

## Graduiertenkolloquium Angewandte Informatik

# Identification of Emerging Scientific Topics in Bibliometric Databases

Dipl.-Inform. Wirt. Carolin Mund

Fraunhofer-Institut für System- und Innovationsforschung ISI

Frühzeitiges Erkennen von aufkommenden Themengebieten in der Wissenschaft unterstützt sowohl Entscheidungen auf individueller als auch öffentlicher Ebene. Obwohl eine entsprechende Suche heutzutage durch die immer stärker zunehmende Digitalisierung von Publikationsdaten erleichtert wird, beschränken sich viele bestehende Verfahren auf eine retrospektive (Zitations-)Analyse der Publikationsdaten. Das Ziel der vorliegenden Arbeit war deshalb die Entwicklung eines Verfahrens, das zeitnah und neutral sogenannte „emerging topic candidates“ aus einer Menge von wissenschaftlichen Publikationen auswählt und einem Endnutzer präsentiert. Die Identifikation solcher aufkommender Themengebiete in einem frühen Entwicklungsstadium kann insbesondere bei der Vergabe von Fördermitteln aber auch bei der – gerade in der Forschung wichtigen - Anerkennung von wissenschaftlichen Leistungen eine entscheidende Rolle spielen. Das entwickelte System besteht aus zwei Ansätzen, die sowohl die „internen“ als auch „externen“ Eigenschaften der Publikationen verwenden. Die externen Eigenschaften bezeichnen in diesem Zusammenhang alle Metadaten, die in den bibliometrischen Datenbanken hinterlegt sind und nicht (vollständig) von den Autoren einer Publikation beeinflusst werden können, wie z.B. die Sichtbarkeit der veröffentlichenden Zeitschrift. Diese externen Eigenschaften wurden in Selektionsregeln für aufkommende Themengebiete umgesetzt.

Ein Kernstück des Ansatzes basiert jedoch auf den internen Faktoren und somit auf den Eigenschaften einer Publikation, die von den Autoren selbst bestimmt werden. Hierzu zählen insbesondere die verwendeten Terme und Referenzen einer wissenschaftlichen Publikation sowie die Autoren selbst. Ein Latent Dirichlet Allocation-Verfahren (LDA) wurde erweitert, um diese verschiedenen Eingabeformate gewichtet in die Berechnung der Themenwahrscheinlichkeit eingehen zu lassen. Zusätzlich wurden Parameter eingeführt, die das LDA-Verfahren besonders flexibel auf die unterschiedlichen Datensätze reagieren lassen. Durch Einsatz des entwickelten Verfahrens kann der für die Selektion von neuen Themen zu sichtende Datensatz drastisch reduziert werden, während sich der darin enthaltene Anteil der Dokumente für aufkommende Themengebiete stark erhöht.

Termin: Freitag, 04. Juli 2014, 14.00 Uhr

Ort: Englerstraße 11, 76131 Karlsruhe  
Kollegiengebäude am Ehrenhof (Geb. 11.40), 2. OG, Raum 231  
(Hinweise für Besucher: [www.aifb.kit.edu/web/Kontakt](http://www.aifb.kit.edu/web/Kontakt))

Veranstalter: Institut AIFB, Forschungsgruppe Wissensmanagement

Zu diesem Vortrag lädt das Institut für Angewandte Informatik und Formale Beschreibungsverfahren alle Interessierten herzlich ein.

Andreas Oberweis, Hartmut Schmeck, Detlef Seese, Wolffried Stucky, Rudi Studer (Org.)