

Einsatz von Text Analytics zur Unterstützung literaturintensiver Forschungsprozesse – Konzeption, Realisierung und Lessons Learned

Frank Bensberg
Fakultät Wirtschafts- und
Sozialwissenschaften
Hochschule Osnabrück
Caprivistrasse 30a
49076 Osnabrück
E-Mail:
f.bensberg@hs-osnabrueck.de

Gunnar Auth
Department Wirtschaft
Hochschule für
Telekommunikation Leipzig
Gustav-Freytag-Str. 43-45
04277 Leipzig
E-Mail:
auth@hft-leipzig.de

Christian Czarnecki
Department Wirtschaft
Hochschule für
Telekommunikation Leipzig
Gustav-Freytag-Str. 43-45
04277 Leipzig
E-Mail:
czarnecki@hft-leipzig.de

ABSTRACT

Das anhaltende Wachstum wissenschaftlicher Veröffentlichungen wirft die Fragestellung auf, wie Literaturanalysen im Rahmen von Forschungsprozessen digitalisiert und somit produktiver realisiert werden können. Insbesondere in informationstechnischen Fachgebieten ist die Forschungspraxis durch ein rasant wachsendes Publikationsaufkommen gekennzeichnet. Infolgedessen bietet sich der Einsatz von Methoden der Textanalyse (Text Analytics) an, die Textdaten automatisch vorbereiten und verarbeiten können. Erkenntnisse entstehen dabei aus Analysen von Wortarten und Subgruppen, Korrelations- sowie Zeitreihenanalysen. Dieser Beitrag stellt die Konzeption und Realisierung eines Prototypen vor, mit dem Anwender bibliographische Daten aus der etablierten Literaturdatenbank EBSCO Discovery Service mithilfe textanalytischer Methoden erschließen können. Der Prototyp basiert auf dem Analyse-System IBM Watson Explorer, das Hochschulen lizenzkostenfrei zur Verfügung steht. Potenzielle Adressaten des Prototypen sind Forschungseinrichtungen, Beratungsunternehmen sowie Entscheidungsträger in Politik und Unternehmenspraxis.

SCHLÜSSELWÖRTER

Text Analytics, Literaturanalyse, Forschungsprozess, EBSCO Discovery Service, IBM Watson Explorer

PROBLEMSTELLUNG

Die Wirtschaftsinformatik ist als wissenschaftliche Disziplin dadurch gekennzeichnet, dass sie über ein hohes Maß an Dynamik in Bezug auf die fokussierten Forschungsthemen verfügt, die letztlich auch durch Moden und Trends geprägt werden [SRRM09]. Als aktuelles Beispiel hierzu ist das Themenfeld der Digitalisierung zu fixieren [BBC18], das mittlerweile auch kritisch reflektiert wird [RBH+17]. Weitere Beispiele für dieses Phänomen gehen auch aus dem Hype Cycle for Emerging Technologies von Gartner hervor, der regelmäßig neue Themenfelder für potenzielle Forschungsaktivitäten im Gegenstandsbereich der Wirtschaftsinformatik identifiziert.

Diese disziplinäre Dynamik stellt besondere Herausforderungen für die wissenschaftliche Bearbeitung solcher

Themenfelder. So erfordert die Bearbeitung eines neuartigen Themenfelds das Erfassen des aktuellen Forschungsstands in der Literatur, um offene Forschungsfragen und Gestaltungsbedarfe identifizieren zu können. Zu diesem Zweck sind systematische Literaturrecherchen durchzuführen, die – insbesondere bei Modethemen – zu einer großen Anzahl zu verarbeitender Literaturquellen führen. Zusätzlich erschwerend wirkt sich dabei der interdisziplinäre Charakter der Wirtschaftsinformatik aus, durch den sich der Suchraum für relevante Literaturquellen weiter vergrößert [SSS15]. Weiterhin sind Literaturquellen nicht ausschließlich akademischen Ursprungs, sondern entstammen bisweilen auch der betrieblichen Praxis und liefern empirische Befunde zur Anwendung und Verbreitung von Informationssystemen sowie deren zugrunde liegenden Informationstechnologien. Wie BORNMANN und MUTZ feststellen, verdoppelt sich generell die Anzahl der wissenschaftlichen Publikationen innerhalb von ca. 24 Jahren bei einer jährlichen Wachstumsrate von etwa 3 % [BoMu15], sodass auch in Zukunft mit weiterhin stark wachsenden Literaturbeständen zu rechnen ist.

Angesichts dieser Entwicklung ist zu konstatieren, dass die Forschungspraxis im Hinblick auf die systematische Verarbeitung des zunehmenden Literaturangebots vor großen Herausforderungen steht. So sind Forschungsprozesse vielfach dadurch gekennzeichnet, dass Aufgaben der Literatursuche und -verwaltung zwar in hohem Maße digitalisiert sind – etwa durch die Nutzung etablierter Literaturdatenbanken und Literaturverwaltungsprogramme – indes die inhaltliche Analyse und Synthese von Fundstellen weitestgehend manuell erfolgen [StSu17].

Als Unterstützung dieser intellektuell anspruchsvollen Aktivitäten zur Verarbeitung von Literaturdaten eignen sich textanalytische Methoden (Text Analytics) [AABG16], die in der Lage sind, bislang unbekannte Muster und Zusammenhänge in Textdaten zu identifizieren [FeSa06]. Das Spektrum textanalytischer Methoden reicht dabei von einfachen lexikometrischen Verfahren (z. B. Frequenz- und Konkordanzanalysen für unterschiedliche Wortformen) bis hin zu komplexen, multivariaten Ansätzen zur Segmentierung und Klassifikation von Textdokumenten mithilfe maschineller Lernverfahren [DHH17].

Im Rahmen dieses Beitrags wird ein Prototyp zur textanalytischen Erschließung von Literaturdaten aus dem

EBSCO Discovery Service [Bart15] vorgestellt, der von zahlreichen Hochschulbibliotheken als Suchumgebung für Literaturrecherchen eingesetzt wird. Zur Analyse der dort verfügbaren Literaturdaten wird das Textanalyse-system IBM Watson Explorer [ZFG+14] verwendet, das Hochschulen lizenzkostenfrei zur Verfügung steht und Endanwender bei der explorativen Analyse von Fachtexten mit unterschiedlichen Methoden interaktiv unterstützen kann. Zur konzeptionellen Fundierung dieses gestaltungsorientierten Ansatzes werden im Folgenden zunächst die zentralen Ziele und Aufgaben der Literaturanalyse vorgestellt. Darauf aufbauend werden die informatorischen Potenziale der Lösung anhand von Realdaten demonstriert. Der Beitrag schließt mit der Diskussion der Lessons Learned, die aus dem Systemeinsatz resultieren.

Ziele und Aufgaben der Literaturanalyse

Literaturanalysen dienen in Forschungsprozessen grundsätzlich dazu, den aktuellen Stand (State of the Art, State of Research) oder die historische Entwicklung der Forschung in einem thematisch abgegrenzten Forschungsfeld widerzuspiegeln [BoDö16]. Zur zweckorientierten Konfiguration von Literaturanalysen hat sich die Taxonomie von COOPER [Coop88] etabliert, die in Abbildung 1 anhand eines morphologischen Kastens dargestellt wird.

Merkmal	Ausprägung			
	Forschungsergebnisse	Forschungsmethoden	Theorien	Praktiken/Anwendungen
Zielsetzung	Integration	Kritik		
Perspektive	Neutrale Darstellung		Einnahme einer argumentativen Position	
Abdeckungsgrad	Vollständig	Vollständig mit selektiver Zitation	Repräsentativ	Zentral
Organisation	Historisch	Konzeptionell	Methodisch	
Zielgruppe	Einzelwissenschaft	Wissenschaft (allgemein)	Praxis oder Politik	Öffentlichkeit

Abbildung 1: Taxonomie für Literaturanalysen nach [Coop88]

Demzufolge sind die folgenden sechs Merkmale zur Klassifikation von Literaturanalysen heranzuziehen:

- Der *Fokus* ergibt sich durch die Ausrichtung auf unterschiedliche Elemente realwissenschaftlicher Forschung. Hierbei können z. B. Ergebnisse, Methoden, Theorien sowie bestimmte Anwendungsfelder von Forschungsergebnissen im Vordergrund stehen.
- Die *Zielsetzung* kann einerseits in der zusammenführenden Darstellung der verfügbaren Literatur bestehen (Integration), auf die Kritik anhand ex ante formulierter Kriterien abstellen und schließlich auch die Identifikation zentraler Problemfelder in der Forschungslandschaft intendieren.
- Die Erarbeitung einer Literaturanalyse kann einerseits aus einer neutralen *Perspektive* erfolgen, andererseits auch durch eine gegebene, argumentative Positionierung geprägt sein.
- Mit dem *Abdeckungsgrad* wird zum Ausdruck gebracht, in welchem Umfang die existierende

Literatur untersucht wird. Eine vollständige Abdeckung ist in der Regel nur bei hochspezialisierten Forschungsfeldern möglich, während allgemeine Forschungsthemen häufig eine Einschränkung auf zentrale Literaturquellen (z. B. A/B-Journals) erfordern.

- Das Merkmal *Organisation* liefert Ankerpunkte für die Gruppierung von Literaturquellen im Zuge der Analyse, z. B. über deren historische Entwicklung, die gemeinsame theoretisch-konzeptionelle Fundierung sowie das eingesetzte Methodenspektrum.
- In Bezug auf die *Zielgruppe* von Literaturanalysen kommen nicht nur wissenschaftliche Akteure infrage, sondern vielmehr auch Praktiker, politische Entscheidungsträger sowie die breite Öffentlichkeit.

Zur operativen Durchführung von Literaturanalysen sind eine Reihe von Aktivitäten zu verknüpfen, die auch Aufgaben der Recherche und Beschaffung von relevanter Literatur zum Gegenstand haben. Abbildung 2 zeigt diese Aktivitäten aufbauend auf der Darstellung in [BoDö16] als BPMN 2.0-Prozessdiagramm.

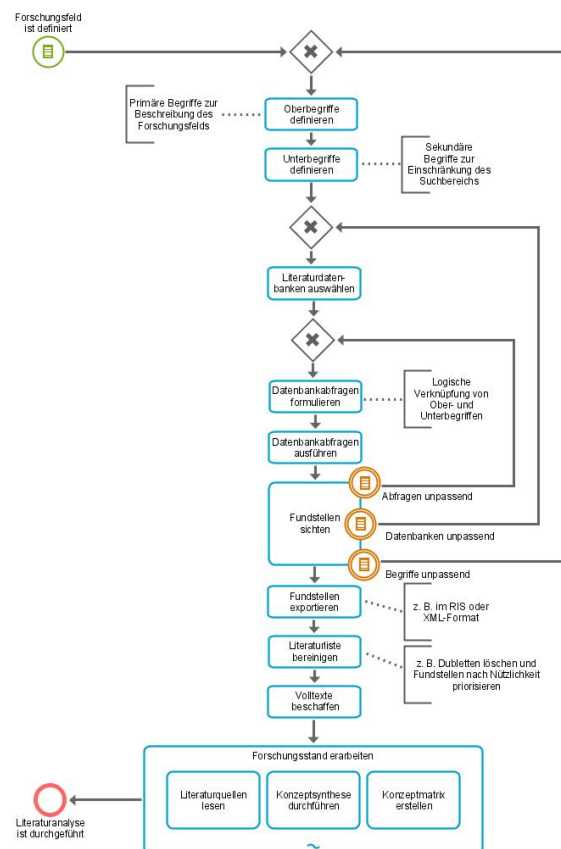


Abbildung 2: Literaturanalyseprozess in Anlehnung an [BoDö16] (BPMN 2.0-Prozessdiagramm)

Der Literaturanalyseprozess beginnt mit der Abgrenzung eines Forschungsfelds, das anschließend mithilfe eines Sets thematischer Schlagwörter bzw. Stichwörter zu beschreiben ist. Oberbegriffe sind primäre Suchbegriffe, die das Forschungsfeld generell beschreiben, während

Unterbegriffe als sekundäre Suchbegriffe das Forschungsfeld einschränken. Anschließend sind geeignete Literaturdatenbanken auszuwählen, in denen die Recherche nach relevanter Literatur erfolgen soll (z. B. Google Scholar, EBSCO Discovery Service, SpringerLink). Für diese Literaturdatenbanken werden geeignete Datenbankabfragen formuliert und anschließend ausgeführt. Die Sichtung der resultierenden Fundstellen kann dazu führen, dass die gewählten Begriffe zur Konzeptualisierung des Forschungsfelds zu überarbeiten oder aber das Set an Literaturdatenbanken bzw. deren Abfragen zu modifizieren sind.

Nach erfolgreicher Ausführung der Abfragen liegen Fundstellen vor, die zur Weiterverarbeitung – z. B. mithilfe gängiger Literaturverwaltungssoftware – exportiert werden. In der Regel werden dabei die zentralen bibliografischen Angaben sowie die Kurzzusammenfassung (Abstract) exportiert, zudem stehen bisweilen auch Volltexte in Literaturdatenbanken zur Verfügung. An den Export schließen sich Aktivitäten zur Bereinigung der Literaturliste (z. B. Dubletteneliminierung, Priorisierung) und zur Beschaffung fehlender Volltexte an. Abschließend ist der Forschungsstand zu erarbeiten und zu dokumentieren, wobei in hohem Maße strukturierende und bewertende Aktivitäten erforderlich sind.

Eine Kernaufgabe bei der Erarbeitung des Forschungsstands besteht in der Identifikation und Interpretation forschungsleitender Konzepte, die zur Artikulation und Präzisierung weiterführender Forschungsbedarfe erforderlich sind. Zur Schaffung eines strukturierten Gesamtüberblicks über die Konzepte des Forschungsfelds schlagen Webster und Watson die Konstruktion einer Konzeptmatrix vor [WeWa02], die mittlerweile auch in Literaturanalysen zur Wirtschaftsinformatik Verbreitung findet (siehe z. B. [SeNi16], [LEZ13]). Eine Konzeptmatrix bildet in ihrer einfachsten Form die Unterstützung unterschiedlicher Konzepte durch die untersuchten Literaturquellen eines Forschungsfelds ab (Abbildung 3).

Literaturquellen	Konzepte			
	A	B	C	[...]
Quelle 1	•	•	•	
Quelle 2		•		•
[...]			•	

Abbildung 3: Struktur einer Konzeptmatrix nach [WeWa02]

Zur Konstruktion der Konzeptmatrix ist ein systematisches Lesen der Literaturquellen durchzuführen, sodass sämtliche relevanten Konzepte identifiziert werden können. Zu diesem Zweck werden Lesetechniken eingesetzt, wie z. B. die SQ3R-Methode [Robi70]. Nach Abschluss der Lesephase findet die Synthese der identifizierten Konzepte statt, die nach fachlogischen Gesichtspunkten in einem Bezugsrahmen zu integrieren sind.

Auf der sprachlichen Ebene handelt es sich bei Konzepten um Fachausdrücke als Elemente von *Fachterminologien*, die durch Substantive bzw. komplexere Substantivkonstruktionen (z. B. Substantivfolgen oder Adjektiv-

Substantiv-Kombinationen wie etwa *Smart Service Engineering*) gebildet werden [HQW12]. Mithilfe textanalytischer Verfahren kann die explorative Identifikation solcher sprachlicher Strukturen unterstützt werden, sodass die Erarbeitung von Konzeptmatrizen durch geeignete Software potenziell effizienter und effektiver realisiert werden kann. Im Folgenden wird daher ein Prototyp präsentiert, der textanalytische Methoden zur Erschließung von Literaturdaten bereitstellt.

Demonstration des Prototypen

Der Prototyp für die Anwendung von Text Analytics zur Unterstützung literaturintensiver Forschungsprozesse basiert auf einer datenbasierten Kopplung der beiden Softwareprodukte EBSCO Discovery Service (EDS) und IBM Watson Explorer (WEX). Im Literaturrechercheprozess kommen die beiden Komponenten nacheinander zum Einsatz. Zunächst wird mit EDS eine themenspezifische Literaturdatenbasis erzeugt, die anschließend mit WEX durch Anwendung von Text Analytics-Funktionalitäten verarbeitet wird.

Konfiguration des Prototypen

Als Datengrundlage für den Prototypen werden Literaturdaten genutzt, die mithilfe von EDS-Abfragen gewonnen werden. Die resultierenden Fundstellen können in unterschiedlichen Formaten (z. B. im Research Information System-Format oder XML-Format) exportiert werden. Als praktisches Anwendungsszenario wird dabei das exemplarische Forschungsfeld *Industrie 4.0 im Straßenverkehr* fixiert. Zur Schaffung einer Datenbasis wurde eine entsprechende Abfrage formuliert, die sämtliche Quellen lokalisiert, die in der Kurzfassung (Abstract) neben dem Oberbegriff *Car* auch die Unterbegriffe *Connected Car*, *Car-to-Car*, *C2C*, *C2X* oder *C2I* enthalten. Die entsprechende Abfrage für die Literaturdatenbank lautet somit:

```
(AB "car") AND (AB "connected " OR AB
"car-to-car"
OR AB "C2C" OR AB "C2I" OR AB "C2X")
```

Auf diese Weise wurden mit EDS 17.926 Fundstellen identifiziert und exportiert. Auf die Beschaffung der Volltexte wurde aufgrund des Volumens der Fundstellen verzichtet.

Zur Analyse der resultierenden Literaturdaten wird das Text Analytics-System WEX eingesetzt. Zum Import der Literaturdaten wird eine einfache XML-Struktur verwendet, die für jede Fundstelle folgende Attribute umfasst:

- Typ der Fundstelle (z. B. Journal, Book)
- Haupttitel
- Buchtitel
- Autor(en)
- Publikationsjahr
- Schlagwörter
- Kurzfassung (Abstract)

Abbildung 4 zeigt eine exemplarische Fundstelle im XML-Format.

Nach Import der Literaturdaten stellt dieses System eine Reihe von Analysemethoden zur Verfügung, die im Folgenden dargestellt werden.

```

1 <?xml version="1.0" encoding="UTF-8" ?>
2 <item>
3 <TY>BOOK</TY>
4 <TI></TI>
5 <AU>Jaeger, Attila</AU>
6 <YI>2016</YI>
7 <KW>Vehicular ad hoc networks (Computer networks)
Automobiles--Electronic equipment COMPUTERS / Online Services
SCIENCE / Environmental Science TECHNOLOGY + ENGINEERING /
Automotive</KW>
8 <AB>Attila Jaeger develops an application which notifies a vehicle's
driver of upcoming road weather dangers. This application maps the
information evaluated by in-vehicle sensors in order to draw
conclusions on the current weather condition. Comprehensive data
basis is gained by sharing information with other vehicles using
Car-to-X communication. In order to prove usability of the presented
approaches, the developed application and selected concepts are
implemented and deployed within the context of large scale Car-to-X
field operational trials simTD and DRIVE C2X. Car-to-X communication
is considered as the next major step towards a significant increase
in road safety and traffic efficiency.</AB>
9 <TI>Weather Hazard Warning Application in Car-to-X Communication :
Concepts, Implementations, and Evaluations</TI>
10 </item>

```

Abbildung 4: XML-Struktur einer exemplarischen Fundstelle

Suchanfragen

Die Literaturdaten können mit WEX durch eine Suchmaschine abgefragt werden. Abbildung 5 zeigt die trunkierte Suche nach dem Schlagwort *car service* mit der entsprechenden Ergebnismenge (n=574) und der markierten Verwendung des Schlagworts im Kontext des Abstracts der jeweiligen Literaturquelle (Keyword in Context, KWIC-Analyse). Dabei können auch komplexere Abfragen mit logischen Operatoren (NOT, AND, OR) formuliert werden, wobei die Suche auf einzelne Felder oder auch Dokumentsprachen (z. B. Englisch, Deutsch) beschränkt werden kann. Mithilfe dieser Analysemethode können Literaturdatenbestände in Bezug auf Begriffe für a priori definierte Konzepte überprüft werden, sodass z. B. eine theoriegeleitete Untersuchung der Literaturdaten möglich wird.

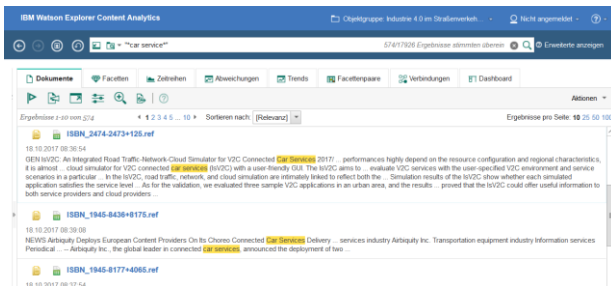


Abbildung 5: Suchanfrage und Trefferliste

Analyse von Wortarten und Subgruppen

Außerdem besteht die Möglichkeit, unterschiedliche Wortarten (z. B. Substantive, Verben) und Satzteile (z. B. Substantivfolgen, Verb-Substantiv- und Adjektiv-Substantiv-Kombinationen) zu identifizieren. Abbildung 6 zeigt einen Ausschnitt der in den Fundstellen enthaltenen Substantivfolgen, in denen der Begriff *car* auftritt, mit Angabe der absoluten Häufigkeiten (Frequenzanalyse). Auf diese Weise können frequente Konzepte (z. B. *car service*, *car platform*) in der Literaturdatenbasis identifiziert werden, sodass ein exploratives Instrument zur induktiven Ableitung potenziell relevanter Konzepte zur Verfügung steht.

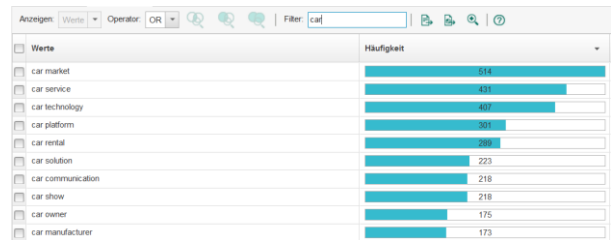


Abbildung 6: Gefilterte Frequenzanalyse von Substantivfolgen

Die Recherche in den Literaturdaten kann auch auf Subgruppen beschränkt werden. In diesem Fall besteht die Möglichkeit, Zusammenhänge zwischen Konzepten in den gebildeten Subgruppen zu identifizieren. Abbildung 7 zeigt Substantivfolgen für sämtliche Literaturquellen (n=301), in denen das Konzept *car platform* auftritt. Diese Liste ist nach der Kennzahl *Korrelation* absteigend sortiert, die mit Werten größer 1 eine besondere Häufung (Dichte) der jeweiligen Substantivfolge in der Subgruppe anzeigt. So wird aus dem Beispiel deutlich, dass das Konzept *car platform* mit Substantivfolgen wie *language understanding*, *voice recognition* und *car service platform* verknüpft ist.

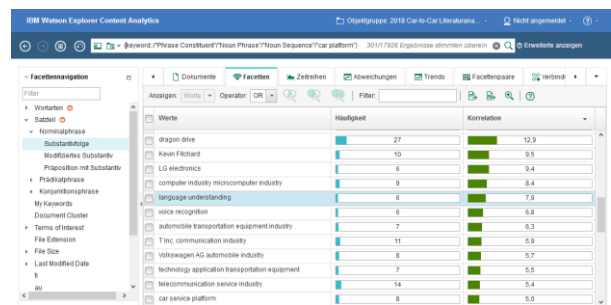


Abbildung 7: Subgruppenanalyse für Literaturquellen mit der Substantivfolge *car platform*

Zeitreihenanalyse

Mithilfe einer Zeitreihenanalyse kann die Fragestellung beantwortet werden, wie sich Publikationen im Zeitablauf entwickelt haben, wobei als Datumsattribut das Publikationsjahr verwendet wird. Dabei kann nach beliebigen Begriffen gefiltert werden, sodass die zeitliche Diffusion einzelner Konzepte transparent wird. Abbildung 8 verdeutlicht als Balkendiagramm die Zunahme an Publikationen für das hier untersuchte Forschungsfeld im Zeitablauf.

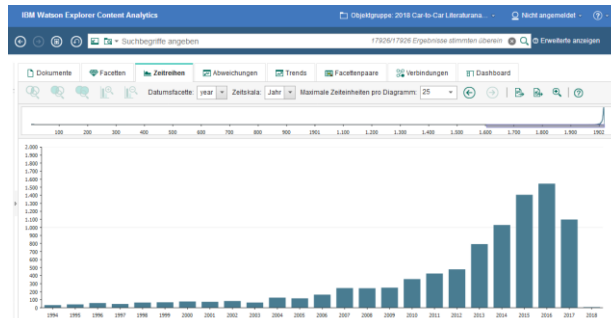


Abbildung 8: Zeitliche Entwicklung der Literaturbasis nach Publikationsjahr

Korrelationsanalyse

Durch Korrelationsanalysen können Zusammenhänge zwischen unterschiedlichen Konzepten der Fachdomäne ermittelt werden. Abbildung 9 zeigt ein Netz hochkorrelierter Substantivfolgen und modifizierter Substantive, aus denen fachliche Zusammenhänge zwischen den Konzepten deutlich werden. So deutet das Tripel *connected car*, *big data*, *smart city* darauf hin, dass die involvierten Konzepte einen thematischen Zusammenhang bilden, der Hinweise zur Formulierung neuer Forschungsfragen liefern kann.

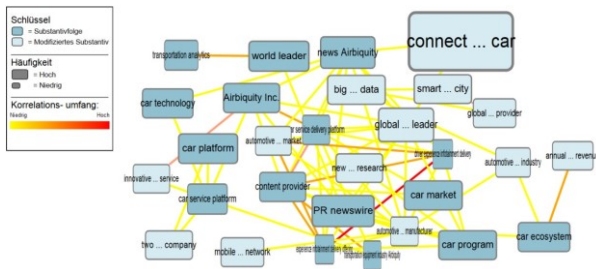


Abbildung 9: Korrelationsanalyse für Substantivfolgen und modifizierte Substantive

Lessons Learned

Der vorgestellte Prototyp zur Literaturanalyse hat sich mittlerweile in mehreren Forschungsprojekten als effektiv und effizient erwiesen, um den Status quo bestehender und emergierender Forschungsfelder in der Literatur zu untersuchen und deren zentrale Konzepte, Methoden und Praxisbezüge zu identifizieren. Mit Referenz auf die in Abbildung 1 dargestellte Taxonomie für Literaturanalysen ist festzuhalten, dass die textanalytische Unterstützung zunächst Potenziale zur Steigerung des *Abdeckungsgrads* von Literaturanalysen freisetzt. So können selbst umfangreiche Kollektionen von Literaturfundstellen mit mehreren hunderttausend Abstracts untersucht werden, womit eine qualitative Steigerung des resultierenden Forschungsoutputs zu erwarten ist. Außerdem gestatten die vorgestellten Funktionalitäten die Analyse der Literaturdatenbasis nach unterschiedlichen *Organisationskriterien*. So kann nicht nur die historische Entwicklung von Konzepten mithilfe der Zeitreihenanalyse nachverfolgt, sondern auch die Referenzierung konzeptionell-theoretischer Fundierungen und Forschungsmethoden in der Literatur detailliert substantiiert werden, z. B. durch Suchanfragen und Exploration von Wortarten. Aufgrund der Möglichkeit, umfangreiche Literaturbestände interaktiv zu sichten, ist das Instrumentarium nicht nur für Adressaten aus der Wissenschaft relevant, vielmehr eröffnen sich auch Potenziale zur Deckung des Informationsbedarfs von Entscheidungsträgern in Politik und Unternehmenspraxis als eine weitere *Zielgruppe*. So kann beispielsweise die analytische Fähigkeit, emergierende Entwicklungen anhand von Literaturdaten zu erkennen, zur Formulierung von Beratungsleistungen beitragen, um strategische Planungs- und Entscheidungsprozesse im Umfeld von Forschung und Entwicklung empirisch zu fundieren. Diese Fähigkeit ist auch vor dem Hintergrund

der digitalen Transformation von Beratungsdienstleistungen von Bedeutung [SeNi16]. Sich schnell verändernde ökonomische und technische Rahmenbedingungen stellen Beratungsunternehmen vor die Herausforderung, Analysen und Empfehlungen kontinuierlich anzupassen und dabei operative Tätigkeiten zunehmend zu virtualisieren.

Wie bisherige Projekterfahrungen belegen, ist die vorgestellte Lösung an Hochschulen für ein breites Spektrum von Anwendern aus unterschiedlichen Fachbereichen von Interesse. Hiermit wird die weiterführende Fragestellung aufgeworfen, wie textanalytische IT-Services für Hochschulakteure in der Breite verfügbar gemacht und in eine bestehende IT-Architektur integriert werden können. Im Hochschulkontext eröffnet sich zudem ein weiteres Anwendungsfeld für den beschriebenen Ansatz: Zur Erfüllung stetig zunehmender Berichts- und Auskunftspflichten gegenüber ihren Aufsichtsbehörden aber auch als Leistungsausweis im Wettbewerb um Drittmittel und Forschungsaufträge benötigen Hochschulen die Möglichkeit, sich jederzeit einen aktuellen und detaillierten Überblick über die eigenen Forschungsaktivitäten und -ergebnisse zu verschaffen [HeSc16]. Zu diesem Zweck werden *Forschungsinformationssysteme* eingesetzt, für die es derzeit allerdings noch kein nennenswertes Angebot an Standardsoftware gibt. Vor diesem Hintergrund bietet der vorgestellte Text Analytics-Ansatz weitere Einsatzpotenziale, insbesondere für die zu favorisierende Datenerfassung aus bereits bestehenden Informationssystemen.

LITERATUR

- [AABG16] Alpar, P., Alt, R., Bensberg, F., Grob, H. L., Weimann, P., Winter, R.: Anwendungsorientierte Wirtschaftsinformatik. Strategische Planung, Entwicklung und Nutzung von Informationssystemen. Springer Vieweg, Wiesbaden, 2016.
- [Bart15] Bartlakowski, K.: Make the library really look more like google: Zur Einführung des Discovery-Systems „scinos“ an der Hochschule Osnabrück. Bibliotheksdienst 6/15, S. 643–648, 2015.
- [BBC18] Bensberg, F., Buscher, G., Czarnecki, C.: Digital Transformation and IT Topics in the Consulting Industry – a Labor Market Perspective. In: Nissen, V. (Hrsg.) Advances in Consulting Research: Recent Findings and Practical Cases, Springer International Publishing.
- [BoDö16] Bortz, N., Döring, J.: Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften, 5. Aufl., Springer, Berlin 2016.
- [BoMu15] Bornmann, L., Mutz, R.: Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. J. Assoc. Inf. Sci. Technol., S. 2215–2222, 2015.

- [Coop88] Cooper, H.M.: Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews. *Knowledge in Society*, 1, 1988, S. 104–126.
- [DHH17] Dann, D., Hauser, M., Hanke, J.: Reconstructing the Giant: Automating the Categorization of Scientific Articles with Deep Learning Techniques. In: Leimeister, J.M. (Hrsg.), *Proceedings 13. Internationale Tagung Wirtschaftsinformatik (WI)*, S. 1538–1549.
- [FeSa06] Feldman, R., Sanger, J.: *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York 2006.
- [HeSc16] Herwig, S., Schlattmann, S.: Eine wirtschaftsinformatische Standortbestimmung von Forschungsinformationssystemen, in: Mayr, H. C., Pinzger, M. (Hrsg.): *INFORMATIK 2016, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2016*, S. 901-914.
- [HQW12] Heyer, G., Quasthoff, U., Wittig, T.: *Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse*. W3L GmbH, Bochum, 2012.
- [SeNi16] Seifert, H., Nissen, V.: Virtualisierung von Beratungsleistungen: Stand der Forschung zur digitalen Transformation in der Unternehmensberatung und weiterer Forschungsbedarf. *Multikonferenz Wirtschaftsinformatik (MKWI) 2016*, TU Ilmenau 09. - 11..03.2016.
- [LEZ13] Labes, S., Ere, K., Zarnekow, R.: Literatürübersicht von Geschäftsmodellen in der Cloud. In: *Wirtschaftsinformatik Proceedings 2013*, S. 443-457.
- [RBH+17] Riedl, R., Benlian, A., Hess, T., Stelzer, D., Sikora, H.: On the Relationship Between Information Management and Digitalization. *Bus Inf Syst Eng*. doi: 10.1007/s12599-017-0498-9.
- [Robi70] Robinson, F. P.: *Effective Study*, 4. Aufl., Harper & Row, New York 1970.
- [SRRM09] Steininger, K., Riedl, R., Roithmayr, F., Mertens, P.: Fads and Trends in Business and Information Systems Engineering and Information Systems Research – A Comparative Literature Analysis. *Bus Inf Syst Eng*. doi: 10.1007/s12599-009-0079-7.
- [SSS15] Sturm, B., Schneider, S., Sunyaev, A.: Leave No Stone Unturned: Introducing a Revolutionary Meta-search Tool for Rigorous and Efficient Systematic Literature Searches. *Proc. of the 23rd European Conference on Information Systems (ECIS 2015)*, Research-in-Progress Papers, Paper 34.
- [StSu17] Sturm, B., Sunyaev, A.: You Can't Make Bricks Without Straw: Designing Systematic Literature Search Systems. *Proc. of the 38th International Conference on Information Systems (ICIS 2017)*.
- [WeWa02] Webster, J., Watson, R.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly* 02/2002, S. XIII–XXIII, 2002.
- [ZFG+14] Zhu, W., Foyle, B., Gagné, D., Gupta, V., Magdalen, J., Mundi, A., Nasukawa, T., Paulis, M., Singer, J., Triska, M.: *IBM Watson Content Analytics: Discovering Actionable Insight from Your Content*, 3. Auflage, IBM, 2014.

KONTAKT

Prof. Dr. Frank Bensberg
Hochschule Osnabrück
Caprivistrasse 30a, 49076 Osnabrück
f.bensberg@hs-osnabrueck.de

Prof. Dr. Gunnar Auth
Hochschule für Telekommunikation Leipzig
Gustav-Freytag-Str. 43-45, 04277 Leipzig
auth@hft-leipzig.de

Prof. Dr. Christian Czarnecki
Hochschule für Telekommunikation Leipzig
Gustav-Freytag-Str. 43-45, 04277 Leipzig
czarnecki@hft-leipzig.de