

Vollständige Erfassung von Patentanmeldungen aus Universitäten

Bericht an das Bundesministerium für Bildung und Forschung (BMBF)

Ulrich Schmoch, Friedrich Dornbusch, Nicolai Mallig,
Carolin Michels, Nicole Schulze

Karlsruhe, Februar 2011

Inhaltsverzeichnis

Zusammenfassung	1
1 Einführung	4
2 Stand der Forschung	6
2.1 KEINS Database on Academic Inventors	6
2.2 Vorarbeiten des Fraunhofer ISI.....	6
3 Forschungsansatz	9
3.1 Zusammenführen von Autoren und Erfindern	9
3.2 Selektionskriterien	9
4 Datenquellen für Patente und Publikationen	16
4.1 PATSTAT	17
4.2 Scopus	18
5 Länderanalysen	21
5.1 Deutschland	21
5.1.1 DPMA-Anmeldungen.....	21
5.1.2 EPA-Anmeldungen	31
5.2 Schweiz, EPA-Anmeldungen.....	33
5.3 Frankreich, EPA-Anmeldungen	36
6 Gesamtergebnisse	41
Literatur, Referenzen	43

Tabellen

Tabelle 3–1: Grobe Konkordanz zwischen Technikfeldern und Wissenschaftsfeldern	14
Tabelle 3–2: Feine Konkordanz zwischen Technikfeldern und Wissenschaftsfeldern	15

Abbildungen

Abbildung 3-1: Anteil der Records in Scopus mit voll ausgeschriebenen Vornamen der Autoren.....	11
Abbildung 4-1: Abdeckung von Artikel und Proceedings in den Feldern Maschinenbau und Informatik in Web of Science und Scopus, 2009	20
Abbildung 5-1: Anmeldungen aus Universitäten am DPMA nach den Kriterien von Universitäten als Anmelder oder Professoren als Erfinder.....	22
Abbildung 5-2: Ermittelte DPMA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien.....	24
Abbildung 5-3: Für das Jahr 2003 ermittelte DPMA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien.....	26
Abbildung 5-4: Ermittelte DPMA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien.....	27
Abbildung 5-5: Für das Jahr 2003 ermittelte DPMA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien.....	28
Abbildung 5-6: Technische Felder der DPMA-Anmeldungen aus deutschen Universitäten.....	30
Abbildung 5-7: Ermittelte EPA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien	32
Abbildung 5-8: Für das Jahr 2003 ermittelte EPA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien.....	32
Abbildung 5-9: Technische Felder der EPA-Anmeldungen aus deutschen Universitäten im Jahr 2006.....	33

Abbildung 5-10:	Ermittelte EPA-Anmeldungen aus Schweizer Universitäten nach verschiedenen Selektionskriterien.....	35
Abbildung 5-11:	Für das Jahr 2003 ermittelte EPA-Anmeldungen aus Schweizer Universitäten nach verschiedenen Selektionskriterien	35
Abbildung 5-12:	EPA-Anmeldungen aus französischen Universitäten nach den KEINS-Analysen	37
Abbildung 5-13:	Ermittelte EPA-Anmeldungen aus französischen Universitäten nach verschiedenen Selektionskriterien	39
Abbildung 5-14:	Für das Jahr 2003 ermittelte EPA-Anmeldungen aus französischen Universitäten nach verschiedenen Selektionskriterien	40

Zusammenfassung

Zu einer korrekten Erfassung von Strukturen und Trends von Patentanmeldungen aus Hochschulen sind Ansätze erforderlich, mit denen auch Anmeldungen erfasst werden können, die nicht formal über Hochschulen selbst eingereicht werden. Denn die Restgröße von Anmeldungen, die über Unternehmen oder Privatpersonen, also nicht von Universitäten selbst an die Patentämter geleitet werden, ist erheblich.

Bislang gibt es keine Ansätze, die mit vertretbarem Aufwand, Patentanmeldungen aus Universitäten zutreffend ermitteln können. In dem sogenannten KEINS-Projekt ist die Ermittlung von Personallisten mit einem großen Zeit- und Arbeitsaufwand verbunden und zudem unvollständig. Der frühere Ansatz des Fraunhofer ISI beruht zu großen Teilen auf einer Abschätzung und nicht eine präzisen Ermittlung relevanter Anmeldungen.

Der Ansatz des vorliegenden Projekts beruht auf der Prüfung auf Namensgleichheiten von Autoren und Erfindern. Gegenüber dem Ansatz des KEINS-Projektes besteht der wesentliche Unterschied darin, dass die Namen von Universitätsmitarbeitern und -mitarbeiterinnen nicht von offiziellen Personallisten der Hochschulen abgeleitet werden, sondern grundsätzlich alle forschungsaktiven Personen erfassen, die wissenschaftlich publizieren. Außerdem können, wenn ein sinnvoller Suchalgorithmus installiert ist, die Autorenlisten regelmäßig mit vertretbarem Aufwand aktualisiert werden und es sind auch Untersuchungen in Ländern möglich, bei denen Personaldaten aus Hochschulen in größerem Maßstab nicht verfügbar sind.

Da das Matching der Namen von Autoren und Erfindern aufgrund der hohen Datenmengen zu Homonymen und damit zur fälschlichen Zuordnung von Patentanmeldungen zu Universitäten führt, ist die Anwendung von Selektionskriterien erforderlich. Diese beziehen sich in der Untersuchung auf

1. Land,
2. Einrichtung/Organisation,
3. Nachname/Vorname,
4. Region,
5. Zeitpunkt,
6. inhaltliches Gebiet.

Die Analysen wurden mit der Datenbank Scopus durchgeführt, weil diese schon seit 1996 die Vornamen von Autoren in ausgeschriebener Form erfasst, weil auch in früheren Jahren eine eindeutige Zuordnung von Autoren zu Einrichtungen möglich ist und weil die Felder Maschinenbau und Informatik breit erfasst sind, die im Kontext von Anmeldungen aus Universitäten von hoher Relevanz sind.

Für die Verbindung von der inhaltlichen Zuordnung von Technikfeldern von Patentanmeldungen und Wissenschaftsgebieten von Publikationen wurden eine "grobe" und eine "feine" Konkordanz entwickelt, deren Anwendung als Selektionskriterien zu sinnvollen Ergebnissen geführt hat.

Im Einzelnen wurden die folgenden Fälle geprüft:

- Anmeldungen deutscher Herkunft am DPMA,
- Anmeldungen deutscher Herkunft am EPA,
- Anmeldungen Schweizer Herkunft am EPA,
- Anmeldungen französischer Herkunft am EPA.

Die Analysen führten in allen Fällen zu sinnvollen Größenordnungen von Anmeldungen aus Universitäten. Die Selektionskriterien führen zu realistischen Abstufungen, wobei offensichtlich die Anwendung der groben Konkordanz oder der 1-stelligen Postleitzahlen zu den besten Ergebnissen führen. Durch die Nutzung von zwei gegenüber einem Publikationsjahr lassen sich höher Erträge erzielen, so dass in Bezug auf die Optimierung von Recall und Precision eine längere Publikationszeitraum verbunden mit strengeren Selektionskriterien am Besten sein dürfte.

Insgesamt ist der Recall am Europäischen Patentamt größer als am Deutschen Patent- und Markenamt, weil am EPA inhaltlich das Schwergewicht auf forschungsintensiven Technologien, deren zugeordnete Wissenschaftsfelder in Scopus breiter erfasst sind. Im Ländervergleich ist der Recall für Frankreich besonders schwach, da offensichtlich französische Autoren in englischen Journals seltener publizieren als deutsche oder Schweizer.

Neben der Abdeckung in Scopus ist für jedes Land außerdem die Granularität der Postleitzahlbezirke zu prüfen, da diese zu spezifischen Wirkungen bei der Anwendung des räumlichen Selektionskriteriums führt. Im vorliegenden Fall gibt es spürbare Unterschiede bei der Wirkung dieses Kriteriums zwischen der Schweiz und Deutschland.

Auch wenn die Analysen zu grundsätzlich sinnvollen Ergebnissen führen, ist der Recall bei einfachen, ungewichteten Selektionskriterien noch unbefriedigend. Um hier zu einer Verbesserung zu kommen, ist es erforderlich, eindeutig definierte Referenzdatensätze zur Verfügung zu haben, um daraus die Maße Recall und Precision berechnen zu können.

Hier wird zum einen für DPMA-Anmeldungen, die von Universitäten selbst angemeldet wurde, die Fälle genauer zu prüfen, die über das Matching-Verfahren nicht bestimmt werden konnten.

Zum zweiten ist für einen Jahrgang der über das Matching-Verfahren bestimmten wurde vorgesehen, alle Erfinder zu befragen, ob sie in zutreffender Weise einer Universität zugeordnet wurden. Auch daraus sollen Folgerungen für die Auswahl und Gewichtung von Matching-Kriterien gezogen werden.

Ziel ist letztlich die Entwicklung eines multivariaten Ansatzes mit unterschiedlicher Gewichtung von Selektionskriterien, um so Recall und Precision optimieren zu können. Diese Arbeiten werden im Rahmen der Eigenforschung bis spätestens Ende 2011 realisiert.

Für den Gesamtrahmen des Kompetenzzentrums Bibliometrie hat die Studie gezeigt, dass mit einer bibliometrischen Version von Scopus umfangreiche Datenverknüpfungen in einer überschaubaren Rechenzeit umsetzbar sind. Von daher ist das Kompetenzzentrum technisch in der Lage, aufwändige Analysen der Auftragsforschung durchzuführen und sich auch an der internationalen wissenschaftlichen Diskussion über die Strukturen bibliometrischer Datenbanken und die Konzeption geeigneter Indikatoren zu beteiligen.

1 Einführung

Der Wissens- und Technologietransfer aus Hochschulen ist in den letzten Jahren als ein wesentlicher Ansatzpunkt zur Modernisierung der Wirtschaftsstrukturen und zur Förderung der wirtschaftlichen Dynamik gesehen worden {Crespi, 2011 6226 /id; Achleitner, 2009 5388 /id; Egelin, 2007 4890 /id}. Ein wichtiger Teilaspekt war dabei, die Patentanmeldungen aus Hochschulen zu fördern, weshalb im Jahr 2002 das sogenannte Hochschullehrerprivileg abgeschafft und gleichzeitig Patentverwertungsagenturen (PVA) aufgebaut wurden, die mittlerweile in der Technologie-Allianz als Netzwerk zusammengeschlossen sind. Die Anmeldung von Patenten aus öffentlichen Forschungseinrichtungen findet auch international zunehmend Aufmerksamkeit, was schon früh im Jahre 2003 in einem entsprechenden Bericht der OECD dokumentiert ist {OECD, 2003 1536 /id}. Zur Analyse der Veränderungen, die aus der zunehmenden Transferorientierung der Hochschulen resultieren, ist es wichtig, die Patentanmeldungen aus Hochschulen in zutreffender Weise erfassen zu können.

Diese Erfassung ist jedoch methodisch komplex, weil trotz der Gesetzesänderung von 2002 nur etwa 40% aller Patentanmeldungen aus Hochschulen von den Hochschulen selbst angemeldet werden. Ein großer Teil wird von Unternehmen registriert und die Hochschulmitarbeiter erscheinen lediglich als Erfinder. In anderen Fällen melden Hochschullehrer ihre Erfindungen immer noch privat an. Nach aktuellen Analysen für die Expertenkommission Forschung und Innovation (EFI) wurden im Jahr 2007 35% der Patentanmeldungen aus Hochschulen von den Universitäten selbst angemeldet, 45% durch Unternehmen und 21% durch Hochschulmitarbeiter privat {Frietsch, 2010 5157 /id}.

Diese Situation für Deutschland, bei der nur ein begrenzter Teil der Anmeldungen aus Universitäten von diesen selbst angemeldet wird, findet sich in noch ausgeprägterer Weise in anderen europäischen Ländern {Lissoni, 2008 3374 /id; Schmoch, 2004 6227 /id}. Auch in den Vereinigten Staaten, in denen nach allgemeiner Auffassung praktisch alle Patentanmeldungen mit universitärer Herkunft von Universitäten angemeldet werden, erweist sich bei genauerer Prüfung die Quote der Anmeldungen über andere natürliche oder juristische Personen mit 40% als überraschend hoch {Audretsch, 2006 6228 /id}.

Zu einer korrekten Erfassung von Strukturen und Trends von Patentanmeldungen aus Hochschulen sind daher verbesserte Ansätze erforderlich, mit denen auch Anmeldungen erfasst werden können, die nicht formal über Hochschulen selbst eingereicht werden. Die im Projekttitel hervorgehobene "vollständige Erfassung" bezieht sich daher

auf diese erhebliche Restgröße von Anmeldungen, die über Unternehmen oder Privatpersonen an die Patentämter geleitet werden.

2 Stand der Forschung

Aufgrund dieser erheblichen Nachfrage aus dem Bereich der Wissenschafts- und Innovationspolitik gibt es verschiedene Ansätze zur Lösung dieser Problematik. In dem folgenden Absatz werden die wichtigsten kurz beschrieben.

2.1 KEINS Database on Academic Inventors

Eine Arbeitsgruppe um Francesco Lissoni bemüht sich seit einigen Jahren die sogenannte KEINS Datenbank mit akademischen Erfindern aufzubauen (KEINS= Knowledge-based Entrepreneurship: Innovation, Networks and Systems). Der Aufbau der Datenbank begann mit den Ländern Frankreich, Italien und Schweden {Lissoni, 2006 6229 /id;Lissoni, 2008 3374 /id;Lissoni, 2009 6230 /id}. Mittlerweile sind Dänemark, Großbritannien und die Niederlande an der Arbeit beteiligt. Dabei werden zunächst Listen des Universitätspersonals zusammengestellt und die daraus resultierenden Namen mit denen von Erfindern aus Patentanmeldungen des jeweiligen Landes verglichen. Im Kern steht also das Matching der Namen aus Personal- und Erfinderslisten. Ausgehend von dieser Methode wurde auch dort festgestellt, dass die große Mehrheit von Patentanmeldungen von Akademikern nicht von Universitäten angemeldet wird. Diese Methode erweist sich durchaus als hilfreich, erfordert aber einen erheblichen Zeit- und Kostenaufwand, da im Zeitverlauf immer neue Listen der Mitarbeiter von Universitäten erhoben werden müssen und in vielen Ländern derartige Daten nur schwer verfügbar sind. Aktuell wird dieses Forschungsnetzwerk von der European Science Foundation (ESF) gefördert, wobei das verbesserte Matching zwischen Personal- und Erfindernamen im Vordergrund steht {Lissoni, 2010 6231 /id}. Abgesehen von dem Kostenaufwand ist es auch ein methodisches Problem, dass in der Regel in den Personallisten der Hochschulen nur die Mitarbeiter mit einer offiziellen Stellung, insbesondere Dozenten, enthalten sind, Erfindungen aber häufig auch von anderen Mitarbeitern stammen.

2.2 Vorarbeiten des Fraunhofer ISI

Das Fraunhofer ISI hat ein Verfahren entwickelt, um für Deutschland eine realistische Abschätzung der Anmeldungen von Patenten aus Hochschulen vornehmen zu können {Schmoch, 2007 5025 /id}. Grundlage der Untersuchung in Patentdatenbanken sind dabei zunächst die Anmeldungen, die von Universitäten registriert werden. Zusätzlich werden in die Recherche solche Patentanmeldungen einbezogen, bei denen der Anmelder oder zumindest einer der Erfinder einen Professorentitel hat. Damit werden auch solche Anmeldungen erfasst, die offiziell über Unternehmen oder Privatpersonen registriert werden. Methodische Grundlage ist dabei, dass es in Deutschland üblich ist,

den Professorentitel auf Dokumenten wie Patentanmeldungen anzugeben, auch wenn dieser juristisch gesehen kein Namensbestandteil wie etwa der Dokortitel ist. Fehlermöglichkeiten dieses Ansatzes resultieren daraus, dass nicht alle Professoren bei einer Anmeldung auch tatsächlich den Professorentitel angeben. Weiterhin können Erfinder mit Professorentitel nicht nur von Universitäten, sondern auch von Fachhochschulen stammen. Es ist allerdings erwünscht, dass auch diese Gruppe bei den Recherchen berücksichtigt wird. Außerdem ist der Anteil von Fachhochschulprofessoren im Gesamtdatensatz relativ gering. Weiterhin ist es möglich, dass ein Mitarbeiter eines Unternehmens einen Titel als Apl-Professor hat und die Erfindung nicht aus dem Kontext der Universität, sondern aus dem des Unternehmens stammt. Hier ist zumindest anzunehmen, dass es einen wissenschaftlichen Hintergrund in Verbindung mit einer Universität gibt.

Ein methodisch größeres Problem ist dagegen, Erfindungen mit universitärer Herkunft zu ermitteln, bei denen nicht Professoren, sondern andere Mitarbeiter und Mitarbeiterinnen die Erfinder sind. Bei den Patentanmeldungen am Deutschen Patent- und Markenamt (DPMA) bei denen Universitäten als Anmelder erscheinen, ist der Anteil der Anmeldungen, bei denen nicht mindestens ein Erfinder mit Professorentitel vertreten ist, etwa 50%. Hier wurde konservativ geschätzt, dass bei den Patentanmeldungen, die nicht über Universitäten laufen, bei denen aber ein Anmelder oder Erfinder einen Professorentitel hat, zusätzlich etwa 40% an Patentanmeldungen hinzugerechnet werden müssen, um die Gesamtzahl der Anmeldungen realistisch zu erfassen. Bei diesem Ansatz kann die Zahl der Patentanmeldungen aus Hochschulen realistisch benannt werden, wobei allerdings die Unsicherheit bei der Quote der universitären Patentanmeldungen ohne Professoren als Erfinder liegt. Außerdem kann hier zwar die Zahl abgeschätzt werden, eine Identifikation einzelner Patentanmeldungen aus dieser Gruppe ist jedoch nicht möglich.

Ein weiterer Nachteil dieses Ansatzes besteht darin, dass die Angabe des Professorentitels nur in Deutschland und Österreich üblich ist. Schon für die Schweiz hat sich dieser Ansatz als nicht tragfähig erwiesen {Schmoch, 2004 4262 /id}. In anderen Ländern ist die Angabe des Professorentitels unüblich. Ein weiteres Problem ist auch, dass bei Anmeldungen am Europäischen Patentamt (EPA) der Professorentitel in der Regel nicht verzeichnet wird, so dass auch für Anmeldungen deutscher Herkunft eine Bestimmung von europäischen Patentanmeldungen aus Hochschulen nicht möglich ist.

Immerhin kann der Datenansatz für Deutschland am DPMA als sinnvolle Referenz für die Größenordnung der zu erwartenden Patentanmeldungen herangezogen werden.

Aufgrund der Beschränkungen des beschriebenen Verfahrens gerade im internationalen Kontext wurde in den Jahren 2002 und 2003 für die Nanotechnologie und ausgewählte Felder der Lebenswissenschaft ein Ansatz verfolgt, bei dem Autoren, deren Publikationen in der Datenbank Web of Science (WoS) registriert sind, mit Erfindern in den jeweiligen Gebieten gematcht wurden {Noyons, 2003 2213 /id;Noyons, 2003 6232 /id;Schmoch, 2004 6227 /id}. Dabei wurde die Recherche sowohl auf der Publikations- als auch auf der Patentseite auf die jeweilige Thematik eingegrenzt, was bei der Zahl der potentiell relevanten Autoren und Erfinder zu einer entsprechenden zahlenmäßigen Beschränkung führte. Bei den Autoren wurden zusätzlich nur solche berücksichtigt, bei denen eine Zugehörigkeit zu Hochschulen in der Publikationsdatenbank ausgewiesen war. Von daher konnten sinnvolle Ergebnisse erzielt werden, wobei aber trotz der Beschränkung der Mengen der Namen eine Unsicherheit in den Fällen blieb, bei denen eine eindeutige Zuordnung zwischen Autoren und Erfindern nicht möglich war. Nach dem damaligen Stand des Web of Science waren bei Autoren jeweils nur die Initialen des Vornamens vermerkt, so dass immer wieder Fälle mit gleichen Nachnamen, aber verschiedenen Vornamen mit gleichem Anfangsbuchstaben auftraten. Für die themenoffene Suche nach Erfindern aus Universitäten ergeben sich demgegenüber völlig andere Verhältnisse, da die Zahl der Namen auf der Seite der Autoren wie auch auf der Seite der Erfinder erheblich größer ist und damit die Unschärfen beim Matching zunehmen. Insbesondere ist es nicht mehr ausreichend, wenn zu den Vornamen der Autoren nur die Initialen zur Verfügung stehen.

3 Forschungsansatz

3.1 Zusammenführen von Autoren und Erfindern

Der Ansatz dieses Projekts beruht, wie schon das frühere Projekt von Noyons et al. {Noyons, 2003 2213 /id /d;Noyons, 2003 6232 /id /d} auf der Prüfung auf Namensgleichheiten von Autoren und Erfindern. Gegenüber dem Ansatz von Lissoni et al. besteht der wesentliche Unterschied darin, dass die Namen von Universitätsmitarbeitern und -mitarbeiterinnen nicht von offiziellen Personallisten der Hochschulen abgeleitet werden, sondern grundsätzlich alle forschungsaktiven Personen erfassen, die wissenschaftlich publizieren. Außerdem können, wenn ein sinnvoller Such-Algorithmus installiert ist, die Autorenlisten regelmäßig mit vertretbarem Aufwand aktualisiert werden und es sind auch Untersuchungen in Ländern möglich, bei denen breite Personaldaten aus Hochschulen in größerem Maßstab nicht verfügbar sind.

Grundsätzliche methodische Probleme, die eine genauere Untersuchung erforderlich machen, bestehen darin, dass forschungsaktive Personen, auch wenn sie Erfindungen generieren, nicht unbedingt auch wissenschaftliche Publikationen veröffentlichen, sodass mit gewissen Ausfällen in den Personenlisten zu rechnen ist. Weiterhin ist es möglich, dass die Publikationsdatenbanken nicht alle Zeitschriften erfassen, in denen Universitätsmitarbeiter und -mitarbeiterinnen publizieren. Schließlich besteht das Problem, dass bei dem Abgleich von großen Mengen von Autoren- und Erfindernamen Namensgleichheiten bei verschiedenen Personen, sogenannte Homonyme, auftreten können, sodass fehlerhafte Zuordnungen von Patentanmeldungen zu Hochschulen auftreten können.

Ziel des vorliegenden Projekts ist es somit, durch den Abgleich von Autoren- und Erfindernamen eine möglichst große Anzahl von Patentanmeldungen aus Universitäten zu identifizieren und dabei die Rate fehlerhafter Zuordnungen möglichst niedrig zu halten. Das Verfahren soll dabei nicht nur für Hochschulpatentanmeldungen deutscher Herkunft, sondern auch für solche aus anderen Ländern anwendbar sein.

3.2 Selektionskriterien

Da bei der Vielzahl von Namen auf Autoren- und Erfinderseite bei einer einfachen Überprüfung der Übereinstimmung von Nachnamen erhebliche Fehler auftreten können, ist es erforderlich, weitere Zuordnungs- oder Selektionskriterien zu berücksichtigen. Ausgangspunkt müssen dabei Informationen sein, die in Patentdatenbanken wie z.B. PATSTAT und Publikationsdatenbanken wie z.B. Web of Science oder Scopus

verfügbar sind. Darunter fallen im Kontext der vorliegenden Fragestellung insbesondere die folgenden Kriterien:

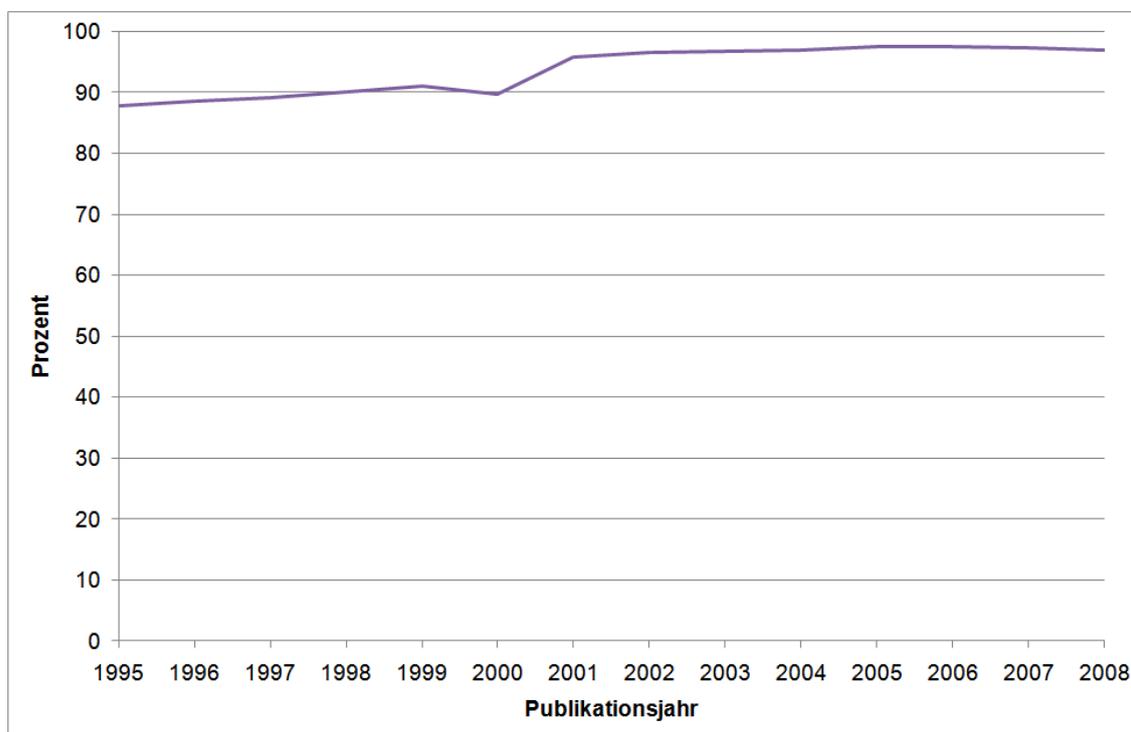
1. Land,
2. Einrichtung/Organisation,
3. Nachname/Vorname,
4. Region,
5. Zeitpunkt,
6. inhaltliches Gebiet.

zu 1: Dieses ist in Publikationsdatenbanken das Sitz-Land der Einrichtung, bei der der Autor beschäftigt ist. Bei Erfindern werden in der Mehrzahl der Fälle die privaten Wohnadressen angegeben; in Einzelfällen handelt es sich auch um die Adresse der Einrichtung, bei der der Erfinder arbeitet. Im Falle multinationaler Unternehmen ist dies in der Regel das jeweilige nationale Tochterunternehmen.

zu 2: Bei Autoren können in Publikationsdatenbanken jeweils diejenigen abgegrenzt werden, die an Universitäten oder anderen Arten von Hochschulen arbeiten. Das ist im deutschen Fall relativ unproblematisch. Bei anderen Ländern muss jeweils sorgfältig geprüft werden, in welchen Bezeichnungen die dortigen Hochschulen in der Datenbank vermerkt sind.

zu 3: In Patentdatenbanken sind grundsätzlich neben den Nachnamen die vollständigen Vornamen verzeichnet, da es sich bei Patentanmeldungen um juristische Dokumente handelt. In der Publikationsdatenbank Web of Science war es lange Zeit üblich, anstelle der vollen Vornamen lediglich die Initialen zu registrieren. Im Wettbewerb mit Scopus hat sich der Hersteller Thomson-Reuters entschlossen, ebenfalls in größerem Umfang die vollständigen Vornamen zu verzeichnen: Ab dem Publikationsjahr 2007 sind für 75% aller abgedeckten Publikationen volle Vornamen verfügbar. Am aktuellen Rand wird die noch höhere Quote von 80% erreicht. In der Datenbank Scopus sind demgegenüber schon ab 1996 vollständige Vornamen verfügbar (vgl. Abbildung 3-1), womit auch für frühere Jahre schon dieses wichtige Kennzeichen verfügbar ist. Prüfungen haben ergeben, dass bei der großen Datenmenge von abzugleichenden Autoren- und Erfindernamen die Initialen des Vornamens keinesfalls ausreichen und in jedem Fall mit vollen Vornamen gearbeitet werden muss. Um auch frühere Jahre abdecken zu können, fiel daher gleich zu Anfang des Projekts die Entscheidung ausschließlich mit der Datenbank Scopus weiterzuarbeiten.

Abbildung 3-1: Anteil der Records in Scopus mit voll ausgeschriebenen Vornamen der Autoren



Quelle: Scopus, eigene Recherchen und Berechnungen

zu 4: Hinsichtlich der Region lassen sich bei Patenten die Postleitzahlen auswerten und auch in Publikationsdatenbanken sind Postleitzahlen verfügbar. Bei Patenten bezieht sich die Postleitzahl in der Regel auf den privaten Wohnsitz des Erfinders, bei Publikationen auf die Adresse der Einrichtung, bei der der Autor beschäftigt ist. Bei der Nutzung von Postleitzahlen als Selektionskriterium wird unterstellt, dass der Erfinder aus einer Hochschule in der regionalen Nähe zu dieser Hochschule wohnt. Da es eine Reihe von Fällen gibt, bei denen Hochschulmitarbeiter in größerer Entfernung von der Hochschule wohnen, ist zu prüfen, welche Quote von Hochschulmitarbeitern davon betroffen ist. Weiterhin ist es möglich, durch eine verschiedene Anzahl von Postleitzahlstellen die Präzision der Zuordnung zu variieren.

zu 5: Grundsätzlich sollte eine zeitliche Nähe zwischen der Publikation eines Artikels und der Anmeldung einer Erfindung bestehen. Allerdings ist zu berücksichtigen, mit welcher Präzision und mit welcher zeitlichen Verzögerung Zeiträume der Veröffentlichungen von Publikationen und Erfindungen angegeben werden. Bei Erfindungen wird grundsätzlich der Zeitpunkt der Einreichung einer Anmeldung korrekt erfasst, da er für das Patentprüfungsverfahren relevant ist. Allerdings werden eingereichte Erfindungen erst 18 Monate nach ihrer Erstregistrierung, dem sogenannten Prioritätsdatum, veröffentlicht und werden auch dann erst in Datenbanken publiziert. Bei Patentanmeldungen

ist davon auszugehen, dass die Forschungsaktivitäten, die zur Entstehung einer Erfindung führten, unmittelbar vor der Einreichung der Patentanmeldung stattfanden, da aus rechtlichen Gründen ein möglichst frühzeitiges Einreichen der Patentanmeldung angezeigt ist.

Publikationen werden üblicherweise nach dem Publikationsjahr analysiert. Für die vorliegende Fragestellung, die auf den Zeitpunkt der Forschungsaktivität zielt, wäre es grundsätzlich sinnvoller, Informationen über das Datum der Einreichung eines Artikels bei einer Zeitschrift zur Verfügung zu haben. Dieses ist jedoch nur bei wenigen Zeitschriftenpublikationen systematisch angegeben und ist weder in Web of Science noch in Scopus erfasst. Hier muss daher eine Zeitverzögerung zwischen dem Zeitpunkt der Einreichung und der Publikation eines Artikels geschätzt werden. Da es sich in beiden Datenbanken überwiegend um hochwertige Zeitschriften handelt, ist von einem Review-Prozess auszugehen, der zu einer deutlichen Zeitverzögerung führt. Die Annahme einer durchschnittlichen Verzögerung von einem Jahr zwischen Einreichung und Publikation scheint angemessen. Als Beispiel seien Erfindungen aus dem Jahr 2005 angenommen. Dem würden Publikationen entsprechen, die im Jahr 2006 publiziert und im Jahr 2005 eingereicht wurden. Im Falle von Proceedings ist allerdings auch eine schnellere Veröffentlichung, etwa ein halbes Jahr nach Präsentation auf einer Konferenz auszugehen. Auch bei einer Konferenz müssen aber die Beiträge vorher eingereicht und ausgewählt werden, so dass auch hier die Annahme von einem Jahr zwischen Einreichung und Veröffentlichung realistisch ist.

Eine weitere Problematik bei der Wahl des Zeitpunkts besteht aber auch darin, dass, wie verschiedene Fallprüfungen ergaben, eine Reihe von Hochschulmitarbeitern die Hochschule verlässt und zu einem Industrieunternehmen wechselt. In vielen Fällen gibt es dann Erfindungen, die unter dem Namen des Unternehmens angemeldet werden. Zu diesem Zeitpunkt ist also der Erfinder nicht mehr Hochschulmitarbeiter. Es kann zwar unterstellt werden, dass in einer Vielzahl von Fällen die Ideen zu diesen Erfindungen noch aus der Hochschule stammen. Es wäre aber prekär, diese Patentanmeldungen als Anmeldungen aus einer Hochschule zu klassifizieren. Von daher sollte sichergestellt sein, dass die Publikation in einem Jahr nach Einreichung der Erfindung erfolgt, um sicherzustellen, dass der Autor immer noch Mitarbeiter der Hochschule ist. Um bei dem Beispiel zu bleiben: Bei einer Erfindung, die im Jahr 2005 als Patent angemeldet wurde, sollte eine zugeordnete Publikation frühestens im Jahr 2006, besser erst im Jahr 2007 und später publiziert worden sein.

zu 6: In der Datenbank Scopus steht für die jeweiligen Publikationen eine Klassifikation von Wissenschaftsgebieten mit 343 Stellen zur Verfügung. Diese stellt keine Klassifikation der einzelnen Artikel dar, sondern ordnet die Zeitschriften, in denen ein Artikel pu-

bliziert wird, wissenschaftlichen Gebieten zu. Diese Klassifikation ist somit nicht präzise, erlaubt aber eine generelle thematische Zuordnung einer Veröffentlichung.

Patentanmeldungen sind einzeln klassifiziert mithilfe der sogenannten internationalen Patentklassifikation (IPK), die etwa 75.000 Einzelstellen hat. Die Klassifikation von Patentanmeldungen ist damit sehr genau.

Grundsätzlich besteht das Problem, dass Klassifikationen der Technik und der Wissenschaft unterschiedlichen Kriterien folgen und somit nur ein sehr allgemeiner Zusammenhang zwischen beiden Bereichen hergestellt werden kann. So ist es relativ eindeutig, dass Patentanmeldungen im Bereich der Chemie mit großer Wahrscheinlichkeit wissenschaftlichen Aktivitäten in diesem Bereich zugeordnet werden können. Im Falle der Mikroelektronik können die wissenschaftlichen Hintergründe in der Elektrotechnik, aber auch in der Materialforschung oder der Physik liegen. Hier ist eine klare Zuordnung somit problematischer. Im vorliegenden Kontext geht es letztlich darum, aus welchen wissenschaftlichen Gebieten mit hoher Wahrscheinlichkeit Patentanmeldungen in bestimmten Technikfeldern zu erwarten sind.

Um hier eine Verbindung herzustellen, wurde als Ausgangspunkt eine Technikklassifikation mit 34 Feldern gewählt, die vom Fraunhofer ISI im Auftrag der World Intellectual Property Organization (WIPO) entwickelt wurde {Schmoch, 2008 5105 /id}. Die einzelnen Felder sind in der Regel über Unterklassen, in Einzelfällen auch über Hauptgruppen der internationalen Patentklassifikation (IPK) definiert. Für die Abgleichung zwischen Autoren- und Erfindernamen wurde eine Konkordanz zwischen den 34 Technikfeldern und zugeordneten Feldern von Scopus entwickelt. Dabei entstand zunächst einmal eine "grobe" Variante, bei der unterstellt wird, dass eine Erfindung in einem spezifischen Technikfeld auf einem relativ breiten Spektrum von wissenschaftlichen Disziplinen stammen kann (Tabelle 3–1). Dieses äußert sich in dieser Variante der Konkordanz darin, dass der Scopus-Code in der Regel auf zweistelligem Niveau angegeben ist, wogegen die detaillierte Scopus-Klassifikation der Felder aus vierstelligen Zahlen besteht. In der Konkordanz ist ein zweistelliger Code als rechtsseitige offene Maskierung zu verstehen. Das heißt, in der grundsätzlich vierstelligen Klassifikation der Scopus-Klassen sind jeweils alle Felder betroffen, die mit den angeführten Zweistellern beginnen. Beispielsweise sind mit 17 alle Felder im Bereich der Computertechnik angesprochen wie etwa 1701 (computer science, miscellaneous), 1702 (artificial intelligence), 1703 (computational theory and mathematics) usw.

Tabelle 3–1: Grobe Konkordanz zwischen Technikfeldern und Wissenschaftsfeldern

Technikfeld	SCOPUS Code
Elektrotechnik, elektrische Energie	21, 22, 3105
Audio-visuelle Technik	17, 22
Telekommunikation	17, 22, 31
Digitale Kommunikation	17, 22, 31
Grundlegende Kommunikationsprozesse	17, 22, 31
Computer-Technologie	14, 17, 18, 22, 20, 31
IT-basierte Managementmethoden	14, 17, 18, 20, 26
Halbleiter	22, 25, 31
Optik	22, 2504, 3107
Messtechnik	15, 16, 1711, 1913, 22, 3105
Analyse biologischer Materialien	13, 16, 2204, 24, 2704
Steuerung	17, 22, 26, 31
Medizintechnik	22, 27, 28, 29, 34, 36
Organische Feinchemie	16
Biotechnologie	11, 13, 1502, 16, 2204, 2304, 24, 27, 28, 30
Pharmazie	13, 15, 16, 24, 27, 28, 30, 34, 36
Polymere	15, 16, 22, 25
Lebensmittelchemie	11, 13, 15, 16, 2204, 27, 28, 30, 34, 36
Grundstoffchemie	15, 16, 25
Materialien, Metalle	15, 16, 22, 25, 3104
Oberflächentechnik	15, 16, 22, 25, 3104, 3110
Verfahrenstechnik	15, 16, 2204
Umwelttechnik	11, 15, 16, 17, 19, 21, 22, 23
Handhabung	17, 22
Werkzeugmaschinen	22
Motoren, Turbinen, Pumpen	21, 22
Textil- und Papiermaschinen	15, 22
Andere Spezialmaschinen	15, 22
Thermische Prozesse und Geräte	15, 21, 22, 31
Maschinenelemente	15, 22
Transport	22
Möbel, Spiele	17, 22
Andere Konsumgüter	11, 17, 22
Bauingenieurwesen	19, 22, 25

Tabelle 3–2: Feine Konkordanz zwischen Technikfeldern und Wissenschaftsfeldern

Inhalt	SCOPUS Code
Electrical machinery, apparatus, energy	2105, 2202, 2203, 2207, 2208, 2214, 3105
Audio-visual technology	1704, 1705, 1707, 2208, 2214
Telecommunications	1705, 1710, 1711, 2208
Digital communication	1705, 1710, 1711, 2208
Basic communication processes	1710, 1711, 2208
Computer technology	17##, 2206, 2207
IT methods for management	1404, 1802, 1803, 2002, 2605
Semiconductors	2208, 2504, 3104
Optics	2202, 2207, 2208, 2209, 2504, 3107
Measurement	1607, 1711, 1913, 2202, 2203, 2207, 2208, 2213, 3105
Analysis of biological materials	1303, 1305, 1308, 1310, 1313, 1502, 1602, 2204, 2402, 2704
Control	1706, 1707, 1710, 1711, 1712, 2202, 2203, 2207, 2208
Medical technology	2204, 27##, 28##, 34##, 35##, 3603, 3604, 3607, 3612, 3614, 3615, 3616
Organic fine chemistry	1605
Biotechnology	1101, 1106, 1305, 1307, 1311, 1312, 1502, 2204, 2304, 27##, 28##, 30##
Pharmaceuticals	27##, 28##, 30##
Macromolecular chemistry, polymers	1605, 2507
Food chemistry	1106, 1303, 2204, 2712,
Basic materials chemistry	1602, 1604, 1605
Materials, metallurgy	1505, 1503, 2202, 2203, 2211, 2215, 25##
Surface technology, coating	1505, 2508, 3110
Chemical engineering	15##
Environmental technology	23##
Handling	1706, 1707, 1708, 1709, 2206, 2207, 2208, 2209, 2210
Machine tools	2201, 2202, 2203, 2209, 2210, 2211
Engines, pumps, turbines	2102, 2104, 2202, 2203, 2210,
Textile and paper machines	22##
Other special machines	22##
Thermal processes and apparatus	1507, 21##, 2202, 2203, 2209, 2210
Mechanical elements	22##
Transport	22##
Furniture, games	17##, 2215, 2216
Other consumer goods	1701, 1704, 1706, 1707, 1712, 22##
Civil engineering	1905, 1906, 1907, 1908, 1909, 2205, 2206, 2211, 2215, 2216, 2503, 2505

In einer zweiten Variante der Konkordanz, der sogenannten "feinen" Version, wurden die Klassifikations-Codes von Scopus fast ausschließlich auf der Viersteller Ebene vergeben, d.h. es wurde eine sehr viel präzisere Zuordnung zwischen dem wissenschaftlichen und dem technischen Feld vorgenommen (Tabelle 3–2). Bei dieser zweiten Variante ist die Wahrscheinlichkeit einer präziseren Zuordnung zwischen Autor und Erfinder höher, grundsätzlich steigt aber auch die Wahrscheinlichkeit, dass zutreffende Zuordnungen wegen einer zu engen Abgrenzung ausgeschlossen werden.

4 Datenquellen für Patente und Publikationen

Für die Analysen in dem Projekt wurden auf Seiten der Patentanmeldungen die Datenbank PATSTAT sowie für Publikationen Scopus verwendet, deren Inhalte im Folgenden genauer beschrieben werden.

4.1 PATSTAT

Die Datenbank PATSTAT wird vom Europäischen Patentamt, Zweigstelle Wien, herausgegeben und deckt Daten von mehr als 80 Ländern ab mit insgesamt 70 Mio. Dokument-Eintragungen (Records). Die Datenbank ist als relationale Datenbank konzipiert und dient ausschließlich statistischen Analysen. In regelmäßigen Abständen von sechs Monaten werden Aktualisierungen zur Verfügung gestellt, um auch neuere Patentanmeldungen erfassen zu können. Die Datenbank muss auf einem eigenen Server installiert werden und kann nur mit üblichen Analysetools für relationale Datenbanken, insbesondere SQL-Befehlen, genutzt werden. Das Fraunhofer ISI hat seit mehreren Jahren eine entsprechende Version aufgebaut und um wichtige Angaben, die vom EPA nicht bereitgestellt werden, ergänzt. Dazu gehören für die vorliegende Fragestellung vor allem Adressdaten für Anmeldungen am Deutschen Patent- und Markenamt (DPMA). Adressdaten zu europäischen Patentanmeldungen liegen dagegen in der EPA-Version von PATSTAT vor.

Grundsätzlich wurden die Analysen in diesem Projekt auf Anmeldungen am Europäischen Patentamt (EPA) beschränkt, da hier üblicherweise Vergleiche zwischen europäischen Ländern vorgenommen werden. Im Falle Deutschlands wurden zusätzlich Anmeldungen am Deutschen Patent- und Markenamt (DPMA) analysiert, da hier aus der Sicht der Innovationspolitik ein besonderes Interesse besteht. Im Übrigen gibt es für DPMA-Anmeldungen aus Universitäten, wie oben in Abschnitt 2.2 beschrieben, einen Referenzdatensatz, der als Benchmark herangezogen werden kann.

In PATSTAT stehen für die vorliegende Fragestellung die folgenden Datenfelder für jede Patentanmeldung zur Verfügung:

1. Herkunftsland: Dabei wird üblicherweise das Land des Erfinders, nicht das Anmelderland genutzt {Hinze, 2004 3545 /id}. Dadurch ist insbesondere sichergestellt, dass bei einem multinationalen Unternehmen nicht das Land des Mutterunternehmens als Anmelderland erscheint, auch wenn die Erfindung bei einer Tochter in einem anderen Land gemacht wurde.
2. Anmelder: Im vorliegenden Fall ist nur bei einem Teil der Erfindungen aus Universitäten auch die Universität als Anmelder verzeichnet.

3. **Namen der Erfinder:** Bei Patenten werden neben den Nachnamen stets die vollen Vornamen der Erfinder verzeichnet.
4. **Erfinderadresse:** Zur Erfassung des Wohnsitzes des Erfinders kann in der Mehrzahl der Fälle die Erfinderadresse genutzt werden, in der auch Postleitzahlen enthalten sind und analysiert werden können. In einzelnen Fällen ist beim Erfinder auch die Adresse des Unternehmens angegeben, bei dem er arbeitet, was im vorliegenden Fall jedoch zu keinen methodischen Problemen führt.
5. **Prioritätsjahr:** Mit dem Prioritätsjahr ist das Jahr der ersten Anmeldung einer Erfindung bezeichnet. Eine Erfindung kann grundsätzlich in der Folge in mehreren Ländern angemeldet werden, was zum Aufbau sogenannter Patentfamilien führt. Maßgebliche Referenz ist dabei aber stets das Jahr der ersten Anmeldung, was bei Anmeldungen deutscher Herkunft in der Regel dem Anmeldedatum am DPMA entspricht. Es gibt allerdings zunehmend Fälle, wo deutsche Erfinder und Unternehmen ihre Erstanmeldung nicht beim DPMA, sondern direkt beim EPA vornehmen. Dieses sind Anmeldungen, die von vornherein auf die Abdeckung mehrerer Märkte zielen. Beim DPMA kann Deutschland wiederum als Zielland benannt werden, sodass auch hier ein Patentschutz erreicht werden kann.
6. **Patentklassifikation:** Auf jeder Patentanmeldung sind eine oder mehrere Patentklassifikationen angegeben, die von den Patentämtern vergeben werden. Die Klassifikationen sind in der Regel auf dem niedrigsten Klassifikationsniveau, den sogenannten Untergruppen, klassifiziert. Recherchen sind aber auch auf höheren Aggregationsniveaus wie Hauptgruppen oder Unterklassen möglich, wie sie zur Umsetzung der oben diskutierten Konkordanz benötigt werden. Grundsätzlich werden die Klassifikationen unmittelbar nach Eingang der Anmeldung beim Patentamt vergeben, sodass auf allen veröffentlichten Patentanmeldungen Klassifikationssymbole verfügbar sind.

4.2 Scopus

An dieser Stelle soll nicht im Detail auf die allgemeinen Inhalte der Datenbank Scopus eingegangen werden. Es werden hier ausschließlich die Kriterien diskutiert, die für das Matching zwischen Autoren und Erfindern relevant sind.

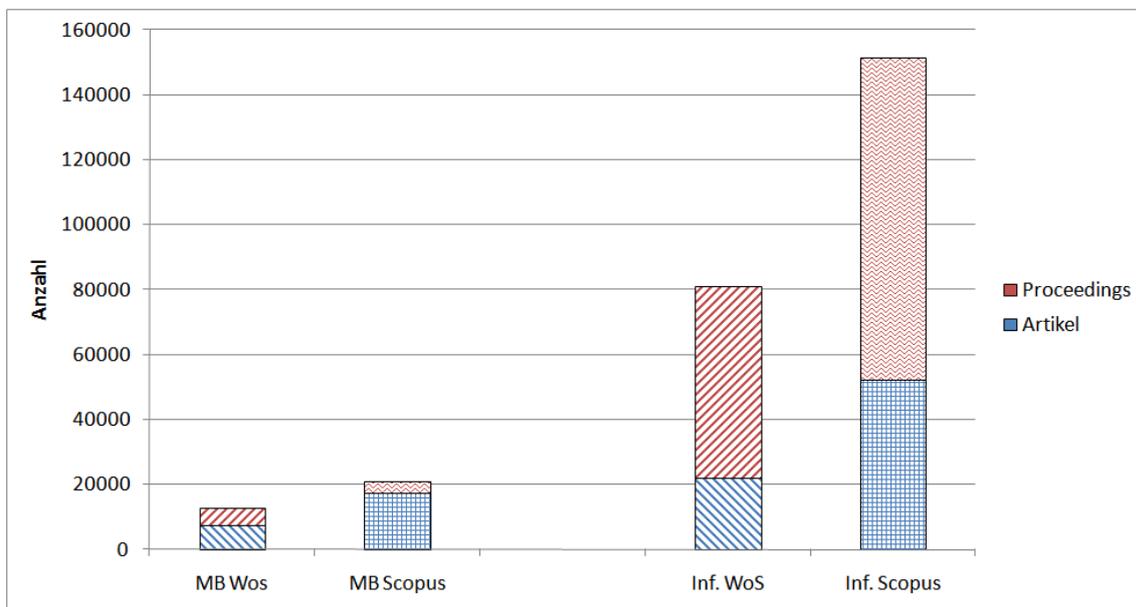
1. In den Scopus-Records gibt es ausschließlich Adressangaben zu der Einrichtung an der der Autor beschäftigt ist. Es fehlen private Adressangaben, da diese bei Publikationen in Fachzeitschriften in der Regel nicht systematisch abgefragt wird. Aktuell fehlen zwar in vielen Scopus-Dokumenten die Länderangaben in Fällen, wo in den Originaldaten aus den Zeitschriften nur die Adressen ohne Länderangaben zur Verfügung stehen. Davon sind vor allem Publikationen amerikanischer und japanischer Herkunft betroffen. Für Publikationen deutscher Herkunft ist dieses nur ausnahmsweise der Fall, sodass die vorliegenden Analysen davon nicht signifikant negativ betroffen sind.

2. Die Angabe der Einrichtung des Autors ist für die vorliegende Analyse relevant, da sich durch die Beschränkung auf Universitäten die Datenmengen erheblich reduzieren. Beispielsweise sind von den 78.000 Artikeln deutscher Herkunft, die für das Jahr 2009 in Scopus erfasst sind, 65.000 von Autoren aus Universitäten, was einem Anteil von 70% entspricht.
3. Bei der Namens Erfassung kann bei Scopus ab 1996 mit einem Anteil von mindestens 88% der Dokumente gearbeitet werden, da diese den voll ausgeschriebenen Vornamen der Autoren enthalten.
4. In der Adresse zu den Einrichtungen sind in Scopus die Postleitzahlen registriert. Im vorliegenden Projekt wurden die Postleitzahlen zur regionalen Abgrenzung eingesetzt, da sie im Falle Deutschlands, der Schweiz und Frankreichs regionale Einheiten gut abbilden und mit den verschiedenen Stellen der Postleitzahl eine zunehmende regionale Untergliederung verbunden ist. Problematischer wäre dies beispielsweise in Großbritannien, wo nur mit speziellen Schlüsseln eine Zuordnung zwischen Postleitzahl und regionaler Feinstruktur herstellbar ist. Mittelfristig ist beabsichtigt, in Scopus eine zusätzliche Tabelle einzubringen, bei der die Postleitzahlen in NUTS-Codes überführt werden (nomenclature des unités territoriales statistiques, Systematik der Gebietseinheiten für die Statistik). Es handelt sich dabei um eine hierarchische Systematik zur eindeutigen Identifizierung und Klassifizierung der räumlichen Einheiten in den Mitgliedstaaten der Europäischen Union. Die NUTS-Klassifikation ist auf verschiedenen Hierarchieebenen verfügbar, so bezeichnet die Ebene NUTS 2 mittlere Regionen, NUTS 3 kleinere Regionen und Großstädte.¹
5. Als zeitliche Referenz steht in Scopus das Publikationsdatum zur Verfügung, während – ähnlich wie bei WoS – das Datum der Einreichung eines Artikels nicht ausgewiesen wird.
6. Hinsichtlich der inhaltlichen Zuordnung können die 4-stelligen Klassifikationen der Zeitschriften verwendet werden, in denen die jeweilige Publikation veröffentlicht wurde. Erste Tests im Rahmen des Kompetenzzentrum Bibliometrie haben gezeigt, dass die bei Zeitschriften unvermeidbaren Mehrfach-Klassifikationen zu verschiedenen Themengebieten in Scopus sehr viel häufiger als in WoS eingesetzt werden. Außerdem gibt es eine Reihe von Fällen, bei denen Zeitschriften unzutreffend oder nicht spezifisch genug klassifiziert sind. Aus diesem Grunde hat der Datenbank-Produzent Elsevier einen Auftrag zur Verbesserung dieser Klassifikation an die Universität Leiden vergeben. Im vorliegenden Projekt musste mit der noch suboptimalen Variante gearbeitet werden. Ein wesentlicher Grund für die Auswahl von Scopus für die vorliegende Fragestellung ist die sehr viel umfangreichere Abdeckung technischer Disziplinen. Während die Abdeckung von Publikationen in der Elektrotechnik in WoS und Scopus in etwa gleichwertig ist

¹ Die früheren Ebenen NUTS-4 bzw. NUTS-5 wurden seit 2003 umbenannt in LAU-1 bzw. LAU-2 (Local Administrative Units).

und im Bauingenieurwesen nach den bisherigen Analysen die Abdeckung in WoS breiter ist als in Scopus, ist der Abdeckungsgrad in den Fächern Maschinenbau und Informatik in Scopus deutlich höher (Abbildung 4-1). Entscheidend ist dabei auch, dass in Scopus in den meisten Disziplinen sehr viel mehr Proceedings als in WoS abgedeckt sind. Dieser Publikationstyp spielt in den technischen Disziplinen eine sehr viel größere Rolle als in den Natur- oder Lebenswissenschaften, sodass für den vorliegenden Kontext die Verfügbarkeit von Namen für technische Bereiche sehr viel breiter als in WoS ist. Gerade bei universitären Forschern in den Disziplinen Maschinenbau und Informatik ist, wie frühere Studien gezeigt haben, das Patentieren besonders häufig {Meyer-Krahmer, 1998 574 /id}, sodass hier eine breite Abdeckung von erheblicher Relevanz ist.

Abbildung 4-1: Abdeckung von Artikel und Proceedings in den Feldern Maschinenbau und Informatik in Web of Science und Scopus, 2009



Quelle: Scopus, WoS, eigene Recherchen und Berechnungen

5 Länderanalysen

5.1 Deutschland

Für Patentanmeldungen aus Universitäten mit deutscher Herkunft wurden im Rahmen der Untersuchung Anmeldungen am Deutschen Patent- und Markenamt (DPMA) und am Europäischen Patentamt (EPA) betrachtet. Beide Ämter sind aus innovationspolitischer Sicht von Interesse, da die DPMA-Anmeldungen die Inlandspatentanmeldungen repräsentieren und das breiteste Bild der Patentanmeldungen aus deutschen Universitäten geben. Nur ein Teil dieser Anmeldungen mit besonders hohem ökonomischem Wert wird auch am EPA angemeldet mit dem Ziel, Patentschutz in mehreren europäischen Ländern zu erlangen. Da die Kosten des EPA-Verfahrens erheblich über denen des DPMA-Verfahrens liegen, ist von einer entsprechend höheren Erwartung der Anmelder in wirtschaftliche Erträge im Ausland auszugehen. Im Übrigen ist aus statistischer Sicht das europäische Patentamt der geeignete Rahmen, um Anmeldungen verschiedener Länder zu vergleichen, da sich die Rechtssysteme bei Inlandsanmeldungen erheblich unterscheiden und damit Vergleiche wenig aussagefähig wären.

5.1.1 DPMA-Anmeldungen

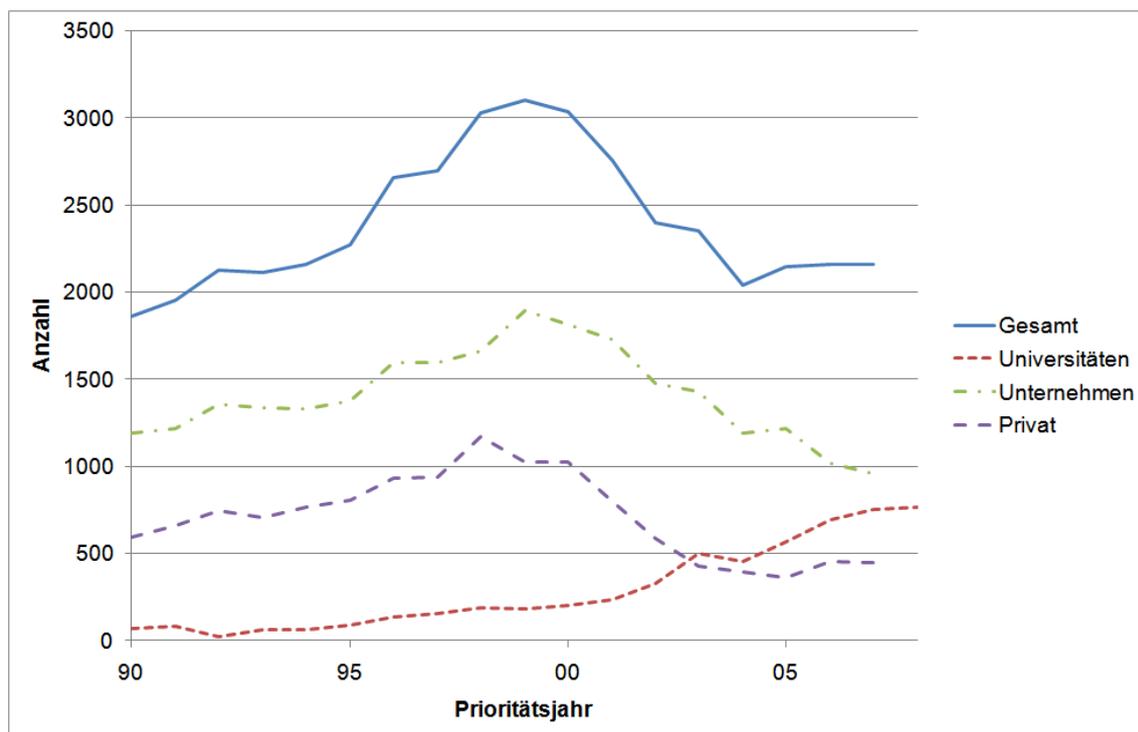
Für DPMA-Anmeldungen gibt es einen früheren Ansatz zur Bestimmung von Patentanmeldungen aus Universitäten. Dabei wurde zum einen nach Patentanmeldungen recherchiert, die von Universitäten als Anmelder hinterlegt wurden. Zusätzlich wurde nach Anmeldungen gesucht, bei denen einer der Erfinder oder Anmelder einen Professoren-Titel angibt. Dieser Titel ist in Datenbanken wie PATDPA recherchierbar. Durch die Recherche nach dem Professoren-Titel konnten auch Anmeldungen identifiziert werden, bei denen insbesondere Unternehmen als Anmelder fungierten. Die wesentliche methodische Schwäche dieses Verfahrens besteht darin, dass viele Erfinder aus Hochschulen keinen Professoren-Titel haben oder aber diesen nicht angeben. Hier kann nur bei Anmeldungen mit Universitäten als Anmelder ermittelt werden, bei welcher Quote keiner der Erfinder einen Professoren-Titel hat, woraus sich ein Anteil von etwa 50% mit leichten jährlichen Schwankungen ergibt. In dem früheren Ansatz wurde konservativ gerechnet, dass diese Quote auch auf Anmeldungen durch Unternehmen und Privatpersonen zutrifft. Es wurde eine entsprechende Hochrechnung vorgenommen, wobei der kritische Punkt die Annahme der Quote der Hochschul-Erfinder ohne Professoren-Titel ist.² Bei aller Unsicherheit ist es über diesen Ansatz möglich, die

² Zu den Einzelheiten dieses Verfahrens vergleiche Frietsch et al. {Frietsch, 2010 5157 /id /d}.

zahlenmäßige Größenordnung der Patentanmeldungen aus Universitäten zu ermitteln und wesentliche Trends zu beschreiben. Die Referenz auf einzelne Dokumente ist nur bei Anmeldungen bei Universitäten als Anmeldern oder aber mit Personen mit Professoren-Titel als Erfinder möglich, sodass keine Aussage gemacht werden kann, welches wirklich die zutreffenden Dokumente im Einzelnen sind.

In Abbildung 5-1 sind die Anmeldestruktur und der generelle Trend der Universitätsanmeldungen nach diesem Ansatz dokumentiert. Er kann hinsichtlich der Größenordnung der Anmeldezahlen und der Trends als Benchmark verwendet werden, nicht aber als exakte Referenz. An diesem Datensatz lässt sich zumindest ablesen, dass trotz steigender Anteile der durch Universitäten angemeldeten Erfindungen die Zahl der durch Unternehmen und Privatpersonen angemeldeten Universitätserfindungen erheblich ist und in keinem Fall die Anmelde-Analyse allein für statistisch relevante Aussagen hinreichend ist.

Abbildung 5-1: Anmeldungen aus Universitäten am DPMA nach den Kriterien von Universitäten als Anmelder oder Professoren als Erfinder



Quelle: Frietsch et al. {Frietsch, 2010 5157 /id /d}

Ziel des in dieser Untersuchung verfolgten Vorgehens einer Abgleichung von Erfinder- und Autorennamen ist die Ermittlung einer möglichst großen Zahl von Anmeldungen aus Universitäten, mit einem möglichst hohen Grad an Genauigkeit, was bei derartigen Retrieval-Verfahren üblicherweise mit den Parametern "Precision" und "Recall"

bewertet wird. Precision gibt den Anteil der relevanten Dokumente unter den gefundenen an, Recall den Anteil der relevanten Dokumente an der Gesamtzahl aller relevanten Dokumente, inklusiv der nicht gefundenen relevanten Dokumente, an. Eine genauere Definition wird weiter unten gegeben.

Für die Analyse von Patentanmeldungen aus Universitäten wurden in dem Projekt die folgenden Fälle betrachtet:

- deutsche Universitätsanmeldungen am DPMA,
- deutsche Universitätsanmeldungen am EPA,
- Schweizer Universitätsanmeldungen am EPA,
- französische Universitätsanmeldungen am EPA.

Für die einzelnen Datensätze wurde der Zeitraum von 1996 bis 2006 untersucht, um langfristige Trends erfassen zu können und den Einfluss der veränderten Datenbankabdeckung von Scopus im Zeitverlauf nachvollziehen zu können. Im Einzelnen wurde der Effekt der folgenden Selektionskriterien geprüft:

1. keine Einschränkung (lediglich Universitäten als Einrichtungen der Autoren),
2. grobe Konkordanz,
3. feine Konkordanz,
4. 1-stellige PLZ,
5. 2-stellige PLZ

sowie die Kombinationen

6. 1-stellige PLZ mit grober und feiner Konkordanz,
7. 2-stellige PLZ mit grober und feiner Konkordanz.

Variiert wurde außerdem die Zahl der Jahre, die für die Ermittlung der Autoren in Scopus zugrundegelegt wurde. Beispielsweise wurde für das Prioritätsjahr 2006 das Publikationsjahr 2007 oder aber die beiden Publikationsjahre 2007 und 2008 als Referenz verwendet. Damit ist vollständig sichergestellt, dass die Autoren auch nach 2006 noch an der Universität beschäftigt waren. Durch das Hinzufügen eines zusätzlichen Publikationsjahres wird der Datensatz durch solche Forscher vergrößert, die selten oder zumindest selten in Zeitschriften publizieren, die von Scopus erfasst sind.

Bei diesem Ansatz kann es also nicht darum gehen, alle relevanten Dokumente zu erfassen, sondern möglichst viele mit hoher Präzision.

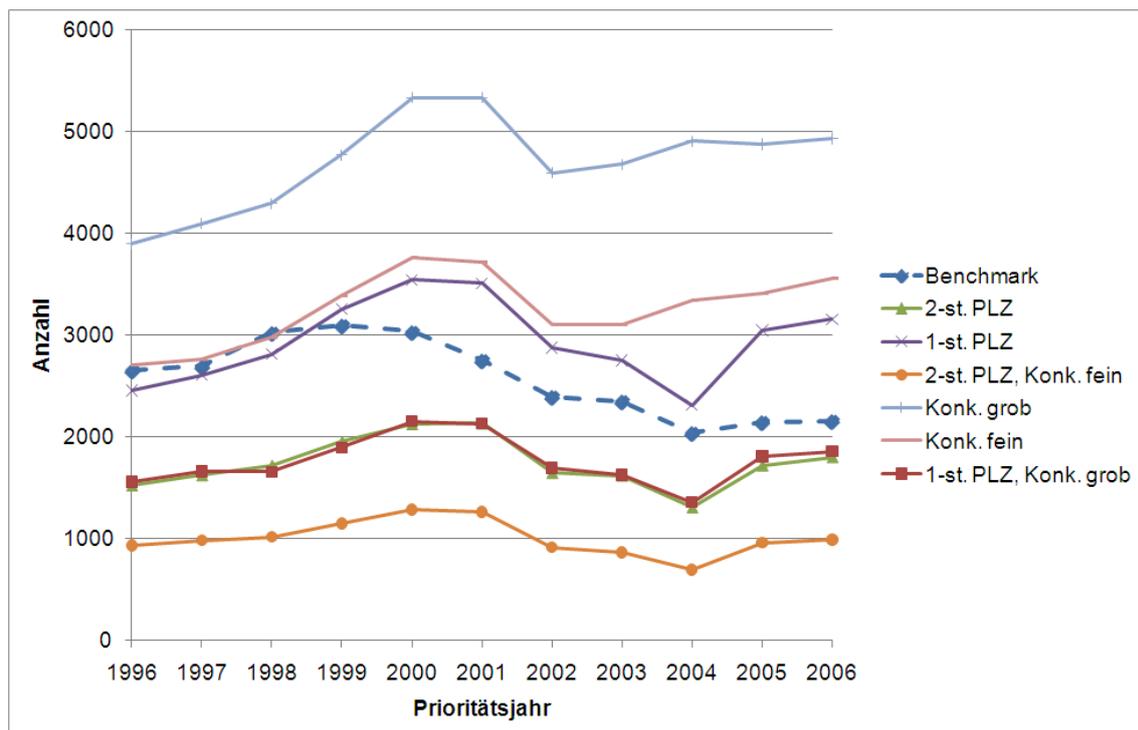
Das Kriterium "ohne Einschränkung" bedeutet, dass lediglich die voll ausgeschriebenen Vor- und Nachnamen von Erfindern und Autoren abgeglichen wurden, wobei aller-

dings auf Seiten der Autoren nur solche aus Universitäten berücksichtigt wurden. Eine allgemeine Verwendung aller Autoren in Scopus ohne diese Einschränkung auf Universitäten, hätte zu einer deutlich größeren Autorenmenge und damit verbunden zu größeren Unschärfen beim Abgleich geführt.

Bei den Selektionskriterien "grobe" und "feine" Konkordanz wurden die in Abschnitt 3.2 dokumentierten Konkordanzlisten zwischen Technikfeldern und Wissenschaftsgebieten benutzt. Durch die Verwendung 1- und 2-stelliger Postleitzahlen wurde ein engerer oder weiterer Radius definiert, innerhalb dessen der Erfinder von der Einrichtung des Autors entfernt wohnen kann.

Es mussten diese verschiedenen Kriterien geprüft werden, da vorab nicht vorhersagbar war, welches Kriterium stärker greift und inwieweit durch die Einführung spezifischer Kriterien adäquate Größenordnungen des Ergebnisdatensatzes erreicht werden.

Abbildung 5-2: Ermittelte DPMA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien (jeweils 2 Publikationsjahre pro 1 Prioritätsjahr)



Quelle: Scopus, PATSTAT, eigenen Recherchen und Berechnungen

Im Ergebnis ergeben sich für Universitätsanmeldungen deutscher Herkunft am DPMA zwischen 1996 und 2006 für die verschiedenen Abgrenzungskriterien die Verläufe nach Abbildung 5-2. Danach zeigen sich im Vergleich zum Benchmark-

Datensatz grundsätzlich ähnliche Trends mit einem Maximum um das Jahr 2000 herum, einem Absinken bis etwa 2004 und einer anschließenden Stabilisierung oder leichten Erholung. Im Detail liegt das Maximum des Benchmark-Datensatz etwas früher als bei den Verläufen nach den verschiedenen Matching-Verfahren zwischen Erfindern und Autoren. Bei Letzteren setzt der Wiederanstieg etwas eher als beim Benchmark-Datensatz ein.

Es zeigt sich eine relativ klare Abstufung der erzielten Mengen nach den verschiedenen Selektionskriterien, was in Abbildung 5-3 beispielhaft für das Jahr 2003 dargestellt ist. Dabei wird bei dem Matching "ohne Einschränkung", bei der nur die Zugehörigkeit der Autoren zu einer Universität Voraussetzung ist, die größte Datenmenge erzielt, bei der aber auch die Zahl der fehlerhaften Identifikationen sehr hoch sein dürfte. Es hat sich bei Stichprobenuntersuchungen gezeigt, dass auch relativ seltene Namen zu Homonymen führen können, also zu Identifikation von verschiedenen Personen mit gleichem Namen. Dieses ist letztlich darauf zurückzuführen, dass in einem Jahr etwa 43.000 Erfindernamen am DPMA mit 160.000 Autorennamen in Scopus abgeglichen werden müssen. In jedem Fall ist dieses einfache Selektionskriterium bei weitem nicht ausreichend. Es ergibt sich eine deutliche Reduzierung der Menge auf fast die Hälfte, wenn zusätzlich die grobe Konkordanz, also ein grobes Matching zwischen Technik- und Wissenschaftsfeldern, eingeführt wird. Die Anwendung der feineren Konkordanz erbringt demgegenüber eine Reduktion von weiteren 50%. Nach dem Kriterium der 1-stelligen Postleitzahl liegt der Ertrag in absoluten Zahlen in etwa derselben Größenordnung, etwa 10% niedriger als für die feine Konkordanz. Es ist von einer erheblichen Überschneidungsmenge zwischen den Kriterien der feinen Konkordanz und der 1-stelligen Postleitzahlen auszugehen, aber bereits die Verknüpfung zwischen den Kriterien der 1-stelligen Postleitzahlen und der groben Konkordanz führt gegenüber dem alleinigen Kriterium der 1-stelligen Postleitzahl zu einer weiteren Einschränkung der Datenmenge von 70%. Von daher impliziert die Einführung jedes zusätzlichen Kriteriums erwartungsgemäß zu einer mehr oder weniger deutlichen Reduzierung des Datensatzes.

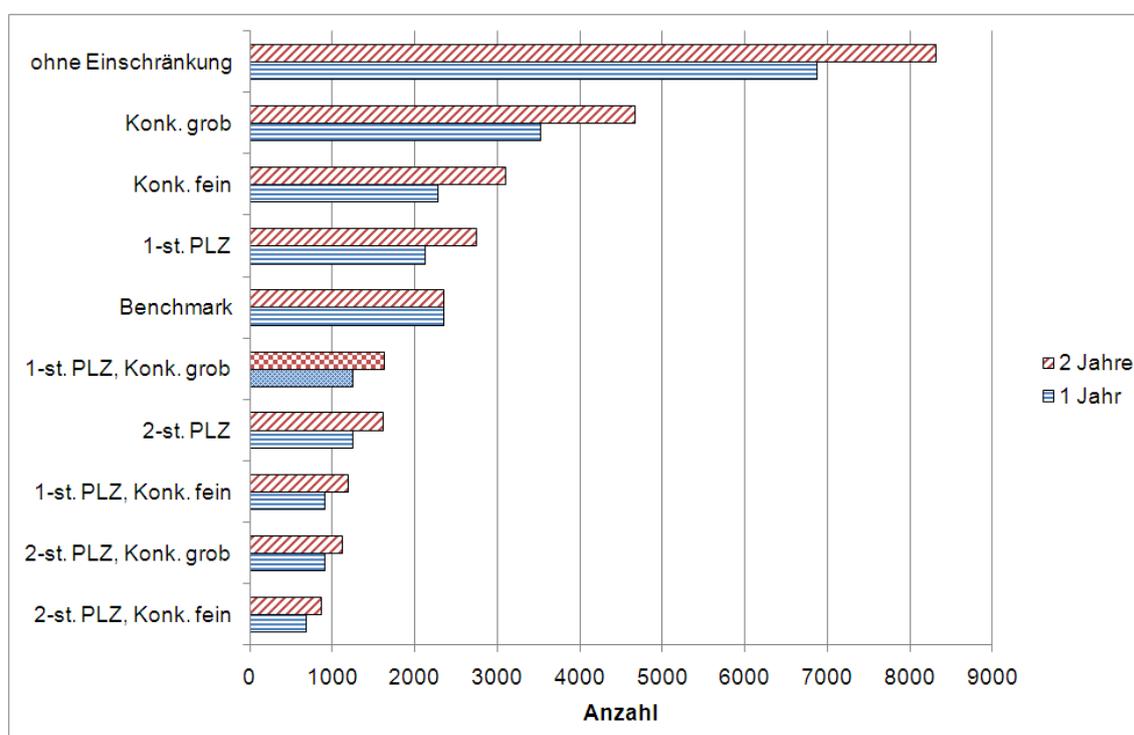
Ein weiterer Aspekt ist schließlich, dass sich bei der Berücksichtigung von Autoren aus zwei Referenzjahren gegenüber einem, ebenfalls eine relevante Erhöhung der Datenmenge ergibt, die je nach Kriterium zwischen 20 und 30% liegt.

Anzumerken ist, dass für die Berechnung der Zeitverläufe nach sieben Selektionskriterien sowie ein oder zwei Referenzjahre für jeden Fall (Anmeldungen der Universitätspatente aus einem Land an einem spezifischen Amt) 154 Datenpunkte zu ermitteln sind, für die jeweils der Abgleich der 43.000 Erfindernamen und der 160.000 Autorennamen verbunden mit dem jeweiligen Selektionskriterium durchgeführt werden muss.

Diese relativ aufwändige Analyse lässt sich mit der installierten bibliometrischen Version von Scopus mit rund einer Stunde Rechenzeit, also in einem überschaubaren Zeitrahmen, durchführen. Die Zeiten liegen bei kleineren Ländern deutlich niedriger.

Aus dem Vergleich der Zeitreihen nach den verschiedenen Selektionskriterien mit dem Verlauf des Benchmark-Datensatzes ergibt sich bei älteren Jahrgängen eine relativ große Nähe zu den Kriterien der 1-stelligen Postleitzahlen sowie der feinen Konkordanz, am aktuellen Rand scheint das Kriterium der 1-stelligen Postleitzahl kombiniert mit der groben Konkordanz treffender zu sein.

Abbildung 5-3: Für das Jahr 2003 ermittelte DPMA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien (mit 1 und 2 Publikationsjahren)

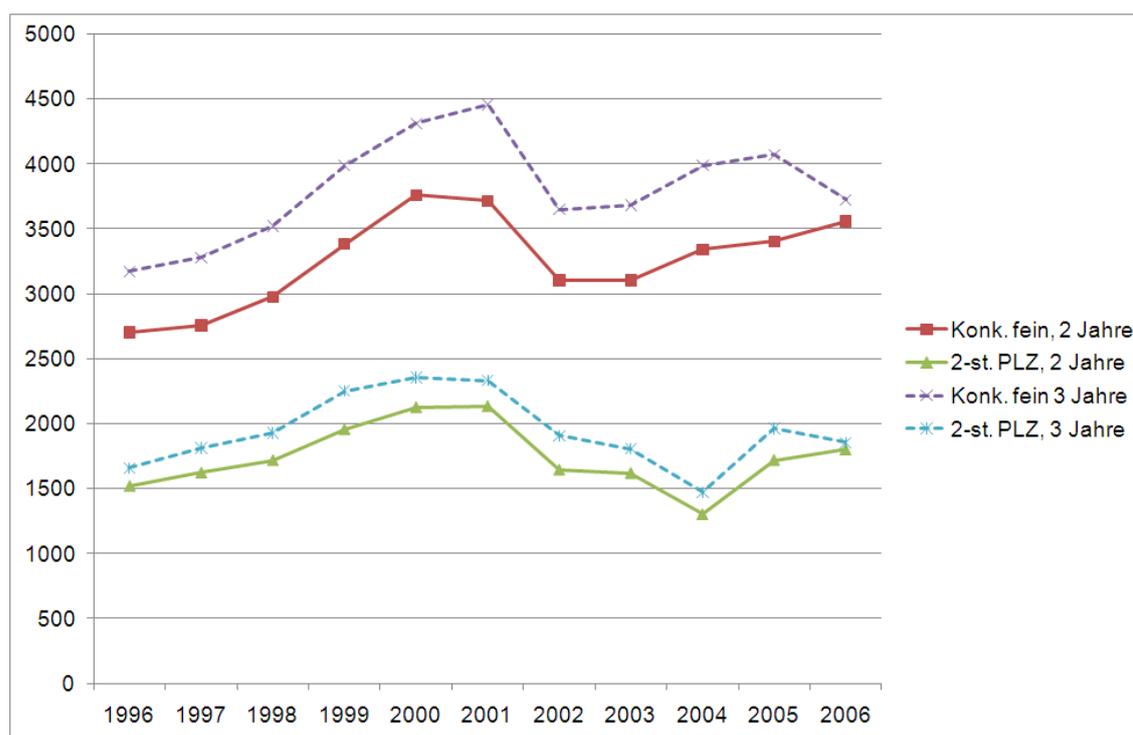


Quelle: Scopus, PATSTAT, eigene Recherchen und Berechnungen

Es wurde für deutsche Anmeldungen am DPMA zusätzlich untersucht, welche Auswirkungen sich ergeben, wenn der Zeitraum für die Erfassung von Autoren in Scopus weiter erhöht wird. Es ist grundsätzlich davon auszugehen, dass die Zahl der Autoren steigt und auch solche eingeschlossen werden, die nur selten in Zeitschriften publizieren, die in Scopus registriert sind. Durch die Erweiterung des Zeitraums steigt auch die Möglichkeit des Auftretens von zusätzlichen Homonymen, also von falschen Zuordnungen zu Universitätsanmeldungen. Die Ergebnisse von einer Zeitreihenanalyse mit

einer Erfassung von drei Jahren im Vergleich zu einer solchen Analyse für zwei Jahre ist in Abbildung 5-4 dokumentiert.

Abbildung 5-4: Ermittelte DPMA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien (jeweils 2 oder 3 Publikationsjahre pro 1 Prioritätsjahr)

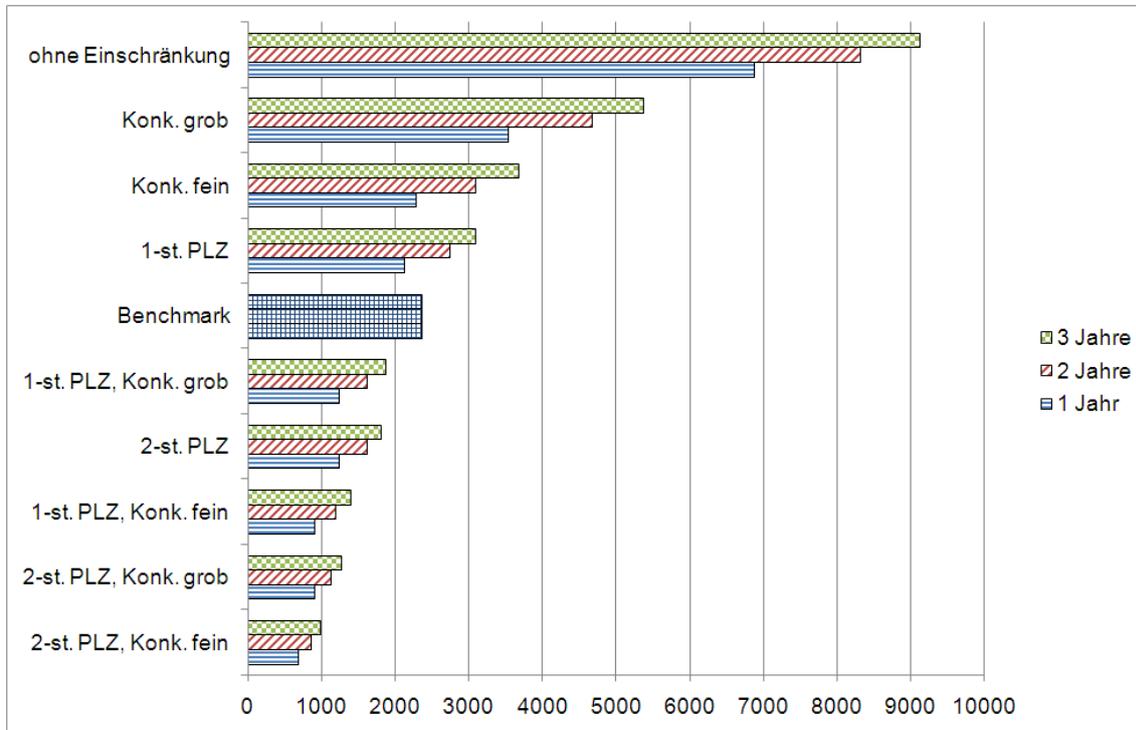


Quelle: Scopus, PATSTAT, eigene Recherchen und Berechnungen

Die Verläufe für die zwei ausgewählten Selektionskriterien der feinen Konkordanz und der 2-stelligen Postleitzahl sind bei dem 2- und dem 3-jährigen Publikationszeitraum sehr ähnlich. Bei der feinen Konkordanz werden im Schnitt etwa 18% an zusätzlichen Dokumenten ermittelt, bei den 2-stelligen Postleitzahlen 12%. Am aktuellen Rand gehen diese Quoten jedoch auf 5 bzw. 3% zurück, was einen besseren Erfassungsgrad von Scopus indiziert, sodass ein zusätzliches Jahr der Autorenerfassung nur noch begrenzte Zusatzerträge erzielt.

Bei einer analogen Darstellung zu Abbildung 5-3 ergibt sich bei Einschluss der 3-Jahres-Variante die erwartete Situation nach Abbildung 5-5, wonach ein größerer Erfassungszeitraum der Autoren zu zusätzlichen Treffern führt, der zusätzliche Ertrag jedoch mit jedem zusätzlich Jahr abnimmt. Insofern ist es vor allem am aktuellen Rand sinnvoller, ausschließlich mit einem 2-Jahresfenster zu arbeiten, weil dadurch die Aktualität der Zeitreihen verbessert werden kann.

Abbildung 5-5: Für das Jahr 2003 ermittelte DPMA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien (mit 1, 2 und 3 Publikationsjahren)



Quelle: Scopus, PATSTAT, eigene Recherchen und Berechnungen

Allerdings ergibt sich aus diesen Überlegungen noch kein Hinweis auf den Recall der verschiedenen Kriterien, also der Identifizierung relevanter Dokumente in Bezug auf einen präzis definierten Zieldatensatz, da der Benchmark-Datensatz ebenfalls mit Unsicherheiten behaftet ist. Ein präziser definierter Zieldatensatz ist zumindest die Zahl der von Universitäten angemeldeten Erfindungen, die sicher als Universitätspatente eingestuft werden können. Im Jahre 2003 waren dies am DPMA 190 Anmeldungen, im Jahre 2006 370, sodass von statistisch relevanten Teildatensätzen auszugehen ist. Hier kann zumindest der Recall bestimmt werden, also der Anteil der identifizierten Dokumente an den relevanten oder korrekten Dokumenten des Zieldatensatzes. Diese Beziehung wird üblicherweise in der folgenden Weise definiert:

$$\text{Recall} = t_p / (t_p + f_n)$$

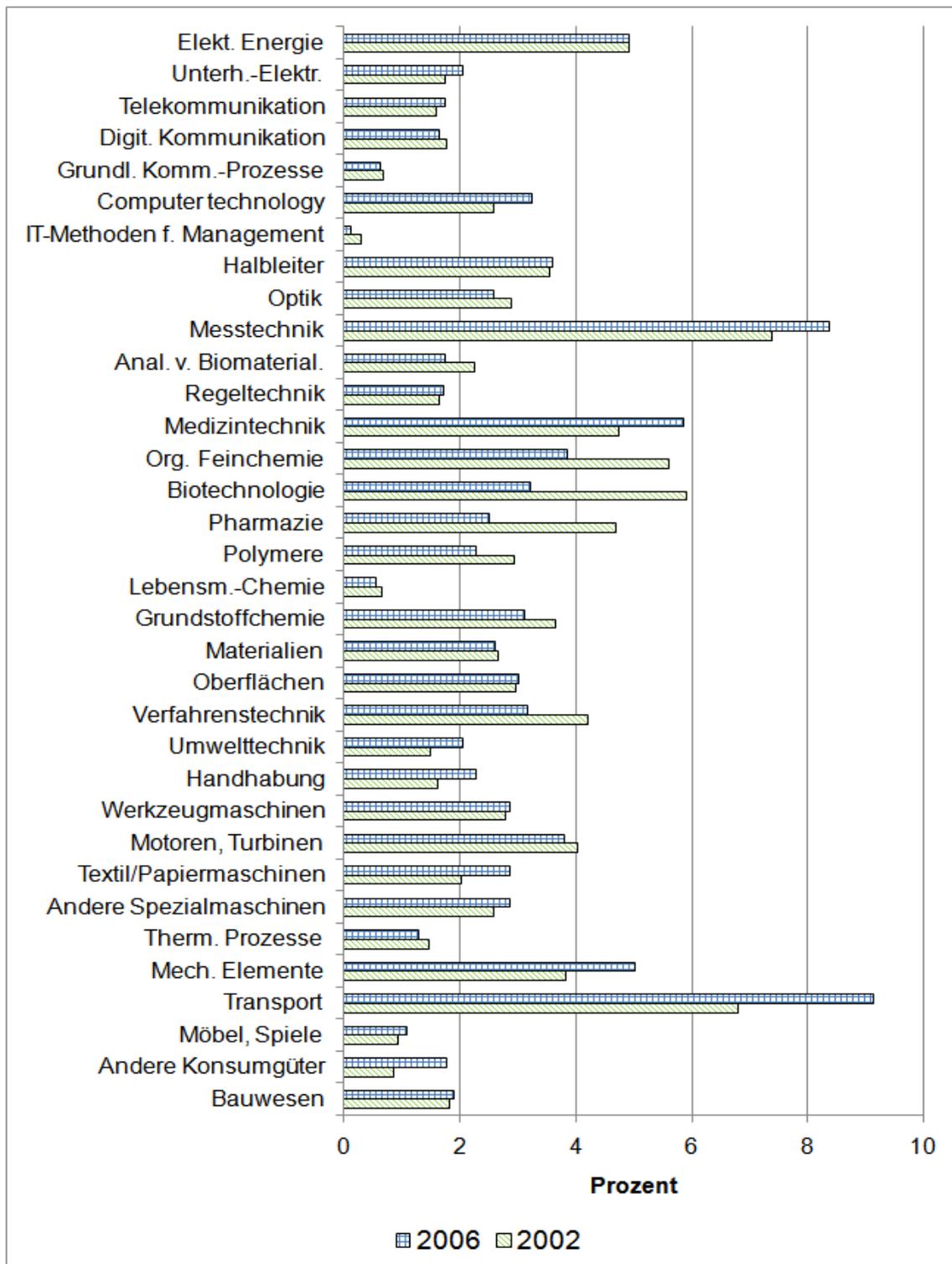
mit t_p (true positive) und f_n (false negative). False negative sind dabei die Elemente, die zutreffend sind, aber nicht identifiziert wurden. $t_n + f_n$ beschreibt also die Menge der zutreffenden, relevanten Dokumente. t_p und f_n könnten deutsch als positive Treffer und fehlende, nicht identifizierte Treffer bezeichnet werden.

Für den Datensatz, der durch Universitäten angemeldeten Erfindungen, ergibt sich zunächst für das Kriterium "ohne Einschränkung" eine relativ gute Recall-Rate von mehr als 80%, wobei aber bei diesem Kriterium insgesamt von einer großen Menge fehlerhaft identifizierter Dokumente auszugehen ist. Bei der Einschränkung durch weitere Selektionskriterien etwa durch eine Postleitzahl oder die feine Konkordanz wird der Recall erheblich reduziert auf 75 bis 66%, was nicht zufriedenstellen kann. Es fällt weiter auf, dass die Recall-Rate für den Datensatz aus dem Jahr 2006 deutlich besser als die für die Menge aus dem Jahre 2003 ist, was auf eine verbesserte Abdeckung relevanter Autoren durch die Datenbank Scopus im Zeitverlauf schließen lässt. In der Tat hat sich die Zahl der in Scopus erfassten Publikationen zwischen 2003 und 2006 um rund 30% erhöht, was auch für die Publikationen deutscher Herkunft gilt.

Als Zwischenfazit bleibt festzuhalten, dass keines der Matching-Kriterien für sich genommen zu einem ausreichenden Recall führt. Hinsichtlich der Abdeckung der Autoren durch Scopus kann zumindest für die Zeit ab etwa 2005 von einer verbesserten Abdeckung ausgegangen werden, die sich durch ein ständiges Wachstum der Datenbank in den kommenden Jahren weiter verbessern wird. Bei den Selektionskriterien erweist sich keines als zufriedenstellend, da schon bei der groben Konkordanz oder der 1-stelligen Postleitzahl die Recall-Rate unbefriedigend ist. Von daher muss es für einen abschließenden Suchalgorithmus zu einer Kombination verschiedener Kriterien kommen.

Eine weitere Möglichkeit, sich einen Eindruck von der Aussagefähigkeit der identifizierten Datensätze zu verschaffen ist es, ein Profil der Technikfelder der Universitätsanmeldungen zu erstellen. Bei einer prozentualen Aufteilung der Anmeldezahlen auf die 34 Felder der in Abschnitt 3.2 WIPO-Klassifikation zeigen sich für das Jahr 2002 mit dem Kriterium der 1-stelligen Postleitzahl Anmeldeschwerpunkte in der Messtechnik, der Medizintechnik, der Organischen Feinchemie, der Biotechnologie, der Pharmazie sowie der Verfahrenstechnik. Aus dem im unteren Diagrammbereich abgebildeten Bereich des Maschinenbaus (im weiten Sinne) sind vor allem Motoren und Turbinen, mechanische Elemente sowie Transport deutlich vertreten. Diese Betonung von eher forschungsintensiven Feldern durch Universitätspatente entspricht den Erwartungen und den Analysen mithilfe des früheren Ansatzes (vgl. Abbildung 5-6).

Abbildung 5-6: Technische Felder der DPMA-Anmeldungen aus deutschen Universitäten (Kriterien: 2 Publikationsjahre, 1-st. PLZ)



Quelle: Scopus, PATSTAT, eigene Recherchen und Berechnungen

Bei dem Profil für die durch den Matching-Ansatz ermittelten Universitätsanmeldungen für 2006 im Vergleich zu 2002 zeigt sich eine relativ starke Betonung der Maschinenbau-Felder "Mechanische Elemente" und "Transport". Diese Beobachtung könnte mit einer breiteren Abdeckung von Maschinenbau-Publikationen durch Scopus zusammenhängen, die wiederum mit einem besonders deutlichen Wachstum der Abdeckung der Ingenieurwissenschaften in Scopus zurückzuführen sein dürfte.

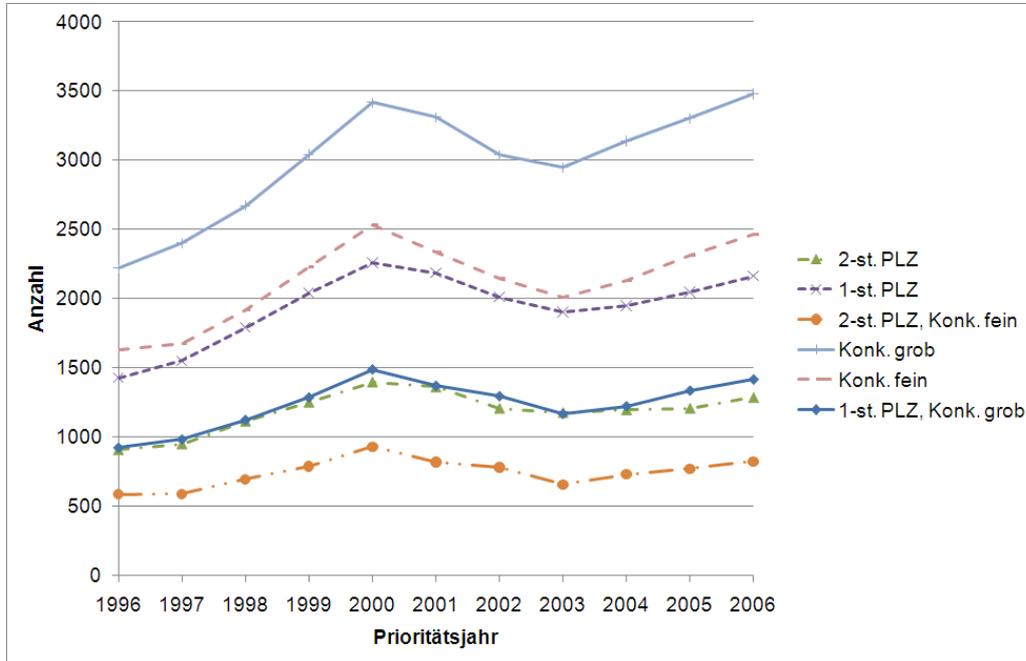
Ein denkbarer Grund für den unbefriedigenden Recall könnte die unzureichende Erfassung von Autoren sein, die Erfindungen zum Patent anmelden. Hier liegt eine denkbare Verbesserung in der Berücksichtigung eines zusätzlichen Publikationsjahres, also statt zwei drei Jahre.

5.1.2 EPA-Anmeldungen

Mit der Recherche von Universitätsanmeldungen deutscher Herkunft am Europäischen Patentamt wird ein Bereich betreten, der mit dem früheren Analyse-Ansatz nicht mehr beschreibbar war, da bei EPA-Anmeldungen der Professorentitel weniger systematisch als bei DPMA-Anmeldungen erfasst wird. Die Frage, ob mit der Matching-Analyse Universitätsanmeldungen in einer sinnvollen Größenordnung ermittelt werden, lässt sich letztlich nur aus der Relation zwischen DPMA- und EPA-Anmeldungen ableiten. Danach werden je nach Selektionskriterium 63 bis 73% nach dem Kriterium der groben Konkordanz oder der 2-stelligen Postleitzahl von den DPMA-Anmeldungen auch am EPA hinterlegt, was um 10 bis 15% über dem Durchschnitt für Deutschland insgesamt liegt. Dieses relativ höhere Niveau der EPA-Anmeldungen ist allerdings bei den forschungsintensiveren Erfindungen aus Universitäten erwartbar, sodass das Niveau der ermittelten EPA-Anmeldungen realistisch erscheint. Bei der Betrachtung des Verlaufs der Anmeldezahlen zwischen 1996 und 2006 ergibt sich erwartungsgemäß ein ähnlicher Trend wie am DPMA, allerdings mit einem stärkeren Anstieg der Anmeldezahlen seit 2003 (Abbildung 5-7).

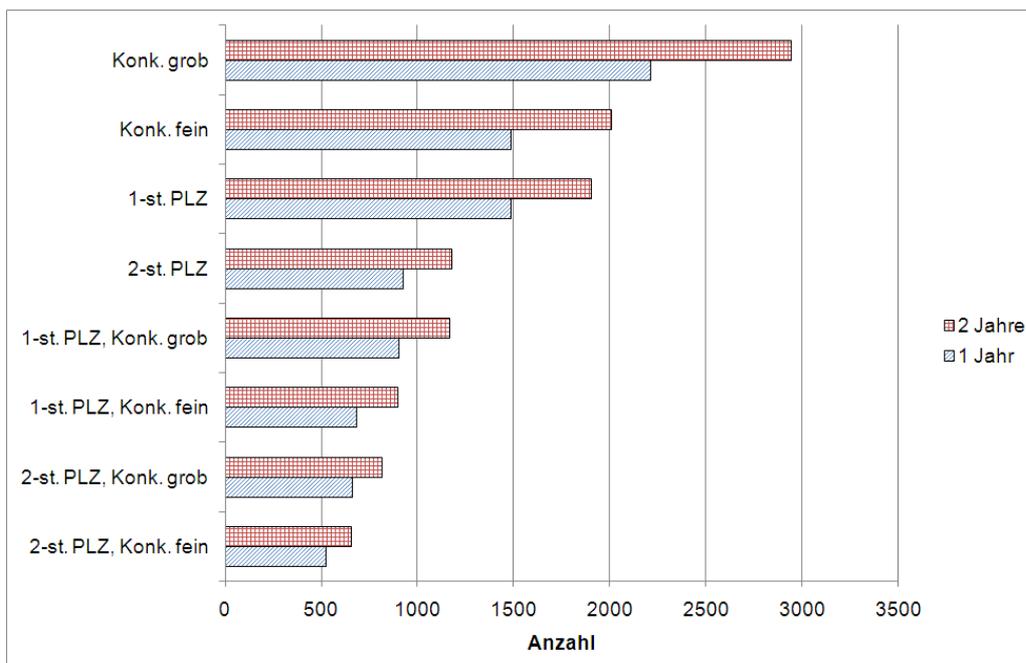
Bei der Wirkung der verschiedenen Selektionskriterien auf die Reduzierung der Ergebnismenge zeigen sich am EPA ähnliche Effekte wie beim DPMA. Die geringste Einschränkung ergibt sich aus der groben Konkordanz, es folgen die feine Konkordanz und der 1-stellige Postleitzahlenbereich (Abbildung 5-8). Beim EPA sind die nächste Einschränkungsstufe die 1-stelligen Postleitzahlen, die in etwa auf der Höhe des kombinierten Kriteriums 1-stellige Postleitzahl und grobe Konkordanz liegen. Auch am DPMA war die Größenordnung dieser beiden Kriterien ähnlich, wenn auch mit einer etwas stärkeren Einschränkung durch die 2-stelligen Postleitzahlen. Im Wesentlichen werden die Ergebnisse für das DPMA am EPA reproduziert, etwa auch beim höheren Erfassungsgrad bei einer Autorenextraktion von zwei Jahren gegenüber einem Jahr.

Abbildung 5-7: Ermittelte EPA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien (jeweils 2 Publikationsjahre pro 1 Prioritätsjahr)



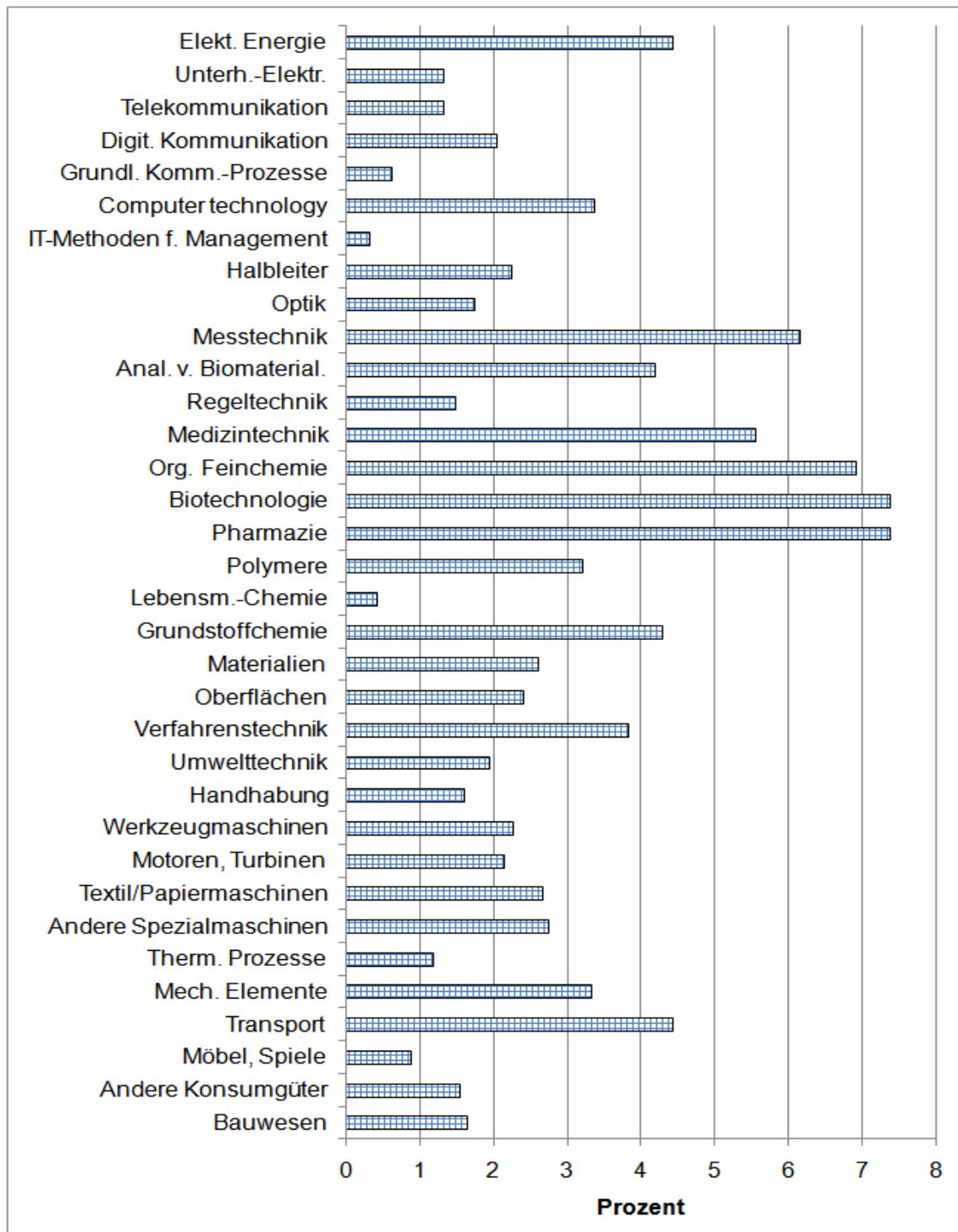
Quelle: Scopus, PATSTAT, eigene Recherchen und Berechnungen

Abbildung 5-8: Für das Jahr 2003 ermittelte EPA-Anmeldungen aus deutschen Universitäten nach verschiedenen Selektionskriterien (mit 1 und 2 Publikationsjahren)



Quelle: Scopus, PATSTAT, eigenen Recherchen und Berechnungen

Abbildung 5-9: Technische Felder der EPA-Anmeldungen aus deutschen Universitäten im Jahr 2006 (Kriterien: 2 Publikationsjahre, 1-st. PLZ)



Quelle: Scopus, PATSTAT, eigenen Recherchen und Berechnungen

Der bemerkenswerteste Unterschied zwischen der DPMA- und der EPA-Analyse für Universitätsanmeldungen deutscher Herkunft ist der höhere Recall bei den Anmeldun-

gen, die durch Universitäten registriert werden. Beispielsweise liegt der Recall bei dem Selektionskriterium grobe Konkordanz und zwei Publikationsjahren in Scopus für das DPMA bei 76%, am EPA bei 84%. Dieses könnte damit zusammenhängen, dass am EPA in der Regel Erfindungen in forschungsintensiveren Gebieten registriert werden als am DPMA. Dieses betrifft insbesondere die Felder Messtechnik, Medizintechnik, Feinchemie oder Biotechnologie, während der Maschinenbau weniger stark vertreten ist (Abbildung 5-9). So haben die Anmeldungen im Bereich Maschinenelemente am DPMA im Jahr 2006 einen Anteil von 5%, am EPA von 3,3%. Offensichtlich ist der Erfassungsgrad des Maschinenbaus durch Publikationen in Scopus wesentlich für den Recall des Matching-Verfahrens.

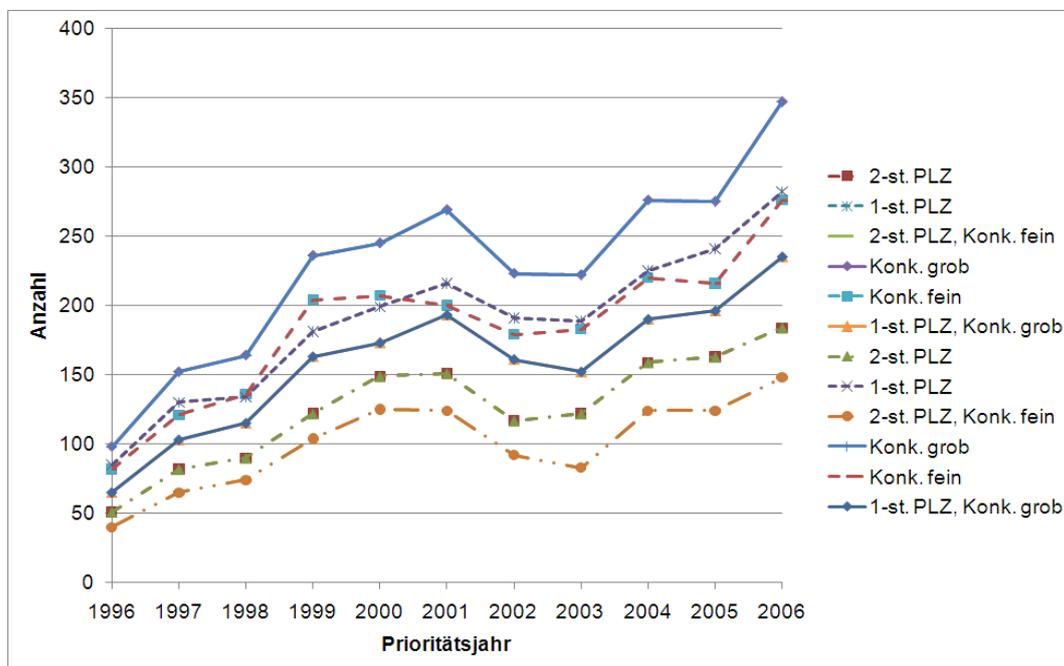
5.2 Schweiz, EPA-Anmeldungen

In diesem Projekt wurden EPA-Anmeldungen Schweizer Herkunft genauer betrachtet, weil sich bei einer Analyse aus dem Jahr 2004 gezeigt hat, dass für dieses Land die Anwendung des Kriteriums des Professorentitels wie für deutsche Anmelder offensichtlich unzureichend war {Schmoch, 2004 4262 /id /ft ": 72ff)}. Durch die vorliegende Untersuchung soll somit geprüft werden, ob mit dem Matching-Ansatz bessere Ergebnisse erzielbar sind.

Auch für die EPA-Anmeldungen Schweizer Herkunft zeigt sich ein ähnlicher Verlauf der Anmeldezahlen für die verschiedenen Abgrenzungskriterien mit einem steigenden Trend bis etwa 2001, einem vorübergehenden Rückgang und einem erneuten Anstieg ab 2003 (Abbildung 5-10). Allerdings ist erwartungsgemäß das absolute Niveau der Schweizer Anmeldungen um ein Vielfaches niedriger als das der deutschen, weshalb die Ausschläge im Zeitverlauf relativ gesehen deutlicher ausfallen.

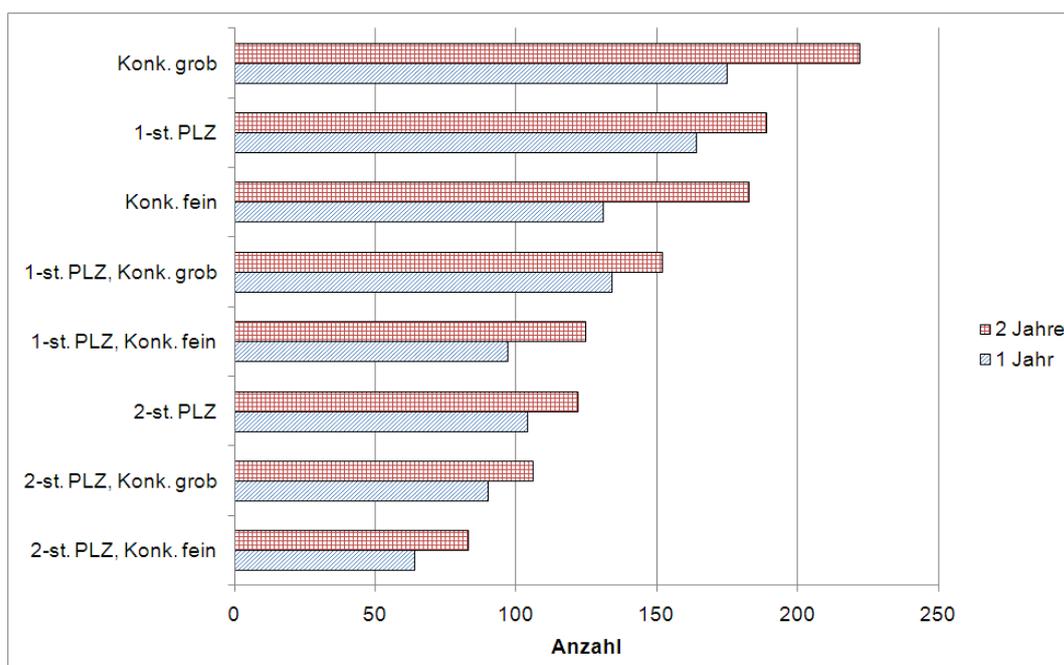
Auch hier ergeben sich ähnliche Abstufungen wie für deutsche Erfinder am EPA, allerdings wirkt die regionale Einschränkung durch die Postleitzahlen in der Schweiz weniger selektiv als in Deutschland (Abbildung 5-11). So wird die Datenmenge in der Schweiz für einen zweijährigen Erfassungszeitraum der Autoren bei der 1-stelligen Postleitzahl gegenüber der groben Konkordanz um 85% reduziert, bei der 2-stelligen um 55%. Bei den deutschen EPA-Anmeldungen liegt die Reduktion bei den 1-stelligen Postleitzahlen hier bei 65% und bei den 2-stelligen bei 40%. Dieser Unterschied dürfte auf die unterschiedliche Definition von Postleitzahlbereichen in der Schweiz gegenüber Deutschland zurückzuführen sein. Auch bei einer Überführung der Postleitzahlkodierung in Regionalkodierungen nach NUTS dürfte es ähnliche Unterschiede geben. In der Konsequenz bedeutet dies, dass das regionale Selektionskriterium in unterschiedlichen Ländern unterschiedliche Effekte hat und jeweils für das einzelne Land zu prüfen ist.

Abbildung 5-10: Ermittelte EPA-Anmeldungen aus Schweizer Universitäten nach verschiedenen Selektionskriterien (jeweils 2 Publikationsjahre pro 1 Prioritätsjahr)



Quelle: Scopus, PATSTAT, eigenen Recherchen und Berechnungen

Abbildung 5-11: Für das Jahr 2003 ermittelte EPA-Anmeldungen aus Schweizer Universitäten nach verschiedenen Selektionskriterien (mit 1 und 2 Publikationsjahren)



Quelle: Scopus, PATSTAT, eigenen Recherchen und Berechnungen

Beim Vergleich der Recall-Raten der Erfindungen, die von Universitäten beim EPA zum Patent angemeldet werden, ergibt sich für die Schweiz ein höherer Wert als für Deutschland. So errechnet sich bei einem zweijährigen Publikationszeitraum und der Anwendung der groben Konkordanz für das Jahr 2006 für die Schweiz ein Recall von 89% gegenüber Deutschland von 84%. Dieses Kriterium ist bei der Schweiz mit einer höheren Unsicherheit belastet, da nur 85 Anmeldungen im Jahr 2006 durch Universitäten getätigt wurden, gegenüber 370 aus Deutschland. Es ist aber davon auszugehen, dass diese Unterschiede im Recall maßgeblich darauf zurückgehen, dass das Schweizer Profil stärker von Anmeldungen im Bereich der Chemie dominiert wird, wo die Abdeckung der Autoren in Scopus breiter als beispielsweise im Maschinenbau ist.

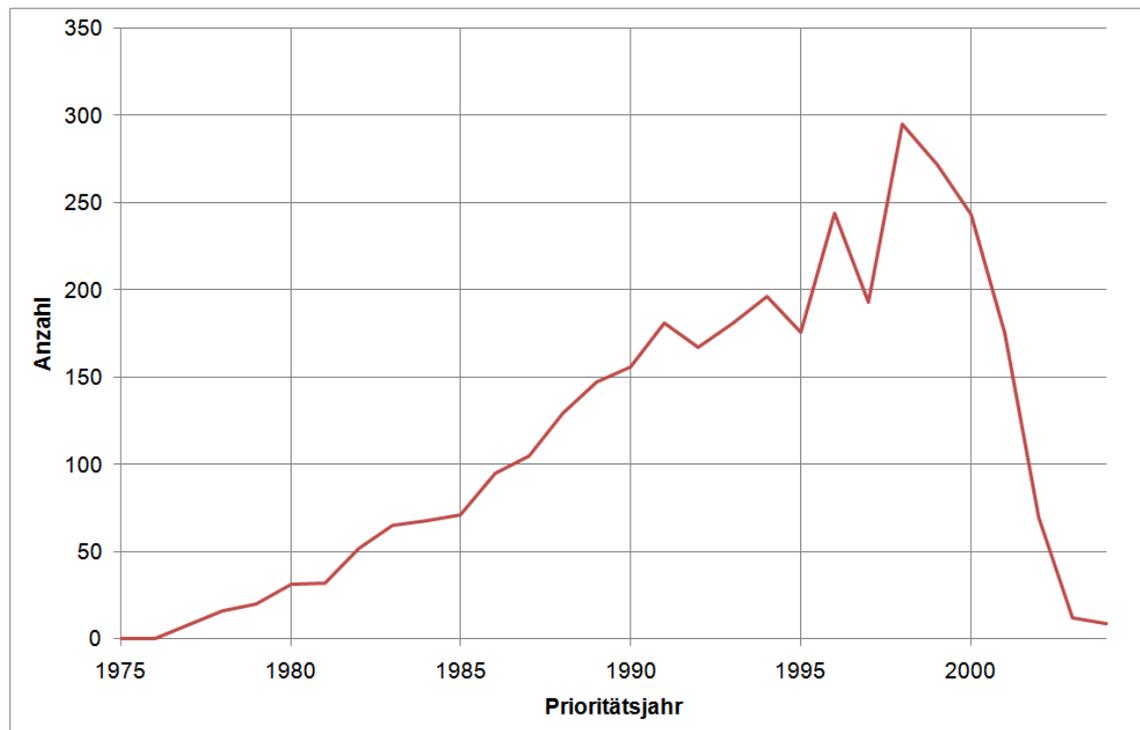
Unter der Annahme, dass die Kriterien der 2-stelligen bzw. der 1-stelligen Postleitzahl kombiniert mit der groben Konkordanz in der richtigen Größenordnung liegen, ist für Schweizer Universitäten für 2002 von einem Niveau von etwa 120 bis 180 Anmeldungen auszugehen. Dieses liegt signifikant über dem Niveau, das in dem Bericht von 2004 nach dem alten Ansatz ermittelt wurde, wo inklusiv der außeruniversitären Einrichtungen für dieses Jahr 130 Anmeldungen identifiziert wurden. Dieses Ergebnis weist darauf hin, dass einerseits der Erfassungsgrad durch den Matching-Ansatz deutlich höher ist als nach dem Verfahren mit dem Professoren-Titel, dass allerdings die Zahlen in einer sinnvollen Größenordnung liegen.

5.3 Frankreich, EPA-Anmeldungen

Die Analyse zu EPA-Anmeldungen französischer Herkunft wurde ergänzend zu der Analyse der deutschen und Schweizer Anmeldungen durchgeführt, wie sie im Antrag für das vorliegende Projekt aufgeführt waren. Der Grund für diese zusätzliche Erhebung liegt darin, dass im Rahmen des bereits in Abschnitt 2.1 diskutierten KEINS-Projektes im Jahr 2010 ein Testdatensatz zu Frankreich publiziert wurde, der dort im Rahmen des sogenannten "Name Game" zum Test von Algorithmen genutzt wurde {Lissoni, 2010 6231 /id}. Der dortige Datensatz beruht auf einer Zusammenstellung des akademischen Personals (Maîtres de Conférences oder Professeurs) aus dem Jahre 2005 {Lissoni, 2008 3374 /id}. Auf der Basis dieser Personalnamen wurden durch ein nicht näher beschriebenes Matching-Verfahren EPA-Anmeldungen aus französischen Universitäten für den Zeitraum 1978 bis 2004 ermittelt. In einem zweiten Schritt wurden die Erfinder der betreffenden Anmeldungen befragt, ob ihre Zuordnung zu einer französischen Universität zutreffend ist. Auf dieser Basis wurden fehlerhafte Zuordnungen eliminiert, sodass ein Datensatz mit EPA-Anmeldungen blieb, der zutreffend französischen Universitäten zugerechnet werden kann. Die Zuordnung der so ermittelten EPA-

Anmeldungen zu Prioritätsjahren, also Jahren der Erstanmeldung, führt zu einem Zeitverlauf entsprechend Abbildung 5-12.

Abbildung 5-12: EPA-Anmeldungen aus französischen Universitäten nach den KEINS-Analysen



Quelle: Lissoni et al. {Lissoni, 2008 3374 /id /d;Lissoni, 2010 6231 /id /d}

Bei diesem Verlauf ist nicht nachzuvollziehen, ob der stetige Anstieg seit 1978, dem Eröffnungsjahr des europäischen Amtes, als realer Anstieg zu werten ist oder aber damit zusammenhängt, dass der Personaldatensatz aus dem Jahre 2005 stammt und viele Inhaber von Universitätspatenten aus den 80er und frühen 90er Jahren im Jahre 2005 nicht mehr an Universitäten tätig waren. Auch der Zeitverzug zwischen dem Prioritätsjahr 1998, dem Maximum der ermittelten Daten, zu dem Erhebungsjahr der Personaldaten 2005 ist immer noch hoch. Eine Prüfung mit unserem Matching-Verfahren mit Autoredaten aus den Jahren 1999 im Vergleich zu 2005 hat allerdings für 1998 nicht zu höheren Ergebnissen geführt. Offensichtlich ist die Personalfluktuations an französischen Universitäten relativ gering. Im deutschen Fall würde ein entsprechend großer Abstand zwischen der Erhebung der Personaldaten und Patentanmeldungen zu einem Verlust von etwa 30% führen. Die in Abschnitt 3.2 hervorgehobene Zeitnähe zwischen Patentanmeldungen und Personaldaten ist zumindest im deutschen Fall relevant.

Das deutliche Absinken der im KEINS-Projekt erhobenen Patentanmeldungen seit 1999 ist darauf zurückzuführen, dass der dortigen Arbeitsgruppe nur eine Datenbank für EPA-Anmeldungen zur Verfügung stand. Es ist aber üblich, vor Anmeldungen im Ausland zunächst eine sogenannte PCT-Anmeldung bei der WIPO vorzunehmen (PCT= Patent Cooperation Treaty, WIPO= World Intellectual Property Organisation). Im Rahmen des PCT-Verfahrens wird eine provisorische, nicht rechtsverbindliche Prüfung der Anmeldungen durchgeführt, und der Anmelder muss sich erst zweieinhalb Jahre nach Priorität entscheiden, ob die Erfindung tatsächlich bei anderen Auslandsämtern angemeldet werden soll. Dieses hat den Vorteil, dass er durch die provisorische Prüfung einen Eindruck von den Erfolgsaussichten eines Patentverfahrens bekommt, und durch die Verzögerung von zweieinhalb Jahren ab Priorität kann er besser abschätzen, ob das mit der Erfindung verbundene Produkt im Ausland realistische Vermarktungschancen hat. Im vorliegenden Kontext bedeutet dies, dass viele EPA-Anmeldungen erst mit großer Verzögerung gegenüber dem Prioritätsjahr aus dem PCT-Verfahren an das EPA übergeleitet und dort nur mit Verspätung registriert werden. Der Rückgang der französischen EPA-Anmeldungen aus Universitäten zeigt somit lediglich, dass in den Jahren ab 2000 nur in begrenztem Maße Überleitungen aus dem PCT-Verfahren erfolgt sind. Der Datensatz ist deshalb nur bis 1999 zuverlässig.

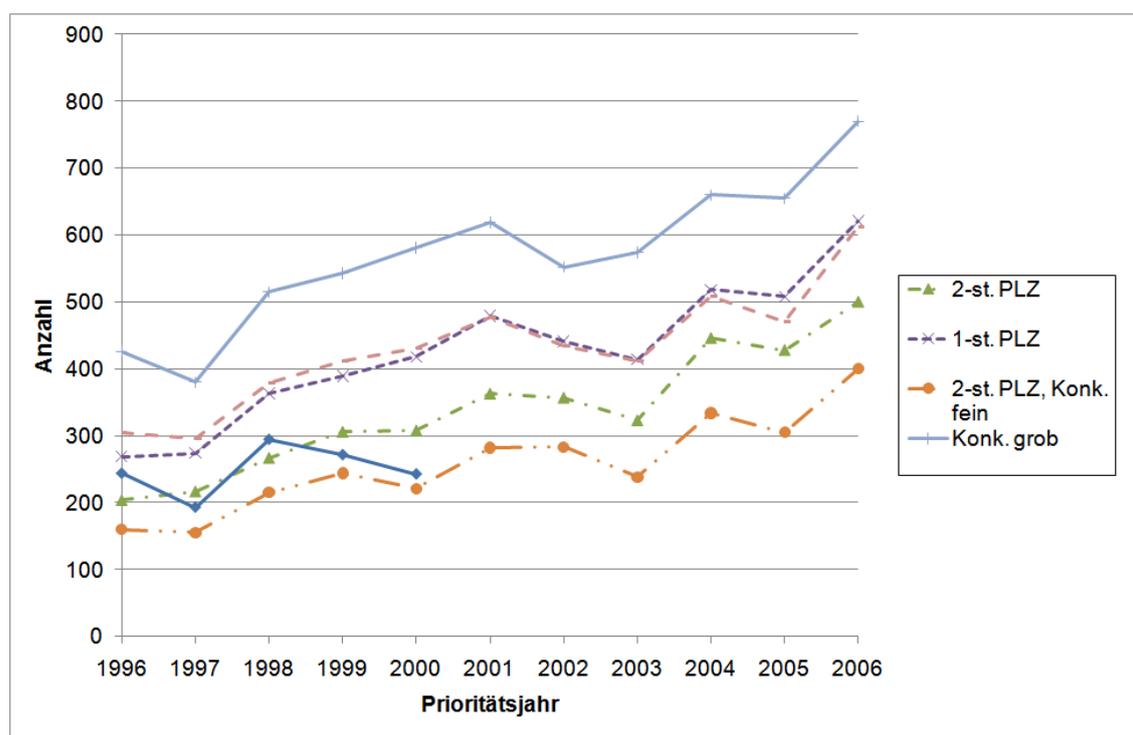
Die Analyse mit dem eigenen Matching-Verfahren, welches analog zu den Untersuchungen für Deutschland und die Schweiz durchgeführt wurde, führt zu einem Verlauf entsprechend Abbildung 5-13. Der Trend ist hier allgemein ansteigend.

Auch im Falle von Frankreich ist das schwächste Selektionskriterium die grobe Konkordanz und die Kriterien der 1-stelligen Postleitzahl und der feinen Konkordanz liegen wieder auf etwa gleichem Niveau /Abbildung 5-14). Analog zu Deutschland führen die Kriterien der 2-stelligen Postleitzahl sowie der 1-stelligen Postleitzahl in Kombination mit der groben Konkordanz zu ähnlichen Niveaus der ermittelten Patentanmeldungen. Damit wirken die 1- und 2-stelligen Postleitzahlbereiche in Frankreich in ähnlicher Weise wie in Deutschland.

Der Vergleich des Testdatensatzes mit den nach unserem Matching-Verfahren ermittelten Verläufen für die Zeit von 1996 bis 2000 ergibt sich in der Höhe ein ähnliches Niveau wie die 2-stelligen Postleitzahlen (Abbildung 5-13). Dieses kann aber nur dahingehend interpretiert werden, dass durch unser Matching-Verfahren Anmeldungen in einer sinnvollen Größenordnung bestimmt werden können. Eine Nutzung des französischen Datensatzes als "korrekte" Referenz für unseren Datensatz ist dagegen nicht möglich, da über die Autorenlisten auch solche Mitarbeiter an Universitäten mit einbezogen werden, die dort keine formale Stellung als Lehrende (Maîtres de Conférences, Professeurs) haben. Grundsätzlich werden durch unser Verfahren also Anmeldungen

aus Universitäten ermittelt, die über Personen mit formalen Positionen an Universitäten nicht ermittelbar sind. Dieser Anteil ist zwar in Frankreich aufgrund der dortigen Strukturen des akademischen Systems gering, dürfte aber eine maßgebliche Rolle spielen. Von daher wurde eine entsprechende Überprüfung nicht unternommen.

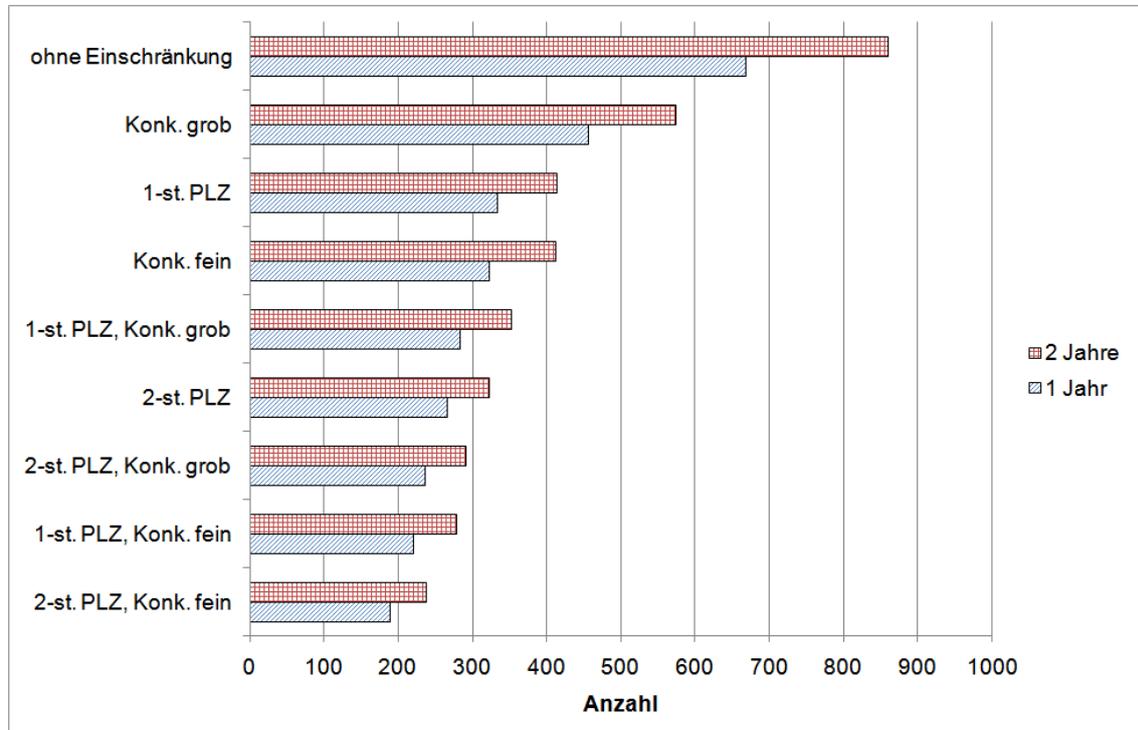
Abbildung 5-13: Ermittelte EPA-Anmeldungen aus französischen Universitäten nach verschiedenen Selektionskriterien (jeweils 2 Publikationsjahre pro 1 Prioritätsjahr)



Quelle: Scopus, PATSTAT, eigenen Recherchen und Berechnungen

Die einzige sinnvolle Prüfung war wiederum die Ermittlung des Recall für EPA-Anmeldungen für 2006 bei Autorenlisten aus zwei Jahren und der Anwendung der groben Konkordanz. Hier ergibt sich für Frankreich der sehr niedrige Wert von 73%, für Deutschland waren es 84% und für die Schweiz 89%. Dieser relativ niedrige Wert für Frankreich dürfte damit zusammenhängen, dass französische Autoren nur in begrenztem Maße in englischen Zeitschriften publizieren und damit nur begrenzt in der Datenbank Scopus erfasst sind. Bei einer Betrachtung des für Frankreich ermittelten Technikprofils zeigt sich ein deutlicher Fokus auf den Feldern Chemie und Biotechnologie, während der Maschinenbau kaum in Erscheinung tritt. Gerade im Maschinenbau dürfte jedoch der Spracheffekt eine besonders große Rolle spielen.

Abbildung 5-14: Für das Jahr 2003 ermittelte EPA-Anmeldungen aus französischen Universitäten nach verschiedenen Selektionskriterien (mit 1 und 2 Publikationsjahren)



Quelle: Scopus, PATSTAT, eigenen Recherchen und Berechnungen

6 Gesamtergebnisse

Die Analysen zu Deutschland, der Schweiz und Frankreich haben gezeigt, dass die verschiedenen Selektionskriterien es ermöglichen, Datenmengen in sinnvollen Größenordnungen für Patentanmeldungen aus Universitäten zu bestimmen. Allerdings hat jedes der analysierten Kriterien Vor- und Nachteile, sodass letztlich nur eine Kombination verschiedener Kriterien zu einem sinnvollen Ergebnis führen kann. Es hat sich darüberhinaus gezeigt, dass insbesondere die Gewichtung des Regional-Kriteriums länderspezifisch geprüft werden muss, da sich die regionalen Bedingungen nach Ländern unterscheiden, wie im vorliegenden Fall das Beispiel der Schweiz gezeigt hat.

Als ein grundsätzliches Problem hat sich erwiesen, dass für die untersuchten Fälle keine abgesicherte Referenzmenge ermittelt werden konnte, die zu einer genaueren Prüfung der Qualität der ermittelten Daten geeignet wäre. Lediglich im Falle von Patentanmeldungen durch Universitäten war es möglich, den Recall zu berechnen, nicht aber die Precision. Die Precision wird nach der folgenden Formel bestimmt:

$$\text{Precision} = t_p / (t_p + f_p)$$

Das heißt, man benötigt zur Berechnung dieser Formel eine Menge, bei der alle Elemente zutreffend entweder als positiv oder aber als negativ eingeordnet werden können (f_p = false positive, fälschlich als positiv eingestufte Elemente). Da eine solche Referenzmenge die Entscheidung über geeignete Selektionskriterien erheblich besser fundieren würde, sollen bis Mitte des Jahres zwei Datensätze generiert werden. Dieses war im Antrag für dieses Projekt nicht vorgesehen; aufgrund der bisherigen Untersuchung erweist sich ein solcher zusätzlicher Schritt jedoch als sinnvoll. Zum einen soll für den Datensatz der von Universitäten am DPMA angemeldeten Erfindungen für ein aktuelles Referenzjahr bestimmt werden, welche Anmeldungen nicht ermittelt wurden. Bei dem berechneten Recall von 84% wären dieses also 16% der durch Universitäten angemeldeten Erfindungen. Für diesen Datensatz sollen im Einzelnen die Gründe bestimmt werden, weshalb bestimmte Anmeldungen nicht ermittelt wurden. Wesentliche Gründe können dabei sein:

- der Erfinder wohnt nicht in räumlicher Nähe zur Universität an der er arbeitet,
- der Erfinder arbeitet in einem wissenschaftlichen Feld, das durch die Konkordanz nicht mit dem Technikfeld assoziiert wurde, dem die Patentanmeldung zuzuordnen ist,
- der Erfinder hat nicht in einer Zeitschrift publiziert, die in Scopus erfasst ist.

Zum zweiten soll für die gesamte Datenmenge, die durch den Matching-Ansatz für ein Jahr ermittelt wurden, die Erfinder/Autoren per E-Mail befragt werden, ob ihre Erfin-

derung zutreffend einen universitären Hintergrund hat. Der wesentliche Punkt für falsche Zuordnungen dürften hier Erfinder mit identischen Namen zu Wissenschaftlern sein, die aber selbst keine Patentanmeldung getätigt haben. In diesen fehlerhaften Fällen, muss genauer hinterfragt werden, ob durch die Anwendung geeigneter Selektionskriterien ein Ausschluss möglich gewesen wäre.

Insgesamt können die beiden so erzeugten Datensätze genutzt werden, um neben dem Recall auch die Precision zu errechnen und die Selektionskriterien zu optimieren. Dabei ist insbesondere geplant, in einem multivariaten Ansatz die verschiedenen Selektionskriterien zu gewichten und miteinander zu kombinieren. Damit ist zwar nicht vollständig sicherzustellen, dass alle relevanten Dokumente gefunden werden und dass es nicht zu einer Identifikation fehlerhafter Dokumente kommt. Es wird aber möglich sein, Precision und Recall zu optimieren, etwa über probabilistische F-Maße {Baeza-Yates 6291 /id /pt "vgl. ";Van Rijsbergen, 1979 6294 /id /pt "oder "}. In den F-Maßen werden Recall und Precision kombiniert, sodass ein Optimum erreicht werden kann.

Das Fraunhofer ISI wird diese ergänzenden Arbeiten im Rahmen der Eigenforschung realisieren.

Literatur, Referenzen

- Achleitner, A.-K./Allmendinger, J./Grupp, H./Luther, J./Soete, L./Llerena, P./Schmoch, U./Stenke, G./Meurer, P./Ulbricht, L. (2009): *Gutachten zu Forschung, Innovation und Technologischer Leistungsfähigkeit, EFI-Gutachten 2009*. Berlin: Expertenkommission Forschung und Innovation (EFI).
- Audretsch, D.B./Taylor Aldridge, T./Alexander Oettl/A. (2006): The Knowledge Filter and Economic Growth: The Role of Scientist Entrepreneurship (= Diskussionspapier Nr: 2006-11 des MPI). Jena.
- Baeza-Yates, R./Ribeiro-Neto, B.: *Modern Information Retrieval*. New York: ACM Press, Addison-Wesley.
- Crespi, G./D'Este, P./Fontana, R./Geuna, A. (2011): The impact of academic patenting on university research and its transfer, *Research Policy*, 40, 55-68.
- Egeln, J./Gehrke, B./Legler, H./Licht, G./Rammer, C./Schmoch, U. (2007): *Bericht zur technologischen Leistungsfähigkeit Deutschlands 2007*, Bundesministerium für Bildung und Forschung (ed.). Berlin: BMBF.
- Frietsch, R./Schmoch, U./Neuhäusler, P./Rothengatter, O. (2010): Patent Applications – Structures, Trends and Recent Developments (= Studien zum deutschen Innovationssystem No. 10-2010), Expertenkommission Forschung und Innovation (EFI) (ed.). Berlin.
- Hinze, S./Schmoch, U. (2004): Opening the black box: analytical approaches and their impact on the outcome of statistical patent analyses. In: Moed, H.F./Glänzel, W./Schmoch, U. (eds.): *Handbook of Quantitative Science and Technology Research, The Use of Publication and Patent Statistics in Studies on S&T Systems*. Dordrecht: Kluwer Academic Publishers, 215-235.
- Lissoni, F./Lotz, P./Schovsbo, J./Treccani, A. (2009): Academic patenting and the professor's privilege: Evidence on Denmark from the KEINS database, *Science and Public Policy*, 36, 595-607.
- Lissoni, F./Maurino, A./Pezzoni, M./Tarasconi, G. (2010): *APE INV's "NAME GAME" Algorithm Challenge: A Guideline for Benchmark Data Analysis & Reporting*. Mailand: Mailand: University Bocconi.
- Lissoni, F./Sanditov, B./Tarasconi, G. (2006): The KEINS database on academic inventors: Methodology and contents (= Working Paper No. 181). Mailand: Università Commerciale Luigi Bocconi.

- Lissoni, F./Llerena, P./McKelvey, M./Sanditov, B. (2008): Academic Patenting in Europe: New Evidence from the KEINS Database, *Research Evaluation*, 17, 87-102.
- Meyer-Krahmer, F./Schmoch, U. (1998): Science-based technologies: university-industry interactions in four fields, *Research Policy*, 27, 835-851.
- Noyons, E.C.M./Buter, R./van Raan, A.F.J./Schmoch, U./Heinze, T./Hinze, S./Rangnow, R. (2003a): *Mapping Excellence in Science and Technology across Europe. Nanoscience and Nanotechnology*, Report to the European Commission. Leiden: University of Leiden.
- Noyons, E.C.M./Buter, R.K./van Raan, A.F.J./Schmoch, U./Heinze, T./Hinze, S./Rangnow, R. (2003b): *Mapping excellence in science and technology across Europe. Life sciences*. Leiden: CWTS.
- OECD (2003): *Turning Science to Business: Patenting and Licensing at Public Research Organisations - Summary*. Paris: OECD.
- Schmoch, U. (2004): The technological output of scientific institutions. In: Glänzel, W./Moed, H./Schmoch, U. (eds.): *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, 717-731.
- Schmoch, U. (2007): Patentanmeldungen an deutschen Hochschulen (= Studien zum deutschen Innovationssystem No. 10-2007). Bonn, Berlin: BMBF.
- Schmoch, U. (2008): Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Office (WIPO), Karlsruhe: Fraunhofer ISI.
- Schmoch, U.; Gauch, S. (2004): Innovationsstandort Schweiz, Studie für das Bundesamt für Berufsbildung und Technologie, Karlsruhe: Fraunhofer-ISI.
- Van Rijsbergen, C.V. (1979): *Information Retrieval*. Boston/London: Butterworth.