

E-Learning und Forschendes Lernen-Diskurse an deutschen Universitäten

Jan Bernoth¹, Julian Dehne¹, Thi Nguyen¹ und Ulrike Lucke¹

Abstract: Mit Hilfe von Webcrawling und quantitativer Inhaltsanalyse wurde eine Übersicht über die Verteilung von E-Learning und Forschendes Lernen-Diskurse an deutschen Universitäten generiert. Dabei ist ein Programm UniDisk entstanden, die für ähnliche Fragestellungen weiterverwendet werden kann. Das Tool liefert einen Beitrag, die unübersichtliche Forschungslandschaft in Deutschland im Bereich E-Learning, des Forschenden Lernens und verwandter Disziplinen verständlicher und einfacher zugänglich zu gestalten.

Keywords: E-Learning, Forschendes Lernen, Crawling, Inhaltsanalyse, Diskurs

1 Einleitung

Mit der zunehmenden Internationalisierung und interdisziplinären Ausrichtung von wissenschaftlicher Forschung ist eine professionelle und umfangreiche Recherche im Forschungsprozess erforderlich. Viele Informationen werden über das Internet der breiten und wissenschaftlichen Öffentlichkeit zugänglich gemacht. Neben der Freitexteingabe in klassische Online-Suchmaschinen besteht auch die Möglichkeit auf anspruchsvolle Portale oder Fachdatenbanken zuzugreifen. Daneben veröffentlichen Universitäten Informationen zu aktuellen Themen und Forschungsergebnissen direkt auf ihren Webseiten.

Um sich zu Beginn eines Forschungsprozesses einen Überblick verschaffen zu können, sind Tools sinnvoll, die bei der thematischen Recherche eingesetzt werden können, um in Erfahrung zu bringen, welche Diskurse aktuell an den Universitäten relevant sind. Zu diesem Zweck entwickelten wir das Programm UniDisk2, die es dem Benutzer erlaubt, die Häufigkeit und Verteilung individuell definierter Stichwörter auf den Universitätswebseiten systematisch zu identifizieren.

Die entwickelte Software (und Methodik) folgt dabei der Logik der quantitativen Inhaltsanalyse. Quantitative Inhaltsanalysen können in Form von einfachen Häufigkeitsanalysen angewandt werden, die bestimmte Textbestandteile heraus greifen und sie dann auszählen [KGW05]. Bei dieser Methode ist es von Vorteil, dass ausgesuchte Merkmale auf einer zahlenmäßig breiten Basis gesammelt werden können. Die

¹ Universität Potsdam, Lehrstuhl für Komplexe Multimediale Anwendungsarchitekturen, August-Bebel-Str. 89, 14482 Potsdam, vorname.nachname@uni-potsdam.de

² <http://fleckenroller.cs.uni-potsdam.de/diskurs>

quantitative Vorgehensweise dient dem Zweck der Reduktion der Datenmassen und damit auch der Komplexitätsreduktion eines Forschungsgegenstandes. In der vorliegenden Studie werden wir das Programm auf ein konkretes Fallbeispiel anwenden. Von Bedeutung sind in diesem Zusammenhang die Themenspektren „E-Learning“ und „forschendes Lernen“.

2 Forschungsfrage

Im Themenbereich des E-Learnings existiert eine große Begriffsvielfalt. Zahlreiche Synonyme wie etwa „E-Lernen“, „electronic learning“, „eLearning“, „E-Education“, „computerbasiertes Training“, „computer-gestütztes Lernen“, „Online-Lernen“, „Internet-basiertes Lernen“ oder auch „multimediales Lernen“ zeigen eine vielfältige aber ungenaue Definition des Begriffs auf. Während Rey unter E-Learning das „Lehren und Lernen mittels verschiedener elektronischer Medien“ [Re09] versteht, betonen Arnold et al. die „subjektive Leistung der Lernenden und Lehrenden in sozialen, kommunikativen, kooperativen und partizipativen Prozessen und entsprechend geformten pädagogischen Verhältnissen“ [AKZ15]. Fischer sieht hingegen in den existierenden Begrifflichkeiten eine Gemeinsamkeit: Die „Verschmelzung von Bildungsprozessen mit digitalen Technologien“ [Fi13].

Die genannten Definitionen beinhalten großen Spielraum für Interpretations- und Definitionsvarianten und damit verbunden unzählige Gestaltungsmöglichkeiten. Die thematische Vielfalt des E-Learnings zeigt sich bundesweit ebenfalls in der wissenschaftlichen Hochschullandschaft. E-Learning ist heutzutage ein fester Bestandteil von Universitäten und unterstützt das Lernen, Lehren und organisatorische Aspekte des Hochschulalltags. Hinter der E-Learning Thematik stehen komplexe technische und didaktische Aspekte. Dementsprechend ist das wissenschaftliche Interesse rund um das Thema groß. Neben hochschuleigenen E-Learning-Institutionen beschäftigen sich informatische, erziehungswissenschaftliche, psychologische und wirtschaftliche und Fachbereiche für die Forschung mit E-Learning-Elementen in den Themen Strategie, Technik, Didaktik und curriculare Integration, Marketing, Rechtsmanagement oder Qualitätssicherung [KW04]. Aufgrund dieser Vielfalt eignet sich das Themengebiet, um die entwickelte IT-Technik UniDisk anzuwenden.

Forschendes Lernen wurde seit der Bundesassistentenkonferenz 2009 [Ba09] als pädagogisch-normatives Konzept verbreitet, um bei Studierenden schon zu Beginn des Studiums Selbstwirksamkeit und Selbstorganisationsfähigkeiten zu fördern [Hu09]. Im Rahmen einer Studie zur Beforschung des Qualitätspakt Lehre unserer Arbeitsgruppe wurden mittels qualitativer Interviews 22 verschiedene Projekte im Bereich Forschendes Lernen untersucht. Dabei wurde der Fragestellung nachgegangen, welche digitalen Medien und E-Learning-Tools für das forschende Lernen relevant sind. Vorläufige Ergebnisse zeigen, dass die beiden Aspekte selten zusammen gedacht werden und der

Einsatz von E-Learning-Tools nicht über die üblichen Werkzeuge zur Kommunikation und Organisation hinausgehen. Als Gründe hierfür wurden drei Thesen aufgestellt:

Die erste These lautet, dass die technisch orientierte E-Learning- und Mediendidaktik-Community innerhalb der Hochschuldidaktik disjunkt von derjenigen ist, die sich mit Forschendem Lernen beschäftigt. Die zweite These ist die, dass die Unterstützungseinrichtungen innerhalb der Universitäten institutionell getrennt sind. So wird bei dem Thema E-Learning eher auf das Rechen- oder E-Learningzentrum verwiesen, wohingegen bei dem Thema des forschenden Lernens andere Akteure (hochschuldidaktische Zentren oder Einzelprojekte) federführend sind. Die dritte These lautet, dass beide Trends bzw. pädagogischen Bewegungen mit unterschiedlichen Diskursen und Traditionen verknüpft sind, so dass es zu einer geographischen Spaltung der Hochschulfokussierungen kommt. Während die ersten beiden Thesen an anderer Stelle überprüft werden, widmen wir uns in diesem Beitrag der dritten These und identifizieren dafür die geographische Verteilung der Forschungsthemen in Deutschland.

3 Methode

3.1 Zur Erstellung der Stichwortliste

Mit der Annahme, dass Forschungsprojekte und -schwerpunkte eine Internetpräsenz aufweisen, kann UniDisk verwendet werden, um aktuelle Forschungen zum E-Learning und zum forschenden Lernen auf bundesweiter Ebene zu lokalisieren und identifizieren. Zunächst muss jedoch eine Grundlage geschaffen werden, um relevante Keywords festzulegen, mit dem das Tool schließlich arbeiten kann. Zu diesem Zweck wurden zum einen Experteninterviews mit sieben E-Learning ForscherInnen geführt, um das entsprechende Sprachfeld in der Forschungspraxis zu explorieren. Des Weiteren wurden die Projektwebsites der uns bekannten Forschenden Lernen Projekte mittels qualitativer Inhaltsanalyse hinsichtlich verwendeter Sprache zum Thema forschendes Lernen untersucht. Da der E-Learning-Diskurs (unter diesem Namen) eine längere Tradition aufweist, konnte bei der Identifizierung relevanter Stichworte zusätzlich auf Lehrbücher zum Thema E-Learning zurückgegriffen werden. Dagegen musste beim forschenden Lernen das Konzept zunächst sprachlich-semantisch angereichert werden. Dabei kann davon ausgegangen werden, dass mit Hinblick auf die geringe Größe der Forschungsgemeinschaft in diesem Feld eine bestmögliche Beschreibung des Konzeptes mittels Stichworte gefunden wurde. Für eine Übersicht über die gewählten Stichworte und deren Gewichtung bei der Analyse verweisen wir auf das Online-Tool³.

³ <http://fleckengeroller.cs.uni-potsdam.de/diskurs>

3.2 Bereinigung der Daten

Die identifizierte Verteilung der Keywords auf den Webseiten der deutschen Universitäten wird durch mehrere Mechanismen optimiert. Dabei wird als Algorithmus der SolrScore verwendet⁴. Die Webseiten werden nach Uni-Domains gefiltert und aggregiert. Die SolrScores für die einzelnen Sites werden mittels Percentile auf 10 minimale und 20 maximale relevante geographische Positionen gefiltert. Percentile beschreiben die Prozentgrenzen für die mehr als 99 % (98, 97 ...) etc. der Daten durch das Modell beschrieben werden. Das technische Vorgehen für die Analyse sieht wie folgt aus: Das Programm folgt den Links auf den Websites und durchsucht diese, wobei ein Solr-Score relevanter Websites erstellt wird. Der Solr-Score setzt sich aus der folgenden Formel zusammen:

```
<lst name="explain">
  <str name="id=archangel,internal_docid=4">
    0.46632254 =
      (MATCH) fieldWeight(text:wings in 4), product of:
      1.7320508 = tf(termFreq(text:wings)=3)
      2.871802 = idf(docFreq=2)
      0.09375 = fieldNorm(field=text, doc=4)
  </str>
  <str name="id=hawkgirl,internal_docid=24">
    0.35897526 =
      (MATCH) fieldWeight(text:wings in 24), product of:
      1.0 = tf(termFreq(text:wings)=1)
      2.871802 = idf(docFreq=2)
      0.125 = fieldNorm(field=text, doc=24)
  </str>5
```

- Tf steht für die Frequenz eines Ausdrucks. Je öfter dieser vorkommt, desto höher ist Tf.
- idf steht für die inverse Dokumentenfrequenz. Seltenere Ausdrücke zählen mehr als häufige.
- coord ist der Koordinationsfaktor – Wenn mehrere Ausdrücke gesucht werden und diese gemeinsam auftreten, wird der Score erhöht.
- lengthNorm – Treffer in einem kleinen Text werden höher gewichtet als in einem großen Text
- index-time boost – nicht verwendet

Die Ergebnisse wurden nach den Geodaten der Universitäten kollabiert, wobei der Solr-Score von den Unterseiten der Universitäten summiert wurde. Dadurch werden die Seiten mit vielen Stichworttreffern auf mehreren Unterseiten höher gewichtet als jene

⁴ <http://lucene.apache.org/solr/>

⁵ https://wiki.apache.org/solr/SolrRelevancyFAQ#How_are_documents_scored

Hochschulseiten, die vom Nutzer möglicherweise als besonders relevant eingestuft werden. Dieser Umstand muss bei der Durchsicht der Stichworttreffer beachtet werden. Aus Gründen der Serverkapazität wird die Suchtiefe auf maximal 8 Sprünge begrenzt.

Der Algorithmus für die Berechnung der relevanten Positionen lautet:

```
minResults = 10;
maxResults = 25;
G = 100; //count percentiles downwards
while (true) {
    1) Berechne das Percentil p basierend auf G.
    2) Schau ob die Ergebnisse e, die von p % der Daten
    beschrieben werden
    if (e > minResults und e < 25) {
        Nimm G als Schwellwert für die Auswahl der
        geographischen Punkte.
    }
}
```

Die Idee bei diesem Algorithmus ist die, dass die Varianz abhängig von den Stichworten und die Homoskedastizität steigen kann. Es sollte daher keine statische Grenze geben (z.B. 95% der Werte), da für die Visualisierung sowohl eine Anzeige von 200 geographischen Positionen als auch der Verweis auf zwei bis drei Hauptbrennpunkte ungeeignet sind. Das hier gewählte Vorgehen hat den Vorteil je nach Datenlage die Auswahl für die Darstellung zu optimieren, nicht jedoch die statistische Signifikanz. Dies hat zur Folge, dass für manche Stichwortmengen die gefundenen Hotspots eventuell nur 70% der Daten beschreiben, für andere jedoch 99%. Wir hatten in unseren Testläufen bislang keinen Fall, bei dem der Percentilcounter unter 90 gefallen wäre; auszuschließen wäre es jedoch nicht. Die Designentscheidung wurde getroffen, damit die Toolchain später ohne statistisches Vorwissen verwendet werden kann. Um den mathematischen Vorgang dennoch nachvollziehbar zu machen, wird der Algorithmus an dieser Stelle in der Tiefe diskutiert. Abbildung 1 zeigt einen Ausschnitt der verwendeten Stichworte, deren Score sowie die Häufigkeit der getroffenen Seiten. Die erste Spalte nennt das Stichwort, die zweite den berechneten SolrScore und die dritte die Anzahl der gefundenen Websites.

| | | | |
|----|------------------------|----------------------|-------|
| 32 | personalized eLearning | 0.000109329296 | 1 |
| 33 | Serious Games | 0.516816903549999 | 689 |
| 34 | Virtuelle Seminare | 0.2618875895459999 | 162 |
| 35 | Web 2.0 | 7.313872188190884 | 12496 |
| 36 | Web Based Training | 0.10793093358099999 | 152 |
| 37 | Wiki | 14.945811192303694 | 21250 |
| 38 | 3D Welten | 0.22475347989000002 | 131 |
| 39 | Blended Learning | 9.955486481774754 | 5185 |
| 40 | digitale Verbreitung | 0.007879945379999998 | 18 |
| 41 | Entdeckend Lern | 0.2168835236230005 | 264 |
| 42 | entdeckendes Lernen | 0.2168835236230005 | 264 |
| 43 | experimental inquiry | 0.02223904389 | 5 |
| 44 | Forschend Lern | 7.284200545555043 | 2750 |
| 45 | Forschendes Lernen | 7.284200545555043 | 2750 |
| 46 | Forschungsbasiert Lern | 0.4706583718429991 | 169 |

Abb. 1: Ausschnitt der verwendeten Stichworte und deren Score

Nach dem automatisierten Filter wurden einzelne Begriffe manuell gelöscht, da sie übermäßigen Einfluss auf die Scores hatten. Zum Beispiel wurde der Begriff „Online Lernen“ entfernt, da dieser auch irreführende Semantik von Sätzen, die mit „Online“ anfangen und „Lernen“ aufhören, inkludiert hätte.

3.3 Visualisierung

Nach der Bereinigung wurden die Ergebnisse mit Hilfe von OpenStreetMap geplottet. Dabei wurden die 10-20 Universitäten mit der höchsten berechneten Score mit Markern belegt. So deutet der Marker in unserem Beispiel auf die Webseiten von Universitäten hin, bei welchen die Themen E-Learning und forschendes Lernen am präsentesten sind.

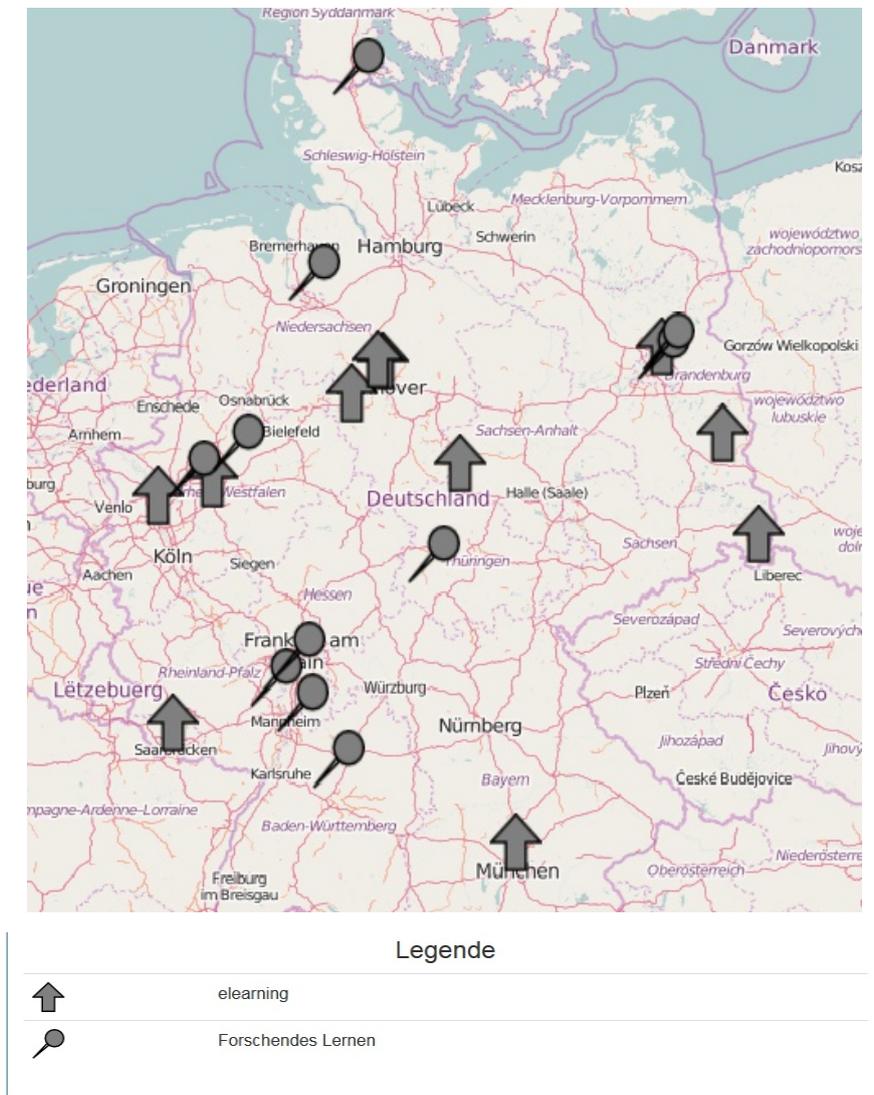


Abb. 2: Geographische Visualisierung der Ballungszentren der E-Learning-Diskurse im Vergleich zum Thema Forschendes Lernen

3.4 Automatisierung der Toolchain

Um die Tools wiederverwendbar zu machen, haben wir die entwickelten Komponenten zu einer Toolchain ausgebaut. Diese ist in Abbildung 3 zu sehen.

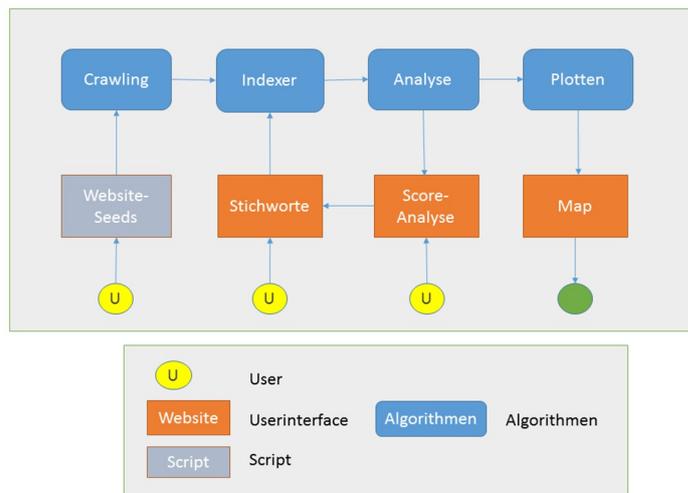


Abb. 3: Aktivitätsdiagramm für die Auswertungskette

Blau markiert sind die entwickelten Teilalgorithmen. Mit Ausnahme der Zusammensetzung der geographischen Visualisierung im letzten Schritt hat der Benutzer die Möglichkeit während des gesamten Prozesses einzugreifen. Zunächst kann mittels der Seeds Einfluss darauf genommen werden, welche Webseiten gecrawled werden. Der Crawler folgt maximal 8 Sprüngen. Dies bedeutet, dass Universitäten, deren Webseiten nicht in den Seed aufgenommen wurde, nur mit einer sehr geringen Wahrscheinlichkeit getroffen werden. Als nächstes gibt der Nutzer die Stichworte entweder über eine Weboberfläche oder direkt in die Datenbank ein. Dieser Vorgang unterstützt auch den parallelen Zugriff von mehreren Usern. Daraufhin werden die Stichworte mit Hilfe eines Indexers innerhalb der heruntergeladenen Universitätswebseiten gesucht, gescored und indiziert. Das Herunterladen der Webseiten hat den Vorteil, dass diese für weitere Analysen oder Forschungsfragen verwendet werden können ohne ein weiteres Mal abgerufen werden zu müssen. Dadurch wird die Last auf den meist stark geforderten Universitätsservern auf ein Minimum reduziert. Nach dem Indexen werden die Ergebnisse aggregiert und statistisch aufbereitet. Dem User wird eine Oberfläche gezeigt, auf der er verschiedene Stichworte und deren Wertung vergleichen kann. Daraufhin kann der User weitere Stichworte hinzufügen oder bestehende löschen (nur in der Datenbank möglich). Ist der Nutzer mit der Qualität der Daten noch nicht zufrieden, kann beliebig oft ein neuer Indexvorgang gestartet werden. Abschließend werden die Daten auf einer Deutschlandkarte geclustered und gezeichnet.

4 Ergebnisse und Diskussion

Abbildung 2 zeigt eine Übersicht über die verschiedenen Ballungszentren nach dem gewählten Verfahren. Dies kann veranschaulichen an welchen Standorten E-Learning und forschendes Lernen auf die Universität verallgemeinert relevant sind. Allerdings ist diese Schlussfolgerung mit Einschränkungen zu treffen. Als Störfaktor ist beispielsweise die Größe der Universität oder des Standortes (z.B. Berlin oder München) einzukalkulieren, denn diese beeinflusst die Wahrscheinlichkeit, dass Stichworte gefunden werden. Auch die lokale Gebundenheit der interviewten Experten kann sich auf die Sprache und damit die Ergebnisse ausgewirkt haben.

Durch das Online stellen der Tool-Chain planen wir, die Stichworte in einem Community-Prozess zu einem reliableren Messinstrument weiterzuentwickeln. Eine tiefer gehende Interviewstudie mit Experten aus einem repräsentativeren Sample bietet Raum für weitere Forschung. Die Frage nach der Verteilung von E-Learning- und Forschendes Lernen-Diskurse an Universitäten in Deutschland kann mit dem vorgestellten Tool beantwortet werden. Die Vermutung der These 3, dass E-Learning- und Forschendes Lernen-Diskurse geographisch getrennt sind, konnte bestätigt werden.

Mithilfe des Tools können weitere beliebige Fragestellungen bezüglich der aktuellen Verbreitung von Themen in der deutschen Hochschullandschaft recherchiert werden. Die Voraussetzungen dafür sind relevante Indikatoren, welche Universitäten auf ihren Webseiten zur Verfügung stellen. Des Weiteren sollte eine systematische Stichwortliste mit einschlägigen Begriffen zum entsprechenden Themenfeld erstellt werden, um mithilfe des technischen Tools ein möglichst präzises Ergebnis erzielen zu können.

Wie bereits im Kapitel 3.1 erwähnt, kann das Tool auch ohne statistische Kenntnisse verwendet werden. Allerdings wird im Zuge der technisch automatisierten Prozesse auf einen reflektierten Umgang mit den gegebenen Daten hingewiesen, damit keine voreiligen Schlüsse auf Grundlage der Crawler-Ergebnisse gemacht werden.

In Folge dieser Publikation haben wir die entwickelte Tool-Chain öffentlich zugänglich gemacht, damit die hier beschriebenen Ergebnisse nachprüft werden und ein Community-Prozess zur Sammlung relevanter Stichworte angestoßen werden können.

Literaturverzeichnis

- [KGW05] Köhler, Reinhard; Ungeheuer, Gerold; Wiegand, Herbert Ernst (Hg.) (2005): Quantitative Linguistik. Ein internationales Handbuch = Quantitative linguistics. Berlin: de Gruyter.
- [Re09] Rey, Günter Daniel (2009): E-Learning. Theorien, Gestaltungsempfehlungen und Forschung. 1. Aufl. Bern: Huber.
- [AKZ15] Arnold, Patricia; Kilian, Lars; Thilloßen, Anne; Zimmer, Gerhard (2015): Handbuch E-Learning. Lehren und Lernen mit digitalen Medien. 4., erw. Aufl. Bielefeld: wbv.
- [Fi13] Fischer, Helge (2013): E-Learning im Lehralltag. Analyse der Adoption von E-Learning-Innovationen in der Hochschullehre. Wiesbaden: Springer Fachmedien Wiesbaden.
- [KW04] Kleimann, Bernd; Wannemacher, Klaus (2004): E-Learning an deutschen Hochschulen. Von der Projektentwicklung zur nachhaltigen Implementierung. Hannover: HIS (Hochschulplanung, 165).
- [Ba09] BAK – Bundesassistentenkonferenz (2009): Forschendes Lernen – Wissenschaftliches Prüfen. Ergebnisse der Arbeit des Ausschusses für Hochschuldidaktik. Bielefeld: Universitätsverlag Webler.
- [Hu09] Huber, L. (2009). Warum Forschendes Lernen nötig und möglich ist. In Huber, L., Hellmer, J., Schneider, F. (Hrsg.). Forschendes Lernen im Studium. Aktuelle Konzepte und Erfahrungen. Bielefeld: Universität Webler.