

„Spieglein, Spieglein an der Wand ...“

Die Bibliometrische Community im Spiegel ihrer eigenen Methodik

Dirk Tunger und Julia Wilhelm

Das Themenfeld „Bibliometrie“ erfreut sich gegenwärtig einer wachsenden Beliebtheit. So zumindest das Bauchgefühl, wenn man die Zahl der einschlägigen Publikationen, der Konferenzbeiträge und der als graue Literatur erscheinenden Studien überschlägt. Will man es genauer wissen, so muss man die bibliometrische Community im Detail unter die Lupe nehmen, beispielsweise mit Hilfe einer bibliometrischen Analyse¹.

1 Vorgehensweise

Der erste Schritt bei der Erstellung einer bibliometrischen Analyse ist das Formulieren einer entsprechenden Suchanfrage. Hierzu ist das Prinzip von Precision und Recall zu beachten, um am Ende nur relevante Treffer, diese aber möglichst vollständig, in der Treffermenge zu vereinen.

Bei der Erstellung der Suchanfrage ist zu beachten, dass es sowohl eindeutig der Bibliometrie zuzuordnende Suchbegriffe gibt als auch solche, die ebenso in anderen Zusammenhängen auftreten können. Aus diesem Grund empfiehlt es sich, die Suchanfrage auf Basis der Topics der Kernzeitschriften (Scientometrics und Journal of Informetrics) zu erstellen. Hierzu wurden die wissenschaftlichen Veröffentlichungen der vergangenen fünf Jahre dieser beiden Zeitschriften auf verwendbare Keywords durchsucht.

Aus diesen Keywords wurden dann zwei Mengen gebildet: Die Menge an Keywords, die thematisch sehr eindeutig (#1) ist und die zweite Menge, die nur in Kombination mit eindeutigen Keywords zu einem brauchbaren Ergebnis führt (#2). Diese beiden Mengen an Keywords bilden das Grundgerüst der verwendeten Suchanfrage. Hinzu kommen alle Veröffentlichungen aus den beiden bibliometrischen Journals „Scientometrics“ und „Journal of Informetrics“ (#3).

¹ Wilhelm, Julia: Bibliometrische Analyse der wissenschaftlichen Community „Bibliometrie und Scientometrie“; Diplomarbeit im Studiengang Bibliothekswesen an der Fachhochschule Köln, Fakultät für Informations- und Kommunikationswissenschaft, vorgelegt bei Prof. Dr. Simone Fühles-Ubach in Zusammenarbeit mit Dr. Dirk Tunger, Forschungszentrum Jülich, Zentralbibliothek, Köln 2011

Dieser Artikel beschreibt den Entwicklungstrend der bibliometrischen Community mit den Methoden, die diese Community ansonsten auf andere Fachgebiete anwendet: Darstellen der Publikationsentwicklung, der Vernetzung innerhalb des Fachgebietes und der Leistung von Einzelautoren. Bibliometrische Standard-Indikatoren werden berechnet und mit ihrer Hilfe werden Aussagen zum Publikationsverhalten der Community insgesamt und dessen Entwicklung in den vergangenen Jahren getroffen.

This article describes development trends in the bibliometrics community with respect to methods which this community usually applies to other subject areas: representation of developments in publications, networking within the subject area and the contribution of individual authors. Bibliometric standard indicators are calculated and used to provide information on the publication behaviour of the community as a whole and the development of this behaviour in the past.

Die beschriebenen Teile der Suchanfrage lauten ausformuliert:

#1 (thematisch sehr eindeutige Suchbegriffe)

```
TS=(self-citation correction or site interlinking or sleeping beaut* or uncitedness or web impact factor* or bibliometric* or scientometric* or citation age data or citation database* or citation distribution* or citation habit* or citation index* or citation metrics or citation network analysis or citation patterns or citation perspective or citations analyse* or co-citation cluster* or co-word analysis* or disciplinary citation impact or egghe* or generalized impact factor* or Google PageRank algorithm* or greater research impact or h-Index sequenc* or hirsch-core or hirsch-index or hirsch-type index* or informetric* or internet citation* or jin a-index or journal influence or journal qualimetric* or journal rank* or journal self-citation* or kosmulski's h((2))-index or long-term citation impact or lotkaian informetric* or Lotka* or matthw core journal* or publication delay* or publication output or scientific-research output))\{0.4cm]
```

Mit dem Field-Tag „TS“ sucht der Nutzer im Web of Science in den thematischen Feldern „Titel“, „Abstract“ sowie „Keywords“.

#2 (thematisch weniger eindeutige Keywords, die in einer Kombination mit spezifischeren Keywords gesucht werden)

TS=(impact factor or peer review or cited papers or cocitation or evaluating research or g-index or journal peer review or journal-citation-report or multiple authorship or publication productivity or research performance or r-index or research trend or scopus or self-citation* or word analys* or zipf's law) **and** TS=(author self-citation* or bibliometric* or citation* or egghe* or hirsch index or hirsch-core or hirsch-type indic* or impact factor* or informetric* or journal influence* or lotka* or „matthew core journal* or patent citation analys* or research output index or scientific research output or scientometric* or sleeping beaut* or uncitedness or webometric* or ranking)

Durch das „and“ in der Suchanfrage wird nach einer Kombination zweier Keywords gesucht: Ein Dokument gelangt nur dann in die Treffermenge, wenn es sowohl ein Keyword aus der ersten Klammer von #2 gemeinsam mit einem Keyword der zweiten Klammer enthält.

#3 (die bibliometrischen Journals)

SO=(„Journal of Informetrics or „Scientometrics“)

Mit #3 werden alle Publikationen der beiden auf den Themenkomplex „Bibliometrie / Scientometrie“ fokussierten Journals erfasst, unabhängig von den verwendeten Keywords.

Mit „or“ verbunden, ergibt sich die gesamte Suchstrategie, die allen weiteren Auswertungen zu Grunde liegt.

2 Vergleich der Suchanfrage von Judit Bar-Ilan mit der erstellten Suchanfrage

Nach dem Erstellen einer so komplexen Suchanfrage ist es ratsam, sich, falls vorhanden, Suchanfragen von Fachkollegen anzuschauen, um die Möglichkeit eines Vergleiches zu haben. So ist es möglich, seine eigene Suchanfrage auf Fehler oder nicht bedachte Aspekte zu prüfen.

Im Fall dieser Arbeit hatte Judit Bar-Ilan in ihrer Publikation „Informetrics at the beginning of the 21st century – A review“² eine ähnliche Suchanfrage erstellt. In dieser Arbeit beschreibt sie ausführlich den Prozess der Erstellung einer Suchanfrage zum Themenkomplex „Bibliometrie“, so dass die von ihr verwendete

Suchbegriffe nachvollzogen werden können.

Die Suchanfrage von Judit Bar-Ilan:

Informetric* OR bibliometric* OR webometric* OR „cocitation analysis“ OR „citation analyses“ OR „cocitation analysis“ OR „cocitation analyses“ OR „co-citation analysis“ OR „co-citation analyses“ OR „link analysis“ OR „link analyses“ OR „link structure“ OR „self citation“ OR „self citations“ OR „self-citation“ OR „self-citations“ OR „S&T indicator“ OR „S&T indicators“ OR „citation map“ OR „citation visuali*“ OR „science policy “ OR „research policy“ OR „impact factor“ OR „impact factors“ OR „h-index“ OR „Hirsch index“ OR „patent analysis“ OR „patent analyses“ OR „Zipf“ OR „Bradford“ OR „Lotka“ OR „collaboration network“ OR „collaboration networks“ OR „coauthorship network“ OR „coauthorship networks“ OR „co-authorship networks“

Diese Suchanfrage beinhaltet eine Reihe von thematisch sehr spezifischen Suchbegriffen. Sie hat die Problematik der Numeri beachtet und hat ihre Suchbegriffe in mehreren Schreibvarianten aufgeführt. Jedoch beinhaltet ihre Suchanfrage Begriffe, die womöglich in anderen Disziplinen auch verwendet werden. So könnte zum Beispiel das Wort h-index ein Problem darstellen, denn dieser ist sowohl in der gesuchten Thematik vertreten, aber auch in anderen Zusammenhängen. So müsste nach diesem Begriff in Kombination mit einer thematisch spezifischen Einschränkung gesucht werden.

Es kann an dieser Stelle festgehalten werden, dass die in Kapitel 1 erstellte und in dieser Arbeit verwendete Suchanfrage durchaus vergleichbar ist mit Suchanfragen, die für vergleichbare Vorhaben erstellt wurden. Durch die sorgfältige Auswahl der Keywords wurden Probleme mit homonym verwendeten Begriffen vermieden.

3 Ergebnisse

3a Publikationsentwicklung

In Abbildung 1 sind die Publikationszahlen der bibliometrischen Community sowie die Gesamtpublikationszahlen der Datenbank „Web of Science“ gegeneinander aufgetragen. Im Jahre 1980 wurden insgesamt 50 Bibliometrie-Publikationen veröffentlicht: In den Jahren 1980 bis 2009 gab es eine Steigerung des Publikationsoutputs um das 17-fache. Es fällt weiterhin auf, dass es im Jahr 2005 eine beinahe Verdoppelung der Publikationszahlen zum Vorjahr gab. Die Publikationszahl des Jahres 2004 zu 2005 steigt von 265 Publikationen auf 453 Veröffentlichungen. Weiterhin kann festgestellt werden, dass es vom Jahr 2005 bis zum Jahr 2009 eine weitere Verdoppelung der Publi-

² Bar-Ilan, Judit: Informetrics at the beginning of the 21st century – A review, in: Journal of Informetrics, (2) 2008, S. 1-52

kationszahlen gab. Verglichen mit der Entwicklung im Web of Science insgesamt (der Gesamtoutput nimmt „nur“ um das 2,5-fache zu) ist also von einem stark überproportionalen Publikationsanstieg der bibliometrischen Publikationen zu sprechen.

Wie ist das zu erklären? Diese Entwicklung könnte im Zusammenhang mit dem verstärkten Augenmerk auf quantitativen Elementen beim Forschungsoutput stehen. So wurde beispielsweise in Deutschland im Jahre 2005 die Exzellenzinitiative ins Leben gerufen. Eine Aufgabe der Exzellenzinitiative ist die Förderung von herausragender Forschung an Universitäten in Deutschland durch Bund und Länder. Die Exzellenzinitiative unterstützt universitäre Spitzenforschung und erhöht ihre internationale Sichtbarkeit, schafft hervorragende Bedingungen für den wissenschaftlichen Nachwuchs an Universitäten, vertieft Kooperationen zwischen Disziplinen und Institutionen, stärkt die internationale Vernetzung der Forschung, fördert die Gleichstellung von Männern und Frauen in der Wissenschaft, intensiviert den wissenschaftlichen Wettbewerb in Deutschland und verbessert die Qualität des Wissenschaftsstandorts in der Breite.³

Zuvor wurde am 12. November 2004 vom Wissenschaftsrat eine Pressemitteilung mit dem Titel „Mehr Transparenz durch Leistungsvergleich in der Forschung“⁴ veröffentlicht. In dieser Pressemitteilung wird Transparenz der Forschungsleistungen der Universitäten und der außeruniversitären Forschungseinrichtungen gefordert. Dies soll mit Hilfe von Forschungsrankings geschehen.

Die angeführten Beispiele sind nur eine kleine Auswahl für das gestiegene öffentliche Interesse am Forschungsoutput und dessen quantitativer Entwicklung. Es gäbe viele weitere Beispiele aus universitären und außeruniversitären wissenschaftlichen Einrichtungen, nicht nur in Deutschland. Ein Erklärungsansatz, warum die Zahl wissenschaftlicher Veröffentlichungen zu bibliometrischen Themen steigt, ist also das gestiegene öffentliche Interesse an der Thematik, rückgekoppelt mit einer größeren Zahl an Akteuren, die nicht nur Auftragsarbeiten durchführen, sondern auch wissenschaftliche Tätigkeit entfalten.

Eine gestiegene Zahl an Personen, die zu bibliometrischen Themen publiziert, ist nicht nur durch die Teilnehmerzahl internationaler Konferenzen sichtbar, sondern auch im Netzwerk der bibliometric community.

3 vgl. hierzu: Wissenschaftsrat zur Exzellenzinitiative. URL: [13.3.2012] <http://www.wissenschaftsrat.de/arbeitsbereiche-arbeitsprogramm/exzellenzinitiative/>

4 Kling-Mathey, Christiane: Mehr Transparenz durch Leistungsvergleich in der Forschung, 2004; URL: [13.3.2012]: http://www.wissenschaftsrat.de/download/archiv/pm_3304.pdf

OBID®
Welcome to RFID

OBID i-scan® HF



RFID-Gate-Antennen für Bibliotheken

**Perfektion in Design,
Funktion und Service.**



reddot design award
winner 2010



OBID® – RFID by FEIG ELECTRONIC

FEIG
ELECTRONIC

FEIG ELECTRONIC GmbH
Lange Straße 4 · D-35781 Weilburg
Phone: +49 6471 3109-0
Fax: +49 6471 3109-99 · www.feig.de

3b Netzwerk der bibliometric Community

Definition von Netzwerken

„Netzwerk: Begriff zur Erklärung zwischenmenschlicher Beziehungen in einem gesellschaftlichen System. Graphisch dargestellt, bezeichnen Punkte Personen, Linien ihre gegenseitigen Beziehungen. Die Gesamtgesellschaft eines definierten Raumes kann in soziale Netzwerke gegliedert und in deren Rahmen analysiert werden.“⁵

Aufgaben und Ziele der Autorennetzwerke

Zu den Aufgaben und Zielen des Autorennetzwerkes gehört es, die Beziehung zwischen Akteuren darzu-

workWorkbench benutzt. Dieses Programm wurde im Jahre 2005 von der Universität Indiana (USA) entwickelt, um Netzwerkanalysen zu erstellen. Es wurde auf der Basis von Cyberinfrastructure Shell (CIShell) entworfen und ist als Open Source Software verfügbar. Die Vorteile von NetworkWorkbench sind, dass man mit ihm aus einer großen Datenmenge sowohl online als auch offline ein Netzwerk in überschaubarer Zeit erstellen kann.

Die nächsten Arbeitsschritte bestehen darin, die heruntergeladene Treffermenge aus der Suchmenge in das Programm NetworkWorkbench zu laden, um aus diesen Rohdaten ein Autorennetzwerk zu generieren. In der Excel Datei werden Informationen wie der Name (label) des Autors, die Anzahl der Zitierungen (timesCited), die Anzahl der Publikationen (numberOfWorks) und eine eindeutige Knotennummer (uniqueIndex) angezeigt. Diese eindeutige Knotennummer ist wichtig bei der Bereinigung der Autorennamen, denn durch die Hilfe dieser Nummer lassen sich falsch geschriebene Autorennamen mit der richtigen Namensschreibweise zusammenfassen, da es vorkommen kann, dass ein und derselbe Autor in der Tabelle sowohl falsch als auch richtig geschrieben ist. Damit die falschen Schreibweisen der Autorennamen mit den richtigen zusammengefasst werden können, wird die eindeutige Knotennummer des falsch geschriebenen Autorennamen mit der richtigen überschrieben. Würde nur der falsche Name korrigiert werden, würden später zwei gleich Namen im Netzwerk auftauchen. Um dem vorzubeugen, müssen die beiden Autoren unter einer eindeutigen Knotennummer zusammengefasst werden. Es wird aber nicht nur der Autorennamen zusammengefasst, sondern alle dazugehörigen Daten.

Nachdem das Netzwerk durch den Algorithmus „Physics“ bearbeitet wurde, ist nun die Struktur im Netzwerk zu erkennen. Als nächstes sollte die Größe der Knoten an die Publikationszahl angepasst werden, dazu wurde die maximale Anzahl der im jeweiligen Zeitraum erschienenen Publikationen bestimmt.

Dieses Netzwerk besteht aus 1805 Knoten, gleichbedeutend mit einer Community aus 1805 Autoren in diesen fünf Jahren. Von diesen 1805 Knoten sind wiederum 232 Knoten isoliert, das bedeutet, dass 13 % der Autoren aus der Community innerhalb des genannten Zeitraumes keinerlei Verbindungen zu anderen Autoren hatten. Darüber hinaus ist an der Entwicklung der Bereitschaft zur Ko-Autorenschaften zu erkennen, dass 87 % der Publikationen mit mindestens einem weiteren Autor entstanden waren. Die im Autorennetzwerk enthaltenen Knoten werden durch Kanten miteinander verbunden. Dieses Netzwerk beinhal-



Abbildung 1:
Publikations-
entwicklung

stellen und zu analysieren. Dabei können die Akteure wissenschaftliche Autoren, Institutionen, Länder usw. sein. Eine Beziehung zwischen den Akteuren kann eine wissenschaftliche Publikation, Kontakte, Mitgliedschaften oder Austauschbeziehungen sein. Es ist möglich, mit einem Autorennetzwerk durch die strategische Positionierung auf eine besondere Relevanz eines Autors in einem Netzwerk zu schließen, sowie diese untersuchte Person in eine thematische Struktur einzugliedern. Durch die Darstellung der Vernetzung erhält man eine Struktur der Informationsverbindungen innerhalb von wissenschaftlichen Disziplinen zu einer bestimmten Zeit. Durch die Akteurzentralität in einem Autorennetzwerk erhält man ebenfalls eine bessere Einschätzung, wer die Experten in der jeweiligen Community sind.

Es wird die zu Beginn des Artikels eingeführte Suchanfrage verwendet und eine Netzwerkanalyse für die zwei Zeiträume 2000–2004 sowie 2005–2009 durchgeführt. Für diese Aufgabe wurde das Programm Net-

⁵ Brockhaus – Enzyklopädie in 30 Bänden, 21. Auflage, Leipzig 2006, S. 13

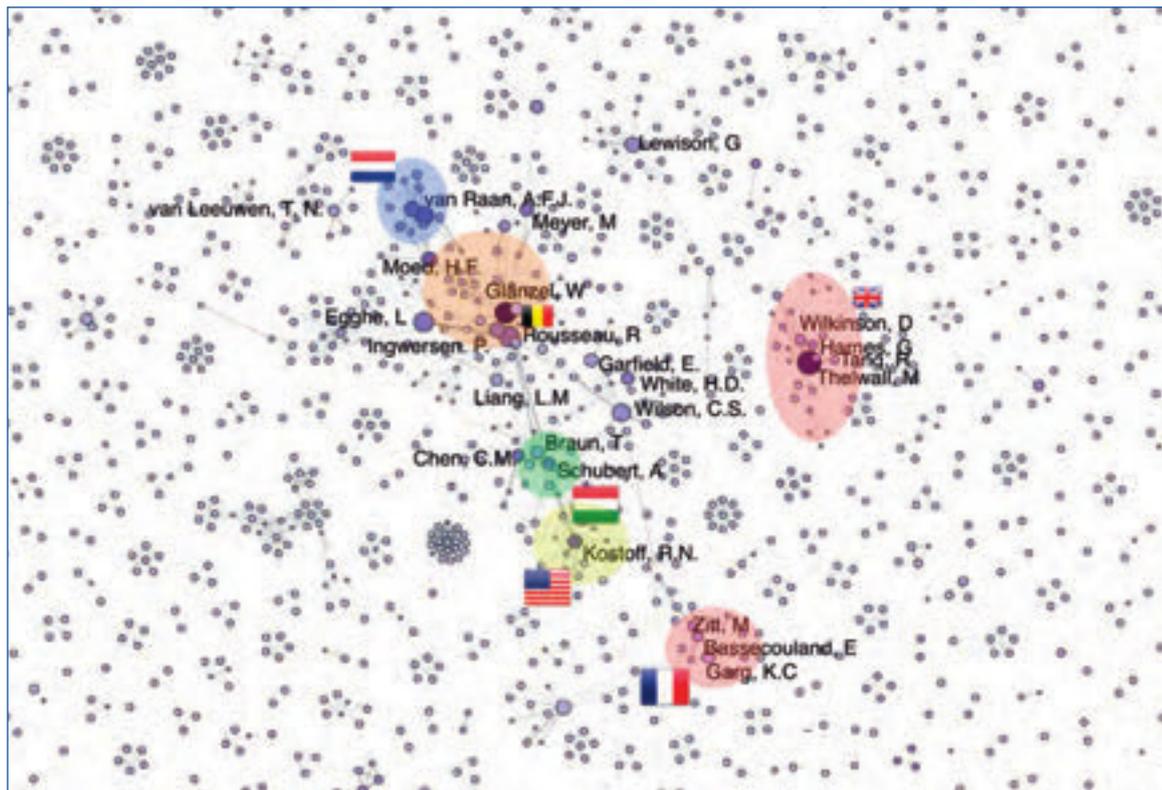


Abbildung 2:
Netzwerkgrafik
für den Zeitraum
2000–2004

tet 2542 Kanten. Dabei ist festzustellen, dass 25 % der Autoren zu einem weiteren Ko-Autor eine Verbindung aufweisen. Ebenso haben 25 % der untersuchten Wissenschaftler eine Kooperation mit zwei Ko-Autoren.

In diesem Autorennetzwerk hatte Michael L. Callahan¹ mit weiteren 21 Ko-Autoren eine Verbindung. Damit ist er der Autor mit den meisten Beziehungen zu anderen Ko-Autoren. Anthony F. J. van Raan und Thed N. van Leeuwen (beide Autoren sind am Centre for Science and Technology Studies in Leiden tätig) haben die intensivste Zusammenarbeit, so haben die beiden im Zeitraum 2000–2004 miteinander zehn Arbeiten veröffentlicht.

Dieses Autorennetzwerk verfügt über 4692 Knoten, gleichbedeutend mit einer Community aus 4692 Autoren. Im gesamten Netzwerk stehen 8,3 % der Knoten in keinerlei Kontakt zu anderen Knoten, das heißt sie sind als isoliert zu betrachten. Somit verfügen nur 4302 Autoren über direkte Kontakte zu anderen Autoren. Diese 4302 Knoten werden von insgesamt 9087 Kanten verbunden. Dabei ist folgendes festzustellen: In diesem Netzwerk haben 21 % der Autoren nur zu einem weiteren Ko-Autoren eine Verbindung. Weitere 23 % hatten im untersuchten Zeitraum mit zwei Ko-Autoren eine Zusammenarbeit.

Zwischen Hans-Dieter Daniel und Lutz Bornmann besteht die stärkste Beziehung in diesem Autorennetzwerk. In dieser Zusammenarbeit entstanden insgesamt 23 Publikationen. Ein weiteres Autorenpaar, welches über eine intensive Beziehung verfügt, ist Wolfgang Glänzel / Bart Thijs. Die beiden Autoren haben gemeinsam 17 Arbeiten in den fünf Jahren 2005–2009 veröffentlicht.

In den beiden Netzwerkgrafiken sind die bekanntesten Autoren der Community farblich gekennzeichnet. Es wird bereits optisch ersichtlich, dass die Autoren im Zeitraum 2005–2009 deutlich enger zusammengerückt sind: Ein Zeichen für eine verstärkte Kooperation untereinander.

Welche weiteren Veränderungen lassen sich insgesamt attestieren?

Zunächst fällt auf, dass das Netzwerk aus den Jahren 2005 bis 2009 wesentlich mehr Knoten enthält als das Netzwerk für die Jahre 2000 bis 2004. So beinhaltet das erste Netzwerk (Zeitabschnitt von 2000 bis 2004) insgesamt 1805 Autoren, während das zweite Netzwerk (Zeitspanne von 2005 bis 2009) zusammenfassend 4692 Autoren beinhaltet. Dies stellt eine 2,5-fache Steigerung dar. Des Weiteren nahm die Anzahl der veröffentlichten Publikationen im Vergleich

Acht gelassen werden. Beachtet man all diese Punkte, ist im Hinterkopf zu behalten, dass die Schwankungen auf Grund niedriger Publikationszahlen deutlich höher sind als bei Einrichtungen. Es ist die Tendenz, die Bibliometrie hier vermitteln kann:

Zwei Bibliometriker, die beide schon sehr lange in der Szene aktiv sind und beide eine sehr zentrale Stellung im Netzwerk haben, verfügen annähernd über eine identische Zahl an Publikationen. Der Anteil nicht zitierter Publikationen liegt ebenfalls fast auf gleichem Niveau und beträgt etwa 20 %. Ein großer Unterschied besteht aber in dem Anteil der Arbeiten, die in Kooperation entstanden sind: Hier kommt der eine auf einen Wert von 85 %, das bedeutet, dass fast jede Arbeit in Kooperation mit mindestens einem weiteren Verfasser entstanden ist. Der andere erreicht einen Anteil von 37 % an Kooperationsarbeiten, also gerade einmal etwas mehr als jede dritte Arbeit. Der erste hat eine doppelt so hohe Zitationsrate als der zweite. Macht sich Kooperation also bezahlt? Es wäre zu weit gegriffen, dies jetzt zu verallgemeinern, geschadet hat sie aber mit Sicherheit nicht.

4 Zusammenfassung

Was ist der Nutzen aus der durchgeführten bibliometrischen Analyse sowie den vorgestellten Ergebnissen? Es konnte gezeigt werden, dass es mit Hilfe thematischer Suchanfragen möglich ist, die Entwicklung eines Themengebietes nachzuzeichnen. Details der Entwicklung konnten mit den entsprechenden Methoden genauer beleuchtet werden und ermöglichen es, das untersuchte Themengebiet besser zu verstehen. Keyplayer können entdeckt und Veränderungen in der Netzwerkstruktur einer Community dokumentiert werden. |



Dr. Dirk Tunger

Forschungszentrum Jülich, Zentralbibliothek
d.tunger@fz-juelich.de

Julia Wilhelm

Fachhochschule Köln, Fakultät für Informations- und Kommunikationswissenschaft



LIBERO v6

Software für Bibliotheksmanagement

Web-basiert

Neue Funktionen

Innovative Oberfläche

**Bibliothekartag Hamburg
Stand H 98**

LIB-IT GmbH • Riedbachstr. 5 • 74385 Pleidelsheim
www.lib-it.de • Tel: 07144 / 8041-0 • info@lib-it.de