

Masterarbeit

zur Erlangung des akademischen Grades
Master

Technische Hochschule Wildau

Fachbereich Wirtschaft, Informatik, Recht

Studiengang Bibliotheks-informatik (M. Sc.)

Thema (deutsch): Konzeption und Entwicklung eines datengetriebenen Unterstützungssystems für Etatplanung und Mittelallokation einer hybriden Spezialbibliothek

Thema (englisch): Design and development of a data-driven support system for budget planning and resource allocation of a hybrid library

Autor/in: Peter Breternitz

Seminargruppe: BIM/17

Betreuer/in: Dipl.-Informatiker Sascha Szott

Zweitgutachter/in: Dr. Frank Seeliger

Eingereicht am: 10.03.2021

Konzeption und Entwicklung eines
datengetriebenen Unterstützungssystems
für Etatplanung und Mittelallokation
einer hybriden Spezialbibliothek

von

Peter Breternitz

Betreuer:

Dipl.-Informatiker Sascha Szott

Dr. Frank Seeliger

ABSTRACT

Scientific documents often use \LaTeX for typesetting. While numerous packages and templates exist, it makes sense to create a new one. Just because.

ZUSAMMENFASSUNG

Scientific documents often use \LaTeX for typesetting. While numerous packages and templates exist, it makes sense to create a new one. Just because.

INHALTSVERZEICHNIS

1	EINFÜHRUNG	1
1.1	Problemstellung	1
1.2	Ziel der Arbeit	1
1.3	Verwandte Arbeiten	2
1.4	Gliederung der Arbeit	2
2	THEORETISCHE GRUNDLAGEN	3
2.1	Bibliothek und Statistik	3
2.2	Datenvisualisierung und Datenvisualisierungstechniken	3
2.3	Business-Intelligence-Systeme	3
3	AUSGANGSSITUATION	5
3.1	Bibliothek, Aufgabe, Kennzahlen	5
3.2	Datengrundlage aus den bibliothekarischen Geschäftsgängen	6
4	KONZEPTION EINER LÖSUNG	7
4.1	Anforderungsanalyse	7
4.1.1	Ziel	7
4.1.2	Funktionale Anforderungen	7
4.1.3	Nicht funktionale Anforderungen	7
4.1.4	Anwendungsfälle	7
5	DISKUSSION DER UMSETZUNG	9
5.1	Design	9
5.1.1	Systemarchitektur	9
5.1.2	Teilsysteme	10
5.2	Implementierung	10
5.2.1	Umgesetzte Anforderungen	10
5.2.2	Funktionsweise	10
5.3	Bewertung	10
6	SCHLUSS	11
6.1	Soll-Ist-Vergleich (Stand der Umsetzung)	11
6.2	Lessons learned	11

Inhaltsverzeichnis

6.3	Welche Themen wurden nicht bearbeitet	11
6.4	Welche Themen sind im Anschluss denkbar	11
TABELLENVERZEICHNIS		13
ABBILDUNGSVERZEICHNIS		14
QUELLCODEVERZEICHNIS		16
AKRONYME		19
GLOSSAR		21
LITERATURVERZEICHNIS		23

1 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.

1.1 PROBLEMSTELLUNG

1.2 ZIEL DER ARBEIT

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 weiteren 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.

¹ Als 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

² Visualisierungen 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).[6]

1.3 VERWANDTE ARBEITEN

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*³, *BibControl* von *OCLC*⁴, *CollectionHq* von *Baker & Taylor*⁵ oder *Libinsight* von *SpringShare*⁶. 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^[2] 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. 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 Erwerbung von Medien unterstützt.^[10] Ebenso finden sich in der Fachliteratur Ansätze, die vorrangig anhand einzelner Fragestellungen hinsichtlich der Bestandsentwicklung^[3] oder anderer bibliothekarischer Servicedienstleistungen^[4, 5, 7] 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.^[8] Fast alle Projekte sind an größeren Universitätsbibliotheken mit ganz unterschiedlichen softwaretechnischen Herangehensweisen^[1, 11] und Zielen^[9] 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.

1.4 GLIEDERUNG DER ARBEIT

³ <https://www.exlibrisgroup.com/products/alma-library-services-platform/alma-analytics> Stand: 26.05.2020

⁴ <https://www.oclc.org/de/bibcontrol.html> Stand: 26.05.2020

⁵ <https://www.collectionhq.com/> Stand: 26.05.2020

⁶ <https://springshare.com/libinsight/> Stand: 26.05.2020

2 THEORETISCHE GRUNDLAGEN

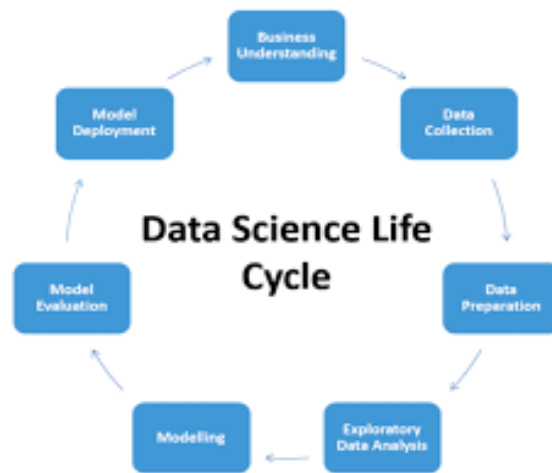


Abbildung 2.1: Data Science Cycle

2.1 BIBLIOTHEK UND STATISTIK

2.2 DATENVISUALISIERUNG UND DATENVISUALISIERUNGSTECHNIKEN

2.3 BUSINESS-INTELLIGENCE-SYSTEME

3 AUSGANGSSITUATION

3.1 BIBLIOTHEK, AUFGABE, KENNZAHLEN

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 *OCCLC* 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.

Geschäftsgang	erhobene Daten	Format	Quelle
Buchservice	Fernleihe	Excel	eigenständig
Buchservice	Scan	Excel	eigenständig
Buchservice	Subito	Excel	eigenständig
Buchservice	Sonstiges	Excel	eigenständig
Buchservice	elektronisch	Excel	eigenständig
Buchservicw	Ausleihen pro Abteilung	Excel	eigenständig
Bibliotheksbenutzung	Benutzer:Innenanzahl	Excel	eigenständig
Bibliotheksbenutzung	nachgefragte Medien	email	LBS
Ausleihe	Ausleihstatistik	Excel	LBS
Erwerbung	monatliche Ausgaben nach Lieferanten	email	LBS
Erwerbung	Budgetübersicht nach Kostenstellen	email	LBS
Bucheingang	monatliche Neuerwerbungslisten	Word	LBS
Elektronische Ressourcen	Counter-Statistiken	tsv	mpdl
Bestand	eigene	csv	LBS

Tabelle 3.1: Eine Liste von Geschäftsgängen zu denen statistische Daten erhoben werden

3 Ausgangssituation

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.¹

3.2 DATENGRUNDLAGE AUS DEN BIBLIOTHEKARISCHEN GESCHÄFTSGÄNGEN

¹ Zwar führt *HeBis* eine Bestandsstatistik, diese ist aber insbesondere für die Evaluation und Optimierung von Geschäftsprozessen einer Spezialbibliothek insuffizient. https://www.hebis.de/de/ueber_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 KONZEPTION EINER LÖSUNG

4.1 ANFORDERUNGSANALYSE

4.1.1 ZIEL

4.1.2 FUNKTIONALE ANFORDERUNGEN

4.1.3 NICHT FUNKTIONALE ANFORDERUNGEN

4.1.4 ANWENDUNGSFÄLLE

5 DISKUSSION DER UMSETZUNG

5.1 DESIGN

5.1.1 SYSTEMARCHITEKTUR

```
class ParetoDemo(AppBase):
    """Example creating a simple Pareto chart."""

    name = 'Car Share Pareto Demo'
    """Application name"""

    data_raw = None
    """All in-memory data referenced by callbacks and plotted. If modified, will impact all
    viewers."""

    chart_main = None
    """Main chart (Pareto)."""

    id_chart = 'pareto'
    """Unique name for the main chart."""

    def initialization(self):
        """Initialize ids with 'self.register_uniq_ids([...])' and application data."""
        super().initialization()
        self.register_uniq_ids([self.id_chart])
        # Format the car share data from plotly express for the Pareto
        self.data_raw = (px.data.carshare()
                        .rename(columns={'peak_hour': 'category', 'car_hours': 'value'}))
        self.data_raw['category'] = [f'H:{cat:02}' for cat in self.data_raw['category']]
```

Quellcode 5.1: Python example

5 Diskussion der Umsetzung

5.1.2 TEILSYSTEME

5.2 IMPLEMENTIERUNG

5.2.1 UMGESETZTE ANFORDERUNGEN

5.2.2 FUNKTIONSWEISE

5.3 BEWERTUNG

6 SCHLUSS

6.1 SOLL-IST-VERGLEICH (STAND DER UMSETZUNG)

6.2 LESSONS LEARNED

6.3 WELCHE THEMEN WURDEN NICHT BEARBEITET

6.4 WELCHE THEMEN SIND IM ANSCHLUSS DENKBAR

TABELLENVERZEICHNIS

3.1	Eine Liste von Geschäftsgängen zu denen statistische Daten erhoben werden	5
-----	---	---

ABBILDUNGSVERZEICHNIS

2.1	Data Science Cycle	3
-----	------------------------------	---

QUELLCODEVERZEICHNIS

5.1	Python example	9
-----	--------------------------------	---

AKRONYME

CBS	Zentralsystem
HeBis	Hessisches Bibliotheksinformationssystem
LBS	Lokalsystem
mpdl	max-planck-digital-library
RDA	Resource Description Access
STM	Science, Technology, and Medicine

GLOSSAR

Business Intelligence	Begriff der Wirtschaftsinformatik, der Verfahren und Prozesse zur systematischen Analyse des eigenen Unternehmens bezeichnet.
Business Intelligence Software	Begriff der Wirtschaftsinformatik, der Verfahren und Prozesse zur systematischen Analyse des eigenen Unternehmens bezeichnet.
\LaTeX	A document preparation system
\mathbb{R}	The set of real numbers

LITERATURVERZEICHNIS

1. J. L. Finch und A. R. Flenner. „Using data visualization to examine an academic library collection“. *College and Research Libraries* 77:6, 2016, S. 765–778. DOI: [10.5860/crl.77.6.765](https://doi.org/10.5860/crl.77.6.765).
2. U. Golas. „Statistische Abfragen mit Alma für die Fachreferatsarbeit“. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB* 5:4, 2018. DOI: [10.5282/o-bib/2018H4S44-57](https://doi.org/10.5282/o-bib/2018H4S44-57). URL: <https://www.o-bib.de/article/view/5386>.
3. M. 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, 2016, S. 152–167. ISSN: 0146-2679. DOI: [10.1080/01462679.2016.1169964](https://doi.org/10.1080/01462679.2016.1169964). URL: <https://doi.org/10.1080/01462679.2016.1169964>.
4. J. E. Knievel, H. Wicht, und L. S. Connaway. „Use of Circulation Statistics and Interlibrary Loan Data in Collection Management“. *2006* 67:1, 2006, S. 15. ISSN: 2150-6701. DOI: [10.5860/crl.67.1.35](https://doi.org/10.5860/crl.67.1.35). URL: <https://crl.acrl.org/index.php/crl/article/view/15771>.
5. A. Kutlay und C. Murgu. „Shiny Fabric: A Lightweight, Open-source Tool for Visualizing and Reporting Library Relationships“. *Code4Lib* 47, 2020. ISSN: 1940-5758. URL: <https://journal.code4lib.org/articles/14938>.
6. Y. Mayank. „Infographics: tools for designing, visualizing data and storytelling in libraries“. 34:5, 2017, S. 6–9. DOI: [10.1108/LHTN-01-2017-0004](https://doi.org/10.1108/LHTN-01-2017-0004). URL: <https://doi.org/10.1108/LHTN-01-2017-0004> (besucht am 16. 05. 2013).
7. A. Meyer. „Using R and the Tidyverse to Generate Library Usage Reports“. *Code4Lib* 39, 2018. ISSN: 1940-5758. URL: <https://journal.code4lib.org/articles/13282>.
8. E. Morton-Owens und K. Hanson. „Trends at a Glance: A Management Dashboard of Library Statistics“. *Information Technology and Libraries* 31, 2012. DOI: [10.6017/ital.v31i3.1919](https://doi.org/10.6017/ital.v31i3.1919).
9. E. Phetteplace. „Effectively Visualizing Library Data“. *Reference & User Services Quarterly* 52:2, 2012-12-20. ISSN: 1094-9054. DOI: <https://doi.org/10.5860/rusq.52n2.93>.

10. E. T. Spielberg. „Der FachRef-Assistent : personalisiertes, fachspezifisches und transparentes Bestandsmanagement“. Thesis. 2017. URL: <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:hbz:79pbc-opus-9888>.
11. L. K. Wiegand und B. Humphrey. „Visualizing Library Statistics using Open Flash Chart 2 and Drupal“. *Code4Lib* 19, 2013. ISSN: 1940-5758. URL: <https://journal.code4lib.org/articles/7812>.

SELBSTÄNDIGKEITSERKLÄRUNG

Ich versichere, dass die vorliegende Arbeit von mir selbständig und ohne unerlaubte Hilfe angefertigt worden ist. Ich habe alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, durch Zitate bzw. Literaturhinweise als solche kenntlich gemacht.

Ort, Datum

Unterschrift