Mario Kubek

Bibliotheksdienste in dezentralen Websuchsystemen

Library Services in Decentralized Web Search Systems

<https://doi.org/10.1515/bd-2022-0022>

Zusammenfassung: Stellt man sich das World Wide Web als riesige Bibliothek vor, bräuchte es einen Bibliothekar oder zumindest Dienste, die die vielfältigen Aufgaben dieser Person in vergleichbarer Weise und Qualität erfüllen können. Aktuelle Websuchmaschinen können diese Aufgaben wie die Katalogisierung von Publikationen sowie die Vermittlung zwischen bibliothekarischen Ressourcen und Nutzern nicht einmal annähernd zufriedenstellend erfüllen. Daher können Bibliothekare insbesondere bei langfristigen und tiefgehenden Recherchen eine viel bessere Unterstützung bieten. Dieser Artikel diskutiert die vielen Vorteile der Dienstleistungen von Bibliotheken und Bibliothekaren und erläutert, wie diese in neuartigen und dezentral organisierten Websuchsystemen realisiert werden können, um Nutzer bei Forschungsaufgaben nachhaltig zu unterstützen.

Schlüsselwörter: Bibliotheksdienste, dezentrale Websuche, WebEngine

Abstract: Imagining the World Wide Web as a giant library implies the proficiency of librarians, or at least qualified library services that accomplish the varied tasks assigned to librarians. Existing web search engines are nowhere near fulfilling such tasks and services as cataloguing publications and mediating between library resources and users in a satisfactory manner. Professional librarians could provide much better support especially in long-term and in-depth research projects. This article discusses the advantages and benefits of library services and the assistance of librarians in research, and illustrates how new, decentrally organ-

Article Note: Aus Gründen der besseren Lesbarkeit wird in diesem Beitrag ausschließlich die männliche Schreibweise verwendet. Sämtliche Personenbezeichnungen gelten gleichwohl für alle Geschlechter.

PD Dr.-Ing. Mario Kubek: mario.kubek@fernuni-hagen.de

ized web search systems can deal with these complex tasks to provide effective and long-term support for users doing scientific research.

Keywords: Library services, decentralized web search, web engine

1 Motivation

Öffentliche Bibliotheken sind heutzutage oft einsame Orte, denn die benötigten Informationen sind häufig leicht aus dem allgegenwärtigen Internet und insbesondere aus dem World Wide Web (WWW, Web) als einem großen Teil davon abrufbar. Vergessen scheinen die Zeiten zu sein, in denen Bibliothekare große Mengen an Büchern sammelten, sie in Katalogkästen mit Tausenden von Katalogkarten wiederauffindbar machten und nach einem (eigenen) speziellen Schema am richtigen Ort in einem Labyrinth von Regalen archivierten. Neben diesen Aufgaben und der Ausleihe von Büchern hatten sie auch noch Zeit, die Bibliotheksnutzer zu unterstützen, indem sie ihnen Tipps gaben, wo sie die gesuchten Informationen schnell finden konnten, und ihnen vielleicht sogar die neuesten Nachrichten und Trends mitteilten.

Derzeit gibt es jedoch kein Informationssystem, das all diese Aufgaben auch nur annähernd in akzeptabler Weise erfüllen kann. Da jedoch die Menge der verfügbaren Textdaten (insbesondere im WWW) stetig wächst und das weltweite Datenverkehrsvolumen für die private Webnutzung und das Versenden von E-Mails auf 27 Exabyte¹ pro Monat im Jahr 2021 prognostiziert wird², besteht ein dringender Bedarf an einem solchen System, das – ähnlich wie der menschliche Bibliothekar in seiner Rolle als Ansprechperson – als aktiver technischer Vermittler zwischen Nutzern und Ressourcen wie Textdokumenten angesehen werden kann. Ein dementsprechend als Intermediär konzipiertes Informationssystem muss daher in der Lage sein, selbstständig

- Informationen in zahlreichen Formaten bereitzustellen, zu archivieren und zu verwalten,
- einen effizienten Informationszugang zu bieten (z. B. Themenvorschläge zu unterbreiten, als „thematischer Wegweiser“ zu agieren sowie Bibliografien zu erstellen),

¹ Ein Exabyte entspricht 10^{18} oder einer Trillion Bytes oder einer Milliarde Gigabyte. Siehe auch die Vergleichstabelle in Wikipedia, <https://de.wikipedia.org/wiki/Byte#Vergleichstabelle> [Anmerkung der Redaktion, Zugriff: 14.12.2021].

² Statista 2021.

- proaktiv Informationen auf der Grundlage des ermittelten Informationsbedarfs (dies ist ein Nachweis von Informationskompetenz³, d. h. „die Fähigkeit, zu erkennen, wann ein Informationsbedarf besteht, sowie die benötigten Informationen zu identifizieren, zu lokalisieren, zu bewerten und effektiv für das jeweilige Thema oder Problem zu nutzen“) zu beschaffen und
- Suchaufgaben durchzuführen und dabei unwichtige oder gar unerwünschte Informationen (man könnte hier auch einfach von Daten sprechen) durch die Anwendung von Klassifizierungsmethoden herauszufiltern.

Dies bedeutet implizit, dass das System in der Lage sein muss, Suchaufgaben selbstständig und im Auftrag des Nutzers durchzuführen, wenn dies gewünscht wird. Ein System, das diese Funktionen bietet, würde vor allem tiefgehende Recherchen nachhaltig erleichtern, die sich stark von kurzfristigen oder Adhoc-Suchaufgaben unterscheiden. Solch eine Tiefenrecherche

- ist ein iterativer und interaktiver Prozess,
- hat einen Kontext und eine Historie,
- besteht aus verschiedenen Suchpfaden und -richtungen,
- bedeutet, aus positiven und negativen Rückmeldungen zu lernen und
- beeinflusst auch die gesuchten Objekte selbst (ein Beispiel: die von Experten auf einem Gebiet am häufigsten angefragten Objekte wären bei entsprechend kategorisierten Anfragen wahrscheinlich von Bedeutung, würden daher zuerst zurückgegeben und wären somit leichter Gegenstand weiterer wissenschaftlicher Untersuchungen).

Diese Punkte implizieren aber auch, dass das erwähnte Informationssystem in der Lage ist, mit sich dynamisch verändernden Kontexten wie (auch kurzfristigen) Verschiebungen im Informationsbedarf und thematischen Veränderungen in der lokalen Dokumentenbasis umzugehen und daraus zu lernen sowie möglicherweise aufkommende neue Trends oder neue Konzepte aus verschiedenen Informationsströmen zu identifizieren. Wenn das System die Historie vergangener und laufender Suchprozesse in Form von Suchpfaden, bestehend aus Anfragen und Ergebnismengen, berücksichtigt, ist eine Navigation in früheren Suchschritten möglich. Auf diese Weise und basierend auf den gelernten Konzepten und deren Beziehungen kann das System auch interaktiv alternative Suchrichtungen sowie weitere passende Themen zur Vertiefung vorschlagen.

Diese Funktionalitäten sind vor allem für Nutzer, die intensive Forschung betreiben, von Vorteil. Sie adressieren aber auch das Problem des Wiederauffin-

³ A. L. A. P. Committee 1989.

dens von Informationen, das durch den so genannten „Google-Effekt“, manchmal auch als „digitale Amnesie“ bezeichnet⁴, verschärft wird. Die wichtigsten Erkenntnisse im Zusammenhang mit diesem Effekt sind, dass Menschen dazu neigen, Informationen zu vergessen, wenn sie davon ausgehen, dass sie mithilfe digitaler Technologie leicht wiedergefunden werden können, und dass sie sich eher daran erinnern, wie sie eine bestimmte Information zuvor mithilfe von Suchmaschinen gefunden haben (den Suchpfad), als an die Information selbst. Dies deutet darauf hin, dass Menschen im Allgemeinen mit den erhaltenen Suchergebnissen zufrieden sind (die relevanten werden zuerst präsentiert); andernfalls würde das Gefühl aufkommen, dass es schwierig sein könnte, sie zu einem späteren Zeitpunkt wiederzufinden, und dass es vorteilhafter sein könnte, sich die entsprechenden Informationen selbst zu merken. Da diese Entwicklung – zumindest bis zu einem gewissen Grad – die Art und Weise beeinflusst, wie heute recherchiert wird⁵, sollten die Funktionalitäten des genannten Systems, Suchpfade zu speichern, abzurufen und vorzuschlagen, auch die Wiederherstellung früherer Suchprozesse und deren Ergebnisse erleichtern.

Der folgende Abschnitt gibt einen Überblick über die zahlreichen Dienstleistungen, die von Bibliotheken angeboten werden, wobei der Schwerpunkt auf den Tätigkeiten der Bibliothekare liegt. Anschließend wird erörtert, wie diese Dienstleistungen durch technische Mittel wie Katalogisierungs- und Informationssysteme unterstützt und erweitert werden. Darauf wird analysiert, welche dieser Tätigkeiten und Dienstleistungen in modernen Websuchsystemen, die sich an den Arbeitsprinzipien von Bibliothekaren orientieren und daher dezentral organisiert sein sollten, technisch realisiert werden können und müssen. Abschließend wird die erste technische Realisierung eines solchen dezentralen und voll integrierten Websuchsystems, genannt „WebEngine“, vorgestellt.

2 Bibliotheksdienste

2.1 Die Aufgaben eines Bibliothekars

Unter einer Bibliothek versteht man gemeinhin eine Sammlung von Büchern oder ein Gebäude, in dem diese gelagert und gepflegt werden. Dieses Verständnis ist zwar im Allgemeinen richtig, gleichzeitig aber auch etwas eingeschränkt.

⁴ Sparrow/Liu/Wegner 2011.

⁵ Varshney 2012.

Es gibt eine große Anzahl von Dienstleistungen, die Bibliotheken anbieten. Die wichtigste Funktion von Bibliotheken ist dabei die Versorgung der Öffentlichkeit und von Institutionen mit Informationen⁶. Um dies tun zu können, sammeln, katalogisieren und stellen sie veröffentlichte Literatur in Form verschiedener Medien zur Verfügung. Dazu zählen Bücher, Zeitschriften, Zeitungen und digitale Speichermedien wie CDs und DVDs. Auch in elektronischer Form wird Literatur bereitgestellt. Hierzu zählen etwa E-Books und Hörbücher. Der Zugang zu der von den Bibliotheken zur Verfügung gestellten Literatur ist offen, uneingeschränkt und wird in der Regel kostenlos oder gegen eine angemessene Gebühr gewährt.

Darüber hinaus ist die Archivierung dieser Ressourcen eine weitere wichtige Aufgabe von Bibliotheken, die den Fortbestand literarischer Werke sichert. Sie umfasst in der Regel zusätzliche Aufgaben, insbesondere bei der Bewahrung von Buchbeständen. So müssen diese zum Beispiel (bei Bedarf) ordnungsgemäß restauriert und besonders gepflegt werden (z. B. Auswahl eines trockenen Lagerorts mit konstanten Temperatur- und Luftfeuchtigkeitsverhältnissen). Auch ihre Digitalisierung kann Teil eines Archivierungsprozesses sein, um die Inhalte in elektronischer Form durchsuchbar und leicht übertragbar zu machen. Diese Aufgaben werden in der Regel von Bibliothekaren und Archivaren durchgeführt, je nach ihrer Spezialisierung.

Die Tätigkeiten von Bibliothekaren lassen sich grob in sammlungs- und nutzerbezogene Tätigkeiten einteilen. Während sich die sammlungsbezogenen Tätigkeiten auf die Verwaltung der gesammelten Medien beziehen und somit

- die Auswahl,
- den Erwerb,
- die Verarbeitung,
- Katalogisierung,
- Pflege und
- Archivierung

der Medien umfassen, beziehen sich die nutzerzentrierten Aktivitäten auf

- die Bereitstellung von Informationen,
- die Erteilung von Ratschlägen und Hinweisen,
- die Organisation und Durchführung von Schulungskursen,
- den Verleih von Medien und
- die Bestandsaufnahme (einschließlich der Verwaltung der örtlichen und interbibliothekarischen Buchausleihe, welche als Fernleihe bekannt ist).

6 Gantert 2016.

In den folgenden Unterabschnitten werden die beiden wichtigsten Tätigkeiten von Bibliothekaren im Detail diskutiert.

2.2 Bibliothekare als Vermittler

Da der Informationsbedarf ständig steigt, besteht die wichtigste Aufgabe der Bibliothekare darin, zwischen den anfragenden Kunden und der geeigneten Literatur sowie den Informationen, die ihren Informationsbedarf decken, zu vermitteln. In dieser Hinsicht werden sie zu aktiven Vermittlern in einem Such- oder Rechercheprozess. Daher müssen sie in der Lage sein, die Bibliotheksnutzer dahingehend richtig anzuleiten, wie sie in der Bibliothek relevante Informationen zum jeweiligen Fachgebiet finden können (Standort der Literatur).

Diese Rolle als „Wissensvermittler“ wird im digitalen Zeitalter⁷ noch wichtiger, da sie nicht nur eine solide Ausbildung und gute Kommunikationsfähigkeiten voraussetzt, sondern auch die Fähigkeit, mit Informationstechnologie umzugehen und entsprechende Werkzeuge für das Datenmanagement und die Datenmanipulation zu nutzen, einschließt. Diese sich verändernde Rolle ermöglicht eine aktive und stärkere Beteiligung an Forschungsprozessen und ist daher für den Beruf des Bibliothekars als solchen in Zukunft von besonderer Bedeutung. Spezialisierte Bibliothekare wie die so genannten Teaching Librarians halten sogar Vorträge über Informationskompetenz⁸.

Zusammenfassend lässt sich sagen, dass die Bereitstellung von Informationen und Informationsquellen aus Sicht der Bibliotheksnutzer die wichtigste Dienstleistung der Bibliothekare ist. Aus diesem Grund sollte die Qualität dieser Dienstleistung anhand der folgenden fünf Indikatoren gemessen werden⁹:

- Ist der Informationsschalter sichtbar und leicht zugänglich?
- Zeigt der Bibliothekar Interesse an der Anfrage des Nutzers?
- Hört der Bibliothekar dem Nutzer aufmerksam zu und fragt bei Bedarf offen nach?
- Nutzt der Bibliothekar die richtigen Informationsquellen und die richtige Recherchestrategie (auch mit entsprechenden Erklärungen für den Nutzer)?
- Werden Nachfragen gestellt, um festzustellen, ob der Nutzer die gegebenen Erklärungen tatsächlich verstanden hat?

⁷ Bell 2016.

⁸ A. L. A. P. Committee 1989.

⁹ American Library Association 2013.

Der Bibliothekar muss also freundlich, hilfsbereit, unterstützend und geduldig sein, um einen guten Service zu bieten. Gleichzeitig muss er in der Lage sein, sich unter Wahrung einer professionellen Distanz auf die Bedürfnisse der Nutzer einzustellen und allen Nutzern den gleichen Service zu bieten (ohne bestimmte Personen zu übervorteilen oder andere zu benachteiligen). Neben einer guten Allgemeinbildung, vertieften Sprachkenntnissen und Kommunikationsfähigkeiten muss ein Bibliothekar auch die Fähigkeit besitzen, strukturiert zu denken und sich für moderne Informationstechnologie interessieren.

2.3 Katalogisierung von Medien

Neben der Bearbeitung von Nutzeranfragen sind Bibliothekare in der Regel auch aktiv an der Katalogisierung von Medien beteiligt, die erforderlich ist, um die Bestände der Bibliothek zu erfassen und letztlich auffindbar zu machen. Diese Aktivität steht daher im Mittelpunkt der genannten sammlungsbezogenen Tätigkeiten. Aus historischer Sicht lassen sich Kataloge in Buchform, Zettelkataloge und moderne Online Public Access Catalogues (OPAC), die die beiden erstgenannten Arten weitgehend ersetzt haben, voneinander unterscheiden.

In der Literatur¹⁰ werden im Wesentlichen zwei Arten von Katalogisierungsansätzen unterschieden: die Formalerschließung (meist einfach als Katalogisierung bezeichnet) und die Sacherschließung. Bei der Formalerschließung werden formale Regeln angewandt, um Bücher und andere Medien anhand formaler Elemente wie Autor und Titel zu beschreiben. Diese Elemente ergeben sich aus den Medien selbst und müssen in eine regelkonforme Form gebracht werden¹¹. Ältere Regelwerke für diese Aufgabe sind

- die RAK (Regeln für die alphabetische Katalogisierung),
- die AACR (Anglo-American Cataloguing Rules) und
- die AACR2.

Der neue Standard RDA (Resource Description and Access) für die Formalerschließung, der 2010 eingeführt wurde, ist breiter angelegt und soll neben Bibliotheken auch von Museen und Archiven angewendet werden. Außerdem bietet dieser Regelsatz¹² umfangreiche Richtlinien zur Extraktion von Attributen von Entitäten, wie z. B. einer bestimmten Ausgabe eines Buches, sowie zur Bestimmung ihrer Beziehungen zu anderen Entitäten, um nachgelagerte Anwendungen zu unter-

¹⁰ Gantert 2016.

¹¹ Eberhardt 2012.

¹² <https://access.rdatoolkit.org> [Zugriff: 14.12.2021].

stützen, die auf derart verknüpften Daten beruhen. Das Ergebnis der Formalschließung ist das so genannte Katalogisat, ein Datensatz, der die erfassten Medien eindeutig beschreibt.

Andererseits bedeutet Sacherschließung, Ressourcen auf der Grundlage ihres Inhalts und inhaltsbezogener Kriterien zu beschreiben, ohne sich auf bibliografische oder andere formale Daten zu stützen. Sacherschließung bedeutet also, Inhalte zu interpretieren und erfordert daher implizit Methoden, die Daten in Informationen umwandeln können. Die beiden gebräuchlichsten Methoden hierfür sind die Verschlagwortung und die inhaltliche Klassifikation. Schlüsselwörter für eine Ressource können direkt aus dieser bezogen werden oder durch Rückgriff auf externe Inhalte wie Rezensionen und Anmerkungen, die von Nutzern vergeben werden. Kategorien ermöglichen es, z. B. zwischen personen-, zeit- und ortsbezogenen Schlagworten zu unterscheiden. Die Inhaltsklassifizierung basiert auf einem vorgegebenen, meist hierarchischen Klassifizierungsschema und zielt darauf ab, Ressourcen Kategorien und Unterkategorien zuzuordnen und damit letztlich nach ihrer thematischen Ausrichtung zu gruppieren. Beide Ansätze können zusammen angewendet werden.

Abgesehen von dieser eher formalen und theoretischen Unterscheidung der Katalogisierungsansätze erfordert der praktische Aufbau einer Bibliothek, in dessen Kern die Umwandlung von Daten in Informationen und letztendlich Wissen steht, von den Bibliothekaren erhebliche Anstrengungen und ist definitiv ein zeitaufwändiger Lernprozess, bei dem die Interaktion mit den Nutzern eine wichtige Rolle spielt. Es ist also ein Prozess mit einer bestimmten Historie. Dies bedeutet implizit, dass zwei Bibliothekare, die Dokumente wie Bücher thematisch einordnen, je nach ihrem eigenen Erfahrungsschatz und dem erfahrenen Prozess des Wissenserwerbs zu völlig unterschiedlichen Klassifikationen kommen können.

Es erfordert in der Regel ein tiefes Studium der Texte (wenn nicht sogar Spezialwissen zu bestimmten Themen), um wichtige Begriffe herauszufinden sowie ihre kontextabhängigen Bedeutungen zu bestimmen, die anschließend bei der Zuordnung von Kategorien zu bisher ungesehenen Inhalten und bei der Bestimmung ihrer Beziehungen untereinander verwendet werden sollen. Dieser Prozess beinhaltet also auch eine Abschätzung der semantischen Ähnlichkeit und Distanz zu anderen lokal verfügbaren Begriffen und Texten. So kann erst nach einer größeren Menge an erlangtem Wissen eine erste Klassifikation von Dokumenten mit dem notwendigen Reifegrad durchgeführt und ein grundlegendes (später erweiterbares) Katalog- und Archivierungssystem aufgebaut werden. Der resultierende Katalog ist eine kleine und kompakte Abstraktion von Details in jedem Buch und in verdichteter Form sogar eine Repräsentation menschlicher Intelligenz, die zum Tragen kam, um inhaltlich verwandte Bücher miteinander zu verbinden und, im

Falle eines Zettelkatalogs, entsprechend über die Platzierung der Karten zu entscheiden.

Technisch gesehen folgt dieser Aufbauprozess – im Gegensatz zu Googles Top-Down-Ansatz – einem Bottom-Up-Ansatz, da die Klassifizierung, Sortierung sowie Ein- und Anordnung der Bücher sukzessive erfolgt und mit einer anfänglich kleinen Menge an Büchern beginnt. Diese Prozesse werden hauptsächlich durch das spezielle (lokale) und allgemeine Wissen des Bibliothekars gelenkt. Da die Bibliothek auf diese Weise wächst, erleichtert das bereits bestehende Klassifikationsschema die Katalogisierung und Einordnung der eingehenden Bücher. Darüber hinaus stellt es – neben dem eigenen Wissen des Bibliothekars – die Wissensbasis für Auskünfte, wo etwa bestimmte Bücher oder Informationen, die für die Bibliotheksnutzer von Interesse sind, zu finden sind, dar. Dieser Ansatz ist wahrscheinlich zweckmäßiger und erfolgreicher als der erwähnte Top-Down-Ansatz von Google und Co., insbesondere dann, wenn Domänen- und Fachwissen benötigt wird, um eingehende Anfragen mit spezieller Terminologie zu bearbeiten, wenn es also (umgangssprachlich formuliert) darum geht, die „Nadel im Heuhaufen“ von Informationen zu finden.

Wie bereits für den Marketing-Bereich¹³ ausgeführt, ist der zweckdienliche Einsatz von „Small Data“-Ansätzen (im vorliegenden Fall u. a. das Fachwissen des Bibliothekars, der die Nutzer zu den gesuchten Informationen führt) oft vorteilhafter als die Verwendung unsachgemäßer Big-Data-Analysen. Darüber hinaus können auf Grundlage dieses lokalen Wissens thematisch ähnliche und verwandte Dokumente schnell identifiziert werden und werden daher in der Regel der gleichen Kategorie und somit konsistent in der Bibliothek zugeordnet.

3 Informationstechnologie in Bibliotheken

3.1 Der elektronische Informationsschalter

Die derzeit wichtigste Form von Bibliothekskatalogen ist der so genannte „Online Public Access Catalog“ (OPAC), eine elektronische bibliografische Datenbank, die die früheren physischen Zettelkataloge weitgehend überflüssig gemacht bzw. ersetzt hat.

Während OPACs den Nutzern den Zugriff auf und die Suche nach Bibliotheksressourcen über die jeweilige Online-Präsenz an jedem Ort und zu jeder Zeit

¹³ Lindstrom 2016.

ermöglichen, hat die Bereitstellung und Nutzung von Integrierten Bibliothekssystemen (ILS)¹⁴ (zu denen OPACs gehören) die Pflege dieser Kataloge (Verwaltung von Metadaten und Informationen) sowie die Erwerbung von Medien und die Verwaltung der Ausleihe auch für Bibliothekare deutlich bequemer gemacht. Insbesondere ermöglichen diese Systeme die Online-Ausleihe von digitalen Publikationen wie E-Books, E-Journals, E-Papers (elektronische Zeitungen und Zeitschriften) sowie von digitalisierten Büchern und elektronischen Lehrmaterialien.

Auch die Zusammenarbeit zwischen Bibliotheken ist durch die Einführung von Datenformaten, die die Nutzung, den Austausch und die Interpretation bibliografischer Informationen in Datensätzen fördern, einfacher geworden. Zu diesem Zweck hat sich der MARC-Standard (MACHine Readable Cataloging) weitgehend durchgesetzt. Auf diese Weise können Bibliotheken ihren Nutzern nicht nur lokale Bestände anbieten, sondern ihnen auch Datensätze von Partnerbibliotheken sowie zusätzliche Dienste wie die Fernleihe von Büchern zur Verfügung stellen. Das bedeutet, dass die Bibliotheksnutzer den Katalog der bevorzugten Bibliothek online und vor Ort einsehen können und landesweite oder sogar globale Informationen anderer Bibliotheken erhalten. Der Begriff „Hybridbibliothek“¹⁵ wurde geprägt, um zu kennzeichnen, dass eine bestimmte Bibliothek sowohl klassische als auch Online-Dienste anbietet.

Durch den Einsatz von ILS können räumliche und zeitliche Beschränkungen klassischer Bibliotheken überwunden werden, da die bereitgestellte elektronische Auskunftsstelle in der Regel rund um die Uhr zur Verfügung steht. Somit ist eine Anpassung an das Kommunikationsverhalten der Nutzer gegeben. Darüber hinaus unterstützen diese Systeme das Bibliothekspersonal durch automatische Analyse von Nutzeranfragen und ihre Weiterleitung an die entsprechenden Assistenten. Diese Hilfestellung wird zudem noch erweitert, indem Standardanfragen, z. B. nach Öffnungszeiten, selbstständig und ohne Beteiligung von Assistenten beantwortet werden. Die Integration weiterer elektronischer Kommunikationsdienste wie Chat- und Instant-Messenger-Dienste, Microblogging-Seiten, soziale Online-Netzwerke und Internet-Telefonie hat die Kommunikation mit den Bibliotheksnutzern erheblich erleichtert. Auch Online-Schulungen können mit diesen Mitteln leicht angeboten und durchgeführt werden. Dennoch ist es hierbei immer notwendig, den Schutz persönlicher und privater Daten zu respektieren.

¹⁴ Baeza-Yates/Ribeiro-Neto 2011.

¹⁵ Oppenheim/Smithson 1999.

3.2 Suche im Web und in OPACs

Aktuelle Websuchmaschinen können hilfreich sein, wenn es darum geht, kurzfristig relevante Dokumente zu finden, vor allem wenn nach bekannten Dingen (z. B. dem Ort eines Geschäfts) gesucht wird. Wenn es jedoch darum geht, eine umfassende Recherche zu einem bestimmten Thema durchzuführen, werden die Nutzer zumeist nicht richtig unterstützt oder gar sich selbst überlassen. In einem solchen Fall, der meist als Themensuche bezeichnet wird, müssen die Nutzer die zurückgegebenen Links zu Webdokumenten selbst untersuchen, ihre Relevanz bewerten und möglicherweise die ursprünglichen oder nachfolgenden Anfragen umformulieren, um ihren Informationsbedarf tatsächlich zu befriedigen. Dieser Prozess ist mühsam und zeitaufwändig, insbesondere wenn der Nutzer mit einem Thema nicht vertraut ist und die passende Terminologie (noch) nicht kennt. Darüber hinaus sind die meisten Webdokumente im Gegensatz zur Literatur in Bibliotheken nicht katalogisiert, und da ihre Vertrauenswürdigkeit nicht als selbstverständlich vorausgesetzt werden kann, muss sie stets aktiv hinterfragt werden.

In diesen Situationen sind Bibliotheken mit ihren Dienstleistungen, die sowohl von Bibliothekaren als auch von OPACs erbracht werden, definitiv von größerer Hilfe. Die Gründe dafür liegen auf der Hand:

1. Die Literatur einer Bibliothek wurde bewusst ausgewählt und erworben.
2. Eine Bibliothek stellt Literatur in gut geordneter und strukturierter Form zur Verfügung, so dass relevante Inhalte zu einem Interessensgebiet schnell gefunden werden können. Ein Suchvorgang kann gezielter durchgeführt werden, als dies bei einer Websuche möglich wäre.
3. Neben Feldern zur formalen Klassifizierung einer Publikation bieten OPACs auch spezielle Felder für Publikationen, die mit Hilfe von Methoden der Sacherschließung gefüllt werden. Die Integration von Optionen zur Facetten-suche ist daher ein gängiges Merkmal von OPACs.
4. OPACs liefern eine Fülle bibliografischer Informationen, die die weitere Suche nach verwandten Materialien vereinfachen. Es werden auch beispielsweise der Name des Autors und der Titel einer Publikation speziellen Feldern oder Elementen mit aussagekräftigen Bezeichnungen im Katalog zugeordnet, was es den Nutzern leicht macht, die bibliografischen Informationen einer Publikation richtig zu interpretieren.
5. Die Informationen, die den Nutzern von Bibliotheken zur Verfügung gestellt werden, sind in der Regel vertrauenswürdig. Dies bezieht sich sowohl auf die gefundene oder vorgeschlagene Literatur als auch auf andere Verweise zu dem gewünschten Thema.

Die Wahrscheinlichkeit, dass eine thematische Suche erfolgreich ist, ist daher größer, wenn man zuverlässige Bibliotheksdienste zu Rate zieht. Darüber hinaus sind ILS in der Lage, automatisch Zitate aus Publikationen zu extrahieren und mit der referenzierten Literatur zu verknüpfen. Der generierte Graph verwandter Materialien kann dann die Grundlage für inhalts- oder merkmalsbasierte Empfehlungsfunktionen sein, die die Nutzer z. B. von Webshops gewohnt sind.

4 Der Bibliothekar des Web

4.1 Dezentrale Websuche mit Bibliotheksdiensten

Bibliotheken sind seit jeher Vorreiter bei der Einführung und Annahme von Informationstechnologien. So setzen Sie Informationssysteme seit den 1950er Jahren ein, einer Zeit, in der auch der Begriff „Information Retrieval“ (IR)¹⁶ geprägt wurde. Damals wurden professionelle Rechercheure als „Suchvermittler“ eingesetzt, um die Anfragen der Nutzer in die Sprache des jeweiligen Systems zu übersetzen¹⁷. Heutzutage ist diese Funktion meist durch Suchmaschinen in verschiedenen Formen ersetzt worden.

Um jedoch eine moderne, von Bibliotheken inspirierte und dezentrale Web-suchmaschine, wie in der Einleitung motiviert, zu realisieren, ist es nötig, eine Transformation der Informationsbedürfnisse der Nutzer in geeignete und erfolgversprechende technische Repräsentationen durchzuführen und diese mit textuellen Ressourcen abzugleichen. Diese Aufgaben müssen autonom und ggf. automatisiert durchgeführt werden. Darüber hinaus müssen diese Repräsentationen in einer dezentralen Umgebung an Peers (Knoten in einem Peer-to-Peer-Suchnetzwerk der dezentralen Websuchmaschine) weitergeleitet werden, die wahrscheinlich in der Lage sind, die genannten Informationsbedürfnisse tatsächlich zu erfüllen. In den letzten Jahren wurden in diesem Rahmen wertvolle Forschungsergebnisse im Bereich des Peer-to-Peer Information Retrieval (P2PIR) erzielt.

Zum einen muss die der Weiterleitung vorausgehende Routing-Entscheidung auf Grundlage semantischer Gesichtspunkte getroffen werden, die auch Bibliothekare (unbewusst) berücksichtigen würden, wenn sie Bibliotheksnutzer zu relevanten Informationen und deren Quellen leiten. Dies ist eine besonders wichtige Aufgabe, da Informationen im Web weitgehend unorganisiert sowie spärlich und

¹⁶ Mooers 1951.

¹⁷ Witschel 2008.

oft nicht konsistent (wenn überhaupt) von Menschen und Maschinen annotiert sind und sich daher von Informationen in katalogisierten Bibliothekspublikationen unterscheiden. Um dazu in der Lage zu sein, muss sich jeder Peer der vorgeschlagenen dezentralen Websuchmaschine auf eine lokale Wissensbasis stützen, deren Organisation eng mit der des menschlichen (in diesem Fall des Bibliothekars) lexikalischen Wissens übereinstimmt. Hierzu muss er wiederum in der Lage sein, wertvolle Informationen aus Textquellen automatisiert zu extrahieren, zu indexieren und in Beziehung zu setzen. Dieser Lern-, Ordnungs- und Katalogisierungsprozess kann durch die Anwendung spezifischer Algorithmen und technischer Lösungen, die aus den Bereichen der Verarbeitung natürlicher Sprache und des Text Mining bekannt sind, realisiert werden. Diese sind insbesondere für Aufgaben wie die automatische und qualitativ hochwertige Schlüssel- und Suchwortextraktion, Textstruktur- und Eigennamenerkennung, sowie das Term- und Dokumentenclustering einsetzbar. Die zuvor beschriebenen Erschließungsansätze können somit in automatisierter Form realisiert werden. Der daraus resultierende Katalog bzw. Index ist jedoch nur bedingt mit eher monolithischen und manuell erstellten OPACs vergleichbar, da er sowohl maschinell und dezentral als auch automatisch erstellt und gepflegt wird.

Zum anderen ist es notwendig, implizite sprachbezogene Dynamiken im Web zu berücksichtigen. Insbesondere in sozialen Online-Netzwerken und Weblogs ist ein Sprachwandel klar erkennbar. Das bedeutet nicht nur, dass das öffentliche Interesse an Themen in einem bestimmten Zeitraum und an bestimmten Orten steigt oder sinkt, sondern auch, dass sich die Formulierungen zu ihrer Beschreibung ändern. Auch auf Jugendsprache und Slang muss dementsprechend eingegangen werden. Während sich insbesondere Bibliothekare aufgrund ihrer ständigen Interaktion mit Nutzern aller Altersgruppen und ihrer wachsenden Kenntnis der fachlichen Entwicklungen im Bibliotheksbestand leicht auf diese Veränderungen einstellen können, werden diese Bedeutungsverschiebungen von den derzeitigen semantischen Ansätzen für die Websuche bisher nicht angemessen berücksichtigt. Insbesondere können (in der Regel) spezialisierte (d. h. domänenbezogene) Ontologien oder Taxonomien diese Sprachdynamik nicht immer angemessen widerspiegeln, da sie normalerweise von Menschen, üblicherweise Domänenexperten, unter Verwendung einer festen Terminologie manuell erstellt werden. Hier ist ein neuer Ansatz für den maschinellen Umgang mit dieser Dynamik erforderlich. Auch ist die Abbildung natürlicher Vergessensprozesse durch technische Maßnahmen und deren Anwendung in diesem Rahmen ein wichtiges zukünftiges Forschungsfeld, um die Relevanzbewertung von Informationen zusätzlich positiv zu beeinflussen.

Wenn es darum geht, im Internet tiefgehende Recherchen durchzuführen, wäre zudem eine von Bibliothekaren inspirierte Mensch-Maschine-Interak-

tion von großer Hilfe. Denn in diesen Fällen wird die Suche zu einem Prozess der Informationssuche, der möglicherweise aus zahlreichen Zwischenschritten besteht, wie z. B. der Analyse der präsentierten Informationen und der Neuformulierung der Suchanfrage. Die vorgeschlagene dezentrale Websuchmaschine sollte in diesen Situationen durch interaktive Unterstützung von ähnlichem Nutzen sein. Das System sollte darum in der Lage sein, den Nutzer bei seiner aktuellen Suchaufgabe dahingehend zu unterstützen, indem es sofortiges Feedback z. B. zur Qualität einer Suchanfrage gibt oder (Gruppen von) thematisch verwandten Suchbegriffen vorschlägt sowie ähnliche und verwandte Websuchergebnisse gruppiert. Dabei sollte das System in der Lage sein, aus der Interaktion mit dem Nutzer zu lernen und somit kontextbasierte Suchwortvorhersagen zu treffen oder als „thematischer Wegweiser“ Empfehlungen für geeignete nächste Suchschritte zu unterbreiten. In diesem Sinne wird ein wichtiger Schritt in Richtung echter Informationskompetenz in Informationssystemen getan.

4.2 Die Realisierung

Ausgehend von diesen Überlegungen und den festgestellten Unzulänglichkeiten aktueller Websuchmaschinen wurde ein neues Konzept für die dezentrale Websuche abgeleitet, das unter dem Namen „Librarian of the Web“ („Bibliothekar des Web“)¹⁸ zusammengefasst wird und neuartige, von Bibliotheken inspirierte Ansätze, Methoden und technische Lösungen zur dezentralen Suche nach Textdokumenten im WWW umfasst. Eine erste Umsetzung in Form eines interaktiven Peer-to-Peer (P2P) Websuchsystems, genannt „WebEngine“¹⁹, wurde bereits veröffentlicht.

Die Client-Software dieses Systems besteht aus mehreren Komponenten, die für die Speicherung, das Retrieval und die semantische Analyse von Textdokumenten, für den Aufbau und die Wartung des P2P-Netzwerks sowie für die Ausführung von lokalen und netzwerkweiten Suchaufgaben zuständig sind. So entsteht ein dezentrales Websuchsystem, das erstmals moderne Textanalysetechniken mit neuartigen und effizienten Suchfunktionen und einem semantisch induzierten P2P-Netzwerkaufbau und -management kombiniert. In einer abstrakteren und allgemeineren Sichtweise nutzt das System Analyse- (Text Mining und Anfrageinterpretation) und Synthesemethoden (Bibliothek- und Netzwerkaufbau), wobei die letzteren von den erstgenannten abhängen.

¹⁸ Kubek 2020.

¹⁹ Kubek/Unger 2018.

Die WebEngine wurde als Java-basiertes P2P-Plug-in für den populären Apache Tomcat²⁰ Servlet-Container und Webserver mit einer grafischen Benutzeroberfläche (GUI) für jeden Standard-Webbrowser realisiert. Durch die Integration in den Webserver nutzt es dessen Laufzeitumgebung und kann auf die angebotenen Webseiten und Datenbanken des Servers mit allen zugehörigen Metainformationen zugreifen. Damit verfolgt das System einen alternativen, integrativen Ansatz zur Websuche unter dem Motto „Das Web ist seine eigene Suchmaschine“ und wurde – wie zuvor motiviert – so konzipiert, dass es die Nutzer bei bestimmten Such- und Rechercheaufgaben inhärent und aktiv unterstützt. Darüber hinaus wird die Struktur des generierten P2P-Netzes direkt durch die Ausnutzung der expliziten Topologie des Web (Links in Webdokumenten) induziert. Das P2P-Netz ist darüber hinaus in der Lage, sich durch Selbstorganisation so umzustrukturieren, dass es ohne eine zentrale Instanz wartbar und durchsuchbar wird und somit tatsächlich rein dezentral arbeitet.

5 Zusammenfassung

In diesem Artikel wurden die wichtigsten von Bibliotheken angebotenen Dienstleistungen und die Aufgaben der dort tätigen Bibliothekare untersucht und klassifiziert. Insbesondere wurden ihre beiden Hauptaufgaben, die Vermittlung zwischen Informationen und Bibliotheksnutzern sowie die Katalogisierung der (eingehenden) Bibliotheksressourcen, detailliert beschrieben. Da diese Aufgaben heutzutage in der Regel durch elektronische Informations- und Katalogisierungssysteme unterstützt werden, wurde auch auf diese eingegangen. Darüber hinaus wurde analysiert, welche bibliothekarischen Tätigkeiten in dezentralen Websuchsystemen technisch realisiert werden können, um tiefgehende Rechercheaufgaben nachhaltig zu unterstützen. Abschließend wurde das neue Konzept des „Librarian of the Web“ sowie dessen erste technische, P2P-basierte Umsetzung, genannt „WebEngine“, skizziert.

6 Literaturverzeichnis

A. L. A. P. Committee: Presidential committee on information literacy: Final Report. Chicago 1989, <https://www.ala.org/acrl/publications/whitepapers/presidential> [Zugriff: 28.11.2021].

²⁰ <http://tomcat.apache.org> [Zugriff: 14.12.2021].

- American Library Association: Guidelines for Behavioral Performance of Reference and Information Service Providers. 2013, <https://www.ala.org/rusa/resources/guidelines/guidelinesbehavioral> [Zugriff: 28.11.2021].
- Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier: Modern Information Retrieval: The Concepts and Technology Behind Search. 2. Aufl. Boston 2011.
- Bell, Jodie: The developing role of librarians in a digital age. 2016, <https://www.infotoday.eu/Articles/Editorial/Featured-Articles/The-developing-role-of-librarians-in-a-digital-age-110185.aspx> [Zugriff: 28.11.2021].
- Eberhardt, Joachim: Was ist (bibliothekarische) Sacherschließung? In: Bibliotheksdienst 46 (2012), S. 386–401.
- Gantert, Klaus: Bibliothekarisches Grundwissen. 9. Aufl. Berlin 2016.
- Kubek, Mario: Concepts and Methods for a Librarian of the Web. Cham 2020.
- Kubek, Mario; Unger, Herwig: The WebEngine – A Fully Integrated, Decentralised Web Search Engine. In: Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval (NLPIR 2018). New York 2018, S. 26–31.
- Lindstrom, Martin: Small Data: The Tiny Clues That Discover Huge Trends. New York 2016.
- Mooers, Calvin N.: Zatoncoding applied to mechanical organization of knowledge. In: American Documentation 2.1 (1951), S. 20–32.
- Oppenheim, Charles; Smithson, Daniel: What is the hybrid library? In: Journal of Information Science 25.2 (1999), S. 97–112.
- Sparrow, Betsy; Liu, Jenny; Wegner, Daniel M.: Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. In: Science 333.6043 (2011), S. 776–778.
- Statista: Monatliches Datenvolumen des privaten Internet-Traffics nach Segmenten in den Jahren 2014 bis 2017 sowie eine Prognose bis 2022 (in Exabyte). 2021, <https://de.statista.com/statistik/daten/studie/152551/umfrage/prognose-zum-internet-traffic-nach-segment> [Zugriff: 28.11.2021].
- Varshney, Lav R.: The Google effect in doctoral theses. In: Scientometrics 92 (2012), S. 785–793.
- Witschel, Hans F.: Global and Local Resources for Peer-to-Peer Text Retrieval. Dissertation. Universität Leipzig 2008.

PD Dr.-Ing. habil. Mario Kubek
 Fakultät für Mathematik und Informatik
 FernUniversität Hagen
 Universitätsstr. 1
 58097 Hagen
 Deutschland
 E-Mail: mario.kubek@fernuni-hagen.de