

Erwartete Zitatraten und Klassifikationen

Bericht an das Bundesministerium für Bildung und
Forschung (BMBF)

Ulrich Schmoch, Carolin Michels, Nicole Schulze, Nicolai Mallig
unter Mitarbeit von
Antje Bierwisch, Esther Bollhöfer, Michaela Gigli, Bärbel Hüsing,
Frank Marscheider-Weidemann, Michael Möckel, Sven Wydra

Karlsruhe, März 2012

Inhaltsverzeichnis

Zusammenfassung	1
1 Einführung	5
2 Grundlegende Arbeiten	6
2.1 Implementation bibliometrischer Versionen von WoS und Scopus	6
2.2 Analyse grundlegender Strukturen der Datenbanken WoS und Scopus	8
3 Wachstumsberechnungen	28
4 Methodische Überlegungen zu Zitatraten	30
4.1 Zeitliche Erfassung von Publikationen und Zitaten	30
4.2 Whole count und fraktionierte Zählweise	31
4.3 Feldspezifische und zeitschriftenspezifische erwartete Zitatraten	34
4.4 Zeitschriftenspezifische Beachtung (ZB) und Internationale Ausrichtung (IA)	36
4.5 Relevanz von Eigenzitaten	42
4.6 Personenspezifische Zitatraten.....	43
4.7 Adäquate Ermittlung der Mittelwerte von Zitatraten.....	46
4.8 Zitatraten von Zeitschriften	48
4.8.1 Impact-Faktor	48
4.8.2 Audience-Faktor.....	50
4.8.3 SNIP.....	51
5 Klassifikation von Zeitschriften	53
5.1 Existierende Klassifikationen	54

5.2	Manuelle/intellektuelle Klassifikation ausgewählter Zeitschriften	59
5.3	Grundlegende Probleme der Klassifikation	63
5.4	Maschinelle Klassifikation von Zeitschriften	66
5.4.1	Datensätze	69
5.4.2	Ergebnisse.....	70
6	Maschinelle Klassifikation einzelner Artikel.....	81
7	Implementierung der Ergebnisse in die KB-Bibliometriedatenbank.....	83
8	Literatur.....	84
	Anhang.....	88

Tabellenverzeichnis

Tabelle 1:	Testabfragen zur Prüfung der Verarbeitungszeiten in WoS	7
Tabelle 2:	Typisches Niveau der Relation von in Scopus abgedeckten Publikationen und akademischem Personal für die Stichprobe im EUMIDA-Datensatz	9
Tabelle 3:	Anteile ausgewählter Länder in SCIE und SSCI an allen Publikationen	16
Tabelle 4:	Anteile ausgewählter Länder in Scopus an allen Publikationen	17
Tabelle 5:	Beobachtete Zitatraten für ausgewählte Länder und Regionen in SCIE und SSCI ohne Eigenzitate	18
Tabelle 6:	Beobachtete Zitatraten für ausgewählte Länder und Regionen in Scopus.....	19
Tabelle 7:	Index der Zeitschriftenspezifischen Beachtung für ausgewählte Länder und Regionen in SCIE und SSCI ohne Eigenzitate	21
Tabelle 8:	Index der Zeitschriftenspezifischen Beachtung für ausgewählte Länder und Regionen in Scopus	22
Tabelle 9:	Index der Internationalen Ausrichtung für ausgewählte Länder und Regionen in SCIE und SSCI ohne Eigenzitate	24
Tabelle 10:	Index der Internationalen Ausrichtung für ausgewählte Länder und Regionen in Scopus	25
Tabelle 11:	Anteile ausgewählter Länder innerhalb aller Publikationen in SCIE und SSCI des Jahres 2010 nach der fraktionierten und der Whole count Zählung und daraus folgende Länderrelationen	33
Tabelle 12:	Indizes der erwarteten Zitatraten für ausgewählte Länder nach unterschiedlichen Definitionen für alle Gebiete von WoS, 2007	37
Tabelle 13:	Aufschlüsselung der Zitate ausgewählter Länder anhand der Herkunft der Autoren des zitierenden Dokuments.....	38
Tabelle 14:	Vergleich der Rangfolge von acht Zeitschriften nach Impact-Faktor und Audience-Faktor im Vergleich	51
Tabelle 15:	Grobsystematik der ISI-Klassifikation	62

Tabelle 16:	Zuordnungsgenauigkeit in der Chemie und dem Maschinenbau nach den Ergebnissen der intellektuellen Klassifikation	63
Tabelle 17:	Übersicht über Gebiete, die in mehreren Disziplinen bearbeitet werden.....	65
Tabelle 18:	Definition von Biologie und Informatik über Subject Codes von WoS.....	69
Tabelle 19:	Eigenschaften der untersuchten Datensätze in Biologie und Informatik.....	70
Tabelle 20:	Einteilung der Dokumentenpaare nach Klassen- und Clusterzugehörigkeit.....	72

Abbildungsverzeichnis

Abbildung 1:	Zahl der Artikel in WoS und in Scopus, 2010	11
Abbildung 2:	Zahl der Artikel und Proceedings in WoS und in Scopus	12
Abbildung 3:	Publikationen in WoS und Scopus für ausgewählte Felder, 2010	13
Abbildung 4:	Publikationen in WoS und Scopus für disziplinäre Gebiete, 2010	14
Abbildung 5:	Artikel ausgewählter Länder in WoS und Scopus, 2010.....	15
Abbildung 6:	Beobachtete Zitatraten für ausgewählte Länder in SCIE und SSCI ohne Eigenzitate.....	20
Abbildung 7:	Beobachtete Zitatrate für ausgewählte Länder in Scopus (ohne Geisteswissenschaften, ohne Eigenzitate).....	20
Abbildung 8:	Index der Zeitschriftenspezifischen Beachtung für vier ausgewählte Länder Regionen in SCIE und SSCI ohne Eigenzitate	23
Abbildung 9:	Index der Zeitschriftenspezifischen Beachtung für vier ausgewählte Länder in Scopus.....	23
Abbildung 10:	Index der Internationalen Ausrichtung für vier ausgewählte Länder in SCIE und SSCI ohne Eigenzitate	26
Abbildung 11:	Index der Internationalen Ausrichtung für vier ausgewählte Länder in Scopus.....	26
Abbildung 12:	Vollständigkeit der Publikationen des Jahres 2009 sowie der Zitate auf das Jahr 2007 mit monatlichen Updates von Januar 2010 bis Juni 2011 in den Datenbanken SSCI und SCIE	31
Abbildung 13:	Anteil von Publikationen in WoS nach Herausgeberland, 2009	35
Abbildung 14:	Beobachtete durchschnittliche Zitatraten für vier ausgewählte Länder in SCIE und SSCI ohne Eigenzitate	39
Abbildung 15:	Index der Zeitschriftenspezifischen Beachtung für vier ausgewählte Länder in SCIE und SSCI ohne Eigenzitate	40
Abbildung 16:	Index der Internationalen Ausrichtung für vier ausgewählte Länder in SCIE und SSCI ohne Eigenzitate	41
Abbildung 17:	Illustration der Ermittlung des h-Indexes bei Sortierung der Publikationen anhand ihrer Zitierungen	44

Abbildung 18:	Ermittlung des h-Indexes für einen älteren und einen jüngeren Forschers des gleichen Fachgebiets.....	45
Abbildung 19:	Fiktive Verteilung der Zitatraten eines Autors mit 30 Publikationen.....	47
Abbildung 20:	"Circle of Science" nach Boyack und Klavans.....	54
Abbildung 21:	Verteilung der Mehrfachklassifikation von Artikeln in den Datenbanken WoS und Scopus, 2008-2010	56
Abbildung 22:	Entropy für den Datensatz Biologie	73
Abbildung 23:	Entropy für den Datensatz Informatik.....	74
Abbildung 24:	Purity für den Datensatz Biologie	74
Abbildung 25:	Purity für den Datensatz Informatik	75
Abbildung 26:	Jaccard-Index für den Datensatz Biologie.....	75
Abbildung 27:	Jaccard-Index für den Datensatz Informatik.....	76
Abbildung 28:	Rand-Statistik für den Datensatz Biologie	76
Abbildung 29:	Rand-Statistik für den Datensatz Informatik.....	77
Abbildung A1:	Referenzabfragen für die bibliometrische Version von WoS.....	88
Abbildung A2:	Datenbankschema von WoS.....	92
Abbildung A3:	Datenbankschema von Scopus.....	94
Abbildung A4:	Eigene Klassifikation von wissenschaftlichen Zeitschriften nach Fachgebieten und -themen.....	95

Zusammenfassung

Der vorliegende Bericht befasst sich im Kern zum einen mit der Normalisierung von Zitaten in Publikationsdatenbanken, um bei diesen als Maß für den Impact oder die Bedeutung einer Publikation eine Normierung einzuführen. Auf diese Weise sollen Unterschiede der Zitathäufigkeit, die auf Unterschieden in Fachgebieten, Zeitschriftengrößen oder unterschiedlichen Herausgeberländern beruhen, ausgeglichen werden. Zum zweiten befasst sich der Bericht mit der Klassifikation von Zeitschriften und Artikeln, da hiervon wesentlich die Qualität der Definition von Gebieten und die Präzision der fachlichen Analyse zu einzelnen Einrichtungen abhängen.

Um einen grundsätzlichen Eindruck in die Strukturen der Datenbanken zu erhalten, werden Scopus und Web of Science (WoS) miteinander verglichen. Dabei zeigt sich, dass der wesentliche Vorteil von Web of Science (WoS) in der zeitlich ausgedehnteren Abdeckung und der höheren Präzision der Klassifikation besteht. Demgegenüber ist die Zahl der jährlich von Scopus abgedeckten Dokumente höher, wobei vor allem die Ingenieurwissenschaften sowie die Schwellenländer, hier insbesondere China, breiter erfasst sind. Bei den Ingenieurwissenschaften ist in erster Linie die sehr viel umfangreichere Abdeckung von Proceedings hervorzuheben, die in diesem Bereich eine besonders hohe Bedeutung haben. Beim Vergleich der Zitatstrukturen zwischen Scopus und WoS ergibt sich für Scopus insgesamt ein etwas niedrigeres Zitatniveau, welches auf die breitere Abdeckung der Schwellenländer sowie der Ingenieurwissenschaften zurückzuführen ist. Insgesamt sind aber im Hinblick auf Ländervergleiche die Rangfolgen bei Industrieländern sowie die Trends sehr ähnlich, vor allem bei Schwellenländern kann es zwischen Web of Science (WoS) und Scopus Unterschiede bei den Rangfolgen geben. Es gibt keine methodischen Hinweise, welches Zitatniveau das objektiv "richtige" ist. Vielmehr ist im Einzelfall zu betrachten, welche Datenbankabdeckung für die jeweilige Analyse die sinnvollere ist.

Bei Web of Science (WoS) und Scopus ist seit dem Jahr 2000 eine erhebliche Ausweitung der abgedeckten Zeitschriften festzustellen, weshalb das Wachstum der Publikationszahlen in einem Gebiet nicht mit einem Wachstum der diesbezüglichen wissenschaftlichen Aktivitäten gleichgesetzt werden kann. Vielmehr ist eine Trennung zwischen Wachstum der Wissenschaft und Ausweitung der Datenbank vorzunehmen, wofür ein geeignetes Verfahren erarbeitet worden ist.

Bei der Analyse von Forschungseinheiten oder auch Ländern stellt sich immer die Frage der Aktualität der Daten. Hierbei ist in Rechnung zu stellen, dass Publikationen in Zeitschriften oder Proceedings nicht unmittelbar an die Datenbankproduzenten gemeldet werden, sondern in vielen Fällen eine deutliche Zeitverzögerung auftritt. Hier wurde

geprüft, zu welchem Zeitpunkt die Daten des Vorjahres als weitgehend vollständig angesehen werden können. Diese Analysen haben gezeigt, dass Ende April eines Jahres die Publikationszahlen des Vorjahres weitgehend vollständig sind, sodass statistisch aussagefähige Analysen möglich sind. Bei der Nutzung eines dreijährigen Zitatfensters sind die Zitatdaten für das drei Jahre zurückliegende Publikationsjahr nahezu vollständig.

Bei der Analyse von Publikationen mit Autoren aus unterschiedlichen Ländern ist es möglich, die Publikationen den Ländern entweder fraktioniert oder jeweils vollständig (Whole count) zuzurechnen. Bei dieser Frage hat sich bei der Betrachtung der Vor- und Nachteile ein Votum zugunsten der Whole count-Zählweise bei internationalen Ko-Publikationen ergeben.

Die Zitatraten sind je nach Disziplin deutlich verschieden. Außerdem gibt es starke Abhängigkeiten vom Herausgeberland der Zeitschrift sowie deren Größe. Um vor allem fachübergreifend zu vergleichbaren Werten zu gelangen, ist es üblich, eine Normierung der Zitatraten mit erwarteten Zitatraten vorzunehmen. Ein weit verbreitetes Verfahren ist dabei die Ermittlung der feldspezifischen erwarteten Zitatraten. Es hat sich allerdings gezeigt, dass die Ermittlung der zeitschriftenspezifischen erwarteten Zitatraten demgegenüber erhebliche Vorteile hat. Dieser Ansatz ist vor allem geeignet, die Nachteile zu kompensieren, die sich für ein Land wie Deutschland mit großem eigenem Sprachraum beim Vergleich mit anderen Ländern ergeben.

Aus den zeitschriftenspezifisch erwarteten Zitatraten lassen sich die Indikatoren *Zeitschriftenspezifische Beachtung* (ZB) und *Internationale Ausrichtung* (IA) ableiten. In ihrer Kombination sind differenzierte Analysen hinsichtlich der Strukturen möglich, die bei den beobachteten Zitatraten ermittelt wurden. In jedem Fall ergeben sich aus der Kombination verbesserte Interpretationsmöglichkeiten.

Die Zahl der Zitate und die auf dieser Basis berechneten Zitatraten werden in der Regel als Impact oder Einfluss einer Publikation interpretiert. In dieser Perspektive macht es keinen Sinn, Eigenzitate bei der Zahl der Zitate mit zu berücksichtigen. Es gibt zwar bei der Ermittlung der Eigenzitate methodische Probleme. Doch erweist sich insgesamt der Ausschluss der Eigenzitate als der empfehlenswertere Weg.

Die Ermittlung der Zitatraten für einzelne Individuen ist methodisch problematisch, da aus statistischer Sicht die Stichprobe sehr klein und damit fehleranfällig ist. Darüber hinaus kann es Besonderheiten geben, dass wichtige Publikationen nicht häufig oder aber sehr spät zitiert werden. Vor diesem Hintergrund ist eigentlich grundsätzlich von der Analyse von Individuen über bibliometrische Analysen abzuraten. Dennoch geschieht dieses in der Praxis, etwa bei Berufungen, sehr häufig. Wenn dieses schon

nicht vermeidbar ist, so sollte auf jeden Fall die Anwendung des sogenannten Hirsch-Index (h-Index) vermieden werden, da hier fachliche Unterschiede nicht ausgeglichen werden und außerdem das Alter des Autors bzw. der Autorin eine erhebliche Rolle spielt.

Bei der Berechnung der Mittelwerte von Zitatraten stehen sich die Verfahren Averages of Ratios und Ratios of Averages gegenüber. Eine genauere Betrachtung zeigt, dass sich hieraus in der Praxis keine nennenswerten Unterschiede ergeben, sodass es sich letztlich nur um einen wissenschaftsinternen Streit handelt. Im Zweifelsfalle ist der Ansatz Ratios of Averages vorzuziehen, da hier vereinzelte Extremwerte besser ausgeglichen werden.

Es ist üblich, die Zeitschriften nach ihrem sogenannten Impact-Faktor zu bewerten, wobei es sich letztlich um abgeleitete Zitatraten handelt. Eine genauere Analyse hat gezeigt, dass die Berechnung sehr zeitaufwändig ist und dass die Impact-Faktoren für die Analyse einzelner Einrichtungen, Länder oder auch Personen wenig hilfreich sind. Von daher wird in dem Bericht die Empfehlung ausgesprochen, die Impact-Faktoren nicht für alle Zeitschriften in den Datenbanken Web of Science (WoS) und Scopus zu ermitteln. Dieselbe Aussage gilt für die verbesserten Varianten des Impact Faktors Audience-Faktor und SNIP, die zwar dem Impact-Faktor in verschiedener Hinsicht überlegen sind, deren Berechnung aber noch aufwändiger ist.

In den großen bibliometrischen Datenbanken WoS und Scopus werden die Artikel nicht einzeln fachlich klassifiziert, sondern letztlich nur die Zeitschriften. Dafür bieten die Datenbankenproduzenten von Web of Science (WoS) und Scopus jeweils eigene Klassifikationsschemata an. Viele Zeitschriften sind dabei mehreren Klassen zugeordnet. Es gibt verschiedene Unzulänglichkeiten dieser Klassifikationen hinsichtlich der fachlichen Beschreibung einzelner Bereiche und hinsichtlich der Präzision der Klassifikation. Vor allem in Scopus ist hier die Zuverlässigkeit der Klassifikation eingeschränkt. Angesichts dieser Situation, die in weniger gravierender Form auch für Web of Science (WoS) gilt, gibt es eine Reihe von Vorschlägen, die Klassifikation aus einer bibliometrischen Sicht zu verbessern. Dabei wurden vor allem Verfahren der bibliometrischen Verknüpfung genutzt. Darüber hinaus gibt es eine Reihe von Klassifikationen der Wissenschaft für unterschiedliche Anwendungszwecke. Im Projekt wurde eine eigene Klassifikation entwickelt, wobei ein Kompromiss zwischen fachlicher Klassifikation und Zeitschriften-Klassifikation erarbeitet wurde. Zu diesem Zweck wurde rund ein Drittel der Publikationen in Web of Science (WoS) und Scopus manuell/intellektuell klassifiziert. Dabei wurde festgestellt, dass ein erheblicher Teil der Klassifikationen in beiden Datenbanken unzutreffend ist, und dass es mit der Zielrichtung einer maschinellen Klassifikation ein grundsätzliches Problem der Zuordnung einzelner Themenbereiche

gibt, die von mehreren Disziplinen behandelt werden. Denn aufgrund Ihrer großen inhaltlichen Ähnlichkeit wird es bei maschinellen Verfahren nicht möglich sein, diesen eine feste Disziplin zuzuordnen. Hier wird der Vorschlag gemacht, diese Gebiete jeweils nur einer der in Frage kommenden Disziplinen zuzuordnen.

Grundsätzlich sollte das Ziel sein, jede Zeitschrift nach ihrem fachlichen Schwerpunkt nur einfach zu klassifizieren. Bei vielen fachübergreifenden Zeitschriften wird dieses allerdings nicht möglich sein. Hier sollte das Ziel sein, nach dem Schwerpunkt-Prinzip maximal drei Klassen zu vergeben.

Für die Fachgebiete Biologie und Informatik wurde ein maschinelles Verfahren zur Klassifikation von Zeitschriften geprüft, wobei sich sehr gute Übereinstimmungen mit der parallel manuell durchgeführten Klassifikation ergaben. Es wird somit möglich sein, bei der Analyse spezifischer Fachgebiete ausgehend von den bereits manuell klassifizierten Zeitschriften als Start-Set, alle relevanten Zeitschriften auf maschinellem Wege zu ermitteln. Bei einer Klassifikation sämtlicher Zeitschriften in den Datenbanken auf maschinellem/automatischem Wege wird es auch nach mehreren Iterationsschritten einen Rest von Zeitschriften geben, der bei diesem Verfahren nicht sinnvoll klassifiziert werden kann. Dieses wird vor allem Zeitschriften betreffen, die Disziplinenübergreifend auf bestimmte Themen orientiert sind. Hier wird es einen Rest von Zeitschriften geben, der nach wie vor manuell klassifiziert werden muss. Die Analysen in diesem Bericht haben allerdings gezeigt, dass dieses mit vertretbarem Aufwand realisierbar ist.

1 Einführung

Der vorliegende Bericht mit dem Titel "Erwartete Zitatraten und Klassifikationen" befasst sich zunächst einmal mit Zitaten, die in bibliometrischen Analysen benutzt werden, um den Impact (Einfluss) von Publikationen zu bestimmen. Annahme ist dabei, dass sehr häufig zitierte Publikationen einen höheren Impact haben als weniger häufig zitierte. In der Regel geht es dabei nicht um den Impact einer einzelnen Publikation, sondern den durchschnittlichen einer Teilmenge. Faktisch wird der Impact vielfach mit Qualität gleichgesetzt. Zitate haben daher eine zentrale Bedeutung in bibliometrischen Analysen. Allerdings gibt es erhebliche Unterschiede bei den Zitatraten abhängig von dem Fachgebiet, der Größe und Herkunft der Zeitschrift und anderen Parametern. Um daher zu sinnvollen, inhaltlich belastbaren Aussagen zu kommen, müssen die ermittelten oder auch die beobachteten Zitatraten in geeigneter Weise standardisiert werden. Diese Standardisierung ist ein zentrales Element bibliometrischer Analysen, weshalb in der geeigneten Nutzung erwarteter Zitatraten eine zentrale Aufgabe des Kompetenzzentrums gesehen wird.

Voraussetzung für den Umgang mit diesen relativ komplexen bibliometrischen Indikatoren ist allerdings, dass die Datenbankstruktur bekannt ist und daraus resultierende Besonderheiten in entsprechender Weise interpretiert werden können. Von daher ist es auch Aufgabe dieses grundlegenden Berichts, die Strukturen der beiden Datenbanken WoS und Scopus, die dem Kompetenzzentrum für Analysen zur Verfügung stehen, näher zu untersuchen.

Neben der Diskussion der grundlegenden Strukturen der Datenbanken und der verschiedenen Varianten von erwarteten Zitatraten ist es ein weiterer Schwerpunkt des vorliegenden Berichts, die Klassifikation der Publikationen und damit letztlich die Klassifikation von Zeitschriften genauer zu untersuchen. Von einer sinnvollen Klassifikation hängt es ab, mit welcher Qualität zu untersuchende Gebiete definiert werden können und ob für untersuchte Einheiten wie Forschungseinrichtungen oder auch Länder, aussagefähige fachliche Profile bestimmt werden können. Außerdem werden erwartete Zitatraten häufig nach Fachgebieten bestimmt, weshalb auch hier die Qualität der Klassifikation zum Tragen kommt. Die Arbeit an den Klassifikationen ist damit ebenso grundlegend wie auch die Analyse der Zitatraten.

2 Grundlegende Arbeiten

2.1 Implementation bibliometrischer Versionen von WoS und Scopus

Für das Projekt *Erwartete Zitatraten und Klassifikationen* mussten große Datenmengen analysiert und verschiedene Teilfelder aus WoS bzw. Scopus zueinander in Beziehung gesetzt werden. Es genügte dafür nicht, begrenzte Teilfelder für die weitere Bearbeitung von den zentral durch FIZ Karlsruhe bereitgestellten Rohdaten herunterzuladen. Da im Jahr 2010 für das Kompetenzzentrum insgesamt keine relationalen bibliometrischen Versionen von WoS und Scopus vorlagen, diese aber für das Projekt zwingend notwendig waren, hat das Fraunhofer ISI eigene bibliometrische Versionen dieser Datenbanken aufgesetzt. Die entsprechenden Datenbankschemata sind in der Anlage dokumentiert (siehe Anhang Abbildung A2 und Abbildung A3). Zum Aufbau dieser Datenbanken wurde auf die beim FIZ Karlsruhe zur Verfügung stehende zentrale Version der Rohdaten zurückgegriffen und ein regelmäßiges Update vorgenommen. *Bibliometrische Version* bedeutet hier, dass nicht alle Felder der Datenbanken erfasst sind, wohl aber die für bibliometrische Analysen üblichen. Es handelt sich somit um Teildatensätze mit dem Ziel, die Antwortzeiten bei Abfragen zu verkürzen und klare Beziehungen zwischen Teilfeldern der Datenbanken herstellen zu können.

Eine erste Version beider bibliometrischer Datenbanken lag im März 2010 vor. Zu diesem Zeitpunkt wurden die Daten auf einem Server implementiert, auf dem gleichzeitig die Patentdatenbank PATSTAT aufgesetzt war, was durch die gleichzeitige Durchführung größerer Patentanalysen durch andere Arbeitsgruppen im Haus zu erheblichen zeitlichen Konflikten führte. Vor diesem Hintergrund hat das Fraunhofer ISI im Rahmen einer Mittelumwandlung einen separaten Server beschafft, auf dem jetzt für beide Datenbanken bibliometrischen Versionen installiert sind.

Für eine effiziente Bearbeitung des Projektes war es erforderlich, dass aufwändigere SQL-Befehle, bei denen große Datenmengen miteinander verknüpft werden, in einem überschaubaren Zeitraum realisiert werden konnten. Nur so war es möglich, zur Bearbeitung eines spezifischen Problems verschiedene Varianten von SQL-Kommandos zu prüfen und auf iterativem Wege zu optimieren. Deshalb wurden einfache Standardabfragen entwickelt und die entsprechende Verarbeitungszeit mit der bibliometrischen Version von WoS getestet. Eine Übersicht findet sich in Tabelle 1, die vollständige Fassung ist in der Abbildung A1 im Anhang dokumentiert.

Auch die zentrale Version der Datenbank WoS, die dem Kompetenzzentrum seit August 2011 über das FIZ Karlsruhe zur Verfügung gestellt wurde, ermöglicht es nicht,

die eigene Version des WoS zu ersetzen. So ist es beispielsweise bei der Recherche nach Universitätspatenten erforderlich, bibliometrische und Patent-Datenbanken miteinander zu koppeln, was innerhalb des Fraunhofer ISI realisierbar ist, bei einer Verbindung mit einem externen Server jedoch erhebliche Probleme bereiten würde. Bei anderen Fragestellungen, etwa beim "Brain Drain Projekt", ist das Anlegen großer Zwischentabellen erforderlich, was ebenfalls auf einem eigenen Server deutlich einfacher zu realisieren ist. Im Übrigen steht nach wie vor die zentrale Bereitstellung einer bibliometrischen Version von Scopus aus. Schließlich erweist es sich als wenig effizient, das Datenbankschema dauerhaft "einzufrieren", da in jedem Projekt neue Fragestellungen auftauchen, die zu einer weiteren Optimierung des Schemas führen können. Von daher sind die Versionen in Abbildung A2 und Abbildung A3 im Anhang erste Versionen, die inzwischen weiter verbessert wurden.

Tabelle 1: Testabfragen zur Prüfung der Verarbeitungszeiten in WoS

Aufgabe	Zeit (hh:mm:ss.ms)
1 Deutsche Publikationen im Jahr 2003 (Ohne Proceedings)	Abgelaufen: 00:00:30.07
2 Publikationen der alten EU-Länder (EU15) im Jahr 2005	00:00:55.04
3 Zitate zu Publikationen nach Befehl 2) im Drei-Jahres_Fenster (2005-2007)	00:03:30.98
4 Zeitreihe 2000 bis 2008 der Artikel (ohne Proceedings) mit dem Stichwort "nano" mit rechtsseitiger offener Trunkierung	00:02:25.65
5 Proceedings der TU Braunschweig im Themenbereich "Computertechnik" im Zeitraum 2003-2007	00:00:29.95
6 Durchschnittliche Seitenzahl der Artikel (ohne Proceedings) in der Physik für das Jahr 2006 mit einer Band/Volumen-Nr von 10 und mehr	00:00:41.69
7 Zitatquote der Artikel (Ohne Proceedings) in der Chemie des Jahres 2004 (5 Jahres-Fenster 2004-2008) differenziert nach Zeitschriften mit us-amerikanischer, englischer und deutscher Herausgeberschaft	00:00:51.54

Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

2.2 Analyse grundlegender Strukturen der Datenbanken WoS und Scopus

Für das Kompetenzzentrum stehen die Datenbanken WoS und Scopus für bibliometrische Recherchen zur Verfügung. Um für Analysen zu inhaltlich belastbaren Ergebnissen zu kommen, ist es erforderlich, die grundsätzlichen Strukturen dieser Datenbanken zu kennen. Dabei geht es zum einen um systematische Kodierungsprobleme, zum anderen auch um die Abdeckung von Fachgebieten und Dokumenttypen, insbesondere auch im Vergleich beider Datenbanken.

Ein grundsätzliches Problem beider Datenbanken besteht darin, dass sie auf Informationen der Zeitschriftenherausgeber zurück greifen müssen und dort nicht alle vollständig und korrekt vorliegen. Das betrifft unter anderem die Erfassung der Sitzländer der Einrichtungen der Autoren, die bei einer Reihe von Artikeln fehlen. Bei WoS (WoS) waren es für die Artikel aus dem Jahr 2008 und 2010 3,0% beziehungsweise 2,3%. Bei Scopus lag der Anteil fehlender Länderangaben bei den Artikeln im Jahr 2008 bei 9,2% und in 2010 bei 6,3%. Die niedrigere Quote bei WoS dürfte darauf zurückzuführen sein, dass bei vielen Artikeln zumindest die Städtenamen oder andere Ortsangaben vorhanden sind und damit Rückschlüsse auf die Länder möglich sind. Insbesondere die Tatsache, dass die 3,0% bzw. 2,3% der fehlenden Länderangaben exakt dem Anteil entsprechen, für die weder Adress-, noch sonstige Einrichtungsinformationen in der Datenbank vorliegen ist besonders auffallend. In WoS dürften also solche Ergänzungen vorgenommen worden sein. In Scopus ist der Anteil fehlender Länderangaben grundsätzlich höher. Interessanterweise ist aber der Anteil der Einträge größer, für die keine sonstigen Adress- oder Einrichtungsinformationen vorhanden sind. Dieser Anteil lag für Artikel in Scopus im Jahr 2008 bei 7,2% und in 2010 bei 5,8%. Generell gibt es folglich zwar Potenzial zur Ergänzung fehlender Ländercodes, der Anteil fehlender Ländercodes wird aber vergleichsweise größer bleiben als bei WoS.

Um vertiefte Erfahrungen mit Recherchen in beiden Datenbanken zu gewinnen, wurde in dem Projekt EUMIDA für die Europäische Kommission, an dem das Fraunhofer ISI mitgearbeitet hat, eine begrenzte bibliometrische Analyse durchgeführt. Im EUMIDA-Projekt ging es darum, eine Reihe von Indikatoren, vor allem Studenten-, Personal- und Budgetzahlen für alle Hochschulen in der Europäischen Union sowie für die der Schweiz und Norwegen zusammenzustellen. Daraus entstand ein Datensatz für rund 2.500 Hochschulen. Die bibliometrische Pilotstudie, deren Ergebnisse in Bonaccorsi et al. (2010) publiziert sind, bezog sich spezifisch auf 57 ausgewählte Hochschulen aus allen beteiligten Ländern mit verschiedenen fachlichen Ausrichtungen. Die Analysen führten zu methodisch interessanten Ergebnissen, insbesondere zu den folgenden Aspekten:

- Länderabdeckung bei Scopus,
- Publikationsintensität in verschiedenen Fachgebieten,
- Institutionen-Kodierung in WoS und Scopus,
- Abdeckung der Ingenieurwissenschaften in WoS und Scopus,
- Niveau der Zitatraten in WoS und Scopus.

Bei der Länderabdeckung ergeben sich auf der europäischen Ebene leichte Verzerrungen aus fehlenden Länderkodierungen in Scopus.

Sowohl in Scopus als auch in WoS zeigen sich erhebliche Unterschiede in der Publikations-Intensität und den Zitatraten nach Fachgebieten. Differenziert wurde in der Untersuchung zwischen Geistes-, Sozial-, Natur- und Ingenieurwissenschaften sowie Landwirtschaft und Medizin. Dabei ergab sich insbesondere eine sehr niedrige Publikations-Intensität in den Geistes- und Sozialwissenschaften, eine überdurchschnittliche in der Landwirtschaft und vor allem in der Medizin. Da in dem Projekt für die untersuchten Hochschulen auch Personalzahlen vorlagen, ließ sich eine Abschätzung der Zahl der Publikationen pro Personal vornehmen. Diese Abschätzung ist allerdings grob, da nicht für alle untersuchten Einrichtungen auch fachdifferenzierte Personalzahlen vorlagen. Die Ergebnisse sind in Tabelle 2 dokumentiert.

Tabelle 2: Typisches Niveau der Relation von in Scopus abgedeckten Publikationen und akademischem Personal für die Stichprobe im EUMIDA-Datensatz

Field	Score
Humanities	0.04
Social sciences	0.06
Natural sciences	0.20
Engineering	0.20
Agriculture	0.60
Medicine	0.60

Quelle: Bonaccorsi u.a. (2010: 81)

Die extrem niedrigen Publikationszahlen für die Geistes- und Sozialwissenschaften haben verschiedene Gründe. Zunächst ist zu bedenken, dass in diesen Fachgebieten neben englischsprachigen Zeitschriften vor allem solche in den nationalen Sprachen intensiv genutzt werden, die jedoch in den beiden Datenbanken in der Regel nicht abgedeckt sind. Zum zweiten sind Publikationen in Sammelbänden oder auch das Verfassen von Monographien in diesen Gebieten häufig auftretende Publikationsformen. Schließlich ist zu beachten, dass es in diesen Fachgebieten grundsätzlich weniger als

in den Naturwissenschaften oder der Medizin publiziert wird, dass es also "Kulturunterschiede" zwischen den Fachgebieten gibt. In der Konsequenz bedeutet dies, dass bei bibliometrischen Analysen stets fachliche Differenzierungen erforderlich sind. Bei bibliometrischen Rankings von Universitäten werden aus diesen Gründen ohne fachliche Differenzierung grundsätzlich Universitäten mit medizinischer oder naturwissenschaftlicher Orientierung bevorzugt, wogegen solche mit geistes- und sozialwissenschaftlicher Ausrichtung nicht maßgeblich in Erscheinung treten, da ihre Publikationszahlen in den Datenbanken aus den genannten Gründen vergleichsweise niedrig liegen. Die fachliche Differenzierung bibliometrischer Analysen erweist sich somit als zentraler Gesichtspunkt.

Bei der Kodierung der Institutionen gibt es wesentliche Unterschiede zwischen WoS und Scopus. In WoS werden durch den Produzenten Thomson Reuters bei den Institutionennamen, die auf den Angaben in den jeweiligen Publikationen beruhen, Standardisierungen vorgenommen wie etwa die Verkürzung von "Universität", "University" oder "Universidad" zu "Univ". Das Ziel dieser Standardisierung ist eine Vereinfachung von Recherchen nach einzelnen Einrichtungen und eine Reduzierung der unterschiedlichen Schreibweisen. In Scopus werden dagegen die Originalangaben aus den Publikationen übernommen und nicht weiter bearbeitet. Dieser Unterschied führt dazu, dass die Recherchen nach Institutionen in WoS auf den ersten Blick einfacher und erfolgreicher als in Scopus sind. Bei genauerer Betrachtung zeigt sich jedoch, dass von den Bearbeitern von Thomson ähnliche Namen unterschiedlicher Einrichtungen oft zu einem Namen zusammengefasst werden. Dieses betrifft vor allem osteuropäische Einrichtungen und dort solche, wo mehrere Hochschulen an einem Ort existieren. Da in WoS die Originalangaben nicht mehr verfügbar sind, lässt sich so nicht rekonstruieren, welche Institutionen hinter einem spezifischen Namen stehen. Dieses betraf in der verwendeten Stichprobe nicht nur wenige Ausnahmefälle, sondern etwa 10% des Datensatzes. Mit Scopus waren dagegen alle Einrichtungen eindeutig identifizierbar. Das bedeutet in der Konsequenz, dass derartige Recherchen nach Institutionen ausschließlich auf der Basis von Scopus durchgeführt werden sollten, da beim Arbeiten mit WoS häufig Fehler auftreten, ohne dass dieses ohne aufwändige Prüfung in Erscheinung träte.

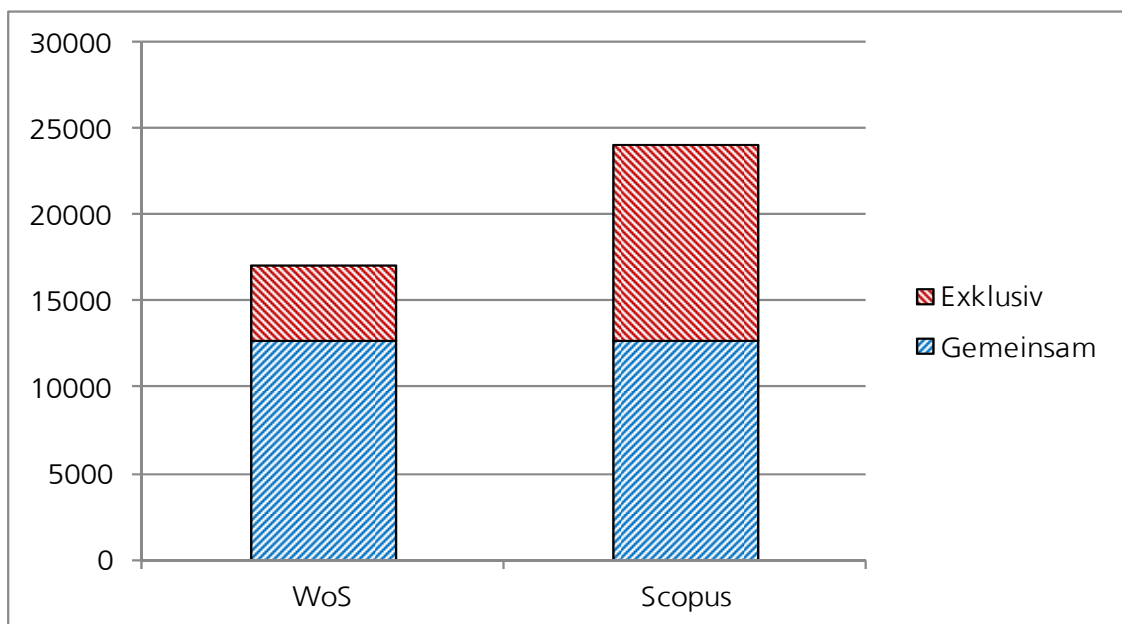
Eine Analyse für die Expertenkommission Forschung und Innovation, die im Februar 2012 erschienen ist, konnte auf den bis dahin vorliegenden Erfahrungen aus diesem Projekt ausbauen. Hierbei kamen auch die aktuellsten bibliometrischen Versionen von WoS und Scopus zum Einsatz.

Im Einzelnen wurde analysiert:

1. Gegenwärtige und frühere Abdeckung der Datenbanken WoS und Scopus,
2. Abdeckung wissenschaftlicher Disziplinen,
3. Abdeckung von Ländern,
4. Zitatniveau.

Um die gegenwärtige Abdeckung von Publikationen in beiden Datenbanken vergleichen zu können, wurde die Zahl der Zeitschriften untersucht, die in beiden Datenbanken, sowie die, die lediglich in einer von ihnen erfasst sind. Die unterschiedliche Abdeckung der Zeitschriften ist in Abbildung 1 dargestellt. Demgemäß ist der Überlappungsgrad von beiden Datenbanken ziemlich groß: etwa 74% der Zeitschriften in WoS sind auch in Scopus abgedeckt. Der Anteil der exklusiven Zeitschriften ist weitaus größer und liegt bei 47% der Scopus Zeitschriften. Im Gegensatz dazu liegt der Anteil der exklusiven WoS-Zeitschriften bei etwa 25%.

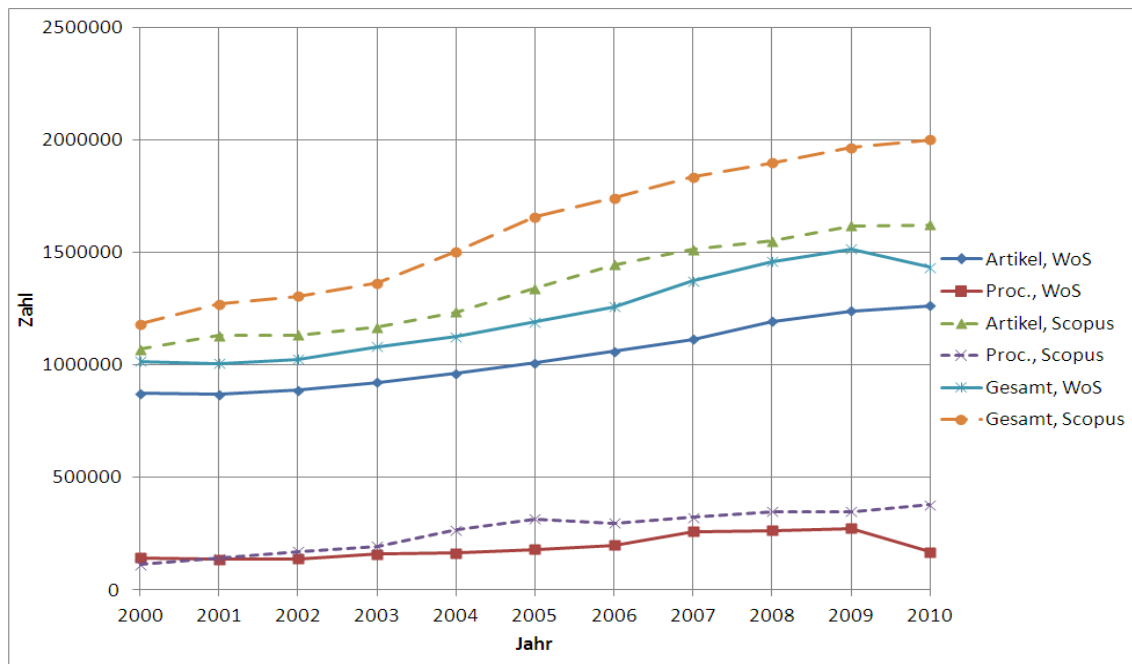
Abbildung 1: Zahl der Artikel in WoS und in Scopus, 2010



Quelle: WoS und Scopus, Recherchen und Berechnungen des Fraunhofer ISI

Beim Vergleich der Entwicklung der Publikationen in Scopus und WoS erweist sich die Zahl der Artikel in Scopus als wesentlich höher als in WoS mit einem Anteil von etwa 28% in 2010 (Abbildung 2). In allen Jahren ist die Zahl der Proceedings in Scopus höher als die in WoS. Seit 2000 ist das Wachstum beider Datenbanken erheblich, offensichtlich gibt es einen Wettbewerb im Hinblick auf die Zeitschriften-Abdeckung, wobei WoS zu keinem Zeitpunkt die Größe von Scopus erreicht hat. Im Jahr 2010 betrug die Zahl der Artikel in WoS etwa 1.264.000, die Zahl der Proceedings 168.000. Die Zahl der Artikel in Scopus erreichte im selben Jahr 1.621.000, die der Proceedings 379.000.

Abbildung 2: Zahl der Artikel und Proceedings in WoS und in Scopus

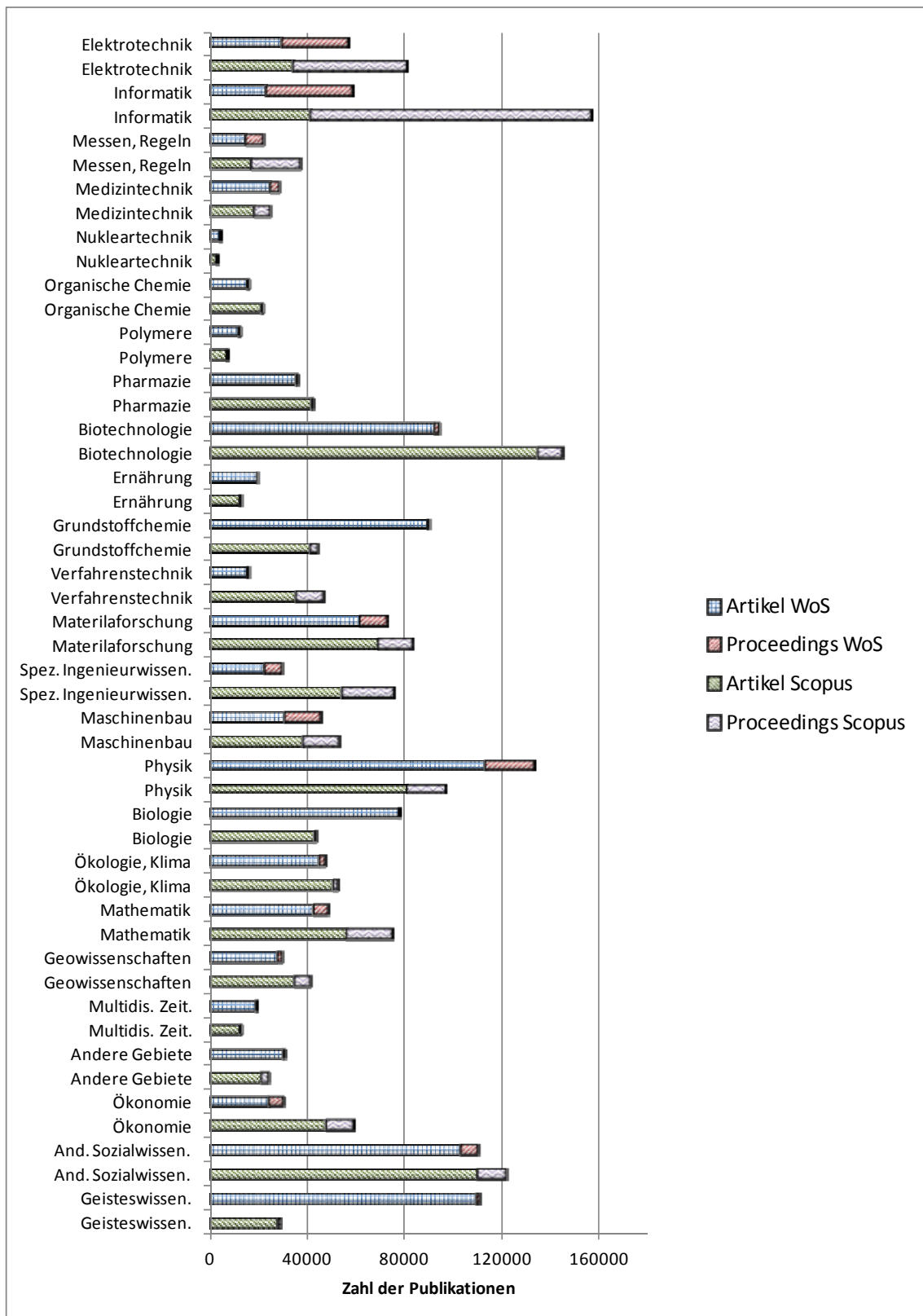


Quelle: WoS, Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Bei der Betrachtung der Unterschiede der Abdeckung von Disziplinen muss die Situation in jedem Feld einzeln betrachtet werden. So ist im Feld "Elektrotechnik" die Zahl der Publikationen als auch die der Proceedings in Scopus größer als in WoS, in "Physik" ist sowohl die Zahl der Publikationen als auch der Proceedings in WoS höher als in Scopus (Abbildung 3)¹. Die bemerkenswertesten Vorteile von Scopus im Hinblick auf die Zahl der Publikationen kann in den Feldern "Computer" und "Biotechnologie" festgestellt werden. Insbesondere in "Computer" ist die Zahl der Proceedings erheblich höher als in WoS. Dieses ist adäquat, da in der Informatik Proceedings sehr viel bedeutsamer als in anderen Feldern sind und daher als Äquivalent zu Artikeln betrachtet werden. In den Sozialwissenschaften ist die Abdeckung von Scopus breiter als in WoS, mit einer Quote von etwas 25%, in den Wirtschaftswissenschaften sogar mit einer Quote von 90%. Im Unterschied repräsentieren die Artikel in Scopus in den Geisteswissenschaften lediglich ein Viertel derer in WoS.

¹ Zur Erstellung von Abbildung 3 wurden die Klassifikationscodes der Hersteller von Scopus und WoS zur Definition spezifischer Felder benutzt. Selbstverständlich ist der Inhalt recht ähnlich, aber nicht identisch. Außerdem war es erforderlich, die Klassifikationen in fraktionierter Form zu zählen, da viele Artikel mehrfach klassifiziert sind, in Scopus erheblich häufiger als in WoS. Deshalb ist ein adäquater Vergleich nur auf der Basis einer fraktionierten Zählung möglich.

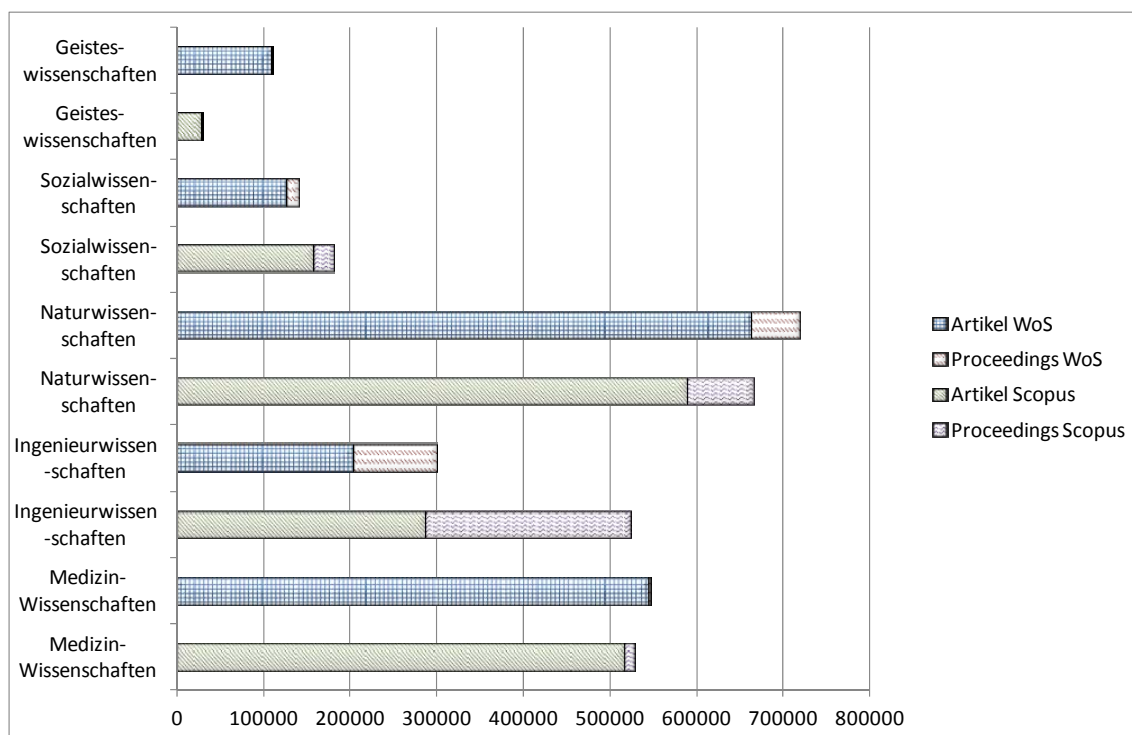
Abbildung 3: Publikationen in WoS und Scopus für ausgewählte Felder, 2010



Quelle: WoS und Scopus, Recherchen und Berechnungen des Fraunhofer ISI

In der breiteren Betrachtung disziplinärer Gebiete sind die Natur- und Medizinwissenschaften in WoS etwas besser repräsentiert als in Scopus (Abbildung 4). In den Ingenieurwissenschaften dagegen ist die Abdeckung von Scopus erheblich besser als in WoS, insbesondere ist die Zahl der Proceedings substantiell höher. Letzteres ist besonders wichtig für Analysen in den Ingenieurwissenschaften, da in diesem Gebiet die Relevanz von Proceedings wesentlich höher als in anderen Gebieten ist. Der entscheidende Unterschied in der Gesamtzahl der Publikationen ist somit im Wesentlichen auf die breitere Abdeckung der Ingenieurwissenschaften in Scopus zurückzuführen.

Abbildung 4: Publikationen in WoS und Scopus für disziplinäre Gebiete, 2010

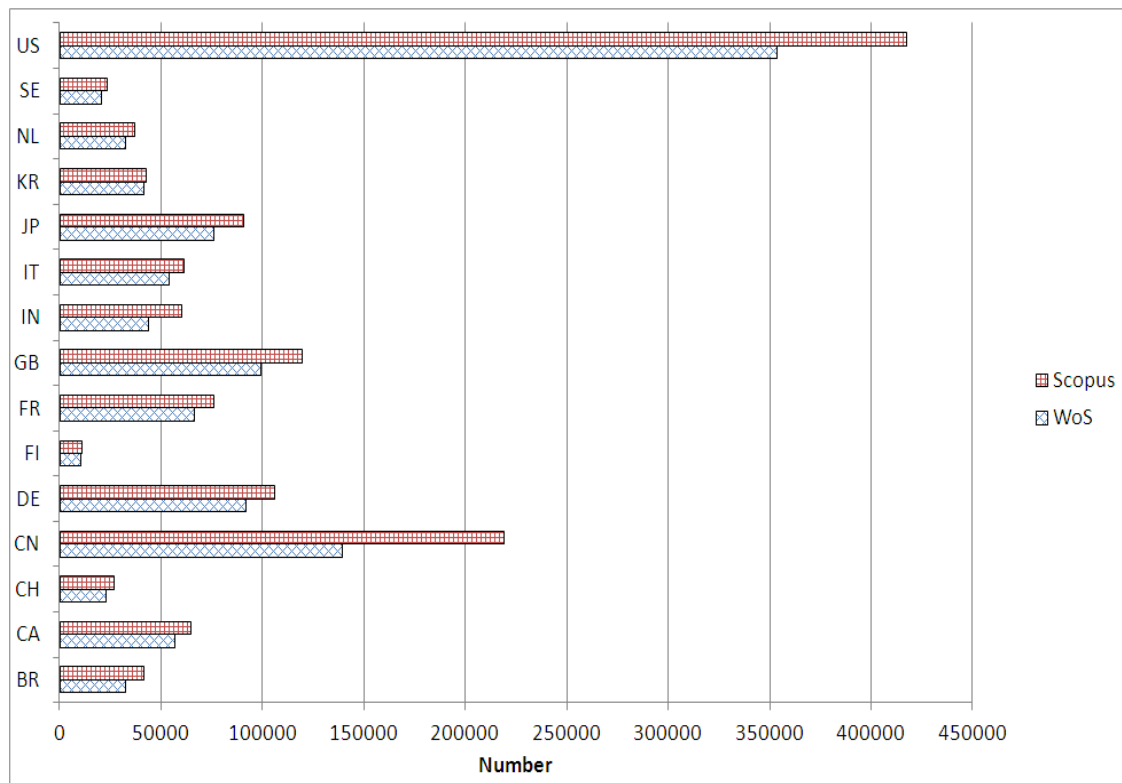


Quelle: WoS, Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Im Hinblick auf die Abdeckung von Ländern in WoS und Scopus äußert sich die höhere Zahl der Artikel in Scopus in einer höheren Zahl von Artikeln pro Land (Abbildung 5). Allerdings sind die Länder in unterschiedlicher Form betroffen. Die industrialisierten Länder erzielen etwa 10 bis 15% mehr an Publikationen, z.B. Deutschland 15%. Die USA erreichen sogar 18% mehr Publikationen und Großbritannien 21%. Die zusätzliche Zahl von Artikeln ist für Entwicklungs- und Schwellenländer noch ausgeprägter. Die zusätzliche Zahl für Brasilien in Scopus ist 28%, für Indien 38%. Der bemerkenswerteste Unterschied ist für China mit 57% in 2010 festzustellen. Die entscheidende Ausnahme in diesem Gesamtbild für Entwicklungs- und Schwellenländer ist Südkorea mit nur 3%. Alles in allem ist der entscheidende Unterschied bei der Länderabdeckung

die erheblich breitere Abdeckung chinesischer Publikationen. Die substantielle Erfassung von Publikationen englischsprachiger Länder wie den USA und Großbritannien ist in Scopus ähnlich wie in WoS, sodass auch hier ein entsprechender Sprach-Bias vorliegt.

Abbildung 5: Artikel ausgewählter Länder in WoS und Scopus, 2010



Quelle: WoS und Scopus, Recherchen und Berechnungen des Fraunhofer ISI

Für einen detaillierteren Vergleich von WoS und Scopus im Hinblick auf Zitationen wurden verschiedene Analysen für ausgewählte Länder in beiden Datenbanken parallel ausgeführt. Die Analysen beginnen erst in 2005, da die ersten Jahre der Abdeckung von Zeitschriften in Scopus recht unsystematisch waren und gemäß dem Hersteller eine wesentliche Verbesserung im Jahr 2004 erfolgte. Um hier auf der sicheren Seite zu sein, begannen die Analysen im Jahr 2005.

Bei der Betrachtung der Länderanteile in den Gesamtdatenbanken ergeben sich Strukturen für WoS und Scopus entsprechend Tabelle 3 und Tabelle 4. Als generelle Beobachtung ist der Anteil der industrialisierten Länder in Scopus niedriger als in WoS, vor allem aufgrund des höheren Anteils von China. Die anderen Schwellenländer haben in Scopus einen ähnlichen Anteil wie in WoS im Jahre 2010, allerdings hatten sie schon im Jahre 2005 in Scopus einen höheren Anteil. In dieser Hinsicht hat WoS seine Abdeckung in den letzten Jahren deutlich erhöht.

Tabelle 3: Anteile ausgewählter Länder in SCIE und SSCI an allen Publikationen (Whole counts)

Land	2005	2006	2007	2008	2009	2010
USA	30,9	30,3	29,4	28,7	28,1	28,0
Japan	7,8	7,4	7,0	6,6	6,3	6,0
Deutschland	7,8	7,6	7,4	7,2	7,2	7,3
Großbritannien	8,3	8,3	8,2	7,9	7,8	7,8
Frankreich	5,5	5,4	5,3	5,3	5,3	5,2
Schweiz	1,7	1,8	1,8	1,7	1,8	1,8
Kanada	4,5	4,5	4,5	4,5	4,5	4,5
Schweden	1,7	1,7	1,7	1,6	1,6	1,6
Italien	4,2	4,2	4,3	4,3	4,3	4,3
Niederlande	2,5	2,4	2,4	2,4	2,5	2,6
Finnland	0,8	0,9	0,8	0,8	0,8	0,8
Südkorea	2,6	2,7	2,7	2,9	3,1	3,3
China	7,1	8,1	8,6	9,2	10,2	11,0
Brasilien	1,7	1,8	2,2	2,5	2,6	2,6
Indien	2,6	2,7	3,1	3,3	3,3	3,4
Welt	100,0	100,0	100,0	100,0	100,0	100,0

Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Tabelle 4: Anteile ausgewählter Länder in Scopus an allen Publikationen (ohne Geisteswissenschaften, Whole counts)

Land	2005	2006	2007	2008	2009	2010
USA	31,7	30,6	29,6	28,7	27,5	26,6
Japan	9,7	9,3	8,8	8,2	7,3	6,0
Deutschland	8,0	7,7	7,6	7,4	7,1	6,9
Großbritannien	8,4	8,3	8,3	7,9	7,7	7,3
Frankreich	5,5	5,4	5,3	5,3	5,2	4,9
Schweiz	2,1	2,1	2,1	2,0	1,9	1,7
Kanada	4,5	4,4	4,5	4,4	4,3	4,1
Schweden	1,7	1,7	1,6	1,5	1,5	1,4
Italien	4,3	4,3	4,4	4,3	4,2	3,9
Niederlande	2,4	2,4	2,4	2,3	2,3	2,3
Finnland	0,8	0,8	0,8	0,8	0,8	0,7
Südkorea	2,4	2,6	2,7	2,7	2,7	2,8
China	11,0	11,9	12,7	14,0	15,5	16,3
Brasilien	2,1	2,4	2,5	2,6	2,6	2,4
Indien	2,9	3,1	3,2	3,4	3,5	3,7
Welt	100,0	100,0	100,0	100,0	100,0	100,0

Quelle: Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Im Hinblick auf die beobachteten Zitatraten sind die durchschnittlichen Werte in Scopus auf einem ähnlichen Niveau wie in WoS (Tabelle 5 und Tabelle 6), allerdings liegen sie in allen Jahren ein wenig niedriger. Z.B. lag der Weltdurchschnitt der Zittrate in Scopus im Jahr 2005 bei 3,0 im Vergleich zu 3,5 in WoS und in 2008 bei 3,4% im Vergleich zu 3,7%. Dieser Unterschied sollte nicht im Hinblick einer unterschiedlichen Qualität oder eines unterschiedlichen Impact interpretiert werden, sondern als Unterschied in der Abdeckung von Zeitschriften. So kann die breitere Abdeckung der Ingenieurwissenschaften in Scopus eine Erklärung sein, da hier das Zitarniveau generell niedriger liegt, oder die breitere Abdeckung von Zeitschriften in Schwellenländern. Außerdem könnte die Erfassung von stärker spezialisierten Zeitschriften eine Erklärung sein, die in der Regel weniger hohe Zitatraten aufweisen infolge einer weniger breiten Leserschaft (Michels/Schmoch 2011).

Tabelle 5: Beobachtete Zitatzahlen für ausgewählte Länder und Regionen in SCIE und SSCI ohne Eigenzitate

Land	2005	2006	2007	2008
USA	5,4	5,4	5,6	5,6
Japan	3,3	3,4	3,5	3,6
Deutschland	4,6	4,6	4,9	5,1
Großbritannien	4,8	4,9	5,2	5,4
Frankreich	4,1	4,2	4,6	4,6
Schweiz	6,3	6,1	6,3	6,7
Kanada	4,5	4,7	4,9	5,0
Schweden	5,0	5,1	5,3	5,5
Italien	4,2	4,2	4,4	4,5
Niederlande	5,4	5,5	5,9	5,9
Finnland	4,3	4,6	4,9	5,2
Südkorea	2,7	2,8	3,0	3,1
Brasilien	2,4	2,3	2,3	2,2
Indien	2,0	2,2	2,2	2,3
China	2,4	2,5	2,9	3,1
EU15-Länder	4,1	4,1	4,3	4,4
EU12-Länder	2,3	2,5	2,5	2,3
EU27-Länder	3,7	3,8	3,9	3,9
Welt	3,5	3,6	3,7	3,7

Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Tabelle 6: Beobachtete Zitratraten für ausgewählte Länder und Regionen in Scopus (ohne Geisteswissenschaften, ohne Eigenzitate)

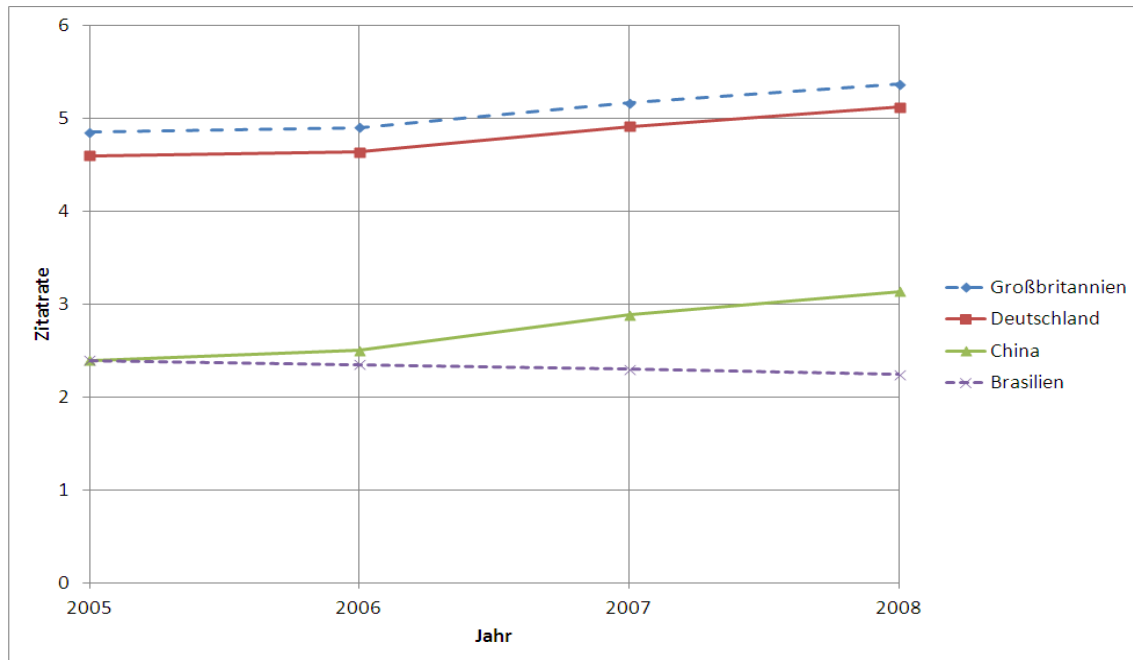
Land	2005	2006	2007	2008
USA	5,2	5,1	5,2	5,5
Japan	3,2	3,1	3,3	3,4
Deutschland	4,4	4,2	4,6	4,7
Großbritannien	4,6	4,4	4,7	5,0
Frankreich	4,1	4,0	4,3	4,4
Schweiz	6,0	5,7	5,9	6,3
Kanada	4,5	4,5	4,8	5,0
Schweden	4,9	4,7	5,1	5,4
Italien	4,2	4,0	4,2	4,7
Niederlande	5,3	5,2	5,6	5,9
Finnland	4,5	4,2	4,6	4,9
Südkorea	3,2	3,2	3,3	3,4
Brasilien	2,4	2,1	2,2	2,3
Indien	1,9	1,9	2,1	2,2
China	1,3	1,5	1,9	2,2
EU15-Länder	4,0	3,9	4,0	4,3
EU12-Länder	2,2	2,2	2,4	2,6
EU27-Länder	3,6	3,5	3,6	3,8
Welt	3,0	3,0	3,1	3,4

Quelle: Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Die Auswirkungen dieser Unterschiede auf einzelne Länder mögen am besten am Beispiel von Deutschland, Großbritannien, Brasilien und China erklärt werden. In diesem Vergleich repräsentieren Deutschland und Großbritannien Industrieländer, das eine mit einem großen eigenen Sprachgebiet, also Deutschland, und Großbritannien als typisches angelsächsisches Land. Der Vergleich ist interessant, da beide Länder ein ähnliches Niveau bei den Publikationszahlen haben. Bei Brasilien und China handelt es sich um zwei Schwellenländer, allerdings aus vollständig unterschiedlichen Gebieten und in Konsequenz mit einem deutlich unterschiedlichen historischen und kulturellen Hintergrund ihrer Innovationssysteme. Im Falle der industrialisierten Länder Deutschland und Großbritannien ist das Niveau der beobachteten Zitate in WoS ein bisschen niedriger

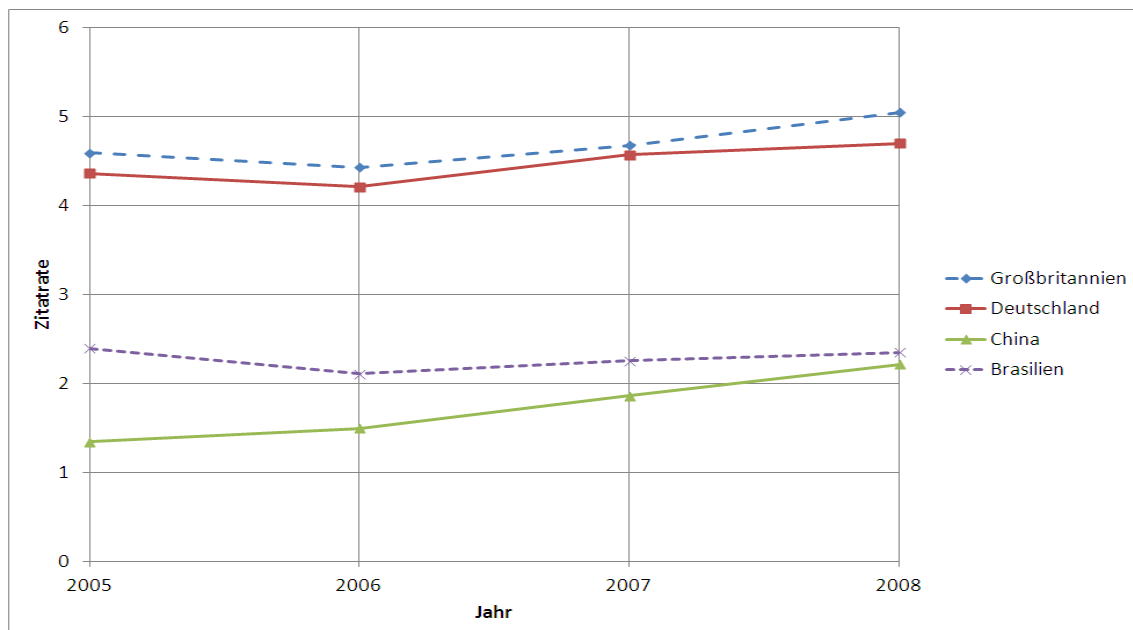
als in Scopus, aber ihre Rangfolge ist identisch und die Trends sind dieselben (Abbildung 6 und Abbildung 7).

Abbildung 6: Beobachtete Zitatraten für ausgewählte Länder in SCIE und SSCI ohne Eigenzitate



Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 7: Beobachtete Zitatrate für ausgewählte Länder in Scopus (ohne Geisteswissenschaften, ohne Eigenzitate)



Quelle: Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Bei Brasilien sind die Zitatraten in Scopus weitgehend identisch zu denen in WoS, aber für China sind die Raten erheblich niedriger, obwohl der Trend eines ansteigenden Niveaus in WoS und Scopus identisch ist. Ohne eine detaillierte Analyse der Gründe für die niedrigen Zitatraten Chinas kann angenommen werden, dass Letztere in Scopus zu einem größeren Ausmaß auf zitierenden Artikeln durch andere chinesische Publikationen in chinesischen Zeitschriften basieren. Im Übrigen erreicht Indien in Scopus ähnliche Zitatraten wie in WoS, Südkorea höhere Raten. Dieses sollte wiederum nicht im Hinblick auf den Impact interpretiert werden, sondern lediglich in Bezug auf die unterschiedliche Zeitschriftenabdeckung. Infolgedessen haben Schlussfolgerungen im Hinblick auf ein "korrektes" Niveau der Zitatraten keine reale methodische Basis. Dasselbe lässt sich für Indien und Südkorea feststellen (Tabelle 7 und Tabelle 8).

Tabelle 7: Index der Zeitschriftenspezifischen Beachtung (ZB) für ausgewählte Länder und Regionen in SCIE und SSCI ohne Eigenzitate

Land	2005	2006	2007	2008
USA	8	7	7	7
Japan	-8	-7	-7	-8
Deutschland	8	7	8	9
Großbritannien	8	9	8	8
Frankreich	2	3	4	3
Schweiz	16	16	14	15
Kanada	5	6	6	8
Schweden	10	11	8	8
Italien	1	1	3	3
Niederlande	9	9	10	9
Finnland	4	10	9	10
Südkorea	-3	-1	-3	-1
Brasilien	-12	-14	-12	-9
Indien	-12	-8	-10	-7
China	3	2	5	7
EU15-Länder	2	2	2	2
EU12-Länder	-9	-7	-3	-9
EU27-Länder	0	0	0	0
Welt	0	0	0	0

Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Tabelle 8: Index der Zeitschriftenspezifischen Beachtung (ZB) für ausgewählte Länder und Regionen in Scopus (ohne Geisteswissenschaften, ohne Eigenzitate)

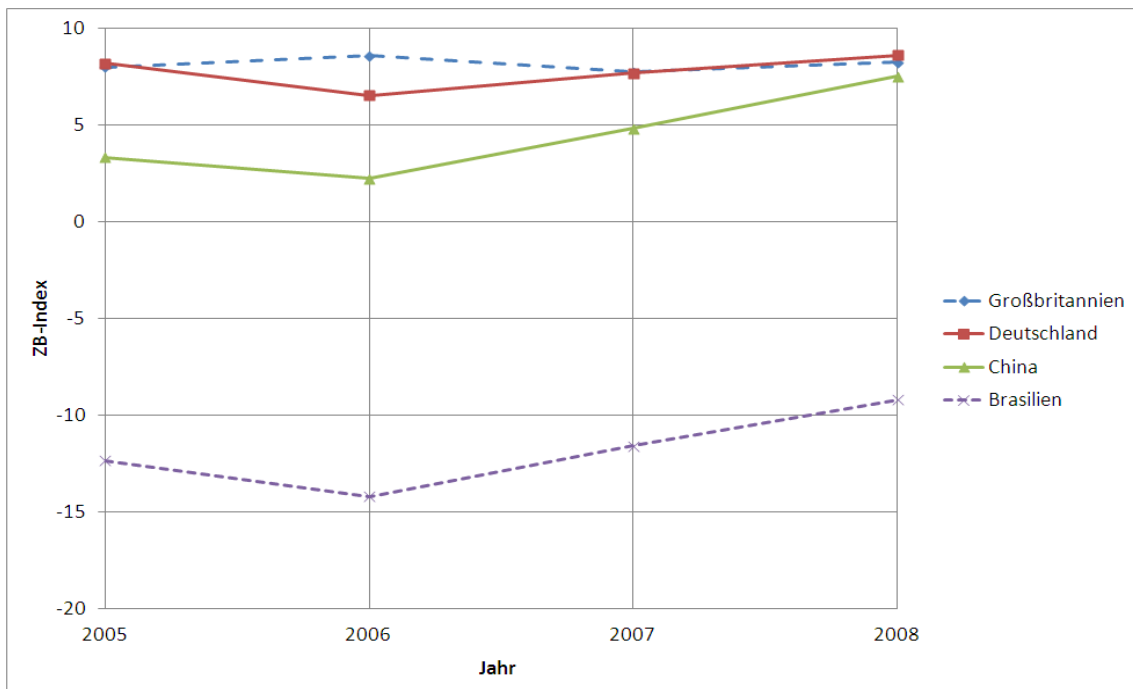
Land	2005	2006	2007	2008
USA	8	8	8	7
Japan	-7	-6	-6	-7
Deutschland	9	8	11	8
Großbritannien	8	9	8	9
Frankreich	5	5	6	5
Schweiz	18	17	16	16
Kanada	7	8	9	9
Schweden	13	10	11	10
Italien	4	2	5	10
Niederlande	10	10	13	12
Finnland	8	6	10	6
Südkorea	-2	-1	-2	-1
Brasilien	-5	-7	-6	-4
Indien	-7	-8	-6	-5
China	5	4	7	10
EU15-Länder	3	3	3	3
EU12-Länder	-4	-4	-1	3
EU27-Länder	2	2	2	1
Welt	0	0	0	0

Quelle: Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Natürlich haben die Unterschiede zwischen Scopus und WoS in Bezug auf die durchschnittlichen Zitatraten und die spezifischen Zitatraten für einzelne Länder auch einen Effekt auf abgeleitete Indikatoren wie die wissenschaftliche Beachtung oder die internationale Ausrichtung². Im Falle der ZB-Indizes, also dem Index für die beobachteten in Relation zu den zeitschriftenspezifischen erwarteten Zitatraten, sind Niveau und Trend für die industrialisierten Länder Deutschland und Großbritannien in Scopus und WoS ähnlich. Für die Schwellenländer sind die Niveaus in Scopus höher, die Trends in beiden Datenbanken jedoch ähnlich (Abbildung 8 und Abbildung 9). Dieselbe Beobachtung gilt auch für Indien und Südkorea (Tabelle 7 und Tabelle 8).

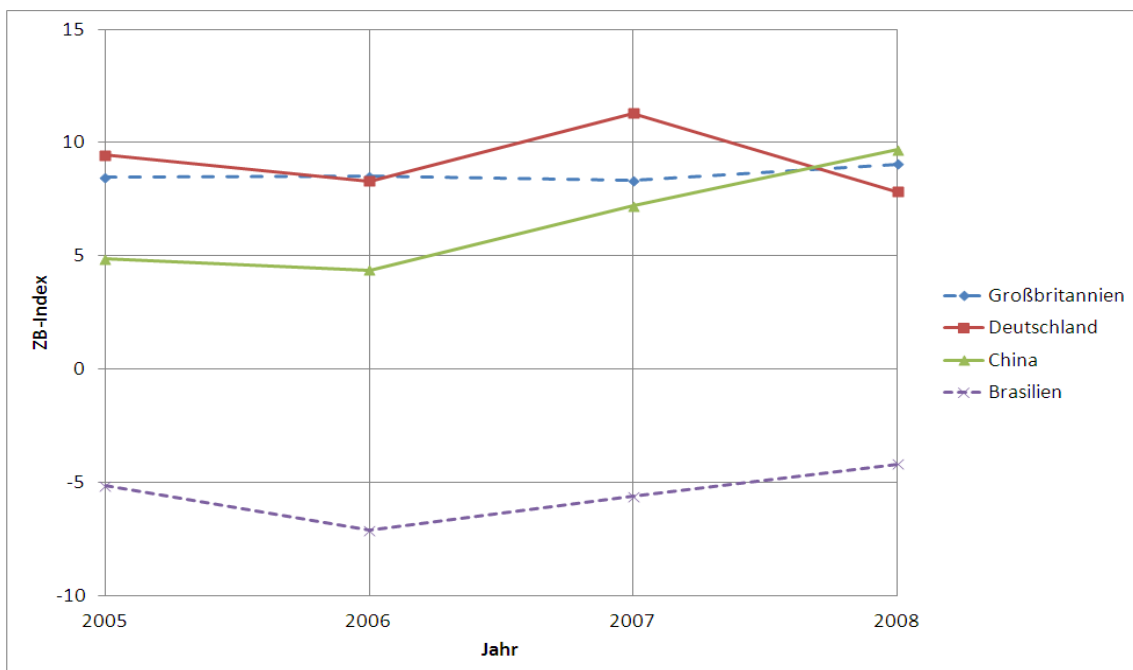
² Die Indikatoren wissenschaftliche Beachtung und IA sind in Abschnitt 4.4 genauer erklärt.

Abbildung 8: Index der Zeitschriftenspezifischen Beachtung (ZB) für vier ausgewählte Länder Regionen in SCIE und SSCI ohne Eigenzitate



Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 9: Index der Zeitschriftenspezifischen Beachtung (ZB) für vier ausgewählte Länder in Scopus (ohne Geisteswissenschaften, ohne Eigenzitate)



Quelle: Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Der IA-Index sagt etwas über die Sichtbarkeit der Zeitschriften aus, in denen die jeweiligen Länder publizieren. In Bezug auf die IA-Indizes sind die Niveaus von Deutschland und Großbritannien in Scopus erheblich höher als in WoS (Abbildung 10, Abbildung 11, Tabelle 9 und Tabelle 10). Dieses war zu erwarten, da die weltdurchschnittlichen Zitaten in Scopus niedriger als in WoS sind, und mit Bezug auf dieses Durchschnittsniveau sind die Zitaten von Deutschland und Großbritannien relativ gesehen höher. Die Rangfolge zwischen Deutschland und Großbritannien ist dieselbe in Scopus wie in WoS. Diese Feststellung trifft auch auf China und Brasilien zu. Die Trends von WoS werden in Scopus reproduziert, aber die IA-Indizes Chinas sind in Scopus erheblich niedriger als in WoS. Dieser Effekt ist mit der breiteren Abdeckung chinesischer Journals in Scopus verbunden, sodass der höhere Anteil chinesischer Publikationen in Zeitschriften mit niedrigerem Impact-Faktor erscheint.

Tabelle 9: Index der Internationalen Ausrichtung (IA) für ausgewählte Länder und Regionen in SCIE und SSCI ohne Eigenzitate

Land	2005	2006	2007	2008
USA	33	33	34	34
Japan	0	0	2	5
Deutschland	18	19	20	23
Großbritannien	23	22	25	28
Frankreich	14	13	17	18
Schweiz	38	36	38	41
Kanada	19	20	22	22
Schweden	23	24	27	30
Italien	15	15	15	17
Niederlande	32	33	35	37
Finnland	16	14	18	23
Südkorea	-25	-24	-16	-16
Brasilien	-26	-28	-35	-39
Indien	-44	-39	-40	-40
China	-40	-37	-29	-24
EU15-Länder	12	12	13	15
EU12-Länder	-31	-29	-33	-36
EU27-Länder	4	4	4	4
Welt	0	0	0	0

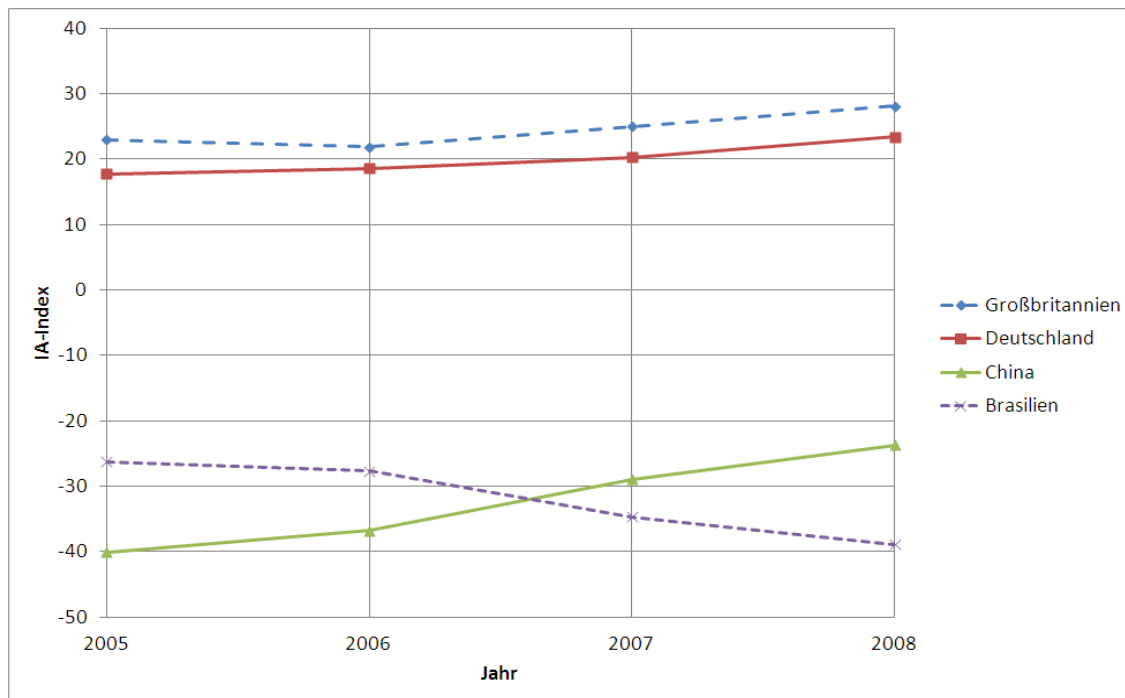
Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Tabelle 10: Index der Internationalen Ausrichtung (IA) für ausgewählte Länder und Regionen in Scopus (ohne Geisteswissenschaften, ohne Eigenzitate)

Land	2005	2006	2007	2008
USA	43	41	41	40
Japan	12	9	10	9
Deutschland	26	24	25	25
Großbritannien	32	29	30	31
Frankreich	26	23	25	23
Schweiz	45	43	44	45
Kanada	30	31	32	30
Schweden	34	32	36	36
Italien	27	25	24	23
Niederlande	42	42	42	42
Finnland	30	27	28	31
Südkorea	8	6	6	3
Brasilien	-19	-28	-27	-31
Indien	-40	-35	-32	-34
China	-70	-63	-54	-47
EU15-Länder	23	21	22	21
EU12-Länder	-27	-26	-25	-29
EU27-Länder	15	12	13	11
Welt	0	0	0	0

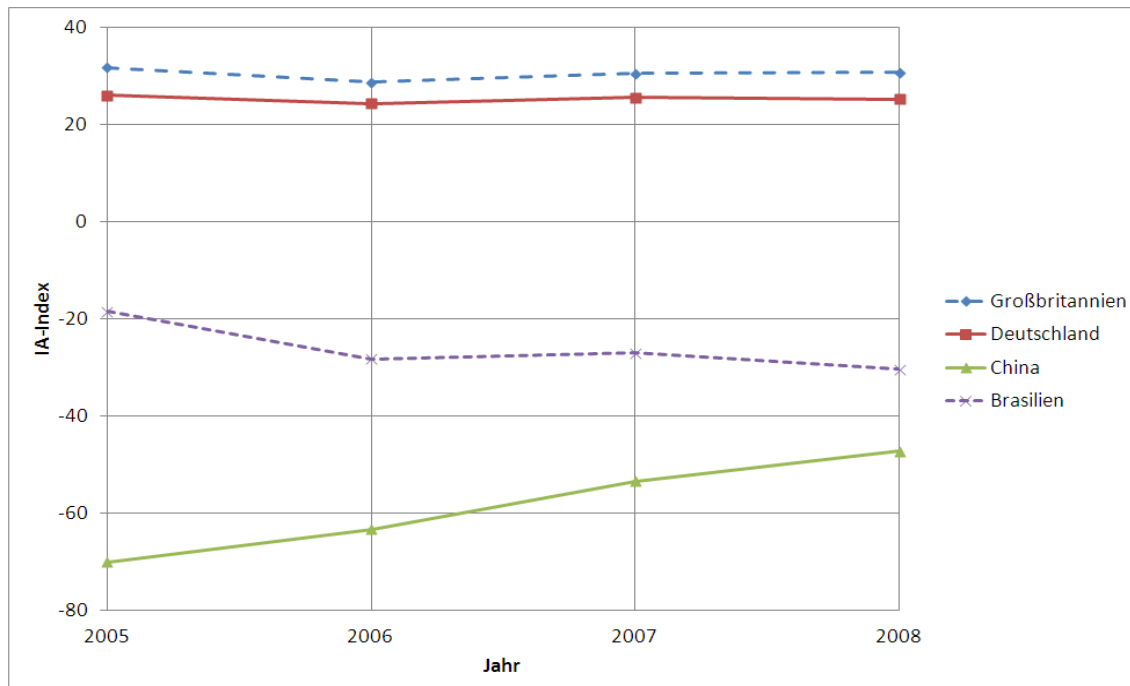
Quelle: Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 10: Index der Internationalen Ausrichtung (IA) für vier ausgewählte Länder in SCIE und SSCI ohne Eigenzitate



Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 11: Index der Internationalen Ausrichtung (IA) für vier ausgewählte Länder in Scopus (ohne Geisteswissenschaften, ohne Eigenzitate)



Quelle: Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Zusammenfassend sind die Niveaus und Trends der Zitatindikatoren in Scopus zu einem großen Teil auch in WoS zu finden. Allerdings ist das Weltdurchschnittsniveau in Scopus etwas niedriger als in WoS, die Rangfolge zwischen den industrialisierten Ländern wird reproduziert, das Niveau der Schwellenländer unterscheidet sich manchmal im Hinblick auf die spezifische Datenbankabdeckung des analysierten Landes.

Insgesamt unterscheidet sich die Abdeckung von Scopus und WoS in erheblichem Maße. Insbesondere deckt Scopus einen großen Anteil von Artikeln exklusiv ab. Die Zahl der Artikel und Proceedings in Scopus ist deutlich höher als die in WoS, wobei die Hauptunterschiede in einer breiteren Abdeckung der Ingenieurwissenschaften in Bezug auf Zeitschriftenartikel und vor allem Proceedings liegen. Weiterhin ist China in Scopus deutlich breiter abgedeckt. Die übrigen Entwicklungs- und Schwellenländer werden in Scopus ebenfalls umfangreicher als in WoS abgedeckt. Diese Unterschiede führen zu unterschiedlichen Zitatraten und daraus abgeleiteten Indikatoren, jedoch sind die Rangfolge der Länder und die wesentlichen Trends in Scopus und WoS weitgehend ähnlich. Dennoch sind einige Abweichungen in Scopus und WoS zu beobachten, insbesondere für Schwellenländer, aber jede Feststellung hinsichtlich eines "korrekten" Wertes ist nicht möglich, denn es ist jeweils eine genauere Betrachtung der jeweiligen Datenbankabdeckung erforderlich. In jedem Fall erweist sich Scopus als wertvolle Ergänzung zu WoS, vor allem sowohl bei Analysen der Ingenieurwissenschaften als auch bei Schwellenländern.

Zur Vervollständigung des Datenbankvergleichs ist darauf hinzuweisen, dass der wesentliche Vorteil von WoS gegenüber Scopus darin besteht, dass sehr viel früher (ab 1980) bereits aussagefähige Zitanalysen möglich sind, während dies in Scopus erst ab dem Jahr 2004 der Fall ist. Demgegenüber wurde in Scopus von Anfang an eine unmittelbare Verbindung der einzelnen Autoren und ihrer Institutionen eingeführt, und es wurden die vollen Vornamen, nicht nur die Initialen, erfasst, was eine Reihe von Analysen erlaubt, die mit WoS nicht möglich sind. Hier hat WoS in jüngster Zeit nachgebessert, wahrscheinlich im Wettbewerb zu Scopus, sodass in Zukunft ähnliche Analysen wie beispielsweise in den Projekten "Universitätspatente" oder "Wissenschaftlermobilität" möglich sein werden.

3 Wachstumsberechnungen

Der Prozess der Globalisierung wird von einem starken Trend hin zur Wissenschaftsgesellschaft begleitet (Stehr 1994), wobei die wachsende Verwissenschaftlichung ein zentraler Aspekt ist. Letztere deckt verschiedene Aspekte ab, wie die Relevanz von wissenschaftlicher Expertise in politischen Entscheidungen oder die Transformation von früher expertenbasierten Berufen hin zu wissensbasierten. Ein weiteres Schlüsselement der Verwissenschaftlichung ist das beschleunigte Wachstum wissenschaftlicher Forschung, vor allen Dingen Forschung in öffentlichen Einrichtungen und Universitäten. Dennoch hat es sich als schwierig erwiesen, das Wachstum der wissenschaftlichen Forschung zu messen und eine objektive Basis für diese "gefühlten" Veränderungen bereit zu stellen. Der nützlichste Indikator wäre die Bereitstellung der Zahl des Personals an öffentlichen Forschungseinrichtungen und Universitäten, doch sind die verfügbaren Statistiken recht lückenhaft, zumindest auf dem internationalen Niveau. Derartige Daten sind für einige Länder zumindest auf aggregiertem Niveau verfügbar, aber die Differenzierung nach wissenschaftlichen Feldern ist noch zu grob.

Vor diesem Hintergrund ist es naheliegend, breite Publikationsdatenbanken wie das WoS (WoS) für die Analyse des wissenschaftlichen Wachstums zu benutzen. De Solla Price (1971) hat bereits die Zahl der wissenschaftlichen Zeitschriften als Beleg für das schnelle Wachstum der Wissenschaft seit dem Beginn des 17. Jahrhunderts herangezogen. Allerdings ist die Zahl der Zeitschriften, die von WoS abgedeckt werden, in ständigem Wachstum begriffen, und es ist unklar, ob diese Beobachtung mit einem Wachstum der Wissenschaft zusammenhängt oder schlicht mit einer zusätzlichen Abdeckung bereits existierender Zeitschriften durch den Datenbankproduzenten. Insbesondere die Zahl der in WoS abgedeckten Zeitschriften wächst in den letzten Jahren mit großer Geschwindigkeit an – vermutlich als eine Konsequenz des Wettbewerbs mit der neueren bibliometrischen Datenbank Scopus. Als eine Reaktion auf Scopus hat der Provider von WoS, früher das Institute of Scientific Information (ISI), nun Thomson Reuters, verschiedene Merkmale verbessert. Insbesondere ist in einem großen Teil der Records der volle Vorname der Autoren anstelle der reinen Initialen aufgeführt und es wurden vor allen Dingen mehr Journals abgedeckt. Deshalb ist wahrscheinlich, dass die wachsende Zahl von Publikationen in WoS vor allen Dingen die veränderte Abdeckung dieser Datenbank reflektiert und nicht das Wachstum der Wissenschaft. Dennoch haben verschiedene Autoren die wachsende Zahl von Publikationen als einen Indikator für das Anwachsen wissenschaftlicher Aktivitäten in spezifischen Feldern herangezogen (siehe etwa Gupta/Dhawan 2008 als eines von vielen Beispielen). Um diesem Problem zu begegnen, haben Leydesdorff et al. (1994) vorgeschlagen, feste Journal-Sets zu nutzen, aber auch neue Zeitschriften mit hinzuzuziehen, um so die Dynamik in neuen wissenschaftlichen Feldern erfassen zu können. Das Ziel der vorliegenden Überlegungen ist es, zu analysieren, ob dieser Ansatz, der bereits zu Beginn

der 1990er Jahre konzipiert wurde, in einer Situation mit relativ langsamem Wachstum in WoS, in gegenwärtiger Situation immer noch angemessen ist, die durch eine markante Änderung der Datenbankabdeckung gekennzeichnet ist.

Um das wissenschaftliche Wachstum in einer Datenbank mit einer zunehmenden Zahl von Zeitschriften analysieren zu können, entschieden wir uns für die Betrachtung der Veränderungen in WoS während der Periode zwischen 2000 und 2008 und unterschieden dabei verschiedene Typen von Zeitschriften:

1. Zeitschriften, die durch WoS im Jahr 2000 erfasst waren, werden als "Basis-Zeitschriften" bezeichnet.
2. Erst kürzlich, in der Zeit von 2001 bis 2008, in WoS eingeführte Zeitschriften. Dieser Typ wird durch eine niedrige Volume-Nummer unterhalb oder gleich 10 Jahren operationalisiert und wird als "neue Zeitschriften" bezeichnet.
3. Zusätzliche, bereits existierende Zeitschriften, die von WoS in der Zeit von 2001 bis 2008 eingeführt wurden. Dieser Typ wurde durch höhere Volume-Nummern von mehr als 11 operationalisiert und "zusätzliche Zeitschriften" genannt.
4. Zeitschriften, die bereits vor dem Jahr 2000 während eines oder mehrere Jahre in der Datenbank enthalten waren, welche dann aber für mehrere Jahre nicht in WoS enthalten waren und dann, nach 2002, wieder aufgeführt wurden. Sie werden "wiederkehrende Zeitschriften" genannt.

Die zugrunde liegende Annahme ist, dass die Datenbank einen Basissatz von Zeitschriften abdeckt, die eine Art Stamm-Set der Datenbank repräsentieren, welcher allerdings seit 2002 ständig ergänzt wurde. In gewissem Ausmaß repräsentieren die hinzugefügten Zeitschriften neue Zeitschriften, die die Notwendigkeit dokumentieren, neue Publikationsmedien zusätzlich zu den bereits existierenden zu installieren, z.B. neue spezifische Zeitschriften zur Nanotechnologie, welche die existierenden in der Physik und der Chemie ergänzen. Diese Typen neuer Zeitschriften reflektieren die Dynamik der Wissenschaft in aufkommenden Feldern, was bereits durch Leydesdorff et al. (1994) vorgeschlagen wurde.

Die Untersuchung hat neben der Unterscheidung von Wissenschaftswachstum und "künstlichem" Wachstum zu einer Reihe von interessanten Teilergebnissen geführt. Dieses trifft insbesondere für das unterschiedliche Wachstum in Teilgebieten zu und den Zusammenhang zwischen Zeitschriftengröße und Zitatraten.

Die Ergebnisse dieser Teiluntersuchung wurden auf der ISSI-Konferenz im Juli 2011 in Durban (Südafrika) als Beitrag angenommen und dort vorgestellt. Weiterhin wurde eine verbesserte Version der dortigen Präsentation bei der Zeitschrift *Scientometrics* eingereicht, die zur Veröffentlichung akzeptiert wurde.

4 Methodische Überlegungen zu Zitatraten

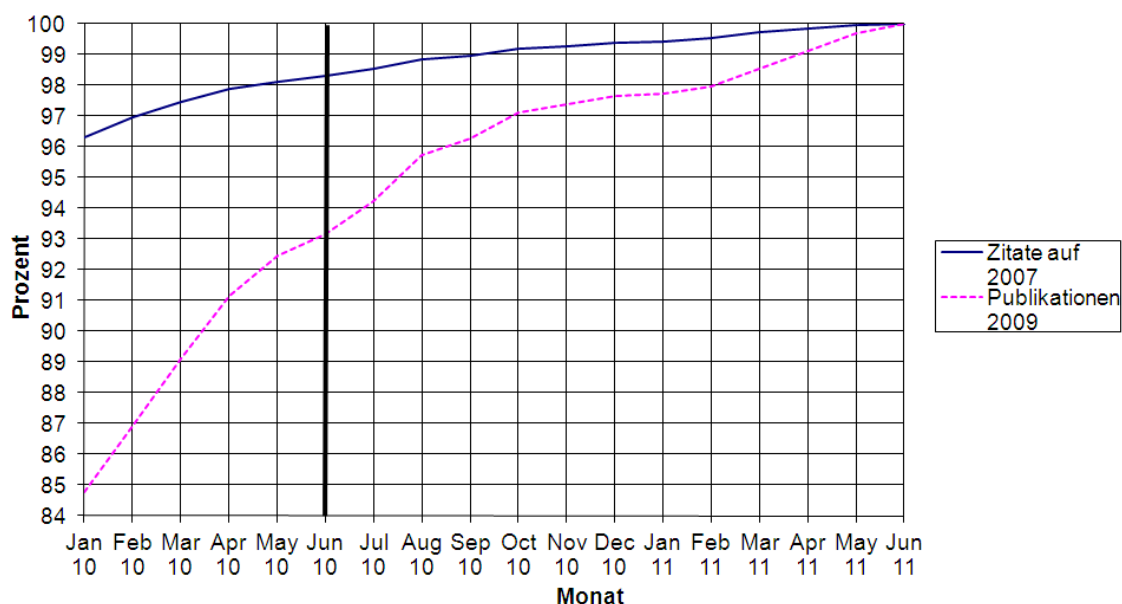
4.1 Zeitliche Erfassung von Publikationen und Zitaten

Für viele Untersuchungen wird neben der absoluten Zahl der Publikationen im Zeitverlauf auch die der Zitate betrachtet. Hier entsteht allerdings das Problem, dass es mehrere Jahre braucht, bis auf einen Artikel reagiert und dieser von anderen zitiert wird. Hier müsste zu einer einigermaßen vollständigen Erfassung der Zitate ein Zeitraum von etwa zehn Jahren abgewartet werden. Hier wäre aber dann im Zusammenhang mit der Analyse aktueller Themen eine inhaltlich aktuelle Erfassung nicht mehr möglich. Von daher ist es üblich, mit Zeitfenstern zu arbeiten, wobei Dreijahreszeitfenster für möglichst aktuelle und 5-Jahresfenster für genauere Analysen üblich sind. Es besteht also stets ein Widerspruch zwischen Aktualität und Genauigkeit, wobei insbesondere für größere Untersuchungseinheiten wie Länder sich ein Zeitraum von drei Jahren in der Regel als hinreichend genau erwiesen hat. Von daher wurden etwa für die Gutachten für den Pakt für Forschung und Innovation oder die Expertenkommission Forschung und Innovation für Ländervergleiche 3-Jahresfenster verwendet, wobei neben dem untersuchten Jahr selber die beiden Folgejahre einbezogen wurden. Wichtig ist dabei, dass für den gesamten Untersuchungszeitraum für alle Jahre einheitlich 3-Jahresfenster benutzt werden, um zu konsistenten Zeitreihen zu gelangen.

Ein Problem ist dabei aber nach wie vor, welche Daten am aktuellen Rand zugrunde gelegt werden müssen und welches Zitatjahr dabei untersucht werden kann. Im Falle der Untersuchungen für den Pakt für Forschung und Innovation und die Expertenkommission Forschung und Innovation standen bei WoS Daten bis zum Juni 2011 zur Verfügung. Beabsichtigt war, zumindest das Jahr 2008 in Bezug auf die Zitate zu untersuchen. Für das Jahr 2008 müssten Zitate aus den Jahren 2008, 2009 und 2010 zur Verfügung stehen. Nun besteht aber das Problem, dass die Daten für 2010 nicht bereits im Dezember 2010 vollständig zur Verfügung stehen, sondern erst mit einem Zeitverzug in die Datenbank eingeführt werden. Von daher wurde untersucht, in welchem Monat des Jahres 2011 genügend Daten zur Verfügung stehen, um das Jahr 2008 im Hinblick auf Zitate untersuchen zu können. Um die Vollständigkeit der Daten im Jahr 2011 prüfen zu können, wurde die Situation für Publikationen für das Publikationsjahr 2009 retrospektiv untersucht. Mit einem Update bis zum Juni 2010 – für die Studie für das Jahr 2010 waren Daten bis Juni 2011 verfügbar – zeigen die Ergebnisse, dass 98% aller Zitate für das Jahr 2007 im Juni 2010 verfügbar waren und 93% aller Publikationen des Jahres 2009. Die verspätete Abdeckung von Publikationen für das Jahr 2009 (manchmal sogar erst im April 2011) ist auf die Tatsache zurückzuführen, dass einige Zeitschriften ihre Daten relativ spät dem Datenbankproduzenten Thomson Reuters zur

Verfügung stellten. Die Zitate für 2007 waren also im Juni 2010 weitgehend vollständig. Die Stichprobe der Publikationen für das Jahr 2009 war sehr groß, sodass statistisch relevante Schlussfolgerungen für 2009 möglich waren. Es kann angenommen werden, dass im Juni 2011 die Situation für Publikationen des Jahres 2010 bzw. für Zitate in Bezug auf das Jahr 2008 ähnlich war. Auch schon Ende April 2010 waren die Zitate für 2007 und die Publikationen für 2009 mit 98% bzw. 91% so breit vertreten, dass damit aussagefähigen Analysen realisiert werden konnten. Das entsprechende Untersuchungsergebnis ist in Abbildung 12 noch einmal dokumentiert.

Abbildung 12: Vollständigkeit der Publikationen des Jahres 2009 sowie der Zitate auf das Jahr 2007 mit monatlichen Updates von Januar 2010 bis Juni 2011 in den Datenbanken SSCI und SCIE



Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

4.2 Whole count und fraktionierte Zählweise

Bei der bibliometrischen Analyse von Ländern ist es eine wichtige methodische Frage, ob die Publikationen fraktioniert gezählt werden sollen oder als sogenannte Whole counts. Denn wissenschaftliche Publikationen werden zunehmend von mehreren Autoren aus unterschiedlichen Ländern erarbeitet. Z.B. wurden im Jahr 2006 44% aller Publikationen mit deutschen Autoren in Kooperationen mit Partnern aus dem Ausland geschrieben (Hinze et al. 2008). Für die Whole count-Analyse erhält jedes Land, das in der Publikation erscheint, den Wert 1, die Beteiligung mehrerer Autoren eines Landes

ist dabei nicht in Rechnung gestellt. Bei einer fraktionierten Zählung ist der Anteil eines Landes an einer Publikation abhängig von der Zahl der beteiligten Länder. Wenn z.B. eine Publikation durch einen deutschen und einen französischen Autor oder Autorin erarbeitet wird, wird bei der fraktionierten Zählweise jedem der beiden Länder 50% der Publikation zugeschrieben. In einer mathematisch, statistischen Perspektive ist die Benutzung der fraktionierten Zählweise naheliegend, da dann die Summe der Länderanteile mit der Zahl der Gesamtpublikationen identisch ist. Summiert man die Länderanteile der Whole count-Zählweise, so liegt die Summe der Anteile über der Anzahl der weltweiten Gesamtpublikationen. Deshalb ist die Berechnung abgeleiteter Indikatoren auf Basis fraktioniert gezählter Werte etwas einfacher als bei der Whole count-Zählweise.

Es zeigt sich allerdings bei einem Vergleich der Ergebnisse für die Whole count und die fraktionierte Zählweise, dass die Anteile der Länder an den Gesamtpublikationen in WoS sich erheblich unterscheiden, wie das Beispiel für das Jahr 2010 in Tabelle 11 verdeutlicht. So ist der Anteil der Vereinigten Staaten an allen Publikationen in der Datenbank von 28,0 auf 23,9% reduziert. Der deutsche Anteil von 7,3 auf 5,4%, der Großbritanniens von 7,8 auf 4,9%. Das kann aus mathematischer Sicht erwartet werden, da bei der Whole count-Zählweise alle Publikationen in der Datenbank als Referenz herangezogen werden.

Allerdings sind die betrachteten Länder in unterschiedlicher Weise betroffen: So wird der Anteil Japans um 14% reduziert, der deutsche Wert um 34% und der der Schweiz sogar um 112%, jedoch der Chinas lediglich um 11%. Generell sind die Länder von einer starken Reduzierung betroffen, wenn es um den Vergleich der Whole count- und der fraktionierten Anteile geht, die einen hohen Anteil von Ko-Publikationen mit Autoren aus dem Ausland haben. Im Gegensatz dazu erzielen eher isoliertere Länder nach der fraktionierten Zählung ein höheres Gewicht. Der Unterschied ist in Tabelle 11 in der Berechnung der Relation der Anteile nach der Whole count- und der fraktionierten Zählweise in Spalte 4 dokumentiert. Beispielsweise ist dieses Verhältnis bei China 1,11, bei Deutschland 1,34 und bei der Schweiz 2,12. Das hat einen erheblichen Effekt auf die Verhältnisse zwischen den Ländern, die in den Spalten 5 und 6 von Tabelle 11 dokumentiert sind. Wenn China als Referenzeinheit herangezogen wird, erhält Deutschland den Wert 0,66 nach der Whole count-Analyse und 0,55 nach der fraktionierten. Für die Schweiz ist die Relation 0,17 für die Whole count-Zählweise und 0,09 für die fraktionierte. Damit ist die Entscheidung für entweder die Whole count- oder die fraktionierte Zählweise von großer Bedeutung.

Für eine adäquate Bewertung der Resultate muss gefragt werden, ob dieser Effekt der fraktionierten Zählung beabsichtigt ist. Denn kann man annehmen, dass Publikationen

in Kooperation mit ausländischen Autoren ein höheres Gewicht haben, als rein nationale Publikationen. Für die Ausarbeitung einer internationalen Publikation sind erhebliche Anstrengungen erforderlich, um eine gute Koordination mit den ausländischen Partnern zu erreichen (Moed 2005: 278). Die fraktionierte Zählweise von Publikationen führt dagegen zu einer Abwertung von in internationalen Kooperationen entstandenen Veröffentlichungen. Um internationale Kooperationen zu honorieren und die implizite Abwertung zu vermeiden, erscheint es angemessener, die Whole count-Zählweise zu verwenden.

Tabelle 11: Anteile ausgewählter Länder innerhalb aller Publikationen in SCIE und SSCI des Jahres 2010 nach der fraktionierten und der Whole count Zählung und daraus folgende Länderrelationen

	Whole count	Fraktioniert	Whole count/ fraktioniert	Whole count	Fraktioniert
USA	28,0	23,9	1,17	2,54	2,41
Japan	6,0	5,3	1,14	0,55	0,53
Deutschland	7,3	5,4	1,34	0,66	0,55
Großbritannien	7,8	4,9	1,59	0,71	0,5
Frankreich	5,2	3,9	1,35	0,48	0,39
Schweiz	1,8	0,9	2,12	0,17	0,09
Kanada	4,5	3,4	1,34	0,41	0,34
Schweden	1,6	1,1	1,47	0,15	0,11
Italien	4,3	3,3	1,28	0,39	0,34
Niederlande	2,6	1,8	1,41	0,24	0,19
Finnland	0,8	0,6	1,41	0,07	0,06
Südkorea	3,3	2,9	1,15	0,3	0,29
Brasilien	2,6	2,3	1,14	0,23	0,23
Indien	3,4	3,1	1,12	0,31	0,31
China	11	9,9	1,11	1	1
Andere Länder	36,1	27,3	1,32	3,28	2,75
Gesamt	126,3	100			

Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

4.3 **Feldspezifische und zeitschriftenspezifische erwartete Zitatraten**

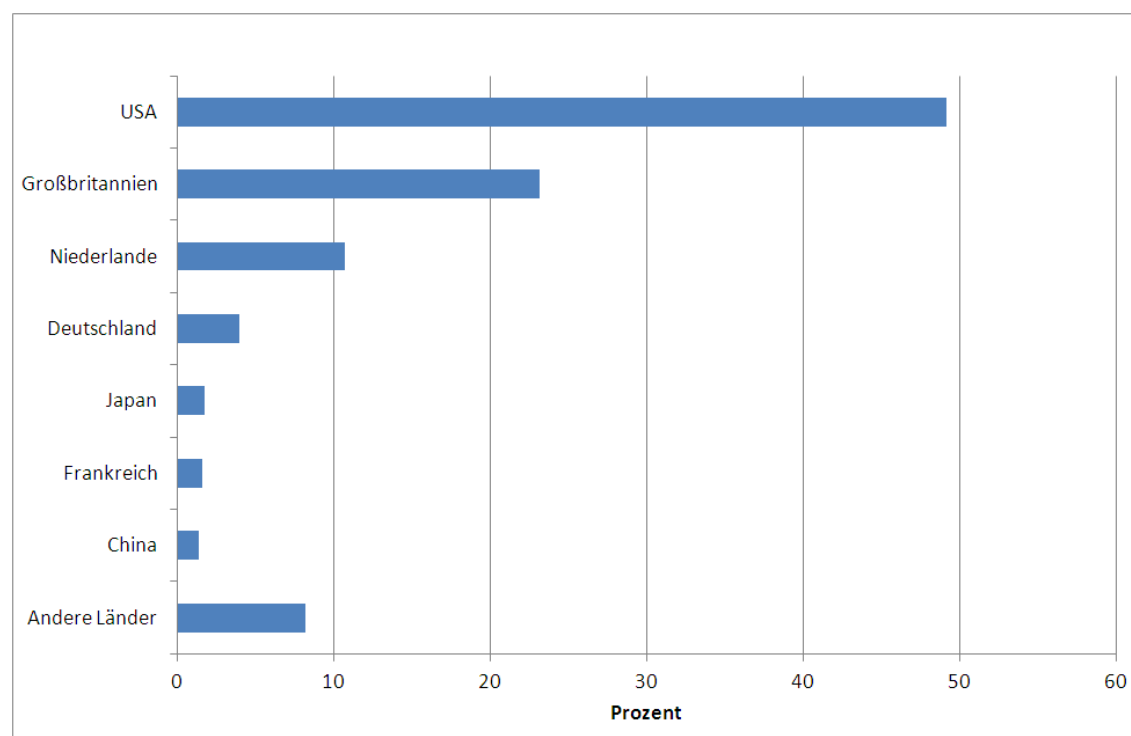
Bei der Untersuchung der Zitatraten für verschiedene Wissenschaftsfelder fallen enorme Unterschiede im Publikations- und Zitatverhalten auf. So ist etwa beispielsweise die Zahl der Zitate in der Biotechnologie erheblich höher als die in der Mathematik. Deshalb dürfen Zitatraten unterschiedlicher Wissenschaftsgebiete nicht unmittelbar miteinander verglichen werden. Vor diesem Hintergrund hat die niederländische Forschungsgruppe CWTS (Center for Science and Technology Studies, Leiden) vorgeschlagen, die einfachen Zitatraten, also das Verhältnis zwischen den Publikationen und den zugehörigen Zitaten, häufig auch beobachtete Zitatraten genannt, mit sogenannten erwarteten Zitatraten zu standardisieren. Dabei kann zwischen feld- oder disziplinspezifischen und zeitschriftenspezifischen erwarteten Zitatraten unterschieden werden. Die feldspezifische Zitatrate wird bestimmt, indem die Durchschnittszitatrate für alle Publikationen eines Feldes bestimmt wird, wobei das Feld durch alle dazugehörigen Zeitschriften definiert wird (Van Raan 2004). Diese feld- oder disziplinspezifische Standardisierung, die manchmal auch "Crown Indicator" genannt wird, ist ohne Zweifel eine erhebliche Verbesserung im Vergleich zur Betrachtung der einfachen beobachteten Zitatraten. Allerdings stellt sie nicht in Rechnung, dass das WoS vor allen Dingen amerikanische Zeitschriften abdeckt, die vergleichsweise hohe Zitatraten erzielen, weil diese Publikationen eine breite Leserschaft erreichen und damit eine gute Sichtbarkeit amerikanischer Journals erzielt wird. Von daher werden die Durchschnittswerte der Disziplinen von Amerikanern dominiert; denn fast die Hälfte der in WoS abgedeckten Zeitschriften wird in den USA herausgegeben (Abbildung 13). Es muss deshalb gefragt werden, ob eine solche Referenz für ein Land wie Deutschland geeignet ist, das einen großen eigenen Sprachraum hat und wo deshalb viele Publikationen in deutscher Sprache veröffentlicht werden.

Ein weiteres Problem besteht darin, dass, wie bereits beschrieben, die Datenbankabdeckung von WoS durch den Hersteller Thomson Reuters erheblich ausgedehnt worden ist und dass sich damit die Zahl kleinerer Journals deutlich erhöht hat. Kleine, spezialisierte Zeitschriften haben eine kleinere Leserschaft und erzielen deshalb niedrigere Zitatraten als die großen Mainstream-Journals (Michels/Schmoch 2011). In der Folge bedeutet dies, dass die feldspezifischen Durchschnittsraten in den letzten Jahren aus einer Mischung von kleinen und großen Zeitschriften bestehen, die deutlich unterschiedliche Zitatraten haben. Es ist deshalb fraglich, ob die Felddurchschnitte als sinnvolle Größe betrachtet werden können. In jedem Fall sind in den Felddurchschnittswerten Effekte der amerikanischen Dominanz und der heterogenen Zeitschriftengrößen in einer unklaren Weise miteinander vermischt.

Vor diesem Hintergrund nutzt das Fraunhofer ISI seit mehreren Jahren bereits sogenannte zeitschriftenspezifische erwartete Zitatraten zur Standardisierung der beobachteten Zitatraten (Grupp et al. 2001). Bei den zeitschriftenspezifischen erwarteten Zitatraten wird jede Publikation mit ihren Zitatraten nur in Referenz zu Zitatraten gesetzt, die dem Durchschnitt der Zeitschrift entsprechen, in der sie veröffentlicht wurde. Es wird damit gemessen, ob die Publikation im Vergleich zum Durchschnitt der jeweiligen Zeitschrift eine über- oder unterdurchschnittliche wissenschaftliche Beachtung erfahren hat. Da außerdem die Zeitschriften immer einem spezifischen Fachgebiet zugeordnet sind, ist implizit auch die feldspezifische Beachtung in den zeitschriftenspezifischen erwarteten Zitatraten enthalten. Auf diese Weise können Publikationen aus Ländern mit großem eigenen Sprachraum oder Publikationen in spezialisierten Fachzeitschriften adäquater bewertet werden. In Kombination mit dem abgeleiteten Indikator IA ergibt sich daraus eine differenziertere Betrachtung, als dies bei der feldspezifischen erwarteten Zitatrate als Referenz möglich ist. Die IA wird als abgeleiteter Indikator im folgenden Abschnitt genauer behandelt.

Aus Sicht der Bestimmung der feldspezifischen erwarteten Zitate ist zu beachten, dass viele Zeitschriften und damit Artikel mehrfach klassifiziert sind. Deshalb müssen die entsprechenden Zitate fraktioniert ermittelt werden.

Abbildung 13: Anteil von Publikationen in WoS nach Herausgeberland, 2009



Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

4.4 Zeitschriftenspezifische Beachtung (ZB) und Internationale Ausrichtung (IA)

Der Indikator ZB, vorgeschlagen von Grupp et al. (2001), beschreibt letztlich die Relation von beobachteten zu erwarteten zeitschriftenspezifischen Zitatraten. Er misst letztlich, ob die Publikationen eines Landes/einer Region oberhalb oder unterhalb des Durchschnitts zitiert werden, im Vergleich zu anderen Artikeln, die in den selben Zeitschriften publiziert werden wie die des Landes/der Region. Auf diese Weise wird in Rechnung gestellt, dass nicht-amerikanische und kleinere Zeitschriften weniger häufiger zitiert werden, sodass in diesem Fall niedrigere Referenzwerte angenommen werden. Umgekehrt ist der Referenzwert für Artikel in hochzitierten Zeitschriften ebenfalls relativ hoch. In diesem Kontext zeigen positive Indikatorenwerte, dass die Zitatraten oberhalb des weltweiten Durchschnitts liegen. Wenn die beobachtete Zitatquote eines Landes in Relation zur erwarteten gesetzt wird, führt dieses bei Gleichheit oder bei der Entsprechung des Zitatwertes mit dem Durchschnitt zu einem Neutralwert von 1. Die beschriebene einfache Relation hat den Nachteil, dass die Verteilung sehr schief ist und dass der Wertebereich zwischen 0 und $+\infty$ liegt. Es gibt Verfahren der Transformation, wo diese Indizes einen Neutralwert von 0 haben und der gesamte Wertebereich zwischen -100 und +100 liegt. Der so abgeleitete Indikator erweist sich als illustrativer für grafische Darstellungen und Interpretationen. Dabei zeigen positive Indexwerte eine Zitierung oberhalb des Weltdurchschnitts der erwarteten Zitatraten, negative Werte ein Niveau unterhalb der erwarteten Zitatrate. Der Indikator wird in folgender Weise berechnet:

$$ZB_k = 100 \tanh \ln (OBS_k/EXP_k)$$

In dieser Formel bezieht sich OBS_k auf die aktuell beobachtete Zitathäufigkeit von Publikationen eines Landes k . EXP_k steht für die erwartete Zitatrate, die aus der Durchschnittszitathäufigkeit der Zeitschriften resultiert, in denen die Autoren dieses Landes ihre Artikel publiziert haben.

Die Unterschiede, die sich aus der Berechnung der feld- und zeitschriftenspezifischen erwarteten Zitatraten ergeben, werden in Tabelle 12 aufgezeigt. In der Analyse mit Selbstzitat³ erhalten die USA einen feldspezifischen Index (F -Index) von 1,41 im Vergleich zu einem zeitschriftenspezifischen Index (Z -Index) von 1,15. Für Deutschland führt dies zu Werten von 1,31 bzw. 1,19. Die Z -Indizes sind meist niedriger als die F -Indizes. Dennoch ist es entscheidend, dass die Länder in unterschiedlicher Weise betroffen sind. Gemäß dem F -Index erhalten die USA einen Rang 3, mit dem Z -Index

³ Die Relevanz der Eigenzitate wird in Abschnitt 4.6 genauer diskutiert.

lediglich einen Rang von 9. Deutschland liegt auf Rang 8 gemäß dem F-Index und auf Rang 4 nach dem Z-Index. Dieses reflektiert die verbesserte Berücksichtigung des Sprach-Bias bei dem Z-Index. Kleinere Länder wie die Niederlande oder Schweden, bei denen die Autoren generell eine stärkere Orientierung auf englischsprachige Zeitschriften haben, sind weniger von den Unterschieden zwischen dem F- und dem Z-Index betroffen.

Tabelle 12: Indizes der erwarteten Zitatraten für ausgewählte Länder nach unterschiedlichen Definitionen für alle Gebiete von WoS, 2007

Land	Z-Index oE	Z-Index mE	F-Index mE	Beob. Zitr.	Rang Z oE	Rang Z mE	Rang F mE	Rang beob. Zitr. oE
Schweiz	1.15	1.26	1.59	6.2	1	1	1	1
Niederlande	1.10	1.20	1.52	5.8	2	3	2	2
Finnland	1.09	1.20	1.34	4.8	3	2	6	8
Schweden	1.09	1.17	1.37	5.2	4	6	5	4
Großbritannien	1.08	1.19	1.37	5.1	5	5	4	5
Deutschland	1.08	1.19	1.31	4.8	6	4	8	6
USA	1.07	1.15	1.41	5.5	7	9	3	3
Kanada	1.06	1.16	1.32	4.8	8	8	7	7
China	1.05	1.10	0.83	2.8	9	11	13	13
Frankreich	1.04	1.16	1.26	4.5	10	7	9	9
Italien	1.03	1.14	1.21	4.4	11	10	10	10
Korea	0.97	1.04	0.88	3.0	12	12	12	12
Japan	0.93	1.01	0.96	3.5	13	13	11	11

Z-Index = Journal-spezifischer Index

F-Index = Feld-spezifischer Index

mE = mit Eigenzitaten

oE = ohne Eigenzitate

Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Zum Zwecke der Illustration sind in Tabelle 12 ebenfalls die beobachteten Zitatraten dokumentiert. Das impliziert ebenfalls ein anderes Ranking der Länder, wiederum vor allem mit Begünstigung für die USA. Bei der Betrachtung der Herkunft der Zitate, die die USA erhalten (mit Eigenzitaten) fällt auf, dass sie im Vergleich zu anderen Ländern einen relativ hohen Anteil an Zitaten von Forschern aus dem gleichen Land erhalten. Tabelle 13 zeigt diese Ergebnisse für ein ausgewähltes Länderset. Für die Berechnung

wurden die zitierenden Autoren whole count gezählt, d.h. für manche Land überschreitet die Summe der Zitierungen aus allen anderen Ländern 100%, da Zitierungen von Autoren aus verschiedenen Ländern mehrfach gezählt werden. Es wird offensichtlich, dass die Forscher aus den USA sich gegenseitig in höherem Maße zitieren als Forscher aus Vergleichsländern. Es muss jedoch in Rechnung gestellt werden, dass die USA zwar einige exzellente Universitäten haben, aber ebenfalls wenig forschungsaktive, sodass die niedrigere Position nach den F- oder Z-Indizes angemessen erscheint.

Tabelle 13: Aufschlüsselung der Zitate ausgewählter Länder anhand der Herkunft der Autoren des zitierenden Dokuments

		Zitierendes Land						
		CN	DE	FR	GB	JP	NL	US
Autoren-Land	CN	52,03%	5,03%	3,32%	4,78%	4,88%	1,29%	18,26%
	DE	7,29%	31,89%	7,56%	10,40%	4,93%	4,19%	29,39%
	FR	6,82%	11,40%	27,50%	10,84%	4,84%	3,77%	28,78%
	GB	6,06%	10,66%	7,20%	28,62%	4,23%	4,50%	32,10%
	JP	10,77%	8,06%	5,13%	6,87%	34,11%	2,09%	25,99%
	NL	5,16%	12,70%	7,51%	12,73%	3,96%	21,61%	31,53%
	US	6,98%	8,44%	5,42%	8,83%	4,47%	3,03%	54,06%

Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Der Indikator IA zeigt, ob die Autoren eines Landes im Vergleich zum Weltdurchschnitt in international mehr oder weniger sichtbaren Zeitschriften publizieren. Durch einen hohen Anteil an Publikationen in international sichtbaren Zeitschriften wird eine intensive Beteiligung an der internationalen wissenschaftlichen Diskussion dokumentiert. Ähnlich zum ZB-Index, zeigen positive Werte eine Orientierung oberhalb des Weltdurchschnitts an. Werte von 0 entsprechen dem Weltdurchschnitt. Der IA-Index wird in folgender Weise berechnet:

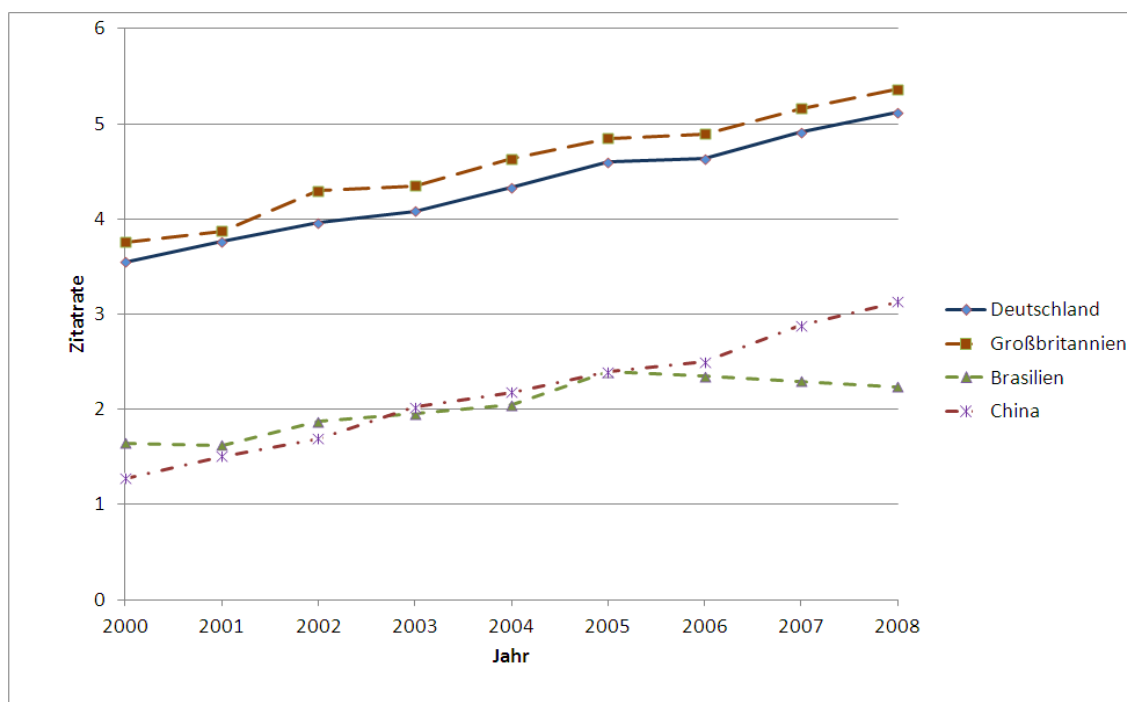
$$IA_k = 100 \tanh \ln (EXP_k / OBS_w)$$

Dieselben Definitionen wie für den ZB-Index sind gültig. Der Index w bezieht sich auf die Welt insgesamt.

Die Entscheidung für zeitschriftenspezifische erwartete Zitatraten erlaubt eine differenzierte Analyse durch die ZB- und IA-Indizes. Es kann so untersucht werden, ob die Zitatraten auf wissenschaftlich hochwertige Publikationen oder aber der Platzierung einer Publikation in einer international gut sichtbaren Zeitschrift zurückzuführen sind.

Die Interpretation der beobachteten und zeitschriftenspezifisch erwarteten Zitatraten sei am Beispiel der Länder Großbritannien, Deutschland, China sowie Brasilien illustriert. Dabei repräsentieren Deutschland und Großbritannien typische Industrieländer, von denen Deutschland einen großen eigenen Sprachraum hat. Der Vergleich ist deshalb interessant, weil beide Länder ein ähnliches Publikationsniveau aufweisen. Brasilien und China wurden für dieses Beispiel ausgewählt, da beide Schwellenländer sind, allerdings aus verschiedenen Regionen und in der Konsequenz mit sehr unterschiedlichem historischem und kulturellem Hintergrund ihres Innovationssystems. In diesem Vergleich konnten Deutschland und Großbritannien bei den beobachteten Zitatraten ihre Quoten ständig verbessern, wobei Großbritannien mit seinen Quoten etwas über Deutschland liegt (Abbildung 14). Auch bei China und Brasilien steigen die beobachteten Zitatraten stetig an, wobei im Falle Brasiliens seit 2005 ein leichter Abfall zu beobachten ist.

Abbildung 14: Beobachtete durchschnittliche Zitatraten für vier ausgewählte Länder in SCIE und SSCI ohne Eigenzitate

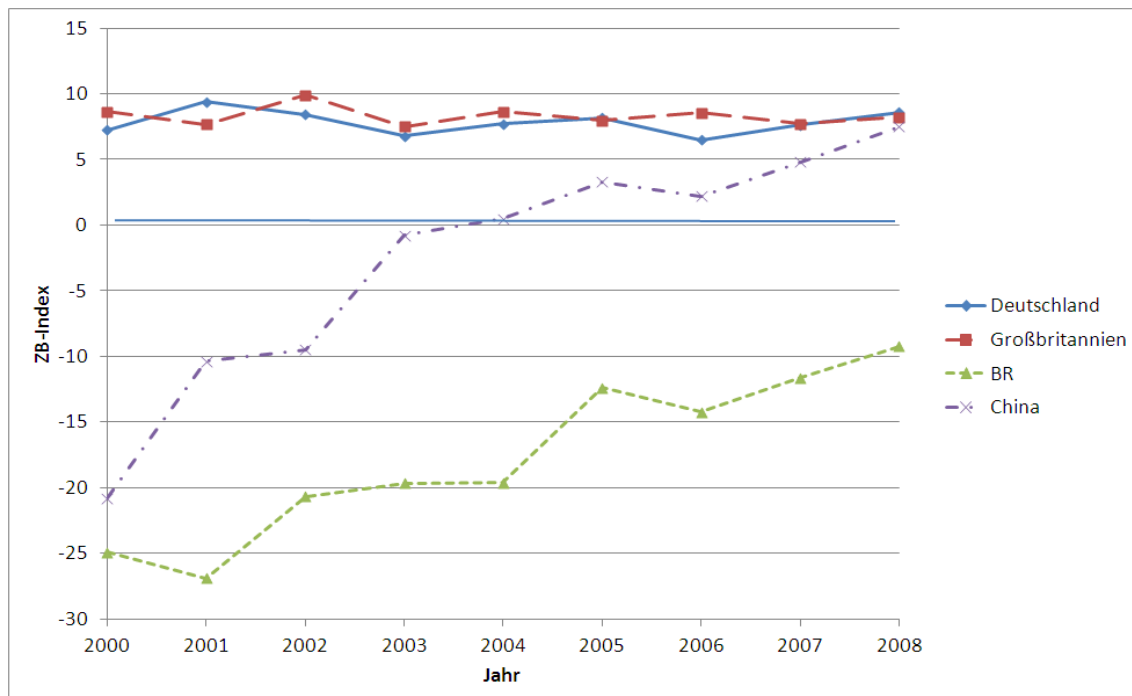


Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Für die beobachteten Zitatraten ergibt sich unter dem Gesichtspunkt der ZB- und IA Indizes ein differenzierteres Bild: Bei den ZB-Indizes zeigt sich, dass die von Deutschland und Großbritannien erheblich über denen Brasiliens und Chinas liegen und im Zeitverlauf sehr stabil sind (Abbildung 15). Die ZB-Werte befinden sich für Deutschland und Großbritannien auf einem ähnlichen Niveau, d.h. die höheren Werte bei den beo-

bachteten Zitatraten zeigen sich im ZB-Index nicht mehr. Bei China beginnen die ZB-Indizes im Jahr 2000 auf einem deutlich unterdurchschnittlichen Niveau von -20, d.h. auf einem Niveau deutlich unterhalb des Weltdurchschnitts. Im Jahr 2004 erreicht es allerdings den Weltdurchschnitt und steigt auch seit dieser Zeit weiter an, allerdings weniger dynamisch. Im Jahr 2008 hat China bei den ZB-Indizes nahezu das deutsche und britische Niveau erreicht. Bei Brasilien beginnen die ZB-Indizes im Jahr 2000 bei einem Wert von -25, also auf einem ähnlich niedrigen Niveau wie China. Seitdem hat sich der Index bis zu einem Wert von -9 verbessert; er steigt ähnlich wie bei China an, aber deutlich weniger stark.

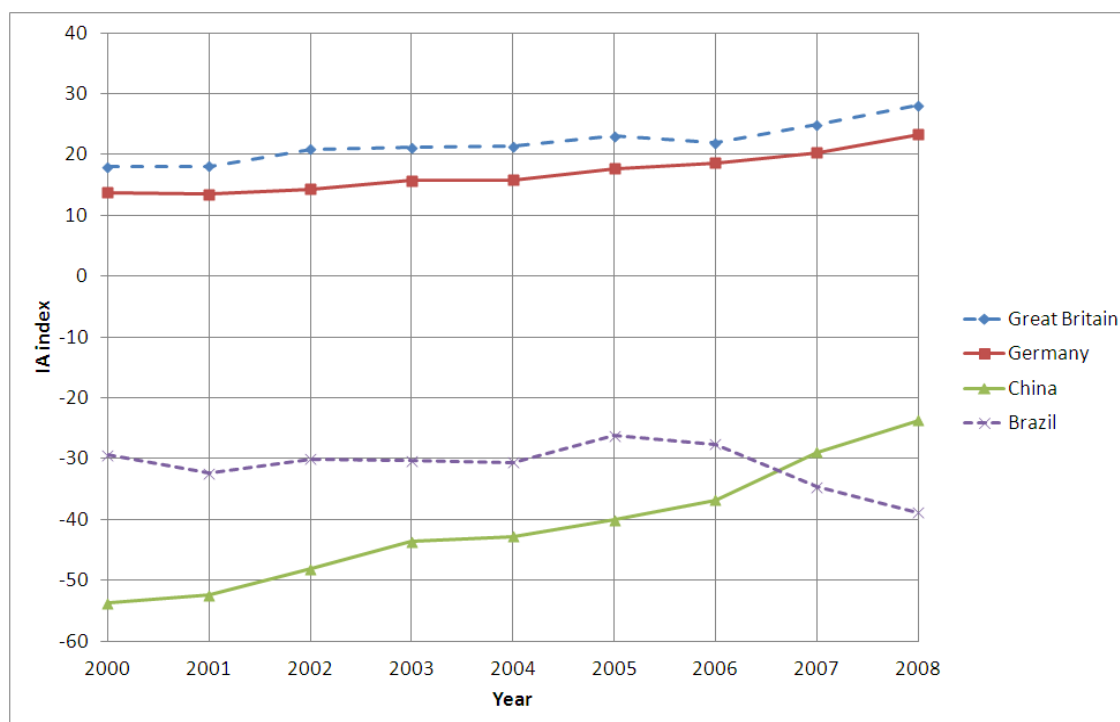
Abbildung 15: Index der Zeitschriftenspezifischen Beachtung (ZB) für vier ausgewählte Länder in SCIE und SSCI ohne Eigenzitate



Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Bei der Berechnung der Internationalen Ausrichtung liegen die deutschen und britischen Werte auf einem positiven Niveau mit leichtem Anstieg im Zeitverlauf (Abbildung 16). Die britischen Werte liegen allerdings stets oberhalb der deutschen, da sie offensichtlich in Zeitschriften mit etwas höherer Sichtbarkeit publizieren. D.h., die etwas besseren beobachteten Zitatraten Großbritanniens im Vergleich zu Deutschland sind in erster Linie auf die Platzierung britischer Artikel in international etwas sichtbareren Zeitschriften zurückzuführen.

Abbildung 16: Index der Internationalen Ausrichtung (IA) für vier ausgewählte Länder in SCIE und SSCI ohne Eigenzitate



Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Im Falle Brasiliens ist der IA-Index im Zeitverlauf relativ stabil und nimmt in den letzten Jahren leicht ab. Das langsame Wachstum des ZB-Indexes im Zeitverlauf kombiniert mit einem Abfall des IA-Indexes in den letzten Jahren erklärt den leichten Abfall der beobachteten Zitatraten in den letzten Jahren, wie sie sich aus Abbildung 14 ergeben. Im Falle Chinas wächst der IA-Index stetig an, aber mit einem aktuellen Wert von -24 ist dieser Wert nach wie vor deutlich unterhalb des Weltdurchschnitts. Da die beobachteten Zitatraten stetig wachsen, wachsen sowohl der ZB- als auch der IA-Index für China an. Der ausgesprochen positive ZB-Index für China in den aktuellen Jahren muss in Kombination mit dem nach wie vor deutlich negativen IA-Index interpretiert werden. Denn der überdurchschnittliche ZB-Index wird durch Publikationen in Zeitschriften erreicht, welche eine begrenzte internationale Sichtbarkeit haben. In dieser speziellen Kombination sollten die mittlerweile positiven ZB-Indizes für China nicht überinterpretiert werden. Dennoch ist festzuhalten, dass die Zahl und der Impact der Publikationen Chinas sich ständig verbessert haben, und dass in Zukunft sich die Rolle chinesischer Beiträge zur Wissenschaft sowohl quantitativ als auch qualitativ verbessern wird.

Die Beispiele der vier Länder dokumentieren deutlich, dass durch die Kombination der IA- und ZB-Indizes eine genauere Interpretation der beobachteten Zitatraten erreicht werden kann.

4.5 Relevanz von Eigenzitat

Eine weitere methodische Frage besteht darin, ob Zitatraten mit oder ohne Eigenzitate berechnet werden sollten. In diesem Kontext werden in der Literatur verschiedene Argumente für und gegen den Einschluss von Eigenzitat angeführt (Costas et al. 2010; Glänzel et al. 2004). Die Argumente gegen den Einschluss von Eigenzitat sind, dass nur Zitationen anderer Wissenschaftler mit dem Konzept des Impact verbunden werden können und dass die regionalen Gewohnheiten bei der Nutzung von Eigenzitat sich unterscheiden. Vor allen Dingen in angelsächsischen Ländern werden Eigenzitate seltener als in vielen anderen Ländern genutzt. Für die Publikationen aus dem Jahre 2006 betrug für deutsche und französische Publikationen der Anteil der Eigenzitate 27%, bei Großbritannien 23% und bei den Vereinigten Staaten 21%. Es ist nicht möglich, daraus irgendwelche sinnvollen inhaltlichen Unterschiede für diese unterschiedlichen Anteile von Eigenzitate anzuführen. Wenn die Eigenzitate ausgeschlossen werden, verschlechtert sich im internationalen Ranking die Position Deutschlands um 2 Positionen, sodass es dann hinter Schweden und Großbritannien liegt (Tabelle 12, Spalten 6 und 7).

Ein Argument für den Einschluss von Eigenzitat besteht darin, dass in neuen Feldern ihre Beachtung wesentlich sei. Wenn ein Artikel von vielen Ko-Autoren veröffentlicht wird, führt dies zum Ausschluss vieler relevanter Zitate. Schließlich führt der Ausschluss von Eigenzitat zu Fehlern aufgrund des Homonym-Problems⁴.

Als einfachster Weg zum Ausschluss von Eigenzitat hat sich die Nutzung der Author Identity erwiesen, die vom Hersteller bereitgestellt wird. Auf diese Weise wird ein eigener ausführlicher lexikalischer Vergleich der Autorennamen der Publikation und der Autorennamen in den zitierten Zeitschriften vermieden. Der Author Identifier ist selbstverständlich nicht immer korrekt, dieser Fehler muss aber in Kauf genommen werden, wenn Eigenzitate ausgeschlossen werden sollen. Das Fraunhofer ISI schließt sich der Empfehlung von CWTS an, Eigenzitate auszuschließen, um damit wirklich den Impact von Publikationen abbilden zu können. Dafür wird notgedrungen ein kleinerer Fehler in Kauf genommen. Insgesamt erweist sich allerdings die Auswirkung auf das Ranking bei Ländervergleichen als relativ begrenzt.

⁴ Im Kontext dieses Berichts wird für den Begriff Homonym die Definition von Aksnes (2008) angewandt.

4.6 Personenspezifische Zitatraten

In der aktuellen Situation werden bibliometrische Analysen in großem Umfang zur Bewertung von individuellen Wissenschaftlern eingesetzt. Insbesondere bei der Berufung von Professoren oder der Besetzung anderer akademischer Positionen wird verlangt, dass die Bewerber ihre Publikationslisten vorlegen. Dabei wird etwa im Falle von Medizinern erwartet, dass sie der jeweiligen Literaturstelle auch den Impact-Faktor der Zeitschrift beifügen, in welcher der jeweilige Beitrag publiziert wurde. Dieses ist allerdings äußerst problematisch, da im Einzelfalle keineswegs damit festgestellt wird, wie hoch die tatsächliche Zitatrate der jeweiligen Publikation ist. Auch in Zeitschriften mit hohem Impact-Faktor ist es möglich, dass die einzelne Publikation überhaupt kein Zitat erhält. In Zeitschriften mit niedrigem Impact-Faktor können dagegen einzelne Publikationen dennoch sehr hoch zitiert werden. Von daher ist von der Verwendung des Impact-Faktors als Maß für die Qualität einer Einzelpublikation dringend abzuraten.

Im Falle von reinen Publikationslisten ohne Angaben von Zitatquoten ist es üblich, dass Mitglieder des Auswahlkomitees selbst die einzelnen Publikationen in Google-Scholar im Hinblick auf ihre Zitathöhe überprüfen. Das ist methodisch äußerst problematisch, da die Zitierungen in Google-Scholar auf einer Auswertung des Internets basieren und das Internet ständig ergänzt wird oder auch Seiten gelöscht werden. Von daher sind die Zitatquoten für einzelne Publikationen äußerst instabil und können sich innerhalb kurzer Zeiträume erheblich ändern. Es ist auch unklar, woher die Zitate tatsächlich stammen. Von daher ist von der Nutzung von Google-Scholar für die Ermittlung von Zitatraten ebenfalls dringend abzuraten, wenn es um die korrekte Bewertung einzelnen Personen geht.

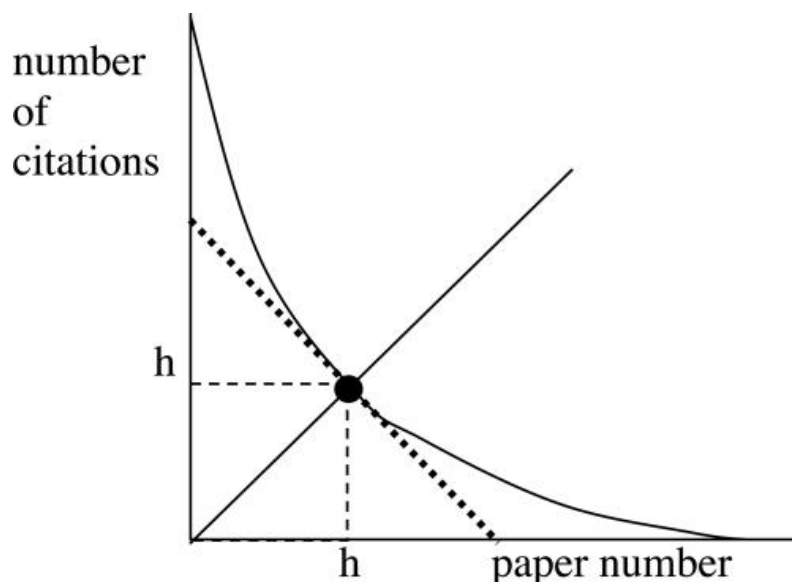
Allerdings ist es ohnehin problematisch, Individuen über bibliometrische Analysen zu bewerten, da die Zahl der Publikationen in der Regel äußerst niedrig und damit in statistischer Sicht die Stichprobe sehr klein ist. In der Regel führt die bibliometrische Bewertung von Wissenschaftlern zwar durchaus zu sinnvollen Ergebnissen. Es sind aber auch Fehlbewertungen möglich, etwa wenn bahnbrechende Publikationen in völlig neuen Gebieten, die eigentlich besonders wichtig sind, erst mit großer Verzögerung zitiert werden, weil die Bedeutung der Publikation aufgrund ihres Neuheitsgrades von Kollegen nicht sofort erkannt wird. Umgekehrt können einzelne Publikationen aufgrund von besonderen Situationen deutlich zu hoch zitiert werden, wenn es um die Qualität des Inhaltes geht. So kann eine ausgesprochen schlechte Publikation häufig als Negativbeispiel zitiert werden, was in jedem Fall kein Indikator für Impact oder Qualität sein kann. Aufgrund der kleinen Stichprobe, die sich für Individuen ergibt, können durch diese Ausreißer erhebliche Fehler entstehen. Von daher wird grundsätzlich von der

bibliometrischen Bewertung von Individuen abgeraten, was der Empfehlung von Gläser et al. (2004) entspricht, die diese Problematik genauer untersucht haben.

Die bibliometrische Bewertung von Individuen hat in jüngster Zeit durch die Einführung des sogenannten Hirsch-Index oder kurz h-Index wieder einen erheblichen Auftrieb erfahren (Hirsch 2005). Mittlerweile wird der h-Index von vielen Einrichtungen benutzt, ohne über dessen Aussagekraft genauer zu reflektieren. Auch beim h-Index gilt, da es um die Bewertung von Individuen geht, die Problematik der Fehlbewertung einzelner Publikationen, die zu erheblichen Falschaussagen bei kleinen Untersuchungseinheiten führen. Darüber hinaus gibt es aber auch spezifische Einwände gegen die Konstruktion des h-Indexes und seiner Aussagekraft.

Der h-Index ist in folgender Weise definiert: Der h-Index eines Wissenschaftlers mit n Publikationen entspricht der Anzahl h seiner Publikationen, die mindestens h -mal zitiert wurden, während die verbleibenden $(n-h)$ Publikationen h oder weniger Zitate erhalten haben. Hirsch spricht hierbei auch von dem "h-core" eines Wissenschaftlers, der besagte h Publikationen umfasst. Die Ermittlung des h-Index kann leicht veranschaulicht werden, wenn man die Publikationen eines Autors anhand der Anzahl ihrer Zitierungen absteigend sortiert. Abbildung 17 zeigt die Ermittlung des h-Indexes bei einer solchen Sortierung: Der Wert h entspricht hier genau dem Schnittpunkt mit der 45°-Linie.

Abbildung 17: Illustration der Ermittlung des h-Indexes bei Sortierung der Publikationen anhand ihrer Zitierungen

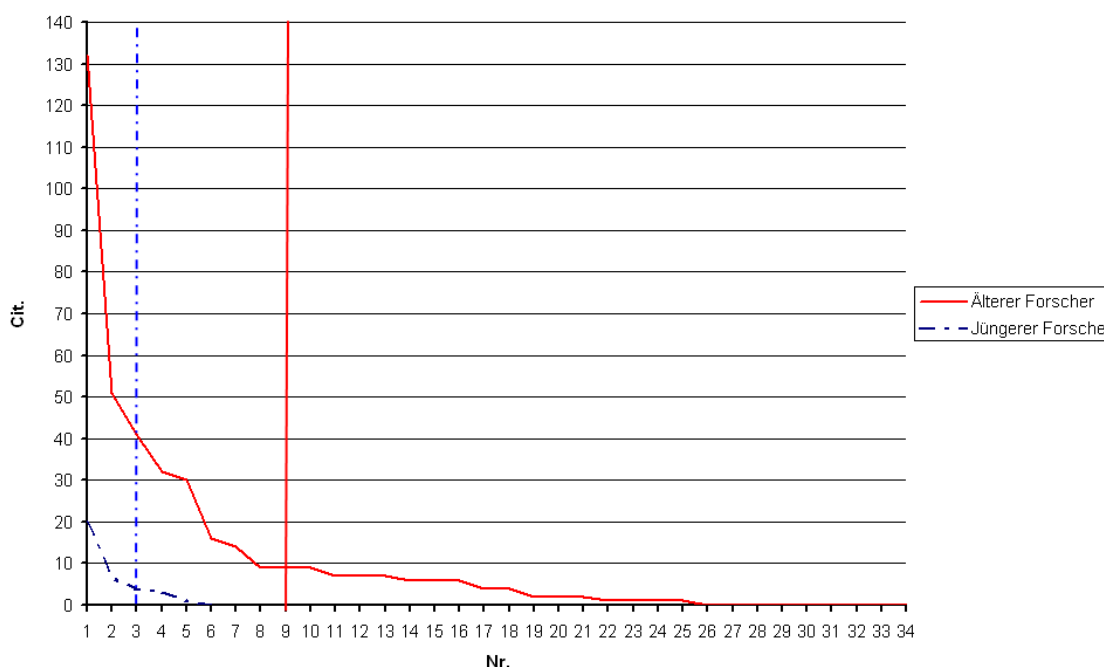


Quelle: Hirsch (2005)

Der h-Index erfreut sich großer Beliebtheit, weil er scheinbar in einer einzigen Zahl die Qualität eines Autors angibt, ohne auf komplexe bibliometrische Indikatoren wie erwar-

tete Zitatraten zurückgreifen zu müssen. Der h-Index hat bei der Bewertung von Individuen in der Tat den Vorteil, dass extreme Ausreißer weniger ins Gewicht fallen, als das bei der üblichen Mittelwertbildung der Fall ist. Letztlich gibt der h-Index an, inwieweit es einem Autor gelungen ist, hoch zitierte Artikel zu publizieren. Das sei am Beispiel eines realen Autors illustriert, bei dem die Zitierungen über alle Publikationen eine erhebliche Schiefverteilung aufweisen. Er erreicht über die Gesamtzahl seiner 45 Publikationen einen h-Index von neun, der etwas über dem Mittelwert seiner Zitatrate liegt (Abbildung 18). In diesem Einzelfall ist natürlich zu fragen, ob die einzelne sehr hoch zitierte Publikation mit 132 Zitaten nicht ein stärkeres Gewicht bekommen sollte.

Abbildung 18: Ermittlung des h-Indexes für einen älteren und einen jüngeren Forscher des gleichen Fachgebiets



Quelle: Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Entscheidender ist jedoch, dass der h-Index nicht in Rechnung stellt, dass die Zitatraten in unterschiedlichen Disziplinen sehr verschieden sind, sodass letztlich nur ein Vergleich von Individuen in demselben Gebiet zu sinnvollen Aussagen führen kann. Darüber hinaus ist zu bedenken, dass der h-Index für die einzelnen Publikationen keine festen Zitatfenster vorsieht, sodass grundsätzlich ältere Wissenschaftler einen höheren h-Index erreichen als jüngere, da ein größerer Zeitraum zum Eingang von Zitaten zur Verfügung steht. Einzelne Beispieluntersuchungen des Fraunhofer ISI haben gezeigt, dass dies zu extremen Unterschieden führen kann (z.B. Abbildung 18). Somit ist der h-Index eigentlich nur zum Vergleich gleichaltriger Wissenschaftler geeignet

(Hirsch 2005). In der Literatur sind zwar vielfache Vorschläge zur Verbesserung des h-Index gemacht worden, wobei es letztlich aber einfacher erscheint, die üblichen bibliometrischen Indizes einzusetzen.⁵

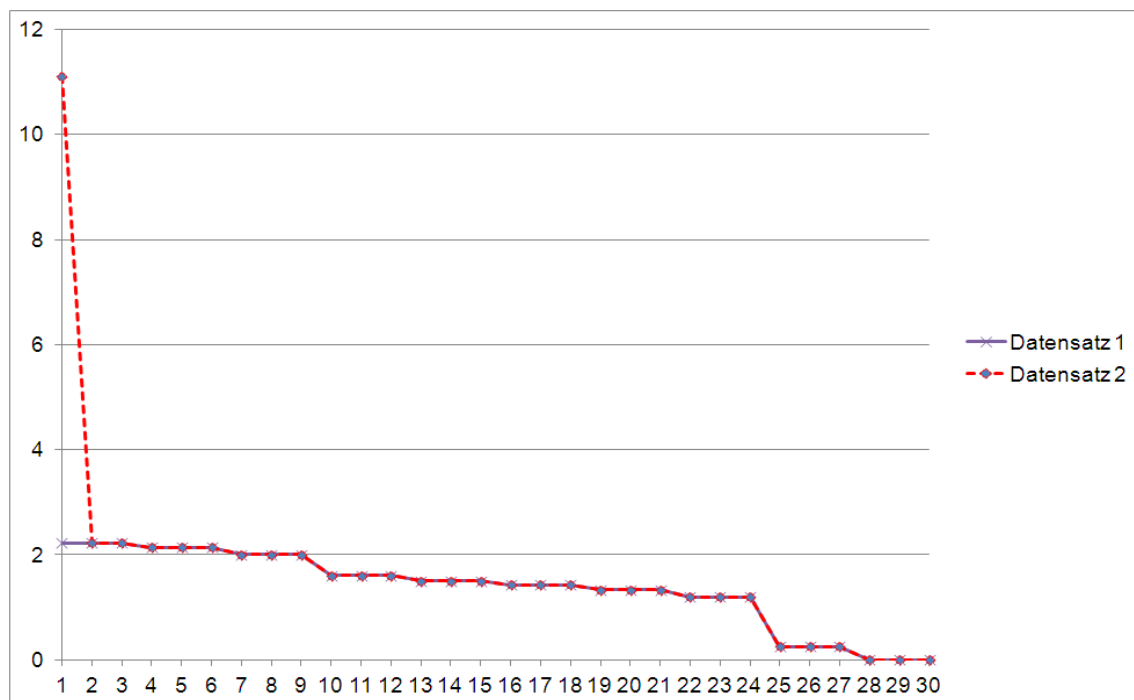
4.7 Adäquate Ermittlung der Mittelwerte von Zitatraten

Bei den Vergleichen von wissenschaftlichen Einheiten oder Ländern mit bibliometrischen Verfahren werden stets Mittelwerte von Zitatraten berechnet und zueinander in Beziehung gesetzt. Dabei gibt es zwei unterschiedliche Möglichkeiten, diese Mittelwerte zu ermitteln. Die niederländische Forschungsgruppe CWTS berechnet die Zitatraten üblicherweise, indem sie sämtliche Zitate einer Einheit durch die Summe der Publikationen dividieren. Dieses pragmatische Verfahren ist von Opthof und Leydesdorff (2012) kritisiert worden mit dem Vorwurf der mathematischen Inkorrektheit. Die Autoren schlagen stattdessen vor, die Zitatraten jeder einzelnen Publikation aufzuaddieren und daraus den Mittelwert zu bilden. Diese Auseinandersetzung um die adäquate Ermittlung der Mittelwerte von Zitatraten wird auch unter der Überschrift "Averages of Ratios vs. Ratios of Averages (AoR vs. RoA)" diskutiert. Van Raan und Kollegen (van Raan et al. 2010) argumentieren damit, dass sich zwischen beiden Ansätzen kaum Unterschiede ergeben und dass nach dem Leidener Vorgehen Extremwerte besser neutralisiert werden. Die faktischen Konsequenzen der verschiedenen Ansätze sind in La Rivière und Gingras (2011) ausführlich dargestellt.

Die Unterschiede der beiden Berechnungsverfahren seien noch einmal anhand eines Beispiels illustriert. Dabei werden die Publikationen eines fiktiven Autors mit 30 Publikationen mit jeweils unterschiedlichen Zitatraten betrachtet. In der Grundverteilung der Artikel liegt die höchste Zitatrate etwas über 2, die niedrigste bei 0 (Abbildung 19, Datensatz 1). Daraus ergibt sich bei der Aufsummierung und anschließender Division der Zähler und Nenner ein Mittelwert von 1,54. Bei der Berechnung der Mittelwerte der jeweiligen Zitatquoten ergibt sich ein Mittelwert von 1,37. Der Unterschied zwischen beiden Werten liegt bei 13%. Sobald ein Ausreißer eingefügt wird, wie in dem Datensatz 2, ergeben sich Mittelwerte von 1,96 bzw. 1,66. Der Unterschied liegt jetzt bei 27%, ist also größer.

⁵ siehe die Übersicht bei Zhang et al. (2011)

Abbildung 19: Fiktive Verteilung der Zitatraten eines Autors mit 30 Publikationen



Quelle: Analysen und Berechnungen des Fraunhofer ISI.

Bei dieser kleinen Datenmenge schlägt der Extremwert bei der Summe der Zähler und Nenner und deren Division stärker zu Buche als bei dem Mittelwert der jeweiligen Raten. Bei sehr großen Stichproben ist das Verhältnis umgekehrt, wie es van Raan et al. (2010) zutreffend dargestellt haben (siehe auch Larivière/Gingras 2011). In größeren Datensätzen sind die Unterschiede von beiden Verfahren demnach auch bei Hinzufügen eines Extremwertes nicht erheblich. Im Vergleich dazu sind die Auswirkungen solcher Extremwerte bei kleinen Datenmengen und bei Einheiten mit niedrigeren Zitatquoten größer. Die Entscheidung für eines der beiden Verfahren ist deshalb von der Datenmenge, der Art der Analyse (z.B. Mikro- oder Makro-Analyse) und anderen Faktoren abhängig. Bei größeren Datenmengen sind die Ergebnisse dennoch nahezu identisch sind.

Innerhalb des Kompetenzzentrums gibt es hier noch keine abschließende Meinungsbildung, da noch nicht genug Analysen zum Vergleich der beiden Verfahren durchgeführt wurden. Im Projekt "Erwartete Zitatraten und Klassifikationen" wurde das AoR-Verfahren angewendet.

4.8 Zitatraten von Zeitschriften

4.8.1 Impact-Faktor

In bibliometrischen Analysen werden nicht nur Zitatraten von einzelnen oder Gruppen von Artikeln benutzt, sondern auch Zitatraten von Zeitschriften insgesamt. Der bekannteste Indikator ist diesbezüglich der Impact-Faktor, der von Eugene Garfield eingeführt wurde und die Bedeutung wissenschaftlicher Zeitschriften indiziert. Der Impact-Faktor ist sicherlich das populärste bibliometrische Maß, welches innerhalb der Bibliometrie benutzt wird. Der Impact-Faktor wird regelmäßig in den Journal Citation Reports (JCR) publiziert, die zyklisch vom WoS-Hersteller Thomson Reuters bereit gestellt werden. Der Impact-Faktor ist in der folgenden Weise definiert:

$$IF(t) = \frac{CIT_t}{Publ_{[t-2:t-1]}}$$

Der Impact-Faktor analysiert, wie viel Zitate eine Zeitschrift in dem analysierten Jahr t erhält, wobei die Zitate durch die Publikationszahlen der vorliegenden Jahre geteilt werden. Der Impact-Faktor stellt somit eine durchschnittliche Zitatrate der betrachteten Zeitschrift dar. Eine gute Übersicht über die Vor- und Nachteile des Impact-Faktors ist in Glänzel und Moed (2002) geboten. Der Impact-Faktor wird in der Literatur sehr intensiv diskutiert, und es werden zur Verbesserung von Problemen verschiedene Alternativen vorgeschlagen (Glänzel/Moed 2002).

Die zentralen Probleme bestehen darin, dass durch die unterschiedlichen Zitatgewohnheiten in verschiedenen Disziplinen die Zitatquoten die Impact-Faktoren der Zeitschriften sehr stark disziplinenabhängig sind. Von daher kann eine breite fachliche Abdeckung von Zeitschriften in WoS über verschiedene Disziplinen nur erreicht werden, wenn als Eintrittskriterium fachspezifische Impact-Faktoren festgelegt werden. Die Bedeutung des Impact-Faktors resultiert vor allem daraus, dass er vom Datenbankproduzenten als zentrales Kriterium genutzt wird, um Zeitschriften in WoS aufzunehmen. Daraus ergeben sich erhebliche jährliche Schwankungen, da in der Folge immer wieder neue Zeitschriften aufgenommen werden und alte herausfallen (Michels/Schmoch 2011). Ein weiterer zentraler Nachteil des Impact-Faktors besteht darin, dass vor allem bei kleineren Zeitschriften der Impact-Faktor von einzelnen sehr hoch zitierten Artikeln abhängt und damit erhebliche jährliche Schwankungen möglich sind. Da Zitatverteilungen stets extrem schief verteilt sind, ist der Mittelwert ohnehin ein problematisches Maß. In der Folge wurde eine Reihe von ergänzenden und alternativen Vorschlägen zur Messung der Zeitschriftenzitationen gemacht. Diese sind insbesondere:

- "– Instead of the mean: other parameters of the distribution (e.g., percentage of uncited papers or quantiles);
- Instead of integer counting of citations: weighting a citation on the basis of the journals in which it is made;
- Instead of applying a single citing year: application of a range of citing years;
- Instead of analysing all ('citable') documents: disaggregate articles on the basis of document type (article + note + review) or content (e.g., theoretical, methodological and experimental);
- Instead of considering only papers 1-2 years earlier: analysing articles from older 'ages'" (Glänzel/Moed 2002: 173).

Bei den Arbeiten des Fraunhofer ISI hat sich herausgestellt, dass es wichtig ist, welche Dokumente bei Zitaten und zitierenden Dokumenten genutzt werden. Wie bereits von Moed und Leeuwen (1995) festgestellt, nutzt Thomson Reuters im Zähler Zitate auf alle Dokumenttypen, also auch Proceedings, bei den Publikationen im Nenner, aber nur die Dokumenttypen Article, Note und Review. Letters, Editorials, Discussions, Abstracts, Book Reviews oder Corrections werden dabei ausgeschlossen. Nach dieser Formel bestehen zwar gewisse Inkonsistenzen zwischen den Dokumenttypen der zitierten Publikationen bzw. deren Zitaten und den Publikationen, zu denen diese ins Verhältnis gesetzt werden. Dennoch ist sicherlich der von Thomson Reuters benutzte Ansatz mit allen Dokumenttypen bei den zitierten Publikationen und nur Articles, Notes und Reviews bei den gezählten Dokumenten ein pragmatischer und sinnvoller Weg. Selbstverständlich ist es möglich, dass Thomson Reuters seine Vorgehensweise diesbezüglich verändert hat, bis dato ist dies aber noch nicht Gegenstand neuerer Untersuchungen geworden.

Am Fraunhofer ISI sind Algorithmen für die Ermittlung des Impact-Faktors für verschiedene Dokumente erstellt und getestet worden. Diese Algorithmen wurden sowohl für WoS als auch für Scopus erstellt.

Allerdings wird der Impact-Faktor für Ländervergleiche oder auch die Bewertung von Einrichtungen nicht benötigt. Die Nutzung des Impact-Faktors zur Beurteilung der Bedeutung einzelner Publikationen ist ohnehin problematisch, da der Impact-Faktor einer Zeitschrift, in der ein Artikel publiziert wurde, nichts über die tatsächliche Zitierung des betrachteten Artikels aussagt. Die entsprechende Publikation kann überhaupt nicht zitiert worden sein oder aber deutlich über dem Durchschnitt der Zeitschrift, sodass diese Verwendung des Impact-Faktors als problematisch zu betrachten ist. Vor diesem Hintergrund wurde bislang nicht weiter verfolgt, Impact-Faktoren flächendeckend zu berechnen. Eine entsprechende Diskussion im Rahmen des Kompetenzzentrums steht noch aus.

Von den vielen Alternativen, die zum Impact-Faktor in der Literatur vorgeschlagen werden und die in Glänzel und Moed (2002) ausführlich dokumentiert sind, seien an dieser Stelle nur zwei Varianten herausgehoben, die am Fraunhofer ISI näher betrachtet wurden. Dabei handelt es sich zum einen um den sogenannten Audience-Faktor, zum anderen um den SNIP-Faktor.

4.8.2 Audience-Faktor

Der Audience-Faktor wurde von Zitt und Small (2008) entwickelt. Der Audience-Faktor gewichtet die Zitate, die eine Zeitschrift bekommt, abhängig von der Zeitschrift, in der der zitierende Artikel publiziert wurde. Das Gewicht w_i einer Zeitschrift i richtet sich dabei danach, wieviele Zitierungen von ihr im Durchschnitt vergeben werden. Damit erhält also ein Zitat aus einer Zeitschrift, die insgesamt wenige Referenzen emittiert, ein höheres Gewicht.

Bei der Berechnung des Audience-Faktors wird ebenfalls berücksichtigt, wieviele Zitate alle Zeitschriften im Durchschnitt vergeben, die die betrachtete Zeitschrift zitieren. Soll für eine Zeitschrift j also der Audience-Faktor berechnet werden, wird zum einen die durchschnittliche Zitatrate für jede Zeitschrift i berechnet, die j zitiert. Dies ergibt für den Zeitraum t bis T m_i . Zum anderen wird m_s , der Durchschnitt über alle m_i für j , berechnet. Das Gewicht einer Zeitschrift i ergibt sich somit als

$$w_i(t, T) = \frac{m_i(t, T)}{m_s(t, T)}$$

Mit Hilfe dieser Gewichtung werden die Zitierungen der zitierenden Zeitschrift i fraktioniert gezählt. Der Audience-Faktor einer Zeitschrift j wird dann wie folgt berechnet:

$$AF - T_j(t) = \frac{\sum_i w_i \times cit_{ij}(t, T)}{a_j(t, T)}$$

Nach den Autoren ist der Audience-Faktor damit eine Variante der fraktionierten Zitat-zählung, allerdings berechnet bezüglich der zitierenden Zeitschrift und nicht des zitierenden Artikels oder der zitierenden Disziplin, und fokussiert damit auf die Strukturen in den zitierenden Zeitschriften. Ein Vergleich mit dem normalen Impact-Faktor von Thomson Reuters führt auf ein verändertes Ranking. Dies ist exemplarisch an einem ausgewählten Set von Zeitschriften in Tabelle 14 gezeigt.

Tabelle 14: Vergleich der Rangfolge von acht Zeitschriften nach Impact-Faktor und Audience-Faktor im Vergleich (Publikationsjahr 2006)

Rang	Impact-Faktor	Audience-Faktor
1	Journal of Molecular Biology	Annals of Internal Medicine
2	Journal of Neuropathology and Experimental Neurology	Lancet
3	Annals of Internal Medicine	Journal of Molecular Biology
4	Lancet	Journal of Neuropathology and Experimental Neurology
5	Trends in Neuroscience	Nature
6	New England Journal of Medicine	Science
7	Science	Trends in Neuroscience
8	Nature	New England Journal of Medicine

Quelle: WoS, Recherchen und Berechnungen des Fraunhofer ISI.

Sicherlich besteht, wie die Autoren des Audience-Faktors richtig herausstellen, der Vorteil, dass von ihren Zitierungen her dominante Zeitschriften weniger hervortreten und die Verteilung gleichmäßiger als bei dem Impact-Faktor ist. Dennoch ergibt sich die gleiche Problematik wie bei der Berechnung des Impact-Faktors, nämlich dass die Berechnung eines Indikators für Zeitschriften keinerlei Rückschlüsse zur Beurteilung der Leistung einzelner Einrichtungen oder Länder zulässt.

4.8.3 SNIP

Der SNIP (Source Normalized Impact Paper), wurde von Moed (2010) vorgeschlagen. Er reagiert auf die Problematik, dass die Zitatgewohnheiten in unterschiedlichen Disziplinen sehr unterschiedlich sind und dass darüber hinaus viele Zeitschriften mehreren Feldern zugeordnet sind, sodass die Definition eines durchschnittlichen Feldgewichtes als Referenz problematisch ist. SNIP ist allgemein definiert als das Verhältnis der durchschnittlichen Zitate pro Artikel in einer Zeitschrift und dem Zitatpotenzial in seinem disziplinären Feld. Das disziplinäre Feld ist dabei über die Papiere definiert, die diese Zeitschrift zitieren. Im Zentrum der Definition stehen somit die zitierenden Artikel, womit eine fairere Betrachtung erreicht wird, als bei der Betrachtung über die von den Datenbankherstellern mitgelieferte Klassifikation einer Zeitschrift in einem oder in mehreren Feldern. Insgesamt führt die Anwendung des SNIP-Faktors zu einer erheblichen Veränderung der Rangfolge der Zeitschriften und sicherlich in eine deutlich verbesserte Richtung. Die Berechnung ist allerdings noch einmal aufwendiger, als das bei dem

Impact- oder Audience-Faktor der Fall wäre, da für jede Zeitschrift ein eigenes disziplinäres Feld berechnet wird.

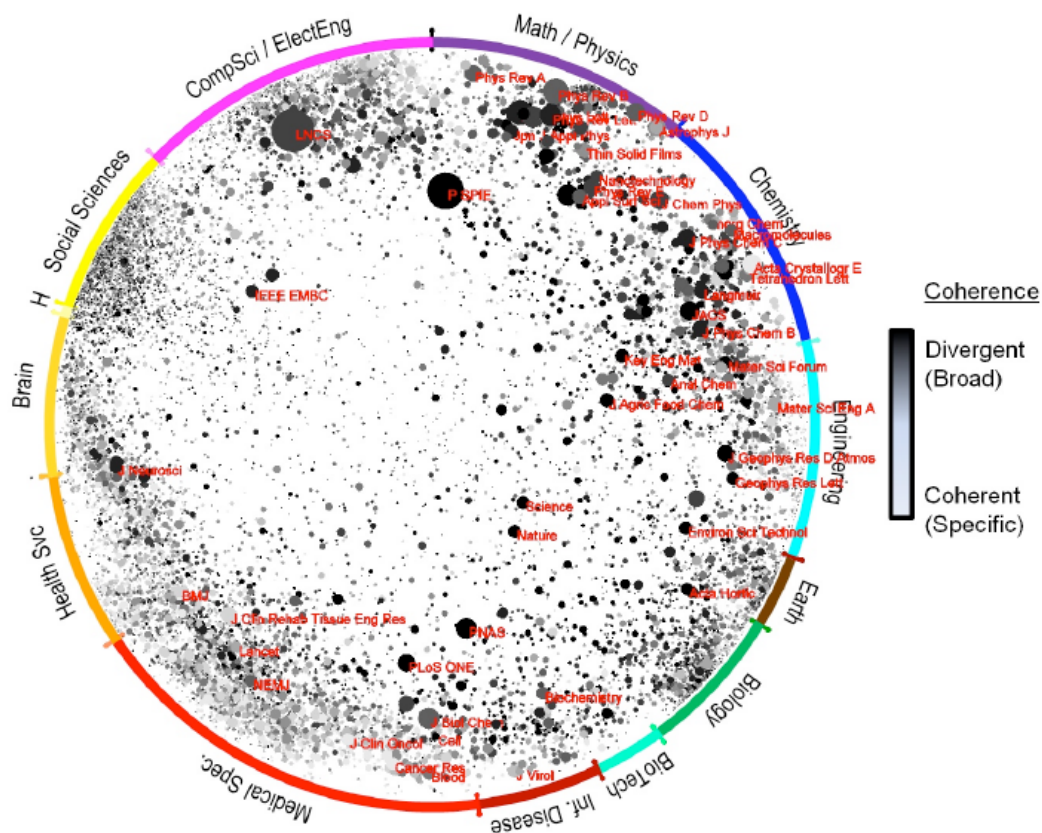
In jedem Fall handelt es sich bei dem SNIP um eine sehr avancierte Definition, die nach Moed (2010) das Verhältnis der durchschnittlichen Zitatrate der Publikationen eines Journals (hier als raw citation impact per paper, RIP, bezeichnet) und dem relativen Zitationspotential in der Datenbank (relative database citation potential) ausdrückt. Es gelten allerdings die gleichen Einschränkungen wie bei den anderen hier vorgestellten Zeitschriften-spezifischen Indikatoren und deren Anwendung und Interpretation.

5 Klassifikation von Zeitschriften

Die Klassifikationen von Artikeln in den beiden großen bibliometrischen Datenbanken Scopus und WoS der Produzenten Elsevier und Thomson Reuters beruhen letztlich auf einer Klassifikation der Zeitschriften, in denen die jeweiligen Artikel publiziert werden. Da die Datenbanken Artikel in wissenschaftlichen Fachzeitschriften abdecken, ist die Klassifikation der Zeitschriften auf der Basis wissenschaftlicher Disziplinen vorgenommen worden. Die Klassifikationen sind hilfreich bei der Konzeption von Suchstrategien nach bestimmten Themen, bei der fachlichen Bewertung der Publikationen von Untersuchungseinheiten oder auch als inhaltliche Grundlage für Forschungsbewertungen. Es hat immer wieder Versuche gegeben, die Klassifikationen von WoS und Scopus zu verbessern. Zu nennen sind hier vor allem verschiedene Ansätze aus bibliometrischer Perspektive wie z.B. Bassecouard und Zitt (1999), Boyack et al. (2005), Carpenter und Narin (1973), Glänzel und Schubert (2003), Katz und Hicks (1995), Klavans und Boyack (2010) oder Science-Metrix (2010). Auch im vorliegenden Projekt soll ein Versuch einer Neuklassifikation vorgenommen werden. Ziel ist es zum einen, die Klassifikationen, die von WoS und Scopus vorgegeben werden, ein Stück weit zu verfeinern, das Klassifikationsschema stärker einer deutschen Auffassung einer Wissenschaftsstruktur anzupassen und vor allem die zahlreichen Fehlklassifikationen von Zeitschriften in WoS und Scopus so weit wie möglich zu eliminieren.

Boyack und Klavans (2011) führen zu Recht aus, dass trotz aller Versuche es nicht möglich sein wird, ein "bestes" Klassifikationssystem zu identifizieren. Sie untersuchen diese Frage mit Hilfe der Analyse von vier Maßen der Spezifität von Zeitschriften und finden, dass diese mehrere Dimensionen hat. Ihrer Auffassung nach sind nur wenige Zeitschriften wirklich im engen Sinne disziplinär. Sie stellen damit die Validität jeglicher Disziplinen-orientierten Zeitschriftenklassifikationen in Frage, insbesondere im Hinblick auf Forschungsevaluationen. Ein Kernbereich ihrer Analyse ist die Untersuchung der textlichen Kohärenz, unter der Annahme, dass Zeitschriften mit einem begrenzten Set von Themen eine hohe Kohärenz haben, während andere mit einem breiten Set von Themen, möglicherweise multidisziplinären Themen, eine geringere Kohärenz haben. Bei einer Kartierung der Journals innerhalb des "Circle of Science" unter dem Gesichtspunkt der Kohärenz zeigt sich, dass nur wenige Zeitschriften eng am Außenrand liegen (Abbildung 20), womit sie - neben anderen Analyseverfahren - begründen, dass eine Disziplinen-orientierte Klassifikation problematisch sei. Gleichzeitig zeigt allerdings ihre Darstellung, dass sich zumindest der überwiegende Teil der Zeitschriften in der Nähe des äußeren Randes befindet und damit zwar keine reine disziplinäre Klassifikation möglich ist, aber eine eindeutige Zuordnung nach dem Schwerpunktprinzip. Allerdings sind für einige Zeitschriften, wie beispielsweise Nature oder Science", Mehrfachklassifikationen unvermeidlich.

Abbildung 20: "Circle of Science" nach Boyack und Klavans (2011)



Quelle: Boyack und Klavans (2011: 129)

5.1 Existierende Klassifikationen

Weiter oben sind bereits die wichtigsten Ansätze der Verbesserung der Klassifikation aus bibliometrischer Sicht aufgeführt worden. In diesem Abschnitt geht es vor allem darum, die wesentlichen Charakteristika der Klassifikationsschemata von WoS und Scopus darzustellen sowie einige Klassifikationsschemata, die außerhalb der Bibliometrie zur Klassifikation der Wissenschaft verwendet werden.

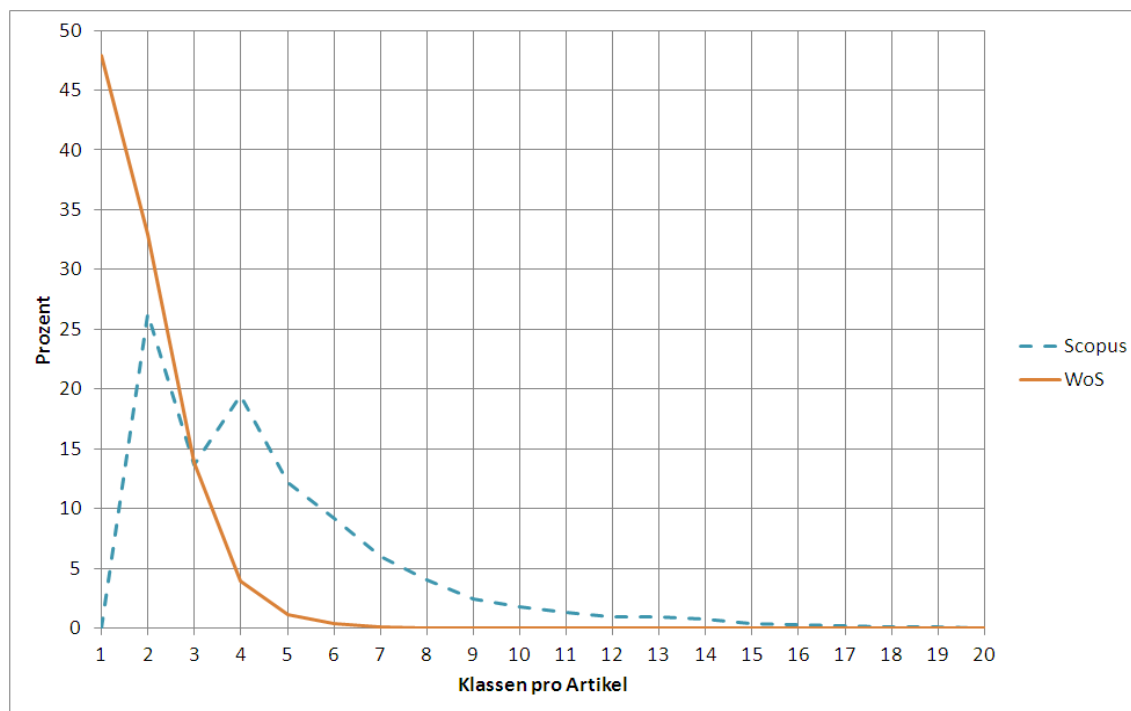
Das in der Bibliometrie bislang am häufigsten diskutierte Klassifikationsschema ist das von WoS, welches vom Institute for Scientific Information (ISI) entwickelt wurde und ständig aktualisiert wird. Die WoS-Klassifikation umfasst aktuell 242 Teilfelder, die alle Bereiche der Wissenschaft abdecken, allerdings mit einer Konzentration auf die Medizin. Alle Teilfelder stehen gleichberechtigt nebeneinander. Die Klassifikation ist also nicht hierarchisch. Es gibt innerhalb von Oberthemen keine Untergliederungen. In der von Thomson Reuters für Inhouse-Datenbanken zur Verfügung gestellten Version hat jedes Feld einen zweistelligen Code, der allerdings keine inhaltlichen Assoziationen zulässt. So entspricht beispielsweise dem Feld "Engineering, mechanical" der Code IU.

Ein grundsätzliches Problem der Klassifikation von WoS besteht darin, dass viele Zeitschriften, und damit in der Folge auch Artikel, mehrfach klassifiziert sind. Dieses ist im Durchschnitt im Zeitraum 2008 bis 2010 1,8 Mal der Fall. Das ist grundsätzlich zu erwarten, da viele Zeitschriften, wie schon von Boyack und Klavans (2011) zutreffend gezeigt, nicht einer einzigen Disziplin zugeordnet werden können. Problematisch ist allerdings, dass einzelne Zeitschriften sehr häufig klassifiziert sind, mit bis zu 9 unterschiedlichen Codes in 2008 bis 2010. Darüber hinaus sind einzelne Zeitschriften, wie sich bei einer intellektuellen Überprüfung ergab, offensichtlich fehklassifiziert.

Als Schlussfolgerung aus der Betrachtung der WoS-Klassifikation sollte eine eigene Klassifikation zumindest innerhalb der einzelnen Felder eine leicht hierarchische Struktur aufweisen. Es sollten also verschiedene Obergebiete mit zugeordneten Teilgebieten existieren. Darüber hinaus sollte die Klassifikation in einzelnen Bereichen etwas feiner sein, da die Klassifikationsstruktur in WoS in vereinzelt Themenfeldern zu grob ist. Anstelle von 242 Klassen wäre eine Größenordnung von 400-500 erstrebenswert. Offensichtlich unkorrekte Klassifikationen sollten korrigiert werden. Bei extrem häufigen Mehrfachklassifikationen sollte nach dem Schwerpunktprinzip die Beschränkung auf einige wenige Codes erfolgen.

Die Klassifikation von Scopus ist mit 343 Teilfeldern etwas feiner als die von WoS. Sie umfasst alle Bereiche der Wissenschaft und hat eine leicht hierarchische Struktur mit 27 Obergebieten. So ist beispielsweise das Obergebiet (Biochemistry, Genetics and Molecular Biology) in 15 Teilfelder wie "Biochemistry", "Biophysics" oder "Biotechnology" aufgegliedert. Jedem dieser Felder ist für Recherchen ein vierstelliger Zahlencode zugeordnet, der diese hierarchische Struktur widerspiegelt. So ist dem Feld "Biochemistry, Genetics..." der Code 1300 zugeordnet, dem Unterfeld "Biotechnology" der Code 1305. Die Code-Struktur ist also insgesamt sinnvoller als die Buchstabenkombination von WoS, könnte aber ebenfalls an einigen Stellen verbessert werden. Es lässt sich relativ leicht eine Konkordanz zwischen den WoS- und Scopus-Codes anlegen, da beide Datenbanken alle Bereiche der Wissenschaft in ähnlicher Weise umfassen. Allerdings gibt es einzelne Felder, bei denen es zu einem WoS-Code keine direkte Entsprechung in Scopus gibt oder umgekehrt. So gibt es beispielsweise für das Feld "Optics" in WoS kein Äquivalent in Scopus, in dem lediglich das aggregierte Feld "Atomic and molecular physics, and optics" existiert. Ein großes Problem von Scopus besteht darin, dass die Mehrfachklassifikation von Artikeln sehr viel häufiger als in WoS genutzt wird, im Schnitt liegt die Klassifikation eines Artikels bei 4,7 mit einem Maximalwert von 68 im Zeitraum 2008 bis 2010. Die höhere Durchschnittsrate der Mehrfachklassifikationen in Scopus im Vergleich zu WoS ergibt sich aus der Verteilung, die in Abbildung 21 dokumentiert ist.

Abbildung 21: Verteilung der Mehrfachklassifikation von Artikeln in den Datenbanken WoS und Scopus, 2008-2010



Quelle: WoS, Scopus, Recherchen und Berechnungen des Fraunhofer ISI.

Darüber hinaus ist die Zahl der fehlklassifizierten Artikel in Scopus erheblich. Dem Hersteller von Scopus ist diese Problematik bewusst, weshalb Elsevier die CWTS-Gruppe mit einer Verbesserung beauftragt hat (Noyons et al. 2010). Bislang wurde diese verbesserte Klassifikation allerdings noch nicht vorgelegt und vor allem noch nicht in dem Schema von Scopus eingeführt.

Obwohl in Scopus die Klassifikation insgesamt feiner strukturiert ist als in WoS, gibt es in vielen Teilbereichen noch Verbesserungsmöglichkeiten im Sinne einer Verfeinerung, wie das Beispiel des Feldes "Optics" zeigt. Auch das legt eine feinere Klassifikation der Größenordnung von etwa 500 Teilstellen nahe.

Eine im deutschen und angelsächsischen Sprachraum kaum bekannte Datenbankklassifikation ist die für die französische Literaturdatenbank PASCAL. Diese ist ähnlich wie WoS oder Scopus multidisziplinär, enthält aber keine Zitate. Die Klassifikation umfasst etwa 6.000 Einzelpunkte auf der untersten Ebene und ist hierarchisch gegliedert. In der Datenbank werden damit nicht Zeitschriften, sondern einzelne Artikel klassifiziert. Bis zum Jahr 2003 wurden die Artikel vollständig einzeln intellektuell klassifiziert, wobei auch Mehrfachklassifikationen vergeben wurden. Wegen des erheblichen Aufwands der Klassifikation von etwas 800.000 Artikeln pro Jahr, die von Mitarbeitern des Instituts INIST (Nancy) durchgeführt wurde, wurde zumindest ein Teil der Klassifikationen

auf ein maschinelles Verfahren umgestellt. In der Konsequenz zeigte sich, dass das maschinelle Verfahren nicht in der Lage ist, eine große Zahl von Artikeln auf der untersten Ebene zu klassifizieren. Im Ergebnis wird somit deutlich, dass eine solche feingliedrige Klassifikation nur dann sinnvoll ist, wenn eine erhebliche intellektuelle Arbeit zur Klassifikation einzelner Artikel geleistet wird. Zur Klassifikation von Zeitschriften ist sie ohnehin zu feingliedrig. Darüber hinaus hat die Klassifikation auch in Einzelpunkten erhebliche inhaltliche Nachteile. So fehlt beispielsweise innerhalb der Biologie das in der heutigen Wissenschaft zentrale Feld der Biotechnologie, die Klassifikation ist offensichtlich deutlich veraltet.

Eine andere im internationalen Bereich häufig genutzte Klassifikation ist "Fields of Science", die von der OECD erarbeitet wurde (OECD 2007). Die Klassifikation ist Teil des Frascati-Manuals. Die Klassifikation umfasst alle Bereiche der Wissenschaft und ist mit 40 Stellen relativ grob. Aufgrund der recht groben Klassifikation hat sie nur eine leichte Hierarchie. Sie bemüht sich um eine Abbildung neuer Gebiete, wobei etwa die Zuordnung der Nanotechnologie zu Ingenieurwesen und Technologie, nicht aber zu Physik oder Chemie etwas problematisch sein dürfte. Zum Bereich Ingenieurwesen und Technologie werden auch die Umweltbiotechnologie und die industrielle Biotechnologie gerechnet, nicht aber zur Biologie. Ein besonders bemerkenswerter Punkt ist die Trennung der Computertechnologie in Computerhardware, die der Elektrotechnik zugeordnet wird, und der Computersoftware, die den Naturwissenschaften und hier der Informatik angegliedert ist.

Die von der OECD international eingeführte Version der Fields of Science darf nicht verwechselt werden mit einer Wissenschaftsklassifikation gleicher Bezeichnung, die lange Zeit im Internet unter gleichem Namen geführt wurde. Inzwischen ist diese Klassifikation, die sehr fein gegliedert ist, in einem Wikipedia-Beitrag unter dem Titel "List of Academic Disciplines" überführt worden⁶. Auch diese Klassifikation umfasst alle Bereiche der Wissenschaft. Eine interessante Abweichung von anderen Klassifikationen ist hier, dass sie die übergeordnete Kategorie "Formale Wissenschaften" eingeführt hat, der die Computerwissenschaft, die Logik, die Mathematik, die Statistik und die Systemwissenschaft untergeordnet sind, die normalerweise den Naturwissenschaften zugerechnet werden.

In Bibliotheken wird in Deutschland, aber auch international, häufig die sogenannte Dewy-Dezimalklassifikation benutzt. Die Dewy-Dezimalklassifikation ist zwar sehr fein gegliedert, hat aber das Problem, dass die Betonung auf den Sozial- und Geisteswis-

⁶ http://en.wikipedia.org/wiki/List_of_academic_disciplines

senschaften liegt, die Technikwissenschaften dagegen nur relativ grob abgebildet werden. Von daher ist sie für die hier intendierten Zwecke weniger geeignet.

Vom kanadischen Unternehmen Science-Metrix, welches im Bereich der Bibliometrie arbeitet, ist eine Wissenschaftsklassifikation vorgeschlagen worden, welche die Klassifikation ausgehend von existierenden Zeitschriftenklassifikationen des Institutes for Scientific Information (ISI), Computer Horizons (CHI) sowie der Klassifikation des Australian Research Council (Evaluation of Research Excellence, ERA). Auf der Basis dieser Klassifikationen wurde mit ontologischem Verfahren eine eigene Klassifikation aufgebaut (Archambault et al. 2011). Die Klassifikation hat eine sehr pragmatische Orientierung mit dem Ziel der Einfachheit und Übersichtlichkeit. Von daher umfasst sie nur insgesamt 176 Teilfelder, und sie ist hierarchisch aufgebaut, d.h. verschiedenen Oberbereichen sind näher aufgeschlüsselte Teilbereiche zugeordnet. Das wesentliche Problem neben der verhältnismäßig groben Struktur ist vor allem, dass alle Zeitschriften mit nur einem Code klassifiziert werden, was angesichts der häufigen Multidisziplinarität von Zeitschriften ohne eindeutigen Schwerpunkt eine recht problematische Entscheidung ist. Das Ziel, die Zahl der Mehrfachklassifikationen zu reduzieren, ist sicherlich zu unterstützen, aber nicht in einer derart radikalen Form. Inhaltlich interessant ist in jedem Fall das von den Autoren eingesetzte ontologische Verfahren zur Klassifikation von Zeitschriften.

Die Klassifikation der amerikanischen National Science Foundation (NSF) umfasst 134 Stellen mit der Betonung auf Medizin. Sie umfasst damit zwar insgesamt alle Bereiche, ist aber relativ grob. Demgegenüber ist die Klassifikation des Australian Bureau of Statistics, die sogenannte ARC-Classification sehr fein und umfasst mehrere hundert Stellen. Die ARC-Klassifikation ist ähnlich die der Datenbank PASCAL für die vorliegenden Zwecke zu fein, kann aber für einzelne Teilbereiche interessante Hinweise zu möglichen Gliederungen geben.

Auch die Joint Academic Coding System-Klassifikation (JACS) der Higher Statistics Agency Großbritanniens ist mit 1.300 Stellen relativ fein. Auch sie ist für den Entwurf einer eigenen Klassifikation durchaus hilfreich.

Eine international wichtige und interessante Klassifikation ist das Common European Research Classification Scheme (CERIF). Diese Klassifikation ist ein Anhang des Common European Research Project Information Format⁷ und umfasst insgesamt 346 Teilfelder. Die Felder sind allerdings relativ grob und könnten oftmals weiter untergliedert werden. Die Klassifikation ist in viele Oberfelder wie Philosophie, Theologie, Ge-

⁷ Daher die Abkürzung

schichte und Kunst, Philologie oder Sozialwissenschaften gegliedert. Allerdings fallen in diese Oberfelder viele Einzelfelder, die genauer strukturiert werden könnten. Von daher ist die Klassifikation nicht wirklich hierarchisch. Ein weiterer Nachteil ist, dass insbesondere die Natur-, Medizin- und Technikwissenschaften unzureichend tief gegliedert sind.

Eine weitere für den deutschen Bereich wichtige Klassifikation ist die DFG-Fachsystematik, genauer *DFG-Systematik der Fächer, Fachkollegien, Fachgebiete und Wissenschaftsbereiche*. Nach dem Stand von Juni 2008 weist diese hierarchisch gegliederte Klassifikation 206 Teilfelder auf. Allerdings zeigen schon die relativ wenigen Teilfelder, dass eine detailliertere Klassifikation sinnvoll wäre. Dieses gilt insbesondere für neuere, moderne Bereiche wie die Biotechnologie und für die Technik- und Ingenieurwissenschaften.

Die DFG-Systematik umfasst die vier großen Bereiche:

- Geistes- und Sozialwissenschaften,
- Lebenswissenschaften,
- Naturwissenschaften und
- Ingenieurwissenschaften.

In dieser Klassifikation wird, anders als bei der List of Academic Disciplines, nicht zwischen Natur- und formalen Wissenschaften unterschieden. Da im deutschen Kontext es eine wahrscheinliche Anwendung der Bibliometrie sein wird, Analysen für die DFG durchzuführen, sollte eine Klassifikation der bibliometrischen Datenbanken die DFG-Systematik zumindest abbilden können.

5.2 Manuelle/intellektuelle Klassifikation ausgewählter Zeitschriften

Angesichts der verschiedenen Vorschläge einer Klassifikation von Publikationen in den Datenbanken WoS und Scopus stellt sich die berechtigte Frage, warum der erneute Versuch einer eigenen Klassifikation unternommen wird. Hierfür sind vor allem die folgenden Gründe maßgeblich:

- Bei Recherchen zu einzelnen Gebieten ist immer wieder aufgefallen, dass einzelne Publikationen in erheblichem Maße in nicht zutreffender Weise klassifiziert wurden. Dieses gilt insbesondere für Scopus, aber in nicht unerheblichem Maße auch für WoS. Durch die intellektuelle Klassifikation sollte herausgefunden werden, ob es sich dabei nur um vernachlässigbare Zufälle, oder aber um ein maßgebliches Problem handelt.

- Viele Klassifikationen für bibliometrische Datenbanken gehen letztlich von Zitatverknüpfungen aus und sind dabei auf die Ergebnisse der automatischen Analysen angewiesen, unabhängig davon, ob sie in dem jeweiligen fachlichen Kontext zu sinnvollen Ergebnissen führen oder nicht. Einer Klassifikation von Zeitschriften sollte aber eine sinnvolle fachliche Systematik zugrundeliegen.
- Die Klassifikationen in den Datenbanken WoS und Scopus sind an vielen Punkten nicht fein genug gegliedert, sodass viele Themen nicht hinreichend präzise beschrieben werden können. Hier sollte eine feinere Untergliederung erreicht werden.
- Schließlich ist davon auszugehen, dass es sich bei der Klassifikation von Fachgebieten und der Klassifikation von Zeitschriften um eine unterschiedliche Herangehensweise handeln muss und nur teilweise Übereinstimmungen gegeben sind. Ziel der intellektuellen/manuellen Klassifikation war es daher, die Besonderheit für eine Klassifikation von Zeitschriften besser zu verstehen.

Das grundsätzliche Interesse bei einer Klassifikation von Zeitschriften besteht darin, soweit wie irgend möglich einer fachlichen/disziplinären Systematik zu folgen. In jedem Fachgebiet wurde eine solche eigene Systematik als Ausgangspunkt genommen. Allerdings zeigt sich in der Realität, dass es zu vielen Teilgebieten keine oder kaum adäquate Zeitschriften gibt. Von daher muss eine eigene Klassifikation ein Stück weit darauf reagieren, welche Zeitschriften zu verschiedenen Bereichen tatsächlich angeboten werden. Von daher gibt es einen Kompromiss zwischen einer sinnvollen fachlichen Systematik und der faktischen Verfügbarkeit von Zeitschriften zu den einzelnen Gebieten. Die Klassifikation entstand daher in einem zweistufigen Verfahren, bei dem im ersten Schritt die Verfügbarkeit von Zeitschriften geprüft wurde und daraufhin eine abschließende Version erstellt wurde.

Da die Autoren dieses Berichts nicht über ausreichende fachliche Kenntnisse zu allen Gebieten der Wissenschaften verfügten, wurden bei Bedarf Spezialisten mit der Klassifikation beauftragt. Das führt zwar zu einer deutlichen fachlichen Verbesserung der Klassifikationssystematik. Es konnte allerdings keine vollständig einheitliche Klassifikations-Philosophie zwischen den verschiedenen Bearbeitern erreicht werden. Insgesamt wurden je nach Gebiet einige Hundert, oder auch einige Tausend Zeitschriften einzeln durchgesehen und klassifiziert. Da sich aus vielen Zeitschriftentiteln kein eindeutiger Hinweis auf den Inhalt ergab, wurde in vielen Fällen auch die Webseite der jeweiligen Zeitschriften im Hinblick auf eine genauere Klassifikation konsultiert. Das Ergebnis dieser Bemühungen ist in Abbildung A4 im Anhang dokumentiert.

Die eigene Klassifikation lehnt sich grundsätzlich an die DFG-Systematik der Fächer an, d.h. auf der obersten Stufe ist die Klassifikation in Geistes- und Sozialwissenschaften, Lebenswissenschaften, Naturwissenschaften sowie Ingenieurwissenschaften eingeteilt. Auch die darunter liegende Struktur, die in der DFG-Systematik als zweistelliger

Code angegeben ist, ist in der Klassifikation weitgehend übernommen worden. Der wesentliche Unterschied besteht darin, dass bei den Ingenieurwissenschaften das Feld Wärmetechnik/Verfahrenstechnik (42) nicht übernommen wurde, da die Energieverfahrenstechnik, die technische Thermodynamik, die Strömungsmechanik sowie die Strömungs- und Kolbenmaschinen nach üblichem Verständnis typische Teilbereiche des Maschinenbaus sind und daher dem Feld Maschinenbau und Produktionstechnik (41) zugeordnet wurden. Auch die Verfahrenstechnik wird in der Regel als Teilgebiet des Maschinenbaus gesehen und ist auch diesem Bereich zugeordnet. Diese Grobstruktur ist in Tabelle 15 dokumentiert. Schon die darunter liegende Ebene der Dreisteller in der DFG-Systematik wurde nur in Einzelfällen übernommen, die Viersteller-Ebene der Felder auf der unteren Ebene wurde völlig neu konzipiert, da sich hier zu viele Widersprüche zwischen einer Systematik der Fächer und einer Systematik der Zeitschriften-Klassifikation ergeben.

Tabelle 15: Grobsystematik der ISI-Klassifikation

1		Geistes- und Sozialwissenschaften
	11	Geisteswissenschaften
	12	Sozial- u. Verhaltenswissenschaften
2		Lebenswissenschaften
	21	Biologie
	22	Medizin
	23	Agrar-, Forstwissenschaften, Gartenbau, Tiermedizin
3		Naturwissenschaften
	31	Chemie
	32	Physik
	33	Mathematik
	34	Geowissenschaften
4		Ingenieurwissenschaften
	41	Maschinenbau u. Produktionstechnik
	43	Materialwissenschaften u. Werkstofftechnik
	44	Elektrotechnik, Informatik u. Systemtechnik
	45	Bauwesen u. Architektur

Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Eine besondere Problematik besteht in der erheblichen inhaltlichen Überschneidung zwischen Elektrotechnik und Informatik, welche vor allen Dingen in den letzten Jahren stark zugenommen hat. Ehemalige Bereiche der Elektrotechnik wie die Signal- oder Bildverarbeitung sind inzwischen selbstverständliche Teilbereiche der Informatik. Auch bei der Computer-Hardware, die häufig der Elektrotechnik zugeordnet wird, ist die Grenze zur Software unscharf. In jedem Fall folgt die eigene Klassifikation der DFG-Empfehlung, die Informatik den Ingenieurwissenschaften zuzuordnen und nicht, wie bei dem OECD-Vorschlag der Field of Science eine Trennung zwischen Hard- und Software vorzunehmen und die Software den Naturwissenschaften, die Hardware den Ingenieurwissenschaften zuzurechnen.

Da in WoS und vor allem in Scopus viele Zeitschriften eindeutig falsch klassifiziert sind, geht es bei der eigenen Klassifikation darum, ein großes Set korrekt klassifizierter Zeitschriften zu ermitteln. Diese können als Referenz für automatisch generierte Klassifikationen herangezogen werden, oder aber als Start-Set für eine maschinelle Klassifikation.

on fungieren, wobei in den noch nicht manuell klassifizierten Zeitschriften nach inhaltlichen Ähnlichkeiten zu bereits klassifizierten gesucht werden kann.

Insgesamt konnten immerhin rund 8.000 Zeitschriften eindeutig klassifiziert werden und können so zu den oben genannten Zwecken eingesetzt werden.

5.3 Grundlegende Probleme der Klassifikation

Ein grundsätzliches Problem bei der Klassifikation von Zeitschriften nach Fachgebieten besteht darin, dass sowohl in WoS als auch in Scopus die Zuordnungen nicht immer korrekt sind. Als Beispiele seien die Gebiete Chemie und Maschinenbau angeführt. In der Chemie wurden die Zeitschriften untersucht, die sowohl in WoS als auch in Scopus enthalten sind. Dabei wurden insgesamt 986 Zeitschriften durchgesehen, wobei eine klare Zuordnung zu Teilgebieten der Chemie in 756 Fällen erreicht werden konnte (Tabelle 16). Das entspricht immerhin einer Quote von 76,7%. Weitere 132 Zeitschriften waren nicht allein mit der Chemie, sondern auch mit anderen Fachgebieten verbunden, weshalb eine eindeutige Verortung innerhalb der Chemie nicht möglich war. Dieses entspricht einem Anteil von 13,4%. Diese breite Interdisziplinarität, also die Abdeckung mehrerer Fachgebiete, ist ein häufiges Phänomen bei Zeitschriften, weshalb der Vorschlag von Science-Metrix, für jede Zeitschrift nur eine Klassifikation zu vergeben, in vielen Fällen irreführend ist. Hier wäre ein Kompromiss denkbar, bei dem maschinell die Verbindung zu verschiedenen Fachgebieten quantitativ ermittelt und dann jeweils die maximal drei größten Klassifikationen herangezogen werden. Das andere Extrem, das teilweise in WoS und noch mehr in Scopus zu finden ist, dass deutlich mehr als fünf Klassen zur Eingruppierung der Zeitschrift vergeben werden, ist ebenfalls nicht hilfreich.

Tabelle 16: Zuordnungsgenauigkeit in der Chemie und dem Maschinenbau nach den Ergebnissen der intellektuellen Klassifikation

Gebiet	Datenbasis	Ausgangsmenge	Klare Zuordnung	Fachübergreifend	Falsch	
Chemie	Schnittmenge WoS/Scopus	986	756	132	98	absolut
Maschinenbau	Scopus	1197	668	102	427	absolut
Chemie	Schnittmenge WoS/Scopus	100,0	76,7	13,4	9,9	%
Maschinenbau	Scopus	100,0	55,8	8,5	35,7	%

Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Ein solcher Bezug zu mehreren Fachgebieten entsteht in der Mehrzahl der Fälle dadurch, dass die Zeitschriften nicht auf eine Disziplin, sondern auf ein Thema orientiert sind. Beispielsweise sind Zeitschriften im Bereich der Kriminologie etwa mit den Themen Recht, Soziologie, Psychologie oder auch Pädagogik verbunden. Statt Klassifikationen für sämtliche dieser Themen, für die es häufig nur wenige Zeitschriften gibt, zu vergeben, erscheint die Mehrfachklassifikation mit maximal drei Stellen als der sinnvollste Weg. Schließlich waren, um beim Beispiel Chemie zu bleiben, 98 Zeitschriften oder 9,9% der Stichprobe falsch zugeordnet, hatten also keine Verbindung zur Chemie. Diese relativ hohe Quote erstaunt, da die Zeitschriften ausgewählt wurden, die sowohl in Scopus als auch in WoS der Chemie zugeordnet werden. Von daher wäre eine größere Treffergenauigkeit zu erwarten gewesen.

Im Falle des Maschinenbaus konnten nur 55,8% der insgesamt 1.197 Dokumente eindeutig zugeordnet werden (Tabelle 16). In diesem Feld war die Fehlerquote mit 35,7% außergewöhnlich hoch. Das mag einerseits mit dem schwer abgrenzbaren Fachgebiet zusammenhängen. Es dürfte aber auch darauf zurückzuführen sein, dass ausschließlich die Datenmenge aus der Datenbank Scopus gewählt wurde, für die eine Verbindung zum Maschinenbau angegeben ist. Diese häufigen Fehlklassifikationen sind offensichtlich charakteristisch für Scopus.

Das Phänomen der thematischen, statt der rein disziplinären Zuordnung ist auch häufig innerhalb der Fachgebiete selbst anzutreffen. So gibt es innerhalb der Medizin die Teilgebiete der Onkologie und der Neurologie, aber auch das spezielle Thema der Neuroonkologie. Entsprechende Zeitschriften, die nur vereinzelt auftauchen, wurden in der Kategorie "Übergreifende Zeitschrift, mehrere Fachgebiete innerhalb Medizin" zugeordnet. Soweit einzelne Themen innerhalb eines Fachgebietes zu einer ausreichenden Menge von Zeitschriften führen, wurden sie auch als eigene Gebiete ausgewiesen, beispielsweise die Getriebetechnik innerhalb des Maschinenbaus.

Ein grundsätzliches Problem der Zeitschriftenklassifikation, die bei der großen Menge der Zeitschriften letztlich maschinell erfolgen muss, besteht darin, dass viele Teilgebiete von mehreren Disziplinen behandelt werden. Ein typisches Beispiel ist die klassische Mechanik, die sowohl ein Teilgebiet der Physik als auch eines des Maschinenbaus ist. Bei einer automatischen Analyse der fachlichen Inhalte ist eine Unterscheidung zwischen der Maschinenbau-Mechanik und der Physik-Mechanik kaum möglich, da die Unterschiede vor allem in der etwas stärkeren theoretischen Ausrichtung in der Physik bestehen, im Prinzip aber dieselben Phänomene behandelt werden. Von daher ist es bei einer Zeitschriftenklassifikation erforderlich, dass eine Entscheidung für eine der Disziplinen getroffen wird, bei der Mechanik beispielsweise zugunsten des Maschinenbaus. In Tabelle 15 ist eine Liste der Fälle von Gebieten, die in mehreren Disziplinen behandelt werden, aufgeführt. Es wird daran deutlich, dass dieses Phänomen sehr

häufig ist und damit einer expliziten Behandlung bedarf. Ein Vorschlag wäre, dass in einer Klassifikation die Gebiete (z.B. Mechanik) jeder übergeordneten Disziplin (im Falle der Mechanik Physik und Maschinenbau) zugewiesen werden, auch wenn dies mehrmals der Fall ist. Zeitschriften, die diesen Gebieten zugeordnet wären, müssten dann zwar nur ein Mal zugeordnet werden (z.B. zur Mechanik), wären aber somit durch Querverweise allen betroffenen Disziplinen (im Falle der Mechanik Physik und Maschinenbau) zugewiesen.

Tabelle 17: Übersicht über Gebiete, die in mehreren Disziplinen bearbeitet werden

Gebiet	Zugeordnete Disziplinen/Fachgebiete
Verhaltensforschung	Psychologie, Biologie
Lebensmittel	Biologie, Agrarwissenschaft, Chemie
Computer	Informatik, Etechnik
Mikrobiologie	Biologie, Medizin
Genetik	Biologie, Medizin
Kriminologie	Recht, Soziologie, Pädagogik
Pharmakologie, Pharmatherapie	Medizin, Pharmazie
Drogenmissbrauch	Medizin, Psychologie, Sozialarbeit
Statistik	Wirtschaftswissenschaften, Mathematik
Thermodynamik	Maschinenbau, Physik
Messtechnik	Maschinenbau, Elektrotechnik, Physik
Akustik, Wärmeübertragung, Klassische Mechanik, Hydraulik, Pneumatik, Strömungsmechanik	Maschinenbau, Physik
Elektronik, Mikroelektronik	Elektrotechnik, Physik
Energie	Maschinenbau, Physik
Physikalische Chemie u. Chemische Physik	Physik, Chemie
Computeranwendungen	Informatik, alle Disziplinen
Bildverarbeitung	Elektrotechnik, Informatik
Medientechnologie	Elektrotechnik, Informatik
Robotik	Maschinenbau, Elektrotechnik, Informatik
Medizintechnik	Medizin, Maschinenbau, Elektrotechnik
Messtechnik	Maschinenbau, Physik, Elektrotechnik

Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Letztendlich wird deutlich, dass trotz der aufgezeigten Probleme durchaus eine sinnvolle Klassifikation der Zeitschriften möglich ist. In den folgenden Abschnitten wird gezeigt, dass auf der Basis von Zitatverknüpfungen auch maschinell Klassen ermittelt werden können, die sich mit den intellektuell erarbeiteten inhaltlich verknüpfen lassen.

Ein grundsätzliches Problem ist allerdings, dass für die Bestimmung der einzelnen Cluster erhebliche Rechenzeiten erforderlich sind, sodass im Rahmen dieses Projekts nur zwei Wissenschaftsgebiete beispielhaft behandelt werden konnten. Grundsätzlich kann die erfolgte manuelle Klassifikation von Zeitschriften genutzt werden, um bei einer Analyse bestimmter Themenbereiche die in Frage kommenden Zeitschriften, ausgehend von dem bereits manuell klassifizierten Set, präzise zu definieren. Bei einer Analyse oder Klassifikation der Zeitschriften sämtlicher Fachgebiete wird es allerdings das Problem geben, dass sich einige Zeitschriften nicht eindeutig zuordnen lassen. Das trifft vor allen Dingen auf thematische Zeitschriften zu, die mehrere Fachgebiete/mehrere Disziplinen betreffen. Hier kann das Ausgangsdaten-Set auf mehrere wahrscheinliche Disziplinen erweitert werden, um das Gewicht der jeweiligen Disziplinen zu bestimmen. Datentechnisch problematisch wird es allerdings, wenn sämtliche Disziplinen zum thematischen Vergleich herangezogen werden müssen. Dieser Ansatz ist möglich, erfordert allerdings erhebliche Rechenzeiten und -kapazitäten. Letztlich wird nach mehrfachen Iterationen eine Restmenge verbleiben, bei der letztlich eine abschließende manuelle Klassifikation erforderlich sein wird.

5.4 Maschinelle Klassifikation von Zeitschriften

Um den Arbeitsaufwand einer manuellen Klassifikation zu vermeiden bzw. zu reduzieren wurde nach der manuellen/intellektuellen Klassifikation eines Startsets ein Verfahren entwickelt, das die Zeitschriften entsprechend ihrer thematischen Relation gruppiert. Das Startset wurde mit Zeitschriften aus dem WoS gebildet, die ebenfalls in Scopus vorhanden sind. Die Berechnungen, auf die im Folgenden eingegangen wird, wurden bisher aber nur in WoS durchgeführt.

Dieses Startset diente im Folgenden nicht nur der Überprüfung der durch das automatische Verfahren erzielten Ergebnisse, sondern stellte auch eine Abgrenzung und Definition der betrachteten Gebiete bereit. Durch eine intellektuelle Überprüfung konnte festgestellt werden, welche Gebiete zu erwarten waren und welche Kategorien der Datenbankhersteller zur Erstellung des Startsets verwendet werden konnten bzw. mussten. Dies ist insbesondere bei Themen nötig, die Kategorien aus verschiedenen Bereichen umfassen können, z.B. weil in der Informatik Verfahren entwickelt werden, die in der Biologie verwendet werden.

Teile des im Folgenden beschriebenen Verfahrens wurden bereits auf dem NBW 2011 in Aalborg präsentiert.

Zur Gruppierung von Zeitschriften wurde ein agglomeratives hierarchisches Clusterverfahren benutzt. Solche Verfahren weisen zu Beginn jeder einzelnen Instanz, in diesem Fall jeder Zeitschrift, ein eigenes, gesondertes Cluster zu. In jedem Iterationsschritt werden zwei Cluster zu einem verschmolzen. Um die Ähnlichkeit zwischen zwei Clustern zu bestimmen, wurde auf eine Analyse der Zitationen zwischen den Zeitschriften zurückgegriffen, die aber über übliche Zitationsanalysen für Clusterverfahren hinausgeht.

$$Sim(c_i, c_j) = \max\{cf(c_i, c_j), cf(c_j, c_i)\} \quad (\text{Formel 1})$$

Formel 1 beschreibt die Berechnung der Ähnlichkeit zwischen zwei Zeitschriften, die auf dem Zitationsfluss (citation flow, cf) beruht. Für zwei Zeitschriften i und j wird der Zitationsfluss von i zu j und von j zu i berechnet. Das Maximum über den Zitationsflüssen entspricht der Ähnlichkeit der beiden Zeitschriften i und j . Die folgende Formel zeigt die Berechnung des Zitationsflusses in eine Richtung (von Zeitschrift i zu Zeitschrift j):

$$Sim(c_i, c_j) = \frac{\sum_{d_i \in c_i} \sum_{d_j \in c_j} cit(d_i, d_j)}{\sum_{d_i \in c_i} \sum_{d \in c_j} cit(d_i, d)} \quad (\text{Formel 2})$$

Die Benutzung einer Maximum-Funktion zur Bestimmung der Ähnlichkeit in Kombination mit der Normalisierung der Zitationen mit der Zeitschriften-spezifischen Zittrate hat einen entscheidenden Vorteil: Kleine Zeitschriften werden nicht diskriminiert. Beispielsweise wird eine kleine Zeitschrift j nur selten von den anderen Zeitschriften in dem Cluster c , zu dem sie thematisch gehört, zitiert. Allerdings zitiert die Zeitschrift j andere Zeitschriften in c aufgrund der thematischen Beziehung öfter als andere Zeitschriften. Durch die Maximum-Funktion wird ausgeblendet, dass die kleine Zeitschrift nur selten von den anderen Zeitschriften zitiert wird, obwohl die häufige Zitation von Zeitschriften in c eine große thematische Verbundenheit suggeriert.

Um während des Cluster-Verfahrens zu bestimmen, ob zwei Cluster mit der größten Ähnlichkeit wirklich zu einem Cluster verschmolzen werden sollen, wird zusätzlich die Intra-Cluster-Ähnlichkeit (intra-clustersimilarity, ics) berechnet:

$$ics(c_i) = \frac{1}{|c_i|} \sum_{d \in c_i} Sim(d, \{c_i \setminus d\}) \quad (\text{Formel 3})$$

Falls die Intra-Cluster-Ähnlichkeit des neuen Clusters, also $ics(c_i \cup c_j)$ größer ist als die Intra-Cluster-Ähnlichkeit beider einzelnen Cluster (s. Formel 4), werden diese zu einem Cluster verschmolzen.

$$\max\{ics(c_i), ics(c_j)\} < ics(c_i \cup c_j) \quad (\text{Formel 4})$$

Zusätzlich zu den Zitierungen wurde noch eine Erweiterung mit einem Keyword-basierten Verfahren getestet. Da mit diesem keine besseren Ergebnisse erzielt werden konnten als mit dem reinen Zitationsverfahren, soll hierauf in diesem Bericht nicht weiter eingegangen werden.

Das Cluster-Verfahren arbeitet nun die folgenden Schritte ab, bis ein Abbruchkriterium (s.u.) erfüllt ist:

1. Berechne für alle Cluster-Paare c_i und c_j die Ähnlichkeit $Sim(c_i, c_j)$
2. Wähle das Cluster-Paar c_i, c_j mit der größten Ähnlichkeit
3. Berechne die Intra-Cluster-Ähnlichkeit für c_i und c_j
 - a. Falls die $ics(c_i \cup c_j)$ die Intra-Cluster-Ähnlichkeit der einzelnen Cluster c_i und c_j übersteigt, erstelle neues Cluster aus c_i und c_j
 - b. Sonst: Wähle Cluster-Paar mit nächstgrößter Ähnlichkeit und gehe zu Schritt 3

Das Cluster-Verfahren stoppt, wenn ein Abbruchkriterium erfüllt ist. In dem hier entwickelten Verfahren kann dies einer der folgenden Fälle sein:

- Alle Instanzen sind in einem Cluster enthalten
- Es gibt keine Cluster mehr, bei denen die Intra-Cluster-Ähnlichkeit durch eine Verschmelzung steigt
- Es gibt keine Ähnlichkeit zwischen zwei Clustern, die einen vorab festgelegten Schwellwert übersteigt
- Ein vorher festgelegtes Limit von Iterationsschritten bzw. Clustern wird erreicht

Die letzten beiden Kriterien sind optional, das heißt, die Grundversion des Algorithmus arbeitet ohne Schwellwert oder Iterationslimit. Diese Kriterien können aber eingesetzt werden, um später Anpassungen an einzelne Datensätze vorzunehmen.

5.4.1 Datensätze

Getestet wurde das Verfahren auf den Gebieten Biologie und Informatik. Hierzu wurden die Publikationen des Jahres 2004 aus dem WoS (WoS) extrahiert. Tabelle 18 zeigt die Kategorien aus dem WoS, die für diese Gebiete benutzt wurden.

Tabelle 18: Definition von Biologie und Informatik über Subject Codes von WoS

Datensatz	WoS-Kategorie-Code	WoS-Kategorie-Name
Biologie	IY	Entomology
	CU	Biology
	QU	Microbiology
	KM	Genetics & Heredity
	CN	Behavioral Sciences
	DE	Plant Sciences
	PI	Marine & Freshwater Biology
	MC	Mathematical & Computational Biology
	DB	Biotechnology & Applied Microbiology
	DR	Cell Biology
	CQ	Biochemistry & Molecular Biology
	CO	Biochemical Research Methods
	WF	Reproductive Biology
	TA	Ornithology
	HT	Evolutionary Biology
	JY	Food Science & Technology
	CT	Cell & Tissue Engineering
	SA	Nutrition & Dietetics
	ZM	Zoology
	HY	Developmental Biology
CX	Biology, Miscellaneous	
DA	Biophysics	
Informatik	EP	Computer Science, Artificial Intelligence
	ER	Computer Science, Cybernetics
	ES	Computer Science, Hardware & Architecture
	ET	Computer Science, Information Systems

Datensatz	WoS-Kategorie-Code	WoS-Kategorie-Name
	EV	Computer Science, Interdisciplinary Applications
	EW	Computer Science, Software Engineering
	EX	Computer Science, Theory & Methods

Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Die Zitierungen dieser Publikationen wurden mit einem 3-Jahresfenster, also für die Jahre 2004 – 2006, berechnet und auf Zeitschriften-Ebene aggregiert. Tabelle 19 zeigt die Eigenschaften dieser beiden Datensätze.

Tabelle 19: Eigenschaften der untersuchten Datensätze in Biologie und Informatik

	Informatik	Biologie
Zeitschriften	264	773
Nicht themenbezogene Zeitschriften	7	120
Manuelle Klassen	31	27

Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Insgesamt ist der Datensatz der Biologie deutlich größer, was lediglich die allgemeine Verteilung der Abdeckung in WoS widerspiegelt. Während der manuellen Klassifikation wurden Zeitschriften aus den in Tabelle 18 genannten Kategorien identifiziert, die nicht den genannten Oberkategorien Informatik und Biologie zugeordnet werden konnten. Hierbei wurden auf diese Weise in der Biologie 15% der Zeitschriften als nicht themenbezogen markiert. Allerdings wurden diese Zeitschriften nicht vom Cluster-Verfahren ausgeschlossen, um ein realistisches Testszenario nachzubilden. Wenn das Verfahren gut funktioniert, sollten solche Zeitschriften allein durch die Tatsache, dass sie zu wenig Verknüpfungen mit themenbezogenen Zeitschriften aufweisen, ausgeschlossen werden und somit auch durch das automatische Verfahren identifiziert werden. In beiden Datensätzen wurden ungefähr gleich viele Klassen in der manuellen Klassifikation erstellt. Das bedeutet, dass der Granularitätslevel der Klassifikationen ungefähr gleich groß ist. Dies widerspricht keineswegs der unterschiedlichen Größe der beiden Datensätze, da diese wie bereits erwähnt lediglich die Datenbank-Abdeckung, aber nicht die Tiefe des thematischen Fokus der beiden Gebiete widerspiegeln.

5.4.2 Ergebnisse

Das oben beschriebene Clusterverfahren wurde auf den beiden Datensätzen Informatik und Biologie angewendet.

Als Verfahren zur automatischen Qualitätsmessung der Cluster, wurden die folgenden Evaluationsmaße berechnet:

- Die Entropy der Cluster, wie nach Tan et al. (2006) definiert,
- die Purity der Cluster,
- der Jaccard-Index und
- die Rand-Statistik.

Diese Maße sind allgemein als Clusterevaluationsmaße anerkannt und alle in Tan et al. (2006) beschrieben. Sie wurden in jedem Iterationsschritt berechnet und den Ergebnissen von vergleichbaren Verfahren gegenübergestellt. Abbildung 22 bis 28 zeigen die Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 29 zeigen die Ergebnisse dieser Evaluationsmaße für die beiden Datensätze. Zusätzlich zu dem in dieser Arbeit neu-entwickelten Verfahren (*cit*), sind die folgenden Verfahren im Vergleich abgebildet, um die Performanz anhand gängiger Verfahren zu messen:

- *bl_cit*: Hierbei werden die absoluten Zitationszahlen zwischen zwei Zeitschriften für das Clustering benutzt und mit der Ähnlichkeit zwischen den zwei Zeitschriften gleichgesetzt. Die Zeitschriften, die sich am häufigsten zitieren, werden also geclustert.
- *bl_cocit*: Hier werden die Ko-Zitationen zweier Zeitschriften benutzt, um ihre Ähnlichkeit zu bestimmen.
- *bl_bcoup*: Einen weiteren Vergleichswert bietet das Clustering nach der bibliographischen Kopplung zweier Zeitschriften. Hierzu wird die Anzahl der Zeitschriften, die von den Zeitschriften A und B gemeinsam zitiert werden, mit der Ähnlichkeit dieser Zeitschriften gleichgesetzt.

Boyack und Klavans (2010) wenden diese drei Verfahren zum Beispiel an, um Artikel zu clustern, wobei nach ihrer Analyse von dieser bibliographische Kopplung am besten funktionierte. Andere Verfahren, die hier in den Vergleich nicht mit einbezogen wurden sind textuelle Verfahren, wie z.B. von Sebastiani (2002), bzw. Mischungen aus Zitations- und Textanalyse wie von He et al. (2009).

Die Entropy ist ein Maß, das angibt, wie sehr die einzelnen Cluster aus reinen Klassen bestehen. Mit Klassen sind in diesem Fall die aus der manuellen Klassifikation erstellten Klassen gemeint, die in dem Testsatz zur Evaluation verwendet werden. Ein geringer Entropy-Wert repräsentiert eine geringe Verteilung der Klassen auf die Cluster und ist somit erstrebenswert. Der Minimal-Wert 0 wird erreicht, wenn die Cluster nur Dokumente enthalten, die jeweils aus derselben Klasse stammen. Wie in Abbildung 22 und

Abbildung 23 dargestellt, ist die Entropy für das hier vorgestellte Verfahren, *cit*, immer am geringsten.

Die Purity für ein einzelnes Cluster bezeichnet die größte Übereinstimmung mit einer der manuellen Klasse. Die Summe der Purity der einzelnen Cluster dividiert durch die Anzahl der Dokumente (Zeitschriften) ergibt den Gesamtwert der Purity für ein Clustering. Auch hier erzielt das neu vorgestellte Verfahren *cit* die besten Ergebnisse unter den vorgestellten Verfahren (Abbildung 24 und Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 25).

Der Jaccard-Index berechnet sich nach der folgenden Formel:

$$J = \frac{f_{11}}{f_{11} + f_{10} + f_{01}} \quad (\text{Formel 5})$$

Wobei sich die einzelnen Werte nach Tabelle 20 bestimmen (vgl. Tan et al. 2006). Es wird hierbei gezählt, bei wie vielen Dokumentenpaaren die Einteilung in die gleichen bzw. in verschiedene Cluster und Klassen vorliegt.

Tabelle 20: Einteilung der Dokumentenpaare nach Klassen- und Clusterzugehörigkeit

	Gleiches Cluster	Verschiedene Cluster
Gleiche Klasse	f_{11}	f_{10}
Verschiedene Klassen	f_{01}	f_{00}

Quelle: Tan et al. (2006).

Analog berechnet sich die Rand-Statistik wie folgt:

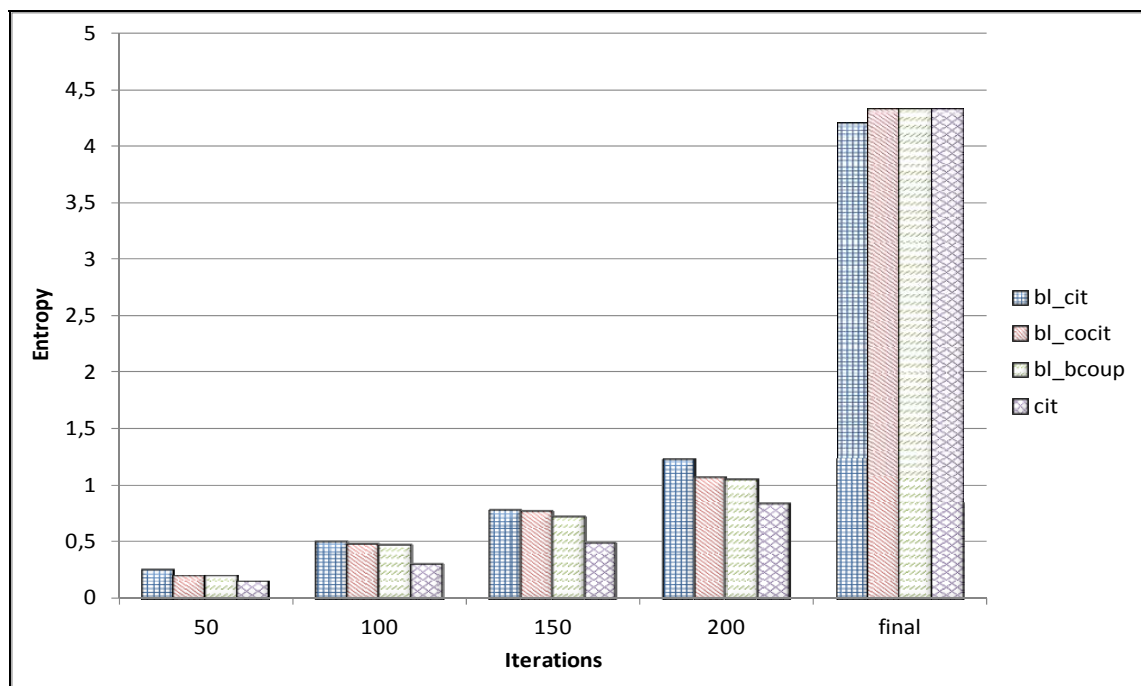
$$R = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (\text{Formel 6})$$

Diese beiden Metriken, Jaccard-Index und Rand-Statistik, bevorzugen im Gegensatz zu den oben genannten Metriken nicht kleinere Cluster, so dass diese weitere gute Indizien zur Performanz des hier vorgestellten Verfahrens liefern. Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 26 zeigt den einzigen Fall, in dem die Vergleichs-Verfahren besser als das neue Verfahren funktionieren. Da die Rand-Statistik für das *cit*-Verfahren im Biologie-Datensatz dennoch die besten Ergebnisse erzielt (Recherchen und Berechnungen des Fraunhofer ISI.

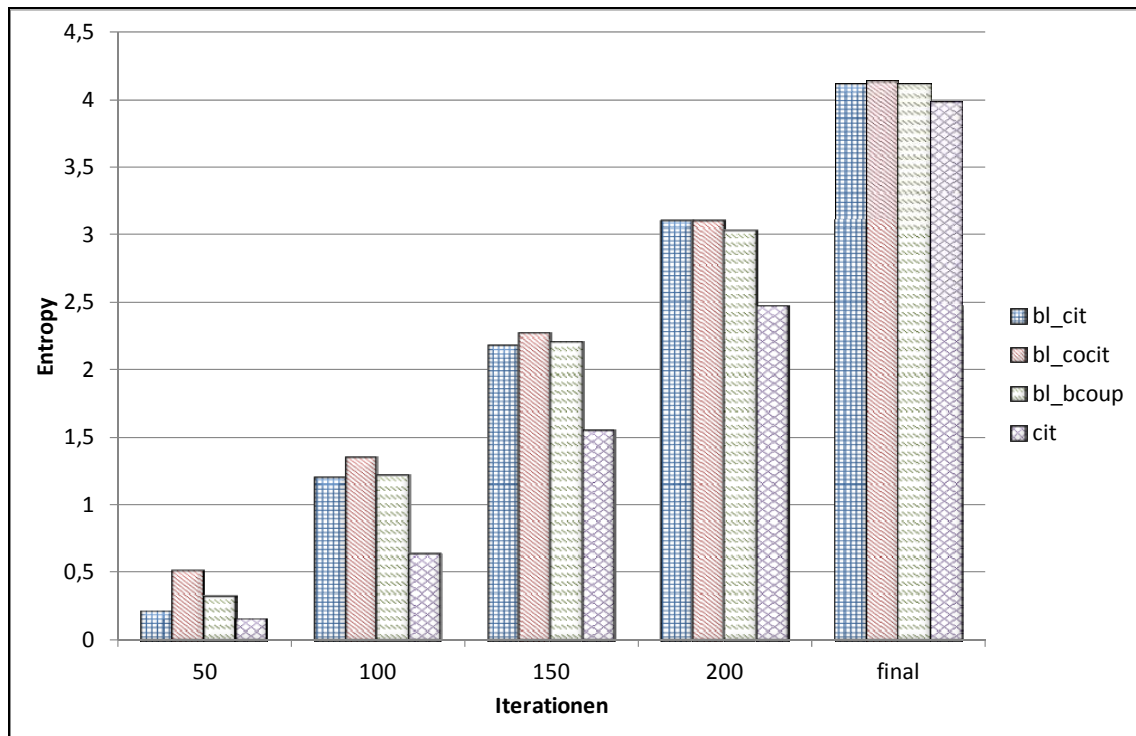
Abbildung 28), scheint das Verfahren hier Probleme mit der Zuordnung von Dokumenten der gleichen Klasse zum gleichen Cluster zu haben, nicht aber mit der Aufteilung von Dokumenten verschiedener Klassen zu verschiedenen Clustern. Es ist zu vermuten, dass das Verfahren aufgrund der Ergebnisse in Purity und Entropy langsamer arbeitet als die vergleichbaren, d.h. mehrere kleinere Cluster bildet, durch die die Dokumente aus gleichen Klassen in getrennten Clustern verbleiben. Die anderen Verfahren hingegen bilden große Cluster, die zwar so relativ viele f_{11} -Dokumente produzieren (da alle Dokumente letztendlich einem Cluster zugeteilt werden), aber eben schlechte Werte in der Purity und insbesondere bei f_{00} erzielen. Ein Problem mit dem Jaccard-Index im Besonderen konnte in der Informatik nicht entdeckt werden, hier ist das *cit*-Verfahren wieder besser als die Vergleichsverfahren (Abbildung 27, Abbildung 29).

Abbildung 22: Entropy für den Datensatz Biologie



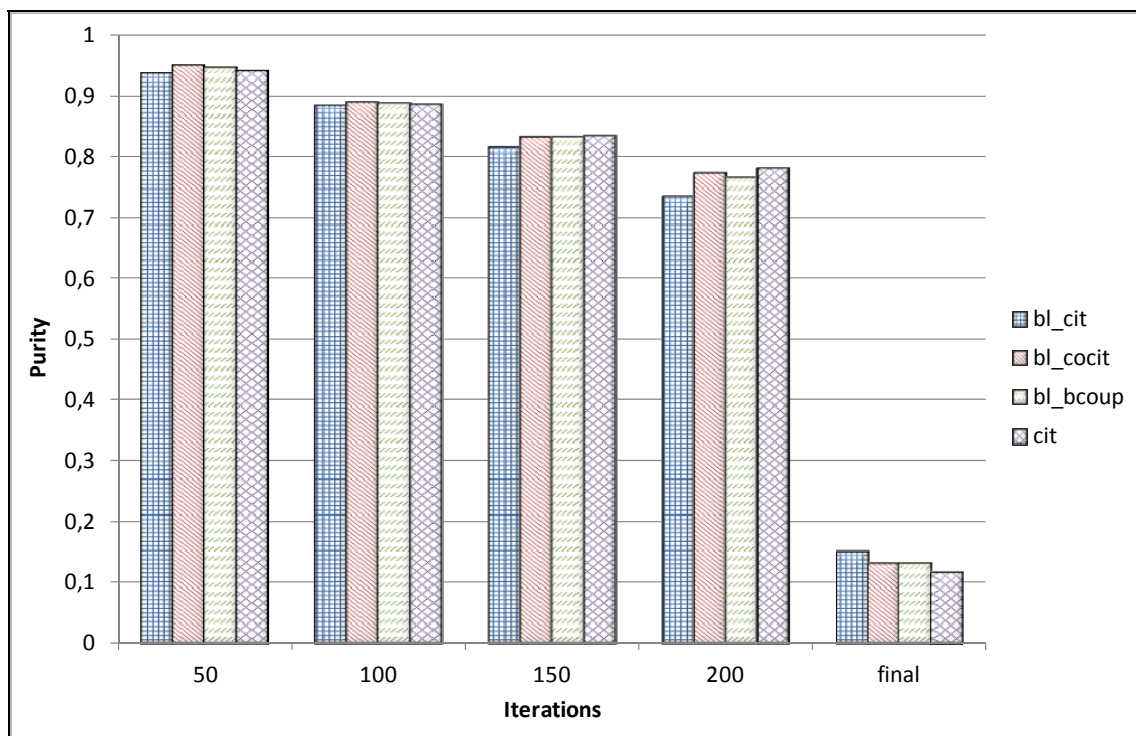
Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 23: Entropy für den Datensatz Informatik



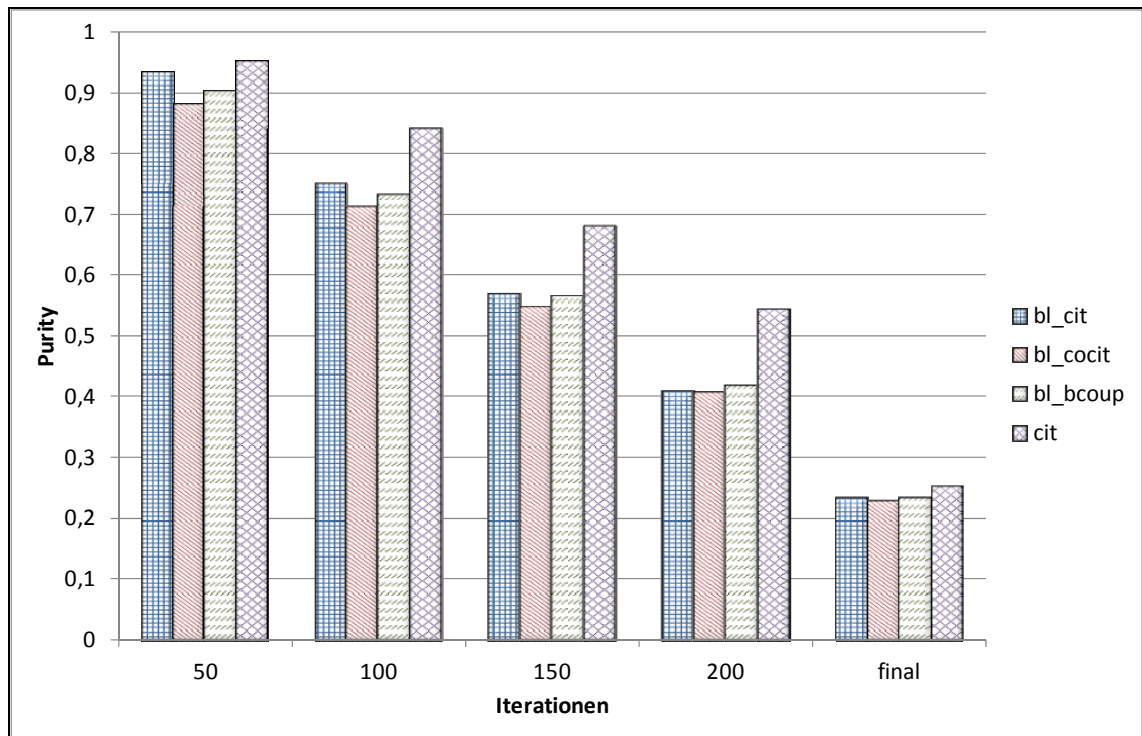
Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 24: Purity für den Datensatz Biologie



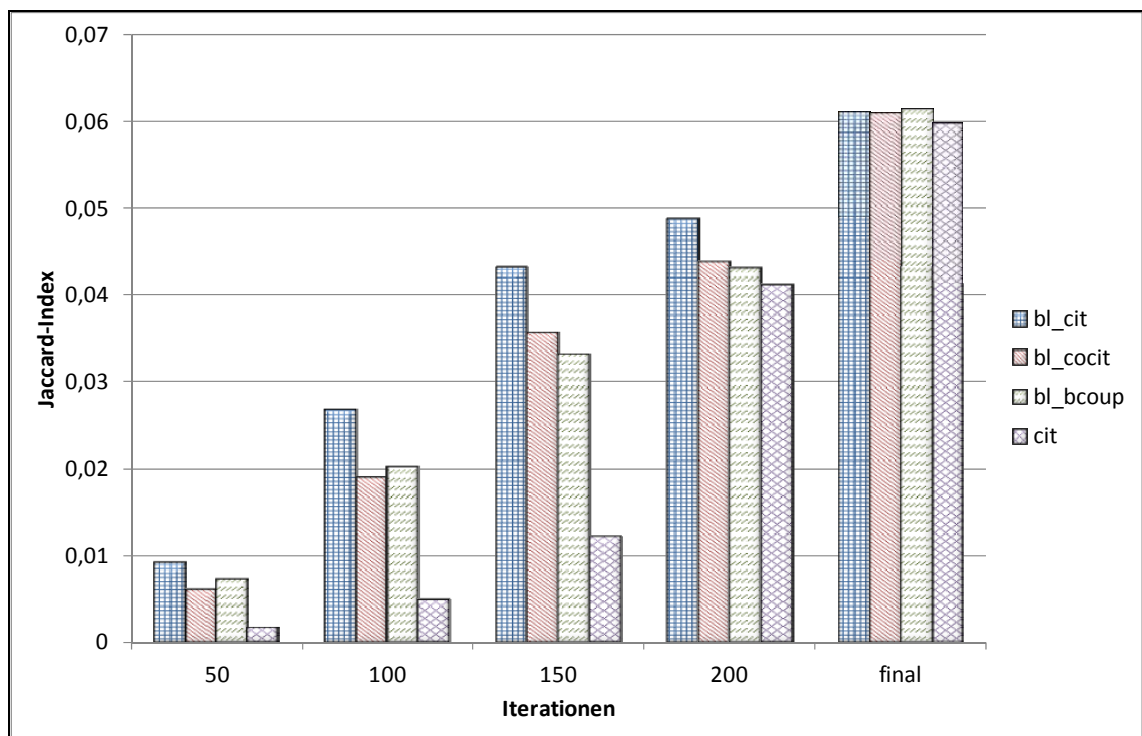
Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 25: Purity für den Datensatz Informatik



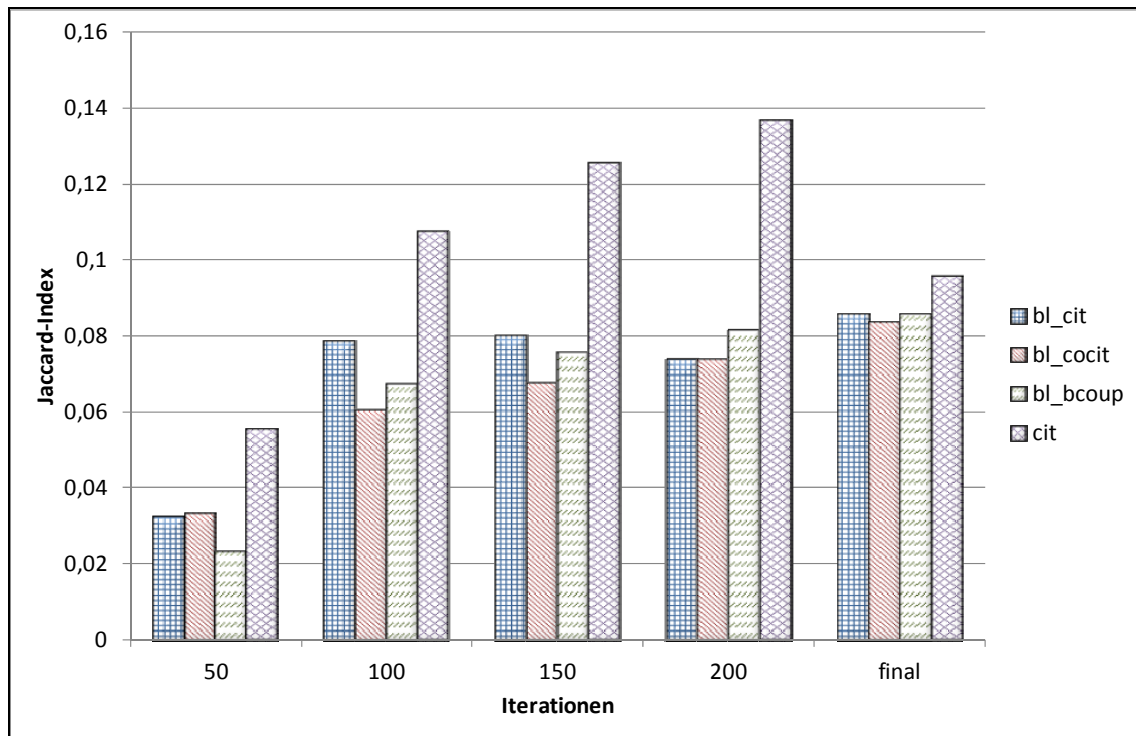
Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 26: Jaccard-Index für den Datensatz Biologie



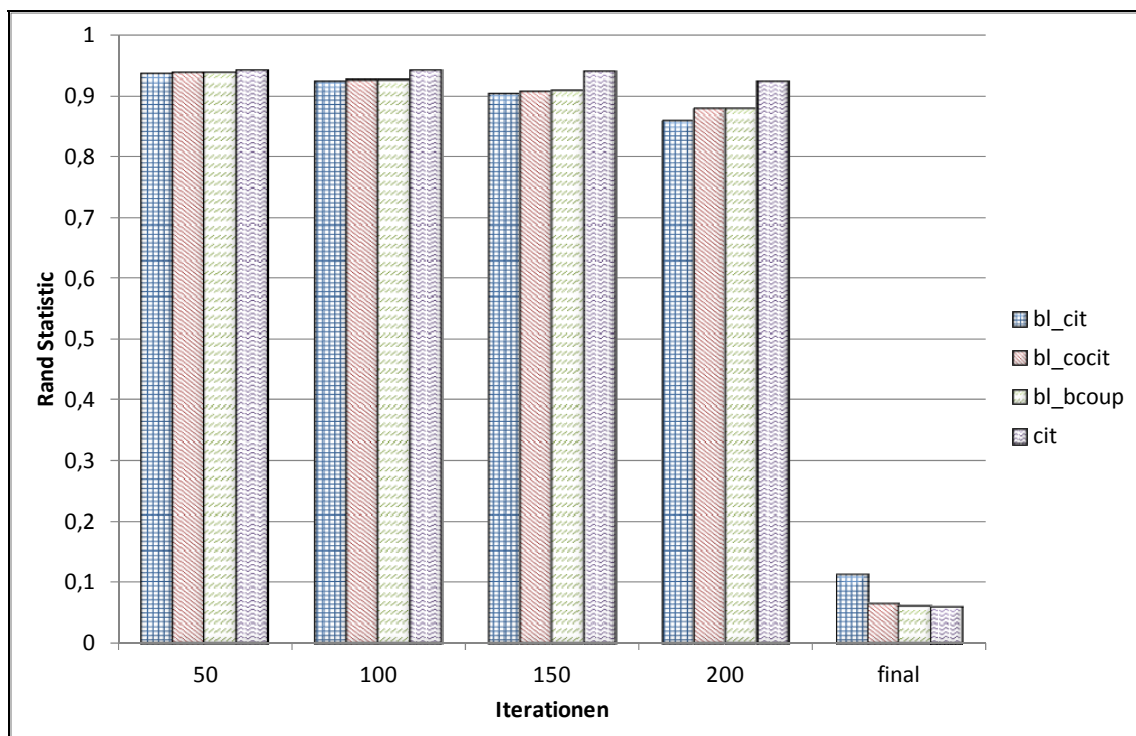
Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 27: Jaccard-Index für den Datensatz Informatik



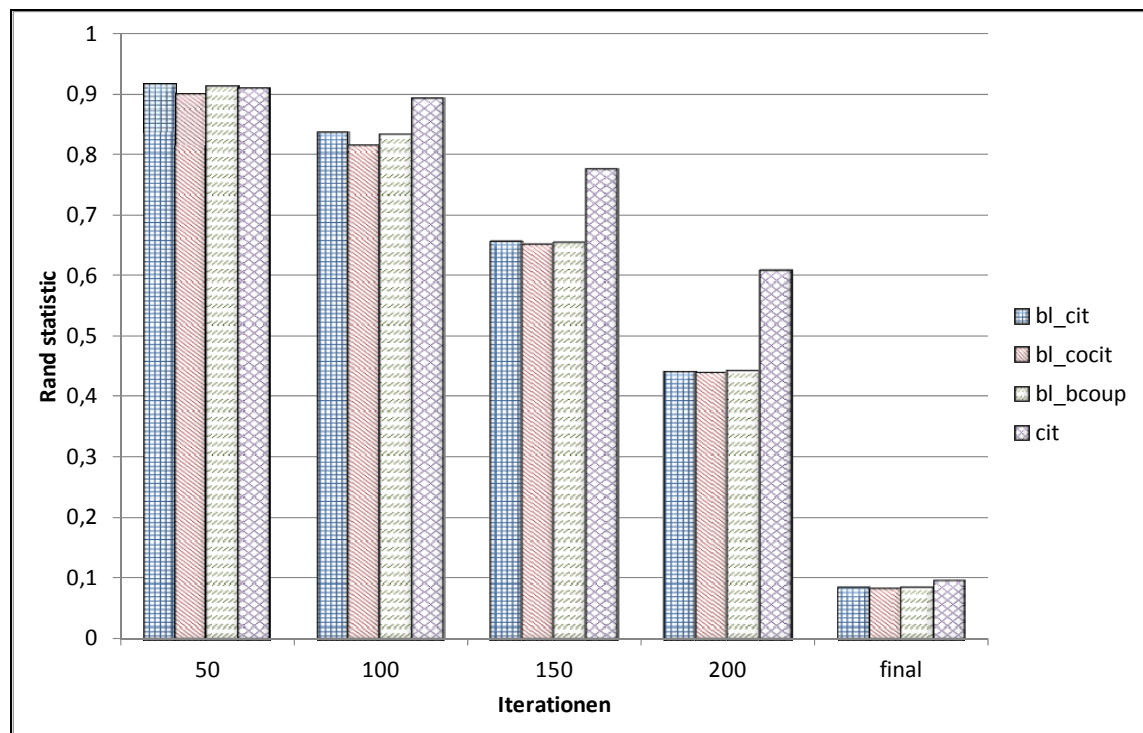
Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 28: Rand-Statistik für den Datensatz Biologie



Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung 29: Rand-Statistik für den Datensatz Informatik



Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Bei der Betrachtung der einzelnen Cluster im Datensatz der Informatik nach 100 Iterationen, lassen sich mit dem *cit*-Verfahren 10 Cluster identifizieren, die zu diesem Zeitpunkt nicht nur aus jeweils einem Dokument bestehen. Im Folgenden werden diese Cluster einzeln betrachtet und diskutiert.

Cluster A: Dieses Clusters besteht aus insgesamt 19 Zeitschriften und enthält vor allem Zeitschriften (15), die in der manuellen Klassifikation zu den Themen Bild- und Musterverarbeitung zugeordnet wurden. Im Vergleich mit der WoS-Klassifikation ist kein direkter Fokus zu erkennen, das heißt, es sind eigentlich alle WoS-Klassen den Zeitschriften zugeteilt und es lässt sich keine Tendenz oder Mehrheit einer oder mehrerer WoS-Klassen erkennen. Beispiele für Zeitschriften in diesem Cluster sind "IEEE Transactions on Image Processing" und "Journal of Visual Communication and Image Representation".

Cluster B: Die acht Zeitschriften aus Cluster B befassen sich laut der manuellen Klassifikation im Speziellen mit Fuzzy Logic und ähnlichem. Allerdings sind auch drei Dokumente enthalten, die in der manuellen Klassifikation als allgemeine Zeitschriften im Bereich Informatik gekennzeichnet wurden. Erneut sind durch die WoS-Klassifikation verschiedene Klassen abgebildet. Im Vergleich mit dieser Klassifikation und mit dem Fokus, den die manuelle Klassifikation impliziert, ist Cluster B aber eindeutig fokussier-

ter und feiner definiert als es die WoS-Klassifikation bieten kann. In diesem Cluster waren z.B. die Zeitschriften "Fuzzy Sets and Systems" und "Soft Computing" enthalten.

Cluster C: Mit der manuellen Klassifikation als Grundlage weist diese Cluster die größte Entropy auf. Das Cluster wird zuerst in Iteration 51 gebildet und enthält zu diesem Zeitpunkt zwei Zeitschriften, die in der manuellen Klassifikation der allgemeinen Informatik zugeordnet sind. Nach und nach werden mehr Zeitschriften aus dieser Kategorie und später auch aus konkreten Klassen hinzugefügt. Der geringe Fokus der Zeitschriften zu Beginn des Clusters scheint hier ein großes Problem darzustellen. Aus diesem Grund zeigen auch die WoS-Klassen kein gemeinsames Thema auf. Der späte Beginn dieses Clusters und der geringe Fokus deuten darauf hin, dass das Cluster-Verfahren eventuell mit 100 Iterationen zu lange läuft. Genauere Untersuchungen müssten gemacht werden, um ein ideales Ende des Verfahrens zu bestimmen. Beispiel für allgemeine Zeitschriften in diesem Cluster sind die Zeitschriften "Computer Journal" und "Journal of Computer and System Sciences".

Cluster D: Dieses Cluster besteht aus acht Zeitschriften, die sowohl in der manuellen als auch in der WoS-Klassifikation jeweils nur einer Klasse zugeteilt sind: "Graphics, visualization" bzw. "EW - Software Engineering". Diese beiden Klassen sind anwendungsbezogen sehr verbunden, so dass diese Verknüpfung sinnvoll erscheint. Auch die Tatsache, dass in WoS keine andere Kategorie in der Informatik vorhanden ist, die sich mit Grafiken beschäftigt, spricht dafür, dass diese Kategorien Sinn machen. Mitglieder dieses Clusters sind z.B. die Zeitschriften "Computer Aided Geometric Design", und "Computer Animation and Virtual Worlds".

Cluster E: Ähnlich wie bei Cluster D ist der thematische Fokus von Cluster E sehr eng und anhand der manuellen und der WoS-Klassifikation abzulesen; alle fünf Dokumente sind in beiden Klassifikationen der Kategorie "Artificial Intelligence" zugeordnet. Im Gegensatz zu Cluster C ist die längere Laufzeit des Cluster-Verfahrens, d.h. mehr als 50 Iteration, für dieses Cluster vorteilhaft, da zwei Zeitschriften erst später hinzugefügt werden (Iterationen 53 und 89). Enthaltene Zeitschriften sind beispielsweise die Zeitschriften "AI Magazine" und "Journal of Artificial Intelligence Research".

Cluster F: Dieses Cluster besteht aus zehn Zeitschriften, die in der manuellen Klassifikation den Klassen "Software Engineering & Programming" und "General Computer Science" zugeteilt wurden. Die Verteilung der Zeitschriften über die WoS-Klassen ist erneut gemischt, was auch Folge der Mehrfachklassifikation in WoS ist. Allerdings wurde – neben anderen Klassen – die Klasse "EW - Software Engineering" allen Zeitschriften in Cluster F zugewiesen. Somit scheint auch das Thema dieses Clusters prägnant herausgearbeitet. Beispiele für Zeitschriften in diesem Cluster sind "Internation-

tional Journal of Software Engineering and Knowledge Engineering" und "ACM Transactions on Software Engineering and Methodology".

Cluster G: Cluster G deckt die manuellen Klassen "General Computer Science" und "Other Applications" in gleichen Teilen ab (jeweils zwei Zeitschriften). In den WoS-Klassen ist immer "ES - Hardware & Architecture" in Kombination mit anderen Klassen vertreten. Dieses Cluster enthielt u.a. die Zeitschriften "IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems" und "ACM Transactions on Design Automation of Electronic Systems".

Cluster H: Dieses Cluster enthält ebenfalls eine Mischung aus manuellen und WoS-Klassen, so dass auf dieser Betrachtungsebene kein Fokus erkennbar ist. Die vier Zeitschriften sind den manuellen Klassen "Human-computer Interaction", "Software Engineering and Programming", "Graphics, visualization" und "Other Informatics" zugeordnet. Es fällt allerdings auf, dass drei Zeitschriften von der ACM (Association for Computing Machinery) herausgegeben werden, so dass eine hohe Zitatrate unter den Zeitschriften trotz einer geringen thematischen Verknüpfung plausibel erscheint.

Cluster I: Cluster I besteht aus lediglich zwei Zeitschriften ("IEEE Micro", "Computer"), die beide in der manuellen Klassifikation der allgemeinen Informatik-Klasse zugeordnet sind. In WoS sind beide Zeitschriften den beiden Klassen "ES - Hardware & Architecture" und "EW - Software Engineering" zugeordnet.

Cluster J: Mit Cluster J wurde wieder ein kleineres, aber sehr klar abgegrenztes Cluster erstellt. Die zwei enthaltenen Zeitschriften sind der manuellen Klasse "Bioinformatics and other biological applications" zugeteilt und haben beide einen klaren thematischen Fokus auf Bildverarbeitung in der Medizin, wie an den Titel erkennbar ist: "IEEE Transactions on Medical Imaging" und "Medical Image Analysis". Die WoS-Klassen weisen ebenfalls in diese Richtung, sind aber wegen der allgemeineren Formulierung und der deutlich höheren Aggregationsebene viel ungenauer als das Cluster und die manuelle Klassifikation. In WoS wurde die Klasse "EV - Interdisciplinary Applications" alleine und in Kombination mit "EP - Artificial Intelligence" zugewiesen. Zwar wird der Vorteil durch die thematische Abgrenzung dieser Zeitschriften deutlich, es müsste aber mit einem größeren Testset und längerer Laufzeit geprüft werden, ob noch weitere Dokumente, die mit diesem Thema verbunden sind, identifiziert werden könnten.

Insgesamt zeigt sich, dass das automatische Verfahren zwar nur wenige größere Cluster bildet, diese aber zum Großteil einen klaren Fokus besitzen, der sich ebenfalls durch thematischen Tiefgang auszeichnen. Das Verfahren ist somit durchaus geeignet, um thematische Klassen automatisch für größere Datensätze zu bilden. Durch eine

vorausgehende manuelle Klassifikation eines Teildatensatzes können die manuellen Klassen der Dokumente auf die Cluster übertragen werden, die diese enthalten.

Angedacht, aber aus Zeitgründen leider nicht realisierbar, war die Verwendung der manuellen Klassifikation als Startset, auf der aufbauend das Clusterverfahren weitere ähnliche Dokumente findet. Am Anfang wären somit nicht alle Dokumente in einzelnen Clustern, sondern einige Zeitschriften bereits in Clustern, die den manuellen Klassen entsprechen, zusammengefasst.

6 Maschinelle Klassifikation einzelner Artikel

Es wurde in Zusammenarbeit mit dem Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB) am Karlsruher Institut für Technologie (KIT) ein Verfahren zur automatischen Klassifikation von Artikeln entwickelt. Hierzu wurden die Klassen aus der manuellen Klassifikation aus der Informatik verwendet, um für einige Artikel eine eindeutige Klassenzuordnung vorzugeben. Mit Hilfe dieser Daten wurde ein vom AIFB entwickelter Klassifikator trainiert.

Der Klassifikator wurde sowohl mit etablierten Evaluationsmaßen (Recall und Precision) als auch manuell evaluiert. Sowohl Recall als auch Precision lagen im Durchschnitt über alle Klassen bei mehr als 60%, wobei einige Klassen (z.B. "Graphics, visualization" und "Natural Language Processing") besonders gut von den anderen Klassen abgegrenzt werden konnten.

Zur besseren Evaluation der Artikel-Klassifikation des entwickelten Systems wurde eine manuelle Klassifikation eines Teils der Artikel durch drei Experten am AIFB durchgeführt. Diese sollte unabhängig von der Klassifikation der Zeitschriften sein und somit eine allgemeine Beurteilung der Geeignetheit des Klassifikators ermöglichen. Bei der Erstellung der Artikel-Klassifikation stießen die Experten auf das Problem, dass eine einheitliche Zuordnung der Artikel in das vorgegebene Klassifikationsschema der Zeitschriften selbst nicht möglich war. Zwar konnten 76% der Artikel von den einzelnen Experten eindeutig einer Klasse zugewiesen werden, aber bei einer Überprüfung der Übereinstimmung mit der Artikel-Klassifikation auf Basis der Klasse der zugehörigen Zeitschrift waren nur 56,9% der 700 betrachteten Artikel richtig zugeordnet. Auf Basis der Zeitschriften-Klassifikation war der Klassifikator aber trainiert worden, um eine ausreichende Menge an Trainingsdaten zur Verfügung zu haben.

Die drei Experten waren sich auch bei nur 36 von 100 Artikeln über die Klassenzugehörigkeit einig. Dementsprechend konnte auch nur eine relativ geringe Übereinstimmung der Experten mit den Ergebnissen des Klassifikators erwartet werden. Der Klassifikator hatte eine gleiche Klassenzuordnung wie einer der Experten in 49%, aber nur in 15% stimmten alle drei Experten und der Klassifikator überein.

Insgesamt zeichneten sich folgende Probleme ab:

- Die Klasse "Computer System Architecture" wird oft mit "Software Engineering and Programming" und "Distributed and Grid Systems" vertauscht.
- Es scheint eine Überschneidung zwischen den Klassen "Mathematics and Statistics" und "Image processing" zu geben. Das könnte daran liegen, dass viele Methoden aus der Mathematik in der Bildverarbeitung eingesetzt werden und die Klassen daher überlappen.

- Besonders schlechte Ergebnisse werden für die Klassen "Computer System Architecture" und "Other Artificial Intelligence" erreicht.

Besonders gute Ergebnisse wurden in den Klassen "Fuzzy Logic", "Other Signal Processing", allen Anwendungsgebieten, "Natural Language Processing", "Other Computer Networks" und "Graphics, visualization" erzielt.

Ausführliche Ergebnisse zu dem Teilprojekt finden sich in dem separaten Projektbericht "Klassifizierung von Artikeln und Proceedings in Publikationsdatenbanken".

7 Implementierung der Ergebnisse in die KB-Bibliometriedatenbank

Die mit diesem Bericht vorgelegten Ergebnisse können wie folgt in die Bibliometriedatenbank des KB implementiert werden:

- Die Algorithmen zur Berechnung der Indikatoren (SNIP, Audience-Faktor, Impact-Faktor) von Zeitschriften sowie der zeitschriften- und feldspezifischen erwarteten Zitatraten wurden bereits an das FIZ Karlsruhe zur Implementierung übergeben.
- Die Ergebnisse der manuellen Klassifikation der Zeitschriften können in Form einer Tabelle aufbereitet werden. Diese Tabelle enthält dann für jede ID einer klassifizierten Zeitschrift (die sogenannte Source-ID) mindestens einen entsprechenden Eintrag zur zugewiesenen Klasse. Über die Source_ID könnte diese Information dann in die bestehende Datenbank des KB integriert und mit anderen Tabellen verknüpft werden.
- Die Ergebnisse der automatischen Zeitschriften- und Artikelklassifikation können ebenfalls durch die Integration der Ergebnistabelle(n) des Klassifikationssystems in die Datenbank integriert werden. Dies könnte analog zur Integration der manuellen Klassifikation umgesetzt werden.

8 Literatur

- Aksnes, D.W. (2008): When different persons have an identical author name. How frequent are homonyms?, *Journal of the American Society for Information Science and Technology*, 59, 838-841.
- Archambault, E./Beauchesne, O.H./Roberge, G. (2011): Towards a Multilingual, Comprehensive and Open Scientific Journal Ontology, 13th Conference of the International Society for Scientometrics and Informetrics (ISSI). Durban, South Africa: University of Zululand.
- Bassecoulard, E./Zitt, M. (1999): Indicators in a research institute: A multi-level classification of journals, *Scientometrics*, 44, 323-345.
- Bonaccorsi, A./Brandt, T./De Filippo, D./Lepori/B./Molinari, F./Niederl, A./Ulrich Schmoch, U./Schubert, T./Slipersaeter, S. (2010): *Feasibility Study for Creating a European University Data Collection. Final Study Report*. Brussels: European Commission.
- Boyack, K.W./Klavans, R./Börner, K. (2005): Mapping the Backbone of Science, *Scientometrics*, 64, 351-374.
- Boyack, K.W./Klavans, R. (2010): Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?, *Journal of the American Society for Information Science and Technology*, 61, 2389-2404.
- Boyack, K.W./Klavans, R. (2011): Multiple dimensions of journal specificity: Why journals can't be assigned to disciplines, Proceedings of the 13th Conference of the International Society for Scientometrics and Informetrics (ISSI), 123-133. Durban, South Africa: University of Zululand.
- Carpenter, M.P./Narin, F. (1973): Clustering of scientific journals, *Journal of the American Society for Information Science*, 24, 425-436.
- Costas, R./Leeuwen, T.N./Bordons, M. (2010): Self-citations at the meso and individual levels: effects of different calculation methods, *Scientometrics*, 82, 517-537.
- De Solla Price, D.J. (1971): *Little Science, Big Science*. Frankfurt/m.: Suhrkamp.
- Glänzel, W./Moed, H.F. (2002): Journal impact measures in bibliometric research, *Scientometrics*, 53, 171-193.

- Glänzel, W./Thijs, B./Schlemmer, B. (2004): A bibliometric approach to the role of author self-citations in scientific communication, *Scientometrics*, 59, 63-77.
- Glänzel, W./Schubert, A. (2003): A new classification scheme of science fields and subfields designed for scientometric evaluation purposes, *Scientometrics*, 56, 357-367.
- Gläser, J./Spurling, T.H./Butler, L. (2004): Intraorganisational evaluation: Are there 'least evaluable units'?, *Research Evaluation*, 13, 19-32.
- Grupp, H./Schmoch, U./Hinze, S. (2001): International alignment and scientific regard as macro-indicators for international comparisons of publications, *Scientometrics*, 51, 359-380.
- Gupta, B.M./Dhawan, S.M. (2008): Condensed matter physics: An analysis of India's research output, 1993-2001, *Scientometrics*, 75, 123-144.
- He, Q./Bi, C./Pei, J./Qiu, B.P.M./Giles, C.L. (2009): Detecting topic evolution in scientific literature: how can citations help?, Proceedings of the 8th ACM Conference on Information and Knowledge Management (CIKM 2009). Hong Kong, China: CIKM 2009.
- Hinze, S./Tang, L./Gauch, S. (2008): *Leistungsfähigkeit und Strukturen der Wissenschaft im internationalen Vergleich 2007. Studien zum deutschen Innovationssystem*, Expertenkommission für Forschung und Innovation (Hrsg.), Studien zum deutschen Innovationssystem. Berlin: Expertenkommission für Forschung und Innovation.
- Hirsch, J.E. (2005): An index to quantify an individual's scientific research output, *Proceedings of the National Academy of Sciences of the United States of America*, 102, 16569-16572.
- Katz, J.S./Hicks, D. (1995): The classification of interdisciplinary journals: A new approach, 5th International Conference of the International Society for Scientometrics and Informetrics. River Forest, IL USA: Rosary College.
- Larivière, V./Gingras, Y. (2011): Averages of ratios vs. ratios of averages: An empirical analysis of four levels of aggregation, *Journal of Informetrics*, 5, 392-399.
- Leydesdorff, L./Cozzens, S./van den Besselaar, P. (1994): Tracking areas of strategic importance using scientometric journal mappings, *Research Policy*, 23, 217-229.

- Michels, C./Schmoch, U. (2011): Growth of Science and Database Coverage (Full paper), Proceedings of the 13th Conference of the International Society for Scientometrics and Informetrics (ISSI), 4.-7. Juli 2011. Durban (ZA).
- Moed, H.F. (2010): Measuring contextual citation impact of scientific journals, *Journal of Informetrics*, 4, 265-277.
- Moed, H.F./Van Leeuwen, T.N. (1995): Improving the Accuracy of the Institute for Scientific Information's Journal Impact Factors, *Journal of the American Society for Information Science*, 46, 461-467.
- Noyons, E.C.M./Waltmann, L./Kähler, O./Van Eck, N.J. (2010): Developing a comprehensive Scopus journal classification scheme using bibliometric data, 11th International Conference on Science and Technology Indicators, 9-11 September 2010. Leiden: Universiteit Leiden.
- OECD (2007): Revised Field of Science and Technology (FoS) Classification in the Frascati Manual, Document JT03222603, Paris: OECD.
- Ophhof, T./Leydesdorff, L. (2012): Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance, *Journal of Informetrics*, 4, 423-430.
- Science-Metrix (2010): Launch of Multi-Lingual Tools to Explore and Measure Science. Online: <http://www.science-metrix.com/eng/news.htm> (abgerufen am: 07.01.2011).
- Sebastiani, F. (2002): Machine Learning in Automated Text Categorization, *ACM Computing Surveys*, 34, 147.
- Stehr, N. (1994): *Knowledge Societies*. London, Thousand Oaks, New Delhi: SAGE Publications.
- Tan, P.-N./Steinbach, M./Kumar, V. (2006): *Introduction to Data Mining*. Boston: Pearson Education.
- van Raan, A.F.J./Van Leeuwen, T.N./Visser, M.S./Van Eck, N.J./Waltman, L. (2010): Rivals for the crown: Reply to Ophhof and Leydesdorff, *Journal of Informetrics*, 4, 431-435.

- Van Raan, A.J.F. (2004): Measuring Science. Capita Selecta of Current Main Issues. In: Glänzel, W./Moed, H./Schmoch, U. (Hrsg.): *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies on R&D Systems*. Dordrecht; Norwell; New York; London: Kluwer Academic Publishers, 19-50.
- Zhang, L./Thijs, B./Glänzel, W. (2011): The diffusion of H-related literature, *Journal of Informetrics*, 5, 583-593.
- Zitt, M./Small, H. (2008): Modifying the Journal Impact Factor by Fractional Citation-Weighting: The Audience Factor, *Journal of the American Society for Information and Technology*, 59, 1856-1860.

Anhang

Abbildung A1: Referenzabfragen für die bibliometrische Version von WoS

Aufgabe	Abfragen WOS	Ergebnis	Zeit (hh:mm:ss.ms)
1 Deutsche Publikationen im Jahr 2003 (Ohne Proceedings)	<pre>SELECT count(distinct p.ut_item_id) FROM published p WHERE p.ut_item_id IN (SELECT pc.ut_item_id FROM productcode pc WHERE pc.pv_product_code not in ('IH', 'IT', 'W')) AND p.ut_item_id IN (SELECT ut_item_id FROM article_countries WHERE country_code='DE') AND p.py_publication_year = 2003;</pre>	85789	Abgelaufen: 00:00:30.07
2 Publikationen der alten EU-Länder (EU15) im Jahr 2005	<pre>SELECT count(DISTINCT tmp.UT_ITEM_ID) AS pub_cnt FROM (SELECT s.SQ_SEQUENCE_NUMBER, p.PY_publication_year, ac.country_code, a.DT_document_type, a.UT_ITEM_ID FROM article a JOIN published p ON a.UT_ITEM_ID=p.UT_ITEM_ID JOIN source s ON p.SQ_SEQUENCE_NUMBER=s.SQ_ SEQUENCE_NUMBER JOIN article_countries ac on a.ut_item_id=ac.ut_item_id WHERE a.ut_item_id in (select pc.ut_item_id from productcode pc where pc.pv_product_code not in ('IH', 'IT', 'W')) AND p.PY_publication_year = 2005 and ac.country_code IN ('AT','BE','DE','DK','ES', 'FI','FR','GB','GR','IE', 'IT', 'LU','NL','SE', 'PT') GROUP BY s.SQ_SEQUENCE_NUMBER, p.PY_publication_year, ac.country_code, a.DT_document_type, a.UT_ITEM_ID) tmp;</pre>	425962	00:00:55.04

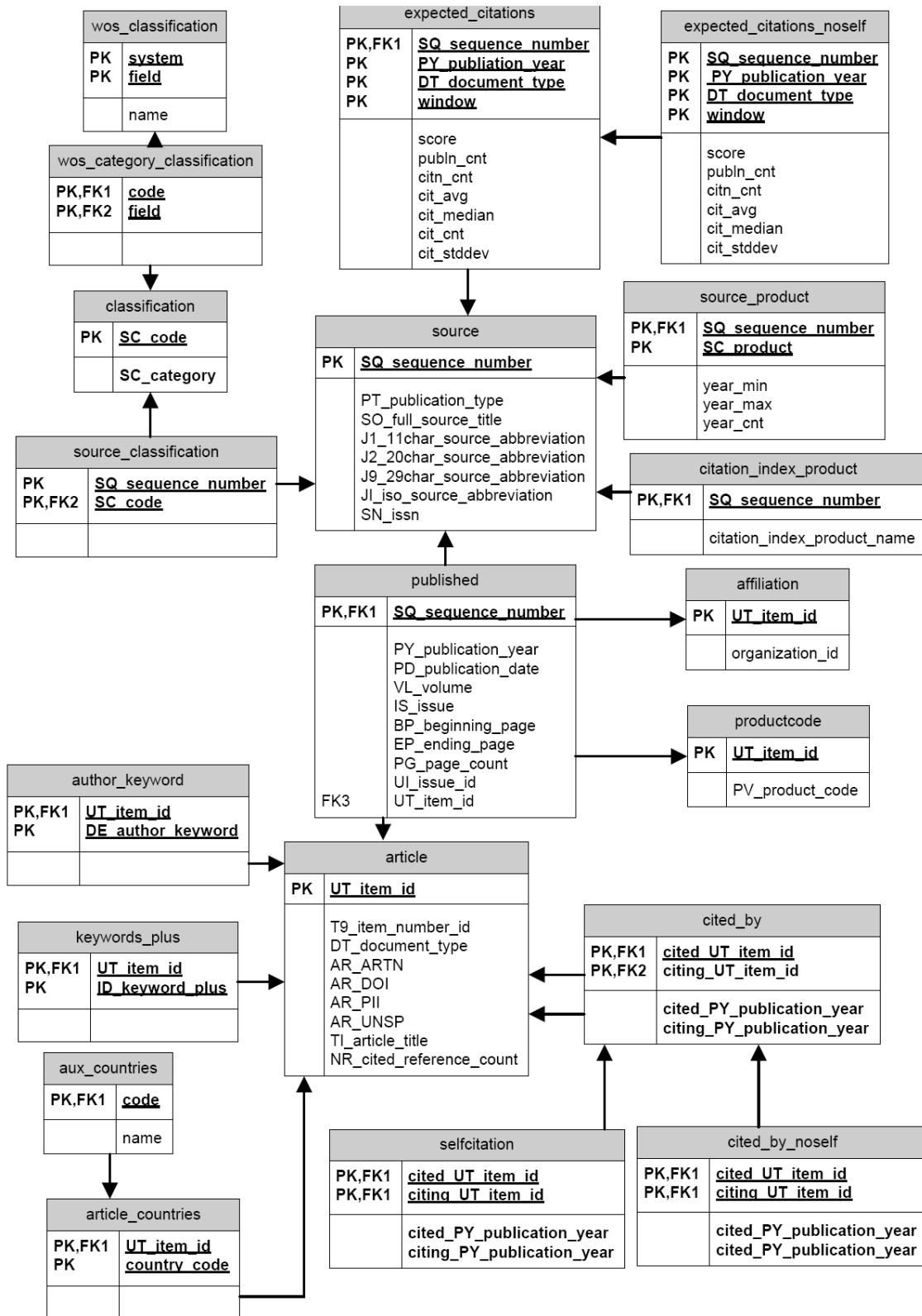
<p>3 Zitate zu Publikationen nach Befehl 2) im Drei- Jahres_Fenster (2005- 2007)</p>	<pre>SELECT count(DISTINCT UT_ITEM_ID) AS pub_cnt, SUM(cit_cnt) AS cit_cnt FROM (SELECT s.SQ_SEQUENCE_NUMBER, p.PY_publication_year, ac.country_code, a.DT_document_type, a.UT_ITEM_ID, COUNT(DISTINCT c.CITING_UT_ITEM_ID) as cit_cnt FROM article a JOIN published p ON a.UT_ITEM_ID=p.UT_ITEM_ID JOIN source s ON p.SQ_SEQUENCE_NUMBER=s.SQ_ SEQUENCE_NUMBER JOIN article_countries ac on a.ut_item_id=ac.ut_item_id LEFT JOIN cited_by c ON a.UT_ITEM_ID=c.CITED_UT_ITEM_ID AND c.CITING_PY_PUBLICATION_YEAR < p.PY_publication_year+3 WHERE a.ut_item_id in (select pc.ut_item_id from productcode pc where pc.pv_product_code not in ('IH', 'IT', 'W')) AND p.PY_publication_year = 2005 and ac.country_code IN ('AT','BE','DE','DK','ES', 'FI','FR','GB','GR','IE', 'IT', 'LU','NL','SE', 'PT') GROUP BY s.SQ_SEQUENCE_NUMBER, p.PY_publication_year, ac.country_code, a.DT_document_type, a.UT_ITEM_ID) tmp;</pre>	<p>PUB_CNT: 425962 CIT_CNT: 2452760</p>	<p>00:03:30.98</p>																				
<p>4 Zeitreihe 2000 bis 2008 der Artikel (ohne Proceedings) mit dem Stichwort "nano" mit rechtsseitiger offener Trunkierung</p>	<pre>SELECT count(distinct p.ut_item_id), p.py_publication_year FROM published p JOIN article a ON a.ut_item_id=p.ut_item_id WHERE p.ut_item_id IN (SELECT pc.ut_item_id FROM productcode pc WHERE pc.pv_product_code not in ('IH', 'IT', 'W')) AND p.py_publication_year BETWEEN 2000 AND 2008 AND LOWER(a.TI_article_title) like 'nano%' GROUP BY p.py_publication_year;</pre>	<table border="1"> <thead> <tr> <th>PY_PUBLICA</th> <th>COUNT(DIST</th> </tr> </thead> <tbody> <tr> <td>2005</td> <td>1761</td> </tr> <tr> <td>2006</td> <td>1967</td> </tr> <tr> <td>2001</td> <td>828</td> </tr> <tr> <td>2007</td> <td>1997</td> </tr> <tr> <td>2000</td> <td>593</td> </tr> <tr> <td>2003</td> <td>1291</td> </tr> <tr> <td>2004</td> <td>1524</td> </tr> <tr> <td>2002</td> <td>960</td> </tr> <tr> <td>2008</td> <td>2195</td> </tr> </tbody> </table>	PY_PUBLICA	COUNT(DIST	2005	1761	2006	1967	2001	828	2007	1997	2000	593	2003	1291	2004	1524	2002	960	2008	2195	<p>00:02:25.65</p>
PY_PUBLICA	COUNT(DIST																						
2005	1761																						
2006	1967																						
2001	828																						
2007	1997																						
2000	593																						
2003	1291																						
2004	1524																						
2002	960																						
2008	2195																						

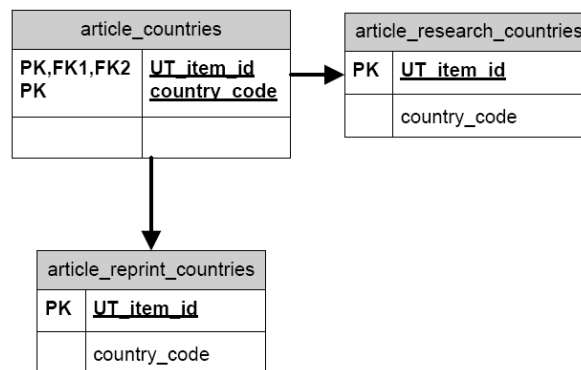
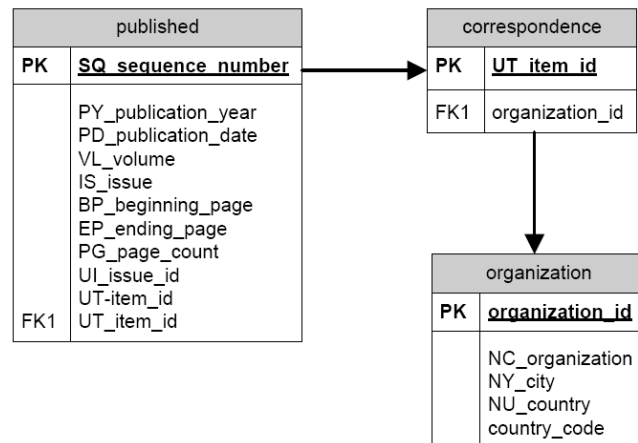
<p>5 Proceedings der TU Braunschweig im Themenbereich "Computertechnik" im Zeitraum 2003-2007</p>	<pre> SELECT COUNT(DISTINCT pc.ut_item_id) FROM productcode pc JOIN affiliation a ON a.ut_item_id=pc.ut_item_id JOIN (SELECT organization_id FROM organization WHERE UPPER(nc_organization) LIKE '%TU BRAUNSCHWEIG%') o ON o.organization_id=a.organization_id WHERE pc.pv_product_code IN ('IH', 'IT') AND pc.ut_item_id IN (SELECT DISTINCT p.ut_item_id FROM published p JOIN source_classification s ON s.sq_sequence_number = p.sq_sequence_number WHERE py_publication_year BETWEEN 2003 AND 2007 AND s.sc_code IN ('EP', 'ER','ES', 'ET', 'EV', 'EW', 'EX')); </pre>	9	00:00:29.95
<p>6 Durchschnittliche Seitenzahl der Artikel (ohne Proceedings) in der Physik für das Jahr 2006 mit einer Band/Volumen-Nr von 10 und mehr</p>	<pre> SELECT AVG(pg_page_count) FROM published p WHERE p.sq_sequence_number in (select sq_sequence_number from source_classification where sc_code in (select code from wos_classification c join wos_category_classification cc on c.field=cc.field where upper(name) in ('OPTICS', 'NUCLEAR TECHNOLOGY', 'PHYSICS') and system='ISI_EFI2010')) and p.vl_volume not in ('1', '2', '3', '4', '5', '6', '7', '8', '9') and p.py_publication_year=2006 and p.ut_item_id in (select pc.ut_item_id from productcode pc where pc.pv_product_code not in ('IH', 'IT', 'W')) group by p.py_publication_year order by p.py_publication_year asc; </pre>	AVG(PG_PAGE_C OUNT): 7,98261275	00:00:41.69

7 Zitatquote der Artikel (Ohne Proceedings) in der Chemie des Jahres 2004 (5 Jahres-Fenster 2004-2008) differenziert nach Zeitschriften mit us-amerikanischer, englischer und deutscher Herausgeberschaft	SELECT SUM(tmp.cit_cnt) AS CIT, tmp.country_code	U	U	00:00:51.54
	FROM (SELECT a.DT_document_type, ac.country_code, p.UT_ITEM_ID, COUNT(DISTINCT c.CITING_UT_ITEM_ID) as cit_cnt	87180	US	
	FROM published p JOIN article a ON a.ut_item_id=p.ut_item_id	18742	GB	
	JOIN article_countries ac on p.ut_item_id=ac.ut_item_id LEFT JOIN cited_by c ON a.UT_ITEM_ID=c.CITED_UT_ITEM_ID AND c.CITING_PY_PUBLICATION_YEAR < p.PY_publication_year+5 WHERE p.py_publication_year=2004 AND a.DT_document_type IN ('@','L','R','N') AND ac.country_code IN ('DE', 'GB', 'US') AND a.ut_item_id IN (SELECT ut_item_id FROM productcode pc WHERE pc.pv_product_code NOT IN ('IH', 'IT')) AND p.sq_sequence_number IN (SELECT sq_sequence_number FROM source_classification WHERE sc_code IN ('EP', 'ER','ES', 'ET', 'EV', 'EW', 'EX')) GROUP BY ac.country_code, a.DT_document_type, p.UT_ITEM_ID) tmp GROUP BY tmp.country_code;	15015	DE	

Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

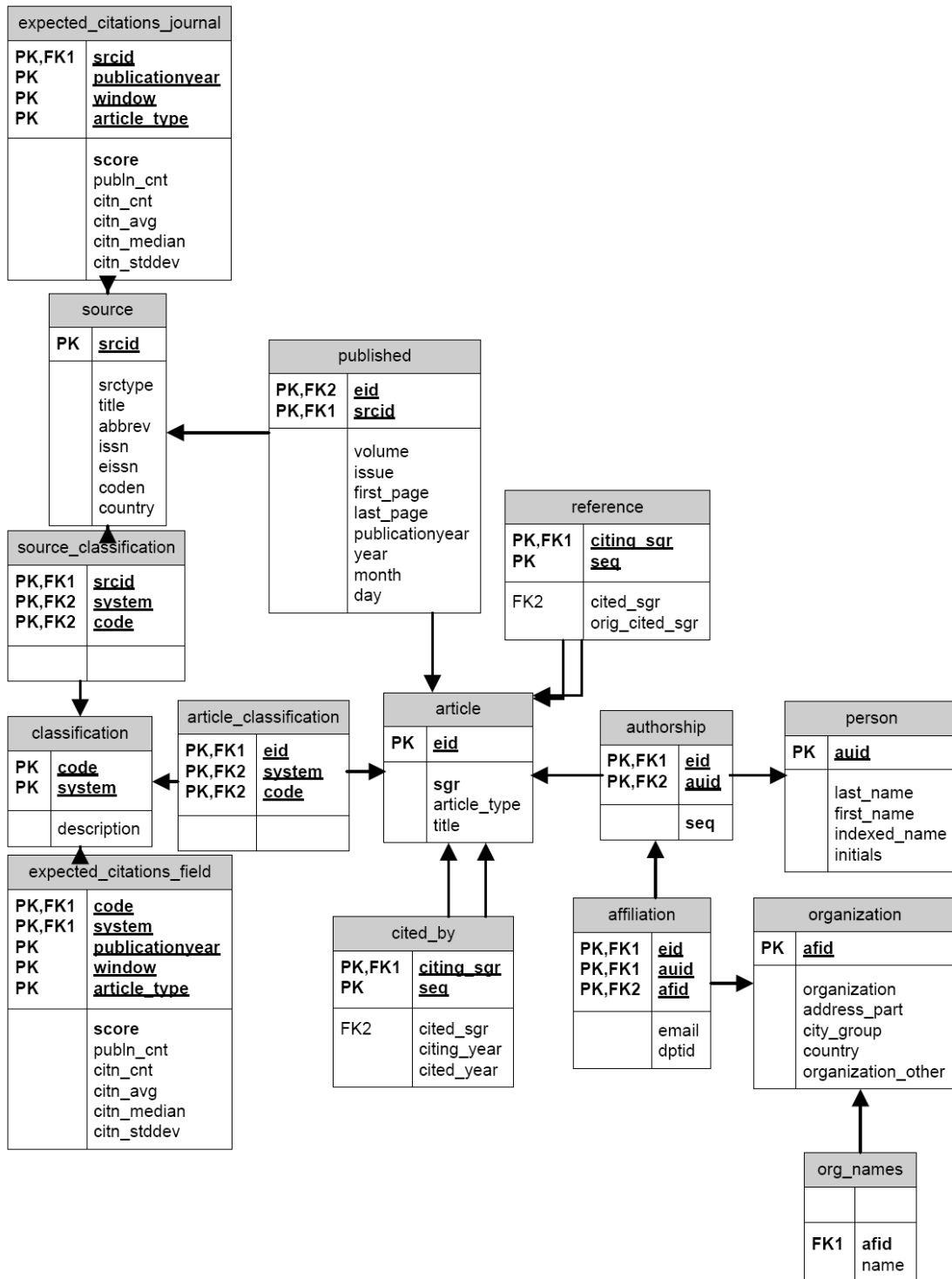
Abbildung A2: Datenbankschema von WoS





Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung A3: Datenbankschema von Scopus



Quelle: Recherchen und Berechnungen des Fraunhofer ISI.

Abbildung A4: Eigene Klassifikation von wissenschaftlichen Zeitschriften nach Fachgebieten und -themen

1				Geistes- und Sozialwissenschaften	N
1	1			Geisteswissenschaften	1426
1	1	1		Alte Kulturen	
1	1	2		Geschichtswissenschaften	
1	1	2	1	Altertum, Vorgeschichte	38
1	1	2	2	Mittelalter, Mediävistik	25
1	1	2	3	Frühe Neuzeit, Aufklärung	29
1	1	2	4	Neuzeit, neuste Geschichte	21
1	1	2	5	Geschichte spezieller Länder und Regionen	73
1	1	2	6	Sonstige Geschichte	156
1	1	3		Kunst-, Musik- Theater- u. Medienwissenschaften	
1	1	3	1	Bildende Kunst	17
1	1	3	2	Alte Musik	18
1	1	3	3	Neuzeitliche Musik	73
1	1	3	4	Tanz	18
1	1	3	5	Theater	43
1	1	3	6	Film	30
1	1	3	7	Sonstige darstellende Kunst (Oper, Medien ...)	22
1	1	3	8	Sonstige Kunst	83
1	1	4		Sprachwissenschaften	
1	1	4	1	Europäische Sprachen	44
1	1	4	2	Sprachen, Linguistik allgemein	171
1	1	5		Literaturwissenschaften	
1	1	5	1	Alte Literatur	35
1	1	5	2	Neuzeitliche Literatur	152
1	1	6		Außereuropäische Sprachen und Kulturen	
1	1	6	1	Außereuropäische Sprachen	9
1	1	6	2	Volkskunde, Ethnologie	30
1	1	6	3	Sonstige Kulturwissenschaft	40
1	1	7		Theologie	108
1	1	8		Philosophie	191

1	2			Sozial- u. Verhaltenswissenschaften	1644
1	2	1		Erziehungswissenschaften	
1	2	1	1	Sportwissenschaften	25
1	2	1	2	Historische Bildungsforschung	4
1	2	1	3	Allgemeine Erziehungswissenschaft	53
1	2	1	4	International und Interkulturell Vergleichende Erziehungswissenschaft	72
1	2	1	5	Empirische Bildungsforschung	22
1	2	1	6	Grundschulpädagogik	10
1	2	1	7	Sekundarstufenpädagogik	46
1	2	1	8	Sonderpädagogik	18
1	2	1	9	Berufs- und Wirtschaftspädagogik	73
1	2	1	10	Sozialpädagogik	66
1	2	1	11	Erwachsenenbildung	6
1	2	1	12	Pädagogische Freizeitforschung und Sportpädagogik	11
1	2	1	13	Frauen- und Geschlechterforschung in der Erziehungswissenschaft	7
1	2	1	14	Medienpädagogik	29
1	2	1	15	Differentielle Erziehungs- und Bildungsforschung	6
1	2	1	16	Pädagogik der frühen Kindheit	26
1	2	1	17	Pädagogische Psychologie	22
1	2	1	18	Sonstige Pädagogik	6
1	2	2		Psychologie	
1	2	2	1	Klinische Psychologie	129
1	2	2	2	Psychotherapie	88
1	2	2	3	Entwicklungspsychologie	68
1	2	2	4	Allgemeine Psychologie	51
1	2	2	5	Biologische Psychologie und Neuropsychologie	36
1	2	2	6	Sozialpsychologie	28
1	2	2	7	Forschungsmethoden und experimentelle Psychologie	27
1	2	2	8	Pädagogische Psychologie	24

1	2	2	9		Angewandte Psychologie: sonstige Subdisziplinen	16
1	2	2	10		Sexualwissenschaft inkl. LGBT	16
1	2	2	11		Arbeits-, (Betriebs-) und Organisationspsychologie	15
1	2	2	12		Differentielle Psychologie und Persönlichkeitspsychologie	13
1	2	2	13		Familienpsychologie, sonstige systemische Familienstudien	10
1	2	2	14		Rechtspsychologie inkl. Forensische Psychologie	8
1	2	2	15		Gesundheitspsychologie	7
1	2	2	16		Counselling Psychology	6
1	2	2	17		Evolutionary Psychology	3
1	2	2	18		verschiedene Psychologische Subdisziplinen	84
1	2	2	19		interdisziplinär, Psychol. + Nicht-Psychologie	79
1	2	3		Soziologie		
1	2	3	1		Soziologische Theorie	15
1	2	3	2		Methoden und Statistik	15
1	2	3	3		Geschlecht, Sexualität und Körper	31
1	2	3	4		Raum-, Stadt-, Regionalforschung	14
1	2	3	5		Migration und ethnische Beziehungen	19
1	2	3	6		Demographie	19
1	2	3	7		Wissenschafts- und Technikforschung	9
1	2	3	8		Medizin- und Gesundheitssoziologie	9
1	2	3	9		Bildung und Erziehung	6
1	2	3	10		Andere spezielle Soziologien	26
1	2	3	11		Soziologie der Kindheit/Jugendsoziologie	6
1	2	3	12		übergreifende Zeitschrift, mehrere Fachgebiete innerhalb Soziologie	46
1	2	3	13		interdisziplinär, mehrere Fachgebiete auch Nicht-Soziologie	143
1	2	3	14		Sonstige	14

1	2	4		Politikwissenschaft		
1	2	4	1		Politische Philosophie, Theorie und Ideengeschichte	22
1	2	4	2		Politische Systeme, Regierungslehre	22
1	2	4	3		Politikfeldforschung	51
1	2	4	4		Vergleichende Politikwissenschaft	11
1	2	4	5		Friedens- und Konfliktforschung	19
1	2	4	6		Öffentliche Meinung, Massenmedien	10
1	2	4	7		Internationale Beziehungen	40
1	2	4	8		Öffentliche Verwaltung	33
1	2	4	9		Sonstige Politikwissenschaft	11
1	2	4	10		übergreifende Zeitschrift, mehrere Fachgebiete innerhalb Politikwissenschaft	44
1	2	4	11		interdisziplinär, mehrere Fachgebiete + Nicht-Politikwissenschaft	76
1	2	4	12		Regionalstudien	23
1	2	4	13		Entwicklungspolitik	16
1	2	5		Publizistik		
1	2	5	1		- Kommunikations-, Medien-, Öffentlichkeitstheorien	27
1	2	5	2		- Spezielle Medienforschung	15
1	2	5	3		- Bibliothekswesen	94
1	2	5	4		- Sonstige Kommunikationswissenschaften	34
1	2	6		Wirtschaftswissenschaften		
1	2	6	1		Volkswirtschaft	
1	2	6	2		Makroökonomie	20
1	2	6	3		Mikroökonomie	18
1	2	6	4		Geld- und Finanzmärkte	45
1	2	6	5		Ökonometrie, Statistik	73
1	2	6	6		Agrar, Umwelt- und Ressourcenökonomik	26
1	2	6	7		Arbeit und Demografie	19
1	2	6	8		Finanzwissenschaft	17
1	2	6	9		Entwicklungsökonomie	19

1	2	6	10		Außenwirtschaft und internationale Wirtschaftsbeziehungen	14
1	2	6	11		Sonstige VWL-Gebiete	58
1	2	6	12		Übergreifende VWL-Zeitschriften	199
1	2	6	13		Betriebswirtschaft	
1	2	6	14		Unternehmensführung und -planung	153
1	2	6	15		Finance, Investition, Taxation and Accounting	73
1	2	6	16		Wirtschaftsinformatik	15
1	2	6	17		Produktionswirtschaft, Fertigungssysteme und Fertigungstechnik, Produktionsplanung	19
1	2	6	18		Marketing, Wirtschaftskommunikation, Unternehmenskommunikation	78
1	2	6	19		Operation Research / Unternehmensforschung	34
1	2	6	20		Logistik and Transport - Materialwirtschaft/ Beschaffung	29
1	2	6	21		Transportation (Railway etc.)	42
1	2	6	22		Personalwirtschaft	28
1	2	6	23		Services - Dienstleistungen - im DL-Sektor und im Industriesektor	53
1	2	6	24		übergreifende BWL Zeitschriften (mehrere BWL-Themen vereint)	57
1	2	6	25		Interdisziplinäre Zeitschriften in BWL	31
1	2	6	26		Sonstige BWL-Gebiete	16
1	2	6	27		Übergreifende VWL-BWL-Zeitschriften	55
1	2	7		Rechtswissenschaften		
1	2	7	1		Rechtstheorie, -geschichte, -philosophie	10
1	2	7	2		Privatrecht	0
1	2	7	3		Wirtschaftsrecht, Wettbewerbs- u. UrhR, IPR	26
1	2	7	4		Medizinrecht	12
1	2	7	5		IT-Recht	6
1	2	7	6		Öffentliches Recht	2

1	2	7	7		Europarecht/ Völkerrecht/ Menschenrechte	9
1	2	7	8		Umweltrecht	19
1	2	7	9		Strafrecht	30
1	2	7	10		Kriminologie	37
1	2	7	11		Sonstige Rechtsgebiete	17
1	2	7	12		Prozessrecht / Gerichte	0
1	2	7	13		Int. Rechtsbeziehungen/ Rechtsvergleich	19
1	2	7	14		Recht einzelner Gebietskörperschaften/ Gebiete/ Länder	12
1	2	7	15		Rechts-Ausbildungszeitschriften	52
1	2	7	16		Gesetzgebung, Verwaltung	7
2				Lebenswissenschaften		2719
2	1			Biologie		
2	1	1	1		Forschungsmethoden, Analytik	25
2	1	1	2		Biochemie	113
2	1	1	3		Biophysik	22
2	1	1	4		Zellbiologie	76
2	1	1	5		Strukturbiologie	14
2	1	1	6		Allgemeine Genetik	41
2	1	1	7		Molekulare Genetik, Genomics und Gentechnik	54
2	1	1	8		Molekular-Biologie	64
2	1	1	9		Entwicklungs- und Reproduktionsbiologie	31
2	1	1	10		Bioinformatik	12
2	1	1	11		Botanik	32
2	1	1	12		Angewandte Botanik	21
2	1	1	13		Zoologie	18
2	1	1	14		Angewandte Zoologie	9
2	1	1	15		Verhaltensforschung	2
2	1	1	16		Ökologie und Ökosystemforschung	10
2	1	1	17		Evolution, Biodiversität, Taxonomie und Systematik	17
2	1	1	18		Mikrobiologie	20
2	1	1	19		Biotechnology	73

2	1	2	20		Synthetische Biologie	0
2	1	2	21		System- Biologie	14
2	1	2	22		Sonstige Biologie	7
2	2			Medizin und Grenzgebiete		
2	2	1		Mikrobiologie, Virologie, Immunologie		
2	2	1	1		Mikrobiologie siehe Biologie	-
2	2	1	2		Virologie	18
2	2	1	3		Immunologie	43
2	2	2		Medizin		
2	2	2	1		Arbeitsmedizin und Sozialmedi- zin	73
2	2	2	2		Physiologie	12
2	2	2	3		Ernährungswissenschaften	nmb
2	2	2	4		Pathologie und Gerichtliche Medizin	36
2	2	2	5		Klinische Chemie und Pathobio- chemie	22
2	2	2	6		Toxikologie	19
2	2	2	7		Anästhesiologie	28
2	2	2	8		Innere Medizin - Kardiologie, Angiologie	76
2	2	2	9		Innere Medizin - Pneumologie	26
2	2	2	10		Innere Medizin - Onkologie	92
2	2	2	11		Innere Medizin - Gastroentero- logie, Stoffwechsel	57
2	2	2	12		Innere Medizin - Endokrinologie, Diabetologie	51
2	2		13		Innere Medizin - Rheumatologie	20
2	2	2	14		Kinderheilkunde	99
2	2	2	15		Frauenheilkunde und Geburts- hilfe	46
2	2	2	16		Dermatologie	41
2	2	2	17		Urologie	52
2	2	2	18		Chirurgie	116
2	2	2	19		Transplantation	12
2	2	2	20		Zahnheilkunde, Mund-, Kiefer- und Gesichtschirurgie	24

2	2	2	21		Radiologie, Nuklearmedizin, Strahlentherapie und Strahlenbiologie	6
2	2	2	22		Gerontologie	15
2	2	2	23		Augenheilkunde	42
2	2	2	24		Anatomie	12
2	2	2	25		Allergologie	4
2	2	2	26		Infektionskrankheiten	39
2	2	2	27		Integrative u. komplementäre Medizin	21
2	2	2	28		Otorhinolaryngologie	27
2	2	2	29		Parasitologie	9
2	2	2	30		Physiologie	12
2	2	2	31		Tropische Medizin	12
2	2	2	32		Periphere Gefäßkrankheiten	15
2	2	2	33		Notfallmedizin	27
2	2	2	34		Drogenmissbrauch	23
2	2	2	35		Gesundheitsversorgung	19
2	2	2	36		Gesundheitspolitik	17
2	2	2	37		Medizinische Ethik	10
2	2	2	38		Medizinische Forschung	15
2	2	2	39		Krankenpflegekunde	37
2	2	2	40		Rehabilitation	26
2	2	2	41		Sonstige Medizin	98
2	2	2	42		Medizinische Informatik	17
2	2	2	43		Medizinische Labortechnik	22
2	2	2	44		Allgemeine Medizin	193
2	2	2	45		Übergreifende Zeitschrift, mehrere Fachgebiete innerhalb Medizin	317
2	2	2	46		Pharmakologie u. Pharmakotherapie	46
2	3			Agrar-, Forstwissenschaften, Gartenbau, Tiermedizin	n nb	
3				Naturwissenschaften	1168	
3	1			Chemie		
3	1	1	1		Anorganische Chemie inkl. Metalle, Organo-Metalle u. Mineralien	70

3	1	1	2		Organische Chemie	58
3	1	1	3		Physikalische u. theoretische Chemie, Katalyse	144
3	1	1	4		Makromolekulare Chemie	81
3	1	1	5		Analytische Chemie und Umweltchemie	91
3	1	1	6		Technische Chemie und Verfahrenstechnik	119
3	1	1	7		Biochemie	45
3	1	1	8		übergreifende Zeitschrift, mehrere Fachgebiete innerhalb Chemie	148
2	3			Physik		
2	3	1	1		Mathematische Physik	14
2	3	1	2		Quanten-Mechanik, Feldtheorie, Spezielle Relativität	--
2	3	1	3		Allgemeine Relativität u. Gravitation	5
2	3	1	4		Statistische Physik, Thermodynamik, nicht-lineare Systeme	21
2	3	1	5		Technische Physik	5
2	3	1	6		Gemeinsame Instrumente u. Apparate von Physik u. Astronomie	16
2	3	1	7		Große Forschungsgeräte (Beschleuniger, Reaktoren)	2
2	3	1	8		Elementarteilchen und Felder	8
2	3	1	9		Kernphysik	10
2	3	1	10		Atom- u. Molekularphysik	10
2	3	1	11		Akustik, Wärmeübertragung, Klassische Mechanik, Druck, Strömungsmechanik	30
2	3	1	12		Elektromagnetismus, Optik	28
2	3	1	13		Festkörper (Struktur-, mechanische u. thermische Eigenschaften)	41
2	3	1	14		Physik der Gase, Plasmen, Elektrische Entladung	6
2	3	1	15		Quanteninformation	3

2	3	1	16		Geophysik, Ozeanografische Physik	8
2	3	1	17		Astronomy, Astrophysik	43
2	3	1	18		Nukleartechnik	28
2	3	1	19		Strahlungsschutz	10
2	3	1	20		Nanotechnology	11
2	3	1	21		Photonik	7
2	3	1	22		Physikalische Chemie u. Chemische Physik	21
2	3	1	23		Biologische u. medizinische Physik	7
2	3	1	24		Geschichte u. Philosophie der Physik	3
2	3	1	25		Ausbildung in der Physik	3
2	3	1	26		Fachübergreifende Zeitschriften innerhalb der Physik	72
3	3			Mathematik		n nb
3	4			Geowissenschaften		n nb
4				Ingenieurwissenschaften		1295
4	1			Maschinenbau u. Produktionstechnik		
4	1	1	1		Produktion, Fertigung	
4	1	1	2		Spanende Fertigungstech., Laserbearbeitung, Produktionsautomat.	13
4	1	1	3		Ur- und Umformtechnik	9
4	1	1	4		Mikro-, Feinwerktechnik, Montage-, Füge-, Trenntechnik, Schweißtechnik	24
4	1	1	5		Konstruktion, Maschinenelemente	21
4	1	1	5		Sonstige Produktion	20
4	1	1	6		Übergreifend Produktion	27
4	1	1	7		Allgemeine Gebiete	
4	1	1	8		Mechanik	56
4	1	1	9		Akustik, Schwingungsmechanik	20
4	1	1	10		Festigkeitsmechanik	8
4	1	1	11		Strömungsmechanik, Hydraulik	23
4	1	1	12		Technische Thermodynamik	28

4	1	1	13		Strömungs- und Kolbenmaschinen	21
4	1	1	14		Fahrzeuge	
4	1	1	15		Verkehrs- und Transportsysteme, Logistik	37
4	1	1	16		Automobiltechnik	9
4	1	1	17		Bahntechnik	9
4	1	1	19		Luft-, Raumfahrttechnik	54
4	1	1	20		Schiffstechnik	34
4	1	1	21		Sonstige Gebiete	
4	1	1	22		Nukleartechnik	21
4	1	1	23		Energietechnik, Verbrennungstechnik	25
4	1	1	24		Offshore, Meerestechnik	13
4	1	1	25		Agrar-, Lebensmitteltechnik	13
4	1	1	26		Umwelttechnik	10
4	1	1	27		Arbeitswissenschaft, Ergonomie, Mensch-Maschine-Systeme	6
4	1	1	28		Verfahrenstechnik	28
4	1	1	29		Getriebetechnik	17
4	1	1	30		Schmierung	6
4	1	1	31		Sicherheits-, Wehrtechnik	15
4	1	1	32		Sonstige Themen	40
4	1	1	33		Übergreifend Maschbau	61
4	3			Materialwissenschaften u. Werkstofftechnik	Werkstoffe	
4	3	1	1		Metalle	23
4	3	1	2		Polymere	8
4	3	1	3		Keramik	3
4	3	1	4		Sinter- und Verbundwerkstoffe	13
4	3	1	5		Oberflächen, Beschichtungen, Funktionsschichten	18
4	3	1	6		Rohstoffe, Bergbau, Hüttenwesen	21
4	3	1	7		Metallurgie und Thermodynamik mehrphasiger metallischer Systeme	9
4	3	1	8		Biomaterialien	11

4	3	1	9		Textilien, Textilmaschinen	65
4	3	1	10		Papier, Papiermaschinen	8
4	3	1	11		Materialprüfung	7
4	3	1	12		Baumaterialien (Bauwesen)	7
4	3	1	13		Sonstige Materialien	66
4	4			Elektrotechnik, Informatik u. Systemtechnik		
4	4	1		Elektrotechnik		
4	4	1	1		Energietechnik	13
4	4	1	3		Elektronik, Mikroelektronik	53
4	4	1	4		Telekommunikation	27
4	4	1	5		Hochfrequenztechnik	18
4	4	1	6		Regelungstechnik, Automatisierung	11
4	4	1	7		Sensoren, Instrumente	8
4	4	1	8		Medizintechnik	43
4	4	1	9		Mikrosysteme	15
4	4	1	10		Sonstiges E-Technik	41
4	4	1	11		Übergreifende Zeitschriften E-Technik	23
4	4	2		Informatik/Computertechnik		
4	4	2	1		Computerlogik u. formale Sprachen	7
4	4	2	2		Computer-Systeme	13
4	4	2	3		Anwendungen (Signalverarb., Bildverarb., CAD, Anw. f. Geow., Bioinformatik, Multimedia, Statistik ...)	12
4	4	2	4		Künstliche Intelligenz (Mustererk., Sprachverarb., maschin. Lernen ...)	83
4	4	2	5		Web- u. Netzwerktechnologien	52
4	4	2	6		Kybernetik u. Robotik	33
4	4	2	7		Grafische Datenverarbeitung, Visualisierung	8
4	4	2	8		Mensch-Maschine-Interaktion	18
4	4	2	9		Informationssysteme	4

4	4	2	10		Software Engineering u. Programmierung	27
4	4	2	11		Super- u. Hochgeschwindigkeitscomputer	28
4	4	2	12		Sonstige Informatik	4
4	5			Bauwesen u. Architektur		
4	5	1	1		Hochbau	10
4	5	1	2		Tiefbau	11
4	5	1	3		Wasserbau, Kanalisation, Küstenbau	13
4	5	1	4		Siedlungsbau, Städtebau	5
4	5	1	5		Landschaftsplanung, -bau	4
4	5	1	6		Bauwesen übergreifend	
4	5	1	7		Bauwesen sonstiges	13
4	5	1	8		Architektur	12
4	5	1	9		Geschichte, hist. Gebäude	5
4	5	1	10		Architektur übergreifend	9

N Zahl der Zeitschriften, die zu den einzelnen Stellen intellektuell identifiziert wurden

nrb = noch nicht bearbeitet

Quelle: Recherchen und Berechnung des Fraunhofer ISI