um die Erstellung tippfehlerfreier Titelaufnahmen nachlassen dürfen. Aber es geht hier nicht um die Erstellung tippfehlerfreier Titelaufnahmen, sondern um die Ermittlung und den Umgang mit tippfehlerbehafteten Titelaufnahmen.³

Die hier dargestellten Verfahren dienen nicht dazu, die Titelaufnahmen absolut fehlerfrei zu machen; sie dienen "lediglich" dazu, die Titelaufnahmen fehlerfreier zu machen. Das Aufspüren von normdateirelevanten Ansetzungsfehlern ist damit nicht zu erledigen; ebenfalls unbeantwortet bleibt die mögliche Frage nach der Ermittlung von denjenigen Fehlern, die nicht *in* einem Wort auftreten, sondern das *ganze* Wort betreffen. Wenn es im Titel z.B. heißt „Karl der Große und Papst Leo II.“ so gibt es m.E. (derzeit?) keine maschinelle Möglichkeit, den Hinweis zu erhalten, dass es sich hierbei wohl um „Papst Leo III.“ gehandelt haben dürfte (weil Leo II. von 682 - 683 Papst war, Karl der Große aber von 747 - 814 lebte; Leo III. war von 795 - 816 Papst); sind „Salier und Ottonen“ in einer Titelaufnahme aufgeführt, so sind - wenn das Buch nicht eine (transkulturelle bzw. zeitübergreifende) Gegenüberstellung, sondern die Darstellung eines Zusammenhanges intendiert - wahrscheinlich „Salier und Ottonen“ gemeint... (Zur Frage, wie mit solchen Fällen umzugehen ist, vgl. Sie bitte Abschnitt II., insbes. Fußnote 6). Eine ohne historisches Wissen falsifizierbare Sequenz wie „a strategie management approach“ statt „a strategic management approach“ lässt sich durch die Analyse des sprachlichen Zusammenhanges, nicht aber mit den hier dargestellten Verfahren ermitteln, da jedes der hier beigespielten Wörter im einzelnen genommen richtig sein könnte: Genau genommen wird einerseits (leider) nicht überprüft, ob die einzelnen Wörter *richtig sind*, sondern nur, ob diese - kontextfrei, d.h. unabhängig von z.B. grammatischen Strukturen - *richtig sein könnten*. Es ist dabei a-

3 Nicht, weil ich an dieser Stelle keine Erfahrungen von Differenzen hätte, aber meine Intention ist weder der komparative, noch gar der quantitative Vergleich von Datenbanken, wie sie sich z.B. bei Dwyer (1991, 81) findet. Meine Intention ist schlicht die Verbesserung der Beschaffenheit und damit effizienten Benutzbarkeit von Bibliothekskatalogen.

ber andererseits egal, ob es sich bei diesen Tippfehlern um Dreher, Auslassungen oder was auch immer handelt.

Grundsätzlich gilt: Je intensiver Titelaufnahmen sachkompetent kontrolliert werden, desto geringer ist die Fehlerquote. Infolgedessen wird zum einen die intensive Nutzung von Fremddaten befürwortet und zum anderen aus den folgenden Ausführungen der gesamte Sektor ausgeklammert, der Normdaten berührt: Diese Elemente der Titelaufnahmen entziehen sich zum einen weitgehend einer maschinellen Kontrolle und sind zum anderen i.d.R. durch ein Normdaten-Update zu reparieren: Der Produktionsprozess von Normdaten ist hochgradig kontrolliert; Normdaten erfreuen sich daher einer hohen Anerkennung. - Im Folgenden soll im Wesentlichen die Ermittlung von Tippfehlern (I.), am Rande der Umgang damit (II.) thematisiert werden.

I. Die Ermittlung von Tippfehlern

Es gibt grundsätzlich drei verschiedene Methoden, Tippfehler in Bibliothekskatalogen zu ermitteln: Zufallstreffer, Recherchen aufgrund eines bestimmten, (etwas) generalisierbaren Anfangsverdachts (für Spezialbibliotheken: in besonders sensiblen Katalogsegmenten) und der Abgleich mit autorisierten Wörterbüchern.

Zunächst zu den *Zufallstreffern*: Wenn bei einer Katalogrecherche eine Titelaufnahme mit einem Tippfehler ermittelt wird, so sollte niemals nur diese eine, sondern alle in gleicher oder ähnlicher Weise fehlbearbeiteten Titelaufnahmen korrigiert werden. Die Wahrscheinlichkeit, dass gleiche oder ähnliche Fehler in Reihe auftreten (z.B. bestimmte Buchstabendreher), ist größer als das Auftreten anders gearteter Fälle. Jeder Zufallstreffer sollte daher für weitere Korrekturbemühungen genutzt werden. Dieses Verfahren hat zwar eine hoffentlich nicht hohe Erfolgsquote, sollte aber dennoch praktiziert und ggf. zu einer trunkierten Recherche und/oder einem gezielten Listebrowsing ausgedehnt werden: Ein Blick ins Wörterbuch lehrt, ob es neben dem gleichen noch ähnliche Tippfehler gibt; das setzt voraus, dass diese Tippfehler am Ende der untersuchten Wörter auftreten.

Hieran anschließend gibt es eine zweite Methode:

Für die *Recherchen aufgrund* eines bestimmten, (etwas) generalisierbaren *Anfangsverdachts* wird jeweils ein bestimmter Abschnitt des Titelwörterbuches durch ein Listebrowsing untersucht. Das ist jedoch erstens aufwendig und setzt zweitens die Definition des zu untersuchenden Abschnittes voraus. Die hierfür notwendige Auswahl kann entweder zufällig oder aufgrund eines Anfangsverdachts bzw. einer für die lokalen Bedürfnisse beson-

dere Sorgfältigkeit erfordernden Titel-Wörterbuch-Abschnittes bestimmt werden (*Randall 1999, 164* wählte z.B. „economy“).⁴

Um ein weiteres Segment von (potentiellen) Tippfehlern in den Blick zu bekommen, ist folgendes Verfahren praktikierbar, mit dem diese Tippfehler in die Zange genommen werden:

Zunächst wird eine Liste der für den Lokalkatalog besonders relevanten Such- einstiege (Begriffe; Namen von Personen und Körperschaften spielen hierbei aus den o.g. Gründen keine Rolle) erstellt. Um die hierbei möglicherweise auftretenden Tippfehler zu ermitteln, werden diese Begriffe möglichst umfassend trunkiert. Um die entstehende Liste nicht unnötig lang werden zu lassen, wird die Recherche mit einer booleschen Nicht-Verknüpfung durchgeführt. Die hierbei gewählten Begriffe der „Nicht“-Liste sind solche, die im zu untersuchenden Katalog zu Recht vorkommen; sie sollten an möglichst später Stelle trunkiert werden. Die „Nicht“-Liste sollte ggf. mehrere Wörter enthalten, um die entstehende Trefferliste möglichst kurz und damit prägnant zu gestalten. Wenn die durch die Recherchestrategie entstehende Trefferliste zahlreiche Titelaufnahmen enthält, die zu unrecht (d.h. nicht aufgrund eines Tippfehlers) ermittelt worden sind, ist die Recherche zu modifizieren, indem z.B. die Wörter der „Nicht“-Liste sachdienlich vermehrt werden. Je besser die Recherche definiert ist, desto kürzer und reichhaltiger an Tippfehlern ist die Liste.

Für die im folgenden angegebenen Beispiele wird unterstellt, dass „?“ genau ein Zeichen (Maskierung), „*“ beliebig viele Zeichen (Trunkierung) ersetzt und dass Leerstellen eine Oder-Verknüpfung bewirken.

suche	und nicht
betriebsw*	betriebswirtschaft* betriebswirt betriebswirte betriebswis- senc*
betr*ft*	betriebswirtschaft* betrifft betriebswirtschaft* betriebsauf- trag* betriebsvorschrift*
e*omi*	economi* ekonomi* economy econometric* ergonom* evo- lutionsoekonomi* entwicklungsoek*

4 Es gibt die Möglichkeit, sich des Problems der Zufälligkeit zu entheben, indem nicht ein bestimmtes Segment, sondern das gesamte Wörterbuch durchforstet wird (vgl. Sie bitte *Ballard 1992, 14ff*), aber den dafür notwendigen Aufwand vermag sicherlich kaum eine Bibliothek zu erbringen. Zudem handelt es sich bei Bibliothekskatalogen ja i.d.R. nicht um abgeschlossene Datenbanken; ein solcher Aufwand wäre also gelegentlich zu wiederholen ...

econ*	economi* economy econometri* econ
env*nt*	environment* environnement* envelopment
intern*	intern interne* internier* internat interna?ion* internal*
inv*on*	investition*
jur*	juristi* jura jurist juristen* juridica juris jurispruden* jurathek juristis jurisdicti* juridique* jurep juri jurij juridicum
kul*	kultur* kultus* kult kultisch
makro*	macro* makrooekonom* macroeconom* macro makro mak- ros makroprogr* macrop*
man*nt*	management* mandant*
mar*ft*	marktwirtschaft* marktfuehrerschaft* marktaufteilung* markgraftum*
mi?roe*	microeconom* micropaedia* microcomput* mikroelektroni* micro mikro
mikro*	mikro mikrocomput* mikrooekonom*
nati*	national* nation nations native nationen nationes nationaux
ök*mi*	oekonomi* oeonomi*
ökol*	oekologie* oekologisch* oekologisierung
ph*ph*	philosoph* philadelphi* photograph* phosphat*
rec*ft*	rechtswissenschaft* rechtskraft* rechtsvorschrift* rechtsg- schaeft* rechtszeitschrift*
wir*ft*	wirtschaft* wirthschaft*
wis*ft*	wissenschaft* wissenschaftgesellschaft*

Eine solche Liste kann jeweils nur anhand des Lokalkataloges entwickelt werden; bei der Ergänzung dieser Liste könnte ich mir die Zuarbeit von z.B. Sprachwissenschaftlern vorstellen. Die hier vorgeschlagene Liste beansprucht nicht, einem standardisierten Test für Datenbankqualität zuträglich zu sein (vgl. Sie bitte *Cahn 1994*).

Für den *Abgleich mit einem autorisierten Wörterbuch* sind aus einem bestimmten Katalogsegment - z.B. einem bestimmten Bereich von Signaturen, Inventarnummern ... am besten in Kombination mit der Definition der Sprache - eine Reihe von Elementen der Titelaufnahmen zu entladen: Auf jeden Fall der Hauptsachtitel samt Zusatz; falls die Felder für Fußnoten, Gesamttitel und Bandangaben indiziert sind, sollten sie hinzu genommen werden; von der Einbeziehung der Parallel- und Einheitssachtitel in dieses Verfahren rate ich

ab. Die so erzeugte Datei ist mit einem autorisierten Wörterbuch abzugleichen. Hierfür ist entweder der Einsatz spezifischer Software (z.B. die linguistische Softwarekomponente PASSAT, vgl. Sie bitte *Knorz 1999*; für Anmerkungen zur CAI-Engine sehen Sie bitte *Hauer 2000*) oder als pragmatische Näherung das Einbinden dieser Datei in ein Schreibprogramm denkbar, das eine möglichst gute Rechtschreibkontrolle (evtl. mit Wörterbüchern in verschiedenen Sprachen) hat. Wenn die verwendeten Wörterbücher nicht gut sind, ist zu überprüfen, ob sich autorisierte Wörterbücher in das Schreibprogramm einbinden lassen.

Auf diese Weise werden die von dem verwendeten Wörterbuch nicht abgedeckten Zeichenketten ermittelt. Die in diesem Wörterbuch nicht vorhandenen Zeichenketten werden vom Schreibprogramm markiert. Natürlich sind nicht alle markierten Zeichenketten Tippfehler. Das festzustellen, zieht einen z.T. erheblichen Aufwand nach sich, denn manche Zeichenketten sind „dudenfest“, aber nicht im Wörterbuch des Schreibprogramms: Insbesondere Katalogdaten wissenschaftlicher Literatur enthalten einen hohen Anteil von Wortunikaten und Neuschöpfungen (vgl. Sie bitte *Eversberg 2000*). In diesem Fall muss das Wörterbuch ergänzt oder die jeweilige Markierung ignoriert werden. (Für große Bibliothekskataloge mag sich an dieser Stelle der Einsatz von aufwendigeren computerlinguistischen Verfahren für die Abarbeitung der erstellten Listen lohnen, vgl. Sie bitte z.B. *Toole, 1999, 122 ff*; ich gehe an dieser Stelle von der intellektuellen Überprüfung dieser Listen anhand der Markierung aus, um sicher zu stellen, dass diese Verfahren in *jeder* Bibliothek praktiziert werden können.) - Bei diesen Verfahren ist es natürlich nicht gewährleistet, dass die nicht-markierten Zeichenketten tippfehlerfrei sind; s.o.

Für den Bearbeiter ist es wahrnehmungspsychologisch ratsam, dafür zu sorgen, dass die entstehenden Listen kurz und die Mengen der darin indizierten Wörter möglichst gering sind. Es empfiehlt sich daher erstens, durch den sachgerechten Aufbau des Wörterbuchs die Menge der indizierten Wörter zu reduzieren und zweitens zur Verkürzung der Listen vorher einige Maßnahmen zu ergreifen: Die Nichtsortierkennzeichen sollten eliminiert werden; zur Reduktion von Komposita sollten die Bindestriche durch Leerstellen ersetzt werden; zudem sollte eine Reihe von Zeichenketten eliminiert werden: alle Satzzeichen und Zahlen, bestimmte und unbestimmte Artikel, häufig wiederkehrende Zeichenketten wie „Bd.“, „Univ.“, „Diss.“, „und“, „als“, „Einführung“, „Grundlagen“, „Handbuch“, „Reihe“, „Beiträge“ usw. Zur Einschränkung des Kontrollaufwandes ist es sinnvoll, diese Datei so zu formatieren, dass in jeder Zeile (je ein Absatz) nur ein Begriff steht und dann eine alphabetische Sortierung durchführen zu lassen. - Ich rate dringend davon ab, den Aufwand für die Durchführung dieser Maßnahmen zu reduzieren, da jede Sparsamkeit an dieser Stelle einen erheblichen Mehraufwand zu einem späteren Zeitpunkt nach

sich zieht (ich habe mir für diesen Zweck ein Makro geschrieben; bei der Durchsicht von Listen gibt es die Wahl zwischen Wörterbuch-Ergänzung, Ignorierung der Markierung oder Aufnahme in das Makro, um die Löschung aus dieser Liste zu bewerkstelligen).

Da das komplettierte Wörterbuch nur auf dem jeweiligen PC ist, ist damit entweder eine Grenze sinnvoller Arbeitsteilung oder aber die Notwendigkeit, die modifizierten Wörterbücher anderen Mitarbeitern zur Verfügung zu stellen, verbunden.

II. Umgang mit ermittelten Tippfehlern

In allen Fällen stellt sich die Frage, wie mit den ermittelten Fehlern (oftmals handelt es sich hierbei ja zunächst lediglich um eine Vermutung) umzugehen ist. Die Sorgfaltspflicht verbietet in jedem Fall die automatische Korrektur, es geht vorerst nur darum, mögliche Tippfehler in den Blick zu bekommen. Es muss dann eine Analyse des Einzelfalls erfolgen. - Es muss zunächst definiert werden, in welchen Fällen überhaupt ein Handlungsbedarf besteht. Bei der Entwicklung einer solchen Definition ist neben der ausschlaggebenden Retrievalrelevanz und dem jeweils gültigen Korrekturrecht die Außendarstellung der Bibliothek zu berücksichtigen. So sind z.B. Verwechslungen von Groß-/Kleinschreibung, die Verwechslung von „ß" und „ss", oder die fälschliche Auflösung von Umlauten („Universitaet") bzw. fälschliche Fusionierung zu Umlauten („Göthe") nicht retrievalrelevant; Tippfehler wie „Göthe" statt „Goethe" im HST ... sind aber m.E. nicht „nur" ein optisches Problem! Abgesehen von besonders groben Tippfehlern und solchen, die sich by the way korrigieren lassen, sehe ich hier im allgemeinen keinen Handlungsbedarf. Bei retrievalrelevanten Tippfehlern muss je nach vorliegendem Fall unterschiedlich vorgegangen werden.

Die einfachsten Fälle sind die, in denen Elemente der bibliographischen Beschreibung zur Disposition stehen, die nicht retrievalrelevant sind; das bezieht sich bspw. auf Dissertationsvermerke, wenn statt „Univ." „Unvi." oder ein Dreher in einem Städtenamen etc. eingegeben worden ist: Die Sorgfaltspflicht ist nicht verletzt, wenn in einem solchen Fall eine „blinde Korrektur" durchgeführt wird. Die Sorgfaltspflicht ist aber m.E. in einem ganz anderen Sinne auch dann nicht verletzt, wenn derlei Korrekturen unterbleiben ...

Die nächste Schwierigkeitsstufe liegt vor, wenn offensichtliche Tippfehler (z.B. „economy", „Steuerbeatung", „Allemange", „Fianzflüsse", „intoduction", „economics", „inTaiwanese", „Anlayse", „arifact", „managment", „Procceedings" oder „Organisastion")⁵ vorliegen, aber nicht - obwohl dieser Sachverhalt ja pro-

5 N.b.: Alle hier beigeispielten Tippfehler sind wirklich ermittelt, keiner ist ausgedacht!

tokolliert werden sollte - ohne weiteres bestimmbar ist, ob diese vorlagengetreu zustande gekommen sind. Schließlich gibt es „scheinbare Schreibfehler, die keine sind, weil die Schreibung nun mal so auf dem Titelblatt steht“ (vgl. Sie bitte *Eversberg 2000*): Ein Hauptsachtitel darf nicht stillschweigend korrigiert werden. In diesen Fällen hilft oftmals eine Analyse der gesamten Titelaufnahme⁶ oder ein Vergleich mit anderen Bibliothekskatalogen - wobei insofern Vorsicht geboten ist, als dass ein irgendwo zugrunde gelegter Fehler durch nicht hinreichend kontrollierte Fremddatenübernahme auch in die zum Vergleich herangezogenen Bibliothekskataloge kopiert worden sein kann. Zudem kann dort aus irgendeinem Grunde genau der gleiche Tippfehler gemacht worden sein: Identische Tippfehler können, müssen aber nicht das Ergebnis von Kopierprozessen sein, sie können unabhängig von einander entstanden sein. Das leitet über zur nächste Stufe:

Die höchste Schwierigkeitsstufe in diesem Zusammenhang liegt dann vor, wenn vergleichend hinzugezogene Bibliothekskataloge entweder keine (hinreichenden) oder nur widersprüchliche Informationen geben. Es bleibt dann nur die Autopsie.

Ich hatte eingangs darauf hingewiesen, dass mit intensivierter Kontrolle der Titelaufnahmen die Fehlerquote sinkt. Infolgedessen ist es erwartbar, dass bei Nutzung von Fremddaten die Fehlerquote sinkt. Dieser Erwartung entsprechen die Erfahrungen der UB Frankfurt (Oder). Da aber auch diese Titelaufnahmen erfahrungsgemäß nicht 100%ig fehlerfrei sind, ist es wünschenswert, wenn die Lieferanten von Titelaufnahmen in ihren Datenbanken eine einfache Möglichkeit für diesbezügliche Rückmeldungen schaffen.

Ein Letztes: Ich halte es für wünschenswert und machbar, wenn eine Bibliothekssoftware bestimmte Elemente der Titelaufnahmen auf Tippfehler untersucht und fordere die Produzenten dieser Software auf, diesbezügliche Schritte einzuleiten.

Literatur

- Ballard, Terry* (1992), Spelling and typographical errors in library databases - In: *Computers in Libraries* 12, no. 6, S. 14-19
Cahn, Pamela (1994), Testing database quality - In: *Database*, 17, no. 1, p. 23-30

6 Eine solche Analyse der gesamten Titelaufnahme kann auch hilfreich in den Fällen sein, in denen nicht einzelne Wörter, sondern deren Zusammenspiel zur Disposition steht. Das o.a. Beispiel mit "Salier und Ottomanen" bzw. "Salier und Ottonen" ließe sich klären, wenn weitere Elemente der Titelaufnahme Schlüsse auf bestimmte geographische, kulturelle, politische ... Bestimmungen zuließen oder den Namen einer Person enthielten, die zweifelsfrei den Ottonen zuzurechnen ist.

- Dwyer, Jim* (1991), The catalogers "invisible college" at work: The case of dirty database test - In: Cataloging and classification quarterly 14, no. 1: 75-82
- Eversberg, Bernhard* (2000), Rechtschreibkontrollprogramm für Daten. - Brief an den Autor - INETBIB, 15.8.2000
- Hauer, Manfred* (2000), Automatische Indexierung. - In: Wissen in Aktion : Wege des Knowledge-Managements ; proceedings / 22. Online-Tagung der DGI, Frankfurt am Main, 2. bis 4. Mai 2000. Hrsg. von Ralph Schmidt. DGI, Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis. - Frankfurt am Main : DGI, 2000 (Tagungen der Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis ; 2) - s.a. <http://www.agimc.de/internet.nsf/251abf1ec3880f42c12567f4004228d8/c3baa4b3dbcaf2c5c12568d80064cdce?OpenDocument&Highlight=2,lotus>
- Knorz, Gerhard* (1999), Automatische Indexierung. - http://www.iud.fh-darmstadt.de/iud/wwwmeth/publ/skript/atuin_d94/paper1.htm
auch in: Hennings, R.-D.; Knorz, G.; Manecke, H.-J.; Reinicke, W.; Schwandt, J.: Wissensrepräsentation und Information Retrieval. Universität - Potsdam, Informationswissenschaft, Modellversuch BETID, Lehrmaterialien Nr. 3, Mai 1994; s.a. http://www.fujitsu-siemens.com/servers/passat/passa_de.htm
- Kuhn, Heinrich C.* (2000), Wie schaden Schlechte Katalogisate? - Vorgetragen: XXIII. Bibliotheksleiter/innenTagung der Max-Planck-Gesellschaft (München, 9. Mai 2000) - Version: Vortragsvorlage, ohne die extemporiert vorgebrachten Ergänzungen - <http://www.phil-hum-ren.uni-muenchen.de/php/Kuhn/ePub/hckBL2000MPG.htm> am 10.5.01
- Randall, Nichols Barbara* (1999), Spelling errors in the database: shadow or substance? - In: Library resources & technical services (LRTS), 43 (1999), S. 161 - 169
- Schulz, Ursula* (1994), Was wir über OPAC-Nutzer wissen : fehlertolerante Suchprozesse in OPACs. - In: ABI-Technik - Wiesbaden. - 14 (1994), 4, S. 299 - 300 ; 302 - 310
- Toole, Janine* (1999), Categorizing unknown words: a decision tree-based misspelling identifier. - In: Advanced topics in artificial intelligence : proceedings .../ Norman Foo (ed.) (Lecture notes in computer science ; Vol. 1747), S. 122 ff. - Berlin ...: Springer
- Weller, Marjorie S./Drabenstott, Karen M.* (1996), Handling spelling errors in online catalog searches - In: Library resources & technical services (LRTS), - 40 (1996), 2, S. 113 - 132

