

# Identifikation der Technologieprofile von FuE-betreibenden Unternehmen anhand eines Matchings von FuE- und Patentdaten

## - Endbericht -

Juni 2014

(Abschluss der Arbeiten: Oktober 2013)

### **Fraunhofer Institut für System- und Innovationsforschung**

Rainer Frietsch, Peter Neuhäusler, Carolin Michels, Nadine Bethke, Oliver  
Rothengatter

### **Wissenschaftsstatistik GmbH**

Verena Eckl, Andreas Kladroba, Gero Stenke

#### **Ansprechpartner:**

Dr. Rainer Frietsch  
Fraunhofer Institut für  
System- und Innovationsforschung ISI  
Breslauer Str. 48  
76139 Karlsruhe  
Tel.: 0721 6809-197  
E-mail: rainer.frietsch@isi.fraunhofer.de

Dr. Gero Stenke  
SV gemeinnützige Gesellschaft für  
Wissenschaftsstatistik mbH  
Barkhovenallee 1, 45239 Essen  
Tel.: 0201 8401-426  
E-Mail: gero.stenke@stifterverband.de

## Inhalt

<b>1</b>	<b>Zusammenfassung .....</b>	<b>1</b>
<b>2</b>	<b>Einleitung und Motive des Projekts .....</b>	<b>2</b>
<b>3</b>	<b>Verwendete Daten .....</b>	<b>4</b>
3.1	FuE-Daten – SV-Datenbank .....	4
3.2	Patentdaten – PATSTAT .....	5
3.2.1	Problematik der Unternehmensnamen.....	6
<b>4</b>	<b>Matching von FuE- und Patentdaten .....</b>	<b>9</b>
4.1	Das Verfahren.....	9
4.1.1	Text-Bereinigung .....	9
4.1.2	Ähnlichkeitsberechnung zwischen Texten.....	11
4.1.3	Auswahl der Matches.....	12
4.1.4	Technische Umsetzung des Matching-Algorithmus.....	13
4.2	Güte des Matchings.....	14
4.2.1	Recall und Precision .....	14
4.3	Ergebnisse des Matchings.....	17
<b>5</b>	<b>Der Mikrodatensatz.....</b>	<b>21</b>
5.1	Verwendete Patentindikatoren .....	21
5.2	Verwendete Indikatoren aus der SV-Datenbank .....	22
5.3	Analyse des Zeitverzugs zwischen FuE-Ausgaben und Patentanmeldungen.....	22
<b>6</b>	<b>Die Konkordanz zwischen Wirtschaftszweigen und Technologiefeldern.....</b>	<b>25</b>

6.1	Gewichtung der Patentdaten.....	31
6.2	Plausibilisierung der Ergebnisse .....	38
6.3	Ein Beispiel für die Möglichkeiten, die sich aus der Umrechnung ergeben .....	40
<b>7</b>	<b>Schlussbemerkungen.....</b>	<b>48</b>
<b>8</b>	<b>Literatur .....</b>	<b>50</b>

## 1 Zusammenfassung

- Ziel der Untersuchung ist die Darstellung der FuE-Aufwendungen nach Technologiefeldern, die bisher nur nach Sektoren ausgewiesen wurden.
- Es werden hierzu Daten der FuE-Erhebung und der Patentanmeldungen der FuE-treibenden Unternehmen auf der Mikroebene zusammengeführt.
- So entsteht ein integrierter Datensatz und daraus abgeleitet eine Konkordanzliste für FuE zwischen Sektoren und Technologien.
- Die Herausforderung besteht in der Validierung der berechneten Ergebnisse. Es sind nur wenige Benchmark-Daten und Untersuchungen/Studien vorhanden, weshalb die Plausibilisierung maßgeblich aufgrund der eigenen Ergebnisse vorgenommen werden musste.

Ausgewählte Ergebnisse:

- Im Technologiefeld Transport werden mit 16,6 Milliarden € im Jahr 2009 die meisten Aufwendungen für Forschung und Entwicklung verausgabt, gefolgt von der Pharmazie (4,9 Milliarden) sowie der Messtechnik, der digitalen Kommunikationstechnik und der grundlegenden Kommunikationstechnik mit jeweils knapp 3 Milliarden €.
- Die Branche 29 (Herstellung von Kraftwagen und Kraftwagenteilen) gibt etwas mehr als die Hälfte ihrer FuE im Technologiefeld Transport aus. An zweiter Stelle stehen Messtechnik, gefolgt von „Motoren, Pumpen und Turbinen“, „Mikro- und Nanostrukturen“, „Elektrische Maschinen und Anlagen“, sowie Kommunikationstechnik, Computertechnik und weiterer Elektronik.
- Umgekehrt verausgaben Unternehmen des Sektors 29 (Herstellung von Kraftwagen und Kraftwagenteilen) knapp 64 % der FuE-Aufwendungen im Technologiefeld Transport. Der sonstige Fahrzeugbau ist für knapp 12 % der Aufwendungen in diesem Technologiefeld verantwortlich, gefolgt vom Maschinenbau (5,5 %), der Herstellung von Datenverarbeitungsgeräten (4,3 %) und der Herstellung von Gummi- und Kunststoffwaren (3,3 %).
- Es zeigt sich eine Spreizung der Technologien über die Sektoren und somit gleichermaßen eine Spreizung der Sektoren über die Technologien. Je nach Fragestellung können in der einen oder der anderen Perspektive zielgerichtete Analysen durchgeführt und damit neue Erkenntnisse gewonnen werden.

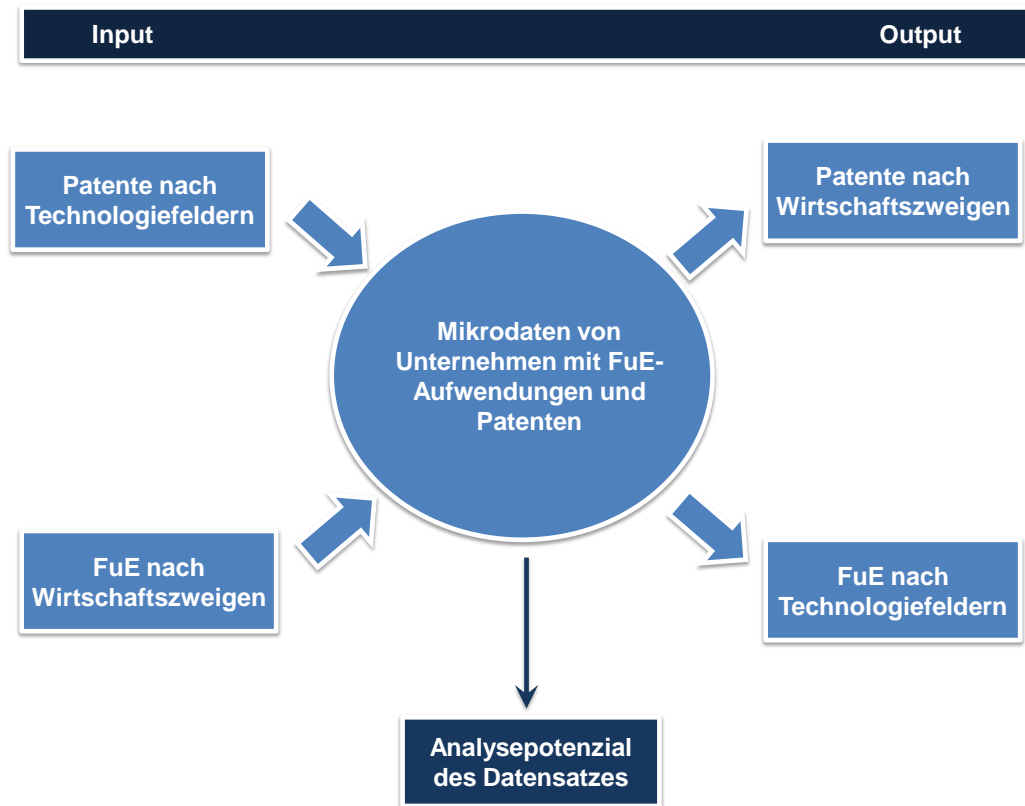
## 2 Einleitung und Motive des Projekts

Das Ziel dieses Projektes ist es, eine Konkordanzliste zwischen Wirtschaftszweigen und Technologien zu erstellen, um die Aufwendungen der Unternehmen für Forschung und Entwicklung nicht ausschließlich nach Wirtschaftszweigen, sondern auch nach Technologien ausweisen zu können. Da insbesondere große, multinationale Unternehmen häufig technologisch heterogen sind, ist die Wirtschaftszweigsystematik nicht in der Lage, die FuE-Aufwendungen der Unternehmen adäquat abzubilden. Zusätzlich liegen zahlreiche andere Indikatoren und Informationen nur auf der Ebene von Technologien vor. Dazu zählen beispielsweise Förderprogramme, wissenschaftsnahe Indikatoren wie Beschäftigte in öffentlichen Einrichtungen, sowie auch Patente. Hinzu kommt, dass Unternehmen selbst selten Angaben über ihre FuE-Aufwendungen nach Technologiefeldern machen können, sondern zumeist nur einzelne Projektaufwendungen oder ihre gesamten FuE-Aufwendungen beziffern können.

Zur Lösung dieses Problems wurden innerhalb dieses Projekts Daten zu den Forschungs- und Entwicklungsausgaben der Unternehmen mit den jeweiligen Daten zu Patentanmeldungen auf der Mikroebene, das heißt auf der Ebene von Unternehmen bzw. Patentanmeldern, verknüpft. Patente werden nach Technologien innerhalb der Internationalen Patentklassifikation (IPC) klassifiziert und sind einer der wichtigsten Indikatoren für den Output von FuE-Prozessen. Sie werden häufig dazu verwendet, die technologische Leistungsfähigkeit von Unternehmen, Technologiefeldern oder ganzen Volkswirtschaften zu beurteilen (Freeman 1982; Grupp 1998). Eine große Anzahl von Patenten deutet auf höhere FuE-Aktivitäten und somit auf einen höheren innovativen Output hin.

Für die Erstellung der Konkordanzliste wurden Patent- und FuE-Daten in einem ersten Schritt aufbereitet, um ein Matching der beiden Datenquellen zu ermöglichen. In einem zweiten Schritt wurden die FuE- und Patentdaten auf Unternehmensebene mit Hilfe eines "String-Matching Algorithmus" zusammen geführt. Auf Basis dieser Zusammenführung (Matching) konnte ein Datensatz erstellt werden, der eine Verknüpfung von FuE- und Patentdaten auf der Unternehmensebene enthält. Durch Aggregation dieses Datensatzes auf der Ebene von Sektoren bzw. Wirtschaftszweigen können die FuE-Aufwendungen somit auf Ebene von Technologiefeldern, bzw. umgekehrt auch Patente auf der Ebene von Wirtschaftszweigen ausgewiesen werden. Jedoch bietet der entstandene Mikrodatensatz weiteres Analysepotenzial, beispielsweise zur Untersuchung des Zeitverzugs zwischen Forschungs- und Entwicklung-Aufwendungen und der Anmeldung eines Patents.

Abbildung 1 Schematische Darstellung des Projekts



Quelle: Fraunhofer ISI – eigene Darstellung

Ein schematischer Überblick des Projekts ist in Abbildung 1 dargestellt. Im nächsten Kapitel werden zunächst die verwendeten Daten und deren Aufbereitung näher erläutert. Kapitel 3 widmet sich dem Matching-Algorithmus und dessen technischer Umsetzung sowie der Analyse der Güte des Matchings. In Kapitel 4 wird der durch die Zusammenführung entstandene Mikrodatsatz näher erläutert, während Kapitel 5 auf die sektorenspezifische Gewichtung der Patentdaten eingeht und die Ergebnisse der Umrechnung von FuE-Daten nach Technologien präsentiert.

### 3 Verwendete Daten

In diesem Kapitel werden zunächst die FuE- und Patentdatenbanken vorgestellt, die für die Erstellung der jeweiligen Datensätze verwendet wurden. Im Anschluss daran wird im Detail das Verfahren dargestellt, das dazu benutzt wird, beide Datenquellen miteinander zu verknüpfen. Aus dieser Verknüpfung resultiert ein neuer Datensatz, mit dessen Hilfe FuE-Ausgaben zu Patenten zugeordnet werden konnten.

#### 3.1 FuE-Daten – SV-Datenbank

Daten über Forschungs- und Entwicklungsaktivitäten der Unternehmen in Deutschland werden systematisch bereits seit den 1970er Jahren vom Stifterverband für die Deutsche Wissenschaft erhoben. Auftraggeber ist das Bundesministerium für Bildung und Forschung (BMBF). Die Befragung basiert auf einem international abgestimmten Regelwerk der OECD, dem Frascati Handbuch. Demnach fallen Tätigkeiten unter die Kategorie „FuE“ wenn sie das Ziel verfolgen

- vorhandenes Wissen zu erweitern und neue allgemeingültige Kenntnisse zu gewinnen,
- neue Anwendungsmöglichkeiten zu finden oder
- neue Produkte, Dienstleistungen oder Verfahren auf experimentelle Weise zu entwickeln oder durch wesentliche Änderungen weiterzuentwickeln.

Forschung und Entwicklung ist nicht an bestimmte Fachdisziplinen gebunden. Wichtig ist lediglich, dass im Ergebnis etwas Neues entsteht und der FuE-Prozess somit zur Lösung von Problemen und Fragestellungen beiträgt.

Die FuE-Erhebung des Stifterverbandes stellt mit Abstand die umfangreichste Datenbasis für wirtschaftsbasierte FuE-Aktivitäten innerhalb Deutschlands dar. Mit Hilfe einer standardisierten schriftlichen Befragung werden im zweijährigen Rhythmus Daten über interne und externe FuE-Aufwendungen, FuE-Personal, verschiedene Formen und Zielsetzungen von FuE, Innovationen als Outputgröße sowie die regionale und sektorale Verteilung der FuE-Tätigkeit ermittelt. Die Erhebung ist als Vollerhebung konzipiert. Die Grundgesamtheit umfasst damit alle FuE-treibenden Unternehmen in Deutschland.

Angesprochen werden Unternehmen, bei denen angenommen wird, dass sie FuE betreiben. Diese Annahme ergibt sich durch Hilfsvariablen wie Branche, Unternehmensgröße, Förderinformationen oder auch aufgrund von Informationen aus früheren FuE-Erhebungen. Berichtseinheit bei den Unternehmen ist in der Regel die kleinste bilanzierende Einheit. Bei Konzernen wird eine Aufgliederung in unterschiedliche Berichtseinheiten oder Geschäftsbereiche angestrebt, um der unterschiedlichen Branchendiffe-

renzung und der Produktvielfalt der Konzerne gerecht zu werden. Die letztendliche Zuordnung eines Unternehmens bzw. der kleinsten bilanzierenden Einheit zu einem Wirtschaftszweig, geschieht nach dem wirtschaftlichen Schwerpunkt. Normative Basis ist die Klassifikation der Wirtschaftszweige des Statistischen Bundesamtes, aktuell die WZ 2008. Eine Abfrage relevanter Technologiebereiche wird im Rahmen der FuE-Erhebung nicht vorgenommen.

Die Variablen FuE-Erhebung sind quantitativ geprägt. Die Abfrage metrischer Werte steht im Vordergrund. Kernindikatoren sind die Aufwendungen für Forschung und Entwicklung sowie das in diesem Bereich tätige Personal. Für die Zusammenführung von FuE- und Patentdaten wurden alle Unternehmen ausgewählt, die FuE-Aufwendungen in den Jahren 2007 und 2009 gemeldet hatten.

Damit eine Verknüpfung von FuE- und Patentinformationen auf Ebene einzelner Unternehmen gelingt ist es erforderlich, dass diese Unternehmen in beiden Datenbanken eindeutig identifiziert werden können. Ein Unternehmen etwa, das sowohl forscht als auch seine Forschungsergebnisse zum Patent anmeldet, wird in beiden Datenbanken registriert sein. Für die einwandfreie Identifizierung dieses Unternehmens existiert jedoch kein einheitliches Identifizierungsmerkmal, etwa eine Registrierungsnummer. Daher müssen andere Daten für eine Zusammenführung der Informationen herangezogen werden. Die größte Wahrscheinlichkeit auf Erfolg bietet ein Verwenden des Unternehmensnamens und der Adresse. Diese Vorgehensweise wird in Kap. 2.2.1 ff näher erläutert.

### **3.2 Patentdaten – PATSTAT**

Die Patentdaten für die vorliegende Untersuchung wurden aus der "EPO Worldwide Patent Statistical Database" (PATSTAT) extrahiert, die Informationen zu veröffentlichten Patenten aus 81 Patentbehörden weltweit zur Verfügung stellt.

Alle Patentanmeldungen werden nach dem Jahr der weltweit ersten Anmeldung, dem sogenannten Prioritätsjahr gezählt. Dies ist der früheste registrierte Zeitpunkt im Patentverfahren und kommt daher dem Zeitpunkt der Erfindung am nächsten. Für das Matching mit der Datenbank des Stifterverbandes aus der FuE-Erhebung wurden zunächst alle deutschen Patentanmelder mit mindestens einer Patentanmeldung am europäischen Patentamt oder dem deutschen Patent- und Markenamt (DPMA) zwischen 2005 und 2009 ausgewählt. Mit Hilfe der Anmelderinformatio n ist es möglich, die Informationen der Patentanmelder mit den FuE treibenden Unternehmen der SV-Datenbank zu verknüpfen.

Nach der Verknüpfung der SV-Datenbank mit PATSTAT wurden zusätzliche Patentindikatoren auf Ebene der Patentanmelder bzw. Unternehmen aus der Datenbank extrahiert und zu einem einheitlichen Datensatz zusammengefasst. Für den erstellten Datensatz wurden alle Patente ausgewählt, die sich auf den deutschen Markt beziehen. Um dies zu erreichen, wurden alle beim Deutschen Patent- und Markenamt (DPMA) direkt angemeldeten Patente, sowie alle Anmeldungen beim EPA gezählt (einschließlich der Patentanmeldungen, die von der World Intellectual Property Organization (WIPO) über das Patent Cooperation Treaty (PCT) Verfahren am EPA angemeldet wurden). Um Doppelzählungen zu vermeiden, wurden alle Patente, die vom EPA an das DPMA weitergeleitet wurden, von der Analyse ausgeschlossen. Die zugrunde liegende Annahme dieses Verfahrens ist, dass alle Patentanmeldungen deutscher Anmelder früher oder später den deutschen Markt erreichen, egal ob diese direkt beim DPMA oder am EPA angemeldet wurden. Die Zählung aller EPA und DPMA Patente unter Ausschluss von Doppelzählungen ermöglicht somit die Auswertung aller Patente, die eine Erfindung auf dem deutschen Markt schützen sollen.

Neben den gesamten Patentanmeldungen pro Unternehmen werden die Patente nach Technologiefeldern differenziert. Die Technologiefelder im finalen Datensatz werden zunächst nach 35 WIPO-Klassen (Schmoch 2008) unterschieden. Dies ist die Basis für die Konkordanzliste. Neben den feldspezifischen Patentanmeldungen wurden zusätzlich weitere, oben erwähnte Patentindikatoren (beispielsweise Vorwärtszitationen) in den Datensatz aufgenommen werden.

### **3.2.1 Problematik der Unternehmensnamen**

Wie im vorigen Abschnitt bereits erwähnt, wird das Matching der Datensätze mit Hilfe der Unternehmensnamen (SV-Datenbank) und der Namen der Patentanmelder (PATSTAT) durchgeführt. Dies bringt einige Herausforderungen mit sich, auf die an dieser Stelle kurz eingegangen werden soll. Das im nachfolgenden Kapitel beschriebene Matchingverfahren muss genau diese Probleme adressieren, um die Namen von Patentanmeldern den Unternehmensnamen zuordnen zu können.

Innerhalb von PATSTAT liegen die Namen der Anmelder in einer Rohdatenversion vor, das heißt, dass unterschiedliche Namensvarianten des gleichen Patentanmelders existieren können, da dieser direkt von der Patentanmeldung übernommen wird. Zudem enthält der Anmeldername Sonderzeichen, Abkürzungen, die Rechtsform, sowie etwaige orthographische Fehler.

Tabelle 1 Namensvarianten – Eine beispielhafte Übersicht anhand der Bayer AG

<b>Originalname</b>	<b>Harmonisierter Name (EEE-PPAT)</b>
<b>Bayer A.G.</b>	BAYER
<b>Bayer AC</b>	BAYER AC
<b>Bayer Adtiengesellschaft</b>	BAYER
<b>Bayer AG</b>	BAYER
<b>Bayer Aktiengesellschaft</b>	BAYER
<b>Bayer Akiengesellschaft</b>	BAYER
<b>Bayer Aktlengesellschaft</b>	BAYER
<b>Bayer Animal Health GmbH</b>	BAYER ANIMAL HEALTH
<b>Bayer BioScience GmbH</b>	BAYER BIOSCIENCE
<b>Bayer Business Services GMBH</b>	BAYER BUSINESS SERVICES
<b>Bayer Chemical Aktiengesellschaft</b>	BAYER CHEMICALS
<b>Bayer Chemicals AG</b>	BAYER CHEMICALS
<b>Bayer Chemicals Aktiengesellschaft</b>	BAYER CHEMICALS
<b>Bayer CropScience AG</b>	BAYER CROPSCIENCE
<b>Bayer CropScience</b>	BAYER CROPSCIENCE
<b>Bayer CropScience GmbH</b>	BAYER CROPSCIENCE
<b>Bayer HealthCare AG</b>	BAYER HEALTHCARE
<b>Bayer Schering Pharma AG</b>	BAYER SCHERING PHARMA
<b>Bayer Schering Pharma Aktien</b>	BAYER SCHERING PHARMA
<b>Bayer Technology Services GmbH</b>	BAYER TECHNOLOGY SERVICES

Um dieser Herausforderung schon vor dem eigentlichen Matching zu begegnen bzw. potentielle Fehlerquellen zu minimieren, wurde statt des Original-Anmeldernamens auf den harmonisierten Anmeldernamen aus der EEE-PPAT Tabelle der KU Leuven zurückgegriffen (Du Plessis et al. 2009; Magerman T. et al. 2009; Peeters B. et al. 2009). Hierbei handelt es sich um eine automatisierte Harmonisierung aller in PATSTAT vorkommenden Anmeldernamen. Bei der Harmonisierung des Anmeldernamens wurde eine schrittweise Validierung vorgenommen, was ein Zeichencleaning (Formatcodes aus HTML, Akzente etc.), ein Interpunktionscleaning, die Überführung der Rechtsform in Sektorinformation (Inc., LTD, GmbH etc. = Company), die Harmonisierung zusätzlicher Unternehmensanzeiger („COMPANY“, „CORP“, „CORPORATION“), eine Harmonisierung der Schreibweisenvarianten („SYSTEM“, „SYSTEMS“, „SYSTEMES“), die Kondensierung irrelevanter Zeichen („3 COM“, „3COM“) sowie eine Umlaut-

Harmonisierung beinhaltet. Ein Vergleich des Anmeldernamens und des harmonisierten Namens, der für das Matching verwendet wurde, ist in

Tabelle 1 dargestellt.

Ein weiteres Problem entsteht dadurch, dass Unternehmen nicht zwingend die anmeldende und damit die im Patent genannte Einheit darstellen, sondern teilweise nur die Konzernmutter oder die Unternehmenszentrale als Patentanmelder fungieren. Zusätzlich verändern sich Unternehmen im Zeitverlauf und kaufen Tochterunternehmen hinzu oder stoßen sie ab. Dieses Problem kann erst nach dem Matching mit Hilfe einer Zuordnung der Technologien zu den Betrieben bzw. Betriebsteilen gelöst werden.

## 4 Matching von FuE- und Patentdaten

### 4.1 Das Verfahren

Das Ziel des verwendeten Verfahrens ist das Auffinden von Anmelderinformationen in PATSTAT, die einem Eintrag in der SV-Datenbank entsprechen bzw. diesem ähneln. Hierfür wird für jeden Eintrag aus der SV-Datenbank die Ähnlichkeit zu allen Anmeldernamen in PATSTAT im Zeitraum 2005 bis 2009 berechnet. Falls ein bestimmter Grad an Ähnlichkeit überschritten wird, wird das jeweilige Paar aus SV- und PATSTAT-Eintrag als sogenannter „Match“ gespeichert. Alle Daten, die dem PATSTAT-Eintrag zugeordnet sind, können nun für den SV-Eintrag übernommen werden. Durch Namensvarianten, Namensänderungen und ähnliches kann es vorkommen, dass mehrere PATSTAT-Einträge einem SV-Eintrag zugewiesen werden. Die damit verbundenen Informationen werden dann aggregiert.

Im Folgenden werden die drei Schritte des Verfahrens im Detail beschrieben. Im ersten Schritt werden sowohl die Anmeldernamen aus PATSTAT als auch aus dem SV-Datensatz auf gleiche Weise bereinigt. Hierbei werden z.B. alle Umlaute in die entsprechenden Vokale umgewandelt. Dies soll soweit wie möglich verhindern, dass eine unterschiedliche Schreibweise zu einer geringeren Ähnlichkeit führt.

Im zweiten Teil des Verfahrens wird dann die Ähnlichkeit zwischen den bereinigten Namenstexten berechnet. Hierfür wurde eine Variante der Levenshtein-Distanz implementiert. Die Levenshtein-Distanz ist ursprünglich ein Maß zur Bestimmung der Unähnlichkeit zweier Zeichenfolgen. Für das hier präsentierte Verfahren wurde sie so weiterentwickelt, dass auf Basis der Distanz die normierte Ähnlichkeit zwischen zwei Texten berechnet werden kann.

Im letzten Schritt werden dann für jeden Eintrag aus dem SV-Datensatz diejenigen Anmelder aus PATSTAT ausgewählt, die die größte Ähnlichkeit aufweisen bzw. einen vorher festgelegten Schwellwert überschreiten. Diese Matches werden in einer extra Tabelle in der Datenbank zur weiteren Verwendung abgespeichert.

#### 4.1.1 Text-Bereinigung

Die Bereinigung der Text-Strings erfolgt, damit die darauf berechnete Ähnlichkeit nicht durch unterschiedliche Schreibweisen unnötig verringert wird. Deshalb wird das gleiche Verfahren auch auf Einträge in beiden Datensätzen gleichermaßen angewendet.

In einem ersten Schritt wird der gesamte Text in Kleinschreibung umgewandelt. Dann werden alle Umlaute und Sonderzeichen entfernt bzw. ersetzt; Umlaute werden durch

den entsprechenden Vokal substituiert, d.h. Ü,Ä und Ö werden zu u, a und o. Sonderzeichen hingegen werden durch ein Leerzeichen ersetzt. Im darauf folgenden Schritt werden alle mehrfachen Vorkommen von Leerzeichen durch ein einzelnes Leerzeichen ersetzt. Sonderzeichen in diesem Verfahren umfasst alle Zeichen, die kein Buchstabe oder keine Zahl sind, also auch Satzzeichen.

Zur Veranschaulichung der Notwendigkeit dieses Vorgehens folgt die beispielhafte Verarbeitung eines genuinen Eintrags aus den SV-Daten - „Werkzeug, Formen- & Systemtechnik Sontra GmbH“. Der Teil „formen- & systemtechnik“ wird durch den vorliegenden Algorithmus zu „formen systemtechnik“ reduziert. Es könnte nun sein, dass die Leerzeichen zwischen den Sonderzeichen und den Textteilen im entsprechenden PATSTAT-Anmeldernamen fehlen. Würde das Verfahren die Sonderzeichen einfach nur entfernen und nicht ersetzen, würde das zu dem Text „werkzeugformensystemtechnik sontra gmbh“ führen. Wie im nächsten Abschnitt deutlicher wird, wären damit Abzüge in der Ähnlichkeit zwischen den beiden Texten („werkzeugformensystemtechnik sontra gmbh“ und „werkzeug formen systemtechnik sontra gmbh“) verbunden. Ebenso würde das mehrfache Vorkommen von Leerzeichen zu einem geringeren Ähnlichkeitswert führen. Die Ersetzung durch Leerzeichen mit anschließender Reduktion auf maximal ein Leerzeichen in Folge umgeht dieses Problem und führt zu einer einheitlichen Schreibweise beider Texte.

Anschließend werden alle Rechts- oder Gesellschaftsformbezeichnungen entfernt. Dies betrifft sowohl abgekürzte wie auch ausformulierte Bezeichnungen. Im oberen Beispiel würde die Bezeichnung „gmbh“ entfernt. In einem vergleichbaren Text würden ebenso alle einzelnen Begriffe aus „Gesellschaft mit beschränkter Haftung“ beseitigt. Analog zu dem Begriff „mit“ wird auch der Begriff „und“ entfernt, um eine größere Ähnlichkeit zu Texten zu sichern, bei denen das „&“-Zeichen im vorangegangenen Schritt entfernt wurde. Tabelle 2 zeigt noch einmal explizit für das oben genannte Beispiel die einzelnen Schritte der Text-Bereinigung.

Tabelle 2 Die verschiedenen Bereinigungsbeispiele am Beispiel zweier Textstrings

	<b>Text in SV-Datensatz (genuin)</b>	<b>Text in PATSTAT (fiktiv)</b>
<b>Initial</b>	„Werkzeug, Formen- & Systemtechnik Sontra GmbH“	„Werkzeug, Formen- und Systemtechnik Sontra Gesellschaft mbH“
<b>Umwandlung in Kleinbuchstaben</b>	„werkzeug, formen- & systemtechnik sontra gmbh“	„werkzeug, formen- und systemtechnik sontra gesellschaft mbh“
<b>Entfernung aller Umlaute und Sonderzeichen</b>	„werkzeug formen systemtechnik sontra gmbh“	„werkzeug formen und systemtechnik sontra gesellschaft mbh“
<b>Entfernung von Gesellschaftsbezeichnungen und sonstiger Begriffe</b>	„werkzeug formen systemtechnik sontra“	„werkzeug formen systemtechnik sontra“

#### 4.1.2 Ähnlichkeitsberechnung zwischen Texten

Für die Ähnlichkeitsberechnung wurde eine Variante der Levenshtein-Distanz in den Algorithmus integriert. Die Levenshtein-Distanz gibt für zwei Zeichenfolgen an, wie viel Editierungen der Texte nötig wären, um beide anzugleichen. Editierungen können hierbei sein: Einfügen eines Zeichens bzw. Löschen eines Zeichens und Ersetzen eines Zeichens durch ein anderes.

Zum Beispiel beträgt die Levenshtein-Distanz zwischen den Texten „siemens“ und „seimens“ 2, da eine beispielhafte Transformation des 1. Strings zum 2. die Löschung und anschließende Einfügung des „i“ wäre.

Wie an diesem Beispiel deutlich wird, gibt es meist mehrere Vorgehen, d.h. Folgen von Editierungen, die zum gleichen Ergebnis führen. Die Levenshtein-Distanz gibt aber immer die minimale Anzahl an Editierungen an, die nötig wären, um die Text-Strings anzugleichen. Verschiedene Varianten, die in der gleichen Distanz resultieren, sind somit nicht von Bedeutung.

Im hier vorgestellten Verfahren wird die Levenshtein-Distanz mit der Anzahl der Zeichen im längeren String normiert. Der so erhaltene Wert zwischen 0 und 1 gibt weiterhin die Distanz an. Um ein Ähnlichkeitsmaß im Bereich 0 bis 1 zu erhalten, wird der vorhergehende Wert deshalb von 1 subtrahiert.

Um die Bedeutung der Normierung und Umwandlung in eine Ähnlichkeit anstatt einer Distanz zu verdeutlichen, vergleichen wir die folgenden (bereits bereinigten) Texte aus dem SV- und PATSTAT-Datensatz: „siemens medical solutionsgsd“ (Länge 28 Zeichen) und „siemens medical solutions diagnostics“ (Länge 37 Zeichen). Die Levenshtein-Distanz beträgt 11, da ein Leerzeichen und 10 Buchstaben („iagnostics“) entfernt bzw. hinzugefügt werden müssen, um beide Strings aneinander anzupassen. Dieser Wert wird nun zum normierten Ähnlichkeitswert von  $1-(11/37)=(26/37)$ .

Die Normierung mit der Anzahl der Zeichen des längeren Strings erscheint sinnvoll, da dies der maximalen Levenshtein-Distanz für die beiden Strings entspricht; bei zwei Strings  $s_1$  und  $s_2$  der Länge  $n_1$  und  $n_2$  mit  $n_1 > n_2$  sind maximal  $n_1$  Editierungen nötig, um beide Strings aneinander anzugleichen. Im schlimmsten Fall, das heißt, wenn alle Zeichen der beiden Texte unterschiedlich sind, müssten  $n_1 - n_2$  Zeichen gelöscht und die verbleibenden  $n_2$  Zeichen durch die Zeichen aus  $s_2$  substituiert werden. Also wären  $n_1 - n_2$  Löschungen und  $n_2$  Substitutionen nötig, was  $n_1$  Editierungen entspricht.

Bei  $n_1$  Editierungen würden wir die normierte Distanz 1 ( $=n_1/n_1$ ) bzw. eine Ähnlichkeit von 0 erhalten. Dies entspricht der Überlegung, dass der gesamte String geändert werden musste.

Sind sowohl für den jeweiligen SV- als auch den PATSTAT-Eintrag PLZ-Informationen enthalten, sieht unser Ansatz vor, dass die ersten 3 Stellen dieser Einträge übereinstimmen müssen. Ansonsten wird die Ähnlichkeit zwischen den beiden Einträgen auf 0 gesetzt. Liegt für mindestens einen der beiden Einträge keine Information zur PLZ vor oder ist diese für beide identisch, entspricht die Ähnlichkeit der oben beschriebenen Berechnung, d.h. die Postleitzahl wird nicht für eine genauere Eingrenzung hinzugezogen.

### 4.1.3 Auswahl der Matches

Nachdem mit Hilfe der Levenshtein-Distanz die Ähnlichkeit eines Eintrags aus dem SV-Datensatz mit allen verfügbaren PATSTAT-Einträgen berechnet wurde, müssen diejenigen Einträge bestimmt werden, die als echte „Treffer“ bzw. Matches gewertet werden sollen. Hierzu wird auf einen vorher festgelegten Schwellwert  $t$  zurückgegriffen. Dieser Schwellwert muss (wie die möglichen Ähnlichkeitswerte) im Bereich  $[0,1]$  liegen. Er bestimmt, wie hoch die Ähnlichkeit zwischen zwei Einträgen mindestens sein muss, damit diese als Match gewertet werden. Ein Schwellwert von  $t=0$  würde dazu führen, dass alle PATSTAT-Einträge, deren Ähnlichkeit größer als der Minimalwert 0 ist, als Treffer zurückgeliefert werden.

Der Algorithmus vermerkt alle PATSTAT-Einträge, die eine Ähnlichkeit größer als  $t$  aufweisen, als Matches für einen SV-Eintrag. Zusätzlich kann noch die maximale Anzahl an Treffern pro SV-Eintrag,  $k$ , angegeben werden.  $k$  ist eine natürliche Zahl größer 0 und soll die Trefferliste beschränken, sofern zu viele Treffer in PATSTAT gefunden werden. Dies kann beispielsweise nötig sein, falls ein Eintrag ambig ist und der allgemein gewählte Parameter  $t$  deshalb die große Trefferliste nicht ausreichend einschränkt. Durch die Beschränkung auf  $k$  Einträge werden die besten Matches ausgewählt, d.h. die mit der höchsten Ähnlichkeit zum SV-Eintrag.  $k$  ist als optionaler Parameter implementiert, er muss also nicht verwendet werden, um die Treffermenge weiter einzugrenzen.

Allgemein kann man sagen: Werden  $n$  Einträge in PATSTAT gefunden, die eine Ähnlichkeit größer als  $t$  aufweisen, wird der Parameter  $k$  überprüft. Gilt  $k < n$ , werden nur die  $k$  besten Treffer aus der Trefferliste als Matches gespeichert. Ist  $k$  nicht definiert oder größer als  $n$ , werden alle  $n$  Einträge übernommen.

#### **4.1.4 Technische Umsetzung des Matching-Algorithmus**

Die Datenbereinigung und das Matching wurden mit einer eigenen Java-Implementierung umgesetzt. Diese objektorientierte Programmiersprache erlaubte eine von einander weitestgehend unabhängige Entwicklung der einzelnen Komponenten (Ähnlichkeitsberechnung, Datenbereinigung etc.). Das Java-Programm wird im Folgenden näher erläutert.

Die Hauptklasse des Programms liest die Daten ein. Dafür müssen zwei getrennte Textdateien mit den zu matchenden Einträgen vorliegen. Das bedeutet, dass sich in dem vorliegenden Fall in einer Textdatei nur die Informationen aus PATSTAT, in der anderen nur die SV-Daten befinden. Außerdem werden Trennzeichen benutzt, um die einzelnen Felder, d.h. ID, Postleitzahl, etc., von einander separieren zu können. Jede Zeile darf nur die Information zu einer Instanz beinhalten. In diesem Fall umfasst dies nur die ID, eventuell die Postleitzahl und den Namen der einzelnen Firma.

In der Bereinigung der Daten werden diese Zeilen zuerst getrennt betrachtet. Zunächst wird der Text in Kleinbuchstaben umgewandelt und alle Zeichen, die keinem Buchstaben und keiner Zahl entsprechen, werden entfernt. Dann werden die Wörter in der Zeile separiert und es wird geprüft, ob sich einzelne Wörter in der Liste mit Rechtsformen und deren Abkürzungen befindet. Ist dies der Fall, wird das Wort aus den Zeileneinträgen entfernt. Ansonsten wird das Wort für die bereinigte Textform verwendet. Die Liste zum Abgleich der Rechtsformen ist hart codiert, das heißt statisch implementiert. Dies hat den Vorteil, dass die Liste nicht bei jedem Aufruf des Programms übergeben und

eingelezen werden muss. Die Inhouse-Entwicklung erlaubt aber eine jederzeitige Anpassung und Erweiterung dieser Liste.

Eine zweite Klasse beschäftigt sich mit der Berechnung der Ähnlichkeiten zwischen den Texteinträgen aus den zwei verschiedenen Textdateien. Hierzu wurden wieder zeilenweise die Text-Strings miteinander verglichen. Die Levenshtein-Distanz wurde wie oben erläutert benutzt, um die Ähnlichkeit der beiden Textzeilen zu berechnen. Bei Bedarf fließt, wie oben beschrieben, die Postleitzahl ebenfalls in die Berechnung mit ein.

Es werden die Ähnlichkeiten zwischen allen Zeileneinträgen (repräsentiert durch deren IDs) gespeichert. Für jeden Eintrag aus der ersten an das Programm übergebenen Textdatei wurden die IDs aus der zweiten Textdatei nach Ähnlichkeit sortiert. Die oben genannten Schwellwert-Parameter wurden benutzt, um eventuell Matches mit zu geringer Ähnlichkeit auszusortieren.

Also Output wurden sowohl eine textuelle Repräsentation der Ergebnisse sowie eine Übertragung in die Inhouse-implementierte PATSTAT-SQL-Datenbank geliefert.

## **4.2 Güte des Matchings**

Je nachdem bei welcher Höhe der Schwellwert  $t$  festgelegt wird, werden beim Matching unterschiedliche Ergebnisse der Zuordnung erzielt. Um die Zuordnung des Matchings zu optimieren, das heißt den "optimalen" Schwellwert  $t$  für das Matching festzulegen, und die Ergebnisse der Datenbankrecherchen validieren zu können, werden zwei Konzepte aus der Informatik verwendet, die bei derartigen Problemen üblicherweise angewendet werden. Dabei handelt es sich um den so genannten Ertrag (Recall) des Matchings und dessen Genauigkeit (Precision). Die Analyse dieser beiden Kennzahlen setzt allerdings voraus, dass eine exakte Referenzmenge bekannt ist, die als Maßstab herangezogen werden kann, gegen den das Ergebnis des Matchings validiert wird. Um diesen Maßstab zu generieren, wurden aus der SV-Datenbank 1.000 Unternehmen zufällig ausgewählt und per Hand zu Patentanmeldern aus PATSTAT zugeordnet, was einem exakten Matching entspricht.

### **4.2.1 Recall und Precision**

Um die Qualität des Matchings zu beurteilen, werden Precision und Recall berechnet. Die Precision ist der Anteil der korrekt identifizierten Elemente an allen identifizierten Elementen, der Recall der Anteil aller korrekt identifiziert Elemente zur Gesamtzahl aller relevanten Elemente.

Der Recall wird nach der folgenden Formel berechnet:

$$\mathbf{REC} = t_p / (t_p + f_n) \quad (1)$$

wobei  $t_p$  true positive und  $f_n$  false negative, also korrekt oder fälschlich als negativ eingestufte Elemente bedeuten. Somit misst der Recall den Anteil der korrekt ermittelten Elemente an der Gesamtheit der korrekten Elemente, wobei die Gesamtheit der korrekten Elemente durch den Maßstabsdatensatz vorgegeben wird. Die Precision wird nach der folgenden Formel bestimmt:

$$\mathbf{PRE} = t_p / (t_p + f_p) \quad (2)$$

wobei  $t_p$  true positive und  $f_p$  false positive, also korrekt oder fälschlich als positiv eingestufte Elemente bedeuten. Precision misst somit den Anteil der korrekt ermittelten Elemente an der Gesamtheit der ermittelten Elemente.

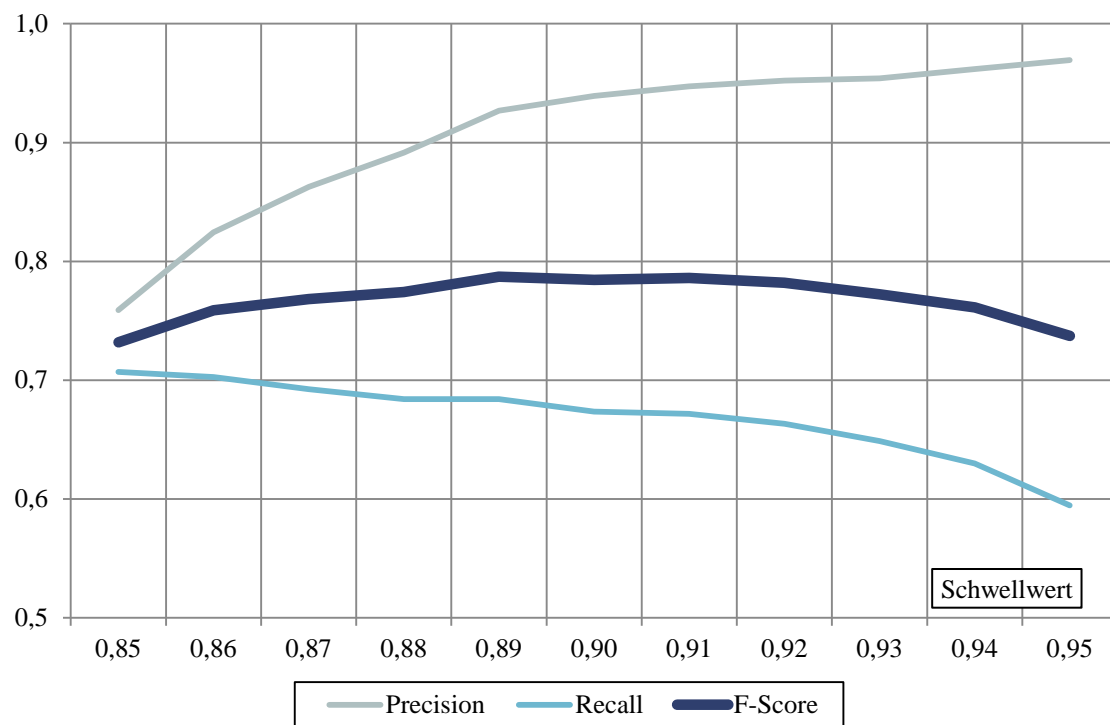
Diese Parameter werden mit zwei verschiedenen Arten von Fehlern verbunden. Wenn ein Typ-I-Fehler (oder false negative) auftritt, verringert diese die Recall-Rate, während ein Typ-II-Fehler (oder false positive) Precision verringert. Je höher Recall und Precision, desto besser ist das Matching (Baeza-Yates, Ribeiro-Neto 2011; Raffo, Lhuillery 2009).

String-Matching-Algorithmen standardisieren die Textteile jedoch, die miteinander verknüpft werden sollen, und reduzieren somit den Informationsgehalt der Daten (in unterschiedlichem Ausmaß) was zu höheren Recall-Raten führt. Zur gleichen Zeit aber verringern sie auch die Precision des Matchings, das heißt, dass sie höhere Typ-II-Fehler-Raten verursachen. Dies wurde von Raffo und Lhuillery (2009) empirisch bestätigt. Erhöht man also den Recall, verringert sich die Precision und umgekehrt. Daher können Matching-Verfahren bei großen Datenmengen nicht gleichzeitig Recall- und Precision-Werte von 1 erreichen, sondern es muss ein optimaler Kompromiss gefunden werden. Bei dieser Fragestellung wird in der Literatur das Maß F-Score verwendet, bei dem es sich um das harmonische Mittel aus Recall und Precision handelt. Bei Gleichgewichtung von Recall und Precision berechnet sich der F-Score wie folgt:

$$\mathbf{F} = 2 (\mathbf{PRE} \times \mathbf{REC}) / (\mathbf{PRE} + \mathbf{REC}) \quad (3)$$

Eine Übersicht über die berechneten Recall, Precision und F-Score Werte bei unterschiedlichen Schwellwerten  $t$  ist in Abbildung 2 dargestellt.

Abbildung 2 Recall, Precision und F-Score bei unterschiedlichen Schwellwerten



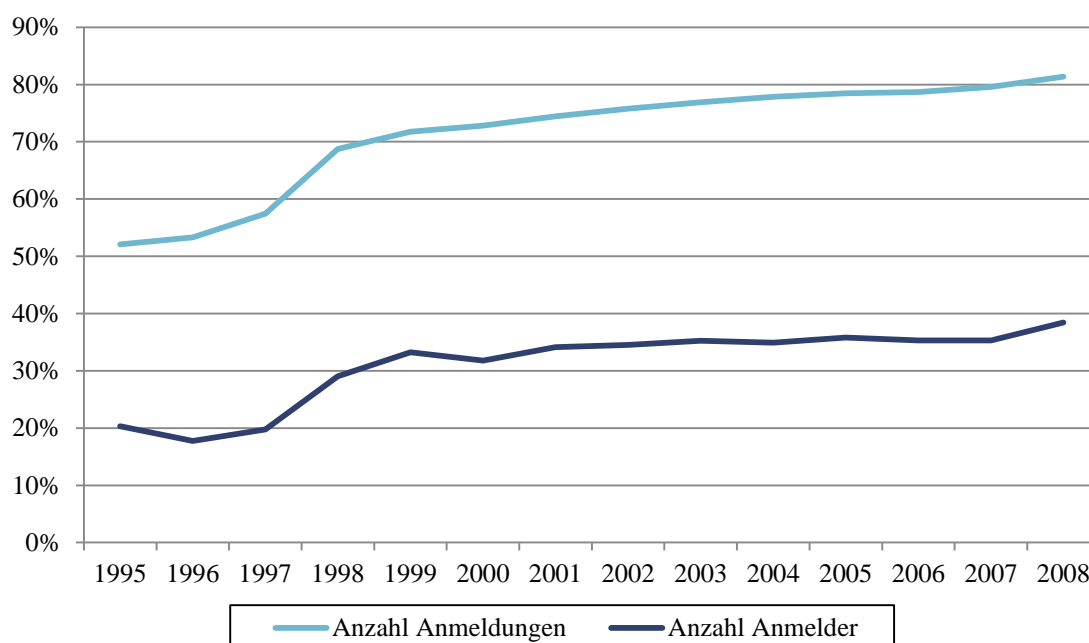
Aus Abbildung 2 wird ersichtlich, dass der F-Score bei einem Schwellwert von  $t=0,89$  am höchsten ist. Somit stellt dieser Schwellwert den optimalen Kompromiss zwischen Recall und Precision dar und wird daher für das Matching der SV-Datenbank mit PATSTAT verwendet.

Die Einschränkung eines optimalen Kompromisses bedeutet natürlich gleichzeitig, dass nicht alle Unternehmen mit Hilfe des automatisierten Verfahrens eindeutig zugeordnet werden können. Vor allem bei nicht mit Hilfe des Algorithmus getroffenen großen Patentanmeldern kann dies einen Verlust einer signifikanten Anzahl von Patentanmeldungen bedeuten. Aus diesem Grund wurden mit Hilfe einer Nachrecherche alle Unternehmen mit mehr als 100 transnationalen Patentanmeldungen im Zeitraum 2005 bis 2009, die in der SV-Datenbank, jedoch nicht im verknüpften Datensatz vorhanden waren, identifiziert. Diese fehlenden Patentanmelder wurden anschließend in einem finalen Schritt von Hand zu den Betrieben in der SV-Datenbank zugeordnet. Auch bei diesem Schritt ergibt sich das Problem, dass Unternehmen/Betriebe nicht zwingend die anmeldende und damit die im Patent genannte Einheit darstellen, sondern teilweise nur die Konzernmutter oder die Unternehmenszentrale als Patentanmelder fungieren. Aus diesem Grund müssen Unternehmen/Patentanmelder (besonders Großunternehmen) in einem letzten Schritt zusammengefasst werden.

### 4.3 Ergebnisse des Matchings

In diesem Abschnitt werden die Ergebnisse des Matchings dargestellt. Abbildung 3 zeigt die Gesamtabdeckung des gematchten Datensatzes in Bezug auf alle deutschen Patentanmelder, die mindestens ein Patent direkt am EPA oder direkt am DPMA angemeldet haben. Aus der Abbildung wird ersichtlich, dass am aktuellen Rand etwa 40% aller Patentanmelder vom Matchingalgorithmus erfasst und entsprechend den FuE-treibenden Unternehmen aus der SV-Datenbank zugeordnet wurden. Da die Abdeckung jedoch bei großen Unternehmen deutlich höher ist als bei klein- und mittelständischen Unternehmen (KMU), werden knapp über 80% der Patentanmeldungen am aktuellen Rand vom Datensatz abgedeckt. Das Matching wurde für die Prioritätsjahre 1997 bis 2009 durchgeführt. Allerdings wurde dieser Datenbestand mit dem in der SV-Datenbank vorhandenen Datenbestand der Jahre 2005 bis 2009 zusammengeführt. Dies resultiert in einer erhöhten Fehlerquote in früheren Jahren, weshalb die Abdeckung besonders in den früheren Jahren insgesamt deutlich geringer ist als am aktuellen Rand.

Abbildung 3 Gesamtabdeckung des Datensatzes, Anteil gematchter Patentanmeldungen an allen Anmeldungen, EPA+DPMA direkt

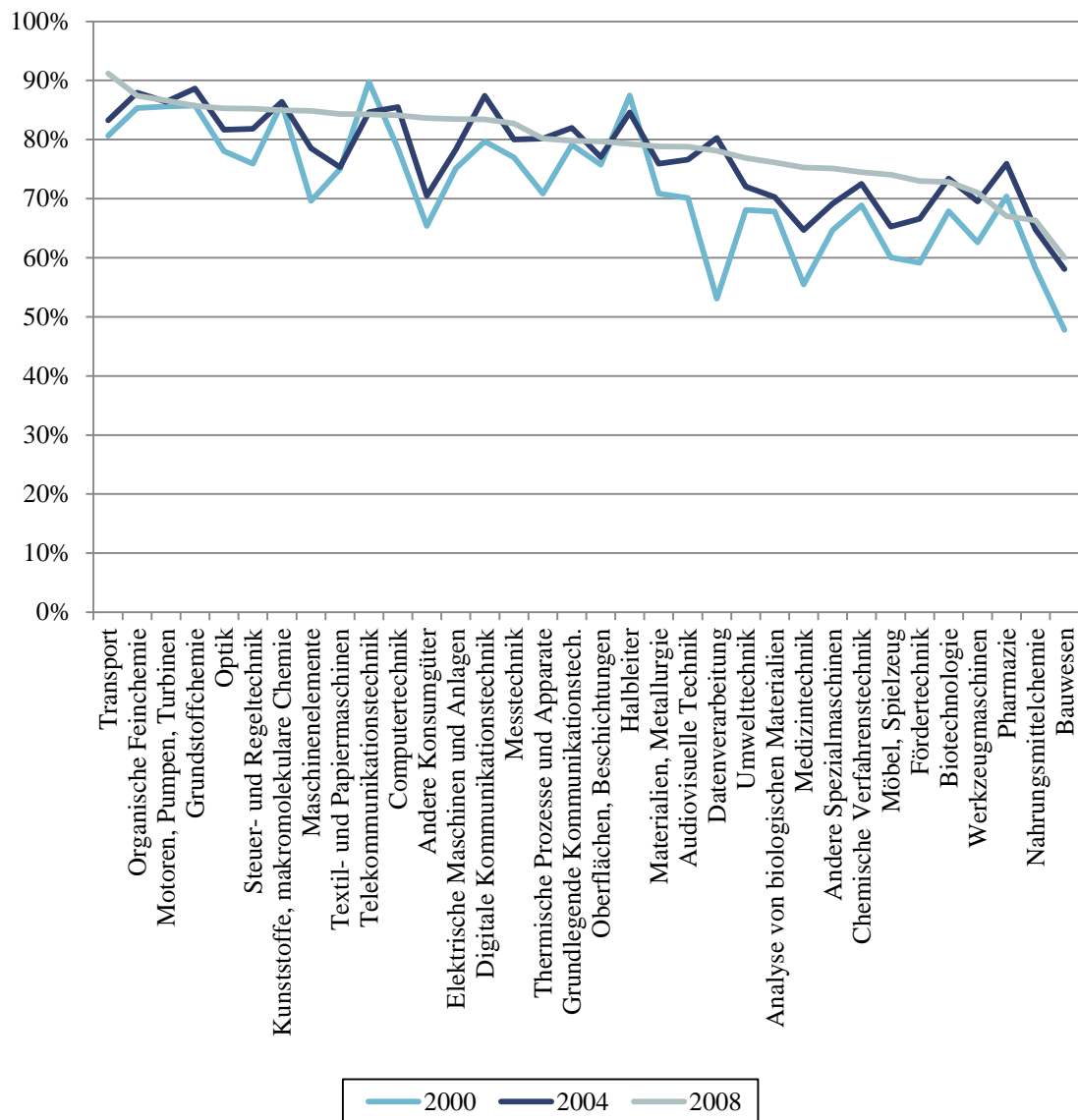


Quelle: EPA – PATSTAT, Berechnungen des Fraunhofer ISI

Abbildung 4 zeigt die feldspezifische Abdeckung des Datensatzes, also den Anteil gematchter Patentanmeldungen an allen Anmeldungen über alle Technologiefelder der 35 WIPO-Klassen (Schmoch 2008) hinweg. Eine besonders hohe Abdeckung von

knapp über 90% wird im Feld Transport erreicht, gefolgt von der organischen Feinchemie, dem Feld "Motoren, Pumpen Turbinen" und der Grundstoffchemie. Insgesamt gesehen wird also eine besonders gute Abdeckung in den Bereichen des Maschinenbaus und der Chemie erreicht. Die geringste Abdeckung mit nur knapp 60% findet sich im Feld "Bauwesen". Auch in der Pharmazie sowie der Biotechnologie ist die Abdeckung etwas schlechter als in den Feldern des Maschinenbaus (außer Werkzeugmaschinen), jedoch können auch hier Anteile über 70% erreicht werden.

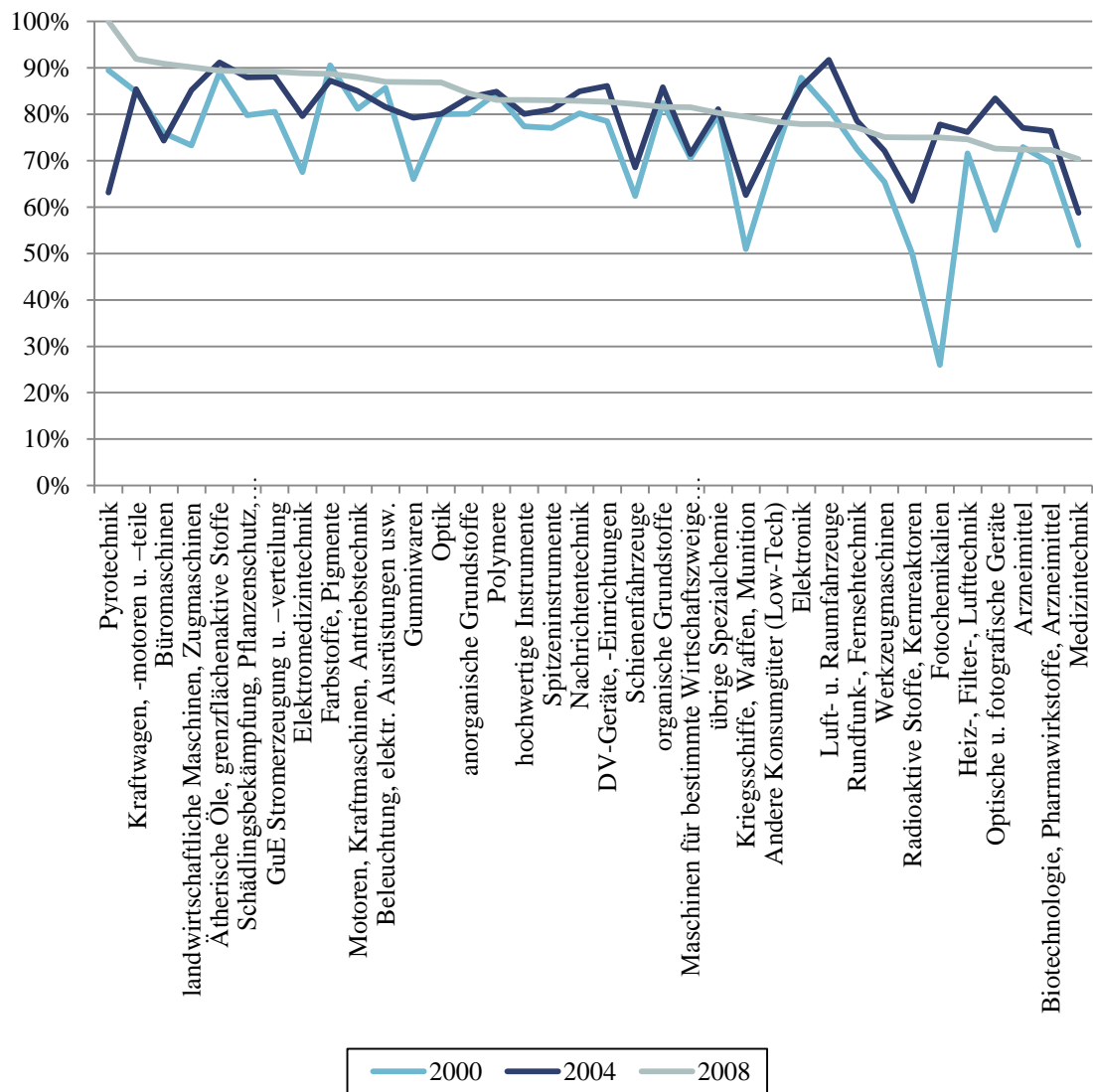
Abbildung 4 Feldspezifische Abdeckung des Datensatzes, Anteil gematchter Patentanmeldungen an allen Anmeldungen, EPA+DPMA direkt, alle Technologien



Quelle: EPA – PATSTAT, Berechnungen des Fraunhofer ISI

In Abbildung 5 ist die technologiefeldspezifische Abdeckung noch einmal in einer alternativen Differenzierung, nämlich nach 35 Hochtechnologie-Bereichen, einschließlich einer residualen "low-tech" Kategorie (Legler, Frietsch 2007) abgebildet. Auch hier zeigt sich, dass eine besonders hohe Abdeckung in den Feldern des Maschinenbaus erreicht werden kann und viele Felder der Chemie eine Abdeckung von über 90% erreichen. In den Teilbereichen der Pharmazie, das heißt bei Arzneimitteln, der Biotechnologie und Pharmawirkstoffen sowie in der Medizintechnik, werden die geringsten Anteile mit knapp über 70% Abdeckung erreicht.

Abbildung 5 Feldspezifische Abdeckung des Datensatzes, Anteil gemachter Patentanmeldungen an allen Anmeldungen, EPA+DPMA direkt, Hochtechnologiefelder

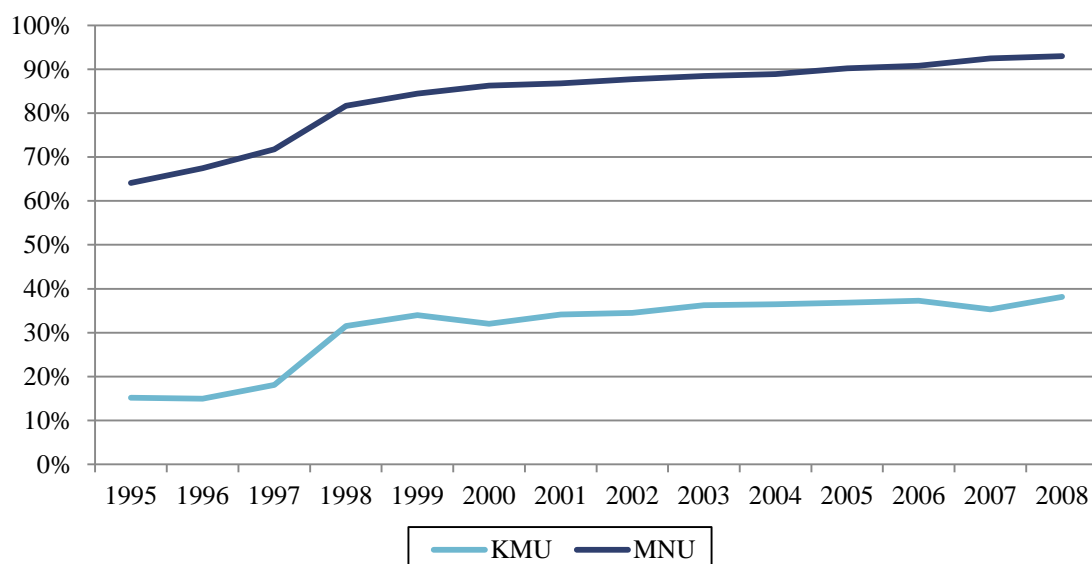


Quelle: EPA – PATSTAT, Berechnungen des Fraunhofer ISI

Wie bereits angesprochen ist die Abdeckung des Datensatzes für große, multinationale Unternehmen mit über 500 Mitarbeitern deutlich höher als für KMU (Abbildung 6). Während bei den Großunternehmen am aktuellen Rand Anteile von über 90% erreicht werden, sind dies für KMU nur knapp 40%.

Für die unvollständige Abdeckung des Matchings in Bezug auf alle Patentanmelder zwischen 1997 und 2009 gibt es mehrere Gründe. Zum einen sind nicht alle patentierenden Unternehmen von der SV-Datenbank abgedeckt bzw. manche Patentanmelder fungieren als Tochterunternehmen einiger größerer Firmen. Dies ist beispielsweise für die Firmen "Osram" und "BSH Bosch und Siemens Hausgeräte" der Fall, die im SV-Datensatz als Töchter der Firma "Siemens" geführt und nicht gesondert ausgewiesen werden. Zusätzlich gibt es auch Patente, die von nicht FuE-treibenden Unternehmen angemeldet werden und somit per Definition nicht vom Matchingalgorithmus zugeordnet werden können. Hinzu kommt, wie bereits erwähnt wurde, dass der Matching-Algorithmus zwar für die Jahre 1997 bis 2009 durchgeführt wurde, die kürzere Periode des SV-Datensatzes jedoch zu einer geringeren Abdeckung in früheren Jahren führt. Der wesentliche Grund für die unvollständige Abdeckung der Patentanmelder ist, dass ein F-Score Matching nicht 100% erreichen kann. Es muss immer ein "optimaler Kompromiss" aus Recall- und Precision-Werten gefunden werden, da ansonsten die Genauigkeit des Matchings leidet und es zu Fehlzugeordnungen kommen kann.

Abbildung 6 Abdeckung des Datensatzes nach Unternehmensgröße, Anteil gematchter Anmeldungen an allen Anmeldungen, EPA+DPMA direkt



Quelle: EPA – PATSTAT, Berechnungen des Fraunhofer ISI

## 5 Der Mikrodatsatz

Wie eingangs bereits erwähnt, wurden die Daten zu FuE-Aufwendungen der Unternehmen auf der Mikroebene zu Daten von Patentanmeldern zugespielt. Dadurch entsteht ein Unternehmenspanel der Jahre 1997 bis 2009, das zwar in erster Linie der Erstellung der Konkordanz zwischen FuE- und Patentdaten dient, jedoch auch den Vorteil hat, dass Analysen zur Verbindung zwischen FuE-Ausgaben und Patentanmeldungen auf der Unternehmensebene durchgeführt werden können. Der vorliegende Mikrodatsatz wird im Folgenden kurz vorgestellt.

### 5.1 Verwendete Patentindikatoren

In diesem Abschnitt werden die im Datensatz beinhalteten Patentindikatoren, die für die weiteren Analysen verwendet werden können, näher erläutert. Neben der absoluten Anzahl der Patente pro Unternehmen und Jahr, sind die Patente zusätzlich nach zwei Technologiefelddefinition, nämlich nach 35 WIPO-Klassen (Schmoch 2008) sowie 35 High-Tech-Bereichen, einschließlich einer residualen "weniger forschungsintensive Industrien (low-tech)" Kategorie (Legler, Frietsch 2007) differenziert. Alle Patente werden nach zwei Ämterdefinitionen unterschieden. Zum einen sind dies transnationale Patentanmeldungen (Frietsch, Schmoch 2010). Transnationale Anmeldungen sind definiert als alle direkten Anmeldungen am EPA und alle PCT-Anmeldungen, egal ob diese zum EPA weitergeleitet werden oder nicht. Doppelzählungen werden hierbei ausgeschlossen. Etwas einfacher ausgedrückt, werden alle Patentfamilien mit mindestens einer PCT- oder EPA-Anmeldung berücksichtigt. Zusätzlich zur Definition der transnationalen Patentanmeldungen beinhaltet der Datensatz alle Patente für Deutschland bzw. den deutschen Markt. Diese sind definiert als alle direkten Anmeldungen am Deutschen Patent- und Markenamt (DPMA) plus alle PCT- und EPA Anmeldungen deutscher Anmelder unter Ausschluss von Doppelzählungen. Mit Hilfe dieser Abgrenzung werden somit alle Patente gezählt, die früher oder später Schutzwirkung in Deutschland erreichen wollen.<sup>1</sup>

Zusätzlich zu den Patentanmeldungen enthält der Datensatz die Anzahl der Zitierungen, die die Patente der Unternehmen pro Jahr innerhalb eines 4-Jahres Zeitfensters erhalten. Diese so genannten Vorwärtszitierungen sind der am weitesten verbreitete Indikator, um den Wert oder die Qualität von Patenten bzw. eines Patentportfolios zu bestimmen. Die Annahme dabei ist, dass die Anzahl der Vorwärtszitierungen den Grad, zu dem ein Patent zur Entwicklung neuer Technologien beiträgt, abbildet. Daher

---

<sup>1</sup> Diesem Konzept liegt die Annahme zugrunde, dass alle Anmeldungen deutscher Anmelder am EPA bzw. über das PCT-Verfahren früher oder später an das DPMA weitergeleitet werden.

zeigen Vorwärtszitationen die technologische Signifikanz eines Patents auf (Deng et al. 1999; Narin et al. 1987; Trajtenberg 1990). Zusätzlich zu den Vorwärtszitationen enthält der Datensatz Informationen zur durchschnittlichen Größe der Patentfamilie pro Unternehmen und Jahr. Eine Patentfamilie ist definiert als die Anzahl der Länder bzw. Patentämter an denen ein Patent angemeldet wurde (Martinez 2009). Die durchschnittliche Familiengröße dient dabei als Internationalisierungsindikator, der die Breite der Marktabdeckung indiziert. Jedoch kann die auch durchschnittliche Familiengröße als Qualitäts- oder Wertindikator interpretiert werden, da Anmelde- und Aufrechterhaltungsgebühren für jedes zusätzliche Land bezahlt werden müssen in dem ein Patent angemeldet wird. Es kann angenommen werden, dass ein Anmelder diese Kosten nur tragen wird, wenn er sich einen entsprechenden Umsatz von der zusätzlichen Marktabdeckung verspricht (Putnam 1996; Schmoch et al. 1988).

## **5.2      Verwendete Indikatoren aus der SV-Datenbank**

Neben den für das Matching erforderlichen Informationen zum Unternehmensnamen und zum Standort des Unternehmens ist der zentrale Indikator aus der SV-Datenbank für die Erstellung der Konkordanzliste die Summe der internen FuE-Aufwendungen. Dazu zählen alle finanziellen Aufwendungen die bei Forschungs- und Entwicklungsaktivitäten anfallen, die innerhalb des eigenen Unternehmens durchgeführt werden. Dies umfasst insbesondere die FuE-Personalaufwendungen, FuE-Sachaufwendungen und Investitionen für FuE. Von den internen FuE-Aufwendungen zu unterscheiden sind externe Aufwendungen, die für FuE-Aufträge an Externe anfallen. FuE-Aufwendungen werden nach OECD-Vorgabe grundsätzlich als nominale Größe erfasst und sind somit nicht inflationsbereinigt. Sie sind der zentrale Indikator, um jahresaktuell quantifizieren zu können, welche Werte als Inputgröße in Forschung und Entwicklung investiert werden. In das hier beschriebene Projekt flossen Angaben zu FuE-Aufwendungen der Jahre 2007 und 2009 ein.

## **5.3      Analyse des Zeitverzugs zwischen FuE-Ausgaben und Patentanmeldungen**

Für die Erstellung der Konkordanzliste ist es essentiell zu erfahren, mit welchem Zeitverzug sich FuE-Ausgaben im Durchschnitt in anschließender Patentierung niederschlagen. Aus diesem Grund wurden mehrere Regressionsmodelle mit transnationalen Anmeldungen bzw. Patentanmeldungen in Deutschland als abhängiger oder zu erklärender Variable berechnet. Die erklärende oder unabhängige Variable pro Modell sind die FuE-Ausgaben der Unternehmen zu den jeweiligen Zeitpunkten, das heißt also im

gleichen Jahr, mit einem Jahr Zeitverzug und so weiter. Zur Kontrolle periodenspezifischer Effekte sind in jedem Modell zusätzlich Jahresdummies enthalten.

Insgesamt wurden drei verschiedene Modellspezifikationen zum Testen des Zusammenhanges zwischen FuE-Ausgaben auf Patentanmeldungen in den verschiedenen Jahren verwendet: ein OLS-Modell, eine negativ-binomial Regression und eine fixed-effects Panelregression (XT (FE)). Da die Anzahl der Patente eine Zählvariable darstellt, kann ein OLS-Modell zu ineffizienten, inkonsistenten und verzerrten Schätzern führen (Long 1997). Die negativ-binomial Regression korrigiert für diesen Effekt und stellt daher den in diesem Fall effizienteren Schätzer dar. Aufgrund der guten Datenlage kann zusätzlich eine Panelregression durchgeführt werden, die der vorliegenden Datenstruktur am stärksten Rechnung trägt.

Tabelle 3 Modellfit ( $R^2$ ) verschiedener Schätzansätze zum Zusammenhang von FuE-Ausgaben und Patentanmeldungen mit Zeitverzug t-x

<i>aV: Transnationale Patentanmeldungen</i>	OLS	Neg. Bin.	XT (FE)
FuE Ausgaben <sub>t-0</sub>	0,138	0,082	0,057
FuE Ausgaben <sub>t-1</sub>	0,149	0,070	0,052
FuE Ausgaben <sub>t-2</sub>	0,137	0,068	0,034
FuE Ausgaben <sub>t-3</sub>	0,146	0,067	0,002
FuE Ausgaben <sub>t-4</sub>	0,143	0,065	0,016
FuE Ausgaben <sub>t-5</sub>	0,140	0,062	0,025
Zeitdummies	JA	JA	JA
<i>aV: Patentanmeldungen in Deutschland</i>	OLS	Neg. Bin.	XT (FE)
FuE Ausgaben <sub>t-0</sub>	0,227	0,104	0,116
FuE Ausgaben <sub>t-1</sub>	0,247	0,085	0,121
FuE Ausgaben <sub>t-2</sub>	0,230	0,083	0,064
FuE Ausgaben <sub>t-3</sub>	0,246	0,083	0,011
FuE Ausgaben <sub>t-4</sub>	0,235	0,079	0,000
FuE Ausgaben <sub>t-5</sub>	0,232	0,077	0,009
Zeitdummies	JA	JA	JA

Quelle: EPA – PATSTAT, Berechnungen des Fraunhofer ISI

Für die berechneten Modelle kann anschließend der Modellfit ( $R^2$ ) verglichen werden (Tabelle 3), um zu untersuchen, in welchem Jahr die FuE-Ausgaben die stärkste Erklärungskraft für die Anzahl der Patentanmeldungen haben. Daraus kann anschließend abgeleitet werden, mit welchem Zeitverzug die FuE-Ausgaben für die spätere Erstellung der Konkordanzliste verwendet werden. Auf Basis der Analyse der  $R^2$ -Werte zeigt sich, dass der Zusammenhang zwischen FuE-Ausgaben und Patentanmeldungen im gleichen Jahr in der negativ-binomial Regression und im FE-Modell am höchsten ist. Das OLS-Modell deutet, mit einem leicht höheren Wert als in t-0 auf einen Zeitverzug von einem Jahr hin. Da das negativ-binomial Modell als auch das FE-Modell jedoch die adäquatere Modellierung für die vorliegende Variable darstellen, wird ein Zeitverzug von t-0 für die weiteren Analysen und zur Erstellung der Konkordanzliste verwendet.

## 6 Die Konkordanz zwischen Wirtschaftszweigen und Technologiefeldern

Kern des Projekts ist die Erstellung einer Konkordanz zwischen Wirtschaftszweigen und Technologiefeldern zur Umrechnung von Aufwendungen für Forschung und Entwicklung, die nach Sektoren bzw. Wirtschaftszweigen (aktuell nach der Klassifikation der Wirtschaftszweige (WZ) 2008) erhoben werden, in Aufwendungen für Forschung und Entwicklung nach Technologien bzw. Technologiefeldern. Da insbesondere große, multinationale Unternehmen häufig technologisch heterogen sind, ist die Wirtschaftszweigsystematik nicht in der Lage, die FuE-Aufwendungen der Unternehmen adäquat abzubilden. Zahlreiche weitere Indikatoren und Informationen liegen nur auf der Ebene von Technologien vor. Dazu zählen beispielsweise Förderprogramme, wissenschaftsnahe Indikatoren wie Beschäftigte in öffentlichen Einrichtungen, sowie auch Patente. Zur Lösung dieses Problems wird das in den vorangegangenen Kapiteln dargestellte Matching von Patentanmeldern und FuE-treibenden Unternehmen auf der Mikro-Ebene verwendet. Die Technologieprofile der Unternehmen, die anhand der Patentaktivitäten bestimmt werden, können dazu verwendet werden, die FuE-Aufwendungen auf die Technologiefelder umzurechnen. Dabei wird zunächst eine Matrix von Wirtschaftszweigen und Technologieklassen erarbeitet, die anschließend auf die aggregierten FuE-Daten nach Wirtschaftszweigen angewendet wird. Die FuE-Aufwendungen der Wirtschaftszweige werden entsprechend der Anteile der Patente auf die Technologiefelder verteilt. Die FuE-Aufwendungen werden anschließend über die Technologiefelder summiert, was dann die technologiefeldspezifischen FuE-Aufwendungen ergibt.

Grundannahme dieser Vorgehensweise ist, dass Patente ein geeigneter Indikator sind, um die technologieorientierten Forschungs- und Entwicklungsausgaben abzubilden. Anders formuliert liegt die Annahme zugrunde, dass Patente ein wesentlicher Output von FuE-Prozessen sind. Es ist allerdings so, dass selbst in technologieorientierten Wirtschaftszweigen nicht alle Ausgaben unmittelbar in einem Patent münden. Unterstellt man jedoch, dass die Anteile von patentierten und nicht-patentierten Forschungsaktivitäten innerhalb von Technologiefeldern – im Wesentlichen – ähnlich bzw. identisch zwischen den Akteuren sind, dann stellen Patente auch für die nicht patentierten bzw. nicht messbaren FuE-Aktivitäten eine geeignete Messgröße dar.

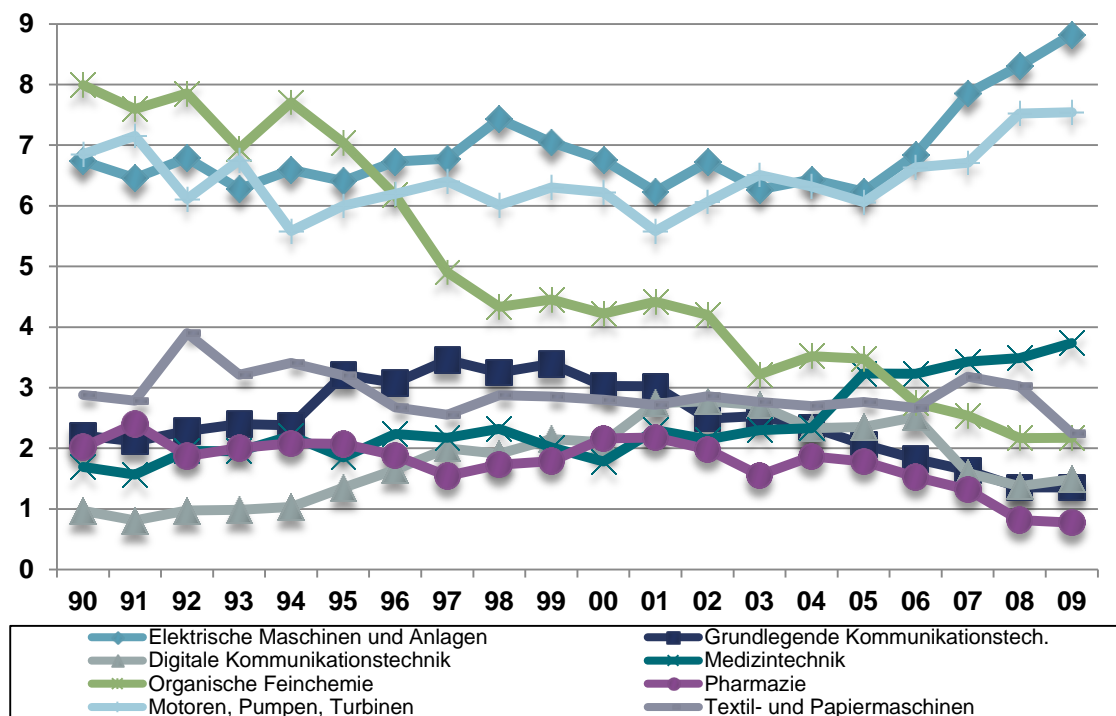
Eine implizite Annahme bei der Erstellung der Umrechnung von Forschungsausgaben nach Wirtschaftszweigen in Technologiefelder ist – und diese Annahme sollte in regelmäßigen Abständen überprüft werden –, dass über die Zeit die Relation zwischen den FuE-Aufwendungen und den Patenten innerhalb von Technologiefeldern mehr oder weniger konstant ist, denn sonst verschieben sich die Relationen zwischen den Technologiefeldern unbemerkt von der Umrechnung durch die Matrix. Anders formuliert,

wenn in einem Technologiefeld mehr (oder weniger) FuE-Aufwendungen pro Patent notwendig werden, in einem anderen aber nicht, dann muss diesem Technologiefeld in der Konkordanzmatrix ein höheres (niedrigeres) Gewicht eingeräumt werden.

Die Anteile der 35 Technologiefelder an allen Patenten der FuE-treibenden Unternehmen in Deutschland ist für die Jahre 1990 bis 2009 in

Tabelle 4 dargestellt. Zur besseren Veranschaulichung sind die Entwicklungen ausgewählter Technologiefelder zusätzlich in Abbildung 7 über die Zeit abgetragen. Es zeigt sich eine nennenswerte Veränderung der Anteile der einzelnen Technologiefelder über die Zeit. Insbesondere die Feinchemie, aber auch die Pharmazie verlieren deutlich an Anteilen, während insbesondere der Bereich Transport seit dem Beginn der 1990er Jahre erheblich angewachsen ist. Auch Motoren, Pumpen und Turbinen sowie Elektrische Maschinen und Anlagen waren zuletzt deutlich gewachsen. Es lässt sich mit dieser Aufstellung auch die Veränderung innerhalb der Kommunikationstechnologien verfolgen. Während die grundlegenden Kommunikationstechnologien ebenso wie die Telekommunikationstechnik abgenommen haben, ist die digitale Kommunikationstechnik deutlich gewachsen. Es finden sich jedoch auch Technologiefelder, deren Anteile sich im Zeitverlauf kaum verändert haben. Hierzu gehören beispielsweise die Optik, die Analyse von biologischen Materialien oder die Spezialmaschinen.

Abbildung 7: Anteile (in %) der Technologiefelder mit besonders großen Abweichungen zwischen 2005 und 2009, Patente in Deutschland der FuE-treibenden Unternehmen



Quelle: eigene Berechnungen auf Basis von EPA-PATSTAT und SV Wissenschaftsstatistik.

Tabelle 4: Anteile der 35 Technologiefelder an allen Patenten in Deutschland der FuE-treibenden Unternehmen 1990-2009

	1990	1995	2000	2005	2009
<b>Elektrische Maschinen und Anlagen</b>	6.7	6.4	6.8	6.2	8.8
<b>Audiovisuelle Technik</b>	3.4	2.6	2.3	2.4	1.7
<b>Grundlegende Kommunikationstechn.</b>	2.2	3.2	3.0	2.1	1.4
<b>Digitale Kommunikationstechnik</b>	1.0	1.4	2.1	2.4	1.5
<b>Telekommunikationstechnik</b>	1.8	1.6	1.3	1.0	0.6
<b>Computertechnik</b>	1.3	2.0	3.1	3.3	2.8
<b>Datenverarbeitung</b>	0.0	0.1	0.3	0.5	0.4
<b>Halbleiter</b>	1.5	1.9	3.0	2.3	2.2
<b>Optik</b>	2.3	1.7	2.1	2.2	1.8
<b>Messtechnik</b>	5.5	5.7	5.6	6.1	6.7
<b>Analyse von biologischen Materialien</b>	0.5	0.5	0.7	0.6	0.5
<b>Steuer- und Regeltechnik</b>	1.6	2.6	2.6	2.3	2.7
<b>Medizintechnik</b>	1.7	1.8	1.8	3.2	3.7
<b>Organische Feinchemie</b>	8.0	7.0	4.2	3.5	2.2
<b>Biotechnologie</b>	1.4	1.0	1.7	1.1	0.5
<b>Pharmazie</b>	2.0	2.1	2.2	1.8	0.8
<b>Kunststoffe, makromolekulare Chemie</b>	6.1	4.3	3.0	2.3	1.4
<b>Nahrungsmittelchemie</b>	0.4	0.3	0.3	0.4	0.2
<b>Grundstoffchemie</b>	7.2	5.6	3.3	3.0	2.4
<b>Materialien, Metallurgie</b>	2.0	2.0	2.0	1.5	1.6
<b>Oberflächen, Beschichtungen</b>	1.8	1.5	2.2	1.6	1.7
<b>Mikro- und Nanonstrukturen</b>	0.0	0.1	0.2	0.4	0.4
<b>Chemische Verfahrenstechnik</b>	4.1	4.3	3.3	2.9	2.6
<b>Umwelttechnik</b>	1.7	1.9	1.6	1.6	2.0
<b>Fördertechnik</b>	2.4	3.0	2.8	2.9	3.1
<b>Werkzeugmaschinen</b>	2.1	2.2	3.1	3.2	3.6
<b>Motoren, Pumpen, Turbinen</b>	6.8	6.0	6.2	6.1	7.5
<b>Textil- und Papiermaschinen</b>	2.9	3.2	2.8	2.8	2.2
<b>Andere Spezialmaschinen</b>	3.0	2.8	3.2	3.5	3.3
<b>Thermische Prozesse und Apparate</b>	1.4	1.7	1.8	1.9	2.8
<b>Maschinenelemente</b>	5.1	5.6	5.9	7.1	7.7
<b>Transport</b>	7.6	9.4	10.6	12.4	12.4
<b>Möbel, Spielzeug</b>	1.2	1.1	1.2	1.3	1.6
<b>Andere Konsumgüter</b>	1.4	1.4	1.3	1.7	2.6
<b>Bauwesen</b>	1.6	2.3	2.3	2.6	2.5

Quelle: eigene Berechnungen auf Basis von EPA-PATSTAT und SV Wissenschaftsstatistik.

Tabelle 5 Die Relationen zwischen den Technologiefeldern, wie sie in Tabelle 4 dargestellt sind, können sich über die Zeit verändern. Es bleibt jedoch die Frage offen, ob die Relationen von FuE-Aufwendungen und Patenten innerhalb der Technologiefelder über die Zeit konstant bleibt. Eine konstante Relation kann dabei nicht erwartet werden. So belegt Tabelle 5, dass im Bereich der Nanotechnologie ein hoher interner FuE-Aufwand geleistet werden muss, um ein Patent anmelden zu können, was jedoch langfristig die Marktdurchdringung und damit den Erfolg einer Technologie gefährdet. Insgesamt finden sich massive Veränderungen im Zeitverlauf, die sich nicht alleine durch die Preisentwicklung bzw. Inflation erklären lassen (die hier angegebenen Werte sind FuE-Ausgaben in nominalen Größen und nicht etwa in realen Größen). Der Strukturbruch in 2007 soll an dieser Stelle zunächst außen vor bleiben. Es zeigen sich zwischen 2001 und 2007 in einigen Bereichen deutliche Veränderungen, wie beispielsweise in „Grundlegender Kommunikationstechnologie“, „Datenverarbeitung“, aber auch in den Bereichen „Analyse von biologischen Materialien“, „Biotechnologie“, aber auch im dem gewichtigeren Feld „Pharmazie“, wenngleich weniger deutlich ausgeprägt. Auch in der Zeit zwischen 2007 und 2009 finden sich die größten Veränderungen in den Bereichen „Biotechnologie“ und „Pharmazie“, sowie bei „Textil- und Papiermaschinen“. Insgesamt geben gerade diese Veränderungen Anlass, die Relationen in Frage zu stellen. Allerdings sind gerade in den oben genannten Bereichen die Effekte der Finanzkrise sehr massiv, die sich auf das Patentierverhalten deutlich stärker auswirken als auf die FuE-Investitionen. Man kann daraus schließen, dass eine Überprüfung in regelmäßigen Abständen notwendig und sinnvoll ist. Allerdings wird die weitere Analyse auch belegen, dass die unmittelbaren Effekte auf die Verteilung der absoluten FuE-Ausgaben weniger deutlich ausgeprägt sind, da dies auf relativen und nicht auf absoluten Werten beruht.

Tabelle 5: Interne FuE-Aufwendungen je Patentanmeldung (FuE-Intensität) der Technologiefelder, 2001-2009

	2001	2003	2005	2007	2007 neu*	2009
<b>Elektr. Maschinen und Anlagen</b>	363	378	370	348	303	308
<b>Audiovisuelle Technik</b>	157	164	165	177	235	294
<b>Grundl. Kommunikationstech.</b>	828	1031	1093	1482	1926	2299
<b>Digitale Kommunikationstechnik</b>	1175	1195	1257	1446	1706	1997
<b>Telekommunikationstechnik</b>	202	209	246	379	750	952
<b>Computertechnik</b>	208	235	256	321	741	887
<b>Datenverarbeitung</b>	217	252	269	428	1315	1470
<b>Halbleiter</b>	417	474	513	573	639	694
<b>Optik</b>	151	174	172	215	276	307
<b>Messtechnik</b>	481	508	464	497	539	580
<b>Analyse von biolog. Materialien</b>	143	198	214	283	325	381
<b>Steuer- und Regeltechnik</b>	254	272	302	333	407	422
<b>Medizintechnik</b>	242	243	211	213	285	329
<b>Organische Feinchemie</b>	474	640	625	801	1002	1493
<b>Biotechnologie</b>	409	601	665	788	1062	1602
<b>Pharmazie</b>	2639	3237	3209	3848	3269	5708
<b>Kunststoffe, makromol. Chemie</b>	336	355	360	475	611	834
<b>Nahrungsmittelchemie</b>	419	544	481	581	656	798
<b>Grundstoffchemie</b>	425	528	460	501	757	1070
<b>Materialien, Metallurgie</b>	724	795	915	1033	955	1057
<b>Oberflächen, Beschichtungen</b>	777	760	840	941	722	761
<b>Mikro- und Nanostrukturen</b>	5845	4737	3087	4419	5367	4710
<b>Chemische Verfahrenstechnik</b>	478	552	604	663	697	896
<b>Umwelttechnik</b>	76	83	84	95	112	115
<b>Fördertechnik</b>	173	184	164	195	117	128
<b>Werkzeugmaschinen</b>	149	145	157	158	110	118
<b>Motoren, Pumpen, Turbinen</b>	380	375	401	409	417	386
<b>Textil- und Papiermaschinen</b>	1090	1103	1073	1117	895	1408
<b>Andere Spezialmaschinen</b>	135	151	140	171	117	135
<b>Therm. Prozesse und Apparate</b>	73	80	84	94	132	127
<b>Maschinenelemente</b>	154	150	154	155	119	127
<b>Transport</b>	2087	1977	1901	2118	1794	1825
<b>Möbel, Spielzeug</b>	95	104	90	103	161	163
<b>Andere Konsumgüter</b>	71	64	68	65	80	94
<b>Bauwesen</b>	58	60	61	78	100	106

\* Die Spalte „2007 neu“ beinhaltet dabei die Abgrenzung nach der neuen Wirtschaftszweigsystematik 2008, während die Spalte „2007“ die Daten auf Basis der Wirtschaftszweigsystematik von 2003 ausweist.

Quelle: eigene Berechnungen auf Basis von EPA-PATSTAT und SV Wissenschaftsstatistik.

## 6.1 Gewichtung der Patentdaten

Patente sind nicht in gleicher Weise für alle Wirtschaftszweige bzw. alle Technologien relevant bzw. nicht in allen Wirtschaftszweigen können alle Erfindungen patentiert werden. Wie bereits beschrieben, ist die Relation zwischen patentierten und nicht-patentierten Forschungsergebnissen zwischen den Technologiefeldern unterschiedlich. Hinzu kommt, dass FuE in manchen Feldern teurer und in anderen weniger teuer ist. Anders formuliert: in einem Technologiefeld muss für Forschung und Entwicklung bis zur Erreichung eines Patents mehr, in einem anderen Technologiefeld deutlich weniger aufgewendet werden. Die Relation der FuE-Aufwendungen je Patentanmeldung variiert folglich stark zwischen den Technologiefeldern. Dies ist bei der Berechnung der FuE-Aufwendungen nach Technologiefeldern mithilfe der Patente zu berücksichtigen. Um dies zu erreichen, wurde eine Gewichtung der Patente je Technologiefeld mithilfe einer Schätzung/Berechnung der Patentintensität vorgenommen.

Wie bereits eingangs zu diesem Kapitel beschrieben, wird bei dieser Vorgehensweise unterstellt, dass die Anteile von patentierten und nicht-patentierten Forschungsaktivitäten innerhalb von Technologiefeldern – im Wesentlichen – ähnlich bzw. identisch sind. Unterschiede zwischen den Technologiefeldern werden durch die Gewichtung berücksichtigt.

Wenn im Maschinenbau beispielsweise nur 50% der Ergebnisse von FuE-Prozessen patentiert werden oder patentierbar sind und in der Biotechnologie beispielsweise 100%, dann sind diese beiden Bereiche zunächst nicht unmittelbar miteinander vergleichbar. Daneben kann beispielsweise die Forschung in der Biotechnologie bis zur Erreichung eines Patents deutlich teurer sein als im Maschinenbau. Insgesamt unterscheidet sich also die Patentintensität in der Biotechnologie von der Patentintensität im Maschinenbau. Hier ein Beispiel: Betreibt ein Unternehmen mit einer Million Euro Forschungsbudget sowohl Forschung im Maschinenbau als auch in der Biotechnologie und meldet in jedem der beiden Bereiche jeweils 5 Patente an, dann bedeutet das nicht, dass sich das Forschungsbudget 50:50 auf die beiden Felder verteilt. Die Verteilung ist eher 40:60 oder gar 30:70. Diese ungleiche Verteilung muss bei der Umrechnung der FuE-Aufwendungen nach Wirtschaftszweigen in Technologien berücksichtigt werden. Da im Allgemeinen Unternehmen zwar einem Wirtschaftszweig zugeordnet werden, diese Unternehmen aber nur selten ein einziges und statistisch überschneidungsfrei definiertes Technologiefeld bearbeiten, muss eine sinnvolle Relation zwischen den Technologiefeldern eines Unternehmens gefunden werden. Da der hier vorgestellte Ansatz zwischen 35 Technologiefeldern unterscheidet, muss auch eine Relation dieser 35 Technologiefelder untereinander gefunden werden. Anders formuliert, die Patentintensität – definiert als FuE-Aufwendungen je Patent – der 35 Felder muss berücksichtigt werden. Wir suchen also eine Gewichtung von Technologien unterei-

inander, bezogen auf die erforderlichen FuE-Aufwendungen. Da wir diese Gewichtung nicht kennen bzw. nicht aus den vorhandenen Daten berechnen können – denn es fehlt ja gerade die Verteilung der Technologien innerhalb der einzelnen Wirtschaftszweige – müssen wir diese mithilfe externer Quellen berechnen bzw. schätzen.

Hierzu wurde auf eine bestehende, manuelle/intellektuelle Konkordanz zwischen Wirtschaftszweigen und 19 Technologieklassen zurückgegriffen (Schmoch et al. 2003; Schmoch, Gauch 2004). Auf Basis dieser Konkordanz von Wirtschaftszweigen und Technologiefeldern wurde zunächst die Patentintensität der Technologiefelder mithilfe der FuE-Daten des Jahres 2007<sup>2</sup> berechnet. Konkret lagen die FuE-Aufwendungen nach Wirtschaftszweigen für das Jahr 2007 vor sowie die Anzahl der Patente in Technologiefeldern im Jahr 2007. Die intellektuelle Konkordanz nach Schmoch und Gauch (2004) unterscheidet 19 Technologiefelder bzw. Wirtschaftszweige. Zunächst wurden sowohl die FuE-Daten des Jahres 2007 als auch die Patentdaten des gleichen Jahres so zusammengefasst, dass sie diese 19 Felder abbilden. Anders formuliert, es wurden die FuE-Aufwendungen und die Patente innerhalb von Wirtschaftszweigen bzw. Technologieklassen aufsummiert. Anschließend wurden die FuE-Aufwendungen durch die Anzahl der Patente in jedem der 19 Felder geteilt und so die Patentintensität berechnet. Es entstand somit ein durchschnittlicher Preis (in FuE-Aufwendungen) pro Patent innerhalb von Technologiefeldern.

Im nächsten Schritt wurden die 19 Technologiefelder den 35 hier verwendeten Technologiefeldern zugeordnet. Jedes der 35 Technologiefelder erhielt dann die zuvor errechnete Patentintensität als Gewicht. Wenn eines der 19 Felder mehr als einem der 35 Felder zugeordnet wurde, dann hat jedes der 35 Felder die gleiche Patentintensität erhalten. Da die Patentintensität bereits ein Durchschnittswert darstellt, muss sie nicht weiter auf mehrere Klassen aufgeteilt werden, sondern kann diesen direkt zugeordnet werden. Die Tatsache, dass zur Berechnung der Gewichte zunächst lediglich 19 Felder verwendet wurden und diese dann den hier verwendeten 35 Feldern zugeordnet wurden, ist für die finale Berechnung der FuE-Aufwendungen nach den 35 Feldern kein nachhaltiger Nachteil. Es werden am Ende ja die 35 und nicht etwa nur die 19 Felder verwendet. Eine Differenzierung nach 35 Feldern ist einer Klassifikation mit 19 Feldern an dieser Stelle vorzuziehen.

Das so entstandene Gewichtungsschema wurde den Mikrodaten zugespielt. Dann wurden die Patentdaten je Wirtschaftszweig einmal gewichtet und einmal ungewichtet berechnet, d.h. jedes Patent innerhalb eines Wirtschaftszweiges, unabhängig vom Technologiefeld, wurde gleich gewichtet. Die Gewichte sind in der folgenden

---

<sup>2</sup> Da diese intellektuelle Konkordanz auf der Wirtschaftszweigklassifikation von 2003 bzw. 1992 beruht, müssen die Daten des Jahres 2007 verwendet werden, die das letzte verfügbare Jahr nach dieser alten Klassifikation darstellen.

Tabelle 6 abgetragen. Die Verteilung der FuE-Aufwendungen nach der ungewichteten Liste wurde lediglich zu Vergleichszwecken berechnet. Es wird empfohlen, lediglich die gewichtete Matrix zur Umrechnung von Aufwendungen für Forschung und Entwicklung nach Wirtschaftszweigen in Technologiefelder zu verwenden.

Abbildung 8: Vorgehensweise bei der Berechnung der FuE-Daten nach Technologiefeldern in der gewichteten und der ungewichteten Variante

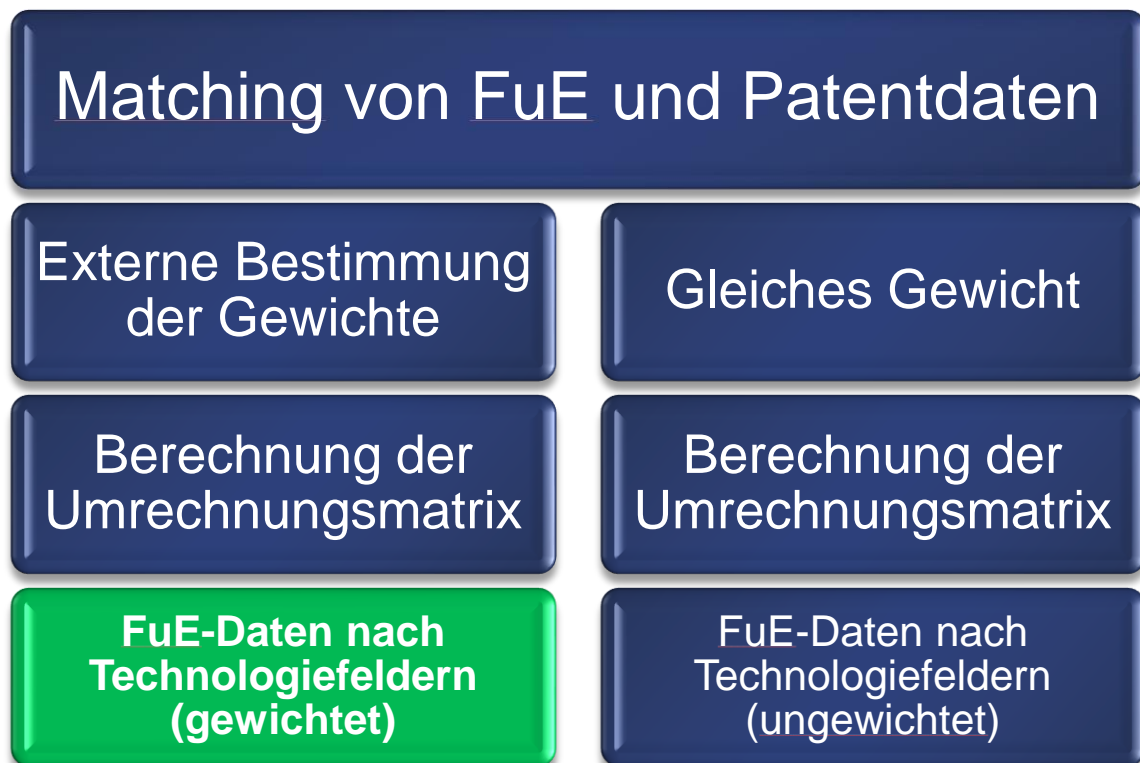


Tabelle 6: Gewichte der Patente in den Technologiefeldern – Indexiert auf das Gewicht im Feld „Transport“ (=100)

<b>Nr.</b>	<b>Bezeichnung</b>	<b>Gewicht Transport =100</b>	<b>Gleichgewichtung</b>
1	Elektrische Maschinen und Anlagen	19.6	1
2	Audiovisuelle Technik	21.5	1
3	Grundlegende Kommunikationstech.	181.5	1
4	Digitale Kommunikationstechnik	181.5	1
5	Telekommunikationstechnik	57.6	1
6	Computertechnik	47.2	1
7	Datenverarbeitung	47.2	1
8	Halbleiter	47.2	1
9	Optik	25.2	1
10	Messtechnik	38.5	1
11	Analyse von biologischen Materialien	38.5	1
12	Steuer- und Regeltechnik	26.7	1
13	Medizintechnik	29.5	1
14	Organische Feinchemie	57.8	1
15	Biotechnologie	116.1	1
16	Pharmazie	261.3	1
17	Kunststoffe, makromolekulare Chemie	57.8	1
18	Nahrungsmittelchemie	57.8	1
19	Grundstoffchemie	57.8	1
20	Materialien, Metallurgie	82.9	1
21	Oberflächen, Beschichtungen	57.8	1
22	Mikro- und Nanonstrukturen	289.0	1
23	Chemische Verfahrenstechnik	57.8	1
24	Umwelttechnik	8.1	1
25	Fördertechnik	8.1	1
26	Werkzeugmaschinen	8.1	1
27	Motoren, Pumpen, Turbinen	20.8	1
28	Textil- und Papiermaschinen	51.7	1
29	Andere Spezialmaschinen	8.8	1
30	Thermische Prozesse und Apparate	8.8	1
31	Maschinenelemente	8.8	1
32	Transport	100.0	1
33	Möbel, Spielzeug	15.2	1
34	Andere Konsumgüter	3.8	1
35	Bauwesen	8.8	1

Quelle: eigene Berechnungen auf Basis von EPA-PATSTAT und SV Wissenschaftsstatistik.

Auf diese Weise wurden zwei Matrizen mit der Anzahl der Patentanmeldungen je Technologiefeld innerhalb von Wirtschaftszweigen berechnet (siehe Tabelle 9 und Tabelle 11). Nun wurde für jeden Wirtschaftszweig der Anteil eines Technologiefeldes an allen Patenten dieses Wirtschaftszweiges berechnet – einmal in der gewichteten Matrix und einmal in der ungewichteten Matrix. Hierzu wurden die Patentzahlen der Jahre 2007-2009 verwendet, um stabile und aussagekräftige Relationen innerhalb der Wirtschaftszweige zu erhalten und den Einfluss kleiner Zahlen möglichst gering zu halten.<sup>3</sup> Abschließend wurden die aggregierten FuE-Aufwendungen je Wirtschaftszweig mithilfe dieser Anteile auf die Technologiefelder verteilt und schließlich wurden die Summen dieser FuE Aufwendungen je Technologiefeld gebildet. Aus diesen Arbeitsschritten resultiert eine Liste mit Ausgaben für Forschung- und Entwicklung je Technologiefeld – einmal in der gewichteten und einmal in der ungewichteten Variante. Die Ergebnisse für die internen FuE-Aufwendungen der Wirtschaft des Jahres 2009 sind in Tabelle 7 dargestellt.

Demnach werden in Deutschland für Transport-Technologien – diese umfassen sowohl KFZ, Schienenfahrzeuge wie auch Schiff und Flugzeugbau – die meisten FuE-Mittel verausgabt, nämlich rund 12,7 Mrd. Euro im Jahr 2009. An zweiter Stelle rangieren Pharma-Technologien mit knapp 4 Mrd. Euro interner FuE-Aufwendungen. Da im Bereich Biotechnologie auch die rote Biotechnologie enthalten ist (d.h. Bio-Pharma), sind die gesamten Aufwendungen für Pharma-Forschung in Deutschland sogar etwas höher als die genannten 4 Mrd. Euro. Die nächsten größeren Gruppen sind Grundlegende und Digitale Kommunikationstechnologien und die Messtechnik, für die jeweils rund 2,3 Mrd. Euro in 2009 aufgewendet wurden. Das kleinste der hier abgegrenzten Technologiefelder ist demnach „Analyse von biologischen Materialien“, wo rund 130 Mio. Euro interne FuE-Aufwendungen verausgabt wurden.

In Tabelle 8 sind die gesamten – internen und externen– FuE-Ausgaben der Unternehmen nach Technologiefeldern dargestellt. Demnach wendete die Wirtschaft für Transport-Technologien sogar rund 16,6 Mrd. Euro auf, gefolgt von der Pharmazie, wo rund 4,9 Mrd. Euro verausgabt wurden. In den Feldern Grundlegende und Digitale Kommunikation sowie der Messtechnik wurden demnach zwischen 2,7 bis 3 Mrd. Euro für Forschung und Entwicklung ausgegeben.

---

<sup>3</sup> Dies wurde nur zur Berechnung der Matrix auf diese Weise durchgeführt, um die Anteile innerhalb von Wirtschaftszweigen berechnen zu können. Für die Analysen auf der Mikro-Ebene, die im weiteren Verlauf durchgeführt wurden, wurden jeweils die FuE- und Patentdaten des gleichen Jahres verwendet.

Tabelle 7: Interne FuE-Aufwendungen der Wirtschaft im Jahr 2009 nach Technologiefeldern

<b>Nr.</b>	<b>Bezeichnung</b>	<b>Gewichtet</b>	<b>Ungewichtet</b>
1	Elektrische Maschinen und Anlagen	1.504.734	3.238.804
2	Audiovisuelle Technik	389.583	838.480
3	Grundlegende Kommunikationstech.	2.202.284	576.986
4	Digitale Kommunikationstechnik	2.386.332	673.374
5	Telekommunikationstechnik	309.352	227.514
6	Computertechnik	1.420.094	1.515.981
7	Datenverarbeitung	371.894	433.450
8	Halbleiter	975.888	1.014.135
9	Optik	343.222	693.893
10	Messtechnik	2.298.411	2.706.702
11	Analyse von biologischen Materialien	133.176	200.725
12	Steuer- und Regeltechnik	624.593	1.034.163
13	Medizintechnik	843.355	1.327.980
14	Organische Feinchemie	1.430.280	1.866.518
15	Biotechnologie	928.879	516.975
16	Pharmazie	3.955.687	1.245.200
17	Kunststoffe, makromolekulare Chemie	573.001	609.367
18	Nahrungsmittelchemie	221.743	209.478
19	Grundstoffchemie	1.323.048	1.523.131
20	Materialien, Metallurgie	1.057.051	603.717
21	Oberflächen, Beschichtungen	854.475	654.541
22	Mikro- und Nanonstrukturen	993.816	168.062
23	Chemische Verfahrenstechnik	1.539.854	1.123.757
24	Umwelttechnik	141.405	799.057
25	Fördertechnik	282.555	1.309.578
26	Werkzeugmaschinen	305.521	1.543.186
27	Motoren, Pumpen, Turbinen	1.563.893	3.429.603
28	Textil- und Papiermaschinen	1.654.091	1.055.627
29	Andere Spezialmaschinen	298.119	1.469.389
30	Thermische Prozesse und Apparate	212.376	943.633
31	Maschinenelemente	632.980	2.850.780
32	Transport	12.762.292	6.020.203
33	Möbel, Spielzeug	249.959	557.044
34	Andere Konsumgüter	140.165	902.321
35	Bauwesen	323.077	1.036.951

Quelle: eigene Berechnungen auf Basis von EPA-PATSTAT und SV Wissenschaftsstatistik.

Tabelle 8: Gesamte FuE-Aufwendungen der Wirtschaft im Jahr 2009 nach Technologiefeldern

<b>Nr.</b>	<b>Bezeichnung</b>	<b>Gewichtet</b>	<b>Ungewichtet</b>
1	Elektrische Maschinen und Anlagen	1.851.049	4.109.213
2	Audiovisuelle Technik	482.407	1.042.977
3	Grundlegende Kommunikationstech.	2.730.352	716.179
4	Digitale Kommunikationstechnik	2.963.248	832.286
5	Telekommunikationstechnik	380.063	280.259
6	Computertechnik	1.697.118	1.804.062
7	Datenverarbeitung	435.133	500.724
8	Halbleiter	1.220.529	1.276.022
9	Optik	413.107	866.413
10	Messtechnik	2.880.518	3.445.193
11	Analyse von biologischen Materialien	166.395	254.055
12	Steuer- und Regeltechnik	770.377	1.303.496
13	Medizintechnik	1.001.225	1.632.027
14	Organische Feinchemie	1.713.663	2.276.137
15	Biotechnologie	1.144.435	644.455
16	Pharmazie	4.933.701	1.575.775
17	Kunststoffe, makromolekulare Chemie	660.088	707.858
18	Nahrungsmittelchemie	248.097	236.662
19	Grundstoffchemie	1.554.572	1.816.982
20	Materialien, Metallurgie	1.273.309	731.799
21	Oberflächen, Beschichtungen	1.028.208	801.933
22	Mikro- und Nanonstrukturen	1.316.374	222.986
23	Chemische Verfahrenstechnik	1.727.901	1.331.950
24	Umwelttechnik	175.333	1.016.653
25	Fördertechnik	321.576	1.512.352
26	Werkzeugmaschinen	372.674	1.925.051
27	Motoren, Pumpen, Turbinen	1.988.154	4.489.604
28	Textil- und Papiermaschinen	1.839.802	1.186.204
29	Andere Spezialmaschinen	352.546	1.769.254
30	Thermische Prozesse und Apparate	266.086	1.192.356
31	Maschinenelemente	747.133	3.606.001
32	Transport	16.660.697	7.961.966
33	Möbel, Spielzeug	269.898	657.272
34	Andere Konsumgüter	164.747	1.082.777
35	Bauwesen	316.611	1.258.195

Quelle: eigene Berechnungen auf Basis von EPA-PATSTAT und SV Wissenschaftsstatistik.

## 6.2 Plausibilisierung der Ergebnisse

Um die Ergebnisse der Umrechnung zu plausibilisieren, wurden externe Informationen über FuE-Aufwendungen in Technologiefeldern herangezogen. Da jedoch kaum Informationen nach Technologiefeldern vorliegen bzw. nur für wenige einzelne Technologiefelder, meist auf Einzeluntersuchungen oder Verbandsangaben beruhend, konnte nur an wenigen Stellen eine solche Plausibilisierung vorgenommen werden. Im Wesentlichen konnten drei Quellen identifiziert werden. Ein Bericht (BMBF 2011) über die Nanotechnologien in Deutschland enthält Angaben über Forschungsaufwendungen von befragten Unternehmen, die dann auch auf die Grundgesamtheit aller in Deutschland forschenden Unternehmen im Bereich Nanotechnologien hochgerechnet wurden. Es werden Forschungsaufwendungen von 1,3 Mrd. € im Jahr 2009 berichtet. Die gesamten von uns errechneten FuE-Aufwendungen der Wirtschaft für „Mikro- und Nanostrukturen“ im Jahr 2009 belaufen sich auf 1,316 Mrd. Euro.

Eine zweite Quelle waren Verbandsangaben, die jedoch häufig lediglich auf Branchenebene und nicht auf Technologieebene vorliegen. Technologiespezifische Angaben konnten für die Biotechnologie und für die Pharmazie in Deutschland gefunden werden. Eine Untersuchung von Ernst & Young (Ernst & Young 2013) weist für die Biotechnologie im Jahr 2009 Forschungsaufwendungen von etwa 780 Millionen € aus. Auf eine ähnliche Größenordnung kommt auch bionity.de, ein Fachportal für Biotechnologie<sup>4</sup>. Bei biotechnologie.de<sup>5</sup> wird eine Summe von 1,05 Milliarden € genannt, die dezidierte Biotechnologie-Unternehmen für Forschung aufwenden. All diesen Angaben ist gemein, dass sie ebenfalls Angaben von wenigen Unternehmen auf die Grundgesamtheit hochrechnen. Im Bereich der Pharmazie werden von verschiedenen Branchenverbänden Summen zwischen 5,1 und 5,4 Milliarden € im Jahr 2009 genannt, wobei es den Anschein hat, dass hier die Gesamtsumme der Forschungsaufwendungen aller in den jeweiligen Verbänden organisierten Unternehmen gebildet wurde und nicht etwa dezidiert nach Pharma-Forschung unterschieden wurde. Die von uns errechneten Ausgaben der Wirtschaft für Forschung und Entwicklung im Bereich „Pharmazie“ belaufen sich auf 4,9 Mrd. Euro. In dem Technologiefeld „Biotechnologie“, wofür 1,1 Mrd. aufgewendet wurden, sind weitere Aufwendungen für „Bio-Pharma“ enthalten, so dass nach unseren Berechnungen faktisch für die Pharmaforschung in Deutschland ein Betrag deutlich oberhalb der 5 Mrd. ausgegeben wurde.

---

4 <http://www.bionity.com/de/news/137508/deutsche-biotech-branche-auf-stabilem-wachstumskurs.html>.

5 <http://www.biotechnologie.de>.

Eine dritte Quelle, die unmittelbar die Relation von Patenten und Forschungsaufwendungen herstellt, ist die jährlich durchgeführte Anmelder-Befragung des europäischen Patentamts (European Patent Office (EPO) 2010). In dieser Befragung wird eine Stichprobe von praktizierenden Unternehmen zu ihrem gegenwärtigen und geplanten Patentverhalten, aber auch zu weiteren Informationen untersucht. Unter anderem wird auch nach den Forschungsaufwendungen innerhalb von 14 Technologie-Clustern des europäischen Patentamtes gefragt. Es liegen zwar keine Angaben zu den Wirtschaftszweigen vor, d.h. es werden auch keine Hochrechnungen auf die Grundgesamtheit innerhalb einer Branche oder eines Technologiefeldes durchgeführt. Vielmehr werden die Patentintensitäten, das heißt die Relation von Forschungsaufwendungen zur Anzahl der Patente innerhalb jedes Clusters (Technologiefeldes) ausgewiesen. In Tabelle 5 auf Seite 30 sind die FuE-Intensitäten unserer 35 Technologiefelder aufgeführt. Um eine Vergleichbarkeit der Intensitäten unserer 35 Technologiefelder und der 14 Technologie-Cluster des europäischen Patentamtes zu erreichen, wurden technologisch ähnlich gelagerte Klassen zu den 14 Clustern zusammen gefasst. Dies bringt einige Unschärfen mit sich, zumal das Europäische Patentamt die genaue Abgrenzung seiner 14 Cluster nicht veröffentlicht hat. Daneben beruhen die Berechnungen auf der Befragung, die auf Grund von fehlenden Angaben gerade bei dieser Frage, in manchen der 14 Clustern auf sehr niedrigen Fallzahlen basieren und werden daher nur in aggregierter Form ausgewiesen. Es wurden uns zwar die nicht-aggregierten Werte zu Vergleichszwecken zur Verfügung gestellt, jedoch mit dem deutlichen Hinweis, dass die statistische Belastbarkeit der Ergebnisse nicht sehr hoch ist.

Vergleichen wir nun unsere Ergebnisse der internen FuE-Aufwendungen der Wirtschaft mit den Ergebnissen des Europäischen Patentamts für das Jahr 2009, dann ergeben sich in den Clustern bezüglich Chemie oder Biotechnologie sowie Polymere ähnliche Relationen, ebenso verhält es sich im Bereich Elektronik. Deutliche Abweichungen gibt es bei „Handling“, „Human Necessities“ und „Civil Engineering“. Im Bereich von Fahrzeugen errechnen wir deutlich höhere FuE-Intensitäten. Allerdings ist die deutsche Automobilindustrie auch dafür bekannt, besonders forschungsintensiv zu sein. Da wir in unseren Analysen ausschließlich die deutschen Unternehmen und das EPA die weltweiten Unternehmen betrachten, lässt sich die Abweichung zumindest zum Teil hierdurch erklären.

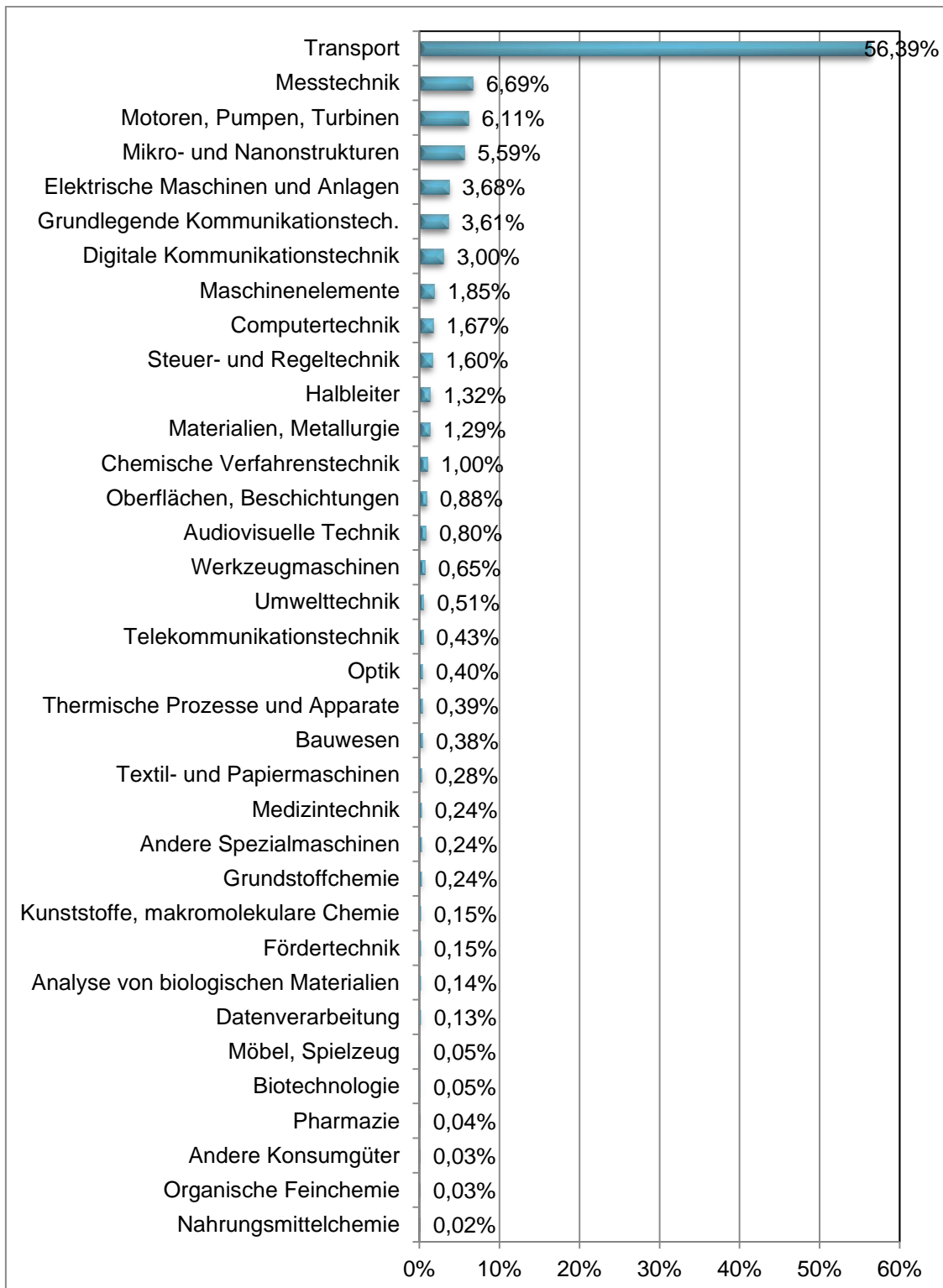
Insgesamt erlauben die zur Plausibilisierung herangezogenen Quellen keine sichere Überprüfung unserer Ergebnisse. Dies ist jedoch auch allein deshalb nicht möglich, da das Fehlen der Daten nach Technologiefeldern den Anlaß dieses Projektes bildete. Die verfügbaren Werte widersprechen jedoch aus unserer Sicht zumindest unseren Ergebnissen nicht.

### **6.3 Ein Beispiel für die Möglichkeiten, die sich aus der Umrechnung ergeben**

Im Folgenden wird ein Beispiel für die Nutzung der Umrechnung und der Daten am Beispiel des Fahrzeugbaus bzw. des Technologiefeldes „Transport“ gegeben. In Abbildung 9 sind die gesamten FuE-Aufwendungen der Branche 29 „Herstellung von Kraftwagen und Kraftwagenteilen“ nach Technologiefeldern abgetragen. Demnach gibt der Fahrzeugbau in Deutschland von den insgesamt rund 18,8 Mrd. Euro gut 56% für die Forschung und Entwicklung von Technologien im Bereich „Transport“ aus. Weitere Teile von jeweils um die 6% werden für „Messtechnik“, „Motoren, Pumpen und Turbinen“, sowie für „Mikro- und Nanostrukturen“ ausgegeben. Auf „Elektrische Maschinen und Anlagen“ und „Kommunikationstechnologien“ (grundlegende und digitale) entfallen jeweils 3-3,7%. Auch für „Computertechnik“, „Halbleiter“ oder „Steuer- und Regelungstechnik“ werden aufgewendet, die zwischen 1,3 und 1,7% der 18,8 Mrd. Euro umfassen. Insgesamt zeigt sich der deutsche Fahrzeugbau in seinen technologieorientierten Aktivitäten sehr breit aufgestellt.

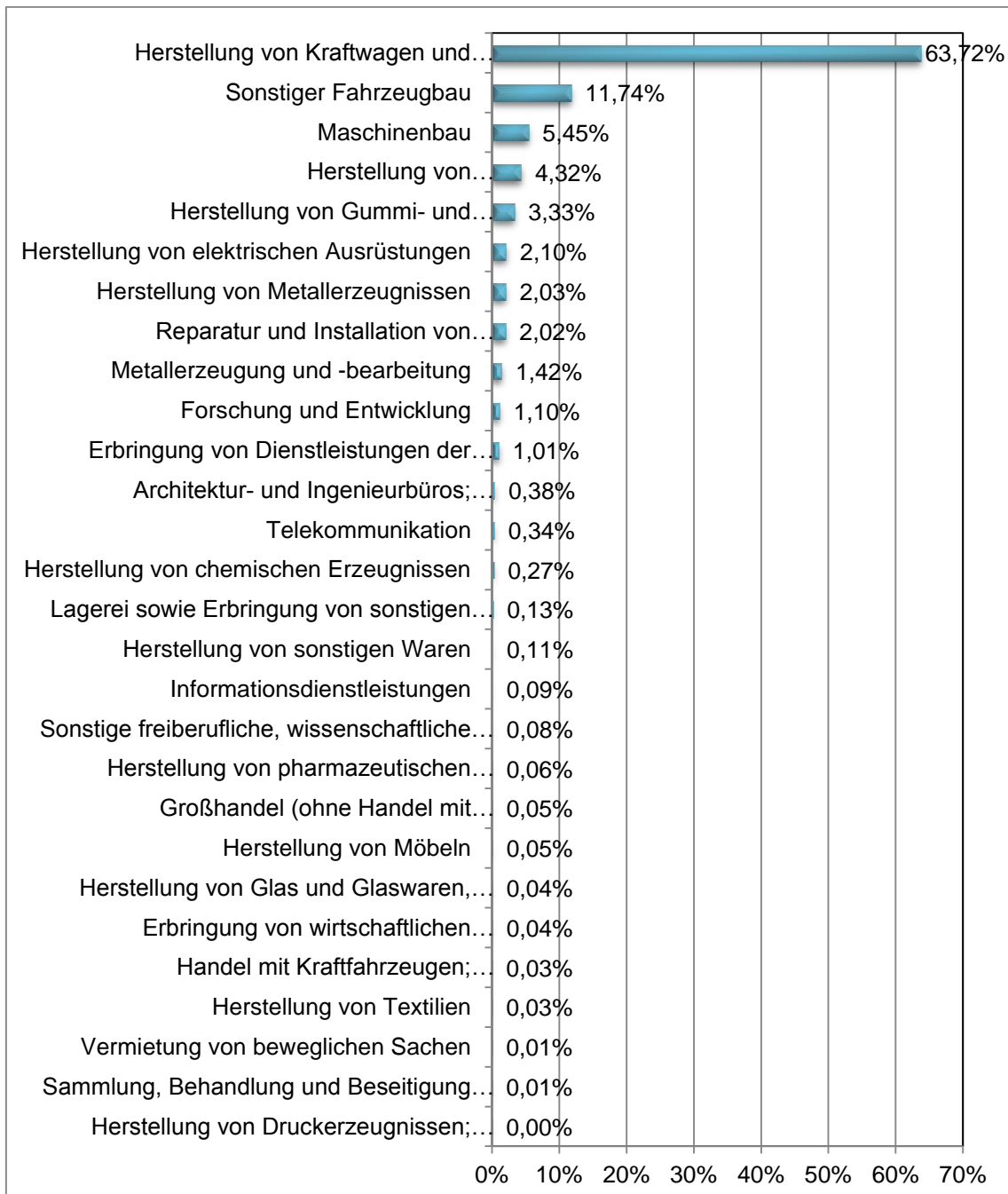
Abbildung 10 zeigt umgekehrt, in welchen Wirtschaftszweigen die Technologien aus dem Bereich „Transport“ beforscht und entwickelt werden. Demnach stammen knapp 64% der Aufwendungen aus der KFZ-Branche und weitere 11,7% aus dem „sonstigen Fahrzeugbau“. Aber auch der „Maschinenbau“, die „Herstellung von Datenverarbeitungsgeräten, elektronischen und optischen Erzeugnissen“ sowie die „Herstellung von Gummi- und Kunststoffwaren“ leisten substantielle Beiträge zu Forschung und Entwicklung rund um Transport-Technologien.

Abbildung 9: Gesamte FuE-Ausgaben der Branche 29 „Herstellung von Kraftwagen und Kraftwagenanteilen“ nach Technologiefeldern



Quelle: eigene Berechnungen auf Basis von EPA-PATSTAT und SV Wissenschaftsstatistik.

Abbildung 10: Gesamte FuE-Ausgaben im Technologiefeld „Transport“ nach Wirtschaftszweigen



Quelle: eigene Berechnungen auf Basis von EPA-PATSTAT und SV Wissenschaftsstatistik.

Tabelle 9: Matrix mit Technologieanteilen nach Wirtschaftszweigen 2008 und Technologiefeldern (basierend auf der gewichteten Verteilung der Patente innerhalb der Wirtschaftszweige)

WZ	Bezeichnung	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16	f17	f18	f19	f20	f21	f22	f23	f24	f25	f26	f27	f28	f29	f30	f31	f32	f33	f34	f35		
01	Landwirtschaft, Jagd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	93	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0		
08	G.v. Steinen etc.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	77	0	0	9	4	0	0	0	0	0	0	0	0	0	0	0	1	
10	H.v. Nahrungs- u. Futterm.	0	0	0	0	0	1	0	0	0	1	1	1	0	7	13	21	2	36	5	2	0	0	4	0	1	0	0	3	2	0	0	0	0	0	0		
11	Getränkeherstellung	0	0	0	0	0	0	0	0	0	37	0	0	0	0	0	0	0	0	0	0	0	0	55	0	8	0	0	0	0	0	0	0	0	0	0		
12	Tabakverarbeitung	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	69	0	0	0	0	0	0	20	0	
13	H.v. Textilien	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	4	0	8	0	0	0	0	0	0	57	1	1	1	7	6	2	4		
14	H.v. Bekleidung	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0		
15	H.v. Leder, Lederwaren	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0		
16	H.v. Holzwaren etc.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	10	14	39	0	0	0	0	7	0	0	1	0	0	0	0	0	1	18		
17	H.v. Papier, Pappe etc.	0	4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	5	4	12	0	7	0	12	0	1	36	1	0	0	0	0	14	1	1		
18	H.v. Druckerzeugnissen etc.	0	2	13	34	0	23	2	0	1	1	0	8	0	0	0	0	1	0	1	0	2	2	0	0	0	0	0	7	0	0	0	0	0	0	2	0	
19	Kokerei und Mineralölverarb.	0	5	44	44	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
20	H.v. chem. Erzeugnissen	0	0	0	0	0	0	0	1	0	0	0	0	1	21	5	21	12	1	18	4	3	1	6	0	0	0	0	1	1	0	0	1	0	0	0		
21	H.v. pharm. Erzeugnissen	0	0	0	0	0	0	0	3	1	1	1	0	2	14	5	60	1	0	8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	H.v. Gummiwaren etc.	2	0	0	0	0	0	0	2	0	2	0	0	1	0	0	1	4	0	2	1	4	0	2	0	1	0	1	3	2	1	5	63	1	0	1		
23	H.v. Glas etc.	3	1	1	0	0	0	9	4	3	0	0	4	0	1	10	1	0	2	37	8	2	4	0	0	0	1	0	3	1	0	2	1	0	2	1	0	1
24	Metallerzeugung u. -bearb.	4	1	0	0	0	0	3	0	2	0	0	1	1	1	0	0	0	1	25	5	0	2	1	0	3	3	0	1	0	2	42	0	0	1			
25	H.v. Metallerzeugnissen	2	0	0	0	0	0	1	0	2	0	1	1	0	0	1	1	0	2	5	3	1	3	0	1	3	1	9	1	1	2	42	5	0	12			
26	H.v. DVgeräten etc.	6	2	15	12	2	7	1	6	2	10	0	2	6	0	1	1	0	0	1	2	1	2	2	0	0	0	2	0	0	1	0	11	1	0	0		
27	H.v. elektr. Ausrüstungen	17	2	11	10	3	3	0	1	1	9	0	4	0	0	0	0	1	0	0	0	1	0	2	0	0	0	1	1	0	2	1	24	4	1	1		
28	Maschinenbau	2	0	2	1	0	1	0	1	0	5	0	1	1	0	1	0	0	1	1	2	5	0	11	0	4	2	4	29	2	0	3	18	0	0	1		
29	H.v. Kraftwagen -teilen	4	1	4	3	0	2	0	1	0	7	0	2	0	0	0	0	0	0	0	1	1	6	1	1	0	1	6	0	0	0	2	56	0	0	0		
30	Sonstiger Fahrzeugbau	2	1	5	4	0	1	0	0	0	4	0	1	0	0	0	0	0	0	1	2	5	0	1	0	0	1	4	1	1	0	1	65	0	0	0		
31	H.v. Möbeln	3	7	14	14	0	7	4	0	0	0	0	4	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	15	21	0	1		
32	H.v. sonstigen Waren	1	1	2	0	0	4	1	1	6	7	1	3	40	0	2	8	2	0	2	1	3	0	3	0	1	1	1	2	1	0	0	4	2	1	0		
33	Rep. u. Inst. v. Masch. etc.	4	1	7	0	1	0	0	7	0	5	0	1	0	0	0	0	2	0	0	12	2	0	1	0	1	1	1	1	3	0	0	49	0	0	0		



Tabelle 10: Wirtschaftszweige ohne Patentanmeldungen

<b>WZ</b>	<b>Bezeichnung</b>
02	Forstwirtsch. U. Holzeinsch.
05	Kohlenbergbau
06	G.v. Erdöl und Erdgas
07	Erzbergbau
09	DL für Bergbau etc.
36	Wasserversorgung
37	Abwasserentsorgung
49	Landverkehr und Transport
50	Schifffahrt
51	Luftfahrt
56	Gastronomie
60	Rundfunkveranstalter
65	Versicherungen
66	Finanz- und Vers. verb. Tät.
68	Wohnungswesen
75	Veterinärwesen
78	Überlassung v. Arbeitskr.
79	Reisebüros
81	Gebäudebetreuung
85	Erziehung und Unterricht
90	Kreative Tätigkeiten
93	Sports, Unterh., Erholung
94	Interessenvertretungen
95	Reparatur von DL-geräten

Tabelle 11: Matrix mit Technologieanteilen nach Wirtschaftszweigen 2008 und Technologiefeldern (basierend auf der ungewichteten Verteilung der Patente innerhalb der Wirtschaftszweige)

WZ	Bezeichnung	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16	f17	f18	f19	f20	f21	f22	f23	f24	f25	f26	f27	f28	f29	f30	f31	f32	f33	f34	f35		
01	Landwirtschaft, Jagd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
08	G.v. Steinen etc.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	50	0	0	8	25	0	0	0	0	0	0	0	0	0	0	0	0	8
10	H.v. Nahrungs- u. Futterm.	0	1	0	0	0	1	0	0	0	1	1	2	0	7	6	5	2	35	5	2	0	0	4	2	9	0	1	3	13	1	0	0	0	2	0	0	
11	Getränkeherstellung	0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	33	0	33	0	0	0	0	0	0	0	0	0	0	0	
12	Tabakverarbeitung	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	17	0	0	0	0	0	0	67	0	
13	H.v. Textilien	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	2	0	4	0	0	0	0	0	0	35	2	2	4	2	13	13	13		
14	H.v. Bekleidung	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	
15	H.v. Leder, Lederwaren	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
16	H.v. Holzwaren etc.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	4	4	15	0	0	0	0	19	0	0	4	0	0	0	0	0	7	44
17	H.v. Papier, Pappe etc.	0	4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	1	5	0	3	1	33	1	1	16	3	0	1	0	20	6	3		
18	H.v. Druckerzeugnissen etc.	1	4	4	10	0	25	2	1	2	1	0	15	0	0	0	0	1	0	1	0	1	0	0	0	2	0	0	7	2	0	0	0	0	0	21	0	
19	Kokerei und Mineralölverarb.	0	29	29	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
20	H.v. chem. Erzeugnissen	1	1	0	0	0	0	1	1	0	0	0	2	23	3	5	13	2	19	3	3	0	7	1	1	1	0	2	6	0	1	1	0	1	0	1	1	
21	H.v. pharm. Erzeugnissen	1	1	0	0	0	0	6	5	1	2	0	7	24	5	23	2	1	13	0	0	0	2	0	2	0	0	1	1	0	1	0	0	0	0	0	0	
22	H.v. Gummiwaren etc.	3	1	0	0	0	0	2	0	2	0	0	1	0	0	0	2	0	1	0	3	0	2	2	4	1	3	3	8	2	22	25	3	3	7			
23	H.v. Glas etc.	7	1	0	0	0	0	8	7	3	0	0	6	0	0	2	1	0	1	19	6	0	3	1	2	2	1	0	15	5	1	1	2	0	4			
24	Metallerzeugung u. -bearb.	8	2	0	0	0	0	3	0	3	0	0	2	1	1	0	0	0	1	12	4	0	2	3	1	17	6	0	2	2	10	18	0	1	3			
25	H.v. Metallerzeugnissen	2	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	1	2	1	0	2	1	4	8	1	4	3	3	7	11	9	2	34			
26	H.v. DVgeräten etc.	14	4	4	3	2	6	1	5	4	11	1	4	9	0	0	0	0	1	1	1	0	2	1	1	1	1	4	0	2	3	2	5	3	4	1		
27	H.v. elektr. Ausrüstungen	29	3	2	2	2	0	1	1	8	0	4	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	0	1	6	3	8	8	10	3		
28	Maschinenbau	3	1	0	0	1	0	1	0	4	0	2	1	0	0	0	0	0	1	0	1	3	0	6	1	13	9	6	17	8	2	11	6	0	2	3		
29	H.v. Kraftwagen -teilen	9	2	1	1	0	2	0	1	1	9	0	3	0	0	0	0	0	0	0	1	1	1	1	3	1	4	14	0	1	2	10	28	0	0	2		
30	Sonstiger Fahrzeugbau	5	2	1	1	0	1	1	0	1	6	0	2	0	0	0	0	0	1	2	5	0	1	1	0	7	10	1	6	1	5	37	1	1	1			
31	H.v. Möbeln	5	10	3	3	0	5	3	0	0	0	5	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	5	46	3	3		
32	H.v. sonstigen Waren	2	1	0	0	0	3	0	1	8	6	1	3	43	0	1	1	1	0	1	0	1	0	1	0	2	3	1	1	3	0	1	1	3	7	0		
33	Rep. u. Inst. v. Masch. etc.	9	2	2	0	1	0	0	7	1	7	0	1	0	0	0	0	2	0	0	7	2	0	1	1	8	5	3	1	14	1	2	23	0	0	2		

35	Energieversorgung	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	8	0	0	23	31	0	0	0	0	8		
38	Abfälle; Rückgewinnung	35	0	0	0	0	0	0	0	4	4	0	8	0	19	8	0	0	4	0	0	0	0	0	4	4	0	0	0	0	8	0	0	4		
39	Beseit. v. Umweltverschm.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0		
41	Hochbau	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	33	0	67			
42	Tiefbau	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	5	0	0	0	9	5	0	0	0	0	68			
43	Vorber. Baustellenarbeiten	2	0	0	0	0	2	0	0	9	0	0	0	0	0	0	0	5	2	2	0	0	2	5	0	2	0	2	25	5	0	0	36			
45	Handel mit KFZ; Instandh.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	13	0	0	13	13	25	13	13	0			
46	Großhandel (ohne KFZ)	2	7	5	20	2	9	1	0	3	6	0	2	3	4	0	1	0	0	0	1	1	0	2	0	2	1	0	1	5	1	3	3	4	2	6
47	Einzelhandel (ohne KFZ)	0	0	0	33	0	0	0	0	0	0	0	0	67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
52	Lagerei u. DL für Verkehr	14	0	0	0	0	0	29	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	14	14		
58	Verlagswesen	0	0	0	13	0	13	0	0	0	25	0	0	0	0	0	0	0	0	13	0	0	0	38	0	0	0	0	0	0	0	0	0	0	0	
61	Telekommunikation	15	2	4	7	1	7	2	3	2	8	1	4	8	0	0	0	0	1	1	1	0	2	1	1	1	6	0	0	6	2	2	4	7	0	
62	DL der IT	2	2	3	9	1	35	17	0	1	5	0	5	0	0	0	0	0	1	0	0	0	2	0	6	1	1	0	0	0	1	5	1	1	0	
63	Informationsdienstleistungen	4	4	7	22	0	15	4	0	0	11	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	22	0	0	0	
64	Finanzdienstleistungen	0	0	0	0	0	0	0	0	0	50	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
70	Verwaltung von Unt.	29	7	0	0	7	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	7	0	0	14	0	0	0	0	0	
71	Ingenieurbüros	4	1	1	1	0	2	0	1	1	5	0	1	1	10	2	5	0	0	18	4	2	0	7	4	3	3	5	1	3	2	5	3	0	0	2
72	Forschung und Entwicklung	3	0	1	0	0	1	0	1	1	4	2	1	4	3	6	6	0	0	1	2	2	0	5	4	0	2	30	0	2	2	10	5	1	1	1
73	Werbung und Marktforsch.	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
74	Sonst. wiss. u. tech. Tätigk.	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	25	25	0	0	
77	Vermietung v. bew. Sachen	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	33	0	0	33	0	0	33		
82	Sonstige DI für Unt.	21	11	0	0	0	0	0	0	16	0	0	5	0	0	0	0	0	5	0	0	0	0	0	0	0	0	5	0	16	5	0	5	11		
86	Gesundheitswesen	0	0	0	0	0	10	0	0	0	0	0	60	0	20	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
96	Sonstige pers. DL	0	13	0	0	0	13	13	13	0	13	0	0	0	13	13	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	

## 7 Schlussbemerkungen

Die Aufwendungen für Forschung und Entwicklung der Unternehmen bilden einen der wichtigsten Indikatoren zur Bewertung der gegenwärtigen und zukünftigen Wettbewerbsfähigkeit von Innovationssystemen bzw. von ganzen Volkswirtschaften. Die Daten werden nach international vergleichbaren Maßstäben erhoben. In Deutschland ist für diese Erhebung die Wissenschaftsstatistik innerhalb des Stifterverbands zuständig, die alle zwei Jahre eine Vollerhebung und in den Zwischenjahren eine Kurzbefragung von FuE-treibenden Unternehmen und Einrichtungen durchführt und so Daten auf der Mikroebene von Unternehmen erfasst. Die Ergebnisse werden in aggregierter Form für Gesamtdeutschland bzw. nach Sektoren entlang der Wirtschaftszweigklassifikation (derzeit WZ 2008) veröffentlicht. Hierzu werden die Daten der Unternehmen entsprechend der jeweiligen Sektorzuordnung zusammengefasst. Um die Aufwendungen für Forschung und Entwicklung jedoch jenseits von Sektoren, beispielsweise nach Technologiefeldern oder einzelnen Technologien auszuweisen, müssten entweder in einer separaten Befragung die Ausgaben nach Technologiefeldern erfasst, oder– wie in dieser Studie geschehen– die bestehenden Daten anhand weiterer Informationen umgerechnet werden.

Es gab in der Vergangenheit bereits verschiedene Versuche die Aufwendungen für Forschung und Entwicklung nach Technologiefeldern bei den Unternehmen direkt zu erfragen. Gerade bei Großunternehmen, die für einen Großteil der Aufwendungen für Forschung und Entwicklung insgesamt verantwortlich sind, liegen diese Informationen jedoch nicht direkt abrufbar vor bzw. können nur schwer berechnet werden. Im Allgemeinen erfassen die Unternehmen ihre Aufwendungen nach Geschäftsfeldern (die nicht notwendigerweise Technologie-homogen sind), nach Projekten oder als Gesamtsumme, nicht jedoch nach Technologien oder Technologiefeldern. Hinzu kommt, dass einheitliche Definitionen von Technologien und Technologiefeldern insbesondere auf der operativen Ebene von Unternehmen nicht vergleichbar und einheitlich vorliegen. Eine klare Abgrenzung auf dieser Basis ist damit nicht möglich. Die verschiedenen Versuche, mit Hilfe von Befragungen die Aufwendungen für Forschung und Entwicklung in einzelnen Technologiefeldern zu erfassen, waren daher nur in Teilen erfolgreich. Eine breite, über alle Unternehmen und insbesondere über die Zeit vergleichbare Erfassung technologiebezogener FuE-Aufwendungen ist bis heute nicht vorhanden und scheint auch nicht möglich. Einzige regelmäßig durchgeführte Erhebung die unter anderem auch Fragen zu den Aufwendungen nach Technologiefeldern beinhaltet, ist die Anmelder-Befragung des europäischen Patentamts (EPO 2010). Allerdings haben diese Fragen auch regelmäßig die höchsten Antwortausfälle zu verzeichnen, was die Schwierigkeit für die Unternehmen belegt, diese Informationen bereitzustellen.

Es bleibt daher als vielversprechender Weg die Umrechnung der jeweiligen Gesamtaufwendungen der einzelnen Unternehmen in Technologiefelder anhand zusätzlicher Informationen. Die hiermit vorgelegte Untersuchung hat dazu einen Beitrag auf Basis der Technologiefelder der Patentanmeldungen von Unternehmen geleistet. Auf der Mikroebene konnte mit hoher Zuverlässigkeit und breiter Abdeckung die Zusammenführung von FuE-Daten und Patentdaten realisiert werden. Dabei können die jeweiligen Anteile der Unternehmen nicht eins zu eins anhand der Patentanteile auf die Technologiefelder verteilt werden, da Patente bzw. Technologien unterschiedlich hohe Forschungsaufwendungen erfordern. Mit anderen Worten, nicht jedes technologische Ergebnis ist mit dem gleichen Aufwand zu erzielen bzw. Technologien sind unterschiedlich forschungsintensiv. Daher musste ein Weg gefunden werden, diese Unterschiede der Forschungsintensität bzw. des notwendigen Aufwands bei der Berechnung der FuE-Aufwendungen nach Technologiefeldern zu berücksichtigen. Hierzu wurden in dieser Studie die Patentintensitäten nach Sektoren aus bereits vorliegenden Untersuchungen (Schmoch und Gauch 2004) verwendet und anhand der verfügbaren Informationen und Daten kalibriert.

Das in methodischer Hinsicht beschrittene Neuland wird in einem weiteren Schritt in den wissenschaftlichen Diskurs eingespeist. Eine weitere Möglichkeit die Ergebnisse vertieft zu validieren wäre, diese mit Hilfe partizipativer Methoden (z.B. Interviews) oder im Rahmen einer Unternehmens- bzw. Expertenbefragung zu überprüfen.

Für die Innovationsforschung ergeben sich Ansatzpunkte insbesondere im Bereich der Netzwerkforschung. Über die Aufschlüsselung von FuE-Aufwendungen einzelner Branchen nach Technologiefeldern wird erkennbar, welche FuE-Kooperationsbeziehungen dazu beitragen, dass sich technologische Schwerpunkte und Kompetenzfelder in der deutschen Wirtschaft entwickeln und festigen. Ebenso ergeben sich interessante Aspekte für die Innovationspolitik, die Schlüsse ziehen kann, welche Akteure, respektive Branchen, für die Entwicklung von Technologien relevant sind. Zu vertiefen wären hierfür insbesondere noch Untersuchungen auf der Mikroebene, um Unternehmen oder Unternehmenstypen zu identifizieren, die für den technologischen Fortschritt eine herausgehobene Relevanz besitzen.

## 8 Literatur

- Baeza-Yates, R.; Ribeiro-Neto, B. (2011): Modern Information Retrieval - The concepts and technology behind, Second Edition: Pearson Education Limited.
- BMBF (Hrsg.) (2011): nano.DE-Report 2011. Status quo der Nanotechnologie in Deutschland. Bonn, Berlin: BMBF.
- Deng, Z.; Lev, B.; Narin, F. (1999): Science and technology as predictors of stock performance. In: Financial Analysts Journal, 55 (3), S. 20-32.
- Du Plessis, M.; Van Looy, B.; Song, X.; Magerman, T. (2009): Data Production Methods for Harmonized Patent Indicators: Assignee sector allocation, EUROSTAT Working Paper and Studies Nr. Luxembourg.
- Ernst & Young (Hrsg.) (2013): Umdenken ... weiter denken, breiter denken. Deutscher Biotechnologie-Report 2013. Mannheim: E&Y.
- European Patent Office (EPO) (2010): Future Filings Survey 2009. Inventions of Applicants Regarding Patent Applications at the European Patent Office and Other Offices: European Patent Office.
- Freeman, C. (1982): The Economics of Industrial Innovation, London: Pinter Publishers.
- Frietsch, R.; Schmoch, U. (2010): Transnational Patents and International Markets. In: Scientometrics, 82 (1), S. 185-200.
- Grupp, H. (1998): Foundations of the Economics of Innovation - Theory, Measurement and Practice, Cheltenham: Edward Elger.
- Legler, H.; Frietsch, R. (2007): Neuabgrenzung der Wissenswirtschaft - forschungsintensive Industrien und wissensintensive Dienstleistungen (NIW/ISI-Listen 2006), Studien zum deutschen Innovationssystem Nr. 22-2007, Hannover, Karlsruhe: NIW, Fraunhofer ISI.
- Long, J.S. (1997): Regression Models for Categorical and Limited Dependent Variables, Thousand Oaks: Sage.
- Magerman T.; Grouwels J.; Song X.; Van Looy B. (2009): Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization, EUROSTAT Working Paper and Studies Nr. Luxembourg.
- Martinez, C. (2009): Insight into different types of patent families, Madrid: CSIC-Institute of Public Goods and Policies.
- Narin, F.; Noma, E.; Perry, R. (1987): Patents as indicators of corporate technological strength. In: Research Policy, 16 (2-4), S. 143-155.

- Peeters B.; Song X.; Callaert J.; Grouwels J.; Van Looy B. (2009): Harmonizing harmonized patentee names: an exploratory assessment of top patentees, EUROSTAT Working Paper and Studies Nr. Luxembourg.
- Putnam, J. (1996): The value of international patent rights, Yale: Yale University.
- Raffo, J.; Lhuillery, S. (2009): 'How to play the "Names Game": Patent retrieval comparing different heuristics. In: Research Policy, 38, S. 1617-1627.
- Schmoch, U.; Grupp, H.; Mannsbart, W.; Schwitalla, B. (1988): Technikprognosen mit Patentindikatoren, Köln: TÜV Rheinland.
- 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.
- Schmoch, U.; Laville, F.; Patel, P.; Frietsch, R. (2003): Linking Technology Areas to Industrial Sectors, Final Report to the European Commission, DG Research, Karlsruhe: Fraunhofer ISI.
- Trajtenberg, M. (1990): A penny for your quotes: patent citations and the value of innovation. In: Rand Journal of Economics, 21 (1), S. 172-187.