

Programmierung eines datengetriebenen Unterstützungssystems für Etatplanung und Mittelallokation einer hybriden Spezialbibliothek

Exposé zur Masterarbeit

Peter Breternitz

TH Wildau, Wildau Institute of Technology

Bibliotheksinformatik

peter.breternitz@th-wildau.de

I. EINFÜHRUNG

Ausgehend von ökonomischen, informationstechnologischen und marktpolitischen Einschnitten in den vergangenen Jahrzehnten¹, sind Bibliotheken dazu veranlasst, ihr Budget hinsichtlich der Informationsbedarfe ihrer Nutzer:innen behutsamer zu planen und sich in zunehmenden Maße gegenüber ihren Unterhaltsträgern zu rechtfertigen.

Die Relevanz von bibliothekarischen Kennzahlen ist in diesem Zusammenhang größer geworden. Deswegen ist es wichtig, Daten aus bibliothekarischen Servicedienstleistungen und Geschäftsprozessen zu aggregieren, zu erheben und statistisch auszuwerten, um auf Basis der daraus erzielten Erkenntnisse handeln zu können.

Ziel der zu entstehenden Arbeit ist die Entwicklung einer interaktiven Business-Intelligence-Applikation als proof-of-concept, mit der systematisch die relevanten Daten einer hybriden Spezialbibliothek aggregiert, statistisch analysiert und mit geeigneten und modernen Datenvisualisierungstechniken² ausgegeben werden sollen. Vor allem soll sich hier auf automatisierte Prozesse zur Gewinnung der Ergebnisse konzentriert werden.

Mit diesen automatisch angefertigten statistischen Datenanalysen sollen zukünftige Entscheidungen im Bibliotheksmanagement wie Erwerbungspolitik, Budgetplanung und Mittelallokation hinsichtlich der weite-

ren Entwicklung der Servicedienstleistungen evidenzbasiert und datengetrieben unterstützt werden.

Darüber hinaus soll die Applikation eine Funktion beinhalten, ausgewählte Resultate automatisiert als *factsheet* zu exportieren, um diese als Rechenschaftsbericht gegenüber Stakeholdern der Bibliothek präsentieren zu können.

Die Notwendigkeit für diese Applikation ist durch das Fehlen eines zentralen Nachweisortes für bibliothekarische Statistiken in der Bibliothek gegeben. Da die Bibliothek zudem verschiedene Recherche-Systeme den Wissenschaftler:innen anbietet, wäre eine Engführung der statistischen Datenerhebung auf eine Plattform begrüßenswert. Des Weiteren ist das Erfordernis, bibliothekarische Geschäftsprozesse zu evaluieren und die Servicedienstleistungen bezüglich der Ziele der Institution noch weiter zu optimieren, von großer Relevanz. Mit dem Ende der Konsolidierungsphase der Bibliothek, die im Zuge des *Max-Planck-Institutes für empirische Ästhetik* 2014 gegründet wurde, tritt sie ein in eine Phase, in der ab dem Jahr 2021 Budgetplanungen eine größere Rolle spielen werden.

Die zu entstehende Applikation könnte hierbei helfen, systematisches Controlling einzuführen und das Bibliotheksmanagement weiter zu professionalisieren. Um künftigen Anforderungen gewachsen zu sein, soll sie modulbasiert programmiert werden und dadurch leicht erweiterbar und eventuell von anderen Bibliotheken nachnutzbar sein.

II. LITERATURDISKUSSION

Ein gutes Beispiel für ein datengetriebenes Unterstützungssystem findet sich in der Literatur bei Spielberg, der sich mit dem Thema der Bestandspflege an der *Universitätsbibliothek Essen* befasst und eine Applikation (weiter-)entwickelt hat, die die Fachreferent:innen bei der Aussonderung und

¹Als Gründe zu nennen wären hier: die Explosion der Zeitschriftenpreise im Bereich der Science, Technology & Medicine (STM), das Aufkommen von E-Publishing und die Konzentration auf wenige Verlage

²Visualisierungen können komplexe Sachverhalte herunterbrechen und so große Datenmengen - im Gegensatz zu großen Tabellen - leicht verständlich darstellen. Im Kontext dieser Arbeit konzentriere ich mich auf Ansätze, die Visualisierungen mittels Visualisierungstechniken algorithmisch aus Daten erzeugen (Informationsvisualisierung, Datenvisualisierung und visuelle Analyse).[5]

Erwerbung von Medien unterstützt.[9] Ebenso finden sich in der Fachliteratur Ansätze, die vorrangig anhand einzelner Fragestellungen hinsichtlich der Bestandsentwicklung[2] oder anderer bibliothekarischer Servicedienstleistungen[4, 3, 6] verschiedene statistische Analysen vollzogen und diese visualisiert haben. Eine Ausnahme bildet die Entwicklung eines Dashboards an der *New York University Health Sciences Libraries*, das versucht, möglichst viele Metriken aus bibliothekarischen Dienstleistungen aufzunehmen.[7] Fast alle Projekte sind an größeren Universitätsbibliotheken mit ganz unterschiedlichen softwaretechnischen Herangehensweisen[1, 11] und Zielen[8] entstanden.

Dennoch fehlen in der gesichteten Literatur Teile, die sich mit der Budgetierung befassen und Auskunft über Mittelallokation geben.

Zudem fehlt ein Beispiel in der Literatur, das holistisch alle relevanten Daten, die in den Geschäftsgängen und Servicedienstleistungen insbesondere einer Spezialbibliothek entstehen, aggregiert, auf diesen Daten automatisch statistische Analysen ausführt und diese mit modernen Visualisierungstechniken interaktiv darstellt.

III. PROBLEMSTELLUNG

Die Bibliothek des *Max-Planck-Institutes für empirische Ästhetik* ist Teil des *hessischen Bibliotheksverbundes (HeBis)*. Die Geschäftsprozesse der Katalogisierung und der Erwerbung finden im Zentralsystem *CBS* und im Lokalsystem *LBS4* von *OCLC* statt. *LBS4* wird gehostet und betreut vom Lokalsystem-Team Frankfurt. Als Service-Leistung werden der Bibliothek besondere Funktionalitäten für das *CBS* bereitgestellt. Außerdem erhält die Bibliothek unter anderem Ausleih-, Budget- und Umsatzübersichten als Text per E-mail zugesandt.

Neben der Verankerung in der deutschen Bibliotheksverbundlandschaft wird die Bibliothek in ihren Aufgaben von der *Max Planck Digital Library (mpdl)* unterstützt. Deren Portfolio umfasst vorrangig die zentrale Lizenzierung von relevanten elektronischen Informationsressourcen, die Bereitstellung von Softwarelösungen, das Betreiben eines Publikationsrepositoriums und das Vorantreiben von Open-Access. Zudem stellt sie den Bibliotheken der einzelnen Max-Planck-Institute *COUNTER*-Statistiken zur Verfügung, die von den Verlagen geliefert werden.

Außer diesen bereitgestellten Daten erhebt die Bibliothek Daten unter anderem über die Frequentation des Lesesaals, die Nutzung des nehmenden Fernleihservices, des Dokumentenlieferdienstes *subito* und des Bestandswachstums vor Ort. Nach den unterschiedlichen Verantwortlichkeiten aufgeteilt, werden diese Daten an verschiedenen virtuellen Orten erhoben. Eine systematische Auswertung der Daten findet nur unzureichend statt. Daher regt sich der Wunsch seitens

der Bibliotheksleitung und der Mitarbeiter:innen nach einem gemeinsamen Tool, mit dem übersichtlich und klar alle notwendigen nutzungs- und sammlungsbezogenen Statistiken einer Spezialbibliothek erfasst und dargestellt werden können.³

Es gibt eine Vielzahl kommerzieller Lösungen für den Bibliotheksbereich, die auf Business-Intelligence-Software basieren. Zu nennen wären *AlmaAnalytics* für das Next-Generation-Library-System *Alma* von *ExLibris*⁴, *BibControl* von *OCLC*⁵, *CollectionHq* von *Baker & Taylor*⁶ oder *Libinsight* von *SpringShare*⁷. Darüber hinaus gibt es Business-Intelligence-Applikationen, die von Bibliotheken für Reporting, Datenanalyse und Datenvisualisierung adaptiert werden, wie zum Beispiel *Tableau* von der Firma *Tableau Software* oder *Crystal Reports* von *SAP*. Diese Applikationen sind entweder an bestimmte Bibliothekssysteme zurückgebunden, limitiert in ihren Funktionen[10] oder zu generisch. Überdies wird sowohl von *HeBis* bzw. von der Lokal-Bibliothekssystembetreuung als auch von der *mpdl* keine Applikation in dieser Richtung angeboten. Ebenso ist ungewiss, wann die Ablösung des schon betagten *CBS/LBS* hin zu einem neuen Next-Generation-Library-System im *HeBis-Verbund* stattfinden wird und ob es ein Modul zur statistischen Datenerhebung liefern wird.

IV. INHALTLICHER ABRISS DER MASTERARBEIT

Im Folgenden wird ein skizzenhafter Überblick gegeben, welche Punkte die theoretische Konzeption und praktische Umsetzung beinhalten könnten.

Theoretische Konzeption

- Identifizierung der sammlungs- und bestandsbezogenen Geschäftsgänge und Servicedienstleistungen und der in diesem Zusammenhang stehenden Daten.
- Sind die Daten ausreichend, um auf ihnen sinnvolle Analysen vollführen zu können? In welchem Format sollten die Daten für die statistischen Analysen vorliegen? Wie müssen diese aufbereitet bzw. bereinigt werden?
- Wie sollen die Daten letztendlich vorgehalten werden?
- Wie sieht die Anbindung an LBS-Datenbank für die Abfrage von Live-Daten aus?

³Zwar führt *HeBis* eine Bestandsstatistik, diese ist aber insbesondere für die Evaluation und Optimierung von Geschäftsprozessen einer Spezialbibliothek insuffizient. https://www.hebis.de/de/lueber_uns/statistik/cbs_statistik.php Auch an der deutschen Bibliotheksstatistik nimmt die Bibliothek nicht teil. Beide bieten zudem nur Zahlenkolonnen und keine weiteren Visualisierungen an.

⁴<https://www.exlibrisgroup.com/products/alma-library-services-platform/alma-analytics> Stand: 26.05.2020

⁵<https://www.oclc.org/de/bibcontrol.html> Stand: 26.05.2020

⁶<https://www.collectionhq.com/> Stand: 26.05.2020

⁷<https://springshare.com/libinsight/> Stand: 26.05.2020

- Identifizierung der statistischen Methoden. Welche Fragen werden an die Daten gestellt, welche Antworten sollen diese liefern?
- Wahl des Frameworks für die Applikation/Dashboard. Begründung warum?
- Identifizierung geeigneter grafischer Visualisierungen. Welche Interaktionen soll die Applikation bieten? Kleine Bedarfsanalyse mit den Kolleg:innen.
- Identifizierung und Beschreibung der funktionalen und nicht funktionalen Anforderungen der Applikation - Lastenheft. Kleine Bedarfsanalyse mit den Kolleg:innen.
- Praktische Überlegungen hinsichtlich der modularen Aufteilung in der Programmierung.

Praktische Umsetzung

- Prototypische Umsetzung ausgewählter Anforderungen des Lastenheftes nach Priorisierung der Anforderungen.
- Programmierung nach modularem Architekturansatz, d.h. die Integration neuer Datenquellen, die Umsetzung neuer Auswertungen und neuer Visualisierungen sollen mit geringem Aufwand erfolgen.
- Programmierung für die automatische Weiterverarbeitung von ausgewählten Daten.
- Programmierung der Algorithmen von ausgewählten statistischen Analysen.
- Programmierung von ausgewählten Datenvisualisierungstechnologien.

LITERATUR

- [1] J. L. Finch and A. R. Flenner. Using data visualization to examine an academic library collection. *College and Research Libraries*, 77(6):765–778, 2016.
- [2] Michael Hughes. A Long-Term Study of Collection Use Based on Detailed Library of Congress Classification, a Statistical Tool for Collection Management Decisions. *Collection Management*, 41(3):152–167, 2016.
- [3] Jennifer E. Knievel, Heather Wicht, and Lynn Silipigni Connaway. Use of Circulation Statistics and Interlibrary Loan Data in Collection Management. 2006, 67(1):15, 2006.
- [4] Atalay Kutlay and Cal Murgu. Shiny Fabric: A Lightweight, Open-source Tool for Visualizing and Reporting Library Relationships. *Code4Lib*, (47), 2020.
- [5] Yuvaraj Mayank. Infographics: tools for designing, visualizing data and storytelling in libraries. 34(5):6–9, 2020/05/26 2017.
- [6] Andy Meyer. Using R and the Tidyverse to Generate Library Usage Reports. *Code4Lib*, (39), 2018.
- [7] Emily Morton-Owens and Karen Hanson. Trends at a Glance: A Management Dashboard of Library Statistics. *Information Technology and Libraries*, 31, 2012.
- [8] Eric Phetteplace. Effectively Visualizing Library Data. *Reference & User Services Quarterly*, 52(2), 2012-12-20.
- [9] Eike T. Spielberg. *Der FachRef-Assistent : personalisiertes, fachspezifisches und transparentes Bestandsmanagement*. Thesis, 2017.
- [10] Golas Ulrike. Statistische Abfragen mit Alma für die Fachreferatsarbeit. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB*, 5(4), 2018.
- [11] Laura K. Wiegand and Bob Humphrey. Visualizing Library Statistics using Open Flash Chart 2 and Drupal. *Code4Lib*, (19), 2013.