

Zeitretrieval in OPACs

Auswirkungen auf Formalerschließung, Datenstrukturierung und Information Retrieval

Vortrag von Martin Fenske auf dem 97. Bibliothekartag in Mannheim
(Informations- und Dokumentationsdienst des Schleswig-Holsteinischen Landtags)

Auch wenn überwiegend von Informatikern zum Zeitretrieval publiziert wird, so handelt es sich dennoch nicht um ein vorrangig technisches Thema. Wenn man bei Zeitraumsuchen im Gemeinsamen Verbundkatalog (GVK) des GBV feststellt, dass im Extremfall bis zu 50 % der gefundenen Titel die gesuchten Jahre nicht enthalten, so ist dies ein bibliothekarisches Problem. In Bibliotheken sollte man sich auch fragen, wie sich die Erschließungsregeln auf Zeitsuchen auswirken und ob regelmäßige Qualitätskontrollen nicht sinnvoll sind.

Während in der Dissertation von Frommeyer (2003) zum Zeitretrieval der Schwerpunkt auf der Inhaltserschließung und der Normierung und Kodierung von Zeitbegriffen liegt, geht es bei diesem Vortrag um die Formalerschließung und die Logik der Zeitsuche. Außerdem wird zum Schluss die Frage gestellt, inwieweit Bibliotheken die Möglichkeiten des Raum-Zeit-Retrievals für sich nutzen können.

Frommeyer (2003) hat gezeigt, dass wir die Bedeutung von Zeitangaben für die Suche nur schwer abschätzen können. Unterschiedliche Zeitangaben spielen aber eine zentrale Rolle bei Datenbanken und Katalogen, z.B. in den Bereichen Erschließung, Arbeitsprozessdokumentation und Steuerung.

Retrievalrelevante Zeitaspekte

Für das Retrieval sind eine Vielzahl von Eigenschaften der Zeit wichtig:

- Zeitrechnung (kalender- und ortsabhängig):
gregorianische oder islamische Zeitrechnung, MEZ oder ...
- Schreibweise: 01.01. oder 1. Januar
- Skalierung: 2008 oder 01.01.08 10 Uhr
- Unschärfe: 2008, ca. 2008, 21. Jahrhundert
- Zeitpunkt oder Zeitraum: 2008 oder 01.01.-31.12.2008
- Bezug: Erscheinungs-, Berichtszeit, Thema
- Wiederkehrende Zeitangaben: Feiertag, 1. Mai, Pfingstmontag
- Universelle Variablen und Konstanten: nie, jetzt, immer; Anfang und Ende des maximal zulässigen Zeitraums

Eine Standardisierung von Zeitangaben erleichtert die Zeitsuche erheblich, führt jedoch auch zu neuen Problemen. Der Wechsel zu einer feineren Skalierung z.B. ist mit einem Zuwachs an Unschärfe verbunden. Eine gröbere Skalierung dagegen ist ohne Zusatzinformationen nicht möglich. Die Einfachheit von Internetsuchmaschinen dagegen führt dazu, dass unterschiedlich strukturierte Zeitangaben bei der Suche "unsichtbar" bleiben und daher auch nicht auf einfache Art gefunden werden können.

Formale Zeitangaben in OPACs

Unsere OPACs kennen als "scharfen" Zeitpunkt vor allem das Erscheinungsjahr. Bei fortlaufenden Sammelwerken werden Zeiträume ebenfalls in Jahren angegeben. Wenn es sich um abgeschlossene Bestände handelt, ist dieser Zeitraum geschlossen (z.B. 2000-2005). Bei noch laufenden Werken ist er offen (z.B. 2000 ff.). Während mit diesen Zeiträumen bei

Zeitschriften auch immer ein bestimmter Bestand dargestellt wird, meint eine vergleichbare Bestandsangabe bei einer Loseblattsammlung nur, dass sie zu einem bestimmten Zeitpunkt begonnen wurde und nun mit dem zuletzt nachgelieferten Stand verfügbar ist. Unschärfe wird bei der Formalerschließung in vielfältiger Form zugelassen (z.B. [vor 1988] oder 20. Jh. bzw. 19XX), nicht ermittelbare Zeitangaben fehlen oder sind entsprechend gekennzeichnet (z.B. o.J.).

Im OPAC erfüllen formale Zeitangaben unterschiedliche Metadatenfunktionen. Sie dienen vor allem der bibliographischen Beschreibung, der Suche und der Sortierung. Auch wenn für alle drei Funktionen im Regelfall nur eine Zeitangabe verwendet wird, so kann im GBV-GVK jeder Funktion ein eigener Wert zugewiesen sein. Dies ermöglicht Differenzierungen, führt aber gleichzeitig auch zu Eingabefehlern und damit zu Inkonsistenzen.

Test der erweiterten Suche des GVK

Um das Zeitretrieval des GBV-GVK, der nach eigenen Angaben über 28 Mio. Titel enthält, zu untersuchen, wurden zwei Tests in der erweiterten Suche durchgeführt. Es wurde nach Titelstichwort und Titelstichwort in Verbindung mit Erscheinungsjahr(en) gesucht. Quantitativ wurden die Trefferzahlen, qualitativ Besonderheiten und Ausreißer ausgewertet. Der erste Test fand im August 2006 statt. Es wurden exemplarisch sechs Suchen mit Endmaskierung durchgeführt, die insgesamt 883 Tsd. Treffer ergaben. 5,12 % dieser Titel waren ohne Jahresangabe. Auch wenn deutlich weniger als 1 % der Titel als Zeitangabe nur eine Jahrhundertangabe hatten, so stellte die Angabe 19. Jh. fast dreiviertel der Treffer, die bei 1900 sortierten. Es zeigt sich damit, dass fehlende Zeitangaben ein Suchproblem und unscharfe Angaben ein Sortierproblem darstellen können. Der zweite Test im Mai 2008 wurde repräsentativer durchgeführt. Es wurde mit allen Buchstaben des Alphabets und den Zahlen 0 bis 9 nacheinander endmaskiert gesucht (s. Abb. 1). Hierbei wurden 143,3 Mio. Treffer analysiert. Mit der Suche nach dem Zeitraum "0000-0000" wurde die Zahl der Treffer ohne Jahr und mit der Suche nach "9999-9999" die Zahl der offenen Zeiträume bestimmt. Eingabefehler wurden mit Zeiträumen gefunden, die im GBV unsinnig sind (zukünftige Jahre: "2009-9999") oder zu denen keine oder nur sehr wenige Treffer zu erwarten sind ("0000-1450").

The screenshot shows the search interface of the GBV-GVK. At the top, there are navigation links: Suchen, Suchergebnis, Erweiterte Suche, Zwischenablage, Benutzerkonto, and Hilfe. Below this is the GBV-GVK logo and the user's session information: Benutzername: GAST | Abmelden | Datenbanktest. The main section is titled 'Suchfilter | Sachgebiete'. Under 'Suchfilter', there is a 'Suchen' button and a prompt: 'Füllen Sie das Formular aus, stellen Sie gegebenenfalls weitere Optionen ein und starten die Suche mit einem Klick auf die Schaltfläche Suche'. The search criteria are: Titel (Stichwort) [TIT] with the value 'a?', Person, Autor [PER], Schlagwörter [SLW], and Ort, Verlag (Stichwort) [PUB]. The sorting options are: sortiert nach: Erscheinungsjahr, Sprache: Alle Sprachen, and Land: Alle Länder. The appearance year is set to '0000-0000' with a note: 'z.B.: 1948-1987, 1996-, 1965, 1984, 2001 oder 2003'. There is also an unchecked checkbox for 'Unschärfe Suche'.

Abb. 1: Erweiterte Suche des Gesamtverbundkatalogs des GBV

Die Logik der GVK-Suche findet mit einem oder mehreren durch Komma getrennten Jahren nur einzelne Erscheinungsjahre und keine Zeiträume. Wird wie in den bibliographischen Beschreibungsdaten mit einer Mischung aus Zahlen und Text (19. Jh.) oder mit Text (o.J.) gesucht, so werden nur vorangestellte Zahlen als Jahr gefunden (statt 19. Jh. 19). Bei einer reinen Text-Suche werden nur sehr wenige Treffer, vermutlich Eingabefehler, gefunden. Die Zeitraumsuche findet neben den Titeln mit "normalen" Erscheinungsjahren auch solche mit offenen und geschlossenen Zeiträumen zuverlässig. Das Suchergebnis wird aber auch noch um Titel ohne Jahr automatisch ergänzt.

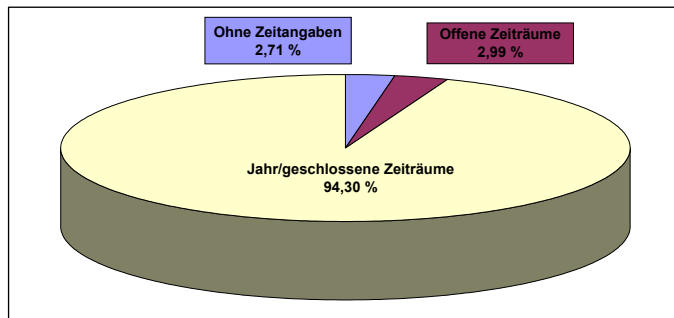


Abb. 2: Ergebnisse im GVK-Gesamtbestand

Im Durchschnitt fanden alle Suchen zu etwa 94 % Titel mit Erscheinungsjahr oder geschlossenem Zeitraum. Titel ohne Zeitangaben und Titel mit offenem Zeitraum waren mit jeweils unter 3 % nur von geringer Bedeutung (s. Abb. 2). Dies dürfte in etwa auch den Verhältnissen im Gesamtbestand

entsprechen. Sucht man jedoch mit dem kurzen Zeitraum "2008-2008" so werden diese Proportionen auf den Kopf gestellt: Erscheinungsjahre und geschlossene Zeiträume stellen nur noch weniger als 5 % der Titel, offene Zeiträume liegen bei fast 50 % und Titel ohne Zeitangabe kommen auf über 45 % (s. Abb. 3). Das Ergebnis einer Zeitraumsuche repräsentiert daher vor allem die Suchlogik und nicht so sehr den Gesamtbestand des GBV, da im GVK ohne direkten Zusammenhang zu den verwendeten Suchwerten unterschiedliche Bestandssegmente miteinander

gemischt werden.

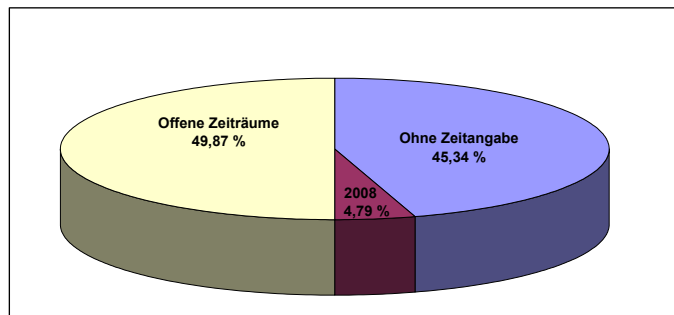


Abb. 3: Ergebnis einer GVK-Suche mit „2008-2008“

Diskussion der Ergebnisse

Es ist bemerkenswert, dass eine in der Internethilfe nicht näher erläuterte Suchsyntax darüber entscheidet, ob ein unvollständiges (Jahressuche findet nur identische Jahre ohne Zeiträume) oder ein mit Treffern ohne Jahr "aufgeblähtes" Suchergebnis (Zeitraumsuche)

gefunden wird. Die nahe liegendere Erwartung, nur solche Treffer zu finden, deren Zeitangaben sich mit den gesuchten Zeitangaben überschneiden, wird nie erfüllt. Die Konzeption der Zeitraumsuche macht nur Sinn, wenn man bedenkt, dass bei der Anzeige des GVK die chronologische Sortierung voreingestellt ist und Treffer ohne Jahr am Ende sortieren. Da voreingestellte Sortierungen selten geändert und lange Listen nicht bis zum Ende durchgeblättert werden, stört der vermeintliche Informationsballast den Nutzer nicht. Gleichzeitig werden Informationsverluste durch Eingabefehler und Frustrationen durch 0-Treffer-Ergebnisse vermieden.

Wir stellen also fest, dass Titel ohne Jahr bei einer Zeitsuche, obwohl dies den meisten Nutzern bewusst ist, entweder nie oder immer gefunden werden. Der Anfang der GBV-Zeitangaben liegt mit dem Jahr 0 bei Christi Geburt, ohne dass mit dieser Zeitangabe Titel jener Zeit nachgewiesen würden. Das Ende der GBV-Zeitangaben liegt im Jahr 9999, also in einer noch nicht vorstellbaren Zukunft, obwohl Erwartungsdaten eigentlich gar nicht im GVK verwaltet werden. Dass bei OPACs Anfang und Ende der Zeitsuchen eigentlich unsinnig definiert werden, wirkt sich nur deshalb nicht weiter aus, weil kein normaler Nutzer im OPAC nach diesen Jahren suchen wird. Wenn wir nun aber in einem historischen Portal OPAC-Daten mit historischen Informationen und, wie von Frommeyer (2003) vorgeschlagen, formale und inhaltliche Zeitangaben mischen, so kann dies durchaus zu Problemen führen.

Zur Verbesserung der Datenqualität und der Datenstruktur bieten sich folgende Maßnahmen an:

- fehlende Zeitangaben werden ergänzt (Qualitätskontrolle) oder durch unscharfe Zeiträume ersetzt,
- unscharfe Zeitangaben werden enger gefasst bzw. als geschätzter Zeitraum angegeben ([vor 1988] => z.B. 1985-1987),
- nur wirkliche Mehr-Jahres-Bestände werden als Zeitraum abgebildet, indem Loseblattsammlungen nur noch ein Erscheinungsjahr erhalten, und zwar jeweils das Jahr der letzten Lieferung (1993 ff.: in 2008 mit 2008 statt 1993-9999), und
- regelmäßige Qualitätskontrollen werden durchgeführt, um gerade auch systematische Erfassungsfehler zu finden, im GVK z.B.:
 - o "P 2000" statt 2000 als Jahr 20,
 - o Islamisch-gregorianische Zeitumrechnung: 1429 [2008] statt 2008 (gregorian.) Jahr 1429 (islam.) und
 - o Erscheinungsjahre formatiert nur im Titel (...||Anno 1630||... s.a.).

OPAC der Landtagsbibliothek

Der Bibliothek des Schleswig-Holsteinischen Landtags hat in ihrem OPAC viele Gemeinsamkeiten und einige Unterschiede zum GVK. Wir haben zusammen mit der GLOMAS Deutschland GmbH auf der Basis der Software STAR ein eigenes Bibliothekssystem entwickelt, um unsere Bibliotheksdaten in die lokale Informationsinfrastruktur besser einbinden und stärker auf unsere lokalen Arbeitsprozesse eingehen zu können. Das Bibliothekssystem ist einer von drei Bereichen des Landtagsinformationssystems Schleswig-Holstein, dessen Konzept bereits von Fenske (2002) beschrieben wurde.

Das Bibliothekssystem besteht aus mehreren Modulen:

- Web-OPAC integriert
 - o im Internet mit Personen und Parlamentsdokumenten,
 - o im Intranet mit einer Regierungsbibliothek,
- Erwerbung, Katalogisierung, Thesaurus,
- Ausleihe,
- Fortsetzungs- und Zugangsverwaltung sowie Haushaltsüberwachung.

Der Internetnutzer kann z.B. über die einfache Portalsuche des Landtagsinformationssystems mit einem Stichwort und einem Zeitraum nach Literatur, Personen und Parlamentsdokumenten suchen (s. Abb. 4 und 5). Das Portal sucht nach Jahren, dem kleinsten gemeinsamen Nenner von Datumsangaben und Erscheinungsjahren. Zeitangaben werden für jeden Bestand unterschiedlich interpretiert. Dadurch werden Probleme einer gemeinsamen Sortierung von Jahr und Datum vermieden. Über einen Link gelangt der Nutzer zur chronologischen Liste der Bibliothekstitel. Im Gegensatz zum GVK gibt es bei uns nur eine Zeitraumsuche. Titel ohne Jahr werden nur gefunden, wenn auch ohne Jahr gesucht wurde.

LIS-SH Landtagsinformationssystem Schleswig-Holstein
Suche im Gesamtbestand

Home Service Infos Hilfe

Suchbegriffe arens

alle Begriffe mind. ein Begriff Begriffe in dieser Reihenfolge

Jahr 1995 bis 2008 Suche

Spezielle Suchen:

- [Parlamentsdokumente und Beratungsvorgänge](#)
- [Amts- und Mandatsträger](#)
- [Literatur \(Landtagsbibliothek\)](#)

Abb. 4: Portalsuche des Landtagsinformationssystems

Hilfe Service Infos

Trefferübersicht

9 Literatur (Landtagsbibliothek)

1 Amts- und Mandatsträger

20 Parlamentsdokumente und Beratungsvorgänge

Abb. 5: Trefferübersicht der LIS-SH-Portalsuche

Die Erfassung der Zeitangaben sollte in unserem Bibliothekssystem einfach sein und dennoch eine ausreichende Abbildung von Zeiträumen erlauben. Mit dem Erscheinungsjahr kann ein Jahr erfasst werden, das zugleich für Anzeige, Suche und Sortierung verwendet wird. Alternativ gibt es aber noch eine sog. pauschale Bestandsangabe. Sie besteht aus einem Textfeld für die Anzeige und zwei Feldern für den Anfang und das Ende eines Zeitraums. Sortiert wird nach dem Anfang des Zeitraums; gesucht werden kann nach allen Jahren eines Zeitraums. Auch offene Zeiträume und unscharfe Zeitangaben sind möglich (s. Abb. 6). Künftig soll bei laufend aktualisierten Loseblattsammlungen das jeweils aktuelle Jahr halbautomatisch als Von-Jahr ergänzt werden. Damit sortieren für unseren Bestand wichtige laufende Loseblattsammlungen am Anfang.

The screenshot shows the STAR Client interface with the following data entered:

Titel			
Personen	Titel/Exemplar	Eingabewerte	
Titeltyp	Medienart	Bearbeitungsnachricht	
2	Zeitschrift		
Titel			
Archiv der Gegenwart : AdG			
Zusatz zum Titel			
Die weltweite Dokumentation für Politik und Wirtschaft			
Ort	Verlag	ISSN/ISSN	
Königswinter	Siegler	ISSN	0003-8865
Erscheinungsjahr	Pauschale Bestandsangabe		
	von	bis	
	15.1946-72.2002	1945	2002

Abb. 6: Eingabe von Zeitangaben als pauschale Bestandsangabe

Weiterentwicklung zu einem Raum-Zeit-Retrieval

Wir wollen uns nun von den Alltagsproblemen der OPAC-Suche wieder entfernen, um der Frage nachzugehen, wie sich Raumsuchen im Zeitalter von GPS und Navigationsgeräten auf unsere Bibliothek und den OPAC auswirken können.

Während wir es bei der Zeit mit einer Dimension zu tun haben, bewegt man sich im Raum in drei Dimensionen. Mit GPS leben wir dagegen nur in einer zweidimensionalen Welt. Die Längen- und Breitengrade unserer Koordinaten können in Dezimalwerte umgewandelt und damit leichter bearbeitet werden. Unsere Raum-Zeit-Suche ist also multidimensional und damit technisch etwas anspruchsvoller. Das Gegenstück zur Zeitvariable "jetzt", die nach belieben als Datum mit Uhrzeit oder z.B. umgerechnet in einen Wochentag ausgegeben werden kann, ist die Raumvariable "hier", die wir über das GPS abfragen können.

Während wir im klassischen Retrieval Schnittmengen bilden (Zeitraumsuche) oder nach identischen Werten suchen (GVK-Jahressuche), bietet das Raumretrieval weitere Operatoren an, die jedoch bei eindimensionalen Zeiträumen technisch wesentlich einfacher sind. Wir können einen Bestandteil oder zu einem Teil das Ganze suchen (z.B. Baden-Württemberg und Deutschland). Das Umgebene und das Umgebende lassen sich ebenfalls unterscheiden (z.B. ein See und sein Ufer). Häufig wird bei der Navigation der kürzeste bzw. der schnellste Weg zwischen zwei Orten gewünscht, von denen ein Ort mobil ist.

Welche Bedeutung hat dies nur für das Zeitretrieval? - Bei einer Suche nach Literatur zu den 1930er Jahren vermuten wir bei einer exakten Übereinstimmung mit dem gesuchten Zeitraum eine Gesamtdarstellung, bei einem Teil des Zeitraums eine Detaildarstellung und bei umgebenden Zeiträumen Abhandlungen, die auch noch einen Bezug zu den beginnenden oder auslaufenden 1930er Jahren haben könnten. Für das Retrieval erlauben entsprechende Operatoren eine wesentlich exaktere Suche als die einfache Bildung von Schnittmengen. Eine allgemein gehaltene Weltgeschichte hätte sicher auch eine Schnittmenge zu den 1930er Jahren, würde unserem Nutzer aber nur wenige und sehr allgemeine Informationen zu diesem Thema bieten. An einer Vielzahl von Operatoren werden jedoch nur spezialisierte Informationsprofis

interessiert sein. Wichtiger ist daher die Nutzung derartiger Operatoren für automatisierte Dienstleistungen (z.B. Linkliste oder Push-Dienst).

Die schlichte Realisierung einer Raumsuche gibt es in vielen kommerziellen Internetportalen als Postleitzahlen- oder Ortsnamen-Suche. Der Internetnutzer muss die Eingabe seines Standorts ("hier") selbst vornehmen. Ausgehend vom "hier" kann meist auch eine maximale Entfernung für die Suche angegeben werden.

Attraktiver ist dagegen eine Raum-Zeit-bezogene Dienstleistung, die über Handy und Navigationsgerät realisiert wird. Das Navigationsgerät liefert die wichtigen Variablen "hier" (Mannheim, CC Rosengarten) und "jetzt" (3. Juni 2008 = Dienstag) unseres Kunden, der gerade zwischen den Vorträgen des Bibliothekartags unterwegs ist. Der Kunde sucht touristische Informationen aus der Umgebung (ca. 20 km) und möchte diese Information wie immer möglichst schnell (Weg+Zeit) haben. Aktualität und Preis sind erst in zweiter Hinsicht für ihn von Bedeutung. Ihm werden daher die Touristeninformation, Buchhandlungen, öffentliche Bibliotheken und Internetadressen sortiert nach seinem persönlichen Profil angeboten. Hierbei wird auch geprüft, ob die Bibliothek geöffnet hat und das Buch innerhalb der vorgegebenen Zeit ausgeliehen werden kann.

Ein derartiges Angebot erfordert eine umfassende Vorarbeit durch die Informationsanbieter:

- Das Gesamtangebot aller Anbieter muss attraktiv genug sein,
- einheitliche numerische Orts- und Zeitangaben müssen kontextfrei genutzt werden können,
- Informationen zur Lage, Öffnungs- und Bereitstellungszeiten sowie Kosten der Institution und der jeweiligen Bestandseinheit müssen genauso verfügbar sein wie die üblichen Metadaten zur Medieneinheit (Ausleiddaten, Thema, ...).

Quellen

1. Fenske, Martin (2002): Integration der Datenbanken des Schleswig-Holsteinischen Landtages zu einem Wissensmanagementsystem.- In: Content in Context – Perspektiven der Informationsdienstleistungen: Proceedings. 24. Online-Tagung der DGI. Frankfurt a.M., 3.-5. Juni 2002. Hrsg.: Schmidt, Ralph. Frankfurt a.M.: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis. 2002. S. 99-108.
2. Frommeyer, Jutta (2003): Zeitbegriffe und Zeitcodierungen in allgemeinbibliographischen Datenbanken. Analyse von RSWK, LCSH und RAMEAU sowie Entwicklung eines theoretischen Modells für ein Zeitretrieval. Berlin: Logos-Verl. (Berliner Arbeiten zur Bibliothekswissenschaft; Bd. 10)
Zugl.: Berlin, Humboldt-Univ., Diss. 2002