

Kompetenzmonitor

OpenBiblioJobs

Idee für die Weiterentwicklung des Stellenanzeigen-Portals OpenBiblioJobs incl. Skizze für ein Einjahresprojekt

| | |
|---|----------|
| Partner und Unterstützer | 1 |
| Kurzbeschreibung | 1 |
| Motivation und Ausgangslage | 1 |
| Ziele des Vorhabens | 2 |
| Material und Methoden | 2 |
| Vorarbeiten | 2 |
| Zeit- und Arbeitsplan | 3 |
| Arbeitspakete | 3 |
| Meilensteine | 4 |
| Ausblick auf mögliche Folge- und parallele Projektaktivitäten, Nutzen für die Community | 5 |
| Referenzen | 5 |

Partner und Unterstützer

- Open Science Lab, Technische Informationsbibliothek Hannover (TIB), Ina Blümel, Lambert Heller, Cedrik Zellmann
- Openbibliojobs, Pascal Ngoc Phu Tu (Admin Openbibliojobs-Portal), SUB Hamburg
- Institut für Digital Humanities (IDH), Jürgen Hermes (Geschäftsführung & Leitung Projekt “Qualifikationsentwicklungsforschung”), Universität zu Köln
- Institut für Deutsche Sprache, Andreas Witt

- VDB - Verein Deutscher Bibliothekarinnen und Bibliothekare, Konstanze Söllner (Vorsitzende), Universitätsbibliothek Erlangen-Nürnberg¹
- BIB - Berufsverband Information Bibliothek, Ute Engelkenmeier (Vorsitzende), Universitätsbibliothek Dortmund²

Kurzbeschreibung

Motivation und Ausgangslage

Digitalisierung und Open Science führen seit einigen Jahren zu einem rapiden Wandel von Forschungsinfrastruktur und wissenschaftlichem Informationsaustausch. Dies drückt sich nicht zuletzt in den sich wandelnden Anforderungen an die 23.000 Mitarbeiterinnen und Mitarbeiter deutscher Bibliotheken aus. Die Stellenangebote der Bibliotheken spiegeln wider, welche Bedarfe für die Qualifikation in Projekten und alltäglicher Arbeit gesucht werden. Die in den Stellenangeboten enthaltenen Informationen haben ein hohes Erkenntnispotential für diejenigen, die den Wandel der Rollen und Funktionen von Bibliotheken (wissenschaftlichen Bibliotheken, aber auch öffentlichen Bibliotheken sowie anderen, ähnlichen Gedächtnisinstitutionen) verstehen wollen. Diese erweiterten und veränderten Funktionen müssen verstanden werden, um z.B. einschlägige Hochschul-Studiengänge sowie die Laufbahnen dieser (in Deutschland überwiegend öffentlich finanzierten) Institutionen (weiter-) zu entwickeln. Neben dieser Beratung von Hochschulen und Bibliotheken kann das in bibliothekarischen Stellenangeboten enthaltene Wissen auch dabei helfen, Open Science und die Digitalisierung der Wissenschaft als einen Aspekt veränderter Arbeitsweisen besser zu verstehen. Welche Infrastrukturen, welche Skills, welche Methoden tauchen auf, nehmen an Bedeutung zu, oder lösen ggf. sogar frühere Arbeitsweisen ab?

Seit 2012 werden bibliothekarische und informationswissenschaftliche Stellenanzeigen über das von Freiwilligen betriebene und inhaltlich gepflegte Portal OpenBiblioJobs³ gebündelt referenziert und

¹ “Die Auswertung des Datensets verbunden mit einem Harvesting der Stellenanzeigen lässt wichtige Ergebnisse hinsichtlich neuer Arbeitsfelder in Bibliotheken und wissenschaftsnahen Infrastrukturen erwarten (z.B. Open Science, Data Stewardship). Mit der Integration in die vom VDB geförderte Plattform Openbibliojobs werden die Ergebnisse für die Community, insbesondere auch für den in der Berufsorientierung befindlichen Nachwuchs unmittelbar sichtbar.” (Konstanze Söllner zur Begründung der Unterstützung, 31.8.2018)

² “Im Zuge des aktuellen Vorhabens, der Kooperation zwischen openbibliojobs und BIB/bibliojobs wird auch überlegt, die Daten aus bibliojobs zu archivieren und zur Verfügung zu stellen.” (Ute Engelkenmeier zur Begründung der Unterstützung, 8.2.2019)

³ <https://jobs.openbiblio.eu/>

hat sich als erste Anlaufstelle für Jobsuchende in benannten Arbeitsfeld etabliert. OpenBiblioJobs ist nicht-kommerziell; die Betriebskosten des Portals in Form einer WordPress-Seite werden vom VDB getragen. OpenBiblioJobs verfolgt den konsequenten Ansatz frei zugänglicher, unter CC-BY zur Verfügung gestellter Inhalte⁴: Die Anzeigen werden mittels eines reduzierten Metadatensets (Einrichtung, Jobbezeichnung, Bewerbungsfrist, Link zur Stellenanzeige) in der Datenbank von OpenBiblioJobs referenziert. Diese Daten wurden im Zuge der Vorarbeiten zu diesem Projekt bereits von Blümel et al (2017) und Zellmann (2018a) bei Zenodo zur Nachnutzung durch Dritte veröffentlicht. Die Volltexte⁵ der Stellen seit 2012 liegen (bzw. lagen) auf den Servern der stellen anbietenden Institutionen, werden aktuell nicht von OpenBiblioJobs archiviert und sind Gegenstand dieses Vorhabens. Die Daten werden im Sinne der Urheberrechtsreform⁶ verarbeitet.

Ziele des Vorhabens

- Aufbau einer offenen Datenbasis strukturierter Informationen aus Stellenanzeigen, die für weitere Forschungs- und Entwicklungsvorhaben nachgenutzt werden kann
- Visualisierung der in bibliothekarischen Stellenangeboten enthaltenen Informationen, insbesondere der gefragten Kompetenzen, um unterschiedliche Nutzergruppen und Fragestellungen zu unterstützen und unter anderem dabei zu helfen, Open Science und die Digitalisierung der Wissenschaft als einen Aspekt veränderter Arbeitsweisen besser zu verstehen
- Konzeption und Erprobung des Workflows von Harvesting, Analyse, Informationsextraktion und Visualisierung
- Langfristige Ermöglichung einer fortlaufenden automatischen Analyse der in Openbibliojobs gelisteten Stellenanzeigen-Volltexte hinsichtlich geforderter Qualifikation sowie Visualisierung der Änderung von geforderten Kompetenzen
- Online-Veröffentlichung mit Visualisierung und Kommentierung der Ergebnisse unter einigen ausgewählten Hauptaspekten, mit dem Ziel, den Wandel der Infrastruktur angesichts sich wandelnder wissenschaftlicher Arbeitsweisen besser verständlich zu machen

Material und Methoden

In diesem als reines Entwicklungsprojekt konzipierten Vorhaben werden ausschließlich schon vorhandene Werkzeuge zum Harvesten, ML und Visualisieren weiterentwickelt und State-of-the-Art-Methoden angewendet. Es gibt keinen reinen Forschungsanteil. Das verwendete Material wird in den Arbeitspaketen referenziert. Die Anschubfinanzierung kann in den am Projekt mittel- und unmittelbar beteiligten Einrichtungen forschungslastige Folgeaktivitäten nach sich ziehen und einzelne Forschungsvorhaben erweitern, wie im Ausblick skizziert.

⁴ <https://jobs.openbiblio.eu/about>

⁵ Ab Juli 2012 bis Ende August 2018 wurden bei OpenBiblioJobs knapp 11700 Stellenangebote veröffentlicht

⁶ *Wissenschaftler erhalten künftig das Recht, Datenbanken, Zeitschriften und andere urheberrechtlich geschützte Werke, zu denen ein legaler Zugang besteht [...] automatisiert auszulesen, zu speichern und auszuwerten (§ 60d UrhG). Die Daten können dabei auch aus verschiedenen Quellen stammen.* (vgl. 24.07.2017 Die Urheberrechtsreform: Das ändert sich für Bibliotheken zum 1. März 2018, dazu auch Kamocki et al. (2018). Information der dbv-Rechtskommission, https://www.bibliotheksverband.de/fileadmin/user_upload/Kommissionen/Kom_Recht/Rechtsinformationen/2017_07_24_Urheberrechtsreform_%C3%84nderungen_zum_01.03.2018.pdf)

Vorarbeiten

- Export von in OpenBiblioJobs aggregierten Metadaten von Stellenausschreibungen, Publikation dieser Daten in Zenodo, vgl. Blümel et al. (2017), Zellmann (2018a)
- Gewinnung strukturierter Metadaten aus Volltexten, vgl. Zellmann (2018b)
- Manuelles Mapping von Stellenanzeigen-Inhalten auf Kompetenzfelder, speziell für IT-lastige Bibliothekstätigkeiten, vgl. Zellmann (2018d), Nutzung als initiales Kompetenz-Vokabular (vgl. AP 3)
- Sechs Jahre Betrieb der Plattform OpenBiblioJobs
- Gemeinsames Projekt TextTransfer⁷ mit vergleichbarer Rollenaufteilung zwischen TIB und IDS und bewährter Zusammenarbeit, Möglichkeit der Nachnutzung einzelner Komponenten, z.B. Text-Preprozessierung (vgl. AP 2)
- erste Experimente an der TIB mit den OS Werkzeugen Elasticsearch/Kibana⁸: Mappen und Indexieren von Daten aus dem gemeinsamen Projekt TextTransfer und Visualisieren im Kibana-Dashboard
- IDH: Projekte "[Informationsextraktion aus Stellenanzeigen](#)" und "[Qualifikationsentwicklungsforschung](#)". Software: quenfo - Framework zur Informationsextraktion aus Stellenanzeigen.⁹

Zeit- und Arbeitsplan

Arbeitspakete

AP 1, M1, konzeptionelle Vorarbeiten: Dieses Arbeitspaket umfasst den konzeptionellen Aufbau der in den folgenden APs beschriebenen Pipeline aus Harvesting, Speicherung, Prozessierung und Zugänglichmachung. Insbesondere dient es zur Prüfung und Entscheidung, welche vorhandenen Workflows und Werkzeuge aus den Vorarbeiten beider Institutionen zur Anwendung kommen.

AP 2, M1-12, Aufbau des Stellenanzeigen-Textkorpus:

1. Alte Anzeigen: Harvesting von Volltexten der noch via Links in den OpenBibliojobs-Metadaten ab 2012 vorhanden sind sowie retrospektive Beschaffung von Anzeigen, deren Volltexte nicht mehr unter der in den Metadaten angegebenen Link-Adresse auffindbar sind unter Nutzung des Frameworks¹⁰ vom Internet Archive und L3S Hannover sowie unter Rückgriff auf das InetBib-Mailarchiv¹¹. Gegebenenfalls können weitere geeignete Quellen identifiziert werden. Ein Großteil der Arbeit wird durch ein selbst entwickeltes Harvesting-Tool erledigt. Manuelle Arbeit ist nötig zur Qualitätskontrolle und zum Finden von Anzeigen, die nicht durch das Tool gefunden werden können.
2. Neue Anzeigen: Harvesting von Volltexten der Anzeigen, die ab Projektstart publiziert werden. In beiden Schritten werden, soweit vorhanden, direkt verlinkte PDF-Dateien, HTML-Seiten und dort verlinkte PDF-Dateien heruntergeladen und archiviert. Seit Januar 2019 wird das [Amber-Plugin](#) für das Harvesting getestet. Dabei werden die Volltexte

⁷ <http://www1.ids-mannheim.de/direktion/fi/projekte/texttransfer.html>

⁸ <https://www.elastic.co/de/products/kibana>

⁹ <https://github.com/spinfo/quenfo>

¹⁰ <https://github.com/helgeho/ArchiveSpark>

¹¹ <http://www.inetbib.de/listenarchiv/>

automatisch zum Zeitpunkt der Veröffentlichung des Stellenangebots auf der OBJ-Seite gespeichert. Als Speicherort kann der OBJ-Server und/oder weitere externe Dienste wie bspw. das Internet Archive angegeben werden. Eine Aus- und Bewertung dieses Harvesting-Verfahrens steht noch aus.

Aufbauend auf die Vorarbeiten von Andreas Witt, Manu Raster, Jana Diesner et al. (2018), werden die Anzeigen für die weitere Verarbeitung im AP 4 in ein Plain-Text-Format überführt. Das Arbeitspaket umfasst ferner die Speicherung der Volltexte und der vorprozessierten Daten.

AP 3, M1-3: Festlegung Kompetenz-Klassen: Grundlage für das Kompetenz-Vokabular sind die Arbeiten von Zellmann (2018) sowie die umfangreiche Kategorisierung des Österreichischen Arbeitsmarkt-Service AMS¹².

AP 4, M2-12: Präprozessierung, Informationsextraktion und Klassifikation: Für die beiden ersten Aufgaben wird auf erprobte Workflows zurückgegriffen, die für die Prozessierung des im Projekt vorgesehenen Stellenanzeigenkorpus genutzt werden können: Für die Präprozessierung der Textdaten und Abschnittsklassifikation Hermes & Schandock (2016), für die Extraktion von Kompetenzen Geduldig (2017).

Für die Kategorisierung der Kompetenzen kann auf die Vorarbeiten von Binnewitt (2017) und Zellmann (2018x) zurückgegriffen werden, um eine Klassifikation in die festgelegten Klassen (AP3) zu realisieren.

AP 5, M6-12: Aufbereitung der Metadaten, Integration und Visualisierung: Dieses Arbeitspaket integriert die mit den extrahierten Kompetenzen aus AP4 angereicherten Metadaten in die bestehende Darstellung in Openbibliojobs. Die Arbeiten umfassen Strukturierung, Mapping auf Elasticsearch-Datenmodell und Indexierung. Zusätzlich soll basierend auf ersten Experimenten an der TIB unter anderem mit Daten aus dem gemeinsamen Projekt TextTransfer für OpenBiblioJobs ein interaktives visuelles Dashboard auf Basis der freien Software Kibana entwickelt werden, indem in den vernetzten Informationen zu den Stellenanzeigen gebrowst werden kann. Dabei können einerseits das limitierte Ausgangs-Metadatenset, als auch die vielfältige Ergebnisse des IDS zur Analyse und Extraktion von Informationen aus den Anzeigen eingebunden werden. Ziel ist ein zeitgemäßes Interface zum intuitiven Erkennen von Zusammenhängen und Trends, sowie das schrittweise Komponieren von Suchanfragen (Drill Down).

In einer Online-Veröffentlichung werden die Daten unter verschiedenen Haupt-Aspekten ausgewertet, visualisiert und kommentiert. Diese eigenständige Publikation wendet sich nicht nur an EntscheiderInnen an Bibliotheken und Informationsmanagement-Studiengänge, sondern stellt auch Bezüge her zum Wandel der wissenschaftlichen Informationsinfrastruktur.

Meilensteine

- M1 (nach 1 Monat): Pipeline-Konzept
- M2 (nach 3 Monaten): erste Version Textkorpus und Kompetenzklassen
- M3 (nach 6 Monaten): erster Klassifikations-Output
- M4 (nach 9 Monaten): erste Integration und Visualisierung in OpenBiblioJobs

¹² <http://www.ams.at/bis/bis/KompetenzstrukturBaum.php>

- M5 (nach 12 Monate): kompletten in AP1 skizzierten Workflow prototypisch implementiert sowie Online-Veröffentlichung der Ergebnisse

Ausblick auf mögliche Folge- und parallele Projektaktivitäten, Nutzen für die Community

- Nutzung des vorklassifizierten Korpus für Extraktion weiterer semantischer Informationen und Überführung Wissensgraphen, vgl. Forschungsaktivitäten an der TIB für einen Open Research Knowledge Graph¹³
- Nutzung des Kompetenzmonitors für die bedarfsgerechte Anpassung der Curricula bibliothekarischer/informationwissenschaftlicher Studiengänge, vgl. Diskussion in Blümel, Walther et al (2018).
- Integration der Softwarekomponenten und Projektergebnisse in das Langzeitprojekt "Qualifikationsentwicklungsforschung" von IDH und dem Bundesinstitut für Berufsbildung (BIBB)
- Weitere Untersuchungen auf der Basis von (Zellmann 2018d) zur tariflichen Einstufung und Befristung der Stellen und die Verbindung zu den geforderten Kompetenzen
- Für die von den beiden großen bibliothekarischen Berufsverbänden in Deutschland (BIB und VDB) betriebene Website library-training.de ist, laut Aussage der VDB-Vorsitzenden Konstanze Söllner, eine inhaltliche Klassifizierung der Bildungsangebote gewünscht, die dem in den Stellenanzeigen ausgedrückten Bedarf entspricht.
- Daten und Pipeline kann für Folgeuntersuchungen verwendet werden
- Nutzen für die Bibliotheks-Community
- Veröffentlichung des Verbundes über Anpassung von Infrastrukturen in Deutschland an sich wandelnde wissenschaftliche Arbeitsweisen

Das Projekt ist explizit ein **Entwicklungsprojekt mit State-of-the-Art-Methoden**, das in den einschlägigen Förderlinien nicht platziert werden kann. Die Ergebnisse des Projektvorhabens kommen direkt der Community zugute, Daten und Code sollen offen zur Nachnutzung bereitgestellt werden und zu weiterer Forschung und Entwicklung anregen.

Referenzen

- Johanna Binnewitt (2017): [Machine-Learning-Verfahren zur Klassifikation von Stellenausschreibungen](#). Bachelor Thesis, Informationsverarbeitung, Institut für Digital Humanities, Universität zu Köln.
- Ina Blümel, Manu Raster, Pascal Ngoc Phu Tu, & Cedrik Zellmann. (2017). OpenBiblioJobs Export 2012-2016 [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.322783>
- Ina Blümel, Tatiana Walther, Cedrik Zellmann, Christian Hauschke, Christian Wartena, Lydia Hahn. (2018). FIS-Curriculum – Bedarfe zur Ausbildung künftiger Forschungsinformationsmanager [Poster]. [Urn:nbn:de:0290-opus4-32963](https://nbn-resolving.org/urn:nbn:de:0290-opus4-32963)

¹³

<https://www.tib.eu/de/service/aktuelles/detail/tib-veroeffentlicht-positions-papier-zu-open-research-knowledge-graph/>

- Markus Caspers. (2016). Der Systembibliothekar an deutschen Bibliotheken (Bachelorarbeit). Fachhochschule Potsdam, FB 5 Informationswissenschaften. [Urn:nbn:de:kobv:525-13906](https://nbn-resolving.org/urn:nbn:de:kobv:525-13906)
- Tanja Erdmenger, Ulrike Krass, Karin Langenkamp, Jens Winalke (2018). Neue Perspektiven im Personalmarketing. Wie Bibliotheken ihre Mitarbeitenden finden - <https://nbn-resolving.org/urn:nbn:de:0290-opus4-157599>
- Alena Geduldig (2017): [Muster und Musterbildungsverfahren für domänenspezifische Informationsextraktion. Ein Bootstrapping-Ansatz zur Extraktion von Kompetenzen aus Stellenanzeigen](#). Master Thesis, Informationsverarbeitung, Institut für Digital Humanities, Universität zu Köln.
- Jürgen Hermes & Manuel Schandock (2016): "[Stellenanzeigenanalyse in der Qualifikationsentwicklungsforschung. Die Nutzung maschineller Lernverfahren zur Klassifikation von Textabschnitten](#)." In: Fachbeiträge in Internet, Bundesinstitut für Berufsbildung.
- Pawel Kamocki, Erik Ketzan, Julia Wildgans, Andreas Witt (2018): "Das neue "Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft" und seine Auswirkungen für Digital Humanities" In *DHd 2018. Kritik der digitalen Vernunft. Konferenzabstracts*. <http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf>
- Mandy Neumann (2015): "[Analyse von Anforderungsprofilen. Eine Studie zur Informationsextraktion aus Stellenanzeigen](#)." Master Thesis, Informationsverarbeitung, Institut für Digital Humanities, Universität zu Köln.
- Konstanze Söllner (2016). Management-Abschluss oder Fachlaufbahn – Wohin entwickeln sich Anforderungsprofile und Karrierewege im wissenschaftlichen Bibliothekswesen? - <https://doi.org/10.5282/o-bib/2016H4S257-270>
- Maximiliane Schwerdtfeger (2017). Forschungsdatenmanagement : Anforderungen des Arbeitsmarktes und Angebote der informationswissenschaftlichen Studiengänge - <https://nbn-resolving.org/urn:nbn:de:0290-opus4-31579>
- Andreas Witt, Manu Raster, Jana Diesner, Diana Steffen, Rezvaneh Rezapour, Jutta Bopp, Norman Fiedler, Christoph Köller, und Jennifer Wockenfuss. (2018). „Impact of Scientific Research beyond Academia: An Alternative Classification Schema“. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, herausgegeben von Jana Diesner, Georg Rehm, und Andreas Witt. Paris, France: European Language Resources Association (ELRA), 7-12. http://lrec-conf.org/workshops/lrec2018/W6/summaries/5_W6.html
- Cedrik Zellmann. (2018a). OpenBiblioJobs Export 2012.07.17-2017.09.01 [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1212306>
- Cedrik Zellmann. (2018b). Metadaten aus Stellenausschreibungen von Bibliothekaren und Informationswissenschaftlern mit IT-Schwerpunkt in Deutschland 2012.07.17-2017.09.01 [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1212330>

- Cedrik Zellmann. (2018c). IT-Inhalte der Modulhandbücher und Studienordnungen bibliotheks- und informationswissenschaftlicher Studiengänge in Deutschland (Stand 07/2018) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1320163>
- Cedrik Zellmann (2018d): Bibliothekare und Informationswissenschaftler mit IT-Schwerpunkt in Deutschland, Bachelorarbeit, <https://doi.org/10.25968/opus-1264>
- Cedrik Zellmann, Ina Blümel. (2018). Systems librarian, IT librarian, data librarian – Demand for graduates in Germany, Austria and Switzerland: a quantitative job advertisement analysis, in *The Future of Education in Information Science, Proceedings from FEIS – International EINFOSE Symposium 10–11 September 2018 Pisa, Italy*, pp 187-197, <http://einfose.ffos.hr/feis-2018/proceedings>