

Eine Analyse der Fakultätsentwicklungspläne mittels NLP

Siegfried Handschuh, Simon Donig, Bernhard
Bermeitinger, Maria Christoforaki, Thomas Huber

Lehrstuhl für Digital Libraries und Web Information Systems

Die Ausgangsfrage, die im Rahmen dieses vorläufigen Berichts verfolgt wird, ist, inwieweit sich mit wissenschaftlichen Instrumenten aus dem Bereich des *Natural Language Processing* (NLP) in den Fakultätsentwicklungsplänen der Universität Passau Gemeinsamkeiten oder fakultätsübergreifende Themen identifizieren lassen.

Dem hier vorliegenden Bericht liegen vier Fakultätsentwicklungspläne (FEP) in der Fassung vom 04. August 2017 zugrunde, im Folgenden auch als das Korpus bezeichnet.

1 Problemaufriss

Der hier angewandte Zugriff kann als eine Form des »fernen Lesens« (*Distant Reading*) im Sinne des Literaturwissenschaftlers Franco Moretti¹ verstanden werden, das uns, ohne den Text als solchen gelesen zu haben, Auskunft über dessen Beschaffenheit vermittelt.

Typische Korpora im Bereich des NLP haben Wortzahlen von mehreren Millionen Worten bis hin zu mehreren Milliarden Worten. So hat beispielsweise

¹Franco Moretti: *Distant Reading*, New York / London: Verso, 2013.

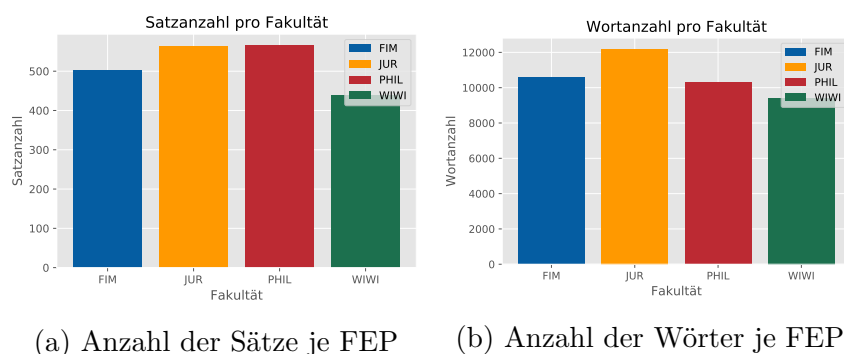


Abbildung 1: Anzahl der Sätze und Wörter je FEP

ein 2014 für Trainingszwecke erstelltes Korpus, das auf der englischen Wikipedia beruht, einen Umfang von 1,9 Milliarden Worten.² Im Vergleich dazu ist das untersuchte Korpus ausgesprochen klein, die Wortzahlen für die einzelnen Berichte liegen zwischen neun- und zwölftausend Worten (Verteilung siehe Abb. 1b), insgesamt rund 42 500. Hinzu kommt, dass dieses relativ kleine Korpus eine sehr spezifische Textgattung und damit Sprache repräsentiert.

Dieser Umstand ist deshalb von Bedeutung, weil die wissenschaftlichen Instrumente im Bereich der Verarbeitung und Analyse natürlicher Sprache vorwiegend mit statistischen Verfahren operieren und deshalb im Regelfall um so effizienter sind, je größer das vorliegende Korpus ist.

Das besondere Potential der Instrumente im Bereich des NLP liegt dabei darin, Beziehungen zwischen den Teilkorpora zu zeigen und Einblicke in deren Beschaffenheit zu vermitteln, die durch einfache Lektüre weitaus schwerer zu erlangen wären.

2 Ontologiebasierte Textanalyse

Da die Datengrundlage für einen rein statistischen Zugriff nicht ideal ist und zugleich eine spezifische Domänensprache vorliegt, bietet sich hier doppelter, Wissen repräsentierender und auf dieser Grundlage datengetrieben analysierender Zugriff an. Deshalb haben wir das an unserem Lehrstuhl entwickelte Instrument *Q-Kalkül* verwendet, das eine formalisierte Wissensrepräsentation in Form einer Ontologie mit der automatisierten Identifizierung der Manifestierung dieser Konzepte als Wortmuster im Text verbindet. Diese Technik erlaubt

²The Wikipedia Corpus, <https://corpus.byu.edu/wiki/>, zuletzt abgefragt am 30. Januar 2018

es, wichtige Themen für die einzelnen Fakultäten zu identifizieren.

Zu diesem Zweck haben wir eine Ontologie erstellt, die den Gegenstandsbereich *Universität* abbildet. Mit einer Ontologie können wir Bestandteile der uns umgebenden Lebenswelt in einer Weise formalisieren, die es einem Computer erlaubt, diese Repräsentation zu verarbeiten und auf dieser Grundlage zu rasonieren. Mittels der Ontologie wird es möglich, in einer sehr kontrollierten Weise die logischen Beziehungen zwischen diesen Bestandteilen der Lebenswelt zu modellieren (z.B. dass eine Universität aus Einheiten und Einrichtungen besteht. Einheiten können etwa die Fakultäten sein, die sich wiederum in Lehrstühle untergliedern usw.). Auf diese Weise kann komplexes Vorwissen strukturiert und für eine Untersuchung von Datenbeständen mit dem Computer zugänglich gemacht werden.

Für den Zweck der vorliegenden Fragestellung haben wir Konzeptfelder gebildet. Eines dieser Felder ist etwa der Bereich *Target*, dem wir beispielsweise Konzepte wie *Ausbildung*, *Empowerment*, *Fortbildung*, *Intersidziplinarität*, *Interkulturalität*, *Internationalisierung*, *Kommunikation*, *Kooperation*, *Mobilität*, *Qualität* und *Transfer* zugeordnet haben. Das Konzeptfeld umschreibt grundlegende Ziele, die die Universität als ganzes erreichen möchte. In ganz ähnlicher Weise haben wir grundlegende Forschungsinteressen aller Fakultäten identifiziert, die in Abb. 4 wiedergegeben werden. Ein weiteres Beispiel wäre der Bereich *Change Management*, dem wir Konzepte wie *Verbesserung*, *Veränderung*, *Wandel*, *Zukunft*, *Entwicklung*, *Neuheit*, *Ausrichtung*, *Vielfalt*, *Förderung*, *Modifikation*, *Entwicklung*, *Einfluss*, *Nachhaltigkeit* und *Maßnahme* zugeordnet haben. Jedes dieser Konzepte wird wiederum durch eine größere Anzahl von lemmatisierten Worten gemessen.

Die Texte können wir nun auf Aussagen hin untersuchen, die einem dieser Felder zugeordnet werden können. Auf diese Weise können wir die logische Struktur des Texts weitaus intensiver durchdringen, als bei einer rein auf einzelnen Worten basierten statistischen Analyse.

Sehr deutlich wird dies beispielsweise im Wortfeld *Change Management* (cf. Abb. 2). Hier zeigt sich zunächst, dass die Philosophische Fakultät von allen vier Fakultäten diejenige ist, die am intensivsten mit Veränderungen befasst ist. Sie hat nicht nur im Schnitt das häufigste Vorkommen der Konzepte aus diesem Wortfeld überhaupt, sondern auch einzelne Konzepte, die exklusiv in ihrem Teilkorpus vorkommen, wie beispielsweise *Modifikation*. Erkennbar befindet sich die Fakultät in einem Prozess der Restrukturierung, der auf die *Zukunft* ausgerichtet ist und etwa durch das häufige Vorkommen von Konzepten wie *Wandel* angezeigt wird. Da dieser Wandel moderiert und implementiert werden muss, kommt dem Konzept *Maßnahme* in der Philosophischen Fakultät ein weitaus größeres Gewicht zu als in allen anderen Fakultäten. Dies setzt die

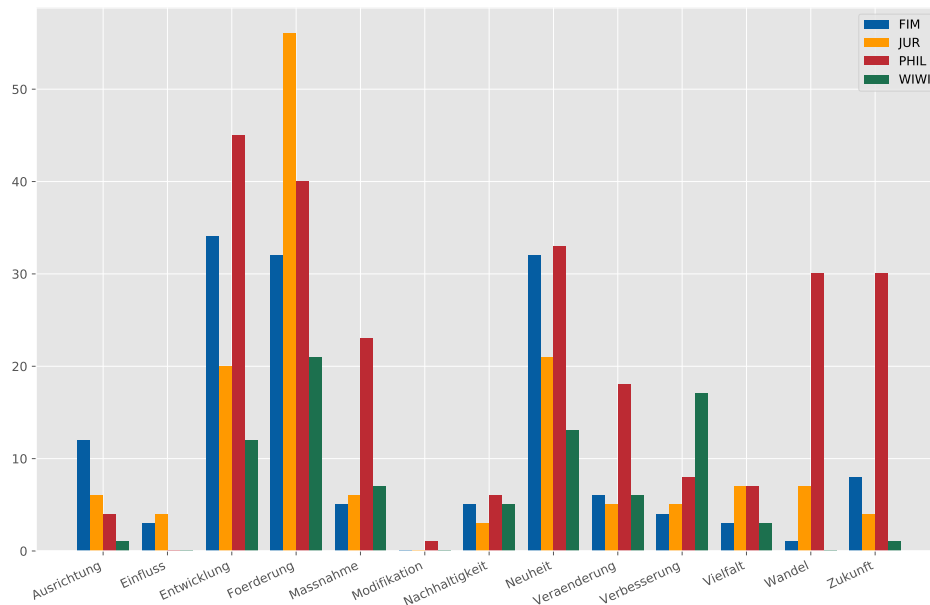


Abbildung 2: Verhältnis zwischen *Change Management* und dessen Objekten

Überwindung und Veränderung bestehender Strukturen und Praktiken voraus, was sich etwa in der Frequenz des Wortes *Neuheit* niederschlägt. Zugleich sollen diese Maßnahmen nachhaltig erfolgen, was das hohe Gewicht des Konzepts *Nachhaltigkeit* im FEP erklärt. Während für die Philosophische Fakultät folglich das Konzept der *Veränderung* eine bedeutende Rolle spielt, stellen andere Fakultäten, die diese Anpassungsleistung nicht erbringen müssen, mehr auf die inkrementelle *Verbesserung* ab.

Dieser Veränderungsprozess der im FEP der Philosophischen Fakultät deutlich wird, kann mit den Zielen korreliert werden, die die einzelnen Fakultäten für sich gesetzt haben. In Abb. 3 können wir neben den selbstverständlichen Kernaufgaben der Fakultäten wie *Forschung* und *Lehre* und deren internationaler Dimension weitere Konzepte erkennen, die zwischen mehreren Fakultäten geteilt werden. So lassen sich zunächst die Wortfelder *Empowerment* und *Nachwuchs* identifizieren, die Konzepte umgreifen, die Aspekte der Diversität und Chancengleichheit sowie der Nachwuchsförderung umfassen. Zugleich werden sekundäre Ziele wie die Sicherung der *Qualität* von Forschung und Lehre adressiert. Bemerkenswert ist die relativ gleichmäßige Verteilung des Wortfelds *Transfer*, in dem der horizontale und vertikale Transfer von Wissen, Praktiken und Technologien zusammengefasst wird.

In ähnlicher Weise können wir eine erste Erkundung der statistisch auffälli-

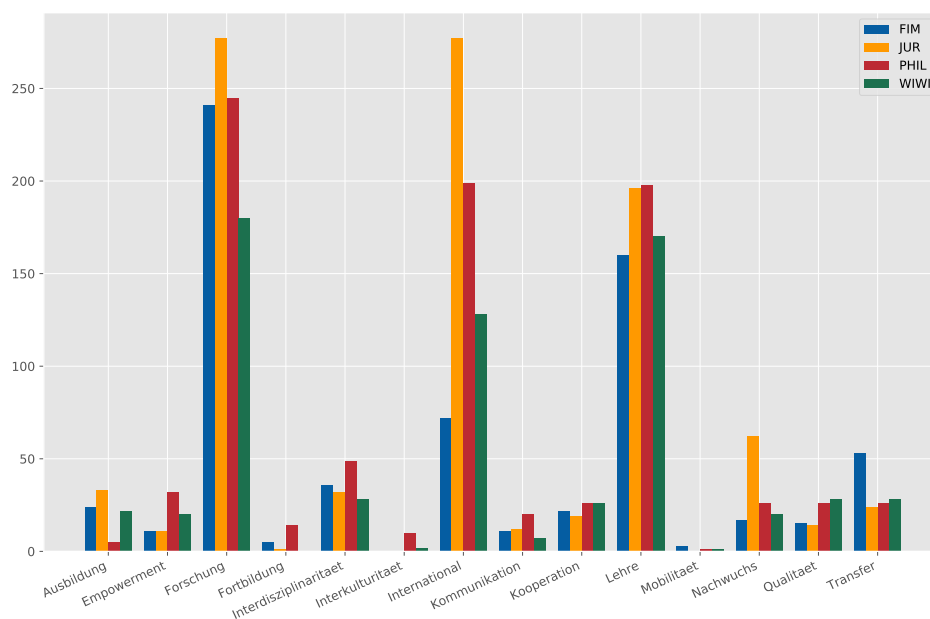


Abbildung 3: Verhältnis zwischen *Target* und dessen Objekten

gen Konzepte durchführen, mit denen die Fakultäten ihre Forschungsinteressen beschreiben. Abb. 4 zeigt einerseits Themen, die primär für einzelne Fakultäten von zentralem Interesse sind, die aber von anderen Fakultäten in begrenztem Umfang geteilt werden, darunter etwa *Kultur* (PHIL), *Wirtschaft* (WIWI), *Europa* (PHIL und JUR) sowie *Gesellschaft* (PHIL). Zugleich zeichnet sich mit *Digitalisierung* aber auch ein Bereich ab, der zwar rein von der Häufigkeit der Erwähnungen in den einzelnen FEPnen nicht mit diesen »Großthemen« mithalten kann, der aber erheblich viel gleichmäßiger gewichtet ist.

Eine Analyse der statistisch auffälliger Wortkombinationen offenbart noch weitere Konzepte, von denen wir annehmen dürfen, dass sie Forschungsfelder repräsentieren und die wir daraufhin befragen können, ob diese mit anderen Fakultäten geteilt werden. Sie umfassen beispielsweise *digital*, *wandel* (PHIL), *kultur*, *wert*, *identität* (PHIL), *vernetzte gesellschaft* (WIWI), *data analytics* (WIWI), *big data* (FIM) und *wissensbasierte bildverarbeitung* (FIM) und lassen sich unschwer dem Wortfeld Digitalisierung zuordnen.

Dies legt nahe, dass sich hier ein potentielles Kooperationsfeld abzeichnet, das in allen Fakultäten gleichermaßen auf Interesse stößt.

Der Wille der Fakultäten zur übergreifenden Zusammenarbeit wird in Abb. 3 deutlich. Die Häufung des Gebrauchs dieses Konzepts in der Philosophischen Fakultät sowie der Fakultät für Informatik und Mathematik legt nahe, dass

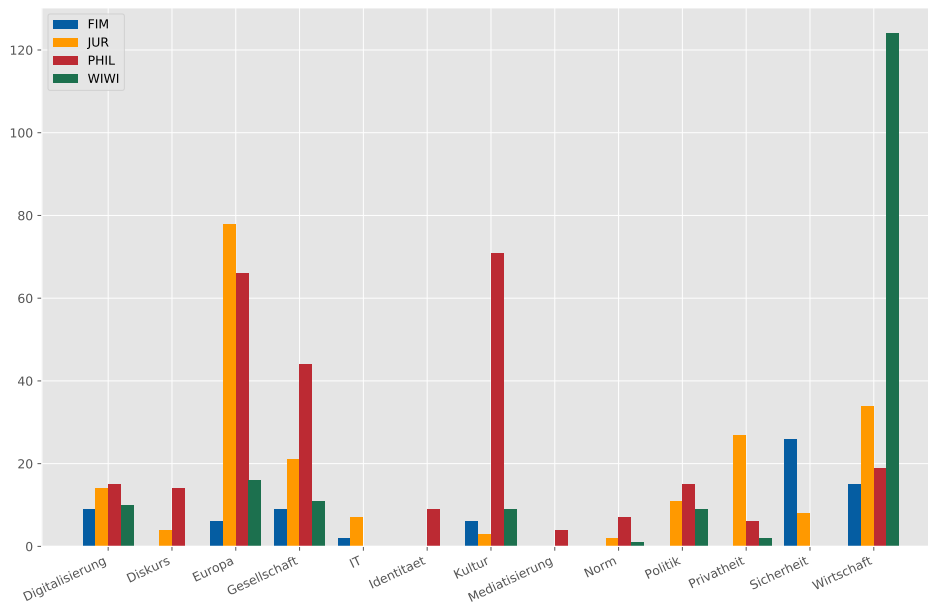


Abbildung 4: Verhältnis zwischen *POI* und dessen Objekten

beide aufgrund struktureller Faktoren besonders an übergreifenden Kooperationsprojekten arbeiten.

Mit dem an unserem Lehrstuhl entwickelten Instrument *Dragn* können Konzepte wie *Interdisziplinarität* in ihren Beziehungen zu anderen Wörtern interaktiv visualisiert und zugleich das Gewicht ihrer Bezüge zu diesen anderen Worten ermittelt werden (cf. Abb. 5). So können wir beispielsweise erfahren, dass *Interdisziplinarität* als ein *Potential* gesehen und erkennbar als *Standortvorteil* wahrgenommen wird, der besondere Bedeutung im Bereich der Förderung des wissenschaftlichen Nachwuchses und der Studiengangsstruktur hat und zugleich auch mit den Forschungsfragen der Fakultäten zusammenhängt.

3 Untersuchungen semantischer Nähe

Für eine Analyse der Bedeutung des Wortfelds *Digitalisierung* in den verschiedenen FEPen berechnen wir mit Hilfe des am Lehrstuhl entwickelten Projekts *Indra* die semantische Nähe dieser Konzepte zu allen anderen im Korpus existierenden Bigrammen, wobei der Grad der Nähe zwischen den Worten auf einer Skala von -1 bis $+1$ abgebildet wird. Wir haben uns hier, um die Komplexität überschaubar zu halten, zunächst auf die Begriffe *Digitaler Wandel* und *Digitalisierung* beschränkt.

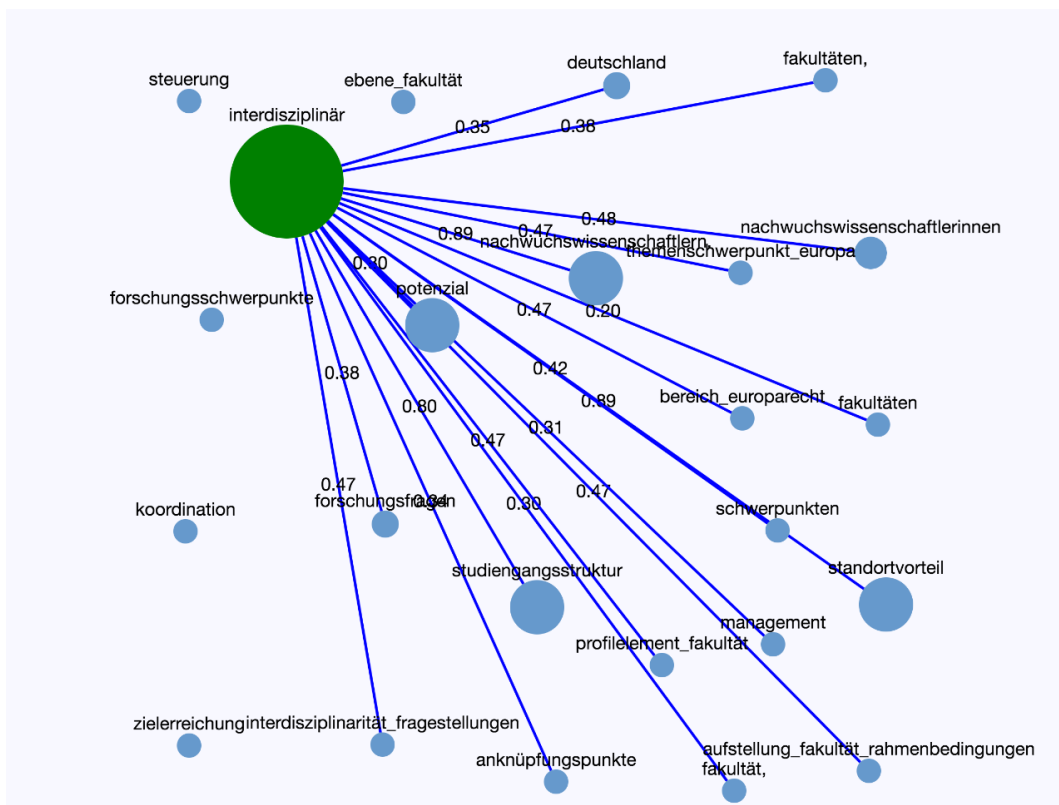


Abbildung 5: Visualisierung der Bezüge zwischen dem Konzept *Interdisziplinarität* und anderen Konzepten über alle FEPE

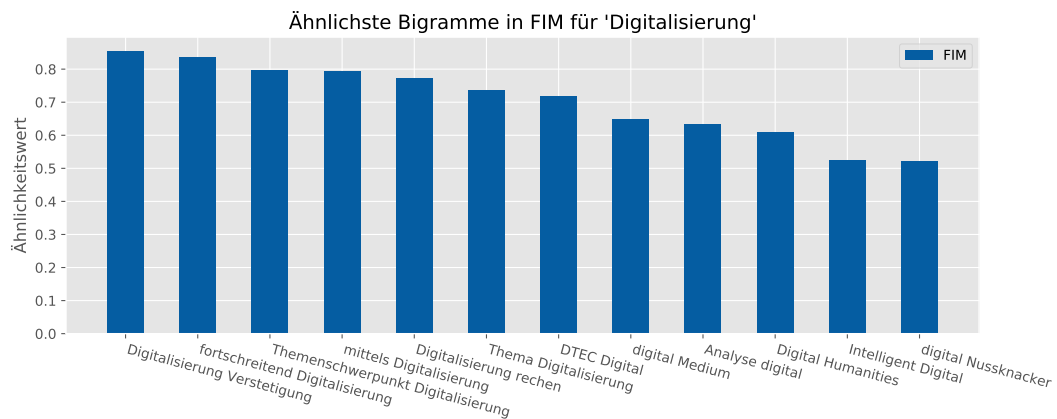


Abbildung 6

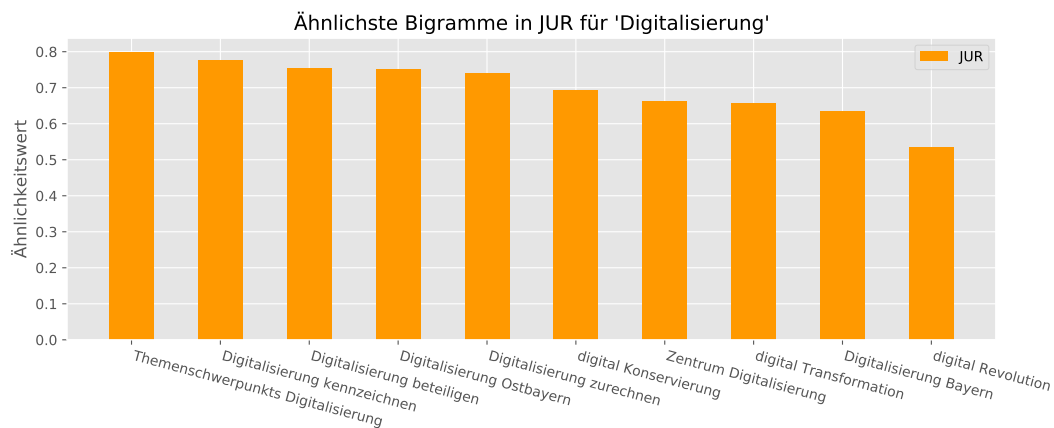


Abbildung 7

Wie Abb. 6 zeigt, wird die Digitalisierung in der Fakultät für Informatik und Mathematik primär als ein fortschreitender Prozess beschrieben, den es zu verstetigen und kontrollieren gilt. Zugleich werden die beiden Kernaufgaben der Disziplin artikuliert. Einerseits der anwendungsorientierte Bezug – Digitalisierung muss tatsächlich »gerechnet« werden – andererseits stellt dieser Prozess neue Herausforderungen an die Gewinnung von Wissen, wie das Bigramm *Analyse digital* nahe legt. Mindestens ebenso interessant ist, dass durch die semantische Nähe auch direkt ein Kooperationspartner erscheint – bieten sich das Feld der *Digital Humanities* hier gleichsam als natürlicher Partner für die Informatik an.

Im FEP der juristischen Fakultät (Abb. 7) finden sich demgegenüber sehr viele Referenzen, die auf die Abrubtheit und den Umbruchcharakter der Digitalisierung als Prozess hinweisen, etwa wenn diese als *digitale Revolution* bezeichnet wird und deutlich passiver als im FEP der FIM das Bemühen der Fakultät verdeutlicht wird, sich am Prozess zu *beteiligen*. Die geringe Zahl von Ausdrücken, die einen direkten Bezug zu Digitalisierung aufweisen, legt nahe, dass hier möglicherweise andere, alternative Beschreibungen angewandt werden. In der Tat offenbart ein traditionelles »Close Reading« des Texts, dass der Wortstamm *internet-* in vielen Fällen fast austauschbar gebraucht wird.

Interessant ist insbesondere eine Analyse der Verbindung der Konzepte *Internet* und *Wandel* mittels *Dragn*, die hier noch einmal das Selbstbild der Fakultät in diesem Prozess illustriert. Die Grafik zeigt zwei relativ disjunkte Graphen, deren einer das Internet als Infrastruktur zeigt, der andere die juristische Dimension des Umgangs mit dem Internet. Verbunden sind beide Graphen durch die Nutzung der Infrastruktur Internet, primär offenbar durch

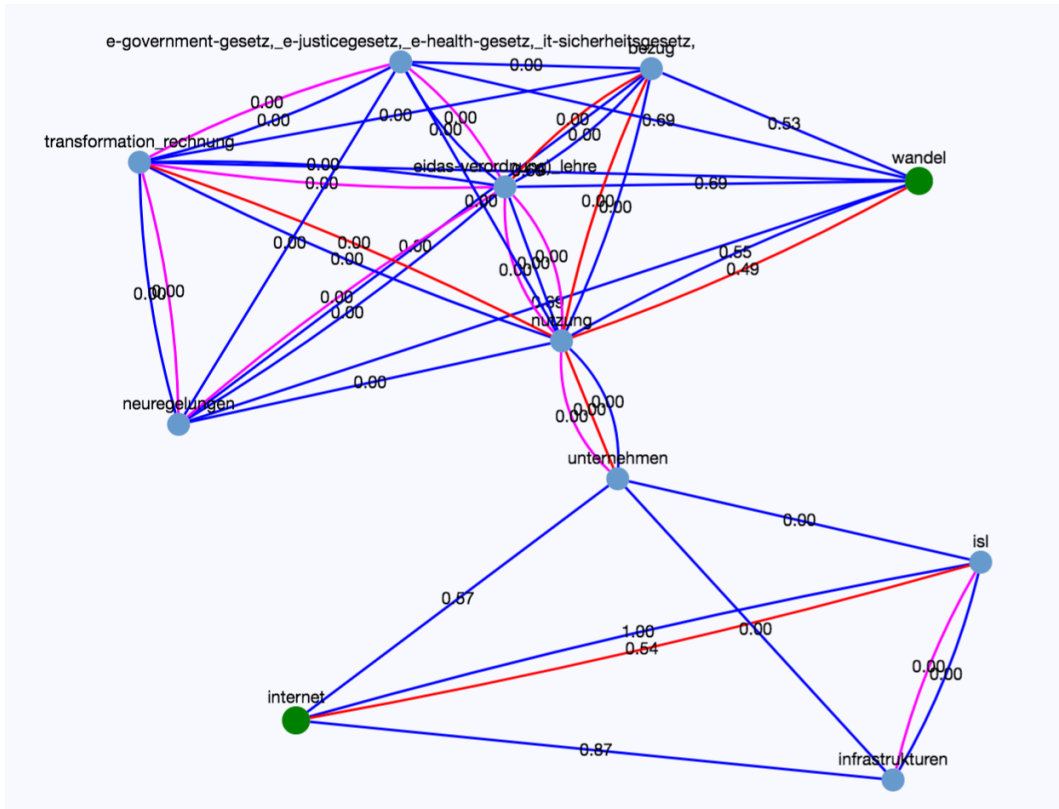


Abbildung 8: Die Verbindung zwischen den Konzepten Internet und Wandel im FEP-JUR

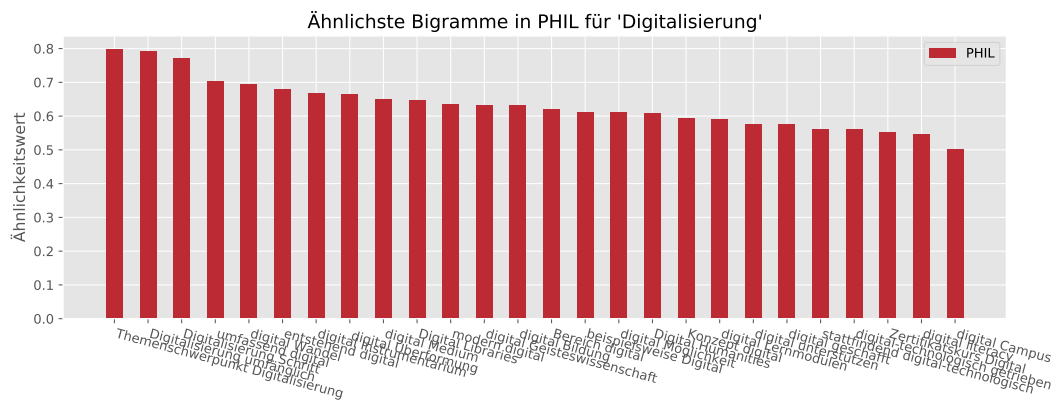


Abbildung 9

Unternehmen. Diese Nutzung stößt, so kann man schlussfolgern, den Bedarf nach Neuregelungen an, wie sie etwa in Themenfeldern wie e-Government, e-Health, IT-Sicherheit etc. juristisch kodifiziert werden.

Auch die Philosophische Fakultät (Abb. 9) verweist auf die Anpassungsleistung, die der Prozess der Digitalisierung ihr abnötigt. Ähnlich wie für die juristische Fakultät hat sie folglich auch einen *Themenschwerpunkt Digitalisierung* ausgebildet. Neben den Chancen stehen hier vor allem auch potentielle Risiken des Prozesses für die Gesellschaft im Mittelpunkt. Deshalb wird in der Philosophischen Fakultät die Reflexionsdimension und der Bildungsauftrag besonders betont, etwa der Aspekt der *digital literacy*. Zugleich wird die grundlegende Transformation von Teilen des eigenen Fachs in Gestalt der *digitalen Geisteswissenschaften* deutlich.

Erhellend ist in diesem Zusammenhang auch ein Blick auf das Wortumfeld des Konzepts »digital« im FEP der Fakultät (Abb. 10). Einerseits wird es durch einen Knoten bestimmt, der die Infrastrukturdimension von Digitalisierung, aber auch ein mögliches Forschungsgebiet markiert (Datenbanken, Softwarestrukturen und Algorithmen). Auf der anderen Seite steht ein eng verflochtener Subgraph, dessen Knoten die gesellschaftliche Dimension der Digitalisierung beschreibt, also die daraus erwachsenden neuen Wissenskulturen, aber eben auch Machtkomplexe. Dieser Prozess wird in Diskursen repräsentiert und involviert eine Vielzahl von Akteuren – Organisationen, Interessengruppen oder Staaten, die sich je spezifischer Begründungslogiken bedienen können. Ein dritter Subgraph repräsentiert offenbar die Fächer der Fakultät und deren epistemologische Bezüge zu den beiden anderen Bereichen des Graphen.

In der wirtschaftswissenschaftlichen Fakultät scheint die Digitalisierung der Gesellschaft als Prozess bereits stärker angekommen und in den eigenen Kanon

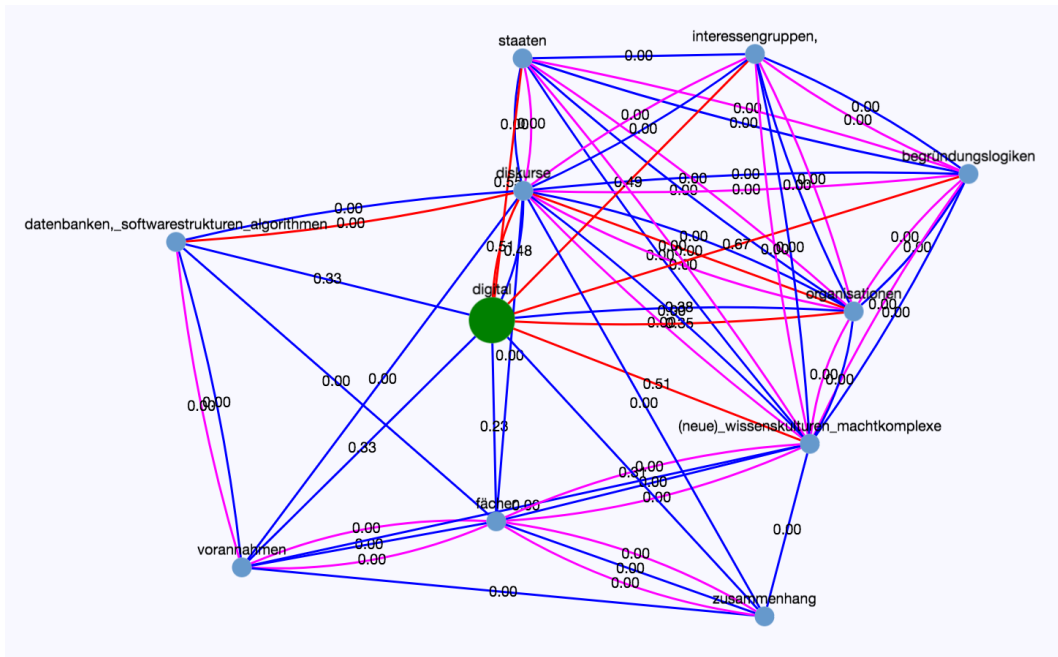


Abbildung 10: Das Wortumfeld von digital im FEP-PHIL

absorbiert worden zu sein. Der Fokus liegt hier primär auf den Chancen und Möglichkeiten, die der Prozess eröffnet. Auch hier wird ein *Themenschwerpunkt Digitalisierung* etabliert und ist der Prozess selbstverständlich ein Forschungsgegenstand. Aspekte wie *Digital Entrepreneurship*, *Digital Marketing* und das *Gründerzentrum Digitalisierung* Niederbayern verweisen darüber hinaus aber klar auf eine anwendungs- und transferorientierte Konnotation des Begriffs.

4 Conclusum

In der hier vorliegenden präliminaren Analyse der FEPE der vier Passauer Fakultäten haben wir zunächst gezeigt, wie mit Mitteln aus dem Bereich des NLP eine Lesung der Texte mittels Computer (*Distant Reading*) möglich ist und wie sich in Teilen die Herausforderung sehr kleiner Korpora durch Domänenmodellierung mittels Ontologien überwinden lässt. Weiter haben wir gezeigt, wie sich mit dem Instrument der semantischen Nähe Aussagen über die Bedeutung des Wortfelds Digitalisierung in verschiedenen Fakultäten treffen lässt. Mit dem Instrumentarium zur Analyse natürlicher Sprache können wir folglich, zunächst basierend auf statistischer Auffälligkeit, Digitalisierung als

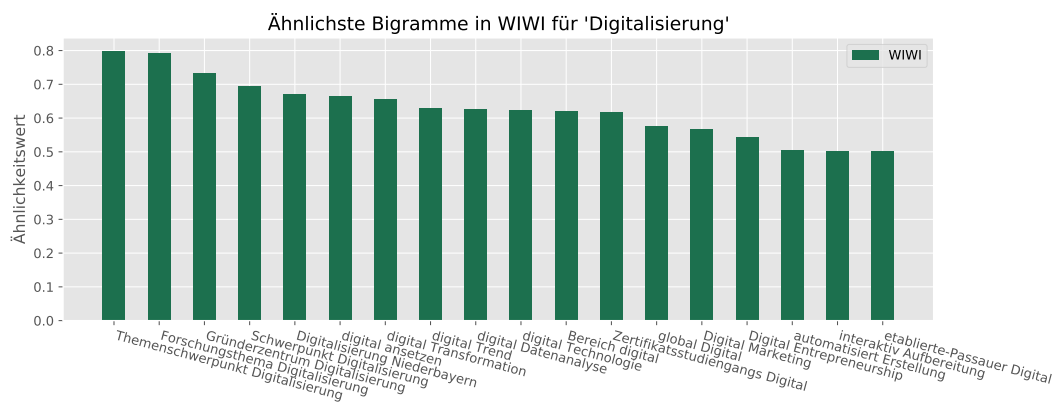


Abbildung 11

ein übergreifendes Kernthema der Fakultätsberichte identifizieren. Wir können weiterhin zeigen, dass und wie der Begriff im Kontext der verschiedenen Entwicklungspläne aufgeladen und gebraucht wird. Dadurch werden einerseits fachliche und fachkulturspezifische Unterschiede im Umgang mit dem Gegenstand deutlich, andererseits zeigen sich aber ebenso Berührungspunkte, die tatsächlich in der interfakultativen Kooperation auch zum Teil schon mit Leben gefüllt werden.