

Abstract

Diese Arbeit beschreibt den Prozess der automatischen Indexierung wissenschaftlich-technischer Kongressschriften aus dem Bereich der Luft- und Raumfahrt. Die Indexierung erfolgt mit dem Open Source Programm Lingo. Lingo ist ein linguistisch basiertes Programm zur Indexierung deutschsprachiger Textdokumente und enthält Wörterbücher, die die allgemeine deutsche Sprache abdecken. Zur erfolgreichen Indexierung wissenschaftlich-technischer Textdokumente ist es notwendig, die vorhandenen Wörterbücher, um das verwendete Fachvokabular zu erweitern. Der Erweiterungsprozess der Wörterbücher wird detailliert beschrieben und umfasst u.a. die Worterkennung, Grundformenerzeugung, das Erkennen und Zerlegen von Wortzusammenfügungen, die Mehrworterkennung und die Synonymzuordnung. Schwerpunkte sind die gezielte Lexikalisierung von Wortstämmen zur Kompositaerkennung und die Behandlung von Binde- und Ergänzungsstrichen.

Schlagworte:

automatische Indexierung, linguistisches Verfahren, wissenschaftlich-technische Information, Lingo

Hiermit versichere ich, die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben.

Ruppicheroth, den 22.06.2007

Automatisches Indexieren technischer Kongressschriften

Diplomarbeit

Studiengang: Informationswirtschaft

Fakultät für Informations- und Kommunikationswissenschaften
Fachhochschule Köln

vorgelegt von:

Ralf Schiffer
Schreckenberger Str. 18a
53809 Ruppichteroth
Matr.Nr.: 11040922

am 22.06.2007 bei Herrn Prof. Dr. Lepsky

Inhaltsverzeichnis

Abbildungsverzeichnis	3
Tabellenverzeichnis.....	3
1 Einleitung	4
1.1 Aufgabenstellung	5
1.2 Zielsetzung.....	5
1.3 Aufbau der Arbeit	6
1.4 Definitionen	6
2 Daten.....	8
2.1 Datenbeschreibung.....	8
2.2 Datenauswahl.....	9
2.3 Datenaufbereitung.....	9
3 Automatische Indexierung	10
4 Lingo, Open Source Software zur automatischen Indexierung	11
4.1 Standard Lingo Konfiguration	14
4.2 Wörterbucherweiterungen.....	15
4.2.1 Lexikalisierung von Wörtern mit einem Basismorphem und deren Flexionen	16
4.2.2 Lexikalisieren von Pluralbildungen	17
4.2.3 Lexikalisieren von Wortstämmen	18
4.2.4 Lexikalisieren von Affixen	21
4.2.5 Lexikalisieren von Komposita	23
4.2.6 Verwendung von Bindestrichen.....	24
4.2.6.1 Bindestrich-Verbindungen (Fall I).....	30
4.2.6.2 Verwendung von Ergänzungsstrichen (Fall VI, Fall VII).....	30
4.2.7 Lexikalisieren von Mehrwortgruppen.....	43

4.2.8	Indexierung von Eigennamen, Akronymen und Abkürzungen	45
4.2.9	Behandlung von Synonymen	47
4.3	Indexierung von englischsprachige Textteilen	48
4.4	Wörter und Mehrwortgruppen anderer Quellen.....	49
4.5	Indexierung der Kontrollmenge	49
5	Indexierungsergebnisse.....	50
5.1	Beschreibung der Indexierungsergebnisse.....	50
5.2	Bewertung der Indexierungen.....	53
6	Zusammenfassung.....	53
7	Literatur- und Quellenverzeichnis	54
8	Anhang	56

Abbildungsverzeichnis

Abb. 1: Fallentscheidung von Binde- und Ergänzungsstrichen.....	29
Abb. 2: Lösungsschema Fall VI, mit <i><und></i> als Token 2	33
Abb. 3: Lösungsschema Fall VI, mit Komma als Token 2.....	37
Abb. 4: Lösungsschema Fall VI, mit <i><bzw></i> als Token 2.....	39
Abb. 5: Lösungsschema Fall VII	42

Tabellenverzeichnis

Tab. 1: Lingo Standardkonfiguration, Kategorien der nicht identifizierten Token	15
Tab. 2: Präfixkombinationen mit <i>-fälle</i>	18
Tab. 3: Übersicht für den Wortstamm <i>-fehl-</i>	19
Tab. 4: Auszug aus Präfixliste	21
Tab. 5: Bindestrich-Verwendungen	25
Tab. 6: Anzahl der Token für die Kombinationen des Fall VI.....	32
Tab. 7: Lösung Fall VI mit <i><und></i> als Token 2	32
Tab. 8: Fall VI, Zusammensetzungen mit <i>raumfahrt</i>	35
Tab. 9: Lösungen Fall VI mit Komma als Token 2	36
Tab. 10: Lösung Fall VI mit Schrägstrich als Token 2.....	38
Tab. 11: Lösung Fall VI mit <i><bzw></i> als Token 2	38
Tab. 12: Fall VI, <i><oder></i> als Token 2	40
Tab. 13: Fall VI, Indexierungsergebnisse ohne Stoppwortliste.....	40
Tab. 14: Fall VI, Indexierungsergebnisse mit Stoppwortliste	41
Tab. 15: Anzahl Token die Kombinationen des Fall VII	41
Tab. 16: Fall VII mit <i><und></i> als vorangestelltes Token	42
Tab. 17: Indexierungsergebnisse der Testmenge.....	50
Tab. 18: Indexierungsergebnisse der Kontrollmenge	51
Tab. 19: Übersicht nicht identifizierter Token pro Dokument	53

Automatisches Indexieren technischer Kongressschriften

1 Einleitung

Die traditionellen - meist papiergebundenen - Publikations- und Informationsformen von wissenschaftlich-technischer Information werden zunehmend durch elektronisch verfügbare digitale Formen ergänzt bzw. ersetzt. Die weltweite Vernetzung erlaubt den direkten Online-Zugriff auf den größten Teil des weltweiten wissenschaftlich-technischen Wissens.¹ Die elektronische Suche nach relevanten Dokumenten erfolgt über so genannte Retrievalsysteme. Sie ermöglichen unter anderem dem Nutzer die Formulierung seines Informationsbedarfs als Suchanfrage. Eine Aufgabe von Retrievalsystemen ist es, einen Vergleich zwischen den gespeicherten Dokumenten und der Suchanfrage herzustellen und die Dokumente auszuwählen, die der Suchanfrage am besten entsprechen. Grundsätzlich vergleichen Retrievalsysteme den als Suchanfrage geäußerten Informationsbedarf nicht direkt mit den Dokumenten, sondern mit Metadaten, die die Dokumente repräsentieren.² Für die Suche sind die, den Inhalt von Dokumenten beschreibenden, semantischen Metadaten von besonderer Bedeutung.³ Metadaten werden nach bestimmten Regeln erstellt, sie können Terme aus einem kontrollierten Vokabular sein, Notationen oder auch Wörter, die direkt aus dem Text der Dokumente entnommen wurden. Die meisten Retrievalsysteme ermöglichen dem Nutzer, seine Anfrage in natürlicher Sprache zu verfassen; dies erfordert eine „Übersetzung“ der Suchanfrage in eine Form, die mit den Metadaten der Dokumente vergleichbar ist. Das System vergleicht anhand festgelegter Methoden die Repräsentanten von Dokument und Suchanfrage und zeigt dem Nutzer die Dokumente, die seinen Suchkriterien entsprechen. Das Ergebnis der Suchanfrage ist damit unter anderem direkt abhängig von der Qualität der Metadaten der Dokumente. Je treffender die Metadaten ein Dokument repräsentieren, desto besser kann ein Dokument den Suchkriterien zugeordnet werden.

Die Methoden und Verfahren, wie Dokumenten inhaltskennzeichnende Bezeichnungen zugeteilt werden, unterscheiden sich nach der Art der Durchführung in maschinelle (automatische) und intellektuelle Verfahren. Der Prozess als solches wird als inhaltliche

¹ Vgl. BMBF, 2002, S. 1

² Vgl. Ferber, 2003, S. 25

³ Vgl. Schütz, 2004, S.340

Erschließung bezeichnet. Unter intellektueller Erschließung versteht man die Zuweisung von Indextermen nach einer von Menschen vorgenommenen Inhaltsanalyse.⁴ Sie strebt die korrekte Repräsentation von Dokumentinhalten auf der Bedeutungsebene an.⁵ Automatische Verfahren können dies nicht leisten; sie bereiten Dokumente so auf, dass sie für eine Suche über Indexterme bestmöglich wieder auffindbar sind.⁶ Sie beinhalten keine Inhaltsanalyse auf der Bedeutungsebene des Dokumentes. Die Analyse ist an statistischen oder linguistischen Kriterien orientiert.⁷ Intellektuelle Verfahren sind in der Regel zeit-, personal- und damit kostenaufwendiger als automatische Verfahren.

Diese Arbeit beschreibt die Erstellung inhaltlicher Metadaten mit dem automatischen Indexierungssystem Lingo.

1.1 Aufgabenstellung

Die Deutsche Gesellschaft für Luft- und Raumfahrt Lilienthal-Oberth e.V. (DGLR) veranstaltet jährlich den Deutschen Luft- und Raumfahrtkongress. Als Dokumentation zum Kongress erscheinen die Kongressschriften in jedem Jahr in schriftlicher Form in drei Bänden. Zusätzlich ist eine CD-ROM Version erhältlich, die jeweils den Band III des Vorjahres und die Bände I und II des aktuellen Jahres enthält. Jede CD enthält eine Volltext-Suchfunktion. Eine thematische oder bibliographische elektronische Suche außerhalb der CD Suchfunktion ist nicht gegeben. Die DGLR beabsichtigt für die Zukunft eine Suchmöglichkeit im Internet bereitzustellen, die sowohl eine bibliographische als auch eine thematische Suche erlaubt. Als Testmenge stehen alle Kongressschriften der Jahre 2004 bis 2006 zur Verfügung.

1.2 Zielsetzung

Ziel dieser Arbeit ist es, die deutschsprachigen Volltexte der Kongressschriften mit Hilfe der Open Source Software Lingo⁸ so zu indexieren, dass der erstellte Index als Grundlage für die thematische Suche dienen kann. Zur Erreichung dieses Ziels ist es notwendig, die in Lingo vorhandenen Lexika um die in den Kongressschriften verwendete Fachsprache zu erweitern.

⁴ Vgl. Bertram, 2005, S. 83

⁵ Vgl. Nohr; 2004, S. 216

⁶ ebd.

⁷ Vgl. Nohr, 2005, S. 34

⁸ www.lex-lingo.de

1.3 Aufbau der Arbeit

Zur erfolgreichen Indexierung der Kongressschriften werden zwei Wege beschritten. Zum einen wird versucht, anhand der Volltexte das Lingo Vokabular gezielt um die verwendeten Terme zu erweitern. Um das Vokabular unabhängiger von den Themen der Kongresse der Testmenge zu gestalten, wird parallel dazu aus anderen Quellen eine Wortliste erstellt, die ebenfalls lexikalisiert wird. Am Ende der Arbeit wurden die Lexikonerweiterungen zusammengeführt, mit diesen eine Indexierung an einer Kontrollmenge durchgeführt und das Ergebnis bewertet.

Zu diesem Zweck werden die Kongressschriften in eine Testmenge (die Jahrgänge 04 und 05), sowie in eine Kontrollmenge (Jahrganges 06) geteilt. Grundlage für die Lexikonerweiterung ist eine Indexierung mit der Standardeinstellung und den Standardlexika von Lingo. Bei dieser Indexierung wird eine Liste von Wörtern generiert, die Lingo nicht identifizieren konnte. Diese Liste wird unter verschiedenen Aspekten untersucht und Kandidaten für die Lexikonerweiterung werden ermittelt. Am Ende der Untersuchung jedes Teilaspektes werden die gefundenen Kandidaten lexikalisiert und eine erneute Indexierung der Textmenge vorgenommen. Zusätzlich wird die Testmenge nach Mehrwortgruppen und Synonymen durchsucht und in die entsprechenden Lexika aufgenommen. Die entstandenen Lexikonerweiterungen werden mit der aus anderen Quellen zusammengetragenen Lexikoneinträgen zusammengeführt und eine erneute Indexierung der Testmenge durchgeführt. Abschließend wird die Kontrollmenge indexiert und im Vergleich der beiden Indexierungen die Lexikonerweiterungen bewertet.

1.4 Definitionen

Algorithmus

„Algorithmen sind Schritt-für-Schritt-Anleitungen zur Lösung mathematischer Aufgaben, deren einzelne Ausführungsschritte so einfach sind, dass sie ohne weitere Hilfsmittel ausgeführt werden können.“⁹

Indexierung

„Beim Indexieren [...] werden dem Dokument einzelne inhaltskennzeichnende Bezeichnungen zugeteilt, um dem Nutzer einen schnellen Zugang zu Dokumenteninhalten zu ermöglichen.“¹⁰

⁹ Grimm, 2005 S. 19

Information Retrieval

Ist im weiteren Sinne die Repräsentation, Speicherung und Organisation von sowie der Zugriff auf Informationen, die in Retrievalsystemen gespeichert werden können.¹¹

Morphem

„Das Morphem ist die kleinste bedeutungstragende Einheit im Sprachsystem.“¹²

Basismorphem (BM)

Sie bilden den Kern eines Wortes und vermitteln die begriffliche Bedeutung. Die Verbindung zweier Basismorpheme wird als Komposition bezeichnet.¹³ Basismorphem wird in der Arbeit synonym mit Kernmorphem verwendet.

Wortbildungsmorpheme (WBM)

Sie werden an Basismorpheme als Präfixe oder Suffixe angefügt. Wortbildungen mithilfe von Präfixen und Suffixen wird als Derivation bezeichnet.¹⁴

Flexionsmorpheme (FM)

Sie kennzeichnen die verschiedenen Wortformen bei Substantiven, Verben und Adjektiven, die durch Flexion entstehen.¹⁵

Precision

Die Precision beschreibt die Genauigkeit der Suche. Mathematisch ausgedrückt ist die Precision der Quotient aus allen gefundenen relevanten Dokumente und allen gefundenen, also allen relevanten und allen nicht relevanten Dokumenten.¹⁶

Recall

Der Recall beschreibt die Vollständigkeit der Suche. Mathematisch ausgedrückt als Quotient aus allen gefundenen relevanten Dokumenten und allen relevanten Datensätzen die vorhanden sind.¹⁷

¹⁰ Bertram, 2005, S. 24

¹¹ Vgl. Bertram, 2005, S. 19

¹² Felgentreu, Friedrich, Huster et al., 2002, S. 109

¹³ Vgl. Ebd., S. 109

¹⁴ Vgl. Ebd., S. 109

¹⁵ Vgl. Ebd., S. 110

¹⁶ Vgl. Stock, 2000, S. 121

Token

Ein Token ist eine Folge von Zeichen die von Lingo als eine zusammengehörige Einheit bearbeitet wird. Token werden im Text kursiv und in spitzen Klammern dargestellt (<token>).

Programmcode/Programmkonfiguration

Um die Lesbarkeit des Textes zu erhöhen, werden Einträge in Lingo-Wörterbücher, Konfigurationsdatei oder anderen Programmteilen kursiv vorgenommen. Lingo berücksichtigt keine Groß- und Kleinschreibung, alle Einträge werden ohne Großbuchstaben vorgenommen.

Die Arbeit basiert zu einem großen Teil auf empirischen Daten. Im Text wird an verschiedenen Stellen auf Tabellen verwiesen, die sich im Anhang befinden. Der Anhang ist aufgrund seines Umfangs in zwei Teile unterteilt, in einen schriftlichen Teil (Anhang A) und eine elektronischen Teil (Anhang B).

2 Daten

2.1 Datenbeschreibung

Die zu indexierenden Dokumente sind ausnahmslos Manuskripte, die für den Deutschen Luft- und Raumfahrtkongress erstellt wurden. Der Kongress findet jährlich statt und behandelt aktuelle Themen der Luft- und Raumfahrt sowie Querschnittsbereiche wie Medizin, Geschichte, Umwelttechnik, Projekt- und Qualitätsmanagement, Dokumentation und Wissensmanagement. Die Manuskripte sind in Englisch oder Deutsch verfasst und haben einen Umfang von bis zu 11 DIN A4 Seiten. Die Dokumente liegen als PDF-Datei vor.

Strukturiert sind die Manuskripte in Überschrift, Autoren Daten, Zusammenfassung (nicht immer) Text mit Grafiken, Tabellen und Bildern sowie einem Literaturverzeichnis. Die deutschsprachigen Manuskripte enthalten oft englischsprachige Textteile.

Formatiert sind die Texte in einspaltige Überschriften und Autoren Daten. Der Text ist zweispaltig angelegt und Grafiken, Tabellen und Bilder sind je nach Größe ein- oder zweispaltig in den Text eingefügt.

¹⁷ Ebd. S. 121

Für die Kongressdokumentation werden allen Kongressschriften eine Kopfzeile mit dem Text „Deutscher Luft- und Raumfahrtkongress 200x“ und der Name des Hauptautors sowie die Programmnummer eingefügt. Alle Dokumente der Jahrgänge 04 und 05 enthalten diese Kopfzeile. Bei den Dokumenten des Jahrgangs 06 wurde diese teilweise entfernt.

2.2 Datenauswahl

Für diese Arbeit wurden alle deutschsprachigen Manuskripte ausgewählt, die der in 2.1 dargestellten Struktur und Format entsprechen. In der Testmenge der Jahrgänge 2004 und 2005 befinden sich 305, und in der Kontrollmenge des Jahrgangs 2006 befinden sich 107 Dokumente.

2.3 Datenaufbereitung

Für die Erstellung des Programms und der Kongressdokumentation werden die Programmnummer, Titel und die Autoren Daten in einer Datenbank gespeichert. Die Jahrgänge 04 und 05 wurden aus dem Acrobat Reader mittels Kopieren und Einfügen in ein Text-Format übertragen. Dabei wurden die Autoren Daten und die Literaturverzeichnisse entfernt. Je nach Dokument sind Absätze bei der Konvertierung erhalten worden oder es wurde an jedem Zeilenende ein Zeilenumbruch eingefügt. Durch die zahlreichen Grafiken, Tabellen und Bilder kommt es vor, dass der Textfluss nicht erhalten wurde und die Sequenz der Absätze in einigen Fällen vertauscht ist. Etwa 30 Dokumente konnten nicht als Text aus den PDF-Dateien entnommen werden. Sie wurden mittels OCR-Bearbeitung¹⁸ in das Text-Format konvertiert.

Die Manuskripte des Jahrgangs 06 wurde mit dem Programm Adobe Acrobat 8 Professional in das Textformat übertragen. Grundsätzlich wurden hier die Textabsätze ohne Zeilenumbrüche am Zeilenende konvertiert. Die Kopfzeilen, die Autoren Daten und die Literaturverzeichnisse wurden entfernt. Bei allen Dokumenten wurden die Inhalte der Tabellen zeilenweise als Text konvertiert, so dass eine Vielzahl von Zahlen und Spaltenbezeichnungen in den Textdateien vorhanden sind.

¹⁸ OCR= Optical Character Recognition (Optische Zeichenerkennung)

3 Automatische Indexierung

Wie im Abschnitt 1 erwähnt, erfolgt bei der automatischen Indexierung keine Inhaltsanalyse im intellektuellen Sinne. Nohr fasst die Verfahrensansätze der automatischen Indexierung in 4 Kategorien zusammen: Die einfache Stichwortextraktion/Volltextinvertierung, statistische Verfahren, Informations- bzw. Computerlinguistische Verfahren und die begriffsorientierten Verfahren.¹⁹ Diese Arbeit befasst sich mit informationslinguistischen Verfahren zur automatischen Indexierung. Sie werden in regelbasierte und wörterbuchbasierte Verfahren unterteilt. Die regelbasierten Verfahren übersetzen die Regeln einer Sprache in Algorithmen. Die im Dokument vorhandenen Terme werden anhand der Algorithmen analysiert und verarbeitet. Diese Verfahren eignen sich für morphologisch wenig komplexe Sprachen wie bspw. das Englische. Für komplexere Sprachen wie die flexionsreiche und kompositumsträchtige deutsche Sprache sind regelbasierende Verfahren nicht ausreichend.²⁰ Hier kommen auf Wörterbücher basierende Verfahren zum Einsatz, bei denen die im Text vorkommenden Terme in Wörterbücher eingetragen sind. Alle Terme, die nicht im Wörterbuch enthalten sind, werden bei der Indexierung nicht berücksichtigt. Die Erstellung und Pflege der Wörterbücher ist aufwendig, liefert aber in der Regel wesentlich bessere Ergebnisse als auf Regeln basierende Verfahren, da die sprachlichen Unregelmäßigkeiten berücksichtigt werden können.²¹

Erfolgt die Indexierung von Texten ausschließlich automatisch, muss ein linguistisches System nach Stock die folgenden Aufgaben bewältigen:

- Erkennung des Schriftsystems
- Spracherkennung
- Erkennen von wenig sinntragenden Worten (Stoppwörter)
- Wortstamm- bzw. Grundformbildung
- Mehrworterkennung
- Eigennamenerkennung
- Identifizierung und Zerlegung von Komposita²²

¹⁹ Nohr, 2004, S.217

²⁰ Nohr, 2005, S. 66ff

²¹ Ebd. S. 66

²² Stock, 2007, S. 216

Die Erkennung des Schriftsystems und der Sprache erfolgt hier manuell durch die Vorselektion der Dokumente in Test- und Kontrollmenge. Die Markierung von wenig sinntragenden Wörtern - so genannte Stoppwörter - wird durch einen entsprechenden Vermerk im Wörterbuch erreicht. Wie mit einem erkannten Stoppwort verfahren wird ist abhängig von den benutzten Programmen. Lingo schließt diese Terme nicht aus dem Indexierungsvorgang aus, da diese für sich genommen zwar keinen Sinn tragen aber im Kontext bspw. bei der Erkennung von Mehrwortgruppen bzw. Eigennamen eine sinntragende Rolle spielen können. Bei der Wortstamm- bzw. Grundformbildung werden die im Text auftretenden Wortformen auf ihren Wortstamm bzw. ihre Grundform reduziert und entsprechend in den Index aufgenommen. Mehrwort- und Eigennamenerkennung erfolgen auch über Wörterbücher, in denen die bekannten Phrasen/Eigennamen eingetragen sind. Für die Zerlegung von Komposita ist es notwendig, dass die Bestandteile des Kompositums im Lexikon enthalten sind; nur dann kann das Kompositum nach festgelegten Regeln in seine Bestandteile zerlegt werden.

Hat der Text eines Dokumentes die Indexierung mit ihren Teilaufgaben durchlaufen, befindet sich eine entsprechende Anzahl von Indextermen im Index. Die Einträge des Index werden vom Retrievalsystem mit den Repräsentanten der Suchanfrage verglichen. Oft kann der Index vom Suchenden eingesehen werden und als alternativer Sucheinstieg genutzt werden. Insgesamt erhöht die automatische Indexierung durch die linguistische Überarbeitung der Indexterme die Anzahl der gefundenen relevanten Dokumente.²³

4 Lingo, Open Source Software zur automatischen Indexierung

An dieser Stelle erfolgt nur einen Überblick über die Grundfunktionen von Lingo. Sie werden im Einzelnen in den entsprechenden Abschnitten näher erklärt und weiterführende Informationen sind unter der Fußnote ²⁴ und im World Wide Web unter www.lex-lingo.de zu finden. Als wörterbuchbasiertes, automatisches Indexierungsprogramm enthält Lingo mehrere Wörterbücher und so genannte „Attendees“. Attendees sind Programmteile für die Bearbeitung von Zeichenketten, dem Errechnen von Termfrequenzen sowie für Eingabe- und Ausgabefunktionen. Wörterbücher und Attendees können über einen Texteditor verändert werden. Der Indexierungsablauf ist flexibel konfigurierbar. Es ist möglich,

²³ Lepsky, Siepmann, Zimmermann, 1996, S. 47ff

²⁴ Lepsky, K; Vorhauer, J, 2006

einzelne Attendees zu- oder abzuschalten, verschiedene Wörterbücher einzubinden oder die Ausgabe der einzelnen Attendees nach bestimmten Kriterien zu filtern.

Lingo benötigt für die Indexierung Text-Dateien und untersucht alle darin vorkommenden Zeichen anhand von bestimmten Zeichensätzen (bspw. ISO-8859-1). In der Regel werden Zeichen, die von Leerzeichen umgeben sind, zu einem Token zusammengefasst und nach bestimmten Regeln mit den Einträgen des Systemwörterbuchs verglichen. Zeichenketten, die Sonderzeichen enthalten, werden so getrennt, dass die Sonderzeichen von der Zeichenkette abgetrennt werden und ein eigenes Token darstellen. Bindestriche und Unterstriche werden nicht abgetrennt. Einträge im Systemwörterbuch erfolgen in folgender Form:

wortform=grundform #wortklasse bspw.: kongress=kongress #s

Da Lingo keine Groß- und Kleinschreibung berücksichtigt, werden alle Systemeinträge ohne Großbuchstaben vorgenommen. Verglichen wird das Token mit der eingetragenen Wortform. Ist das Token erkannt, wird die Grundform mit der Wortklasse in die Ergebnismenge aufgenommen. Das Systemwörterbuch kennt im Wesentlichen 4 Wortklassen: Adjektive (*#a*), Substantive (*#s*) Verben (*#v*) und Eigennamen (*#e*). Der Vergleich von Token und Wortform erfolgt in Abhängigkeit von der Wortklasse nicht als direkter Vergleich sondern erlaubt bei Adjektiven, Substantiven und Verben Flexionen, die in Suffixlisten vermerkt sind. Das Beispiel *kongress* hat als Substantiv unter anderem ein *-e* und *-en* als mögliche Suffixendung vermerkt. Erkannt wird nicht nur das Token <kongress> sondern auch <kongresse> und <kongressen> und alle werden mit *kongress #s* indexiert. Mit Eigennamen, die aus einem Wort bestehen, wird in gleicher Weise verfahren, für sie ist nur eine mögliche Endung mit *-s* vorgesehen. Eigennamen, die aus mehreren Token bestehen, fallen unter die Mehrworterkennung. Bei der Identifizierung von Token verfolgt Lingo den Grundsatz des „longest matching“. Der Vergleich von Token und Wortform beginnt immer von rechts und es wird nach dem längsten Eintrag im Wörterbuch gesucht, der dem Token entspricht. So wird bspw. <kohle> dem Lexikoneintrag *kohle* zugeordnet und nicht *kohl* mit dem erlaubten Suffix *-e*.

Bei zusammengesetzten Wörtern wird ähnlich verfahren. Sie können grundsätzlich nur als Kompositum erkannt werden, wenn alle Bestandteile im Systemwörterbuch lexikalisiert

sind. Zwischen den einzelnen zusammengesetzten Wörtern sind Fugungen möglich, die sich ebenfalls in einer Liste befinden. Wird ein Kompositum erkannt, wird dieses als Ganzes und zusätzliche alle Bestandteile in den Index aufgenommen. So führt das Token <himmelsmechanik> zu den Einträgen: *himmelsmechanik*, *himmel* und *mechanik*. Auch für die Mehrworterkennung müssen alle Einzelwörter zuerst über das Systemwörterbuch identifiziert werden. In einem zweiten Schritt kontrolliert Lingo die Sequenz der identifizierten Token mit den Einträgen im Mehrwortwörterbuch. Identifizierte Mehrwortgruppen werden gesondert ausgegeben und können bspw. invertiert in den Index aufgenommen werden. Besteht für die Suche die Möglichkeit den Index einzusehen, ermöglicht die invertierte Darstellung eine thematische Unterscheidung von Indexeinträgen. So werden die drei Mehrwortgruppen

effektiver anstellwinkel
...
induzierter anstellwinkel
...
kritischer anstellwinkel

bei alphabetischer Sortierung an verschiedenen Stellen im Index aufgeführt und vom Suchenden nur gefunden, wenn die 3 Varianten des Anstellwinkels bereits bekannt sind. Werden die Mehrwortgruppen invertiert alphabetisch sortiert

anstellwinkel, effektiver
anstellwinkel, induzierter
anstellwinkel, kritischer

so werden die unterschiedlichen Varianten an einer Stelle des Indexes sichtbar und ermöglichen damit bei der Suche nach Anstellwinkel eine thematische Unterscheidung.

Zusätzlich verfügt Lingo über eine algorithmische Mehrworterkennung, bei der die Abfolge von Token in Abhängigkeit von ihrer Wortklasse untersucht wird. So wird eine Liste von Kandidaten für das Mehrwortlexikon erzeugt, die gezielt nach Mehrwortgruppen untersucht werden kann. Die Muster nach denen die Kandidaten ausgewählt werden befinden sich in einer Liste, die durch den Nutzer veränderbar ist. So erzeugt bspw. der Eintrag [[AS, "2, I"]] eine Liste von Adjektiv – Substantiv Kombinationen, die nach Mehrwortgruppen untersucht werden kann.

Zusätzlich hat Lingo noch einen Attendee zur Relationierung von Synonymen. Auch die setzt eine Wortidentifizierung über das Systemwörterbuch voraus. Im Synonymwörterbuch werden äquivalente Wortpaare oder auch Mehrwortgruppen in der Form von *Deutsches Institut für Normung=DIN* eingetragen. Ist eine Zeichenkette im Text vorhanden, die dem Synonymeintrag entspricht, wird der Eintrag rechts vom Gleichheitszeichen zusätzlich in den Index aufgenommen und ermöglicht einen Sucheinstieg über beide synonyme Einträge.

Indexterme werden je Dokument nur einmal vergeben. Über den Lingo Attendee „vector_filter“ besteht die Möglichkeit, die Ausgabe von Indexierungsergebnissen zu beeinflussen. So ist bspw. die Ausgabe auf bestimmte Wortklassen einzuschränken oder es können Termfrequenzen wie die einfache oder relative Termhäufigkeit den Indextermen hinzugefügt werden.

Als letzter Attendee sei der „variator“ erwähnt, der nicht identifizierte Token einer weiteren Untersuchung unterzieht. Er wurde für Dokumente entwickelt, die mittels OCR digitalisiert wurden und systematische Zeichenvertauschungen aufweisen. Über eine Liste werden bei nicht erkannten Token bestimmte Zeichenkombinationen durch andere ersetzt und das Token erneut mit den Einträgen des Systemlexikons verglichen. Diese Funktion ermöglicht auch, einige Änderungen der neuen deutschen Rechtschreibung zu berücksichtigen. So ersetzt der Eintrag [*ss, ß*] das *ss* von <*kongress*> durch ein *ß* und identifiziert das Token über den Eintrag *kongreß=kongreß #s*.

Grundgedanke von Lingo ist, die möglichst weitgehende Identifizierung der möglichen Wortformen und Komposita auf eine einfache Weise zu erreichen. Durch die Lexikalisierung von einfachen Wörtern und deren Ableitungen kann mithilfe von wenigen Wortklassen, Suffixlisten und Fugungen im Baukastenprinzip die potentiell unendliche Anzahl von Zusammensetzungen weitgehend erkannt und indexiert werden.

4.1 Standard Lingo Konfiguration

Im Internet ist derzeit die Lingo Version 1.63 verfügbar.²⁵ In der Konfigurationsdatei ist der folgende Indexierungsablauf eingestellt:

- Einlesen des Textes und Erstellen der Token

²⁵ www.lex-lingo.de

- Worterkennung mit dem Systemwörterbuch und dem Wörterbuch für Abkürzungen
- Erkennung und Zerlegung von Komposita mit dem Systemwörterbuch
- Mehrworterkennung mit dem Mehrwortwörterbuch
- Synonymzuordnung mit dem Synonymwörterbuch
- Ausgabe der indexierten Wörter
- Ausgabe aller nicht identifizierten Token
- Ausgabe aller Mehrwortgruppen
- Ausgabe aller Synonyme

Das Systemwörterbuch enthält 62029 Wortformen aus verschiedenen offenen Quellen und das Abkürzungswörterbuch verfügt über 67 eingetragene Abkürzungen. Die Wörterbücher für die Mehrworterkennung und die Synonymrelation enthalten nur beispielhafte Einträge für den Buchstaben A.

4.2 Wörterbucherweiterungen

Um zukünftige Lingo Wörterbucherweiterungen nutzen zu können, ohne dass die eigenen Erweiterungen verloren gehen, werden alle Einträge für das Systemwörterbuch in einem separaten Nutzerwörterbuch vorgenommen. Das Nutzerwörterbuch wird über die Konfigurationsdatei in den Indexierungsprozess einbezogen. Änderungen im Mehrwort- und Synonymwörterbuch werden direkt in den Wörterbüchern vorgenommen.

Bei der ersten Indexierung der Testmenge mit dem Standard Systemwörterbuch ergibt eine Liste von erkannten Wörtern, die in einem oder mehreren Dokumenten vorkommt, 44103 Einträgen. Die Liste der nicht identifizierten Token enthält 28871 Einträge. Die nicht identifizierten Token der Lingo Standardindexierung wurden in 12 Kategorien gruppiert.

Nr.	Anzahl unterschiedlicher Token	Kriterium
1	55	Ein Zeichen
2	835	Zwei Zeichen
3	213	Abkürzungen Akronyme
4	462	Eigennamen
5	6543	Fragmente Fehler
6	7281	Mehrere Kernmorpheme
7	783	Unregelmäßige Pluralbildung
8	3626	Englische Sprache
9	1079	Buchstaben- Zahlenkombinationen
10	5111	Token mit <->
11	2502	Nur ein Kernmorphem
13	381	Mehrere Kernmorpheme mit fremdsprachigen Anteil
	28871	

Tab. 1: Lingo Standardkonfiguration, Kategorien der nicht identifizierten Token

Auffällig ist die hohe Anzahl von Fragmenten und Fehlern, die sich in der Mehrzahl aus dem zweispaltigen Textaufbau, Tabellen und der Datenkonvertierung vom PDF- in das Text-Format ergeben. Da es sich bei den Kategorien 1 und 2 fasst ausschließlich um Wortfragmente und Abkürzungen aus Tabellen handelt, werden sie im Folgenden nicht weiter untersucht. Die anderen sind Grundlage für die systematische Bearbeitung und Erschließung der in den Manuskripten verwendeten Fachsprache. In der Regel werden nur Token betrachtet, die mindestens in zwei Dokumenten vorkommen.

4.2.1 Lexikalisierung von Wörtern mit einem Basismorphem und deren Flexionen

Die Verwendung der Bezeichnungen Kern- und Basismorphem (BM) erfolgt synonym. Sie bilden den Kern eines Wortes und vermitteln die begriffliche Bedeutung. Am Anfang der systematischen Lexikalisierung werden Wörter, die nur ein Kernmorphem besitzen sowie deren Ableitungen und Flexionen behandelt. Die Liste der Token wird nach Häufigkeit sortiert und die Token erneut kategorisiert. Die Token <von>, <als>, <vor> und <vom> wurden als hochfrequente Terme (in mindestens 205 von 305 Dokumenten) als Wortklasse #t lexikalisiert. Durch die Lexikalisierung als Wortklasse #t wird erreicht, dass die Einträge nur als Teil eines Kompositums oder einer Mehrwortgruppe in den Index gelangen. Für den Fall, dass sie als Wort alleine auftreten, können sie durch den Attendee *vector_filter* herausgefiltert werden und gelangen nicht in den Index. Zusätzlich werden 8 Token (bspw. aero, turbo, ultra) als Wortklasse #u lexikalisiert. Mit den Wortklassen #w und den für diese Arbeit eingefügten Wortklassen #u und #g wird nach dem gleichen Muster verfahren wie mit der Wortklasse #t und dienen der Übersichtlichkeit. Alle anderen zu lexikalisierenden Wörter wurden als Grundform oder als Flexion einer Grundform markiert und bei der Eintragung in das Lexikon zusammengeführt. Die Token:

<axial>, <axiale>, <axialen>, <axialer> und <axiales>

werden als ein Lexikoneintrag: *axial=axial #a* lexikalisiert. Über die Suffixliste der Adjektive werden alle 5 Token als *axial* indiziert.

Bei einigen Ausnahmen wurde der Wortformeintrag in das Lexikon verkürzt. Das Adjektiv <kompressible> wird als *kompressibl=kompressibel #a* eingetragen. Durch das Auslassen der Endung -e wird erreicht, dass <kompressibler> oder <kompressiblen> über die Suffixliste identifiziert werden kann und mit *kompressibel* zusammengeführt wird.

Bei der Untersuchung der Liste fällt auf, dass oft Token mit demselben Kernmorphem mit verschiedenen Wortbildungsmorphemen kombiniert werden und nicht im Wörterbuch vorhanden sind. So ist *strömung #s* lexikalisiert, die Wortformen An-, Um- und Zuströmung werden aber nicht erkannt obwohl *zu #t* sich im Lexikon befindet. Die Zerlegung von Zusammenfügungen in Lingo ist von verschiedenen Parametern abhängig. In der Standard-Einstellung werden nur die Komposita erkannt und zerlegt, die mindestens 7 Zeichen lang sind, aus nicht mehr als 5 Einzelworten bestehen, deren einzelne Bestandteile mindestens 3 Zeichen haben und die durchschnittliche Zeichenanzahl der einzelnen Bestandteile mindestens 4 beträgt. Um bspw. <zuströmung> zu indexieren reicht es nicht aus, *zu* und *strömung* zu lexikalisieren, da die Kriterien für die Erkennung und Zerlegung der Zusammenfügung nicht erfüllt sind. Zuströmung muss als ein Eintrag in das Lexikon aufgenommen werden. Diese Problematik betrifft alle Wortformen die aus weniger als 4 Zeichen bestehen, da sie in Kombination miteinander die notwendige Zeichenlänge von 7 Zeichen für die Kompositumserkennung nicht erreichen. Es empfiehlt sich, alle durch Ableitung gebildeten Wortformen mit Wortbildungsmorphemen, die aus weniger als 4 Zeichen bestehen, zu lexikalisieren. Das Thema wird bei der Bearbeitung der Affixe wieder aufgegriffen. Von den 873 Token der Liste wurden 468 lexikalisiert. Durch die Zusammenführung der Flexionen führte das zu 322 Lexikoneinträgen.²⁶

4.2.2 Lexikalisieren von Pluralbildungen

Die deutsche Sprache verwendet für die Pluralbildung Umlaute und Pluralendungen. Lingo erkennt Plural-Wortformen, die über Pluralendungen gebildet werden, über die Suffixlisten. <flaschen> wird über *flasche=flasche #s* mit der Endung *-n* aus der Suffixliste indexiert. Die Pluralbildung über Umlaute mit oder ohne Pluralendung benötigt einen separaten Eintrag im Lexikon: <häuser> wird über *haus=haus #s* nicht identifiziert. Im Lexikon wird *häuser=häuser #s haus #s* eingetragen, um neben dem Plural auch den Singular zu indexieren. Wird nach Haus gesucht so werden auch die Dokumente, die den Plural Häuser enthalten, gefunden.

Um die im vorherigen Abschnitt beschriebene Problematik der Indexierung von Ableitungen mit kurzen Wortbildungsmorphemen zu berücksichtigen, werden die Wortformen nach möglichen Wortbildungen mit entsprechenden Präfixen untersucht. Zu

²⁶ Anhang B-1 EinBasismorphem.xls
Weitere Beispiele befinden sich im Anhang B-1-1 bis B-1-14

diesem Zweck wurde eine Liste von heimischen Präfixen gebildet.²⁷ So ergeben sich bspw. für den Plural <fälle> die Kombinationen in Tab. 2

Wortform	Grundform1	Wortklasse1	Grundform2	Wortklasse2
<i>fälle</i>	<i>Fälle</i>	<i>#s</i>	<i>fall</i>	<i>#s</i>
<i>abfälle</i>	<i>Abfälle</i>	<i>#s</i>	<i>abfall</i>	<i>#s</i>
<i>anfälle</i>	<i>Anfälle</i>	<i>#s</i>	<i>anfall</i>	<i>#s</i>
<i>ausfälle</i>	<i>Ausfälle</i>	<i>#s</i>	<i>ausfall</i>	<i>#s</i>
<i>einfälle</i>	<i>Einfälle</i>	<i>#s</i>	<i>einfall</i>	<i>#s</i>
<i>unfälle</i>	<i>unfälle</i>	<i>#s</i>	<i>unfall</i>	<i>#s</i>
<i>vorfälle</i>	<i>vorfälle</i>	<i>#s</i>	<i>vorfall</i>	<i>#s</i>
<i>zufälle</i>	<i>zufälle</i>	<i>#s</i>	<i>zufall</i>	<i>#s</i>

Tab. 2: Präfixkombinationen mit *-fälle*

Bei Token, die aus mehreren Wörtern zusammengesetzt sind, wurde nur der Teil mit der Pluralform berücksichtigt. Die Grundformen werden dann bei der Lexikalisierung mit den Singular-Grundformen zusammengeführt. So ergeben sich 213 Einträge in das Nutzerlexikon.²⁸

4.2.3 Lexikalisieren von Wortstämmen

Viele der nicht erkannten Token enthalten Wortstämme ohne Flexionsendungen. In der verwendeten Fachsprache sind Wortbildungen, die Wortstämme ohne Flexionsendungen enthalten, verhältnismäßig häufig anzutreffen. Lingo versucht, die in den deutschen Texten vorkommenden Wörter auf die Grundform zu reduzieren und bildet keine Wortstämme.²⁹ Daraus ergibt sich, dass sich in den Lexika grundsätzlich keine Wortstämme befinden. Es wird die Möglichkeit untersucht, das Systemlexikon gezielt um Wortstämme zu erweitern. Dabei wird nicht die Absicht verfolgt, Wortformen auf Wortstämme zurück zuführen sondern lediglich Wortstämme so zu lexikalisieren, dass sie als Bestandteil von Komposita für deren Erkennung zur Verfügung stehen. Alle lexikalisierten Wortstämme und bestimmte Präfixe und präfixartig verwendete Elemente (siehe unten) werden als Wortklasse *#u* eingetragen. Dadurch wird ein als Fragment auftretender Wortstamm nicht lexikalisiert und es erfolgt kein Wortformenvergleich mit Hilfe der Suffixlisten.

²⁷ Siehe Anhang A-1, S. 63

²⁸ Anhang B-22 Lexikoneinträge der Pluralbearbeitung.xls

²⁹ Mit der Ausnahme von identischen Wortstämmen und Grundformen.

In der Liste im Anhang A-2 sind die gefundenen Wortstämme zusammengefasst. Besonders häufig werden bspw. die Wortstämme *-brenn-* im Bereich der Triebwerkstechnik, *-prüf-* und *-mess-* im Prüf- und Messwesen sowie *-ström-* für die Strömungslehre verwendet. Die in der Liste grau hinterlegten Einträge wurden einzeln untersucht.

Das Vorgehen wird am Beispiel des Wortstamms *-trag-* beschrieben (Tab. 3):

Häufigkeit	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ
11	messfehler	komp			fehl	er			1
5	messfehlers	komp			fehl	ers			1
4	fehlströmung				fehl			komp	3
3	fehlstarts				fehl			komp	3
2	fehl				fehl				-
2	fehlalarmen				fehl			komp	3
2	fehler-								-
2	fehlereinflüsse				fehl	er		komp	2
2	fehlinterpretation				fehl			komp	3
2	fehlsteuerung				fehl			komp	3
2	messfehlern	komp			fehl	ern			1

Typ	Anzahl	BM			BM	FM		BM	
1	3	x			x	x			
2	1				x	x		x	
3	5				x			x	

	Lexikoneintrag	Findet						
	messfehler	1;2;11;						
	Fehl #u	3;4;6;9;10;						
		2	8					

Tab. 3: Übersicht für den Wortstamm *-fehl-*

Im oberen Teil der Tabelle sind die nicht identifizierten Token mit der Anzahl der Dokumente in denen sie vorkommen aufgelistet sowie eine Aufschlüsselung ihrer Kombination mit anderen Basis-, Wortbildungs- und Flexionsmorphemen. Darunter eine Übersicht die zeigt, dass *-fehl-* in den Manuskripten meist ein Kompositum mit anderen Basismorphemen bildet und keine Wortformen mit Wortbildungsmorphemen in der Liste enthalten sind. Die relativ geringe Anzahl der Token und das Fehlen von Wortformen mit Wortbildungsmorphemen ist mit einer bereits erfolgten umfangreichen Lexikalisierung (106 Wortformen mit *-fehl-*) zu erklären.³⁰ Der untere Teil der Tabelle zeigt die aufgrund der Liste erstellten Lexikoneinträge. Das Token mit dem Ergänzungsstrich wird nicht lexikalisiert und im Abschnitt 4.3.6 behandelt. Das Token *<fehl>* wird über *fehl=fehl #u*

³⁰ Anhang A-3; Lingo Systemwörterbuch, Einträge mit *-fehl-*, S. 66

identifiziert aber nicht indexiert. <fehlereinflüsse> wird identifiziert über *fehler #s* im Systemwörterbuch und *flüsse=flüsse #s fluss #s* im Nutzerwörterbuch dessen Eintrag bei der Untersuchung der Pluralformen erfolgte. Der Eintrag *messfehler* wird jedoch nicht in das Lexikon aufgenommen, da *fehler #s* bereits im Lexikon vorhanden ist und *mess #u* als weiterer lexikalisierte Wortstamm dazu führt, dass <*messfehler*> identifiziert wird.

32 weitere Wortstämme wurden in gleicher Weise bearbeitet und führten insgesamt neben einigen anderen Einträgen zu 55 Wortstämmen die in das Lexikon aufgenommen wurden.³¹ Die Lexikalisierung erfolgte nach folgendem Muster:

- Wortstämme mit angehängter Flexionsendung oder Wortbildungsmorphem werden als Wortableitung lexikalisiert. (*strömen #v*; oder *strömung #s*)
- Wortstämme mit vorangestelltem und angehängtem Wortbildungsmorphem werden ebenfalls als Wortableitung lexikalisiert. (*zuströmen #v*; oder *anströmung #s*)
- Wortstämme mit folgendem und oder vorangestelltem Basismorphem (Kompositum) können als Wortstamm in das Lexikon eingetragen werden. (<*verdichterdrehzahl*> als *verdichter #s*, *dreh #u*, und *zahl #s*)
- Wortstämme mit angehängtem Basismorphem und vorangestelltem Wortbildungsmorphem sind bei gebundenen Wortbildungsmorphemen und bei freien Wortbildungsmorphemen von weniger als 4 Zeichen Länge als Einheit in das Lexikon aufzunehmen (*anstöm #u*, *winkel #s* oder *einstell #u*, *winkel #s*).

Zur Kontrolle, ob die Lexikalisierung von Wortstämmen negative Auswirkungen auf das Indexierungsergebnis hat, wurde mit den Lexikonerweiterungen eine weitere Indexierung durchgeführt. Aus der Liste aller erkannten Wörter sind alle diejenigen ausgewählt worden, die einen entsprechenden Wortstamm enthalten und auf eine fehlerhafte Indexierung untersucht. Die Zeichenkette ‚fehl‘ ist in 237 Indexeinträge enthalten. In der Liste befinden sich keine falschen Indexierungen, die auf die Lexikalisierung des Wortstamms zurückzuführen sind.³²

Die Überprüfung der anderen Wortstämme bestätigt das Ergebnis. Die Tabellen der lexikalisierten Wortstämme befinden sich im Anhang A-4 und die Tabellen der Wortstammüberprüfung im Anhang B-2.

³¹ Die Tabellen der Wortstämme befinden sich im Anhang A-4, S. 67ff

³² Anhang A-5; Indexierte Wörtern mit Zeichenfolge *-fehl-*, S.108

4.2.4 Lexikalisieren von Affixen

Affixe sind an den Wortstamm vorne (Präfixe) oder hinten (Suffixe) angefügte Wortbildungsmorpheme. Bisher wurden Präfixe nur bei der Untersuchung der Pluralbildung systematisch berücksichtigt. Die nächsten beiden Abschnitte befassen sich mit den nicht identifizierten Token, die anhand einer Präfix- und Suffixliste sortiert und untersucht werden.

Präfixe

Die Präfixliste (Ausschnitt in Tab. 4) zeigt in der ersten Spalte die Anzahl der Einträge des Lingo Systemwörterbuchs für das entsprechende, in der zweiten Spalte aufgeführte Präfix.³³ Die Spalte Klasse enthält die Wortklasse des Präfixes sofern es bereits im Systemwörterbuch enthalten ist. Die nächste Spalte zeigt die Anzahl der nicht identifizierten Token (ohne doppelte Nennungen) für die Indexierung die bereits alle vorherigen Lexikonerweiterungen enthält. Die Spalte Lingo Standard zeigt die Anzahl der nicht identifizierten Token der Standard Indexierung. Die letzte Spalte enthält Anmerkungen bzw. Lexikoneinträge für Präfixe, die im Folgenden nicht einzeln aufgeführt sind. Die grau hinterlegten Präfixe werden einzeln untersucht.

Lingo	Präfix	Klasse	stamnon >1	Lingo Standard >1	Anmerkungen/Lexikalisierung
1975	ab-		32	50	
1767	an-		62	84	
1249	auf-	t	10	17	
1653	aus-	t	18	22	
2245	be-	t	67	86	
216	bei-	t	3	4	beiwert, beibehält
569	durch-	w	3	4	durch-fluss
1600	ein-	t	22	30	
838	ent-		20	24	
1077	er-	t	47	49	

Tab. 4: Auszug aus Präfixliste³⁴

Es fällt auf, dass von den 127 Präfixen bereits 41 als eigenständiger Eintrag lexikalisiert sind. Es handelt sich ausschließlich um freie Morpheme wie auf, aus, über, zu, usw., die der Wortklasse #t (meist bestimmte und unbestimmte Artikel und Pronomen) oder der Wortklasse #w (u.a. Numeralien, Präpositionen) zugeordnet wurden. Die Lexikalisierung der verwendeten Präfixe orientiert sich an dem im Abschnitt Wortstämme beschriebenen

³³ Anhang A-6, Präfixliste, S. 111

³⁴ Die Präfixliste basiert auf Reinhardt, Köhler, Neubert, 1992, S. 35ff

Muster. Präfixe wie bspw. *ultra*, *trans*, *mega*, *mikro*, und Wortstämme mit Präfixen wie bspw. *anpress*, *aufheiz*, *umlenk* werden als Wortklasse #u in das Wörterbuch aufgenommen. Neben Wortfragmenten, englischen Termen und Ergänzungsstrichverwendungen sind die Mehrzahl der Token Ableitungen von Wörtern, die als Wortklasse #a, #s oder #v lexikalisiert werden.³⁵

Bei dem Präfix *pro-* sind 37 Token nicht erkannt worden. Darin enthalten sind 6 Bindestrichverbindungen, 23 englische Wörter, 3 Fragmente/Fehler, ein Eigenname, 3 Flexionen von <programmierbar> und ein Kompositum mit Profil. Für das Präfix *pro-* erfolgte lediglich der Lexikoneintrag: *programmierbar* #a.³⁶

Bei dem Präfix *in-* sind von 91 Token 17 Bindestrichverbindungen, 42 englische Wörter, 13 Fragmente/Fehler die nicht lexikalisiert werden. Aus den übrigen Token ergeben sich 5 Substantive, 3 Adjektive, 3 Wortstämme und zwei Eigennamen, die in das Nutzerwörterbuch eingetragen werden.³⁷ Insgesamt werden bei der Untersuchung der Präfixe 285 Lexikoneinträge vorgenommen.³⁸

Suffixe

Die Suffixliste ist ähnlich aufgebaut wie die Präfixliste. Die vierte Spalte zeigt die Indexierungsergebnisse mit den Lexikonerweiterungen der vorherigen Abschnitte. Alle nicht identifizierten Token wurden nach den Suffixen gruppiert und untersucht. Suffixe, die eine entsprechende Häufung von Token haben, sind grau markiert und werden einzeln untersucht.³⁹ Bei den heimischen Suffixen sind Regelmäßigkeiten in der Wortbildung festzustellen, für die eine eigenständige Lexikalisierung untersucht wurde. Als Beispiel dient das Suffix *-bar*, welches im Systemwörterbuch als Wortklasse #a und #s enthalten ist. Mit *-bar* werden mittels verbalen Basen Adjektive gebildet (*schwenkbar*, *aufklappbar*).⁴⁰ Die relativ hohe Zahl von 70 nicht identifizierten Token ergibt sich aus der Tatsache, dass bei der Wortbildung von der Basis nur der Wortstamm verwendet wird. Da im Lexikon kaum Wortstämme enthalten sind, und bei der Wortbildung die Wortstämme oft mit Präfixen kombiniert werden, können die gebildeten Adjektive nicht über die

³⁵ Anhang A-7, S. 114ff

³⁶ Anhang A-7-20, S. 139

³⁷ Anhang A-7-16, S. 134

³⁸ Anhang B-21 Lexikoneinträge der Präfixbearbeitung

³⁹ Anhang A-8, S. 149

⁴⁰ Vgl. Reinhardt, Köhler, Neubert, 1992, S. 45

Kompositumserkennung identifiziert werden. Ein Beispiel: das Token <skalierbar> wird nicht erkannt, obwohl *skalieren* #v und *bar* #a im Wörterbuch vorhanden sind. Für den Fall, dass *skalier* #u nach dem Muster der Wortstämme lexikalisiert wäre, würde das zu einer Indexierung von *skalierbar* und *bar* führen. Wird nach *bar* im Sinne von Theke gesucht, würden auch Dokumente die *skalierbar* enthalten gefunden, was die Precision reduzieren würde. Somit unterbleibt eine Lexikalisierung von *skalier* #u.

Für die Lexikalisierung von Suffixen bietet sich die Wortklasse #u nicht an, da keine Suffixendungen berücksichtigt werden. So wird bei *handwerk*=*handwerk* #s und *lich*=*lich* #u <*handwerklich*> lexikalisiert, <*handwerkliche*> bzw. <*handwerklicher*> jedoch nicht und müssten gesondert in das Lexikon aufgenommen werden. Um dies zu vermeiden, müssten die Suffixe als Wortklasse #a, #v oder #s lexikalisiert werden, was dazu führen würde, dass Suffixe wie *-lich*, *-keit* oder *-mäßig* über die Kompositumserkennung im Index erscheinen würden; dies ist jedoch nicht beabsichtigt. Hinzu kommt, dass diese Lösung nur für Suffixe möglich ist, die bei der Wortbildung nur eine Wortklasse bilden. Die Lexikalisierung hätte bei einigen heimischen Suffixen (bspw. *-keit*, *-ung* und *-lich*) den Vorteil, dass die Lexikalisierung des Suffixes einige weitere Lexikoneinträge ersparen würde. Dieses Vorgehen ist aber nur bei sehr wenigen Suffixen und suffixartig verwendeten Elemente möglich. Die Suffixe, die als eigenständige Wörter vorkommen können, sind bereits im Lingo Systemwörterbuch enthalten.⁴¹ Somit wird eine Sonderregelung für einige wenige Suffixe nicht weiter verfolgt und die Token, sofern sie in mindestens 2 Dokumenten vorkommen, in der im Text vorkommenden Wortform lexikalisiert. Insgesamt wurden die Token auf 265 verschiedene Suffixendungen untersucht, 15 mit relativ unbekanntem Token wurden im Anhang 9 einzeln dokumentiert. Die Bearbeitung der Suffixliste ergab 108 Lexikoneinträge.

4.2.5 Lexikalisieren von Komposita

In der Standard Lingo Indexierung ist die Gruppe der nicht identifizierten Token, die mehrere Kernmorpheme enthalten, mit fast 7300 die größte Gruppe (Tab. 1). Hinzu kommen 381 Zusammensetzungen, die ein oder mehrere Fremdwörter enthalten. Wie bereits beschrieben, müssen alle Wortbestandteile eines Kompositums lexikalisiert sein; die verwendeten Wortformen können nur über die Suffixlisten und Fugungen identifiziert

⁴¹ Anhang A-8, Spalte: Wortklasse, S. 149

werden. Die in der Standard Lingo Einstellung vorhandenen Parameter wurden nicht verändert.

Die alphabetisch sortierte Liste der Token wurde systematisch nach den nicht lexikalisierten Kompositumsbestandteilen untersucht. Durch die Sortierung fallen häufig verwendete Präfixe und Präfix-Wortstamm-Kombinationen (bspw. *leucht-*, *quetsch-*, *setz-* oder *versprüh*) auf, die zu weiteren 53 Lexikoneinträgen der Wortklasse #u führen. Zusätzlich werden alle Zusammensetzungen mit *öl* aufgrund der Mindestzeichenlänge von 3 für die Kompositumsbestandteile lexikalisiert. Insgesamt werden 219 Lexikoneinträge nach dem bereits beschriebenen Muster vorgenommen. Darunter befinden sich 20 fremdsprachige Einträge, auf die im Abschnitt 4.5 eingegangen wird.

4.2.6 Verwendung von Bindestrichen

Der Duden beschreibt 11 verschiedene Fälle, in denen Bindestriche verwendet werden können oder müssen.⁴² Betrachtet man die verschiedenen Verwendungsmöglichkeiten auf der Zeichenebene, so sind -abgesehen von der Worttrennung am Zeilenende- nur 3 Varianten zu beobachten:

- Bindestriche innerhalb einer Zeichenkette (Umsatzsteuer-Tabelle),
- Bindestriche am Anfang einer Zeichenkette (Kameraposition und -brennweite),
- Bindestriche am Ende einer Zeichenkette (Ein- und Ausgang).

Wie beschrieben, werden Bindestriche bei der Erstellung der Token nicht von den Zeichenketten abgetrennt. Wörter mit Ergänzungsstrichen werden mit dem führenden oder angehängten Ergänzungsstrich als Token bearbeitet.

Die Indexierung von Token die Bindestriche enthalten, erfolgte nach folgenden Regeln:

Für Token, die einen Bindestrich innerhalb der Zeichenkette (nicht an erster oder letzter Stelle) haben, genügt die Lexikalisierung eines Teils der Zeichenkette. Das Token *<navier-stokes-gleichung>* ist der im Lexikon vorhandene Eintrag von *gleichung #s* ausreichend, um das Token als Kompositum zu erkennen und zu indexieren. Auch bei Bindestrichen gilt die Regelung, dass eine Zeichenkette aus mindestens 3 Zeichen bestehen muss, um bearbeitet zu werden. Das Token *<x-band>* wird dementsprechend nicht indexiert, obwohl *<band> #s* sich als Eintrag im Lexikon befindet. Somit ergibt sich als Regel für die Lexikalisierung von Bindestrichverbindungen, dass alle Verbindungen die einen

⁴² Duden, Die deutsche Rechtschreibung, S.31-34

Bestandteil von weniger als 3 Zeichen haben, als Ganzes in das Lexikon eingetragen werden. Für $\langle x\text{-band} \rangle$ erfolgt der Eintrag $x\text{-band}=x\text{-band} \#s$. Bei Token, die mehr als einen Bindestrich enthalten, ist es ausreichend, dass ein Teil über das Lexikon identifiziert wird und ein weiterer Teil mindestens aus 3 Zeichen besteht. Token, die einen Ergänzungsstrich (einen Bindestrich an erster oder letzter Stelle) haben, werden wie alle anderen Token auch mit den Lexikoneinträgen verglichen. Da es keine Einträge gibt, die einen Ergänzungsstrich enthalten, werden alle Token in die Liste der nicht identifizierten Token aufgenommen.

In der Ausgabe der nicht identifizierten Token der Jahrgänge 04 und 05 befinden sich: 8081 Token, die mindestens einen Bindestrich enthalten; werden identische Token nur einmal gezählt, befinden sich 5111 Token in der Kollektion.⁴³

Zur Ermittlung der Verwendungsmöglichkeiten von Bindestrichen und deren Häufigkeit wurden 5 Dokumente mit jeweils 8 bis 11 DIN A4-Seiten auf das Vorkommen von Bindestrichen untersucht. Die Tabelle 5 beschreibt die 10 gefundenen Kategorien von Bindestrich-Verwendungen.

		Beispiele:	Anzahl	
I.	Bindestrich Verbindung	Avionik-Umgebung	229	
I.a	I. am Zeilenende	Avionik-Umgebung	50	
II.	Silbentrennung	Um-Gebung	753	
III.	Bindestrich Verbindung mit Sonderzeichen	Synchro-/Resolver Signale	1	
IV.	Bindestrich mit Leerzeichen	Produkte – zurückgegriffen	Insg.	
V.	Aufzählungen	- erstens... - zweitens...	13	
VI.	Bindestrich mit folgender Auslassung	Einweg- versus Mehrwegsystem Geräteein- und Ausbau	24	
VII.	Bindestrich mit vorgestellter Auslassung	Signalverteilung und -Filterung Programmabläufe- oder Programmergebnissen	1	
VIII.	VI. oder VII. mit mehreren Auslassungen	Rumpfoberschalen, -unterschalen oder -seitenschalen	In VI. und VII enthalten	
IX.	Bindestriche in Kombination mit Zahlen	-10, 10 ⁻²	Wurden nicht mitgezählt	
X.	Identifikationen, ISBN, Dokumentidentifikationen	IAC-05-B4.2.08	Sehr viele im Literaturverzeichnis	

Tab. 5: Bindestrich-Verwendungen

⁴³ Anhang B-3 Token mit Bindestrich.xls

Die Auswertung der einzelnen Fälle ergibt:

Fall I

Diese Token werden bereits von Lingo verarbeitet. Sie wurden nicht identifiziert, weil weder der Text vor und nach dem Bindestrich lexikalisiert ist. Veränderungen im Verarbeitungsablauf von Lingo können hier keine Abhilfe schaffen. Von den 5111 Token sind 2460 dem Fall I zuzuordnen und zählt man nur die Token, die mindestens zwei mal vorkommen, reduziert sich die Anzahl auf 219. Diese werden weiter unten getrennt betrachtet. Eine Anpassung der Lingo Verarbeitungsroutine ist nicht notwendig.

Fall Ia

Siehe Fall II

Fall II

In den Manuskripten befinden sich, bedingt durch die Zweispaltigkeit der Textgestaltung, relativ viele Bindestriche an den Zeilenenden. Diese sind den Fällen Ia oder II zuzuordnen. Diese Zuordnung ist selbst intellektuell zum Teil schwierig. Es ist nicht immer deutlich, ob es sich um ein durch die automatische Silbentrennung getrenntes Kompositum handelt oder um eine durch den Autor manuell getrennte Bindestrichverbindung. Bei der Umwandlung der Manuskripte vom PDF-Format in ein Text-Format werden die automatischen Silbentrennungen nicht einheitlich zurück genommen. Wie oben beschrieben, wurden die Manuskripte der Jahre 04 und 05 mithilfe des Acrobat Readers konvertiert. Dadurch wurde in den meisten Fällen am Zeilenende ein Zeilenumbruch eingefügt. Bindestriche am Zeilenende wurden dabei nicht einheitlich verarbeitet. Je nach Dokument bleiben die Bindestriche erhalten und es wird anschließend ein Zeilenumbruch eingefügt, oder es werden die Bindestriche entfernt, die beiden Textteile zusammengefügt und der Zeilenumbruch hinter dem zusammengefügten Wort angehängt. In einigen Fällen bleibt der Absatz erhalten und es wird nur am Ende des Absatzes ein Zeilenumbruch eingefügt.

Aber auch in diesem Fall werden die Bindestriche am Zeilenende nicht einheitlich behandelt. Eine richtige Bearbeitung der Fälle Ia und II kann nicht auf der im Text-Format vorliegenden Dateien erfolgen, da keine Muster vorhanden sind, anhand dessen die beiden Fälle zu unterscheiden sind. Um die Zeichen richtig in das Text-Format zu konvertieren, müssen je nach benutzter Software entsprechende Einstellungen bei der Erstellung der Dokumente, der Konvertierung in das PDF-Format und noch einmal bei der Konvertierung

in das Text-Format vorgenommen werden. Liegen die Daten in einer fehlerfreien Form vor, bearbeitet Lingo die Token bei denen die Silbentrennung zurückgenommen wurde, wie alle anderen Token ohne Bindestrich und die Token mit Bindestrich wie beschrieben. Somit werden keine Veränderungen in der Lingo Verarbeitungsroutine angestrebt.

Fall III

Token vom Typ „TEXT-/TEXT“ werden vom Tokenizer in drei Token zerlegt: <TEXT-> </> <TEXT>. Bei der Analyse der Token stellt dies damit ein Sonderfall des Falles VI dar und wird dort behandelt.

Fall IV

Bei den Textpassagen dieses Falls wird der Bindestrich meist zur Hervorhebung eines besonderen Textteils oder als Gedankenstrich benutzt. Lingo verarbeitet den Text vor und nach dem Gedankenstrich als separate Token. Da keine inhaltliche Beziehung zwischen den beiden Token besteht, ist keine Veränderung der Lingo Bearbeitung nötig.

Fall V

Hier wird der Bindestrich als Aufzählungszeichen benutzt und stellt keine Bedeutung für den inhaltlichen Zusammenhang der Token vor und nach dem Bindestrich dar. Dieser Fall wird nicht weiter betrachtet.

Fall VI, VII, Fall VIII

Der Fall VIII ist ein Sonderfall der Fälle VI und VII und wird nicht eigenständig sondern innerhalb der Fälle VI und VII betrachtet. Trifft Lingo bei der Bearbeitung der Token auf eines, das mit einem Ergänzungsstrich beginnt oder endet, so kann dieses Token aufgrund der fehlenden Lexikoneinträge nicht erkannt werden.

Fall IX

Token, die nur aus Zahlen und einem Bindestrich bestehen, werden nicht lexikalisiert und werden vom Lingo Nonworder nicht in die Datei der nicht identifizierten Tokens geschrieben. Der Fall IX wird nicht weiter betrachtet.

Fall X

Textstellen des Falles X befinden sich nur in den Literaturverzeichnissen der 5 betrachteten Dokumente. Werden andere Sonderzeichen als <-> oder <_> verwendet, zerlegt Lingo die Zeichenketten an den Stellen der Sonderzeichen in separate Token die dann, aufgrund der beliebigen Zusammenstellung von Text und Sonderzeichen, nicht mehr als Einheiten erkennbar sind. Bei allen Manuskripten wurde das Literaturverzeichnis bei der Datenkonvertierung entfernt. Fall X wird nicht weiter betrachtet.

Fall XI

Bei der Auswertung der Fälle I bis X wurde ein weiterer Fall der Bindestrichverwendung entdeckt. In seltenen Fällen werden vorangestellte- und folgende Bindestrichauslassungen kombiniert. Beispielsweise: Scheiben-Zug- und -Druckbelastung. Diese müsste in Scheiben-Zugbelastung und Scheiben-Druckbelastung aufgelöst werden. Eine Lösung dieses Falls wird als Sonderfall der Fälle VI und VII betrachtet.

Die Lingo Bearbeitung von Bindestrichen/Gedankenstrichen der Fälle I, II, IV und V ist ausreichend. Die Einbeziehung der Fälle XI und X in das Indexierungsergebnis ist nicht beabsichtigt. Verwendete Ergänzungsstriche werden von Lingo nicht berücksichtigt. Sollen die Fälle VI, VII und deren Sonderfälle III und VIII in das Indexierungsergebnis einbezogen werden, so muss die Funktionalität von Lingo erweitert werden. Eine mögliche Lösung für die Einbeziehung von Ergänzungsstrichen in die Indexierung wird in den folgenden Abschnitten dargestellt.

Für die Indexierung der Token des Falls I und der Erarbeitung einer Bearbeitungsroutine für die Fälle VI und VII wurden die Token des Jahrgangs 05 nach dem Schema in Abb. 1 selektiert.

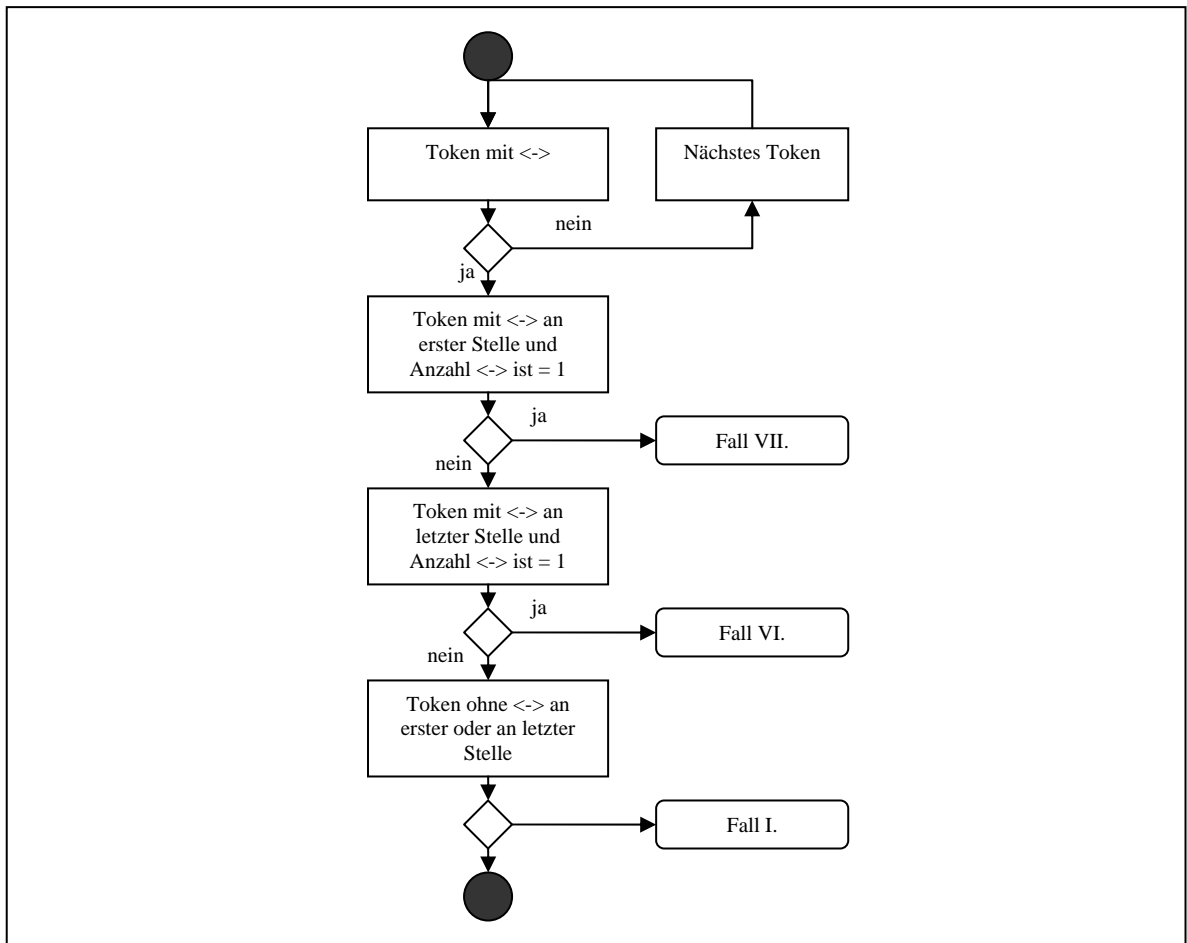


Abb. 1: Fallentscheidung von Binde- und Ergänzungsstrichen

Die Prozedur basiert auf der Annahme, dass bei einer seriellen Prüfung der Token alle die, welche mindestens einen Bindestrich enthalten, potentielle Kandidaten für den Fall I, VI oder VII sind.

Bei der Programmierung einer Routine, die die gewünschten Tokens auswählt, stellte sich die Frage, ob Token, die mehrere Bindestriche oder einen Ergänzungsstrich und zusätzlich einen oder mehrere Bindestriche enthalten, gesondert zu behandeln sind.

Fallzuordnung von Token mit mehreren Bindestrichen

Zu diesem Zweck wurden zwei Testmengen generiert: Eine mit Token, die Bindestriche und einen Ergänzungsstrich am Ende enthalten sowie eine, die Bindestriche und einen Ergänzungsstrich am Anfang des Token enthalten. Zur Analyse wurden beiden Testmengen die 8 folgenden bzw. vorangestellten Token hinzugefügt.

Für den Fall VI wurden 165 Token gefunden und untersucht. 146 Token bildeten mit dem darauf folgenden Token eine sinnvolle Zeichenkette, die dem Fall I zuzuordnen ist. Davon

ausgehend, dass die beiden Token zusammengehören, wurden diese als ein Token erneut einer Lingo Indexierung unterzogen und alle bis auf 5 wurden identifiziert. Von den 5 nicht identifizierten bestehen drei aus englischen Wörtern, eine ist eine Buchstaben-Zahlenkombination *<ti4-st6-st6>* und die letzte beinhaltet zwei Sonderzeichen. Lediglich 7 Token Kombinationen sind dem Fall VI zuzurechnen.⁴⁴

Für den Fall VII befinden sich 55 Token in der Testmenge. Lediglich 5 Token Kombinationen sind dem Fall VII zuzurechnen, und alle 5 kommen aus demselben Dokument. 31 Token sind Bindestrichverbindungen wie *<6-dof-analysen>*, bei denen der Tokenizer die führende Zahl abtrennt und als eigenes Token weitergibt.⁴⁵

Da nur 10 Bindestrichverbindungen mit mehreren Bindestrichen für die Fälle VI und VII gefunden wurden, werden diese bei der Lösungssuche ausgeschlossen und die Prozedur in Abb. 1 entsprechend angepasst.

4.2.6.1 Bindestrich-Verbindungen (Fall I)

Nach den vorherigen Erweiterungen der Lingo Systemlexika verbleiben in der Datei der nicht identifizierten Token 219 Bindestrichverbindungen, die mindestens in zwei Dokumenten vorkommen.⁴⁶ Bei 150 Token besteht ein Teil der Bindestrich-Verbindung aus einem oder zwei Zeichen. Lingo indexiert keine Bindestrich-Verbindungen, bei denen ein Teil der Verbindung aus weniger als drei Zeichen besteht. Es verbleiben 69 Token, die bei der Indexierung nicht erkannt wurden. Insgesamt werden von den 219 Token 29 lexikalisiert. Die Token, die aus englischen Wörtern bestehen, werden als Wortklasse *#g* und alle anderen als Wortklasse *#s* in das Nutzerwörterbuch aufgenommen. Alle Token werden als Bindestrich-Verbindung lexikalisiert, da es sich um Token handelt, bei denen ein Teil der Verbindung aus nur einem oder zwei Zeichen besteht (*<x-band>*) oder es sich um fest stehende Bezeichnungen handelt *<spalart-allmaras-modell>*.

4.2.6.2 Verwendung von Ergänzungsstrichen (Fall VI, Fall VII)

Alle Zahlen, Testmengen und Übersichten in diesem Abschnitt beziehen sich auf die Manuskripte des Jahrgangs 05. Ausgangsmenge für alle besprochenen Fälle ist die Ausgabe des Lingo Tokenizers (Rund 608000 Token).

⁴⁴ Anhang B-4 Mehrfach-.xls

⁴⁵ Anhang B-5 -Mehrfach.xls

⁴⁶ Anhang B-6 Fall I.xls

Lösungsansatz für die Identifizierung von Ergänzungsstrichen

Jeder Ergänzungsstrich verweist auf eine Zusammenfügung (mit oder ohne Bindestrich), die dem Ergänzungsstrich in einem nicht festgelegten Abstand von Token vorangestellt, ist oder folgt. Ziel ist es, diese Zusammenfügung zu finden, an der richtigen Stelle zu trennen und dem Token mit dem Ergänzungsstrich anzufügen oder voranzustellen. Hat die Zusammenfügung ein Bindestrich, so wird sie indexiert und am Bindestrich getrennt. Der abgetrennte Teil wird mit dem Token mit dem Ergänzungsstrich zusammengefügt und ebenfalls indexiert. Enthält die Zusammenfügung keinen Bindestrich, so wird der erste Teil von rechts, der lexikalisiert ist, abgetrennt. Wieder wird die Zusammenfügung in den Index übernommen und im Fall VI wird dem Token mit dem Ergänzungsstrich dieser entfernt und mit dem rechten Teil der Zusammenfügung indexiert. Für den Fall VII wird der linke Teil der Zusammenfügung mit dem Token mit dem Ergänzungsstrich zusammen indexiert. Zwei Beispiele: Angenommen die Textsstelle ‚Studien- und Diplomarbeiten‘ wird vom Tokenizer in drei Token zerlegt: <studien-> <und> <Diplomarbeiten>. Das Token <studien-> wird als Fall VI identifiziert und das Token <diplomarbeit> als zugehörige Zusammenfügung. Im Lexikon ist Diplomarbeit nicht enthalten sondern nur <diplom> und <arbeiten>. Das System muss Diplomarbeit als Zusammenfügung erkennen und trennen. In den Index wird <diplomarbeit> und <studienarbeit> aufgenommen.

Die Textstelle ‚Fehlstellenfläche und -breite‘ wird nach dem gleichen Schema zugeordnet. Im Lexikon befindet sich <fehl> <stelle> und <fläche>. Abgetrennt wird der erste erkannte Wortteil von rechts also <fläche>. Indexiert wird Fehlstellenbreite und Fehlstellenfläche.

Mit jedem Token, mit dem die Zusammenfügung weiter von dem Ergänzungsstrich entfernt ist, steigt die Komplexität einer möglichen Lösung. In den folgenden Abschnitten werden die gefundenen Ergänzungsstriche nach Typ und Häufigkeit analysiert und überprüft, inwieweit der beschriebene Lösungsansatz erfolgreich ist.

Fall VI <Text->

Im Jahrgang 05 befinden sich 3156 Token, die dem Fall VI wie oben definiert entsprechen.⁴⁷ Tab. 6 zeigt die Anzahl der Token-Kombinationen in Abhängigkeit von dem Token, das dem Token mit dem Ergänzungsstrich folgt.

⁴⁷ Anhang B-7 FallVI.xls

Token 1	Token 2	Anzahl
<text->	<und>	2310
<text->	<,>	110
<text->	</>	56
<text->	<bzw>	68
<text->	<oder>	76

Tab. 6: Anzahl der Token für die Kombinationen des Fall VI

Andere sinnvolle Token-Kombinationen, beispielsweise mit <als> oder <zu> sind so selten, dass sie nicht betrachtet werden.

Fall VI, <und> als Token 2

Die Kombination von Ergänzungsstrich (Token 1) und <und> als Token 2 ist mit 2309 die häufigste Variante. Es befinden sich 2233 sinnvolle Kombinationen in der Testmenge. Diese lassen sich in 18 Kategorien einteilen.⁴⁸

Für die praktische Umsetzung wurden nur die ersten beiden Varianten zu einer zusammengefasst (Tab. 7).

Lösung	Token1	Token2	Token3	Findet
1.	<text->	<und>	<kompositum> <text-text>	2003

Tab. 7: Lösung Fall VI mit <und> als Token 2

Abb. 2 zeigt die logische Umsetzung der Lösung.

Wird die Testmenge nach dem Schema aus Abb. 2 untersucht stellt man fest, dass 4 nicht sinnvolle Kombinationen indexiert werden. Die Indexierung erfolgt aufgrund von Fehlern im Token 1, bspw. <-litäts-> <und> <flugeigenschaften>. Hier werden *-litäts-* und *flugeigenschaften* in den Index übernommen. Das Schema orientiert sich an den Regeln für Bindestrichverbindungen. Wenn das dritte Token als Kompositum erkannt wird, ist es nicht notwendig, zusätzlich das Token 1 über das Lexikon zu identifizieren.

⁴⁸ Anhang B-7-1 FallVI und.xls; Tabellenblatt: Tok2 Und Kombinationen

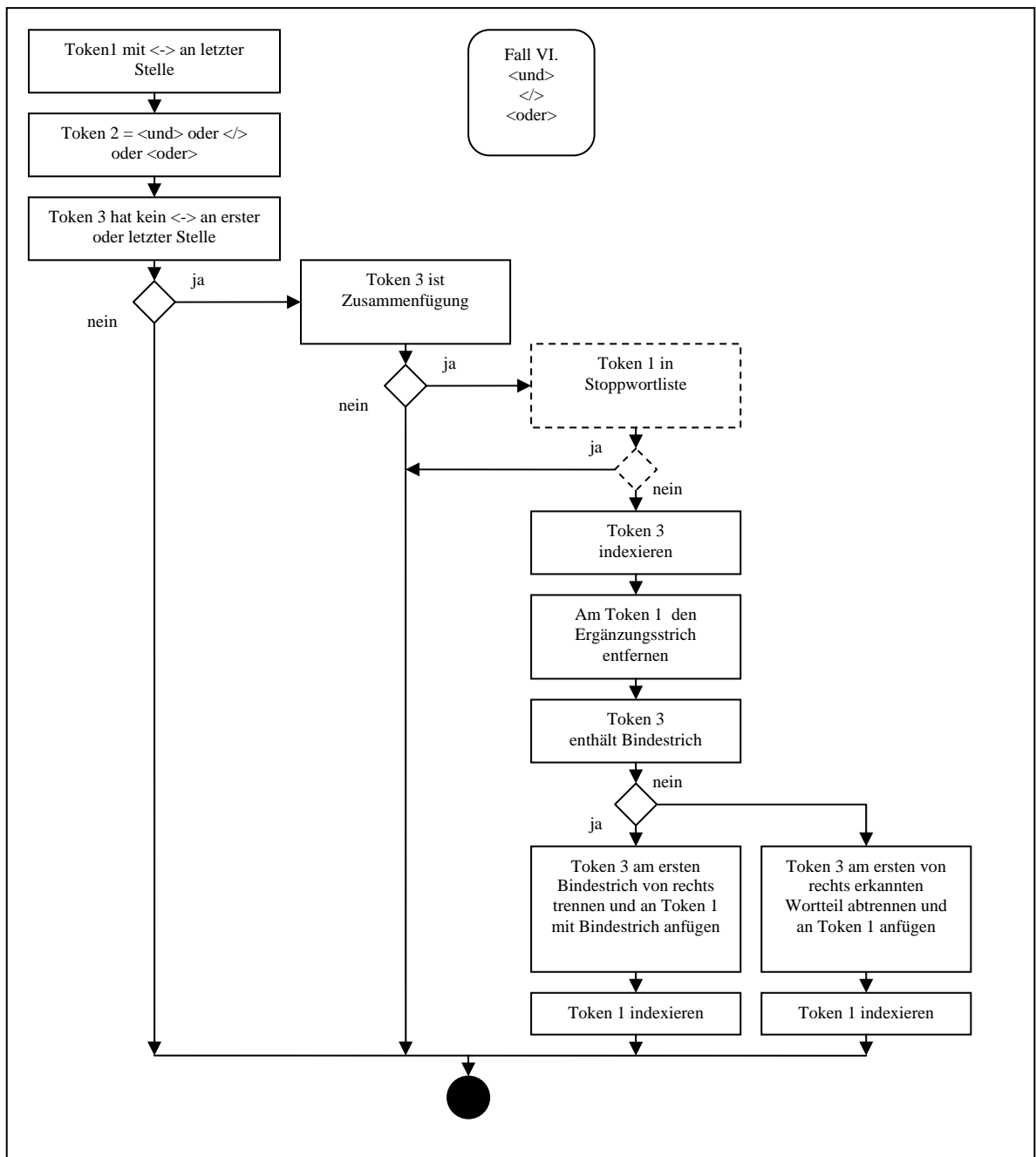


Abb. 2: Lösungsschema Fall VI, mit *<und>* als Token 2

Von den 2233 sinnvollen Kombinationen sind 2053 dem Schema in Abb. 2 zuzuordnen. Von diesen werden 998 richtig und 1055 falsch indexiert. Die falsche Indexierung hat im Wesentlichen zwei Gründe:

- Der dem Token 1 zuzuordnende Wortteil der Zusammenfügung ist als eigenständiger Teil erkannt worden, befindet sich aber nicht am Ende der Zusammenfügung. Bei

<kohle-> <und> <aramidfaserverstärktem>

wird *<kohle>* dem *<verstärkt>* und nicht *<faserverstärkt>* zugeordnet weil dieses nicht am Ende der Zusammenfügung steht. Die Wahrscheinlichkeit einer falschen Zuordnung steigt mit der Anzahl der Bestandteile des 3. Tokens. Die Häufigkeit des Auftretens wirkt dem aber entgegen. Es ist nur eine falsche Indexierung in der Testmenge, bei der das Token 3 aus 5 Bestandteilen besteht und bei 4 Bestandteilen sind es 7. Eine Einschränkung der Bearbeitung auf Zusammenfügungen, die nur aus 2 oder 3 Bestandteilen bestehen, wird als nicht notwendig erachtet.

- Der dem Token 1 zuzuordnende Wortteil wird durch das Longest-Matching von einem längeren Lexikoneintrag überdeckt.

<luft-> <und> <raumfahrttechnik>

Hier wird *<raumfahrttechnik>* in *<raumfahrt>* und *<technik>* zerlegt und *<luft->* der *<technik>* zugeordnet. Zudem befindet sich *<fahrt>*, selbst wenn es als eigener Bestandteil gefunden würde, nicht an der letzten Stelle der Zusammenfügung, so dass die Entfernung von Raumfahrt aus dem Lexikon keine Abhilfe schafft. Auch eine Lexikalisierung von *fahrttechnik #s* ist nicht sinnvoll. Die Einführung eines Algorithmus für die Suche und die Erkennung des richtigen Wortbestandteils der Zusammenfügung in Token 3 wird aufgrund der zu erwartenden Komplexität nicht betrachtet.

Zu beachten ist, dass sich in den 1055 falsch indexierten Kombination 863 mal die Token-Kombination *<luft-> <und> <raumfahrtkongress>* befindet. Entfernt man doppelte Nennungen so verbleiben 73 falsche Kombinationen.⁴⁹ Zur Reduzierung der Fehler wurden diese weiter untersucht. Betrachtet man die Token 1 fällt auf, dass die Präfixe *an-* und *ein-* häufig in den falsch indexierten Kombinationen vorkommen. Der Grund liegt in der Lexikalisierung von Wörtern, den *<an>* oder *<ein>* vorangestellt ist. Auch hier ist die Parametereinstellung für die Kompositumserkennung für die fehlerhafte Indexierung verantwortlich. Das führt dazu, dass alle Zusammenfügungen, die einen Bestandteil mit nur 2 Zeichen enthalten, nicht erkannt werden. Aus diesem Grund müssen wie schon beschrieben Zusammenfügungen wie *Ölfass* als ein Wort lexikalisiert werden. Ein Beispiel aus der Fehlermenge ist *<an-> <und> <abflug>*. Diese Kombination kann nicht erkannt werden da *abflug #s* als Substantiv lexikalisiert ist und nicht als Kompositum erkannt wird.

⁴⁹ Anhang B-7-1 FallVI und.xls, Tabellenblatt: VI-Und Fehler

Diese Praxis führt dazu, dass alle Ergänzungsstriche mit bspw. *an-* oder *ab* nicht richtig zugeordnet werden können, weil entweder die folgende Zusammenfügung als Ganzes lexikalisiert wurde oder an die Zusammenfügung ein weiterer Wortteil angehängt wurde und der Teil, auf den sich der Ergänzungsstrich bezieht, im linken Teil des zerlegten Kompositums befindet. In der Fehlermenge trifft diese Problematik für *ab-*, *an-*, *aus-*, *be-* und *ein-* zu. Zusätzlich wird *<luft->* oft mit *<raumfahrt>* als Teil von Komposita benutzt; auch bei *<kurz->* und *<ziel>* überwiegt die fehlerhafte Erkennung.⁵⁰

Für die Testmenge ist die Einführung einer Stoppwortfunktion eine Lösung zur Reduzierung der Fehlerhäufigkeit. Alle Kandidaten der Stoppwortliste haben eine Fehlerquote von 50% oder höher. Um die Stoppwortfunktion flexibel zu gestalten, könnte diese über eine Liste in der Konfigurationsdatei von Lingo aktiviert und mit den gewünschten Stoppwörtern gefüllt werden. Wird die Liste von Stoppwörtern auf die Testmenge angewendet, entsprechen 1042 Token Kombinationen den Kriterien und werden bearbeitet. 47 werden falsch indexiert, was einer Quote von 4,51 % entspricht. Als Alternative zur Stoppwortliste ist es möglich, mit einer konsequenten Lexikalisierung der falschen Kombinationen mit *<luft->* und *<raumfahrt...>* diese aus der Bearbeitung auszuschließen. Notwendig ist die Lexikalisierung der 15 Lexikoneinträge aus Tab. 8.

<i>raumfahrt</i> Zusammensetzungen
Raumfahrtabsolventen
Raumfahrt-Absolventen
Raumfahrtaustellung
Raumfahrtbereich
Raumfahrtbranche
Raumfahrttechnik
Raumfahrtforscher
Raumfahrtindustrie
Raumfahrtkongress
Raumfahrtnutzung
raumfahrtorientierter
Raumfahrtstrukturen
Raumfahrtsysteme
Raumfahrttechnik
Raumfahrttechnologien

Tab. 8: Fall VI, Zusammensetzungen mit *raumfahrt*

So wird eine Fehlerquote von 8,2 % erreicht.

⁵⁰ Anhang B-7-1 FallVI und.xls, Tabellenblatt: VI-Und Stoppwort

Fall VI, mit Komma als Token 2

Folgt dem Ergänzungsstrich (Token 1) ein Komma, so wird damit eine Aufzählung eingeleitet, die sich über weitere Token erstreckt. Für diesen Fall wurden bis zu 6 folgende Token berücksichtigt. In den 110 Token Kombinationen befinden sich 75 sinnvolle Kombinationen. Diese lassen sich in 16 Kategorien einteilen.⁵¹ Diese 16 Kategorien wurden zu zwei möglichen Lösungen für die Identifizierung zusammengefasst:

Lösung	Token1	Token2	Token3	Token4	Token 5	findet
1.	<text->	<,>	<text->	<und> <oder> <,>	<kompositum> <text-text>	61
2.	<text->	<,>	<kompositum> <text-text>			4

Tab. 9: Lösungen Fall VI mit Komma als Token 2

Werden die beiden Lösungen auf die Testmenge angewendet, so werden 65 der 75 sinnvollen Token Kombinationen verarbeitet. Abb. 3 zeigt den logischen Ablauf der kombinierten Lösung.

⁵¹ Anhang B-7-2 FallVI Komma.xls, Tabellenblatt: Tok2 Komma Kombinationen

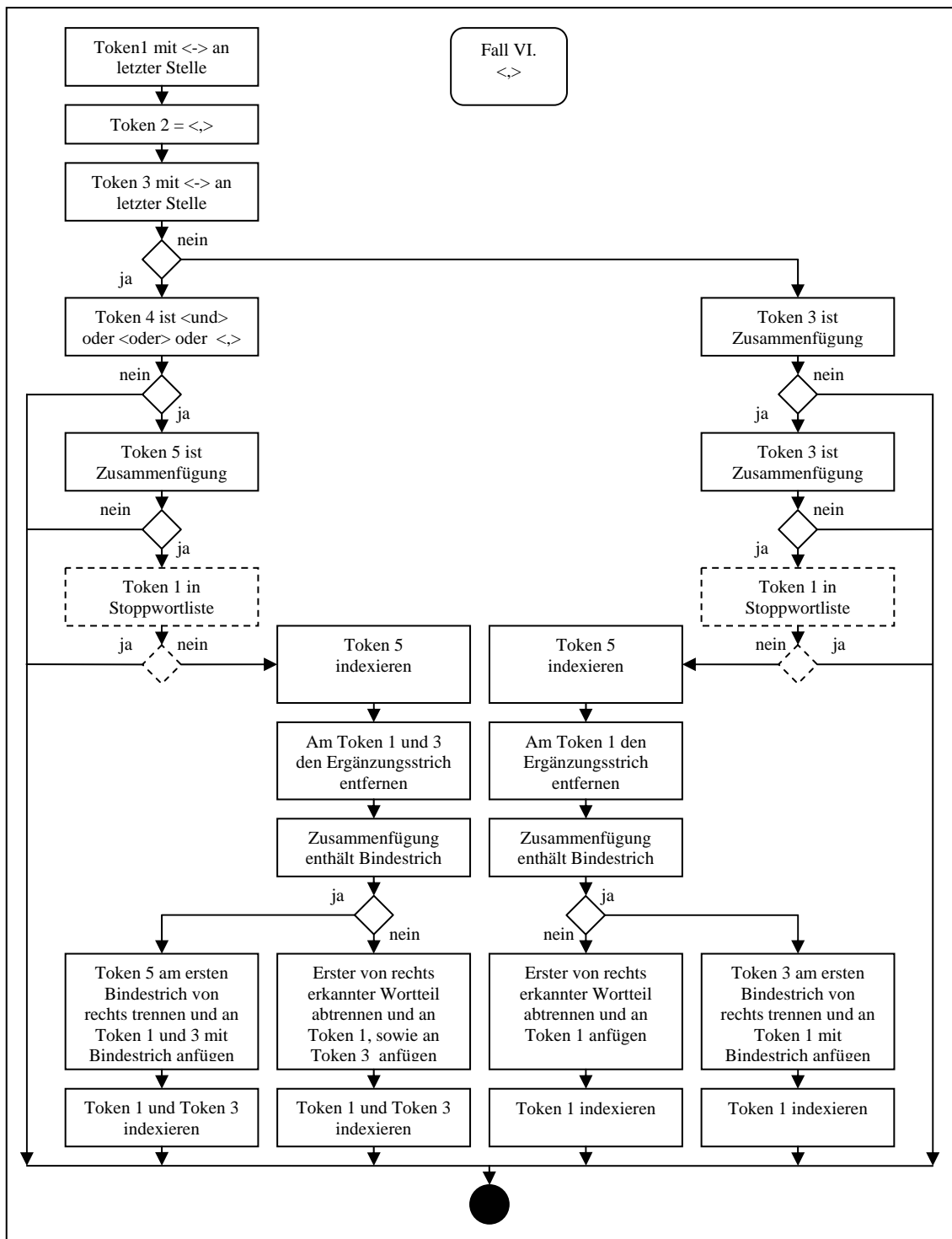


Abb. 3: Lösungsschema Fall VI, mit Komma als Token 2

Es werden 61 von dem linken Strang und 4 von dem rechten bearbeitet. Die Quote für die Fehlindexierung liegt bei 6.6 %, nachdem 2 Fehlindexierungen von der Stoppwortliste gestrichen wurden. *<seiten->* und *<querruderausschläge>* würden als *seitenausschläge* und *querruderausschläge* und nicht als *seitenruderausschläge* indexiert.

Fall VI, </> als Token 2

Diese Kombinationen entsprechen den im Abschnitt 4.2.6 beschriebenen Fall III. Es befinden sich 54 sinnvolle Kombinationen in der Testmenge.⁵² Sie kommen in 4 verschiedenen Zusammensetzungen vor, von denen die beiden häufigsten zu einer Lösung zusammengefasst wurden.⁵³

Lösung	Token1	Token2	Token3	Findet
1.	<text->	</>	<kompositum> <text-text>	50

Tab. 10: Lösung Fall VI mit Schrägstrich als Token 2

Das logische Schema zur Bearbeitung dieser Kategorie ist mit der von Abb. 2 identisch. Mit der Lösung werden 48 der 54 sinnvollen Kombinationen gefunden. Auch hier könnten 3 falsche Einträge mit <an-> als Token 1 über die Stoppwortfunktion entfernt werden. Zusätzlich fällt auf, dass auch <auf-> als Token 1 eine Fehlerquote von 100 % hat und somit ein Kandidat für die Stoppwortliste ist. <auf> ist weder in der Menge mit <und> noch mit <,> als Token 2 enthalten.

Fall VI, <bzw> als Token 2

Für diese Kategorie befinden sich 57 sinnvolle Kombinationen in der Testmenge.⁵⁴ Sie gruppieren sich in 8 unterschiedliche Kombinationsmöglichkeiten.⁵⁵ Davon werden zwei ausgewählt und zur Indexierung zusammengefasst.

Lösung	Token1	Token2	Token3	Token 4	Findet
1.	<text->	<bzw>	<.>	<kompositum> <text-text>	45

Tab. 11: Lösung Fall VI mit <bzw> als Token 2

Die Abb. 4 zeigt das logische Schema für diese Kategorie. Auch dieses Schema hat das von Abb. 2 als Grundlage hier wurde der <.> als Token 3 eingefügt.

⁵² Anhang B-7-3 FallVI Schrägstrich.xls, Tabellenblatt: VI-Schrägstrich Alle

⁵³ Anhang B-7-3 FallVI Schrägstrich.xls, Tabellenblatt: Tok2 Schrägstrich Kombinationen

⁵⁴ Anhang B-7-4 FallVI bzw.xls, Tabellenblatt: VI-bzw Alle

⁵⁵ Anhang B-7-4 FallVI bzw.xls, Tabellenblatt: Tok2 bzw Kombinationen

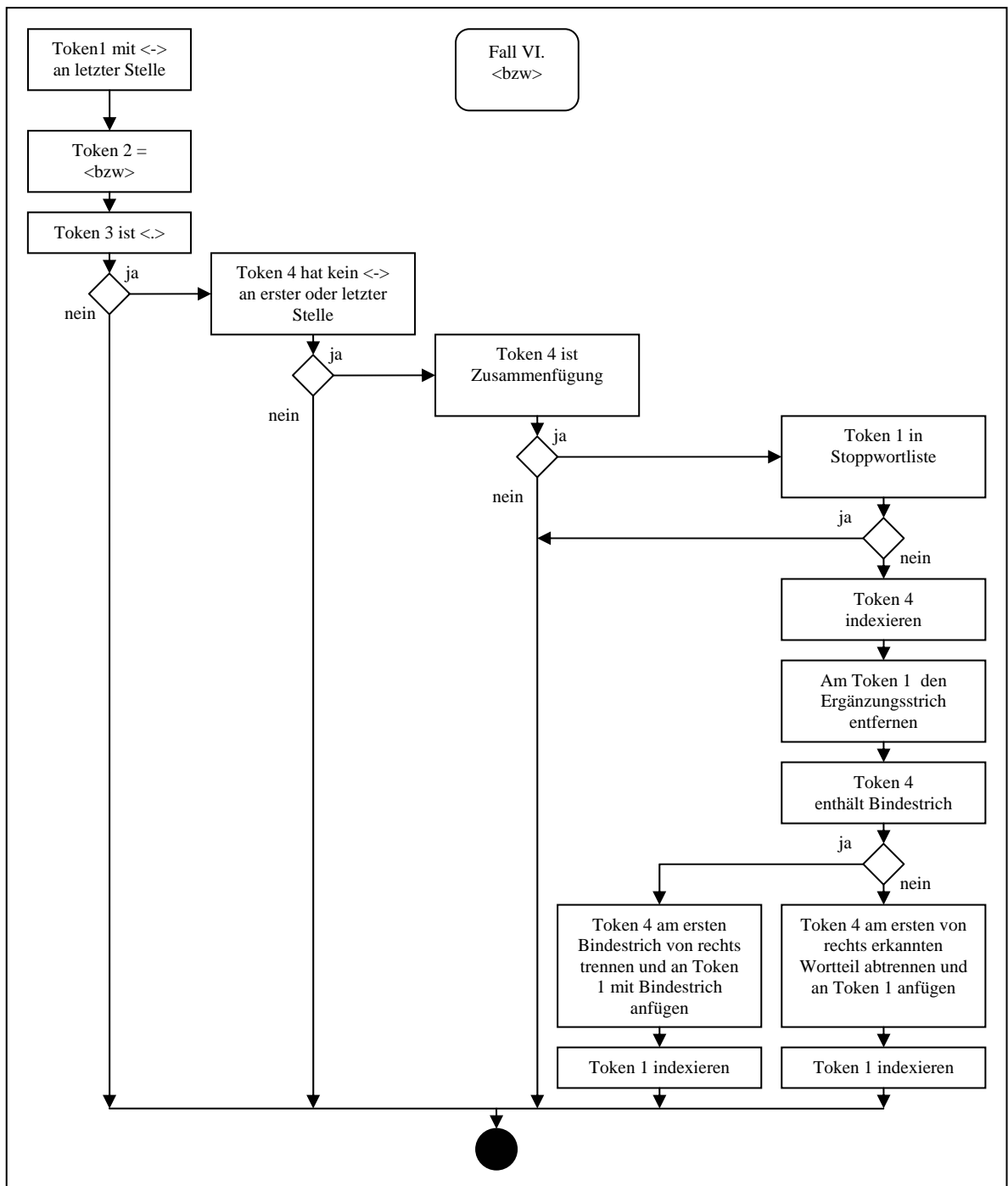


Abb. 4: Lösungsschema Fall VI, mit <bzw> als Token 2

Von den erkannten 45 Kombinationen werden 39 richtig indexiert. Entfernt man die Einträge mit Stoppwörtern, beträgt die Fehlerquote 9,3%.⁵⁶

Fall VI, <oder> als Token 2

In dieser Kategorie ergeben 75 der 76 Kombinationen einen Sinn.⁵⁷ Sie teilen sich in 9 Kategorien auf, von denen die ersten beiden zur Indexierung zusammengefasst werden.⁵⁸

⁵⁶ Anhang B-7-3 FallVI Schrägstrich.xls, Tabellenblatt: VI-bzw Alle

Lösung	Token1	Token2	Token3	Findet
1.	<text->	<oder>	<kompositum> <text-text>	68

Tab. 12: Fall VI, <oder> als Token 2

Das Lösungsschema entspricht der Abb. 2. Das Schema indexiert 59 Token-Kombinationen, von denen ohne Stoppwortliste 8 falsch indexiert werden. 5 falsche Indexierungen werden durch die Stoppwortliste verhindert, was zu einer Fehlerquote von 5,5 % führt.

Fall VI Zusammenfassung

Für alle 5 Varianten des Tokens 2 konnten praktikable Lösungen gefunden werden. Für 3 (<und>, </> und <oder>) ist die Lösung identisch und relativ leicht realisierbar. Die Lösung mit <bzw> als Token 2 entspricht vom Ablauf auch der ersten Lösung, es muss aber der folgende <.> als eigenes Token berücksichtigt werden. Lediglich die Version mit einem Komma als Token 2 ist aufwändiger. Es handelt sich hier um Aufzählungen, die meistens 2 Token mehr einbeziehen als die erste Lösung. Tab. 13 zeigt die Zusammenfassung der Indexierungsergebnisse für alle 5 Varianten des Falles VI.

ohne Stoppwortliste			
	indexiert	Falsch	Fehlerquote
<und>	2003	992	49,53%
<,>	63	6	9,52%
</>	50	5	10,00%
<bzw>	45	6	13,33%
<oder>	59	8	13,56%
insgesamt	2220	1017	45,81%

Tab. 13: Fall VI, Indexierungsergebnisse ohne Stoppwortliste

Die hohe Fehlerquote für <und> als Token 2 ist beschrieben worden.

Es wurden zwei Wege dargestellt, die eine Reduzierung der Fehlerquote ermöglichen. Tab. 14 zeigt die Fehlerquoten bei Verwendung der Stoppwortliste und mit der Lexikalisierung der Terme aus Tab. 8.

⁵⁷ Anhang B-7-5 FallVI oder.xls, Tabellenblatt: VI-Oder Alle

⁵⁸ Anhang B-7-5 FallVI oder.xls, Tabellenblatt: VI-Oder Kombi

Ohne Stopwortliste (Mit Lexikalisierung der Terme aus Tab. 8)				mit Stopwortliste		
	indexiert	falsch	Fehlerquote	indexiert	falsch	Fehlerquote
<und>	1097	90	8,20%	1042	47	4,51%
<,>	63	6	9,52%	61	4	6,56%
</>	50	5	10,00%	48	3	6,25%
<bzw>	45	6	13,33%	43	4	9,30%
<oder>	59	8	13,56%	54	3	5,56%
insgesamt	1314	115	8,75%	1248	61	4,89%

Tab. 14: Fall VI, Indexierungsergebnisse

Für die Testmenge ist die Verwendung der Stopwortliste effektiver. Insgesamt werden durch die Stopwortliste 10 richtige und 62 falsche Einträge verhindert.⁵⁹

Fall VII <-text>

In der Testmenge befinden sich 502 Token, die einen Ergänzungsstrich an erster Stelle des Tokens haben.⁶⁰ Tab. 15 zeigt die Anzahl der Token Kombinationen in Abhängigkeit von dem Token, das dem Token mit dem Ergänzungsstrich vorangestellt ist.

Token 1	Token -1	Anzahl
<-text>	<.>	100
<-text>	<und>	91
<-text>	<,>	11

Tab. 15: Anzahl Token Kombinationen für den Fall VII

Es befinden sich 174 Kombinationen in der Testmenge, bei denen das dem Ergänzungsstrich vorangestellte Token aus einer oder mehrerer Zahlen besteht, bspw. <6> <-dof>. Auf diese Problematik wurde bereits oben eingegangen und wird hier nicht behandelt. Der dritte Fall mit dem Token <,> wird aufgrund der geringen Anzahl ebenfalls nicht berücksichtigt.

Fall VII, <.> als vorangestelltes Token

Es sind 89 sinnvolle Kombinationen vorhanden. Diese treten in 6 Token-Kombinationen auf, von denen nur 2 eine Häufigkeit größer 10 haben.⁶¹ Die Kombination <dok> <.> <ing> 46 mal und <dipl> <.> <ing> 35 mal. Beide sind Abkürzungen, die auf die Titel von Autoren hinweisen. Da die Autoren von der Indexierung ausgeschlossen wurden,

⁵⁹ Anhang B-7-6 FallVI Stopwortliste.xls

⁶⁰ Anhang B-8 FallVII.xls

⁶¹ Anhang B-8-1 FallVII Punkt.xls, Tabellenblatt Tok -1 Punkt Kombinationen

werden auch deren Titel nicht indiziert. Somit wird keine Variante mit <.> als vorangestelltes Token in die Lösung aufgenommen.

Fall VII, <und> als vorangestelltes Token

Hier ergeben 5 Tokenvarianten 87 sinnvolle Kombinationen.⁶² Die zwei häufigsten Kombinationen wurden zu einer Lösung zusammengefasst (Tab. 16).

Lösung	Token1	Token -1	Token -2	Findet
1.	<-text>	<und>	<kompositum> <text-text>	68

Tab. 16: Fall VII mit <und> als vorangestelltes Token

Somit hat der Fall VII nur eine Lösung, die in Abb. 5 dargestellt ist.

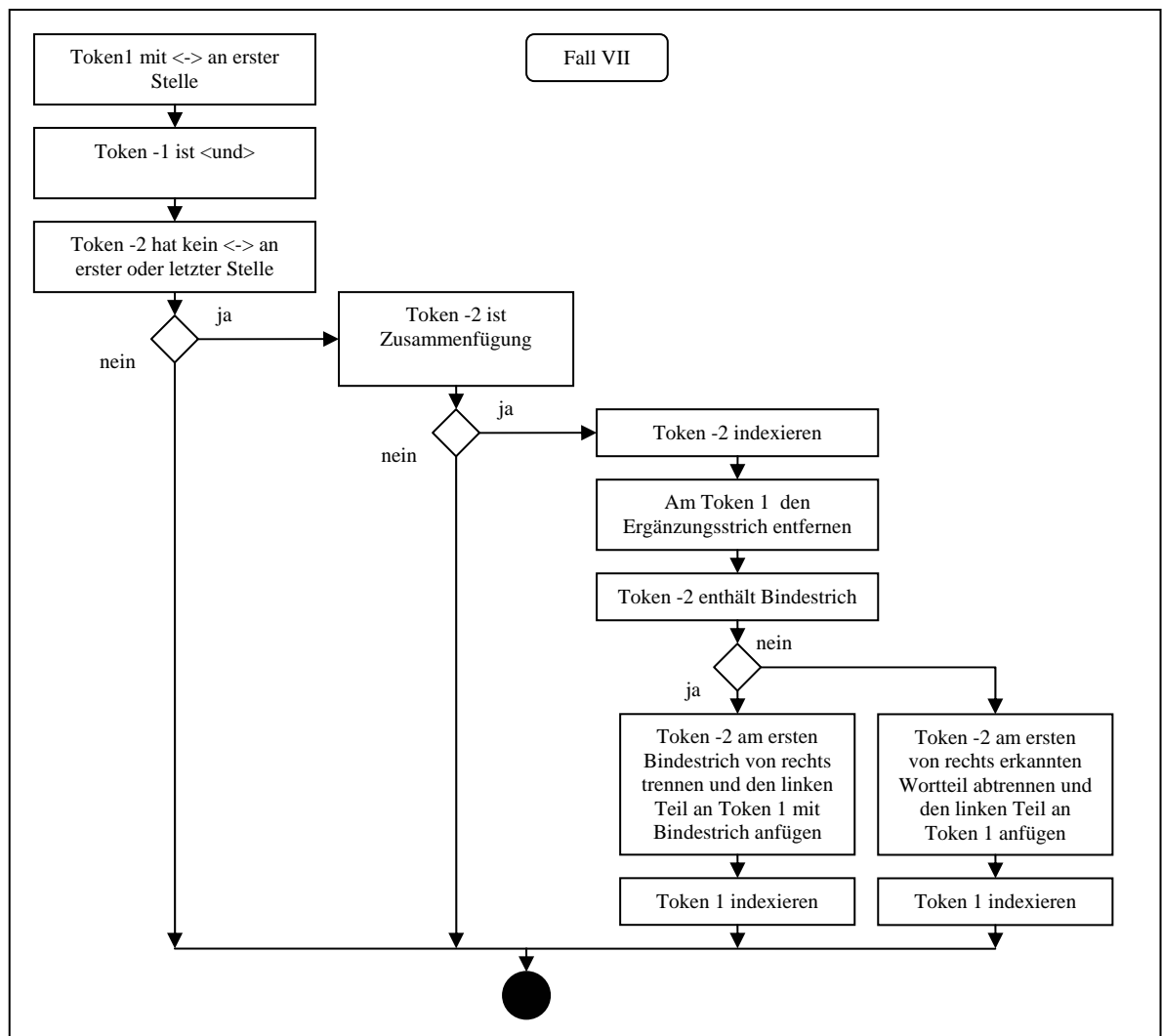


Abb. 5: Lösungsschema Fall VII

⁶² Anhang B-8-2 FallVII und.xls, Tabellenblatt Tok -1 und Kombinationen

Mit diesem Schema werden 77 der 87 sinnvollen Kombinationen bearbeitet. Nur 2 werden falsch indexiert. Die Fehlerquote beträgt 2,6 %.

Die Verwendung einer Stoppwortliste ist hier nicht notwendig, da in diesem Fall der linke Teil der Zusammenfügung als Ganzes dem Token mit dem Ergänzungsstrich zugeordnet wird.

Der Abschnitt zeigt, dass die Implementierung einer Ergänzungsstrichverarbeitung in Lingo möglich ist. Eine Stoppwortliste ist nicht zwingend notwendig und kann durch die Beachtung von Regeln bei der Lexikalisierung von Termen vermieden werden. Das zeigte die Lexikalisierung der 15 Wortkombinationen mit *raumfahrt* (Tab. 8). So konnte durch den beschriebenen Lösungsansatz für die Ergänzungsstriche (Fall VI und Fall VII) eine Indexierung von 1274 Token bei nur 117 falschen Indexierungen, ohne die Verwendung einer Stoppwortliste, erreicht werden.

4.2.7 Lexikalisieren von Mehrwortgruppen

Mehrwortgruppen bestehen aus mehreren Wörtern, die zusammen eine semantische Einheit bilden. Bei Mehrwortgruppen werden Substantive meist durch Adjektive in ihrer Bedeutung präzisiert: *kinetische Energie, feuerhemmender Baustoff*.⁶³ Dabei wird in technischen Fachsprachen im Wesentlichen ein Vorgang, eine Eigenschaft oder Gegenstand durch die Adjektive näher bestimmt.

Lingo als lexikonbasiertes Indexierungssystem benötigt für die Mehrworterkennung Lexikoneinträge von Mehrwortgruppen. Da Mehrwortgruppen eine semantische Einheit bilden wird ein gewisses Maß an Domainwissen benötigt um die semantischen Zusammenhänge der einzelnen Token zu beurteilen. Dieser Abschnitt beschränkt sich auf die Lexikalisierung der in den Manuskripten befindlichen Mehrwortgruppen. Dabei liegt das Hauptaugenmerk auf die Identifizierung von Unterscheidungsmerkmalen für die durch das Substantiv oder Kompositum bestimmten Vorgänge, Gegenstände oder Eigenschaften. Ein Beispiel: die Mehrwortgruppe *Erstes Faradaysches Gesetz* wird im Index durch die drei Einträge *erstes*, *faradaysches* und *gesetz* vertreten, sofern alle drei Bestandteile lexikalisiert sind. Bei der Suche nach einem der drei Wörter würde das Dokument aus dem der Eintrag stammt in die Treffermenge aufgenommen. Die Lexikalisierung der Mehrwortgruppe und die Erzeugung eines Indexeintrags aus allen drei Wörtern (*erstes*

⁶³ Vgl. Reinhardt, Köhler, Neubert, 1992, S. 18

faradaysches gesetz) würde nur dann zu einem Sucherfolg führen, wenn die Mehrwortgruppe als Ganzes gesucht würde. Dem Suchenden helfen indexierte Mehrwortgruppen erst dann, wenn diese für Suchenden einsehbar sind. Findet der Suchende mit dem Sucheinstieg *gesetz* Indexeinträge wie:

...
gesetz, coulomb
gesetz, faraday erstes
gesetz, faraday zweites
gesetz, newton
...

bietet der Index eine Übersicht über die in den Dokumenten enthaltenen Wörter und erleichtert die Präzisierung der Suchanfrage.

Mit Hilfe der algorithmischen Mehrworterkennung wurden zwei Listen erzeugt. Die erste Liste selektierte Tokenkombinationen nach dem Muster Adjektiv, Adjektiv Substantiv/Kompositum, die zweite Liste nach dem Muster Adjektiv Substantiv/Kompositum. Die zweite Liste wurde gezielt nach Adjektiven durchsucht, die das Substantiv/Kompositum differenziert zu anderen, in der Liste befindlichen Adjektiven darstellt. So wurde bspw. *aktuator* einmal mit dem Adjektiv *hydraulisch* und einmal mit *elektrisch* als Mehrwortgruppe lexikalisiert. In der ersten Liste wurde analog dazu nach einer weiteren Differenzierung von *aktuator, hydraulisch* gesucht. Obwohl die erste Liste 3961 Einträge enthält konnten keine Einträge gefunden werden, die eine weitere inhaltliche Differenzierung durch das zweite Adjektiv enthielten. In den meisten Fällen wurde das zweite Adjektiv zur Verstärkung oder Relativierung des ersten Adjektivs benutzt. Die Adjektiv Substantiv Kombination *thermische bedingung* kommt insgesamt 5-mal vor. Es wurde mit den Adjektiven: *real, typisch, gut* und *stationär* verwendet, die allerdings keine weitere Differenzierung bedeuten. Ein weiteres Beispiel ist die Substantiv Adjektiv Kombination: *experimentell daten* die mit den Adjektiven *eigen, belastbar, existierend, zugänglich* und *verfügbar* kombiniert wurden und ebenfalls keine weitere Differenzierung ermöglichen. In der Liste konnten keine für die Suche hilfreiche Differenzierung der Adjektiv-Substantiv-Kombinationen gefunden werden.⁶⁴ Eine mögliche Ursache hierfür könnte ein unzureichender Umfang der Datenbasis sein. Die zweite Liste wurde nach unterscheidenden Adjektiven untersucht und insgesamt 572 Einträge für das

⁶⁴ Anhang B-9-1 Mehrwortgruppen AdjAdjSub.xls

Mehrwortwörterbuch ausgewählt.⁶⁵ So sind bspw. für das Substantiv *wirkungsgrad* die Mehrwortgruppen

wirkungsgrad, dynamisch

wirkungsgrad, elektrisch

wirkungsgrad, mechanisch

wirkungsgrad, statisch

wirkungsgrad, thermisch

ausgewählt worden. Auch hier fällt auf, dass thematisch unterscheidende Adjektive oft nur bei sehr häufig auftretenden unspezifischen Substantiven (bspw. *antrieb, daten, dehnung* und *feld*) gefunden wurden. Für die Lexikalisierung von Mehrwortgruppen ist zusätzlich zu beachten, dass alle Wörter die in das Mehrwortwörterbuch aufgenommen werden, einer lexikalisierten Wortklasse angehören. Komposita, die bei dem Indexierungsvorgang als Kompositum mit der Klasse *k* markiert werden, sind nicht als Ganzes lexikalisiert und werden bei der Mehrworterkennung nicht berücksichtigt. Der Eintrag *wirkungsgrad* aus dem letzten Beispiel ist kein Lexikoneintrag sondern setzt sich aus *wirkung #s* und *grad #s* als Kompositum zusammen. Für die Mehrworterkennung muss *wirkungsgrad #s* lexikalisiert werden. Das hat zur Folge, dass alle Mehrwortgruppen vor dem Eintrag in das Lexikon auf nicht lexikalisierte Worte überprüft werden müssen. Für die 572 Einträge in das Mehrwortwörterbuch sind 40 Einträge in das Nutzerwörterbuch notwendig.

4.2.8 Indexierung von Eigennamen, Akronymen und Abkürzungen

Unter der Bezeichnung Eigennamen werden hier Namen von Personen, Firmen und Gegenständen sowie deren Abkürzungen und Akronyme verstanden. Für die Bearbeitung der Eigennamen werden die Listen der Eigennamen und Abkürzungen/Akronyme aus der Standard Indexierung (Tab. 1) zusammengefasst. Die 675 Token werden nach Firmenbezeichnungen, Projektnamen, Namen von Gegenständen und Persönlichkeiten untersucht. Nicht berücksichtigt werden Autorennamen, die bereits gesondert in einer Datenbank für die Suche zur Verfügung stehen, Abkürzungen von physikalischen Einheiten, die in großer Zahl aus den im Text befindlichen Tabellen stammen und Ortsnamen.

⁶⁵ Anhang B-9-2 Mehrwortgruppen AdjSub.xls

Eigennamen von Personen

In Lingo werden Eigennamen, die aus einem Wort bestehen in das Systemwörterbuch mit der Wortklasse *#e* eingetragen. Die Suffixliste der Wortklasse *#e* hat nur ein *-s* als Eintrag. Für die Indexierung von Eigennamen ist ein Eintrag wie: *einstein=einstein #e* notwendig, der die Erkennung der Token *<einstein>* und über die Suffixliste *<einsteins>* ermöglicht. Soll nicht nur Einstein sondern auch Albert identifiziert werden muss, zusätzlich *albert=albert #e* in das Lexikon aufgenommen werden. Die beiden Lexikoneinträge sorgen dafür, dass sowohl Albert als auch Einstein in den Index aufgenommen werden. Da ein Index in der Regel alphabetisch sortiert ist, würde Albert unter *a* zu finden sein und Einstein unter *e* was eine Indexierung von Albert nicht hilfreich erscheinen lässt. Ein Indexeintrag von *albert einstein* oder invertiert *einstein, albert* kann nur über eine Mehrworterkennung erfolgen. Da die Autorennamen der Manuskripte bereits in einer Datenbank vorliegen werden diese nicht lexikalisiert. Somit ist das Ziel der Eigennamenerkennung von Personennamen, Persönlichkeiten in den Dokumenten zu finden, die Namensgeber für Gesetze, Verfahren usw. sind. Am Beispiel von Newton wird deutlich, dass diese in der Regel ohne Vornamen verwendet werden.

1. Newton
2. Newtonverfahren
3. Newton-Verfahren
4. Newton-Raphson-Verfahren
5. Newtonsches Reibungsgesetz
6. Newton-Euler Bewegungsgleichung

Newton wird als einzelnes Wort, als Zusammenfügung mit oder ohne Bindestrich, als Teil einer Mehrwortgruppe oder als eine Kombination von Zusammenfügung und Mehrwortgruppe verwendet. Da *verfahren #s*, *bewegung #s*, *reibung #s*, *gesetz #s* und *gleichung #s* lexikalisiert sind ermöglicht die Lexikalisierung von *newton #e*, dass die Nummern 1, 2, 3, 4 und 6 bei der Indexierung identifiziert werden. Um *<newtonsche>* in der Nr. 5 zu erkennen muss die Suffixliste für Eigennamen um den Eintrag *sches* erweitert werden. Es ist notwendig die Nr. 5 und 6 in das Mehrwortwörterbuch aufzunehmen, um zu erreichen, dass die Einträge an einer Stelle im Index stehen. So ergeben sich 29 Namen und 16 Mehrwortgruppen, die lexikalisiert werden.⁶⁶

⁶⁶ Anhang B-10 Eigennamen.xls, Tabellenblatt: Persönlichkeiten

Organisationsnamen, Abkürzungen und Akronyme

Namen von Organisationen die aus einem Wort bestehen (Eurocopter), sowie Abkürzungen (DLR) und Akronyme (ICAO) werden als ein Wort ohne Rechtsformen (bspw. GmbH) in der Form *eurocopter=eurocopter #e, ihk=ihk #e* lexikalisiert. Insgesamt werden 65 Einträge vorgenommen. Zusätzlich werden alle Organisationsnamen, die aus Mehrwortgruppen bestehen, in das Mehrwortwörterbuch aufgenommen (38 Einträge, bspw. *motoren und turbinen-union*). Organisationen, bei denen die ausgeschriebene und abgekürzte Schreibweise üblich ist, werden für das Synonymwörterbuch vorgemerkt.⁶⁷

Projektnamen und Gegenstände

Auch hier befinden sich Abkürzungen, Akronyme und Mehrwortgruppen in der Liste. Eine Besonderheit stellen die Luftfahrzeug- und Triebwerksbezeichnungen dar. Typisch für diese ist eine Buchstabenkombination, die den Hersteller oder eine bestimmte Serie kennzeichnet, gefolgt von einer Zahlenkombination (*a320*). Die einzelnen Variationen des Typs werden bei fast allen Herstellern durch das Anhängen von einem Bindestrich und weiteren Zahlen gekennzeichnet (*a320-200*). Durch die Bindestrichverbindung ist es ausreichend *a320=a320 #e* zu lexikalisieren da die Identifizierung eines Teils der Bindestrich-Verbindung zur Indexierung genügt um das Token *<a320-200>* als Ganzes zu indexieren. Die Mehrwortgruppen und Akronyme werden wie bei den Organisationsnamen beschrieben, lexikalisiert. Es werden 35 Mehrwortgruppen lexikalisiert, 99 Einträge in das Nutzerwörterbuch vorgenommen und 44 Synonymrelationen für den nächsten Abschnitt vorgemerkt.⁶⁸

4.2.9 Behandlung von Synonymen

Der Lingo Attendee „synonymer“ ermöglicht Lexikoneinträgen Worte mit synonyme Bedeutung zuzuordnen. Anstatt der Mehrwortgruppe aus der Bearbeitung der Eigennamen *internationale luftfahrt ausstellung* wird oft das Akronym *ila* benutzt. Durch den Eintrag *internationale luftfahrt ausstellung=ila* kann das Dokument zusätzlich mit dem Akronym indexiert werden. So ist es möglich, das Dokument über beide der Indexterme zu finden. Für Lexikoneinträge sind Regeln zu beachten. Der Eintrag auf der linken Seite des Gleichheitszeichens muss sich im Text des Dokumentes befinden um den synonymen

⁶⁷ Anhang B-11 Fremdspraches.xls; Tabellenblatt: Organisationsnamen

⁶⁸ Anhang B-11 Fremdspraches.xls; Tabellenblatt: Gegenstände Projekte

Eintrag rechts vom Gleichheitszeichen dem Dokument zuzuordnen. Auf beiden Seiten des Gleichheitszeichens können beliebige Kombinationen von alphanumerischen Zeichen, Mehrwortgruppen und der Bindestrich als Sonderzeichen verwendet werden. Werden Mehrwortgruppen verwendet, müssen diese in das Mehrwortwörterbuch eingetragen werden. Da kein deutschsprachiger Fachthesaurus zur Verfügung stand, wird die Synonymfunktion für die Auflösung von Akronymen und Abkürzungen verwendet. Lexikalisiert wurden alle vorgemerkten Kandidaten aus den vorherigen Abschnitten. Um eine Indexierung mit beiden synonymen Einträgen sicher zu stellen wurden alle Einträge zusätzlich in vertauschter Reihenfolge vorgenommen, so ist die Synonymrelation *luftfahrt bundesamt=lba* auch als *lba=luftfahrt bundesamt* in das Lexikon eingegeben worden. Möglich ist auch die Zuweisung von chemischen Formeln zu den entsprechenden Bezeichnungen. In den Manuskripten kommt bspw. der Name *invar* vor, der die Bezeichnung für eine Eisen-Nickel Verbindung mit der Formel FeNi ist. *invar* wurde als Wortklasse *#e* lexikalisiert und als Synonym *feni* in das Synonymwörterbuch eingetragen. Insgesamt werden chemische Formelzeichen selten in den Manuskripten verwendet, so dass von einer systematischen Lexikalisierung abgesehen wurde.

4.3 Indexierung von englischsprachigen Textteilen

In der Liste der nicht identifizierten Token (Tab. 1) befinden sich 4007 Token, die meist englischsprachig sind oder englischsprachige Anteile haben.⁶⁹ Die englischsprachigen Wörter können nicht in den Lingo Wortklassen *#a*, *#s* und *#v* lexikalisiert werden, da die zugehörigen Suffixlisten nicht der englischen Sprache entsprechen. Englischsprachige Wortklassen mit den dazugehörigen Regeln zu implementieren und die englischsprachigen Wörter diesen zuzuordnen würde den Rahmen dieser Arbeit sprengen und unterbleibt. Um zumindest einen Teil der benutzten englischen Wörter zu indexieren, wurden die Einträge des englischen Open Office Wörterbuchs für die Rechtschreibung als separates Nutzerwörterbuch in die Indexierung einbezogen. Alle Einträge wurden als Wortklasse *#e* lexikalisiert. Bei einer Indexierung nur mit dem englischen Wörterbuch werden von den 4007 Token 813 nicht erkannt. Werden alle Lexikonerweiterungen in die Indexierung einbezogen, reduziert sich die Zahl auf 764.⁷⁰ Alle in den bisherigen Lexikonerweiterungen lexikalisierten englischsprachigen Wörter wurden als Wortklasse *#g*

⁶⁹ Anhang B-11 Fremdspraches.xls

⁷⁰ Ebd.

in das Nutzerwörterbuch eingetragen. Diese werden als Wortklasse #e dem englischen Nutzerwörterbuch hinzugefügt und als Wortklasse #g im Nutzerwörterbuch gelöscht.

4.4 Wörter und Mehrwortgruppen anderer Quellen

Für die Erweiterung des Nutzerwörterbuchs wurden drei Quellen benutzt. Die erste Quelle ist das Fachthemenverzeichnis des Lexikon der Luftfahrt von Klußmann und Malik.⁷¹ Aus etwa 2100 Einträgen wurden 50 Mehrwortgruppen und 95 Einzeleinträge ausgewählt. Alle Wörter wurden vor einer Lexikalisierung mit Lingo indexiert. Von den 96 Einzeleinträgen werden 68, und von den Wörtern der Mehrwortbegriffe werden alle ohne vorherige Lexikalisierung erkannt. Somit ergeben sich nur 28 Lexikoneinträge aus den Einzelwörtern und zusätzliche 17 aus den Mehrwortgruppen, da diese, wie im Abschnitt 4.2.7 beschrieben, keine Zusammenfügungen enthalten dürfen.⁷² Als weitere Quellen wurden die Einträge des Inhaltsverzeichnisses des Online Lexikon von www.thgweb.de⁷³ und des elektronischen Registers von Duden Physik⁷⁴ und Duden Mathematik verwendet.⁷⁵ Alle Einträge wurden einer Indexierung unterzogen und alle nicht erkannten Worte für ihre Verwendbarkeit im Nutzerwörterbuch untersucht. Für die beiden Bücher des Duden-Verlages wurden 31 Einträge für das Nutzerwörterbuch und 11 Einträge für das Mehrwortwörterbuch ausgewählt.^{76 77} Aus dem Inhaltsverzeichnis des www.thgweb.de - Lexikon wurden 87 Einträge⁷⁸ für das Nutzerwörterbuch und 241 Mehrwortgruppen ausgewählt.

4.5 Indexierung der Kontrollmenge

Zum Vergleich mit der Testmenge, wird die Kontrollmenge insgesamt 4-mal indexiert. Zuerst wird eine Indexierung nur mit den Standard Lingo Wörterbüchern durchgeführt, dann mit allen Lexikonerweiterungen bis auf das englischsprachige Open Office Lexikon und Einträgen aus fremden Quellen. Anhand dieser beiden Indexierungen werden die anhand der Testmenge erarbeiteten Lexikonerweiterungen bewertet. Zusätzlich wird noch eine Indexierung mit dem englischen Open Office Lexikon, und eine mit allen Lexikon Erweiterungen durchgeführt.

⁷¹ Klußman, Malik, 2004, S. 294-313

⁷² Anhang B-12 Lexikon der Luftfahrt.xls

⁷³ Vgl. TG Publishing AG Homepage, Lexikon, Kategorie: Technik

⁷⁴ Meyer, Schmidt, 2001, CD-ROM

⁷⁵ Bossek, Eichler, Engelmann et al., 2001, CD-ROM

⁷⁶ Anhang B-13 DudenPhysik.xls

⁷⁷ Anhang B-14 DudenMathematik.xls

⁷⁸ Anhang B-15 THG Lexikon.xls

5 Indexierungsergebnisse

5.1 Beschreibung der Indexierungsergebnisse

Für die Beschreibung der Indexierungsergebnisse der Testmenge wurde die Tab. 17 erstellt. Zeilenweise werden die einzelnen Indexierungen, beginnend mit der Standard Lingo Indexierung, beschrieben. In jeder Zeile wird die aufgeführte Lexikonerweiterung hinzugefügt und die Werte der Indexierung angegeben. Zur Übersicht wird die Anzahl aller Token ohne doppelte Nennungen und die Anzahl aller Token, die in mindestens zwei Dokumenten vorkommen, angegeben. In der letzten Spalte wird der Quotient von Token (ohne doppelte Nennung) mit der Anzahl der hinzugekommenen Lexikoneinträge aufgeführt.

Die Tabelle 17 zeigt, dass die letzte Indexierung rund 15600 Token nicht identifizieren kann. Darin enthalten sind die 6543 Token die Fragmente oder Fehler enthalten (siehe Tab. 1), 890 Token die aus einem oder zwei Zeichen bestehen und 2652 Token die einen Ergänzungsstrich enthalten.⁷⁹ Alle diese Token können und sollen nicht indexiert werden. Somit verbleiben 5499 Token die Kandidaten für eine mögliche Indexierung sind. Zur Kontrolle wurden 2745 Token dieser Menge, die in mindestens zwei Dokumenten vorkommen erneut untersucht. Lediglich 30 Token wurden in das Nutzerwörterbuch nachgetragen.⁸⁰

Indexierung der Testmenge (Jahrgang 04 und 05)				
Lexikonerweiterung	Anzahl Lexikoneinträge	Nicht identifizierte Token		Gefundene Token (ohne doppelte) je Lexikoneintrag
		ohne doppelte	mindestens in 2 Dokumenten	
Lingo Standard	-	27919	6962	-
+ Ein Kernmorphem	323	26215	6289	5,3
+ Plural	213	25600	6111	2,9
+ Wortstämme	90	23289	5360	25,7
+ Präfixe	262	22583	5069	2,7
+ Suffixe	105	22231	4934	3,4
+ Komposita	211	21618	4829	2,9
+ Bindestrichverbindungen	25	21580	4800	1,5
+ Eigennamen/Akronyme	195	20776	4598	4,1
+ Sprache	-	15742	2773	-
Alle Erweiterungen	238	15584	2745	

Tab. 17: Indexierungsergebnisse der Testmenge⁸¹

⁷⁹ Anhang B-16 Token mit Ergänzungsstrich.xls

⁸⁰ Anhang B-17 Testmenge Nonwords Alle.xls

⁸¹ Die Zahlen der nicht identifizierten Token befinden sich im Anhang B-18

Zudem wurden 241 Mehrwortgruppen in 150 Dokumenten indexiert, sowie synonyme Terme 51 Dokumenten zugewiesen.⁸²

Für die 4 Indexierungen der Kontrollmenge wurde analog zur Indexierung der Testmenge eine Tabelle angelegt (Tab. 18). Der Quote aus gefundenen Token je Lexikoneintrag wurde hier nicht berechnet, da aufgrund der geringeren Anzahl der Dokumente und Token, die Zahlen mit der Testmenge nicht vergleichbar sind. Vergleicht man die 885 verbleibenden nicht identifizierten Token, die in mindestens zwei Dokumenten vorkommen mit der Testmenge, so stellt man fest, dass 395 Token in beiden Mengen vorkommen.⁸³ Es verbleiben also 477 Token, die nicht in der Testmenge vorhanden sind. Diese 477 Token beschreiben den Aufwand, der bei der Dokumenterweiterung durch die Kontrollmenge (Jahrgang 06), neben der Suche nach Mehrwortgruppen und Synonymen, zu leisten wäre.

Indexierung der Kontrollmenge (Jahrgang 06)		
Lexikonerweiterung	Nicht identifizierte Token	
	ohne doppelte	mindestens in 2 Dokumenten
Lingo Standard	9303	2031
Alle Erweiterungen außer dem englischen Open Office Lexikon und der fremden Quellen (entspricht der „Eigennamen/Akronyme“ Indexierung aus Tab. 17)	7243	1331
Alle Erweiterungen außer den fremden Quellen (entspricht der „Sprache“ Indexierung aus Tab 17)	5226	898
Alle Erweiterungen	5176	885

Tab. 18: Indexierungsergebnisse der Kontrollmenge⁸⁴

Für die Kontrollmenge konnten in 40 Dokumenten 55 Mehrwortgruppen identifiziert werden und bei 17 Dokumenten wurden insgesamt 18 Synonyme zugeordnet.⁸⁵

5.2 Bewertung der Indexierung

Festzustellen ist, dass die Zahl der nicht identifizierten Token nach der letzten Indexierung mit allen Lexikonerweiterungen immer noch relativ hoch ist. Einer der Hauptgründe dafür ist die, mit über 6500, sehr hohe Zahl der Wortfragmente und Fehler. Die fragmentierten Token sind innerhalb des Indexierungsvorgangs von Lingo nicht zu reparieren. Die Lösung

⁸² Anhang B-8-10 Testmenge NonAlles.xls

⁸³ Anhang B-20 Vergleich von Test- und Kontrollmenge.xls

⁸⁴ Die Zahlen der nicht identifizierten Token befinden sich im Anhang B-19

⁸⁵ Anhang B-19-4 Kontrollmenge alles.xls

dieses Problems ist bei der Erstellung der Dokumente und bei der Zeichenkonvertierung vom PDF-Format in das Text-Format zu suchen.

Vergleicht man die Ausgangsindexierung mit der Indexierung, die die Eigennamen einschließt, so ergibt eine Quote von fünf identifizierten Token (ohne doppelte Nennung) die sich aus einem Lexikoneintrag ergeben. Diese Quote wird durch das Konzept der Lexikalisierung von einfachen Wörtern und deren Ableitungen erreicht, die es ermöglicht die Mehrheit der Zusammenfügungen über die Kompositaerkennung zu identifizieren. Diesem Grundkonzept von Lingo arbeitet nur das Verfahren der Erkennung von Mehrwortgruppen entgegen, die eine Lexikalisierung von Komposita benötigt. Besonders erfolgreich hat sich die systematische Lexikalisierung von Wortstämmen und Präfix-Wortstammverbindungen herausgestellt. Hier findet jeder Lexikoneintrag durchschnittlich 25 bisher nicht identifizierte Token ohne erkennbare negative Nebenwirkungen. Die vorhandenen Token mit Ergänzungsstrichen konnten nicht in die Indexierung einbezogen werden. Aber es wurde aufgezeigt, dass eine Einbeziehung in Lingo grundsätzlich möglich ist, da die hierfür notwendige Voraussetzung - eine erfolgreiche Kompositazerlegung - durch Lingo gegeben ist. Die Mehrworterkennung und die Behandlung von Synonymen konnten aufgrund der eingeschränkten Kenntnisse der Fachsprache nur exemplarisch erfolgen und bedürfen weitere Anstrengungen. Die Einbindung des englischen Open Office Lexikons wird nur als „Notlösung“ angesehen, da bspw. keinerlei Flexionen berücksichtigt werden. Auch hier ist es notwendig - mit den entsprechenden Kenntnissen über die Fachsprache - die englischen Wörter, die in den deutschen Sprachgebrauch übernommen wurden, in die Lingo Wörterbücher aufzunehmen. Die Lexikonerweiterung aus fremden Quellen haben wenig zur Erkennung von bislang nicht identifizierter Token beigetragen. Dies ist aber keine überraschende Feststellung, da diese Erweiterungen erst nach erfolgter, weitgehender Lexikalisierung der in der Testmenge vorkommenden Token, erfolgte. Insgesamt ist die Indexierung als erfolgreich zu bewerten. Betrachtet man die Indexierung ohne das englischsprachigen Open Office Lexikon und ohne die Wörter aus fremden Quellen, so ist es gelungen die deutschsprachigen Texte und die darin verwendete Fachsprache mit „nur“ 1424 zusätzlichen Einträge für das Systemwörterbuch weitgehend zu erschließen.

Die Test- und die Kontrollmenge ist insofern vergleichbar, dass der Quotient aus nicht identifizierten Token und der Anzahl der Dokumente, für beide Indexierungen (Tab. 19) gleich verhält.

Indexierung	Testmenge (305 Dokumente)		Kontrollmenge (107 Dokumente)	
	Anzahl Token ohne doppelte Nennungen	Nicht identifizierte Token pro Dokument	Anzahl Token ohne doppelte Nennungen	Nicht identifizierte Token pro Dokument
Lingo Standard	27919	91	9303	87
Ohne Sprache und fremde Quellen	20776	68	7243	68
Ohne fremde Quellen	15742	52	55226	49
alle	15584	51	5176	48

Tab. 19: Übersicht nicht identifizierter Token pro Dokument

Zudem zeigte der Betrachtung, der nach der letzten Indexierung verbliebenen nicht identifizierten Token, dass durch die 107 Dokumente der Testmenge nur 477 bislang unbekannte Token hinzugekommen sind. Der benötigte Aufwand der Lexikonpflege ist somit überschaubar.

6 Zusammenfassung

Es ist festzustellen, dass Lingo als linguistisch basiertes Programm zur automatischen Indexierung geeignet ist, wissenschaftlich-technischer Volltexte erfolgreich zu indexieren. Durch die erarbeiteten Lexikonerweiterungen ist es gelungen, in den Texten vorkommende Terme weitgehend zu erschließen und somit die Grundlage für eine thematische Suche in einem Retrievalsystem zu schaffen, deren Pflege eines geringen Aufwands bedarf. Die dargestellten Änderungen in der Mehrworterkennung und Verarbeitung von Ergänzungsstrichen würden die Indexierungsergebnisse von Lingo weiter verbessern.

Für die Zukunft steht neben der Einbettung der Indexierungsergebnisse in ein Retrievalsystem, die Suche nach einer möglichen Erschließung der englischsprachigen Kongressschriften an. Bisher war der Anteil von deutsch- und englischsprachigen Manuskripten in etwa ausgeglichen, tendenziell steigt der Anteil der englischsprachigen Manuskripte an und würde eine englischsprachige Version von Lingo erfordern.

7 Literatur- und Quellenverzeichnis

Bertram, Jutta: Einführung in die inhaltliche Erschließung. Grundlagen – Methoden – Instrumente. Würzburg: Ergon 2005, S. 83.

Felgentreu, Simone; Friedrich, Anne-Cathrin; Huster, Sonja; Langermann, Detlef; Lindner: Duden Deutsch. Basiswissen Schule. Berlin: PAETEC, Mannheim: Brockhaus 2002, CD-ROM.

Ferber, Reginald: Information Retrieval. 1.Aufl. Heidelberg: dpunkt 2003, S. 25.

Grimm, Rüdiger: Digitale Kommunikation. München, Wien: Oldenbourg 2005, S. 19.

Klußmann, Niels; Malik, Arnim: Lexikon der Luftfahrt. Berlin, Heidelberg, New York: Springer 2004, S. 294-313.

Lepsky, Klaus: Automatische Indexierung. Linguistischen & statistische Verfahren. Vorlesungsskript der Lehrveranstaltung: Automatisches Indexieren. Köln: Fachhochschule Köln. Institut für Informationswissenschaft, 2006. Online: <http://www.iws.fh-koeln.de/institut/personen/lepsy/skript-3d-automatische-indexierung-linguistik-statistik-06.pdf>

Lepsky, K., Siepmann, J; Zimmermann, A: Automatische Indexierung für Online-Kataloge: Ergebnisse eines Retrievaltests. In: Zeitschrift für Bibliothekswesen und Bibliographie. 43(1996) H.1, S.47-56.

Lepsky, Klaus; Vorhauer, John: Lingo – ein open source System für die Automatische Indexierung des Deutschen, 2006. Online: <http://www.iws.fh-koeln.de/institut/personen/lepsy/lingo-uebersichtsartikel.pdf> (15.02.2007)

Meyer, Lothar; Schmidt, Gerd-Dietrich: Duden Physik. Basiswissen Schule. Berlin: PAETEC, Mannheim: Brockhaus 2001, CD-ROM.

Bossek, Hubert; Eichler, Klaus-Peter; Engelmann, Lutz; Fanghänel, Günter; Lehmann, Karlheinz; Oberländer, Franz; Paulin, Gerhard; Sill, Hans-Dieter; Stamm, Reinhard: Duden Physik. Basiswissen Schule. Berlin: PAETEC, Mannheim: Brockhaus 2001, CD-ROM.

Nohr, Holger: Theorie des Information Retrieval II: Automatische Indexierung. In Grundlagen der praktischen Information und Dokumentation. Hrsg. von Kuhlen, Seeger, Strauch. 5. Ausg. München: Saur 2004, S. 216, S. 217.

Nohr, Holger: Grundlagen der automatischen Indexierung. Ein Lehrbuch. 3.Aufl. Berlin: Logos 2005, S. 34, und S. 66ff.

o.A.: Duden. Die deutsche Rechtschreibung. 22.Aufl. Mannheim, Leipzig, Wien, Zürich: Duden 2000, S. 31-34.

- o.A.:** „Zukunft der wissenschaftlichen und technischen Information in Deutschland.“
Hrsg. Bundesministerium für Bildung und Forschung (BMBF), Bonn 2002, S. 1.
Online: http://www.bmbf.de/pub/zukunft_der_wti_in_deutschland.pdf (02.06.2007)
- Reinhardt, W.; Köhler, C.; Neubert, G.:** Deutsche Fachsprache der Technik. Hrsg. Arntz, Reiner und Wegner, Norbert. Hildesheim, Zürich, New York: Olms 1992, S. 35ff
- Schütz, Thomas:** Dokumentenmanagement. In: Grundlagen der praktischen Information und Dokumentation. Hrsg. von Kuhlen, Seeger, Strauch. 5. Ausg. München: Saur 2004, S. 340.
- Stock, Wolfgang G.:** Information Retrieval. Information suchen und finden. München, Wien: Oldenbourg 2007, S. 216.
- TG Publishing AG Homepage:** Tom´s Hardware Lexikon, Kategorie: Technik. Online: <http://www.thgweb.de/lexikon/Kategorie:Technik> (02.04.2007)

8 Anhang

Anhangsverzeichnis

Anhang A

Anhang A-1	Liste heimischer Präfixe.....	63
Anhang A-2	Übersicht der Wortstämme	64
Anhang A-3	Lingo Systemwörterbuch, Einträge mit <i>-fehl-</i>	66
Anhang A-4-1	Wortstamm <i>-blas-</i>	67
Anhang A-4-2	Wortstamm <i>-brems-</i>	68
Anhang A-4-3	Wortstamm <i>-brenn-</i>	69
Anhang A-4-4	Wortstamm <i>-bring-</i>	71
Anhang A-4-5	Wortstamm <i>-dehn-</i>	72
Anhang A-4-6	Wortstamm <i>-dreh-</i>	73
Anhang A-4-7	Wortstamm <i>-fahr-</i>	74
Anhang A-4-8	Wortstamm <i>-fließ-</i>	76
Anhang A-4-9	Wortstamm <i>-grenz-</i>	77
Anhang A-4-10	Wortstamm <i>-kalibrier-</i>	79
Anhang A-4-11	Wortstamm <i>-kenn-</i>	80
Anhang A-4-12	Wortstamm <i>-kerb-</i>	81
Anhang A-4-13	Wortstamm <i>-koppl-</i>	82
Anhang A-4-14	Wortstamm <i>-last-</i>	83
Anhang A-4-15	Wortstamm <i>-misch-</i>	85
Anhang A-4-16	Wortstamm <i>-nick-</i>	86
Anhang A-4-17	Wortstamm <i>-nutz-</i>	87
Anhang A-4-18	Wortstamm <i>-press-</i>	88
Anhang A-4-19	Wortstamm <i>-prüf-</i>	89
Anhang A-4-20	Wortstamm <i>-roll-</i>	91
Anhang A-4-21	Wortstamm <i>-rück-</i>	93
Anhang A-4-22	Wortstamm <i>-saug-</i>	95

Anhang A-4-23	Wortstamm <i>-schalt-</i>	96
Anhang A-4-24	Wortstamm <i>-scher-</i>	97
Anhang A-4-25	Wortstamm <i>-steig-</i>	98
Anhang A-4-26	Wortstamm <i>-stell-</i>	99
Anhang A-4-27	Wortstamm <i>-ström-</i>	101
Anhang A-4-28	Wortstamm <i>-such-</i>	103
Anhang A-4-29	Wortstamm <i>-trag-</i>	104
Anhang A-4-30	Wortstamm <i>-trimm-</i>	105
Anhang A-4-31	Wortstamm <i>-wirk-</i>	106
Anhang A-4-32	Wortstamm <i>-wölb-</i>	107
Anhang A-5	Indexierte Wörtern mit Zeichenfolge <i>-fehl-</i>	108
Anhang A-6	Präfixliste	111
Anhang A-7-1	Token mit Zeichenfolge <i>ab-</i>	114
Anhang A-7-2	Token mit Zeichenfolge <i>ad-</i>	115
Anhang A-7-3	Token mit Zeichenfolge <i>an-</i>	116
Anhang A-7-4	Token mit Zeichenfolge <i>auf-</i>	118
Anhang A-7-5	Token mit Zeichenfolge <i>aus-</i>	119
Anhang A-7-6	Token mit Zeichenfolge <i>be-</i>	120
Anhang A-7-7	Token mit Zeichenfolge <i>de-</i>	122
Anhang A-7-8	Token mit Zeichenfolge <i>di-</i>	124
Anhang A-7-9	Token mit Zeichenfolge <i>ein-</i>	125
Anhang A-7-10	Token mit Zeichenfolge <i>en-</i>	126
Anhang A-7-11	Token mit Zeichenfolge <i>erd-</i>	128
Anhang A-7-12	Token mit Zeichenfolge <i>eu-</i>	129
Anhang A-7-13	Token mit Zeichenfolge <i>ex-</i>	130
Anhang A-7-14	Token mit Zeichenfolge <i>ge-</i>	131
Anhang A-7-15	Token mit Zeichenfolge <i>im-</i>	133
Anhang A-7-16	Token mit Zeichenfolge <i>in-</i>	134
Anhang A-7-17	Token mit Zeichenfolge <i>iso-</i>	136
Anhang A-7-18	Token mit Zeichenfolge <i>kon-</i>	137
Anhang A-7-19	Token mit Zeichenfolge <i>mikro-</i>	138

Anhang A-7-20	Token mit Zeichenfolge pro-	139
Anhang A-7-21	Token mit Zeichenfolge rück-.....	140
Anhang A-7-22	Token mit Zeichenfolge trans-	141
Anhang A-7-23	Token mit Zeichenfolge über-.....	142
Anhang A-7-24	Token mit Zeichenfolge ultra-	143
Anhang A-7-25	Token mit Zeichenfolge um-.....	144
Anhang A-7-26	Token mit Zeichenfolge un-.....	145
Anhang A-7-27	Token mit Zeichenfolge ver-.....	146
Anhang A-7-28	Token mit Zeichenfolge vor-	147
Anhang A-7-28	Token mit Zeichenfolge zu-	148
Anhang A-8	Suffixliste	149
Anhang A-9-1	Token mit Zeichenfolge -ation.....	155
Anhang A-9-2	Token mit Zeichenfolge -bar.....	156
Anhang A-9-3	Token mit Zeichenfolge -chen	157
Anhang A-9-4	Token mit Zeichenfolge -ent.....	158
Anhang A-9-5	Token mit Zeichenfolge -ig.....	159
Anhang A-9-6	Token mit Zeichenfolge -ion.....	160
Anhang A-9-7	Token mit Zeichenfolge -it.....	162
Anhang A-9-8	Token mit Zeichenfolge -keit.....	163
Anhang A-9-9	oken mit Zeichenfolge -kel	164
Anhang A-9-10	Token mit Zeichenfolge -lich.....	165
Anhang A-9-11	Token mit Zeichenfolge -ment.....	166
Anhang A-9-12	Token mit Zeichenfolge -ns	167
Anhang A-9-13	Token mit Zeichenfolge -o.....	168
Anhang A-9-14	Token mit Zeichenfolge -on.....	169
Anhang A-9-15	Token mit Zeichenfolge -ung.....	172
Anhang A-9-16	Token mit Zeichenfolge -werk.....	174

Anhang B⁸⁶

Anhang B-1-1	Kontrolle -blas-.xls
Anhang B-1-2	Kontrolle -brems-.xls
Anhang B-1-3	Kontrolle -brenn-.xls
Anhang B-1-4	Kontrolle -bring-.xls
Anhang B-1-5	Kontrolle -dehn-.xls
Anhang B-1-6	Kontrolle -dreh-.xls
Anhang B-1-7	Kontrolle -fahr-.xls
Anhang B-1-8	Kontrolle -fehl-.xls
Anhang B-1-9	Kontrolle -fließ-.xls
Anhang B-1-10	Kontrolle -grenz-.xls
Anhang B-1-11	Kontrolle -kalibrier-.xls
Anhang B-1-12	Kontrolle -kenn-.xls
Anhang B-1-13	Kontrolle -kerb-.xls
Anhang B-1-14	Kontrolle -koppl-.xls
Anhang B-1-15	Kontrolle -last-.xls
Anhang B-1-16	Kontrolle -mess-.xls
Anhang B-1-17	Kontrolle -misch-.xls
Anhang B-1-18	Kontrolle -nick-.xls
Anhang B-1-19	Kontrolle -nutz-.xls
Anhang B-1-20	Kontrolle -press-.xls
Anhang B-1-21	Kontrolle -prüf-.xls
Anhang B-1-22	Kontrolle -roll-.xls
Anhang B-1-23	Kontrolle -rück-.xls
Anhang B-1-24	Kontrolle -saug-.xls
Anhang B-1-25	Kontrolle -schalt-.xls
Anhang B-1-26	Kontrolle -scher-.xls
Anhang B-1-27	Kontrolle -steig-.xls
Anhang B-1-28	Kontrolle -stell-.xls
Anhang B-1-29	Kontrolle -ström-.xls
Anhang B-1-30	Kontrolle -such-.xls

⁸⁶ Der Anhang B ist ein elektronischer Anhang und befindet sich auf der beigelegten CD-ROM

Anhang B-1-31	Kontrolle -trag-.xls
Anhang B-1-32	Kontrolle -trimm-.xls
Anhang B-1-33	Kontrolle -wirk-.xls
Anhang B-1-34	Kontrolle -wölb-.xls
Anhang B-2	Token mit Bindestrich.xls
Anhang B-3	Mehrfach-.xls
Anhang B-4	-Mehrfach.xls
Anhang B-5	Fall I.xls
Anhang B-6	FallVI.xls
Anhang B-6-1	FallVI und.xls
Anhang B-6-2	FallVI Komma.xls
Anhang B-6-3	FallVI Schrägstrich.xls
Anhang B-6-4	FallVI bzw.xls
Anhang B-6-5	FallVI oder.xls
Anhang B-6-6	FallVI Stoppwortliste.xls
Anhang B-7	FallVII.xls
Anhang B-7-1	FallVII Punkt.xls
Anhang B-7-2	FallVII und.xls
Anhang B-7-3	FallVI Schrägstrich.xls
Anhang B-7-4	FallVI bzw.xls
Anhang B-7-5	FallVI oder.xls
Anhang B-7-6	FallVI Stoppwortliste.xls
Anhang B-8	FallVII.xls
Anhang B-8-1	FallVII Punkt.xls
Anhang B-8-2	FallVII und.xls

Anhang B-9-1 Mehrwortgruppen AdjAdjSub.xls

Anhang B-9-2 Mehrwortgruppen AdjSub.xls

Anhang B-10 Eigennamen.xls

Anhang B-11 Fremdspraches.xls

Anhang B-12 Lexikon der Luftfahrt.xls

Anhang B-13 DudenPhysik.xls

Anhang B-14 DudenMathematik.xls

Anhang B-15 THG Lexikon.xls

Anhang B-16 Token mit Ergänzungsstrich.xls

Anhang B-17 Testmenge Nonwords Alle.xls

Anhang B-18-1 Testmenge NonKernmorphem.xls

Anhang B-18-2 Testmenge NonPlural.xls

Anhang B-18-3 Testmenge NonWortstamm.xls

Anhang B-18-4 Testmenge NonPräfix.xls

Anhang B-18-5 Testmenge NonSuffix.xls

Anhang B-18-6 Testmenge NonKomposita.xls

Anhang B-18-7 Testmenge NonBindestrich.xls

Anhang B-18-8 Testmenge NonEigennamen.xls

Anhang B-18-9 Testmenge NonSprache.xls

Anhang B-18-10 Testmenge NonAlles.xls

Anhang B-19-1 Kontrollmenge Standard.xls

Anhang B-19-2 Kontrollmenge Eigennamen.xls

Anhang B-19-3 Kontrollmenge Sprache.xls

Anhang B-19-4 Kontrollmenge alles.xls

Anhang B-20 Vergleich von Test- und Kontrollmenge.xls

Anhang B-21 Lexikoneinträge der Präfixbearbeitung.xls

Anhang B-22 Lexikoneinträge der Pluralbearbeitung.xls

Anhang A-1

Liste heimischer Präfixe

Präfixliste
ab
an
auf
aus
be
ein
ent
er
ge
mit
um
un
ur
ver
vor
zer
zu

Anhang A-2

Übersicht der Wortstämme

		0405statnon	lexikalisiert	Kontrolle	Anz. Index
beul	40	-/- 4	j		
blas	58	-/- 5	n		
brems	94	-/- 10	j	ja	112
brenn	204	-/- 44	j	ja	295
bring	14	-/- 5	n		
brück	3	-/- 3	n		
dehn	43	-/- 12	j		
dick	29	-/- 5	n		
dien	42	-/- 7	n		
dreh	161	-/- 33	j	ja	229
dring	5	-/- 1	n		
fahr	269	-/- 48	n		
fang	28	-/-4	n		
fehl	60	-/- 11	j	ja	237
flamm	27	-/- 3	j		
fließ	18	-/- 5	j		
fliess	9	-/- 2	n		
fuhr	4	-/- 1	n		
führ	33	-/- 5	j		
füll	46	-/- 5	n		
gleit	85	-/- 13	j	ja	78
heiz	26	-/- 1	j		
hitz	41	-/- 5	n		
kalibrier	34	-/- 2	j		
kenn	87	-/- 29	j		
kipp	10	-/- 1	j		
kerb	30	-/- 5	j		
klebe	8	-/- 2	n		
knet	4	-/- 0	n		
kodier	4	-/- 1	n		
kriech	15	-/- 4	j		
lapp	32	-/- 8	n		
last	323	-/- 40	n	ja	523
lenk	79	-/- 12	n		
lösch	17	-/- 2	j		
mess	364	-/- 103	j	ja	514
misch	84	-/- 40	j		
nenn	15	-/- 3	j		
nick	55	-/- 17	j		
nutz	100	-/- 17	j		
pass	51	-/- 16	n		
plank	7	-/- 1	n		
prell	1	-/- 1	n		
press	45	-/- 13	n		
prüf	174	-/- 40	j		
pump	29	-/- 3	j	ja	44
reiss	5	-/- 1	n		
reiß	9	-/- 3	j		

Fortsetzung Anhang A-2

Übersicht der Wortstämme

	regio	5 -/- 0	n		
	renn	175 -/- 43	n		
	repetier	1 -/- 0	n		
	roll	216 -/- 50	j		
	rück	158 -/- 33	j		
	rüst	9 -/- 3	j		
	saug	45 -/- 12	j		
	schäl	6 -/- 0	j		
	schalt	34 -/- 5	j		
	scher	102 -/- 23	j	ja	78
	schraub	17 -/- 4	n		
	schrumpf	9 -/- 0	n		
	schub		n		
	schul	9 -/- 1	j		
	schwad	2 -/- 1	n		
	schweb	3 -/- 0			
	schwing	58 -/- 6	j		
	setz	46 -/- 6	n		
	sicht	26 -/- 4	n		
	sink	19 -/- 5	j		
	sperr	16 -/- 4	j		
	spül	3 -/- 1	j		
	stanz	1 -/-0	n		
	stäub	5 -/- 3	n		
	steig	35 -/- 7	j		
	stell	217 -/- 48	j		
	strom	53 -/- 5	j		
	ström	196 -/- 45	j		
	stütz	49 -/- 3	j		
	such	104 -/- 20	j	ja	426
	tank	24 -/- 2	n		
	tast	26 -/- 3	n		
	trag	115 -/- 28	j		
	trenn	23 -/- 5	j		
	treib	20 -/- 5	n		
	trimm	44 -/- 8	j		
	wälz	9 -/- 1	j		
	wirk	47 -/- 8	j		
	wölb	49 -/- 12	j	ja	33
	zieh	20 -/- 2	j		
	zünd	20 -/- 2	n		

Anhang A-3

Lingo Systemwörterbuch, Einträge mit *-fehl-*

lingo		
abwehrfehler	fehlersuchprogramm	hauptfehler
anempfehlen	fehlertyp	hörfehler
befehl	fehlerzustand	japanempfehlung
befehlen	fehlfunktion	lötfehler
befehlend	fehlgebeten	meßfehler
befehligen	fehlgegriffen	parityfehler
befehligend	fehlgeleitet	programmierfehler
befehligt	fehlgeschlagen	rechtschreibfehler
befehlshaber	fehlgeschossen	schießbefehl
befehlshaberisch	fehlgewicht	schreibbefehl
denkfehler	fehlgreifend	schreibfehler
durchhaltebefehl	fehlhieb	sehfehler
empfehlen	fehlinvestieren	setzfehler
empfehlend	fehlinvestierend	stimmfehler
empfehlenswert	fehlinvestiert	strafbefehl
empfehlung	fehlleitend	suchbefehl
empfehlungsliste	fehlmaß	unfehlbar
erbfehler	fehlmenge	unfehlbarkeit
erstempfehlung	fehlmessung	verfehlen
fehlanzeige	fehlpass	verfehlend
fehlbarkeit	fehlschießen	verfehlt
fehlbestand	fehlschießend	verfehlung
fehlbitte	fehlschlag	verkabelungsfehler
fehlbitten	fehlschlagen	weiterempfehlend
fehlbittend	fehlschlagend	
fehlbrand	fehlschlägt	
fehldruck	fehlschlug	
fehleinschätzung	fehlspekulation	
fehlen	fehlstelle	
fehlend	fehlt	
fehlentscheid	fehlte	
fehler	fehlten	
fehleranfällig	fehlurteil	
fehlerbit	fehlverhalten	
fehlerfall	fehlversuch	
fehlerliste	fehlweg	
fehlerlos	fehlwurf	
fehlernummer	führerbefehl	
fehlerquelle	gefehlt	
fehlerrate	gegenbefehl	
fehlersuche	geheimbefehl	

Anhang A-4-1

Wortstamm *-blas-*

Häufigkeit	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
5	ausblasung			aus	blas		ung		1	abblasen
2	aufblasbaren			auf	blas		bar		1	abblasend
2	ausblasrate			aus	blas			komp	2	abblasen
2	einblasung			ein	blas		ung		1	abblasend
2	kühlufteinblasung	komp		ein	blas		ung		3	abgeblasen

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		
1	3			x	x		x			ablassen
2	1			x	x			x		ablassend
3	1	x		x	x		x			anblasen

	Lexikoneintrag	Findet	
	ausblasung	1;	
	einblasung	4;	
	aufblasbar	2;	

ablass
ablassen
ablassend
anblasen
anblasend
angeblasen
aufblasen
aufblasend
aufgeblasen
aufgeblasenheit
aufzublasen
ausblasen
ausblasend
ausgeblasen
blase
blasen
blasend
blasiert
blasinstrument
blaskapelle
blasmusik
blasorchester
blasphemie
blasphemisch
blass
blassend
blastogenese
durchblasend
durchgeblasen
einblasen
einblasend
eingebblasen
erblasen
erblasend
erblasser
erblasung
fortblasen
fortblasend
fortgeblasen
geblasen
herablassung
niederblasen

Anhang A-4-2

Wortstamm *-brems-*

Häufigkeit	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
3	bremsleistung				brems			komp	1	abbremsen
2	brems-				brems				0	abbremsend
2	bremsansätze				brems			komp	1	abbremsung
2	bremsklappen				brems			komp	1	abbremsverhalten
2	bremskoeffizient				brems			komp	1	abgebremst
2	bremskolben				brems			komp	1	anbremsen
2	bremsmoment				brems			komp	1	anbremsend
2	bremschirme				brems			komp	1	angebremst
2	bremsysteme				brems			komp	1	ausbremsen
2	bremszylinder				brems			komp	1	ausbremsend

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		
1	9				x			x		ausgebremst

	Lexikoneintrag	Findet	
	brems #u	1;4;5;6;7;8;9;10	

bremsaktion
 bremsbelag
 bremsseffekt
 bremsen
 bremsend
 bremsler
 bremsfunktion
 bremsklotz
 bremskraft
 bremskurs
 bremslicht
 bremsmanöver
 bremspedal
 bremsprobe
 bremsschuh
 bremsspur
 bremsstrecke
 bremsung
 bremsventil
 bremsversuch
 bremsvorgang
 bremsvorrichtung
 bremswagen
 bremsweg
 bremswirkung
 feststellbremse
 fußbremse
 gebremst
 inflationsbremse
 luftdruckbremse
 verbremsen
 verbremsend
 verbremst
 vierradbremse
 vorderradbremse

Anhang A-4-3

Wortstamm *-brenn-*

Häufigk.	Token	BM	WBM	BM	FM	WBM	BM	Typ	lingo
28	brennkammer			brenn			komp	1	abbrennen
18	brennkammern			brenn			komp	1	abbrennend
8	brennweite			brenn			komp	1	anbrennen
7	brennraum			brenn			komp	1	anbrennend
5	brennkammerdruck			brenn			komp	1	atombrennstoff
5	brennkammerwand			brenn			komp	1	aufbrennen
5	brennstoff-							0	aufbrennend
5	raketenbrennkammern	komp		brenn			komp	2	ausbrennen
4	brenndauer			brenn			komp	1	ausbrennend
4	brennraums			brenn			komp	1	bobrennen
4	modellbrennkammer	komp		brenn			komp	2	branntweinbrenne rei
3	ausbrenngrad		aus	brenn			komp	4	brennbarkeit
3	brenngas			brenn			komp	1	brennen
3	brennkammer-							0	brennend
3	brennkammeraustritt			brenn			komp	1	brenner
3	brennkammerentwicklung			brenn			komp	1	brennerei
3	brennkammergeometrie			brenn			komp	1	brennerpass
3	brennkammerkomponenten			brenn			komp	1	brennessel
3	brennkammersegmente			brenn			komp	1	brennstoff
3	brennkammertechnologie			brenn			komp	1	brennstoffversorg ung
3	brennkammerwände			brenn			komp	1	durchbrennend
3	ringbrennkammer	komp		brenn			komp	2	einbrennen
3	vormischverbrennung	komp	ver	brenn		ung		3	einbrennend
2	brenner-							0	einbrennlackieren
2	brennkammerakustik			brenn			komp	1	entbrennen
2	brennkammerauslegung			brenn			komp	1	entbrennend
2	brennkammerbetrieb			brenn			komp	1	ketzerverbrennun g
2	brennkammerdesign			brenn			komp	1	kornbrennerei
2	brennkammerdrücken			brenn			komp	1	nachverbrennung
2	brennkammergehäuse			brenn			komp	1	niederbrennend
2	brennkammergröße			brenn			komp	1	schnapsbrennerei
2	brennkammerkonzepte			brenn			komp	1	trabrennen
2	brennkammerprozesse			brenn			komp	1	verbrennen
2	brennkammerprüfstand			brenn			komp	1	verbrennend
2	brennkammerschwingungen			brenn			komp	1	verbrennung
2	brennkammertemperatur			brenn			komp	1	verbrennungsofen
2	brennkammervolumen			brenn			komp	1	wegbrennen
2	brennkammerwandkühlung			brenn			komp	1	wegbrennend
2	brennraumes			brenn			komp	1	weinbrennerei
2	brennschluss			brenn			komp	1	
2	-brennweite							0	
2	gasturbinenbrennkammer	komp		brenn			komp	2	
2	lpp-brennkammer							0	
2	raketenbrennkammer	komp		brenn			komp	2	

44

Fortsetzung Anhang A-4-3

Wortstammen *-brenn-*

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
1	32				x			x	
2	5	x			x			x	
3	1	x		x	x		x		
4	1			x	x			x	
0	5								

	Lexikoneintrag	Findet							
	brennkammer	1;2;5;6;8;11;15;16;17;18;19;20;21;22;25;26;27;28; 29;30;31;32;33;34;35;36;37;38;42;43;44;							
	brenn #u	3;4;9;10;12;13;39;40;							
		2	39						

Anhang A-4-4

Wortstamm *-bring-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
5	aufbringung			auf	bring		ung		1	abbringen
5	einbringung			ein	bring		ung		1	abbringend
3	lastaufbringung			auf	bring		ung		1	abringen
3	verbringung			ver	bring		ung		1	abringend
2	nutzbringend	komp			bring	end			2	abzubringen
										anbringen
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		anbringend
1	4			x	x		x			anbringung
2	1	x			x	x				anzubringen
										aufbringen
										aufbringend
										aufzubringen
										ausbringen
										ausbringend
										beibringen
										beibringend
										beizubringen
										bringen
										bringend
										bringschuld
										darbringend
										darbringung
										darzubringen
										durchbringend
										einbringen
										einbringend
										einzubringen
										entgegenbringend
										entgegenzubringen
										erbringen
										erbringend
										erbringung
										fertigbringend
										fortbringend
										glückbringend
										heimbringend
										herausbringen
										herauszubringen
										herbeibringen
										herbeibringend
										herbeizubringen
										herbringen
										herbringend
										herunterbringen
										herunterbringend
										herunterzubringen
										hervorbringend
										hervorzubringen
										herzubringen
										Lexikonliste gekürzt !!

3

5

Anhang A-4-5

Wortstamm *-dehn-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
6	dehnmessstreifen				dehn			komp	1	ausdehnbar
3	dehnrate				dehn			komp	1	ausdehnen
3	dehnungs-				dehn		ung		0	ausdehnend
3	kriechdehnung	komp			dehn		ung		3	ausdehnung
3	spannungs-dehnungs-				dehn				2	ausgedehnt
2	dehngrenze				dehn			komp	1	auszudehnen
2	dehnraten				dehn			komp	1	dehnbar
2	dehnratenabhängige				dehn			komp	1	dehnbarkeit
2	dehnratenabhängigen				dehn			komp	1	dehnen
2	dehnratenabhängigkeit				dehn			komp	1	dehnend
2	dehnungsverläufe				dehn		ung	komp		dehnung
2	dehnverhalten				dehn			komp	1	gedehnt
12										überdehnend
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		überdehnt
1	8				x			x		
2	1				x		x	x		
3	1	x			x		x			

	Lexikoneintrag	Findet			
	dehn #u	1;2;6;7;8;9;10;12;			
	kriech #u	4;			
	messstreifen	1;			
	mess #u	1;			

3 11

Anhang A-4-6

Wortstamm *-dreh-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
15	drehmoment				dreh			komp	1	abdrehen
11	drehraten				dreh			komp	1	abdrehend
9	drehbewegung				dreh			komp	1	abgedreht
8	nennndrehzahl	komp			dreh			komp	2	abzudrehen
6	drehmomente				dreh			komp	1	andrehen
6	drehratensensoren				dreh			komp	1	andrehend
6	drehrichtung				dreh			komp	1	angedreht
6	verdrehwinkel			ver	dreh			komp	3	aufdrehen
5	drehflügler				dreh			komp	1	aufdrehend
5	drehgeschwindigkeit				dreh			komp	1	aufgedreht
5	drehmomenten				dreh			komp	1	aufzudrehen
5	drehsinn				dreh			komp	1	ausdrehen
5	drehwinkel				dreh			komp	1	ausdrehend
4	drehgeschwindigkeiten				dreh			komp	1	ausgedreht
3	drehbewegungen				dreh			komp	1	beidrehen
3	drehmomentmeßwelle				dreh			komp	1	beidrehend
3	drehzahl-				dreh				0	beigedreht
3	verdichterndrehzahl	komp			dreh			komp	2	drehachse
2	-achsen-drehtisch				dreh				0	drehbar
2	ausgangsrehmoment	komp			dreh			komp	2	drehen
2	dreh-				dreh				0	drehend
2	drehfrequenz				dreh			komp	1	dreher
2	drehmatrix				dreh			komp	1	drehimpuls
2	drehmomentes				dreh			komp	1	drehleier
2	drehmomentreserven				dreh			komp	1	drehspiegelung
2	drehmoments				dreh			komp	1	drehstrom
2	drehrate				dreh			komp	1	dreht
2	drehraten-				dreh				0	drehte
2	drehtisch				dreh			komp	1	drehten
2	drehtisches				dreh			komp	1	drehtür
2	drehwinkels				dreh			komp	1	drehung
2	mindestdrehzahl	komp			dreh			komp	2	drehzahl
2	verdrehwinkeln			ver	dreh			komp	3	durchdrehend
33										durchgedreht
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		durchzudrehen
1	23				x			x		eindrehen
2	4	x			x			x		eindrehend
3	2			x	x			x		eingedreht
	29									erdumdrehung
	Lexikoneintrag	Findet								gedreht
	verdichter	18;								herausdrehen
	drehmoment	1;5;11;18;20;24;25;26;								herausgedreht
	dreh #u	2;3;6;7;10;12;13;14;15;19;22;23;27;29;30;31								herdrehen
	verdreh #u	8;33;								herumdrehen
	nenn #u	4;								herumgedreht
	5	28								langdrehend
										langgedreht
										Lexikonliste gekürzt !!

Anhang A-4-7

Wortstamm *-fahr-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
17	fahrwerk				fahr			komp	1	abfahren
16	messverfahren	komp			fahr	en			2	abfahrend
8	fahrwerke				fahr			komp	1	abfahrt
7	fahrwerks				fahr			komp	1	abfahrtshang
6	luftfahrt-								0	abfahrtsläufer
6	prüfverfahren	komp		ver	fahr	en			3	abfahrtsrennen
5	fügeverfahren	komp		ver	fahr	en			3	abgefahren
4	fahrwerken				fahr			komp	1	abschleppfahrer
4	hauptfahrwerk	komp			fahr			komp	4	anfahren
4	messverfahrens	komp		ver	fahr	ens			3	anfahrend
4	panelverfahren	komp		ver	fahr	en			3	anfahrt
3	bugfahrwerk	komp			fahr			komp	4	anfahrtermin
3	fahrwerkes				fahr			komp	1	anfahrverzögerung
3	fahrwerksbein				fahr			komp	1	angefahren
3	fahrwerksgeometrie				fahr			komp	1	anzufahren
3	handlaminierverfahren	komp		ver	fahr	en			3	auffahren
3	hauptfahrwerkes	komp			fahr			komp	4	auffahrend
3	herstellverfahren	komp		ver	fahr	en			3	auffahrt
3	suchverfahren	komp			fahr	en			2	aufgefahren
2	anfah-								0	ausfahren
2	anfahrwirbel			an	fahr			komp	5	ausfahrend
2	anfahrwirbels			an	fahr			komp	5	ausfahrt
2	ausfahrwinkel			aus	fahr			komp	5	ausgefahren
2	dreibeinfahrwerk	komp			fahr			komp	4	austauschverfahren
2	fahrgestell				fahr			komp	1	auswerteverfahren
2	fahrt-								0	auszufahren
2	fahrwerks-								0	autobahnfahrt
2	fahrwerksbeine				fahr			komp	1	autofahren
2	fahrwerksmasse				fahr			komp	1	autofahrt
2	fahrwerksschächte				fahr			komp	1	bankerfahrung
2	fahrwerkstüren				fahr			komp	1	befahrbar
2	fahrwerkssysteme				fahr			komp	1	befahren
2	flugzeugfahrwerken	komp			fahr			komp	4	befahrend
2	gefahren-								0	beifahrer
2	hammesfahr-								0	berufserfahren
2	hauptfahrwerken	komp			fahr			komp	4	berufsradfahrer
2	hauptfahrwerks	komp			fahr			komp	4	betfahrt
2	hauptfahrwerksbeinen	komp			fahr			komp	4	binnenschifffahrtsstreit
2	Injektionsverfahren	komp		ver	fahr	en			3	bobfahren
2	panelverfahrens	komp		ver	fahr	ens			3	bootfahren
2	preformverfahren								0	branchenerfahrener
2	pvdverfahren								0	bühnenerfahrung
2	raumfahrtrückstände	komp			fahr	t			2	busfahrt
2	traglinienverfahren	komp		ver	fahr	en			3	dahinfahren
2	triangulationsverfahren	komp		ver	fahr	en			3	dahinfahrend
2	vefahren								0	dahingefahren
2	-verfahren								0	darüberfahren
2	volumendiskretisierungsverfahren	komp		ver	fahr	en			3	darüberfahrend
48										davonfahren

Fortsetzung Anhang A-4-7

Wortstamm *-fahr-*

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
13	1				x			x	davonfahrend
					x				davongefahren
2	2	x			x	x			dazwischenfahrend
10	3	x		x	x	x			dazwischengefahren
8	4	x			x			x	disziplinarverfahren
5	5			x	x			x	drüberfahren
	Lexikoneintrag	Findet							
	mess #u	2;10;							durchfahrend
	rück #u	43;							durchgefahren
	fahrwerk	1;3;4;8;9;10;12;13;14;15;17;24;28;29;31;32;33;36;37;38;							einfahren
	fahrgestell	25;							einfahrend
	prüf #u	6;							einfahrt
	panel	11;40;							eingefahren
	such #u	19;							eingefahren
	anfah #u	21;22;							einzufahren
	ausfahr #u	23;							entfahren
	triangulation	45;							entgegenfahrend
	herstellverfahren	18;							entgegengefahren
									entlangfahren
									entlangfahrend
									entlanggefahren
									erfahren
									erfahrend
									erfahrenheit
									erfahrung
									erfahrungssache
									fahrbar
									fahrbericht
									fahren
									fahrend
									fahrenheit
									Lexikonliste
									gekürzt !!

11 32

Anhang A-4-8

Wortstamm *-fließ-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
4	fließgrenze				fließ			komp	1	abfließen
2	fließstrecken				fließ			komp	1	abfließend
2	fließverhalten				fließ			komp	1	auffließe
2	fließverhaltens				fließ			komp	1	ausfließen
2	fließwege				fließ			komp	1	ausfließend
5										durchfließend
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		einfließen
1	5				x			x		einfließend
										fließband
	Lexikoneintrag	Findet								fließbänder
	fließ #u	1;2;3;4;5								fließen
	1	5								fließend
										herabfließen
										herabfließend
										herauffließen
										herauffließend
										hindurchfließen
										hindurchfließend
										hinfließen
										hinfließend
										hinunterfließen
										hinunterfließend
										überfließend
										umfließen
										umfließend
										verfließen
										verfließend
										vorbeifließen
										vorbeifließend
										wegfließen
										wegfließend
										weiterfließen
										zerfließend
										zufließen
										zufließend
										zurückfließend
										zusammenfließend

Anhang A-4-9

Wortstamm -grenz-

Häuf.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
30	grenzschicht				grenz			komp	3	abgegrenzt
18	grenzwerte				grenz			komp	3	abgrenzen
16	grenzwert				grenz			komp	3	abgrenzend
14	grenzschichten				grenz			komp	3	abgrenzung
11	grenzwerten				grenz			komp	3	abschaltgrenze
8	grenzfall				grenz			komp	3	abschreibegrenze
8	grenzfläche				grenz			komp	3	abtastgrenze
8	pumpgrenze	komp			grenz	e			1	abzugrenzen
6	grenzschichtverhalten				grenz			komp	3	aneinandergegrenzt
5	fließgrenze	komp			grenz	e			1	aneinandergrenzend
5	grenzschichtparameter				grenz			komp	3	angegrenzt
5	grenzschichtrand				grenz			komp	3	angrenzen
4	grenzfrequenz				grenz			komp	3	angrenzend
4	grenzschichtdicke				grenz			komp	3	ausgegrenzt
4	grenzwertes				grenz			komp	3	ausgrenzen
3	beulgrenze	komp			grenz	e			1	ausgrenzend
3	grenzschichtablösung				grenz			komp	3	ausgrenzung
3	grenzschichtmessungen				grenz			komp	3	begrenzen
3	grenzschichtströmung				grenz			komp	3	begrenzend
3	grenzschichttheorie				grenz			komp	3	begrenzer
3	grenzschichttransition				grenz			komp	3	begrenzt
3	grenzschichtumschlag				grenz			komp	3	begrenztheit
3	grenzschichtzustand				grenz			komp	3	begrenzung
3	profilgrenzschicht	komp			grenz			komp	2	begrenzungsmauer
3	schaufelgrenzschicht	komp			grenz			komp	2	eingegrenzt
2	dehngrenze	komp			grenz	e			1	eingrenzen
2	dosisgrenzwerte	komp			grenz			komp	2	eingrenzend
2	emissionsgrenzwerte	komp			grenz			komp	2	eingrenzung
2	gehäusegrenzschicht	komp			grenz			komp	2	einzugrenzen
2	grenz-								0	gegrenzt
2	grenzfällen				grenz			komp	3	grenzabkommen
2	grenzflächen				grenz			komp	3	grenzabstand
2	grenzkurve				grenz			komp	3	grenzausschuss
2	grenzkurven				grenz			komp	3	grenzbach
2	grenzschichtablösungen				grenz			komp	3	grenzbereich
2	grenzschichtabsaugung				grenz			komp	3	grenzen
2	grenzschichtentwicklung				grenz			komp	3	grenzend
2	grenzschichtgrößen				grenz			komp	3	grenzenlosigkeit
2	grenzschichtparametern				grenz			komp	3	grenzer
2	grenzschichtrechnung				grenz			komp	3	grenzfrage
2	grenztemperatur				grenz			komp	3	grenzgebiet
2	grenztemperaturen				grenz			komp	3	grenzgewässer
2	grenzüberwachung				grenz			komp	3	grenzlandausschuss
2	grenzzyklusschwingungen				grenz			komp	3	grenzoffizier
2	lärmgrenzwerte	komp			grenz			komp	2	grenzpatrouille
2	magerverlöschgrenze	komp			grenz	e			1	grenzpolizist

Fortsetzung Anhang A-4-9

Wortstamm *-grenz-*

2	pumpgrenzabstand	komp			grenz			komp	2	grenzposten
2	sperrgrenze	komp			grenz	e			1	grenzschutzamt
2	streckgrenze	komp			grenz	e			1	grenzschützer
2	wandgrenzschicht	komp			grenz			komp	2	grenzsperre
50										grenzstelle
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		grenzstreifen
7	1	x			x	x				grenzstreitigkeit
8	2	x			x			x		grenzübertritt
34	3				x			x		grenzverkehr
										grenzvertrag
	Lexikoneintrag	Findet								grenzwächter
	grenzschicht	1;4;9;11;12;14;17;18;19;20;22;23;24;25;29;35;37;38;39;40;50;								grenzzoll
	fließ #u	10;								grenzzwischenfall
	dehn #u	26;								hörbarkeitsgrenze
	grenz #u	2;3;5;6;7;13;15;27;28;31;32;33;34;41;42;43;44;45;								hörgrenze
	beul #u	16;								lärmgrenze
	sperr #u	48;								milliardengrenze
	pump #u	8;47;								millionengrenze
	absaugung	36;								rheingrenze
		8	46							schamgrenze
										schmerzgrenze
										steilheitsbegrenzt
										strombegrenzt
										umgrenzend
										umgrenzt
										unbegrenzt
										unbegrenztheit
										untergrenze
										warngrenze

Anhang A-4-10

Wortstamm *-kalibrier-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
3	kalibriermessungen				kalibrier			komp	1	kalibrieren
2	kalibrierkurven				kalibrier			komp	1	kalibrierung
1	durchsatzkalibriereinrichtung									
1	kalibrierdüse									
1	kalibrierende									
1	kalibrierfaktor									
1	kalibrierfaktoren									
1	kalibrierfehlern									
1	kalibrierfunktionen									
1	kalibriergenauigkeit									
1	kalibriergitter									
1	kalibrierkanal									
1	kalibrierkoeffizienten									
1	kalibrierkonstante									
1	kalibriermatrix									
1	kalibriermechanismus									
1	kalibriermesspunkte									
1	kalibriermethode									
1	kalibrierphase									
1	kalibrierpunkte									
1	kalibrierquellen									
1	kalibriersignale									
1	kalibrierstand									
1	kalibrierstrahl									
1	kalibrierstrategie									
1	kalibrierstrecke									
1	kalibrierverfahren									
1	kalibriervorgänge									
1	kalibriervorrichtung									
1	kalibrierzelle									
1	kalibrierzyklus									
1	prüfstandskalibrierung									
1	sondenkalibrierstand									
1	unkalibrierte									

2 -/- 35

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
1	2				x			x	

Lexikoneintrag	Findet
kalibrier #u	33 von 35

Anhang A-4-11

Wortstamm *-kenn-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
23	kenngrößen				kenn			komp	1	aberkennen
23	kennwerte				kenn			komp	1	aberkennend
14	kennfeld				kenn			komp	1	aberkennung
13	kennzahlen				kenn			komp	1	abzuerkennen
13	kennzeichnet				kenn			komp	1	anerkennen
11	kennlinie				kenn			komp	1	anerkennend
9	kennzeichnen				kenn			komp	1	anerkennenswert
8	kennlinien				kenn			komp	1	anerkennnis
7	kennwert				kenn			komp	1	anerkennung
6	ähnlichkeitskennzahlen	komp			kenn			komp	2	anzuerkennen
6	kenngröße				kenn			komp	1	auskennen
5	kenndaten				kenn			komp	1	auskennend
5	kennfelder				kenn			komp	1	balkennummerierung
5	kennfeldes				kenn			komp	1	bekennen
5	kennzeichen				kenn			komp	1	bekennend
5	verdichterkennefeld	komp			kenn			komp	2	bekennerschreiben
4	kennfeldern				kenn			komp	1	bekennnis
4	kennwerten				kenn			komp	1	detaillkenntnis
4	werkstoffkennwerte	komp			kenn			komp	2	einbekennen
3	materialkennwerte	komp			kenn			komp	2	einbekennend
3	werkstoffkennwerten	komp			kenn			komp	2	einbekennung
2	festigkeitskennwerte	komp			kenn			komp	2	erkennbar
2	flugzeugkenngrößen	komp			kenn			komp	2	erkennbarkeit
2	grundkennndaten	komp			kenn			komp	2	erkennen
2	kenn-								0	erkennend
2	kennfeldmessungen				kenn			komp	1	erkenntlich
2	kennfeldwerten				kenn			komp	1	erkenntlichkeit
2	kennwertermittlung				kenn			komp	1	erkenntnis
2	kennwertfunktionen				kenn			komp	1	erkennung
2	kennwerts				kenn			komp	1	erkennungsmelodie
2	kennzahl				kenn			komp	1	gekennzeichnet
2	kennzahldefinitionen				kenn			komp	1	hochbaukenntnisse
2	kennzahlensystem				kenn			komp	1	kennedy
2	kennzahlensystems				kenn			komp	1	kennen
2	personenkenziffern	komp			kenn			komp	2	kennend
2	werkstoffkenngrößen	komp			kenn			komp	2	kennengelernt
2	Werkstoffkennwert- funktionen	komp			kenn			komp	2	kennenlernend
37										kenner
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		kennt
1	25				x			x		kenntlich
2	11	x			x			x		kenntlichkeit
	1									kenntnis
										kenntnisnahme
										kenntnisreich
										kennung
										kennzeichnend
										kennzeichnung

	Lexikoneintrag	Findet		
	kenn #u	alle		

Anhang A-4-12

Wortstamm *-kerb-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
2	kerbformen				kerb			komp	1	ackerbau
2	kerbspannung				kerb			komp	1	eingerkerbt
2	kerbspannungen				kerb			komp	1	einkerben
2	kerbspannungsfaktor				kerb			komp	1	einkerbend
2	kerbtiefen				kerb			komp	1	einkerbung
										gekerbt
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		
1	6				x			x		kerbe
										kerbel
										kerben
										kerbend
										knickerbocker
										zuckerbrotmethode
										zurückerbeten
										zurückerbittend

	Lexikoneintrag	Findet	
	kerb #u	alle	

Anhang A-4-13

Wortstamm *-koppl-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
46	kopplung				koppl		ung		1	ankopplung
7	kopplungen				koppl		ung		1	entkopplung
7	rückkopplung				koppl		ung		1	koppler
5	einkopplung			ein	koppl		ung		2	mitkopplung
3	kopplungseffekte				koppl		ung	komp	4	optokoppler
3	kopplungsmechanismen				koppl		ung	komp	4	
2	kopplungsalgorithmus				koppl		ung	komp	4	
2	kopplungsfläche				koppl		ung	komp	4	
2	kopplungsprozesse				koppl		ung	komp	4	
2	rückkopplungen	komp			koppl		ung		3	
2	torsionskopplung	komp			koppl		ung		3	
2	verkopplung			ver	koppl		ung		2	

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
1	3				x		x		
2	2			x	x		x		
3	2	x			x		x		
4	5				x		x	x	

	Lexikoneintrag	Findet		
	kopplung	1;2;4;5;6;7;8;9;11		
	rück #u	3;10;		

Anhang A-4-14

Wortstamm *-last-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
39	nutzlast	komp			last				1	achslast
15	nutzlasten	komp			last	en			2	altlast
10	aeroelastische	komp			elast		isch		3	angelastet
10	aeroelastischen	komp			elast		isch		3	anlasten
9	nutzlastmasse	komp			last			komp	4	anlastend
8	aeroelastik	komp			elast		ik		3	aufgelastet
8	auslastung			aus	last		ung		5	auflastend
4	aeroelastischer	komp			elast		isch		3	aufzulasten
4	beullast	komp			last				1	ausgelastet
4	impactbelastung	komp		be	last		ung		0	auslasten
4	last-								0	auslastend
4	nutzlastkapazität	komp			last			komp	4	ballast
3	belastungszustände			be	last		ung	komp	5	baulast
3	duroplastischen	komp			plast		isch		3	belastbar
3	lastaufbringung				last			komp	6	belastbarkeit
3	nutzlastanteil	komp			last			komp	4	belasten
3	nutzlastverkleidung	komp			last			komp	4	belastend
3	sekundärnutzlast	komp			last				1	belastet
3	thermoplaste	komp			plast	e			2	belastung
3	thermoplasten	komp			plast	en			2	blastogenese
3	überlastung	komp			last		ung		3	bürdelast
2	aeroelastie	komp			elast		ic		0	eingeglast
2	aeroelastisch	komp			elast		isch		3	elastik
2	belastungs-			be	last		ung		0	elastisch
2	duroplastische	komp			plast		isch		3	elastizität
2	elastomers				last				0	entlasten
2	entlastungs-				last				0	entlastend
2	inelastischen			in	elast		isch		5	entlastet
2	lasteinleitungs-				last				0	entlastung
2	lastumverteilung				last			komp	6	fernlastzug
2	lastzuständen				last			komp	6	folgelast
2	nennlast	komp			last				1	geglast
2	nutzlast-				last				0	gelastet
2	nutzlastfunktion	komp			last			komp	4	gepflastert
2	nutzlastverhältnisse	komp			last			komp	4	glasteile
2	nutzlastvolumen	komp			last			komp	4	gouverneurspalast
2	thermoplast	komp			plast				1	kongresspalast
2	viskoelastische	komp			elast		isch		3	königspalast
2	zuglast	komp			last				1	kopflastig
2	zuglasten	komp			last	en			2	last
40										lasten
										lastend

Fortsetzung Anhang A-4-14

Wortstamm *-last-*

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		
6	1	x			x					laster
4	2	x			x	x				lasterhaft
9	3	x			x		x			lastfrei
8	4	x			x			x		lastzug
3	5			x	x		x			leukoplast
3	6				x			x		linkslastig
8	0									linkslastigkeit
										mitbelastet
										nervenbelastung
	Lexikoneintrag	Findet								palast
	nutz #u	1;2;5;12;16;17;18;34;35;36								pflaster
	nenn #u	32;								pflastern
	beul #u	9;								pflasterstein
	aufbringung	15;								pilaster
	aero #u	3;4;6;8;23;								plaste
	auslastung	7;								plastik
	duro #u	14;25;								plastikboot
	thermo #u	19;20;								plastikeimer
	überlastung	21;								plastikkarte
	zuglast	39;40;								plastikschirm
	umverteilung	30								plastisch
		11	27							plastizität
										rechtslastig
										scholastiker
										schwanzlastig
										sonderbelastung
										soziallasten
										thermoplastisch
										überlastend
										überlastet
										überlastfest
										überlastgeschützt
										unbelastet
										verglast
										vorbelastung
										wechsellastfestigkeit
										weiterbelastet
										zugepflastert
										zulasten

elastisch und plaste sind bereits lexikalisiert

Anhang A-4-15

Wortstamm *-misch-*

Häuf..	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
78	aerodynamischen	komp			dynamisch	en				abgemischt
69	aerodynamische	komp			dynamisch	e				abmischen
61	dynamische				dynamisch	e				akademisch
24	aerodynamischer	komp			dynamisch	er				algorithmisch
19	aerodynamisch	komp			dynamisch					anämisch
6	aerodynamisches	komp			dynamisch	es				anatomisch
3	elektrochemische	komp			chemisch	e				arhythmisch
3	vormischverbrennung			vor	misch			komp	3	astronomisch
3	zumischung			zu	misch		ung		2	beigemischt
2	aerodynamischem	komp			dynamisch	em				beimischen
2	aerothermodynamische	komp			dynamisch	e				beimischend
2	aerothermodynamischen	komp			dynamisch	en				beizumischen
2	elektrochemischen	komp			chemisch	en				binomisch
2	Magnetoplasma- dynamischen									blasphemisch
2	mischluft				misch			komp	4	blutmischung
2	piezokeramische	komp			keramisch	e				böhmisch
2	piezokeramischer	komp			keramisch	er				bremisch
2	vorgemischte		vor	ge	misch	t			1	chemisch
2	vormischstrecke			vor	misch			komp	3	durchmischen
2	zugemischt		zu	ge	misch	t			1	dynamisch

20

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		
1	2		x	x	x	x				einmischen
2	1			x	x		x			einmischend
3	2			x	x			x		einmischung
4	1				x			x		einzumischen
										epidemisch

	Lexikoneintrag	Findet								
	aero #u	1;2;4;5;6;10;11;12;								ergonomisch
	misch #u	15;								filmisch
	elektro #u	13;								flämisch
	piezo #u	16;17;								garmisch
	vorgemischt	18;								gastronomisch
	zugemischt	20;								gemisch

6 14

heimisch
hydrodynamisch
islamisch
kältemischung
keramisch
komisch
kosmisch
logarithmisch
mimisch
mischanlage
**Lexikonliste
gekürzt !!**

Anhang A-4-16

Wortstamm *-nick-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
9	nick-				nick				0	abgeknickt
9	nickmoment				nick			komp	1	abknicken
8	nickachse				nick			komp	1	abknickend
6	nickbewegung				nick			komp	1	eingeknickt
6	nickwinkel				nick			komp	1	einknicken
5	nicklage				nick			komp	1	einknickend
3	nickdynamik				nick			komp	1	geknickst
3	nickel-				nick				0	geknickt
3	nickelbasislegierung				nick			komp	1	genick
3	nickmomentenbeiwert				nick			komp	1	genickbruch
3	nickrate				nick			komp	1	genickschuss
2	nick				nick				0	genickt
2	nickbewegungen				nick			komp	1	karnickel
2	nickdämpfung				nick			komp	1	knick
2	nickelbasis-				nick				0	knicken
2	nicksteuerung				nick			komp	1	knickend
2	nickwinkels				nick			komp	1	knickerbocker
17										knickerig
										knickrig
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		knicks
1	13				x			x		knicksen
										knicksend
										knickt
	Lexikoneintrag	Findet								knickte
	nick #u	2;3;4;5;6;7;10;11;13;14;16;17;								knickten
		1	12							köpenick
										nickel
										nicken
										nickend
										nicki
										picknick
										pumpernickel
										schnickschnack
										schnickschnacklos
										umgeknickt
										umknicken
										umknickend
										vernickeln
										zerknicken
										zerknickt
										zugenickt
										zunicken

Anhang A-4-17

Wortstamm *-nutz-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
39	nutzlast				nutz			komp	1	abgenutzt
15	nutzlasten				nutz			komp	1	abnutzen
9	nutzlastmasse				nutz			komp	1	abnutzend
4	nutzlastkapazität				nutz			komp	1	abnutzung
3	nutzlastanteil				nutz			komp	1	ausgenutzt
3	nutzlastverkleidung				nutz			komp	1	ausnutzen
3	nutzsignal				nutz			komp	1	ausnutzend
3	sekundärnutzlast				nutz			komp	1	ausnutzung
2	benutzer-				nutz				0	auszunutzen
2	nutzbringend				nutz			komp	1	benutzbar
2	nutzeffekt				nutz			komp	1	benutzbarkeit
2	nutzer-				nutz				0	benutzen
2	nutzlast-				nutz				0	benutzend
2	nutzlastfunktion				nutz			komp	1	benutzer
2	nutzlastverhältnisse				nutz			komp	1	benutzerdefiniert
2	nutzlastvolumen				nutz			komp	1	benutzerebene
2	nutzsignals				nutz			komp	1	benutzereigen
17										benutzerorientiert
										benutzt
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		benutzung
1	17				x			x		endbenutzer
										genutzt
	Lexikoneintrag	Findet								innennutzung
	nutz #u	alle								mitbenutzen
		1	14							mitbenutzend
										mitbenutzt
										mitbenutzung
										mitzubenutzen
										nichtsnutzig
										nießnutzer
										nutzbar
										nutzbarkeit
										nutzen
										nutzend
										nutzer
										nutzlos
										nutzlosigkeit
										nutzmachung
										nutznießen
										nutznießler
										nutzung
										nutzwild
										unbenutzt
										ungenutzt
										zunutze

Anhang A-4-18

Wortstamm *-press-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
17	pressure								0	abgepresst
8	inkompressiblen			in	kompres		ible		1	abpressen
6	inkompressible			in	kompres		ible		1	abpressend
5	inkompressibel			in	kompres		ible		1	akupressur
5	kompressible				kompres		ible		3	anpressdruck
4	inkompressibler			in	kompres		ible		1	anpressen
3	kompressiblen				kompres		ible		3	anpressend
3	kompressibler				kompres		ible		3	aufgepresst
2	anpresskraft			an	press			komp	2	ausgepresst
2	inkompressibilität			in	kompres		ibilität		1	auspressen
2	kompressibel				kompres		ible		3	auspressend
2	suppression								0	bahnexpress
2	verpresst			ver	press	t			1	computerpresse
13										dekompression
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		depression
1	5			x	x		x			depressiv
2	1			x	x			x		durchpressen
3	4				x		x			durchpressend
4	1			x	x	x				einpressen
										einpressend
										erpressbarkeit
										erpressen
										erpressend
										erpresser
										erpresserisch
										erpressung
										espresso
										express
										expressionismus
										expressionist
										expressionistisch
										expressis
										expressiv
										freipressend
										fruchtsaftpresserei
										gepresst
										hinauspressen
										hineinpressen
										impression
										impressionismus
										impressionist
										impressionistisch
										impressiv
										impresum
										Lexikonliste gekürzt !!

7 11

Anhang A-4-19

Wortstamm *-prüf-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
25	prüfstand				prüf			komp	1	durchgeprüft
9	prüfstandes				prüf			komp	1	durchprüfen
8	prüfstands				prüf			komp	1	durchprüfend
6	prüfkörper				prüf			komp	1	Facharbeiter- prüfung
6	prüfständen				prüf			komp	1	geprüft
6	prüfverfahren				prüf			komp	1	leidgeprüft
5	höhenprüfstand	komp			prüf			komp	2	nachgeprüft
5	prüfmaschine				prüf			komp	1	nachprüfbar
5	verdichterprüfstand	komp			prüf			komp	2	nachprüfen
4	prüfstände				prüf			komp	1	nachprüfend
4	zugprüfmaschine	komp			prüf			komp	2	nachprüfung
3	prüfstandsversuche				prüf			komp	1	nachzuprüfen
3	prüftechnik				prüf			komp	1	prüfen
2	axialverdichterprüfstand	komp			prüf			komp	2	prüfend
2	brennkammerprüfstand	komp			prüf			komp	2	prüfer
2	crashprüfstand	komp			prüf			komp	2	prüfling
2	Hochgeschwindigkeits- prüfeinheit								0	prüfung
2	Hochgeschwindigkeits- prüfmaschine								0	Rechtschreib- prüfung
2	komponentenprüfstand	komp			prüf			komp	2	risikoprüfung
2	prüf-								0	schulprüfung
2	prüfanforderungen				prüf			komp	1	überprüfbar
2	prüfdruck				prüf			komp	1	überprüfend
2	prüfeinheit				prüf			komp	1	überprüft
2	prüfeinrichtung				prüf			komp	1	überprüfung
2	prüfgestell				prüf			komp	1	vorprüfung
2	prüfkomponenten				prüf			komp	1	wahlprüfer
2	prüfkonzept				prüf			komp	1	
2	prüfkraft				prüf			komp	1	
2	prüfmaschinen				prüf			komp	1	
2	prüfmethoden				prüf			komp	1	
2	prüfstandsaufbau				prüf			komp	1	
2	prüfstandseinrichtungen				prüf			komp	1	
2	prüfstandskomponenten				prüf			komp	1	
2	prüfsystem				prüf			komp	1	
2	prüfsystems				prüf			komp	1	
2	prüfzylinder				prüf			komp	1	
2	triebwerksprüfstand	komp			prüf			komp	2	
2	verdichterprüfstandes	komp			prüf			komp	2	
2	verschleissprüfstand								0	
2	vprüf								0	

40

Fortsetzung Anhang A-4-19

Wortstamm *-prüf-*

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
1	26				x			x	
2	9	x			x			x	

	Lexikoneintrag	Findet							
	prüf #u	alle außer 11;14;18;20;39;40							
	axial	14;							

2

35

Anhang A-4-20

Wortstamm *-roll-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
13	roll-								0	abgerollt
11	rolls-royce								0	abrollen
11	rollwinkel				roll			komp	1	abrollend
9	roll				roll				0	abrollwiderstand
7	rollmoment				roll			komp	1	angerollt
6	rollachse				roll			komp	1	anrollen
6	rollrate				roll			komp	1	anrollend
6	rollsteuerung				roll			komp	1	aufgerollt
5	kontrollvolumen				kontroll			komp	1	aufrollen
4	controlling				control		ling		5	aufrollend
4	kontrollpunkte				kontroll			komp	1	aufzurollen
4	kontrollsystem				kontroll			komp	1	ausgerollt
4	rollage								0	ausrollen
4	rollbewegung				roll			komp	1	ausrollend
4	rolls								0	controller
4	rolls-								0	drollig
3	kontroll-								0	eingerollt
3	kontrollflächen				kontroll			komp	1	einrollen
3	kontrollpunkt				kontroll			komp	1	einrollend
3	kontrollraum				kontroll			komp	1	entrollen
3	rollbahnen				roll			komp	1	fernkontrolliert
3	rollbewegungen				roll			komp	1	gepäckkontrolle
3	rollmomente				roll			komp	1	gerollt
3	rollmoments				roll			komp	1	groll
3	rollraten				roll			komp	1	grollen
2	abrollwege			ab	roll			komp	4	hauptrolle
2	bodenkontrollstation	komp			kontroll			komp	3	hechtrolle
2	bodenkontrollstationen	komp			kontroll			komp	3	heraufgerollt
2	controlled								0	heraufrollen
2	kontrolleinheit				kontroll			komp	1	heraufrollend
2	kontrollkräfte				kontroll			komp	1	heruntergerollt
2	kontrollmaßnahmen				kontroll			komp	1	herunterrollen
2	kontrollmechanismen				kontroll			komp	1	herunterrollend
2	kontrollmöglichkeiten				kontroll			komp	1	kasserolle
2	kontrollpunkten				kontroll			komp	1	kontrollaktion
2	kontrollrechnungen				kontroll			komp	1	kontrollamt
2	microcontroller								0	kontrolle
2	mikrocontroller								0	kontrolleur
2	passkontrolle	komp			kontroll				2	kontrollfluss
2	rollbahn				roll			komp	1	kontrollflussanalyse
2	rollage								0	kontrollgewalt
2	rollmanöver				roll			komp	1	kontrollierbar
2	rollmomentenbeiwert				roll			komp	1	kontrollierbarkeit
2	rollmomentes				roll			komp	1	kontrollieren
2	rollsteuerbedarf				roll			komp	1	kontrollierend
2	rollverkehr				roll			komp	1	kontrolliert

Fortsetzung Anhang A-4-20

Wortstamm *-roll-*

2	rollwiderstand				roll			komp	1	kontrollrat
2	rollwinkels				roll			komp	1	kontrolltechnik
2	rollzeiten				roll			komp	1	kontrollzentrum
2	umlenkrollen				roll	en				kopfkontrolle
50										korollar
										losrollen
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		
33	1				x			x		messdatenkontrolle
1	2	x			x					nachkontrollieren
2	3	x			x			x		nachkontrollierend
1	4			x	x			x		nachkontrolliert
1	5				x		x			rechtschreibkontrolle
1	6				x	x				rolladen
10	0									rolle
										rollen
										rollend
Lexikoneintrag	Findet									
rolls-royce #e	2;									rollenkern
pass	39;									roller
roll #u	3;5;6;7;8;9;11;12;14;18;19;20;21;22;23;24;25;27;28;30;31;32;									rolli
	33;34;35;36;40;41;42;43;44;45;46;47;48;49									rolling
kontroll #u	9;11;12;18;19;20;27;28;30;31;32;33;34;35;36;									rollkragenpullover
										rollladen
										rollstuhl
										scrollen
										sprechrolle
										staatskontrolle
										statistenrolle
										troll
										trollen
										überrollend
										überrollt
										unkontrolliert
										videocontroller
										vormachtrolle
										weggerollt
										wegkontrolle
										wegrollen
										wegrollend
										zeilenvorschubkontrolle
										zeitkontrolle
										zugerollt
										zurollen
										zurollend
										zurückgerollt
										zurückrollen
										zurückrollend
										zusammengerollt
										zusammenrollen
										Lexikonliste gekürzt !!

Anhang A-4-21

Wortstamm *-rück-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
14	rückschlüsse				rück			komp	1	abdrücken
10	rückseite				rück			komp	1	abdrückend
8	rückführung				rück			komp	1	abgedrückt
7	rückkopplung				rück			komp	1	abgerückt
5	rückströmung				rück			komp	1	abrücken
5	rückwirkung				rück			komp	1	abrückend
4	rückgang				rück			komp	1	abzudrücken
4	rückstellkraft				rück			komp	1	abzurücken
4	rückwirkungen				rück			komp	1	angerückt
4	überbrückung	komp			brück		ung		2	anrücken
3	berück								0	anrückend
3	flammenrückschlag	komp			rück			komp	3	aufdrücken
3	rückstellmoment				rück			komp	1	aufdrückend
2	berücksichtigung			be	rück			komp	0	aufgedrückt
2	brennkammerdrücken	komp			drück	en			0	aufgerückt
2	brückner								0	aufrücken
2	dickenrücklage	komp			rück			komp	3	aufzudrücken
2	messbrücken	komp			brück	en			0	aufzurücken
2	raumfahrtrückstände	komp			rück			komp	3	ausdrücken
2	rückblick				rück			komp	1	ausdrückend
2	rücken-								0	ausdrücklich
2	rückgangs				rück			komp	1	ausdrückstift
2	rückgewinnung				rück			komp	1	ausgedrückt
2	rückkopplungen				rück			komp	1	ausgerückt
2	rückmeldungen				rück			komp	1	ausrücken
2	rückschlag				rück			komp	1	ausrückend
2	rückschluss				rück			komp	1	auszudrücken
2	rückstände				rück			komp	1	autobahnbrücke
2	rückstellmechanismus				rück			komp	1	autobrücke
2	rückstromgebietes				rück			komp	1	bedrücken
2	rücktransformationen				rück			komp	1	bedrückend
2	rückverfolgbarkeit				rück			komp	1	bedrückt
2	rückwärtspeilung				rück			komp	1	bedrückung
33										berücksichtigen
										berücksichtigend

Fortsetzung Anhang A-4-21

Wortstamm *-rück-*

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
23	1				x			x	berücksichtigt
1	2	x			x		x		berücksichtigung
3	3	x			x			x	brücke
2	5	x			x	x			brückenbaupolier
									brückenprojekt
									draufdrücken
									drückeberger
	Lexikoneintrag	Findet							drücken
	überbrückung	10;							drückend
	messbrücke	18;							drücker
	rück #w	2;3;4;5;6;7;9;12;17;19;20;22;23;24;25;26;28;30;31;							drückt
	stell #w	8;13;29;							drückte
	pfeilung	33;							drückten
	schluss	27;							durchdrückend
		6	26						durchgedrückt
									durchzudrücken
									eindrücken
									eindrückend
									eindrücklich
									eingedrückt
									ingerückt
									einrücken
									einrückend
									einrückung
									einzu drücken
									entrücken
									entrückend
									entrückt
									entrückung
									erdrücken
									erektorbrücke
									exportrückgang
									flachdrücken
									flachdrückend
									flachgedrückt
									flussbrücke
									gedrückt
									gedrücktheit
									gerückt
									geschäftsrückgang
									handrücken
									hängebrücke
									herabdrücken
									herabgedrückt
									herangerückt
									Lexikonliste
									gekürzt !!

Anhang A-4-22

Wortstamm *-saug-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
2	grenzschichtabsaugung	komp		ab	saug		ung		1	abgesaugt
2	profilsaugseite	komp			saug			komp	2	absaugen
7	saug-				saug				0	absaugend
19	saugseite				saug			komp	3	angesaugt
2	saugseiten				saug			komp	3	ansaugen
2	saugseitig				saug			komp	3	ansaugend
4	saugseitige				saug			komp	3	argusaugen
3	saugseitigen				saug			komp	3	aufgesaugt
6	saugspitze				saug			komp	3	aufsaugen
3	saugspitzen				saug			komp	3	aufsaugend
2	saugwindkanal				saug			komp	3	ausgesaugt
2	statorsaugseite	komp			saug			komp	2	aussaugen
										aussaugend
										blutsaugerei
										durchgesaugt
										durchsaugen
										durchsaugend
										ingesaugt
										einsaugen
										einsaugend
										einzaugen
										gesaugt
										saugbohnern
										saugen
										saugend
										sauger
										saugfähig
										saugfähigkeit
										saugflasche
										saugkraft
										saugkräftig
										saugnapf
										saugpapier
										saugpumpe
										saugrohr
										saugwirkung
										staubgesaugt
										staubsaugend

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
1		1	x		x	x		x	
2		2	x			x			x
3		8				x			x

Lexikoneintrag	Findet
saug #u	2;4;5;9;10;11
absaugung	1;
saugseitig	6;7;8;
stator	12;

Anhang A-4-23

Wortstamm *-schalt-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
3	mikroschalter	komp			schalt	er			1	abgeschaltet
2	schaltflächen				schalt			komp	2	abschalten
2	schaltschwelle				schalt			komp	2	abschaltend
2	schaltventil				schalt			komp	2	abschaltgrenze
3	vorgeschaltet		vor	ge	schalt	et			3	abschaltrelais
3	vorgeschalteten		vor	ge	schalt	eten			3	abschaltsystem
										abschaltung
										abschaltvorgang
										abzuschalten
										angeschaltet
										anschalten
										anschaltend
										ausgeschalt
										ausgeschaltet
										ausschaltend
										ausschalter
										ausschaltfunktion
										ausschaltung
										auszuschalten
										beschalten
										beschaltung
										blockschaltbild
										dazugeschaltet
										doppelventilbeschaltung
										durchgeschaltet
										durchzuschalten
										eingeschaltet
										einschalten
										einschaltend
										einschaltung
										einzuschalten
										ersatzschaltbild
										fangschaltung
										freigeschaltet
										freischaltung
										freizuschalten
										geschalt
										geschaltet
										gleichgeschaltet
										gleichschalten
										gleichschaltend
										gleichschaltung
										gleichzuschalten
										hergeschaltet
										herzuschalten
										innenbeschaltung
										Lexikonliste gekürzt !!

3 6

Anhang A-4-24

Wortstamm *-scher-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
24	aerodynamischer									abgeschert
9	scherschicht				scher			komp	1	abgezwitchert
5	scherschichten				scher			komp	1	abscheren
4	aeroelastischer								0	abscherend
4	transsonischer								0	abscherung
3	scherfestigkeit				scher			komp	1	abzwitchern
3	scherkräfte				scher			komp	1	abzwitchernd
3	scherrate				scher			komp	1	äscher
3	wärmetauschers								0	äschern
2	axialsymmetrischer								0	äschernd
2	elektrooptischer								0	ausgeschert
2	piezokeramischer								0	ausrutscher
2	prismatischer								0	ausscheren
2	probabilistischer								0	ausscherend
2	scher-								0	beherrscher
2	scherbeanspruchung				scher			komp	1	bescheren
2	scherraten				scher			komp	1	bescherend
2	scherung				scher		ung		2	beschert
2	scherwind				scher			komp	1	bescherung
2	scherwinde				scher			komp	1	dolmetscher
2	stöchiometrischer								0	drescher
2	thermoakustischer								0	durstlöscher
2	windscherungen	komp			scher		ung		3	effekthascherei
23										einäschern
										einäschern
										einäschern
										eingeschert
										eingeschert
										einscheren
										einscherend
										erforscher
										escher
										fälscher
										fischer
										fischerdorf
										fischerei
										fischerin
										fischerring
										fleischerei
										fleischern
										forscher
										fortgeschert
										fortscheren
										fortscherend
										gartenschere
										Lexikonliste gekürzt !!

4 13

Anhang A-4-25

Wortstamm *-steig-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
9	steig-				steig					absteige
7	steigflug				steig			komp	1	absteigen
2	steigfluges				steig			komp	1	absteigend
2	steiggeschwindigkeit				steig			komp	1	absteiger
2	steigleistung				steig			komp	1	abzusteigen
2	steigvermögen				steig			komp	1	aktivitätssteigerung
2	steigwinkel				steig			komp	1	ansteigen
7										ansteigend
										anzusteigen
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		aufsteigen
1	6				x			x		aufsteigend
										aufsteiger
	Lexikoneintrag	Findet								aufzusteigen
	steig #u	alle								aussteigen
										aussteigend
										aussteiger
										aussteigewarnung
										bergbesteigung
										bergsteiger
										bergsteigerei
										besteigen
										besteigend
										besteigung
										durchsteigend
										durchsteiger
										einsteigen
										einsteigend
										einsteiger
										einzusteigen
										elektrosteiger
										emporsteigen
										emporsteigend
										entsteigen
										entsteigend
										erstbesteigung
										ersteigen
										ersteigern
										ersteigernd
										ersteigert
										ersteigerung
										exportsteigerung
										gasteig
										gehsteig
										gesteigert
										Lexikonliste
										gekürzt !!

Anhang A-4-26

Wortstamm -stell-

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
13	stützstellen	komp			stell	en			1	abbestellen
12	meßstellen	komp			stell	en			1	abbestellend
11	anstell-								0	abbestellt
6	feststellbar	komp			stell		bar		2	abbestellung
6	herstellkosten			her	stell			komp	3	abgestellt
5	messstelle	komp			stell	e			1	abschussstelle
5	verstellbaren			ver	stell		bar		4	abstellen
4	herstellungs-								0	abstellend
4	rückstellkraft	komp			stell			komp	5	abstellfläche
4	v-stellung								0	abstellgleis
3	einstellwinkel			ein	stell			komp	3	abstellkammer
3	herstellbar			her	stell		bar		4	abstellplatz
3	herstellprozess			her	stell			komp	3	abstellraum
3	herstellprozesse			her	stell			komp	3	abstelltisch
3	herstellverfahren			her	stell			komp	3	abstellung
3	rückstellmoment	komp			stell			komp	5	abstellzeit
3	stellodynamik				stell			komp	6	abwehrdienststelle
3	stellflächen				stell			komp	6	abzubestellen
3	stellgenauigkeit				stell			komp	6	abzustellen
3	stellkräfte				stell			komp	6	alleingestelltsein
3	stützstelle	komp			stell	e			1	angestellt
3	trennstellen	komp			stell	en			1	angestellte
3	verstellbarer			ver	stell		bar		4	angestellten- krankenkasse
3	verstellbarkeit			ver	stell		bar keit		4	anheimgestellt
3	verstellmechanismus			ver	stell			komp	3	anheimstellend
3	verstellpropellern			ver	stell			komp	3	ankerstellung
2	anstellung								0	anlaufstelle
2	constellation								0	anlegestelle
2	constellations								0	anstelle
2	darstellbarkeit			dar	stell		bar keit		4	anstellen
2	einstell-								0	anstellend
2	einstellparameter			ein	stell			komp	3	anstellig
2	fahrgestell	komp		ge	stell				7	anstellung
2	feststellmechanismus	komp			stell			komp	5	anstellwinkel
2	herstellerseitig			her	stell	er		komp	8	anzustellen
2	herstellprozesses			her	stell			komp	3	aufgestellt
2	herstellzulassung			her	stell			komp	3	aufstellen
2	isoliniendarstellung	komp		dar	stell		ung		7	aufstellend
2	prüfgestell	komp		ge	stell				7	aufsteller
2	rückstellmechanismus	komp			stell			komp	5	aufstellung
2	stell-								0	aufzustellen
2	stellkommandos				stell			komp	6	ausbaugestell
2	stellkraft				stell			komp	6	ausbaustellung

Fortsetzung Anhang A-4-26

Wortstamm *-stell-*

2	stellsystem				stell			komp	6	ausgabestelle
2	stellsysteme				stell			komp	6	ausgestellt
2	stellsystemen				stell			komp	6	ausstellen
2	verstellbare			ver	stell		bar		4	ausstellend
2	verstellbarem			ver	stell		bar		4	aussteller
48										ausstellerguppe
										ausstellfenster
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		ausstellung
1	5	x			x	x				ausstellungsmöglichkeit
2	1	x			x		x			ausstellung
3	10			x	x			x		ausweichstelle
4	7			x	x		x			auszustellen
5	4	x			x			x		autoausstellung
6	9				x			x		bankstelle
7	3	x		x	x					beigestellt
8	1			x	x	x		x		beistellen
0	8									beistellnetzteil
										beistellung
	Lexikoneintrag	Findet								beizustellen
	herstellbar	12;								bereitgestellt
	mess #u	6;								bereitstellend
	fahrgestell	33;								bereitstellung
	herstellprozess	13;14;36;								bereitzustellen
	herstellverfahren	15;								bestelländerung
	prüf #u	39;								bestellbedingung
	stell #u	4;5;11;17;18;19;20;32;34;37;42;43;44;45;46;								bestellbezeichnung
	rück #u	9;16;40;								bestellbuch
	stütz #u	1;21;								bestelleingang
	meß #u	2;								bestellen
	verstellbar	7;23;47;48;								bestellend
	verstellbarkeit	24;								besteller
	darstellbarkeit	30;								bestellformular
	verstell #u	25;26;								bestellliste
	trenn #u	22;								bestellnr
		15	38							bestellnummer
										bestellschein
										bestellt
										bestellzettel
										bewerkstelligen
										bewerkstelligend
										bewerkstelligt
										bittsteller
										bloßgestellt
										Lexikonliste gekürzt
										!!

Anhang A-4-27

Wortstamm *-ström-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
31	<i>anströmung</i>		an		ström		ung		1	abgeströmt
28	<i>umströmung</i>		um		ström		ung		1	abströmen
16	<i>anströmgeschwindigkeit</i>		an		ström			komp	2	abströmend
15	<i>zuströmung</i>		zu		ström		ung		1	ausgeströmt
11	<i>strömungs-</i>				ström					ausströmen
10	<i>strömungslöser</i>				ström		ung	komp	3	ausströmend
9	<i>angeströmt</i>		an	ge	ström	t			4	ausströmung
7	<i>angeströmten</i>		an	ge	ström	en			4	durchströmend
6	<i>abströmwinkel</i>		ab		ström			komp	2	durchströmt
6	<i>umströmten</i>		um		ström	ten			5	eingeströmt
5	<i>anströmbedingungen</i>		an		ström			komp	2	einströmen
5	<i>rückströmung</i>	rück			ström		ung		6	einströmend
5	<i>strömungsvorgänge</i>				ström		ung	komp	3	entströmen
5	<i>zuströmgeschwindigkeit</i>		zu		ström			komp	2	entströmend
4	<i>abströmung</i>		ab		ström		ung		1	entströmt
4	<i>fehlansströmung</i>	fehl		an	ström		ung		7	geströmt
4	<i>strömungslösers</i>				ström		ung	komp	3	herbeigeströmt
4	<i>strömungsvorgängen</i>				ström		ung	komp	3	herbeiströmend
3	<i>anströmrichtung</i>		an		ström			komp	2	hineinströmen
3	<i>anströmwinkel</i>		an		ström			komp	2	hingeströmt
3	<i>grenzschichtströmung</i>	komp			ström		ung		8	hinströmend
3	<i>profilumströmung</i>	komp			ström		ung		8	hinüberströmen
3	<i>strömungszustände</i>				ström		ung	komp	3	strömen
3	<i>strömungszuständen</i>				ström		ung	komp	3	strömend
3	<i>umströmt</i>		um		ström	t			5	strömung
3	<i>umströmungslärm</i>		um		ström		ung	komp	10	übergeströmt
3	<i>verdichterströmung</i>	komp			ström		ung		8	überströmend
3	<i>zuströmbedingungen</i>		zu		ström			komp	2	warmluftströmung
3	<i>zuströmmachzahl</i>		zu		ström			komp	2	zugeströmt
3	<i>zuströmwinkel</i>		zu		ström			komp	2	zurückgeströmt
2	<i>anströmbedingung</i>		an		ström			komp	2	zurückströmen
2	<i>anströmmachzahl</i>		an		ström			komp	2	zurückströmend
2	<i>anströmmzustands</i>		an		ström			komp	2	zusammengeströmt
2	<i>flügelumströmung</i>	komp		um	ström		ung		9	zusammenströmend
2	<i>heißgasströmung</i>	komp			ström		ung		8	zuströmen
2	<i>spoilerumströmung</i>	komp		um	ström		ung		9	zuströmend
2	<i>strömungsabl</i>				ström					
2	<i>strömungskanäle</i>				ström		ung	komp	3	
2	<i>strömungsmachzahlen</i>				ström		ung	komp	3	
2	<i>Turbomaschinenströmungen</i>	komp			ström		ung		8	
2	<i>überschallanströmung</i>	komp		an	ström		ung		9	
2	<i>umströmter</i>		um		ström	ter			5	
2	<i>Zuströmgeschwindigkeiten</i>		zu		ström			komp	2	
2	<i>zuströmmachzahlen</i>		zu		ström			komp	2	
2	<i>zuströmrichtung</i>		zu		ström			komp	2	

Fortsetzung Anhang A-4-27

Wortstamm *-ström-*

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
1	4			x	x		x		
2	15			x	x			x	
3	6				x		x	x	
4	2		x	x	x	x			
5	2			x	x	x			
6	1	x			x		x		
7	1	x		x	x		x		
8	4	x			x	x	x		
9	3	x		x	x		x		
10	1			x	x		x	x	

	Lexikoneintrag	Findet						
	abgeströmt							
	abströmen							
	abströmung	14;						
	anströmen							
	anströmung	1;40						
	angeströmt	7;8						
	umströmen	10;24;						
	umströmt	41;						
	umströmung	2;21;25;33;35;						
	zuströmen							
	zuströmung	4;						
	zugeströmt							
	abström #u	9;						
	anström #u	11;18;19;30;31;32						
	zuström #u	13;27;28;29;42;43						
	grenzschicht	20;						
	verdichter	26;						
	heißgas	34;						
	machzahl	38;						
	turbo #u	39;						
	fehl #u	15;						
	rück #u	12;						

Total 23

Total 36

Anhang A-4-28

Wortstamm *-such-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
6	vorversuchen		vor	ver	such	en			1	abgesucht
5	such									absuchen
5	vorversuche		vor	ver	such	e			1	absuchend
4	versuchs-									abzusuchen
4	vorversuch		vor	ver	such				2	angesucht
3	prüfstandsversuche	komp		ver	such	e			1	ansuchen
3	such-									ansuchend
3	suchrichtung				such			komp	3	attentatsversuch
3	suchverfahren				such			komp	3	aufgesucht
3	versuchsflügen			ver	such			komp	4	aufsuchen
3	versuchsverdichters			ver	such			komp	4	aufsuchend
2	Impact- untersuchungen	komp			such		ung		5	aufsuchung
2	impactversuche	komp		ver	such				6	aufzusuchen
2	rigversuch	komp		ver	such				6	ausgesucht
2	Schnellzerreiß- versuchen	komp		ver	such				6	aussuchen
2	schwingversuchen	komp		ver	such				6	aussuchend
2	suchkopf				such			komp	3	auszusuchen
2	suchkopfes				such			komp	3	bergungsversuch
2	suchraums				such			komp	3	bestbesucht
2	versuchsverdichter			ver	such			komp	4	besuch
20										besuchen
										besuchend
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		besucher
1	3		x	x	x	x				besucherinnen
2	2		x	x	x					besucht
3	5				x			x		bittgesuch
4	3			x	x			x		blitzbesuch
5	1	x			x		x			brechsucht
6	4	x		x	x					bremsversuch
										damenbesuch
	Lexikoneintrag	Findet								durchgesucht
	prüf #u	6;								durchsuchend
	verdichter	11;20;								durchsucht
	vorversuch	1;3;5;								durchsuchung
	such #u	8;9;11;18;19;20;								eifersucht
	schwing #u	16;								einzelersuchen
	reiß #u									entlassungsgesuch
	zerreiß #u	15;								erholungsuchende
										ersuchen
										ersuchend
										ersucht
										fehlersuche
										fehlersuchprogramm
										fehlversuch
										freßsucht
										freundschaftsbesuch
										Lexikonliste
										gekürzt !!

7 14

Anhang A-4-29

Wortstamm -trag-

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
23	tragflügel				trag			komp	1	abgetragen
15	tragfläche				trag			komp	1	abtrag
11	tragflächen				trag			komp	1	abtragen
9	tragflügels				trag			komp	1	abtragend
7	tragflügeln				trag			komp	1	abtragung
6	auftragung			auf	trag		ung		2	angetragen
5	traggas				trag			komp	1	antrag
5	tragkraft				trag			komp	1	antragen
3	trag-								0	antragend
3	tragflügelmodell				trag			komp	1	anzutragen
3	tragflügelstruktur				trag			komp	1	atomsperrvertrag
3	tragflügelstrukturen				trag			komp	1	aufgetragen
3	tragflügeltheorie				trag			komp	1	aufklärungsauftrag
3	traggasvolumen				trag			komp	1	auftrag
3	tragverhalten				trag			komp	1	auftragen
2	traganteil				trag			komp	1	auftragend
2	tragflächen-								0	auftragseinwerbung
2	tragflächenenden				trag			komp	1	auftragswerk
2	tragflug-								0	aufzutragen
2	tragflügel-								0	ausgetragen
2	tragflügelprofil				trag			komp	1	austragen
2	tragflügelsehne				trag			komp	1	austragend
2	traggases				trag			komp	1	austragung
2	traggasversorgung				trag			komp	1	bauantrag
2	traglinien-								0	bausparvertrag
2	traglinienverfahren				trag			komp	1	beantragen
2	tragrohr				trag			komp	1	beantragend
2	tragstruktur				trag			komp	1	beantragt
28										beantragung
										beauftragen
Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM		beauftragend
1	22				x			x		beauftragt
2	21			x	x		x			beauftragte
										beauftragung
										beigetragen
	Lexikoneintrag	Findet								beitrag
	trag #u	1;2;3;4;5;8;10;11;12;13;15;16;18;21;22;23;26;27;28;								beitragen
	auftragung	6;								beitragend
	traggas	7;14;24;								beitragsfinanziert
3	23									beitragshöhe
										beizutragen
										betrag
										betragen
										betragend
										betragsmäßig
										bundesbeauftragte
										datenschutzbeauftragte
										datenübertragungsrate
										Lexikonliste gekürzt !

Anhang A-4-30

Wortstamm *-trimm-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
7	trimmung				trimm		ung		1	getrimmt
3	trimmzustand				trimm			komp	2	nachgetrimmt
2	trimm-								0	nachzutrimmen
2	trimmbare				trimm		bar		1	trimmen
2	trimmrechnung				trimm			komp	2	trimmend
2	trimmrechnungen				trimm			komp	2	vertrimmen
2	trimmwiderstand				trimm			komp	2	vertrimmend
2	trimmwiderstandes				trimm			komp	2	vertrimmt
8										

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
1	2				x		x		
2	5				x			x	

	Lexikoneintrag	Findet						
	trimmung	1;						
	trimmbar	4;						
	trimm #u	2;5;6;7;8;						

3 7

Anhang A-4-31

Wortstamm *-wirk-*

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
5	rückwirkung	komp			wirk		ung		1	ausgewirkt
4	rückwirkungen	komp			wirk		ung		1	auswirken
3	entgegenzuwirken	komp		zu	wirk	en			2	auswirkend
3	wirkebenenkriterium				wirk			komp	3	auswirkung
2	verdichterwirkungsgrad	komp			wirk		ung		1	bewirken
2	wirkebenenbezogene				wirk			komp	3	bewirkend
2	wirkebenenbezogenen				wirk			komp	3	bewirkt
2	wirkleistung				wirk			komp	3	bremswirkung
8										dagegengewirkt
										dagegenwirken
										dagegenwirkend
										durchgewirkt
										durchwirkend
										eingewirkt
										einwirken
										einwirkend
										einwirkung
										einzuwirken
										entgegengewirkt
										entgegenwirken
										entgegenwirkend
										erwirken
										erwirkend
										erwirkt
										fortgewirkt
										fortwirkend
										Gegen-
										einanderwirken
										gewirkt
										hineinwirken
										hineinzuwirken
										hingewirkt
										hinwirken
										hinwirkend
										hinzuwirken
										mitbewirken
										mitgewirkt
										mitwirken
										mitwirkend
										mitwirkung
										mitzuwirken
										nachgewirkt
										nachwirken
										nachwirkend
										nachwirkung
										rückwirken
										Lexikonliste
										gekürzt !!

3 7

Anhang A-4-32

Wortstamm -wölb-

Häufigk.	Token	BM	WBM	WBM	BM	FM	WBM	BM	Typ	lingo
5	verwölbung			ver	wölb		ung		1	flüstergewölbe
3	wölbklappen				wölb			komp	2	gewölbe
3	wölbkrafttorsion				wölb			komp	2	gewölbt
2	verwölbt			ver	wölb	t				vorgewölbt
2	wölbefekte				wölb			komp	2	vorwölben
2	wölbfrei				wölb			komp	2	vorwölbend
2	wölbklappenausschlag				wölb			komp	2	wölben
2	wölbkrafteinfluss				wölb			komp	2	wölbend
2	wölbspannung				wölb			komp	2	wölbung
2	wölbspannungen				wölb			komp	2	
2	wölbspannungsverteilung				wölb			komp	2	
2	wölbstörungen				wölb			komp	2	

12

Typ	Anzahl	BM	WBM	WBM	BM	FM	WBM	BM	
1	1			x	x		x		
10	2				x			x	
1				x		x			

	Lexikoneintrag	Findet						
	wölb #u	2;3;5;6;7;8;9;10;11;12;						
	verwölbung	1;						
	verwölbt	4;						

3 12

Anhang A-5

Indexierte Wörtern mit Zeichenfolge *-fehl-*

Häufigkeit	Identifizierte Token	Häufigkeit	Identifizierte Token
90	fehler	2	gestaltungsempfehlung
35	fehlend	2	fehlerquadratsumme
33	fehlen	2	datenfehler
31	empfehlen	2	folgefehler
21	empfehlung	2	implementierungsfehler
17	fehlerhaft	2	fehlerschrank
15	messfehler	2	fehlsteuerung
13	fehlerquelle	2	fehlersuche
12	fehlerfall	2	fehlerursache
7	befehl	2	fehleranalyse
7	positionsfehler	2	fehleranfällig
7	fehlerquadrat	2	fehleranfälligkeit
6	fehlerfällen	2	fehlerbehandlung
6	fehlermeldung	2	modellfehler
6	befehlen	2	nullpunktfehler
6	empfehlenswert	2	fehlerbild
6	fehlertolerant	2	fehlerdetektion
5	fehlererkennung	2	fehlerdiagnose
5	fehlstelle	2	temperaturmessfehler
4	fehlertoleranz	2	fehlerindikator
4	fehlerfrei	1	fehlerfindung
4	fehlfunktion	1	fehlerausbreitungsund
4	fehlermodell	1	fehlerausbreitung
4	fehlerfreie	1	temperaturfehler
4	fehlerbehaften	1	fehleraufnahme
4	fehlerabschätzung	1	fehleraufkommen
4	einzelfehler	1	fehlerauswertung
3	fehlergröße	1	fehleranzeige
3	fehlerquadratminimierung	1	fehleranzahl
3	fehlersignal	1	fehlerfläche
3	fehlalarm	1	systemischemessfehler
3	fehlinterpretation	1	syntaxbefehl
3	fehlstart	1	fehlerart
3	gesamtfehler	1	fehlerdynamikregelung
3	lagefehler	1	fehlerbeurteilung
3	einbaufehler	1	fehlerbewertung
3	fehlerfreien	1	fehlerbewußtsein
3	steuerbefehl	1	fehlerbestimmung
3	systemfehler	1	fehlerbudget
3	winkelfehler	1	fehlerbedingt
2	fehlerschätzung	1	fehlerdynamik
2	fehlermöglichkeit	1	fehlerbaum
2	fehlerbetrachtung	1	fehlerdynamikregler
2	übertragungsfehler	1	fehlerbehaftet
2	sensorfehler	1	testdatenfehler
2	fehleranteil	1	fehlerbeträge
2	zeitfehler	1	fehlergeschichte

Fortsetzung Anhang A-5

Indexierte Wörtern mit Zeichenfolge *-fehl-*

Häufigkeit	Identifizierte Token	Häufigkeit	Identifizierte Token
1	fehlererfassung	1	byte-fehler
1	fehlerbeseitigung	1	blockfehler
1	fehlerbeschreibung	1	biasfehler
1	baufehler	1	befehlszeile
1	fehlerabwicklung	1	befehlsgebend
1	echtzeitfehlersuche	1	befehlsdaten
1	zeichenbefehl	1	fehlstellenlänge
1	einbauwinkelfehler	1	fehlstellengeometrie
1	bauteilfehler	1	fehlstellenfrei
1	einsatzbefehl	1	fehlstellenfläche
1	batteriefehler	1	fehlstellencharakterisierung
1	bahnwinkelfehler	1	fehlstellenbreite
1	auswertungsfehler	1	fehlstellenbereich
1	ausweichempfehlung	1	fehlstellenausbildung
1	ausrichtungsfehler	1	flugwindhängewinkelfehler
1	anzeigefehler	1	mag-fehlstelle
1	anstellwinkelfehler	1	passfehler
1	amplitudenfehler	1	öffnungsfehler
1	aktuator-fehlverhalten	1	nullpunktfehler-drift
1	absolutfehler	1	nicklagewinkel-fehler
1	bedienungsfehler	1	navigationsfehlerkorrektur
1	fehlauslagerung	1	navigationsfehler
1	fehlerangabe	1	modifikationsempfehlung
1	trägheitsfehler	1	modellierungsfehler
1	fehlerhäufigkeit	1	piv-messfehler
1	fehlendend	1	interpretationsfehler
1	verfahrensfehler	1	maßempfehlung
1	verfehlen	1	fehlmessung
1	fehlein-schätzung	1	laufzeitfehlerfall
1	fehlbergschema	1	laufzeitfehler
1	fehlbedienungstoleranz	1	längslagewinkelfehler
1	windrichtungsfehler	1	prognosefehler
1	fehlausrichtung	1	ladebefehl
1	fehlerannahme	1	konstruktionsfehler
1	verfehlt	1	komponentenfehlervektor
1	eulerwinkelfehler	1	komponentenfehler
1	entwurfsfehler	1	kalibrierfehler
1	energiefehler	1	materialfehler
1	vertikalfehler	1	fehlerinformation
1	wellenfrontfehler	1	fehlerordnung
1	empfehlend	1	fehlermodus
1	windfehler	1	sensorfehler-modell
1	einzelfehlerquelle	1	fehlermerkmal
1	windgeschwindigkeitsfehler	1	seu-fehlerrate
1	fehlausrichtungswinkel	1	fehlerkorrektur
1	startpunktfehler	1	fehlerquadratmethode
1	fehlversuch	1	fehlerkomponente
1	skriptbefehl	1	fehlerklasse

Fortsetzung Anhang A-5

Indexierte Wörtern mit Zeichenfolge *-fehl-*

Häufigkeit	Identifizierte Token
1	fehlverhalten
1	fehlerfunktion
1	fehlerkatalog
1	richtungsfehler
1	fehlerfreiheit
1	regelfehler
1	fehlgeschlagen
1	reglerfehler
1	fehlerwahrscheinlichkeit
1	fehlervermeidung
1	fehlerverlauf
1	fehlerverhalten
1	fehlervektor
1	fehlerprognose
1	rekonstruktionsfehler
1	abbildungsfehler
1	fehlersumme
1	fehlerspeicherung
1	fehlersimulation
1	rms-fehler
1	fehlersicherheit
1	fehlersequenz
1	fehlerrechnung
1	fehllrate
1	rollwinkelfehler
1	radienfehler
1	fehlerträchtig
1	fehlerfunktionswert
1	include-befehl
1	icao-empfehlung
1	hauptfehlerquelle
1	gps-fehler
1	gierwinkel-fehler
1	geschwindigkeitsfehler
1	prozessfehler
1	formatierungsfehler
1	iterationsfehler
1	fertigungsfehler
1	itu-empfehlung
1	fehlerkennungsrat
1	quantisierungsfehler
1	fehlerinjektion
1	fehlerquadratminimum
1	fehlerhäufung
1	phasenfehler
1	fehlerhandhabung
1	fehlergrenze

Anhang A-6

Präfixliste

1	2	3	4	5	6
lingo			Stamnon >1	Lingo Standard >1	Anmerkungen/Lexikalisierung
1975	ab-		32	50	
1767	an-		62	84	
1249	auf-	t	10	17	
1653	aus-	t	18	22	
2245	be-	t	67	86	
216	bei-	t	3	4	beiwert, beibehält
569	durch-	w	3	4	durchfluß
1600	ein-	t	22	30	
838	ent-		20	24	Teilmenge von en-
1077	er-	t	47	49	
59	fehl-		1	3	fehl + #u
2958	ge-		58	69	
111	gegen-	w	1	1	
24	grund-	s	1	2	grundzügen
35	haupt-	s	6	6	hauptfahrwerk-
104	hinter-	w	0	3	
39	miss-	w	2	2	
37	miß-		1	1	
368	mit-	w	5	7	mittelung, mittelständisch
643	nach-	w	4	4	nachgiebig, nachhaltig, nachvollziehbarkeit
37	neben-	w	0		
64	nicht-	w	1	1	
70	rück-	w	12	12	
525	über-	w	13	12	
687	um-		9	24	
1329	un-		43	68	
416	unter-	w	12	12	
118	ur-		0	0	
2789	ver-		40	102	
890	vor-		46	54	
64	wider-	w	2	2	
322	zer-		6	6	zerstäubung, zerspanbarkeit
1424	zu-	t	19	34	
309	zurück-	w	0	0	
48	zwischen-	w	0	0	
94	ad-		18	18	
8	allo-		0	0	englisch
2	amphi-		0	0	
0	ampho-		0	0	
44	ana-		3	3	analysierbar
6	ante-		2	2	englisch
65	anti-		1	1	
22	apo-		1	1	apoapsis apochromatisch
2	äqui-		1	1	
3	archäo-		0	0	
10	archi-		1	1	englisch
79	auto-	s	9	9	englisch, autoklav, automav #s

Fortsetzung Anhang A-6

Präfixliste

242	bi-		13	13	
31	bin-	t	0	0	
0	cis-		0	0	
570	de-		71	79	
57	des-	t	8	8	Teil von de-
374	di-		38	39	
44	dia-	s	2	2	
94	dis-		20	20	
3	dys-		0	0	
0	eiso-		0	0	
13	ek-		1	1	
0	ekto-		0	0	
943	en-		60	60	
6	endo-		0	0	
3	ento-		0	0	
24	epi-		1	1	
1	equi-		2	2	
55	eu-		12	12	
187	ex-		22	22	
7	exo-		1	1	exobiologie, exomar
12	extra-	s	0	0	
0	fero-		0	0	ferogramm, interferometer, interferometrie, interferometrisch
17	hemi-		0	0	hemisphäre
3	hetero-		0	0	
16	homo-		1	1	homographie, homonuklear
3	homöo-		0	0	
14	hyper-		0	0	
12	hypo-		0	0	hypoxie
122	im-		28	28	
618	in-	t	91	91	
8	infra-		3	3	infraschall, infrastrukturell
258	inter-		21	21	siehe in-
10	intra-		1	1	intralaminar
1	intus-	a	0	0	
21	iso-		6	6	
26	kata-		0	0	katadioptrisch
0	kato-		1	1	
367	kon-		16	29	
19	kontra-	w	0	0	bereits in Lingo vorhanden
12	makro-	s	0	0	bereits in Lingo vorhanden
2	maximal-	a	0	0	1x -verbindung, 1x fragment, gleitzahl
3	mega-		2	1	
3	meso-		0	0	
33	meta-		2	2	
20	mikro-		17	17	
50	mini-	a	5	5	
11	multi-	s	4	4	multiple, multiplikation
12	neo-		1	1	neopren
12	non-		1	1	englisch
90	ob-	t	12	12	oberfläche, englisch

Fortsetzung Anhang A-6

Präfixliste

0	palin-		0	0	
43	pan-	s	7	7	21 panel
65	para-		6	6	21
92	per-		4	5	40 englisch und peri-
9	peri-		2	2	12 periapsis, perigäums
27	poly-		1	1	18
37	post-	w	2	3	10 englisch
90	prä-		0	0	11 prädiktive, präferiert, prädierte
308	pro-	w	37	39	160
16	pros-		0	0	2
7	pseudo-		2	2	12
3	quasi-	w	1	1	1 fragment
2	retro-		1	1	10 retrospektiv #a (fehler lingo)
0	rheo-		0	0	2
11	semi-		2	2	13
1	strato-		2	2	10 stratosphärisch#a,
46	sub-		6	6	40 Beispiel für #u lösung
15	super-	w	3	3	13 supersonisch
2	supra-		0	0	0
26	syn-		4	4	19 synchronisierung
40	tele-		7	7	29
17	thermo-		3	9	41 thermoplast
53	trans-	w	11	11	45
7	ultra-		7	7	37
37	uni-	a	5	5	13 englisch
0	vario-		2	2	5 variometer
0	zirkum-		0	0	0

Erklärung:

Spalte 2 enthält die Präfixe

Spalte 1 die Anzahl von Einträgen im Systemwörterbuch mit dem Präfix

Spalte 3 enthält die Wortklasse wenn das Präfix bereits lexikalisiert ist

Spalte 4 die Anzahl der Token die nach der letzten Indexierung nicht erkannt wurden

Spalte 6 die Anzahl der Token die bei der Standard Lingo Indexierung nicht erkannt wurden und ggf. Anmerkungen für die Lexikalisierung einzelner Terme

Spalte 5 wie Spalte 6 ohne Doppelnennungen

Die grün gekennzeichneten Präfixe werden gesondert untersucht

Anhang A-7-1

Token mit Zeichenfolge ab-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
213	ab								
12	abbildet				abbilden				
9	abtastfrequenz	komp							
6	abströmwinkel	komp							
5	abgeschätzten			abgeschätzt					
4		#e firma						abaqus	
4	abgeschätzte	s.o.							
4	abgemindert			abgemindert					
3	abzumindern			abzumindern					
3	ab-	- verbindung							
3	abu-ghannam	#e autor							
3	above	englisch							
3	abgewägt			abgewägt					
3	abschätzbar			abschätzbar					
2	abfangmanöver					abfang			
2	abgelängt			abgelängt					
2	abgegebene			abgegeben					
2	abfertigungs-	- verbindung							
2	abatement	englisch							
2	abge	f							
2	ability	englisch							
2	abrasiven			abrasiv					
2	abzusenken								
2	abreißgewebe	f							
2	abrollwege	f							
2	abs	f							
2	abschliessen	f							
2	abschn	f							
2	abschwimmen				abschwimmen				
2	abschwimmenden								
2	abtastwerte					abtast			
2	abtastwerten	s.o.							
2	abw	f							
2	about	englisch							

Anhang A-7-2

Token mit Zeichenfolge ad-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
6	ad	f							
5	ada							ada	
9	adaptiver	adaptiv (lingo)							
2	adaptronik		adaptronik						
3	adc	f							
2	adcs	f							
2	add	englisch							
2	addressing	englisch							
3	adequate	englisch							
5	adiabaten			adiabat					
2	adiabatic	englisch							
3	adif	f							
2	adjustment	englisch							
2	adkins	#e ?							
3	ads	f							
2	ads-b	-verbindung							
21	advanced	englisch							
7	advisory	englisch							

Anhang A-7-3

Token mit Zeichenfolge an-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
304	an	-							
21	an-	-verbindung							
2	ana	f							
6	analyse-	-verbindung							
2	analytical	englisch							
101	and	englisch							
2	ande-	-verbindung							
72	anderem	-							
228	anderen	-							
71	anderer	-							
20	anderes	-							
2	andrich	f €							
3	anemometer		anemometer						
2	anfachtung		anfachtung						
2	anfah-	-verbindung							
2	anfänge	-							
8	anfängen	-							
4	anfangs-	-verbindung							
2	anforde	f							
2	ange	f							
2	angewen	f							
9	angle	englisch							
2	angleich	f ?							
2	angles	englisch							
2	angusskonzepte		anguss						
2	anhängigkeit		anhängigkeit						
5	anisotroper			anisotrop					
3	anm	f							
4	annähert	-							
2	annealing	englisch							
6	annex	englisch							
2	another	englisch							
2	anpresskraft					anpress			
2	anregungs-	-verbindung							
2	anriß	lingo anriss							
4	ans	f							
2	anschaffungs-	-verbindung							
6	anschliessend	f							
27	anschließende	-							
4	anschliessenden	f							
35	anschluss		anschluss						
2	anschlusses	s.o.							
2	anschlussflüge	s.o.							
2	anschlussmöglichkeit	s.o.							
11	anstell-	-verbindung							
2	anstellund	f							
13	ansys	f							

Fortsetzung Anhang A-7-3

Token mit Zeichenfolge an-

2	ant	f							
4	antenna	englisch							
2	antennen-	-verbindung							
2	anthropometrische	s.o.							
2	anthropometrischen			anthropometrisch	anthropo				
4	anti-	-verbindung							
3	antonov	#e firma							
5	antriebs-	-verbindung							
2	anund	f							
2	anwendungs-	-verbindung							
2	any	englisch							
3	anzeige-	-verbindung							
2	anzufliegen	-							
2	anzunehmende	-							
2	anzuwachsen	-							

Anhang A-7-4

Token mit Zeichenfolge auf-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
18	aufschluss	aufschluß (lingo)							
11	auftriebs-	-verbindung							
8	auftraten	-							
8	auf-	-verbindung							
6	auftragung		auftragung						
2	aufstaugebiet		aufstau						
2	aufklärungs-	-verbindung							
2	aufheizrate					aufheiz			
2	aufhält	-							
2	aufgepr	f							

Anhang A-7-5

Token mit Zeichenfolge aus-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
9	aus-	-verbindung							
4	ausge	f							
4	ausserhalb	f							
4	ausserdem	f							
4	ausschluss	ausschluß (lingo)							
4	ausgangs-	-verbindung							
3	ausbrenngrad					ausbrenn			
2	ausblasrate					ausblas			
2	ausfahrwinkel					ausfahr			
2	ausfall-	-verbindung							
2	ausfallsichere	ausfall (lingo)							
2	ausfallsicheren	s.o.							
2	ausklappwinkel					ausklapp			
2	auszuregeln	-							
2	auslegungs-	-verbindung							
2	ausmass	f							
2	ausschliesslich	f							
2	ausser	f							
2	auswertbar			auswertbar					
2	ausgasphase	f?							

Anhang A-7-6

Token mit Zeichenfolge be-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
20	be	-							
10	be-	-verbindung							
4	beam	englisch							
2	beams	englisch							
3	bearings	englisch							
3	because	englisch							
3	bed	englisch							
2	bedford	#e						bedford	
3	bedien-	-verbindung							
3	bediengerät					bedien			
2	bedienkonsole	s.o.							
2	bedienkonsolen	s.o.							
3	bedrohungen	lingo							
4	beein	f							
4	been	englisch							
5	befanden	-							
2	befuerung		befuerung						
2	befuerungselemente	komp							
3	before	englisch							
2	befundung		befundung						
6	begannen	-							
3	begin	englisch							
2	beginning	englisch							
2	begleiterscheinung		begleiterscheinung						
4	behält	-							
2	behütung		behütung						
3	beibehält	-							
2	beeinflußt				beeinflußt				
2	beißer	-							
3	bekamen	-							
2	belastungs-	-verbindung							
2	below	englisch							
4	bench	englisch							
2	benda	f							
2	bending	englisch							
2	benutzer-	-verbindung							
2	berech	f							
3	berechen	f							
7	berechnungs-	-verbindung							
2	berechnung	f							
3	bergner	#e autor							
3	berück	f							
2	berücksichtigung	f							
2	besaßen	-							
2	besch	f							

Fortsetzung Anhang A-7-6

Token mit Zeichenfolge be-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
2	Beschleunigungs- messwerte	f							
3	beschloss	beschlossen (lingo)							
2	beschluss	beschluß (lingo)							
3	beschreibbar			beschreibbar					
2	beson	f							
5	best	englisch							
2	betr	f							
2	betrafen	-							
2	betreuungs-	-verbindung							
2	betriebene				betrieben				
6	betriebenen	s.o.							
3	betriebener	s.o.							
6	betriebs-	-verbindung							
2	betritt	-							
2	bettung		bettung						
6	between	englisch							
2	betz	#e autor							
2	beulform					beul #u			
3	beulgrenze	s.o.							
4	beullast	s.o.							
3	beulverhalten	s.o.							
4	bewegungs-	-verbindung							
3	bewer	f							
2	bewicklung		bewicklung						
4	bez	f							
2	bezog	-							

Anhang A-7-7

Token mit Zeichenfolge de-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
22	de	f							
3	de-	- verbindung							
2	de-arrest	- verbindung							
3	deboarding	englisch	deboarding						
2	deboardingzeiten	englisch							
5	debriefing	englisch	debriefing						
5	debris		debris						
3	debugging		debugging						
2	dec	f							
2	decay	englisch							
3	deck-	- verbindung							
2	decomposition	englisch							
2	decrease	englisch							
3	dedizierte			dedizierte					
3	dedizierten	s.o.							
8	deep	englisch							
3	defects	englisch							
6	defence	englisch							
4	defense	englisch							
4	deformations-	- verbindung							
3	deg	f							
3	dehnungs-	- verbindung							
2	deinert	#e autor							
3	dekodierung		dekodierung						
3	delay	englisch							
2	delays	englisch							
2	deliverables	englisch							
3	delivery	englisch							
2	demand	englisch							
2	demands	englisch							
2	denkansätze					denk			
5	density	englisch							
2	denton	#e							
5	department	englisch							
4	departure	englisch							
2	departures	englisch							
2	depend	englisch							
4	dependent	englisch							
2	depends	englisch							
3	depicted	englisch							
2	deployable	englisch							
2	deployed	englisch							
4	derivativa		derivativa						

Fortsetzung Anhang A-7-7

Token mit Zeichenfolge de-

2	derivativs	englisch							
2	derived	englisch							
2	derwert	f							
7	descent	englisch							
3	described	englisch							
6	description	englisch							
2	descriptor	englisch							
7	design-	-verbindung							
3	desired	englisch							
3	dessau	#e							
3	destabilisierende			destabilisierend					
3	det	f							
4	detailed	englisch							
2	detailgetreues			detailgetreu					
3	detailliert	f							
3	detaillierte	f							
6	detection	englisch							
2	detector	englisch							
5	determination	englisch							
5	determine	englisch							
5	determined	englisch							
3	deutschsprachigen			deutschsprachig					
2	dev	f							
11	development	englisch							
4	deviation	englisch	deviation						
2	deviations	englisch							
3	deviationsmomente	s.o.							
2	dez	f							

Anhang A-7-8

Token mit Zeichenfolge di-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
12	di	f							
2	diag	f							
3	diamond	englisch							
3	dichte-	- verbindung							
5	diehl	#e firma							
2	dielektrische			dielektrische					
4	dielektrischen	s.o.							
2	diff	f							
4	difference	englisch							
2	differences	englisch							
3	different	englisch							
2	dij	f							
8	din	Abk.							din
3	dir	f							
6	direct	englisch							
5	direction	englisch							
3	directly	englisch							
2	director	englisch							
4	dis	f							
3	discovery	englisch							
3	discrete	englisch							
2	discussed	englisch							
3	diskontinuierliche			diskontinuierlich					
2	diskontinuität		diskontinuität						
14	diskretisiert			diskretisiert					
2	diskretisierte	s.o.							
3	diskretisierten	s.o.							
23	diskretisierung		diskretisierung						
2	dispersive			dispersiv					
4	displacement	englisch							
8	dissipation	englisch	dissipation						
3	dissipationsrate	s.o.							
2	dissoziieren				dissoziieren				
8	distance	englisch							
2	distances	englisch							
2	distortion	englisch							
2	distributed	englisch							
2	disturbance	englisch							

Anhang A-7-9

Token mit Zeichenfolge ein-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
43	ein-	-verbindung							
4	eindringtiefe					eindring			
2	einflussgrößen	einfluß (lingo)							
3	eingangs-	-verbindung							
3	eingangsgrößen	f							
3	eingef	f							
2	eingef	f							
3	einlox	f							
2	einpendelt	-							
2	einsatzszenarien		szenarien						
2	einschaligen			einschalig					
2	einschl	f							
2	einschliesslich	f							
5	einschnürung		einschnürung						
2	einschwingvorgang					einschwing			
26	einsetzbar			einsetzbar					
6	einsetzbaren	s.o.							
3	einsetzbares	s.o.							
4	einsparpotenzial					einspar			
2	einstell-	-verbindung							
2	einstellparameter					einstell			
3	einstellwinkel	s.o.							
4	eintritts-	-verbindung							
4	einund	f							
6	einzel-	-verbindung							

Anhang A-7-10

Token mit Zeichenfolge en-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
22	en	f							
4	enable	englisch							
2	encoder	englisch							
8	end	englisch							
8	endanflug		endanflug						
2	endanwender		endanwender						
2	endauswahl		endauswahl						
2	endbearbeitung		endbearbeitung						
2	endergebnis		endergebnis						
2	endgerät		endgerät						
2	endgeschwindigkeit		endgeschwindigkeit						
5	endgültig			endgültig					
11	endgültige	s.o.							
12	endgültigen	s.o.							
3	endkunden		endkunde						
5	endlicher	-							
2	endmontage		endmontage						
2	endmontiert			endmontiert					
4	endphase		endphase						
2	endprodukt		endprodukt						
3	endprodukte	s.o.							
2	endpunkt		endpunkt						
3	endscheiben		endscheibe						
3	endurance	englisch							
3	endzustand		endzustand						
3	enemy	englisch							
2	ener	f							
7	energie-	-verbindung							
11	energy	englisch							
17	engine	englisch							
16	engineering	englisch							
12	engines	englisch							
16	engl	f							
3	engstem	-							
6	enhanced	englisch							
5	ent	f							
3	entfernungs-	-verbindung							
2	entformbarkeit		entformbarkeit						
2	entformungstests		entformung						
3	entgegenzuwirken	-							
2	entgeltiges			entgeltig					
3	enthalpy	-?							
85	enthält	-							
2	entire	englisch							
2	entladevorgängen					entlade			

Fortsetzung Anhang A-7-10

Token mit Zeichenfolge en-

2	entlastungs-	-verbindung							
2	entnehmbar			entnehmbar					
4	entry	englisch							
2	entschloß	-							
8	entsprachen	-							
3	entspre	f							
26	entstand	f							
19	entwicklungs-	-verbindung							
2	entwickelt	f							
7	entwurfs-	-verbindung							
4	envelope	englisch							
9	environment	englisch							
5	environmental	englisch							
2	environmentally- friendly	englisch							
5	envisat	#e						envisat	

Anhang A-7-11

Token mit Zeichenfolge erd-

Anz.	Token		#s	#a	#v	#u	#w	#e
2	erd					erd		
7	erdbeobachtung	s.o.						
5	Erdbeobachtungsmissionen	f						
2	erdbeobachtungssatellit	beobachtung						
3	erdboden		erdboden					
3	erdkundung		erdkundung					
2	erdkundungssatelliten	s.o.						
2	erdfesten			erdfesten				
3	erdgebundene			erdgebunden				
2	erdmagnetfeld			erdmagnetfeld				
2	erdmagnetfeldes	s.o.						
4	erdorbit			erdorbit				
2	erdradius			erdradius				
3	erdschwerfeldes			erdschwerfeld				
2	ere	f						
3	erfah	f						
4	erflogen				erflogen			
2	erforder	f						
4	erforderte	erfordert (lingo)						
5	erforderten	erfordert (lingo)						
2	erfahren	-						
41	ergaben				ergaben			
2	ergeb	f						
2	ergebenen	-						
2	ergingen	-						
3	erhielten	-						
2	erhö	f						
4	ermangelung			ermangelung				
3	ermöglich	f						
4	ermüdungs-	-verbindung						
2	err	f						
7	error	englisch						
2	erschien	-						
166	erste						erste	
3	erstem						erstem	
3	erstere						erstere	
3	erstes						erstes	
17	erstes						erstes	
2	erstreckung		erstreckung					
2	erwachsenen	erwachsen						
2	erwartbaren			erwartbar				
2	erwei	f						
3	erweiterbar			erweiterbar				
2	erweitere	f						
11	erwies	-						
4	erzielbare			erzielbar				
10	erzielbaren	s.o.						

Anhang A-7-12

Token mit Zeichenfolge eu-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
16	eu	#e							eu
4	eu-projektes	-verbindung							
2	eu-projekts	-verbindung							
2	eu-rahmenprogramm	-verbindung							
2	euranus	englisch							
2	eurocae	#e firma						eurocae	
4	eurockot	#e firma						eurockot	
6	eurocontrol	#e firma						eurocontrol	
11	eurocopter	#e firma						eurocopter	
4	eurofighter	#e						eurofighter	
3	europe	englisch							
21	european	englisch							

Anhang A-7-13

Token mit Zeichenfolge ex-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
8	ex	-							
7	exakter	exakt (lingo)							
2	exceed	englisch							
5	excel	#e						excel	
3	excellent	englisch							
2	excessive	englisch							
2	excimerlaser	englisch							
2	execution	englisch							
2	executive	englisch							
2	existence	englisch							
3	existing	englisch							
2	exomars	#e projekt							
10	exp	f							
2	expected	englisch							
3	experi	f							
2	experiment-	-verbindung							
34	experimente	experiment (lingo)							
2	explorer		explorer						
2	exposure	englisch							
4	extended	englisch							
4	external	englisch							
3	extrema	-							

Anhang A-7-14

Token mit Zeichenfolge ge-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
17	ge	-							
4	ge-	-verbindung							
5	gear	englisch							
2	gearbox	englisch							
2	gefahren-	-verbindung							
2	gefitteten	-							
4	gegen-	-verbindung							
2	gehäusegrenzschicht		gehäuse						
4	gehgeschwindigkeit	?							
6	gekühlten	gekühlt (lingo)							
2	geld-	-verbindung							
2	gem	f							
3	gemäss	f							
10	gen	f							
2	genauerem	-							
2	genauestens	-							
2	generally	englisch							
2	generated	englisch							
3	generic	englisch							
2	genetic	englisch							
6	geo					geo			
6	geodäsie		geodäsie						
2	geom	f							
2	geomet	f							
2	geometric	englisch							
9	geometrie-	-verbindung							
2	geometry	englisch							
2	georeferenzierung	geo #u							
3	geostationäre	geo #u							
7	geostationären	geo #u							
11	gepfeilten			gepfeilt					
6	gerader	-							
2	geratenen	-							
9	geringerem	-							
3	german	englisch							
3	germany	englisch							
9	ges	f							
5	gesamt-	-verbindung							
2	geschäfts-	-verbindung							
2	geschwindi	f							
12	geschwindigkeits-	-verbindung							
2	geschwindigkeitskoe	f							
2	gesintertem			gesintert					
4	gesonderte	gesondert (lingo)							
2	gestal	f							

Fortsetzung Anhang A-7-14

Token mit Zeichenfolge ge-

2	gestrakten			gestrakt					
5	gestufte			gestuft					
2	gestwa	f							
4	get	englisch							
3	getempert	f							
2	getriebene	getrieben (lingo)							
4	getriggert			getriggert					
5	gew	f							
8	gewichts-	-verbindung							
31	gewisse	-							
2	gewissem	-							
7	gewisser	-							
4	gewisses	-							
2	gezoomt			gezoomt					

Anhang A-7-15

Token mit Zeichenfolge im-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
302	im	-							
8	ima	f							
6	imaging	englisch							
2	imc	f							
2	iml	f							
12	impact	englisch							
3	impact-	-verbindung							
4	impactbelastung	komp englisch							
5	impactor		impactor						
2	impactorgeschwindigkeit	s.o.							
3	impactors	s.o.							
4	impacts	englisch							
3	impactschäden	komp englisch							
3	impactschädigung	komp englisch							
2	impactuntersuchungen	komp englisch							
3	impactverhalten	komp englisch							
2	impactversuche	komp englisch							
2	impactwerkzeug	komp englisch							
2	impakt	f							
2	implemented	englisch				impact			
2	importance	englisch							
2	important	englisch							
2	impose	englisch							
2	improved	englisch							
2	improvement	englisch							
3	impuls-	-verbindung							
11	imu	f							
2	imu3	f							

Anhang A-7-16

Token mit Zeichenfolge in-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
7	in-	-verbindung							
2	inboard	englisch							
4	inc	f							
2	incidence	englisch							
3	inclination	englisch							
6	including	englisch							
3	inclusive	englisch							
2	inconel	f							
2	increase	englisch							
4	increased	englisch							
3	increasing	englisch							
2	increment	englisch							
2	ind	f							
4	indicator	englisch							
3	industrial	englisch							
2	industry	englisch							
2	inert		inert						
4	inertia	englisch							
2	inf	f							
6	in-flight	-verbindung							
4	influence	englisch							
2	influenced	englisch							
7	informations-	-verbindung							
2	infrared	englisch							
4	infrarot-	-verbindung							
6	infrastructure	englisch							
2	infusions-	-verbindung							
3	ing	f							
2	ingress	englisch							
2	init	f							
5	injection	englisch							
2	injector	englisch							
3	injektions-	-verbindung							
2	injury	englisch							
5	inlet	englisch							
3	inm	f							
3	inmarsat	#e						inmarsat	
5	innen-	-verbindung							
3	innendruck-	-verbindung							
3	in-orbit	-verbindung							
2	in-plane	-verbindung							
15	input	englisch							
4	inputs	englisch							
56	ins	-							

Fortsetzung Anhang A-7-16

Token mit Zeichenfolge in-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
3	insbe	f							
3	inserts	englisch							
5	in-situ	-verbindung							
3	inspection	englisch							
2	inspektions-	-verbindung							
4	instantan	f							
3	instantanen	f							
3	instantaneous	englisch							
2	instationary	englisch							
2	instru	f							
3	instrument-	-verbindung							
3	instrumenten-	-verbindung							
2	insurance	englisch							
2	inta	f							
2	integr	f							
9	integrated	englisch							
4	integrity	englisch							
2	intensity	englisch							
3	inter					inter			
3	inter-	-verbindung							
3	interaction	englisch							
4	interactive	englisch							
2	interconnect	englisch							
2	interconnection	englisch							
2	interference	englisch							
6	interferometer		interferometer						
2	interferometers	s.o.							
3	interglas	#e firma						interglas	
2	intergral		intergral						
2	interior	englisch							
3	interlaminare			laminar		inter			
3	interlaminaren								
3	intermediate	englisch							
4	internal	englisch							
2	internationalisierung		nationalisierung						
2	internet-	-verbindung							
5	interoperabilität		interoperabilität						
2	interplanetaren		interplanetar						
2	interpretierbar			interpretierbar					
2	in-the-loop	-verbindung							
8	into	englisch							
2	intralaminaren					intra			
4	introduction								
4	invar							invar	feni
2	invest	englisch							
2	investigation	englisch							
2	inviskosen			inviskos					

Anhang A-7-17

Token mit Zeichenfolge iso-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
10	iso							iso	iso
2	isochor			isochor					
2	isochoren	s.o.							
4	isolinien		isolinie						
2	isoliniendarstellung	s.o.							
2	isotropic	englisch							

Anhang A-7-18

Token mit Zeichenfolge kon-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
2	kon	f							
2	konf	f							
3	konfig	f							
2	konfigu	f							
2	konfigurateurs	-							
2	konfigurations-	-verbindung							
2	kong	f							
2	konservativität			konservativ					
3	konst	f							
3	konstruktions-	-verbindung							
3	kontroll-	-verbindung							
3	konturierte			konturiert					
2	konturierten	s.o.							
2	konturierter	s.o.							
2	konzept-	-verbindung							
2	konzessionär	-							

Anhang A-7-19

Token mit Zeichenfolge mikro-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
3	mikro-	-verbindung							
2	mikrocontroller					mikro			
2	mikroelektronik	s.o.							
6	mikroflugzeug	s.o.							
6	mikroflugzeuge	s.o.							
4	mikroflugzeugen	s.o.							
5	mikroflugzeuges	s.o.							
4	mikroflugzeugs	s.o.							
3	mikromechanische	s.o.							
4	mikromechanischen	s.o.							
2	mikrometer	s.o.							
3	mikroprozessor	s.o.							
2	mikroprozessors	s.o.							
6	mikrorisse	s.o.							
2	mikrorissen	s.o.							
3	mikroschalter	s.o.							
3	mikrostruktur	s.o.							

Anhang A-7-20

Token mit Zeichenfolge pro-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
3	pro-	-verbindung							
2	probable	englisch							
5	procedure	englisch							
9	procedures	englisch							
7	process	englisch							
15	processing	englisch							
2	processors	englisch							
3	procfk	f							
2	pro-cfk	-verbindung							
2	produce	englisch							
2	produced	englisch							
2	product	englisch							
5	production	englisch							
2	produkt-	-verbindung							
3	produktions-	-verbindung							
9	prof	f							
3	profil-	-verbindung							
2	profilstrak		strak						
6	program	f							
2	programmable	englisch							
2	programmierbar			programmierbar					
6	programmierbare	s.o.							
8	programmierbaren	s.o.							
7	project	englisch							
2	projects	englisch							
2	proof	englisch							
2	property	englisch							
2	proposed	englisch							
6	propulsion	englisch							
10	protection	englisch							
2	protex	#e firma							
2	protocol	englisch							
3	prototyping	englisch							
2	provided	englisch							
3	provides	englisch							
3	proximity	englisch							
4	prozess-	-verbindung							

Anhang A-7-21

Token mit Zeichenfolge rück-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
1	rück-	-verbindung							
1	rückkehrkapsel	f							
2	rücken-	-verbindung							
1	rückenlehne-	-verbindung							
1	rückftihrzweig	f							
1	rückgabety		rückgabe						
1	rückkehrfahr	f							
1	rückschliessen	f							
2	rückschluss	schluß (lingo)				rück			
2	rückverfolgbarkeit		verfolgbarkeit						
1	rückwärts-	-verbindung							
1	rückzugs-	-verbindung							

Anhang A-7-22

Token mit Zeichenfolge trans-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
2	transceiver		transceiver						
5	transiente			transient					
7	transienten								
2	transientes								
2	transonischen	englisch							
2	transportables	englisch							
10	transportation		transportation						
5	transsonische			transsonisch					
14	transsonischen	s.o.							
4	transsonischer	s.o.							
2	transsonisches	s.o.							

Anhang A-7-23

Token mit Zeichenfolge über-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
8	über-	-verbindung							
4	überbrückung		überbrückung						
2	übergangs-	-verbindung							
3	überlappung		überlappung						
2	überlappungen	s.o.							
3	überlastung		überlastung						
2	überlebbar			überlebbar					
4	übermittlung	übermittlung (lingo)	übermittlung						
13	überschreitung		überschreitung						
4	überschreitungen	s.o.							
8	übersichtlichkeit		übersichtlichkeit						
2	überwachbar	-							
7	überwachungs-	-verbindung							

Anhang A-7-24

Token mit Zeichenfolge ultra-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
3	ultra					ultra			
2	ultraleichten			ultraleicht					
2	ultraleichtes	s.o.							
2	ultraleichtflugzeuge	s.o.							
2	ultraleichtflugzeugen	s.o.							
2	ultraviolett		ultraviolett						
2	ultravioletten	s.o.							

Anhang A-7-25

Token mit Zeichenfolge um-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
296	um	-							
3	umfangs-	-verbindung							
2	umgebenen	umgeben (lingo)							
2	umlauf-	-verbindung							
2	umlenkrollen					umlenk			
2	umlenkverhalten					umlenk			
2	umriss	umriß (lingo)							
3	umwelt-	-verbindung							
2	umzurüsten	-							

Anhang A-7-26

Token mit Zeichenfolge un-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
7	un	-							
3	unabh	f							
2	unacceptable	englisch							
3	unbeeinflusst			unbeeinflusst					
4	unbestritten			unbestritten					
2	unbestrittenen	s.o.							
2	unbeträchtlichen			unbeträchtlich					
3	under	englisch							
2	underestimated	englisch							
2	understanding	englisch							
3	undertaking	englisch							
2	ung	f							
2	ungedämpften			ungedämpft					
2	ungekühlt			ungekühlt					
2	ungemittelten			ungemittelt					
3	ungeschädigte			ungeschädigt					
2	ungeübte			ungeübt					
2	uninteressant	-							
30	unit	englisch							
6	units	englisch							
2	universitäre			universitär					
3	university	englisch							
2	unlocked	englisch							
5	unmanned	englisch							
3	unsymmetrie	-							
8	unter-	-verbindung							
24	untere						untere		
71	unteren						unteren		
3	unterer						unterer		
2	unteres						unteres		
2	unterhält	-							
2	unternehmens-	-verbindung							
3	unterschall-	-verbindung							
2	unterschied-	-verbindung							
3	unterste	-							
4	untersten	-							
2	untersu	f							
2	until	englisch							
11	unumgänglich	-							
2	unveränderlich			unveränderlich					
2	unverwundene	-							
2	unverwundener	-							
2	unzweifelhaft			unzweifelhaft					

Anhang A-7-27

Token mit Zeichenfolge ver-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
12	ver	-							
8	ver-	-verbindung							
2	verdampferstrecke		verdampfer						
6	verdichter-	-verbindung							
7	verformungs-	-verbindung							
23	verfügbare			verfügbar					
32	verhält	-							
4	verification	englisch							
2	verifikations-	-verbindung							
2	verkaufs-	-verbindung							
4	verkehrs-	-verbindung							
2	verkehrsszenarien		szenarien						
2	verkokung		verkokung						
9	verlässt	-							
20	verläuft	-							
4	verlief				verläuft				
2	verließen	-							
3	vermag	-							
3	vernachl	f							
35	vernachlässigbar			vernachlässigbar					
3	vernachlässigbare	s.o.							
4	vernachlässigbaren	s.o.							
4	verschiebbar			verschiebbar					
4	verschiede	f							
2	verschleiss	f							
2	verschleiß-	-verbindung							
3	verschleisschutz	f							
2	verschuß	verschuss							
2	versehentlich								
2	verspiegelung		verspiegelung						
3	versprach	-							
13	verspricht	-							
2	verstärkungs-	-verbindung							
4	versuchs-	-verbindung							
2	vert	f							
2	vertical	englisch							
2	vertritt	-							
2	verwen	f							
2	verwunderlich			verwunderlich					
10	very	englisch							

Anhang A-7-28

Token mit Zeichenfolge vor-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
255	vor	-							
23	vor-	-verbindung							
10	vorauslegung	-							
3	vordefinierter			vordefiniert					
9	vorder-	-verbindung							
29	vorderen	-							
2	vorderes	-							
2	vorderkanten-	-verbindung							
2	vordimensionierung	-							
6	vorentwicklung		vorentwicklung						
2	vorfilter		vorfilter						
5	vorflügel		vorflügel						
2	vorflügeln	s.o.							
2	vorformling		vorformling						
3	vorge	f							
2	vorgekrümmten								
2	vorgemischte								
3	vorgeschädigte			vorgeschädigt					
2	vorgeschädigten	s.o.							
3	vorgeschaltet			vorgeschaltet					
3	vorgeschalteten	s.o.							
4	vorhersagbar			vorhersagbar					
2	vorhersagbaren	s.o.							
2	vorimprägnierten			vorimprägniert					
2	vorkörper			vorkörper					
2	vorkrümmung			vorkrümmung					
3	vorleitrad			vorleitrad					
2	vorleitrades	s.o.							
2	vormischstrecke					vormisch			
3	vormischverbrennung								
7	vorraussetzung	f							
7	vorraussetzungen	f							
3	vorsteuerung			vorsteuerung					
12	vortex			vortex					
3	vortex-lattice	-verbindung							
2	vortices	englisch							
2	vorticity	englisch							
2	vorund	f							
4	vorverdampfung			vorverdampfung					
3	vorverformung			vorverformung					
4	vorversuch			vorversuch					
5	vorversuche	s.o.							
6	vorversuchen	s.o.							
2	vorwärmung			vorwärmung					
2	vorwarnzeit			vorwarnzeit					
2	vorzuhalten	-							

Anhang A-7-28

Token mit Zeichenfolge zu-

Anz.	Token		#s	#a	#v	#u	#w	#e	syn
7	zu-	-verbindung							
19	zug-	-verbindung							
2	zugemischt			zugemischt					
2	zuglast		zuglast						
2	zuglasten	s.o.							
4	zugprüfmaschine		prüfmaschine						
2	zukünf	f							
6	zul	f							
3	zuläßt	-							
2	zulieferfirmen					zuliefer			
2	zulieferindustrie					zuliefer			
2	zulieferkette					zuliefer			
3	zulieferketten					zuliefer			
3	zumischung		zumischung						
2	zusam	f							
2	zusammenh	f							
3	zusammenschluss	schluß (lingo)							
2	zuschneide-	-verbindung							
4	zustands-	-verbindung							

Anhang A-8

Suffixliste

Lingo	Suffix	Wortklasse	Präfix non (>1-/-alle)	Anmerkungen Lexikalisierungen
4	-ähnlich		0	
30	-arm		1	
21	-artig		0	
244	-bar		8 -/- 70	
1019	-chen		20 -/- 140	
57	-echt		0	
4	-		0	
25	-fähig		1	manövrierfähig
7	-fertig		0	
18	-fest		0	
20	-förmig		2	spiralförmig
26	-frei		0	
7	-gemäß		0	
9	-gerecht		0	
7	-treu		0	
1361	-haft		2	botschaft
6	-haltig		0	
355	-heit		1	sende #u
	-ig		18 -/- 105	
1194	-isch		1	elektromechanisch
928	-keit		21 -/- 152	
1	-leer		0	
727	-lich		9 -/- 22	
81	-ling		12	englisch
184	-los		0	
26	-mäßig		0	
105	-nis		0	
30	-recht		0	
61	-reich		3	meist -bereich
156	-rig		1	englisch
13	-sal		2	
64	-sam		1	
147	-schaft		1	botschaft
28	-seitig		5	siehe -ig
71	-sel		1	
6	-sicher		0	
	-t		-	-
11	-technisch		0	
7	-treu		0	
72	-tum		0	skriptum
2861	-ung		72 -/- 721	
22	-voll		0	
107	-weise		0	weise #u
39	-werk		3 -/- 22	
59	-wert		3	reibwert,
17	-wesen		0	
1	-widrig		0	
16	-würdig		0	
20	-zeug		1	flugzeug, werkzeug
				Ende heimischer Suffixe

Fortsetzung Anhang A-8

Suffixliste

Lingo	Suffix	Wortklasse	Präfix non (>1/-alle)	Anmerkungen Lexikalisierungen
49	-abel	e	1	Beginn fremder Suffixe
52	-ade		3	
152	-age		14	englisch
3	-aille		0	
11	-ain		9	englisch
2	-aire		0	englisch
6	-akel		0	
327	-al		58	meist englisch
24	-ale		8	siehe -al
27	-alisch		0	
29	-alistisch		0	generalistisch
66	-alität		2	kritikalität, dimensionalität
13	-ama		0	
151	-an		25	kalman, karman
4	-ana		2	
237	-and		12	
33	-auer		0	
60	-anisch		1	
5	-ans		3	
129	-ant		7	englisch
63	-anz		1	
374	-ar		30	englisch
83	-är		1	
243	-at		15	englisch, f
258	-ation		46 -/ 206	
94	-ator		8	
23	-do		1	
2	-eell		0	
0	-eill		0	
1	-ekel	s,a	0	
95	-elle	s,a	4	
6	-ematik		0	
30	-ement		9	englisch
47	-ende	s	3	
106	-enz		4	-frequenz
19	-eole		0	
46	-erie		0	
786	-ern		5	lenkflugkörper
0	-ernal		2	englisch
17	-erne		0	
52	-ese		0	
0	-esis		0	
3	-esk		0	
3	-eske		0	
1	-estrisch		0	
3	-etät		0	
1	-etes		1	quell
13	-etik		0	
6	-etion		1	
32	-etisch		0	

Fortsetzung Anhang A-8

Suffixliste

Lingo	Suffix	Wortklasse	Präfix non (>1/-alle)	Anmerkungen Lexikalisierungen
5	-ett		0	
63	-ette		0	
2	-emma	e	0	
1	-eumatik		0	
46	-eur		1	
6	-euse		1	reuse
213	-ent		38 -/- 115	
4	-eutik		0	
5	-eutisch		0	
45	-ial		10 -/- 35	englisch
16	-ain		9	englisch
1	-aina		0	
7	-ianer		0	
0	-iase		0	
0	-iasis		0	
17	-ibel		0	kompressibel,
2	-ica		0	englisch
8	-ice	w	3	englisch
71	-id		13	meteorid, oxid, monoxid, dioxid, nitrid, rfid
584	-ie		12	orthoskopie
40	-iell		0	vektoriell
0	-iens		0	
2	-iente		0	transiente
5	-ienz		0	
177	-ier		5	meist #e
16	-iere		0	
1210	-ieren		1	
29	-ifizieren		0	
24	-ika		0	charakteristika
7	-ikel		2	
102	-iker		0	
4	-ikon		0	
16	-ikum		1	charakteristikum
3	-ikus		0	
96	-il		16	email
9	-ill		1	englisch
29	-ille		0	
14	-ima		1	
9	-iment		0	
597	-in	t	32	viel englisch
96	-ine		14	subroutine
38	-iner		2	
23	-ikisch		0	
22	-io		3	
3	-iole		0	
694	-ion	s	87 -/- 376	viel englisch
5	-ios		1	helio
20	-iös		0	
3	-iosität		0	

Fortsetzung Anhang A-8

Suffixliste

Lingo	Suffix	Wortklasse	Präfix non (>1-/-alle)	Anmerkungen Lexikalisierungen
154	-ise		4	englisch
215	-ismus		0	
166	-ist	t	3	
258	-istik		1	
24	-istisch		0	
133	-it		39 -/- 271	
1468	-ität		7	
224	-itiden		0	
0	-itik		0	
16	-itis		0	
9	-itisch		0	
18	-ium		2 -/- 19	astrium,
91	-iv		3	
187	-ive		12 -/- 35	englisch
22	-ivisch		0	
5	-ivität		9	emissivität
26	-ivum		0	
1	-izismus		0	
0	-izität		0	
10	-kel		6 -/- 29	
82	-ktion		1	
103	-ma		6	snecma
80	-mat		0	
11	-matik		0	1
22	-matisch		0	
34	-matismus		0	
4	-men		7	
543	-ment		11 -/- 65	
69	-min		2	
13	-minal		0	
4	-minös		0	
2	-mos		0	
2	-mus		0	vulkanismus, vandalismus
182	-ns		24 -/- 146	
56	-nt		51	komp mit brems, reib sonst englisch
645	-nz		6	
181	-o		45	
292	-od		7	
11	-ode	s	8	
31	-oden		2	
13	-oid		2	meteorid
11	-oir		0	
3	-oire		0	
4	-olle		1	
4	-om		11	englisch
34	-oma		0	
5	-omat		0	
3	-omatik		0	
7	-on		109 -/- 485	siehe -tion

Fortsetzung Anhang A-8

Suffixliste

Lingo	Suffix	Wortklasse	Präfix non (>1/-alle)	Anmerkungen Lexikalisierungen
888	-one		3	
28	-ont		5	
27	-or		35 -/- 140	<vor> sonst alles englisch
275	-ör		0	
8	-os		12	helios
215	-ös		0	
44	-ose		6	englisch
57	-osis		0	
2	-osität		1	
8	-ostik		0	
2	-ot		13	eurokot
75	-ote		1	
22	-oten		0	
25	-otisch		0	
15	-se		34	meist Flexion -e
4	-sil		0	
14	-sis		1	
31	-siv		0	korrosiv, inklusiv, adhäsiv
27	-sor		3	
3	-sorisch		0	
1	-sorium		0	
102	-sion		8	englisch
14	-sur		0	
230	-tät		7	duktilität
2	-tätisch		0	
542	-ter		39 -/- 175	
14	-til		2	
563	-tion		87	siehe -ion
148	-tiv		1	konvektiv
195	-tor	s	20	siehe -or
33	-torisch		0	
12	-torium		0	
2	-trice		0	
5	-trix		3	Komposita mit -Matrix
5	-tron		0	englisch
2	-tude		3	magnitude, sonst englisch
4	-tüde		0	
79	-tur		4	kühlgas, heißgas
19	-uell		0	
119	-uend		0	
20	-ul		6	
4	-ula		2	
0	-üle		0	
8	-ulle		0	
2	-ulum		0	
316	-um		10	astrium, perigäum,
10	-ument		1	
38	-un		4	meist Fragmente
52	-unde		0	

Fortsetzung Anhang A-8

Suffixliste

Lingo	Suffix	Wortklasse	Präfix non (>1/-/alle)	Anmerkungen Lexikalisierungen
0	-uole		0	
194	-ur		11	Komp und - Verbind. Mit temperatur + struktur
12	-üre		0	
2	-urn		3	englisch
1	-urne	s	0	
0	-urno		0	
383	-us		15	
34	-use		4	englisch
0	-usis		0	
3	-ustik		0	
0	-ustr		0	Fragment
0	-ustrisch		0	
13	-ution		2	
5	-uum		1	englisch
4	-xion		0	
0	-xur		0	
15	-yse		1	Komp und -verbindung mit analyse
2	-ysis		0	englisch

Erklärung:

Spalte 1: Anzahl Lingo Wörterbucheinträge mit dem entsprechenden Suffix

Spalte 2: Suffixe, grün hinterlegte Suffixe werden einzeln untersucht

Spalte 3: Wortklasse bereits lexikalisierten Suffixe

Spalte 4: Anzahl nicht erkannte Token die mindestens in 2 Dokumenten vorkommen und für die grün hinterlegten Suffixe zusätzlich alle Token

Spalte 5: Anmerkungen und Lexikalisierungen

Anhang A-9-1

Token mit Zeichenfolge -ation

Anz.	Token		#s	#a	#v	#u	#e
20	aviation	englisch					
8	communication	englisch					
7	validation	englisch					
7	estimation	englisch					
6	configuration	englisch					
6	application	englisch					
6	elevation	englisch					
5	duration	englisch					
5	coordination	englisch					
5	location	englisch					
5	determination	englisch	determination				
5	calculation	englisch					
4	acceleration	englisch					
4	corporation	englisch					
4	verification	englisch					
4	certification	englisch					
4	computation	englisch					
4	qualification	englisch					
4	fe-simulation	-verbindung					
4	identification	englisch					
3	bwbkonfiguration	f					
3	classification	englisch					
3	telecommunication	englisch					
3	inclination	englisch					
3	condensation	englisch					
3	equation	englisch					
3	combination	englisch					
3	calibration	englisch					
3	optimization	englisch					
2	representation	englisch					
2	backpropagation	englisch					
2	authorisation	englisch					
2	utilization	englisch					
2	actuation	englisch					
2	specification	englisch					
2	cross-validation	-verbindung					
2	concentration	englisch					
2	consolidation	englisch					
2	constellation	englisch					
2	cooperation	englisch					
2	investigation	englisch	investigation				
2	hilsimulation	f					
2	d-simulation	-verbindung					
2	foundation	englisch					
2	elongation	englisch					
2	orientation	englisch					

Anhang A-9-2

Token mit Zeichenfolge -bar

Anz.	Token		#s	#a	#v	#u	#e
5	mbar	abkürzung					
3	handhabbar			handhabbar			
2	hinnehmbar			hinnehmbar			
2	reduzierbar			reduzierbar			
2	kommandierbar			kommandierbar			
2	überwachbar			überwachbar			
2	skalierbar			skalierbar			
2	identifizierbar			identifizierbar			

Anhang A-9-3

Token mit Zeichenfolge -chen

Anz.	Token		#s	#a	#v	#u	#e	#w
78	solchen							solchen
10	akustischen			akustisch				
9	elektromagnetischen					elektro		
8	entsprachen	-						
8	schen	-						
5	-fachen	-verbindung						
4	volumetrischen	-						
4	elektrostatischen					elektro		
3	berechnen	f						
3	reynoldsschen						reynold	
3	duroplastischen					duro#		
2	schaltflächen					schalt		
2	robotischen			robotisch				
2	elektromechanischen					elektro		
2	machschen						mach	
2	transonischen	f						
2	elektrochemischen					elektro		
2	hydromechanischen					hydro		
2	monokristallinen					mono		
2	magnetoplasmadynamischen					magneto		
1	elastoplastischen			plastisch		elasto		

Anhang A-9-4

Token mit Zeichenfolge -ent

Anz.	Token		#s	#a	#v	#u	#e	#w
12	equipment	englisch						
11	development	englisch						
10	measurement	englisch						
9	environment	englisch						
9	component	englisch						
8	fluent	englisch						
7	descent	englisch						
6	event	englisch						
6	efficient	englisch						
5	department	englisch						
5	ent	f						
5	concurrent	englisch						
4	document	englisch						
4	displacement	englisch						
4	compartment	englisch						
4	assessment	englisch						
4	dependent	englisch						
3	treatment	englisch						
3	excellent	englisch						
3	agreement	englisch						
3	requirement	englisch						
3	placement	englisch						
3	different	englisch						
2	convenient	englisch						
2	adjustment	englisch						
2	attachment	englisch						
2	bremskoeffizient						brems	
2	bremsmoment						brems	
2	reibmoment						reib	
2	qomponent	f						
2	content	englisch						
2	equivalent	englisch						
2	current	englisch						
2	judgement	englisch						
2	increment	englisch						
2	improvement	englisch						
2	abatement	englisch						
2	present	englisch						

Anhang A-9-5

Token mit Zeichenfolge -ig

Anz.	Token		#s	#a	#v	#u	#e
9	fig	f					
5	vielfältig	-					
4	bordseitigen			bordseitig			
4	saugseitige			saugseitig			
4	druckseitigen			druckseitig			
4	heißgasseitigen			gasseitig			
3	modellseitig	-		modellseitig		seitig	
3	ängig	f					
3	spannweitig			spannweitig			
3	konfig	f					
3	bodenseitig			bodenseitig		seitig	
3	bordseitig			bordseitig		seitig	
3	s-förmig	-verbindung					
3	saugseitigen	s.o.					
3	druckseitige	s.o.					
3	modellseitig			modellseitig			
3	bordseitige	s.o.					
3	bordseitig	s.o.					
3	bodenseitig			bodenseitig			
2	spiralförmig			spiralförmig		seitig	
2	saugseitig	-					
2	herstellenseitig					seitig	
2	mehrzellig			mehrzellig			
2	manövrierfähig			manövrierfähig			
2	danzig						danzig
2	kleinräumig	-					
2	big	englisch					
2	hochwertig			hochwertig			
2	herstellenseitig			herstellenseitig			
2	gasseitigen	s.o.					
2	saugseitig	s.o.					
2	flugzeugseitigen			flugzeugseitig			

Anhang A-9-6

Token mit Zeichenfolge -ion

Anz.	Token		#s	#a	#v	#u	#e
20	aviation	englisch					
14	function	englisch					
10	protection	englisch					
8	communication	englisch					
7	validation	englisch					
7	estimation	englisch					
7	section	englisch					
6	elevation	englisch					
6	detection	englisch					
6	description	englisch					
6	precision	englisch					
6	propulsion	englisch					
6	configuration	englisch					
6	application	englisch					
5	condition	englisch					
5	prediction	englisch					
5	injection	englisch					
5	duration	englisch					
5	location	englisch					
5	direction	englisch					
5	determination	englisch					
5	coordination	englisch					
5	calculation	englisch					
5	production	englisch					
5	collision	englisch					
4	certification	englisch					
4	acceleration	englisch					
4	fe-simulation	-verbindung					
4	verification	englisch					
4	computation	englisch					
4	identification	englisch					
4	corporation	englisch					
4	correction	englisch					
4	qualification	englisch					
4	reduction	englisch					
4	introduction	englisch					
3	criterion	englisch					
3	conversion	englisch					
3	condensation	englisch					
3	combination	englisch					
3	competition	englisch					
3	attention	englisch					
3	acquisition	englisch					
3	telecommunication	englisch					
3	solution	englisch					
3	calibration	englisch					

Fortsetzung Anhang A-9-6

Token mit Zeichenfolge -ion

3	selection	englisch					
3	equation	englisch					
3	optimization	englisch					
3	bwbkonfiguration	f					
3	inspection	englisch					
3	inclination	englisch					
3	classification	englisch					
2	backpropagation	englisch					
2	assumption	englisch					
2	completion	englisch					
2	ascension	englisch					
2	combustion	englisch					
2	concentration	englisch					
2	actuation	englisch					
2	collection	englisch					
2	authorisation	englisch					
2	hilsimulation	f					
2	utilization	englisch					
2	tension	englisch					
2	suspension	englisch					
2	suppression	englisch					
2	specification	englisch					
2	richtfunktion	-					
2	representation	englisch					
2	orientation	englisch					
2	motion	englisch					
2	investigation	englisch					
2	distortion	englisch					
2	ignition	englisch					
2	consolidation	englisch					
2	foundation	englisch					
2	execution	englisch					
2	elongation	englisch					
2	x-position	-verbindung					
2	d-simulation	-verbindung					
2	decomposition	englisch					
2	cross-validation	-verbindung					
2	cooperation	englisch					
2	construction	englisch					
2	constellation	englisch					
2	interconnection	englisch					

Anhang A-9-7

Token mit Zeichenfolge -it

Anz.	Token		#s	#a	#v	#u	#e
30	unit	englisch					
13	it	f					
10	fit	englisch					
9	steuerbarkeit		steuerbarkeit				
7	mbit	abk					
6	manövrierbarkeit		manövrierbarkeit				
5	manövrierfähigkeit		manövrierfähigkeit				
4	-bit	-verbindung					
4	gegeschwindigkeit					geh	
4	positioniergenauigkeit	-					
3	krit	f					
3	wiederzündfähigkeit		zündfähigkeit				
3	gambit	-					
3	hit	englisch					
3	in-orbit	-verbindung					
3	wiederholgenauigkeit	-					
3	kbit	abk					
3	lochleibungsfestigkeit		lochleibung				
3	keit	f					
3	stabilit	f					
3	wait	englisch					
3	nachhaltigkeit		nachhaltigkeit				
3	-abhängigkeit	-verbindung					
2	init	f					
2	spirit	englisch					
2	crit	f					
2	darstellbarkeit		darstellbarkeit				
2	dünnwandigkeit		dünnwandigkeit				
2	fertigbarkeit	-					
2	sendeeinheit					sende	
2	rüstzeit	-					
2	langlebigkeit		langlebigkeit				
2	oberflächenrauigkeit		rauigkeit				
2	mindestfluggeschwindigkeit	-					
2	maßhaltigkeit	-					
2	-zeit	-verbindung					
2	welligkeit		welligkeit				
2	lernfähigkeit		lernfähigkeit				
2	regelbarkeit	-					

Anhang A-9-8

Token mit Zeichenfolge -keit

Anz.	Token		#s	#a	#v	#u	#e
9	steuerbarkeit		steuerbarkeit				
6	manövrierbarkeit		manövrierbarkeit				
5	manövrierfähigkeit		manövrierfähigkeit				
4	gegeschwindigkeit	-					
4	positioniergenauigkeit	-					
3	nachhaltigkeit		nachhaltigkeit				
3	wiederholgenauigkeit	-					
3	wiederzündfähigkeit		zündfähigkeit				
3	-abhängigkeit	-verbindung					
3	lochleibungsfestigkeit		leibung				
3	keit	f					
2	darstellbarkeit		darstellbarkeit				
2	dünnwandigkeit		dünnwandigkeit	dünnwandig			
2	oberflächenrauigkeit		rauigkeit				
2	regelbarkeit			regelbar			
2	mindestfluggeschwindigkeit	-					
2	fertigbarkeit	-					
2	lernfähigkeit	-					
2	langlebigkeit	-					
2	welligkeit	-					
2	maßhaltigkeit	-					

Anhang A-9-9

Token mit Zeichenfolge -kel

Anz.	Token		#s	#a	#v	#u	#e
14	schiebewinkel					schiebe	
6	hängewinkel					hänge	
4	elevationswinkel		elevation				
2	heinkel						heinkel
2	schlackepartikel		schlacke				
2	kleinstpartikel		kleinstpartikel				

Anhang A-9-10

Token mit Zeichenfolge -lich

Anz.	Token		#s	#a	#v	#u	#e
11	unumgänglich	-					
5	schlussendlich	-					
3	schliesslich	f					
3	ermöglich	f					
3	massgeblich	f					
2	lich	f					
2	ausschliesslich	f					
2	einschliesslich	f					
2	versehentlich	-					
1	ändlich						
1	ausschließlich						
1	zerfreundlich						
1	ersichtlich						
1	luftverkehrlich						
1	ächlich						
1	verdeutlich						
1	zusätzlich						
1	verschiedentlich						
1	uglich						
1	unabdinglich						
1	verglich						
1	jülich						

Anhang A-9-11

Token mit Zeichenfolge -ment

Anz.	Token		#s	#a	#v	#u	#e
12	equipment	englisch					
11	development	englisch					
10	measurement	englisch					
9	environment	englisch					
5	department	englisch					
4	document	englisch					
4	displacement	englisch					
4	compartment						
4	assessment	englisch					
3	placement	englisch					
3	treatment	englisch					
3	agreement	englisch					
3	requirement	englisch					
2	bremsmoment					brems	
2	increment	englisch					
2	attachment	englisch					
2	abatement	englisch					
2	improvement	englisch					
2	judgement	englisch					
2	adjustment	englisch					
2	reibmoment					reib	

Anhang A-9-12

Token mit Zeichenfolge -ns

Anz.	Token		#s	#a	#v	#u	#e
56	ins	-					
11	ns	f					
7	communications	englisch					
7	conditions	englisch					
5	equations	englisch					
5	applications	englisch					
4	ans	-					
3	rands	f					
3	gains	englisch					
2	pylons		pylon				
2	constellations	englisch					
2	considerations	englisch					
2	means	englisch					
2	genauestens	-					
2	pins		pin				
2	sections	englisch					
2	cns	f					
2	comparisons	englisch					
2	locations	englisch					
2	lins	abk					
2	lessons	englisch					
2	adkins	#e autor					
2	turns	englisch					
2	truderverhaltens		truderverhalten				

Anhang A-9-13

Token mit Zeichenfolge -o

Anz.	Token		#s	#a	#v	#u	#e	#t
291	so	-						so
72	o	f						
42	to	englisch						
21	icao	abk						
18	no	englisch						
16	co	f						
16	mio	f						
10	iso	abk						
9	leo	-						
8	do	englisch						
8	into	englisch						
8	lufo	f						
7	galileo						galileo	
6	lo	f						
5	cargo	englisch						
5	carlo	-						
5	zero	englisch						
5	h2o	abk						
5	two	englisch						
4	ko	f						
4	bo	f						
4	mta	f						
3	fo	f						
3	gto	f						
3	gyro	englisch						
3	too	englisch						
3	meo	f						
3	recaro						recaro	
2	electro	englisch						
2	eo	f						
2	faero	f						
2	who	englisch						
2	aktuato	f						
2	vsaero	f						
2	go	englisch						
2	uoo	f						
2	po	f						
2	io	f						
2	aervico	-						
2	meco	f						
2	soho	-						
2	n2o	abk						
2	ro	f						
2	oo	f						
2	visio	-						

Anhang A-9-14

Token mit Zeichenfolge -on

Anz.	Token		#s	#a	#v	#u	#e	#t
302	von							von
31	on	englisch						
20	aviation	englisch						
14	function	englisch						
12	common	englisch						
10	protection	englisch						
8	communication	englisch						
7	section	englisch						
7	validation	englisch						
7	estimation	englisch						
6	description	englisch						
6	propulsion	englisch						
6	configuration	englisch						
6	elevation	englisch						
6	detection	englisch						
6	application	englisch						
6	precision	englisch						
6	carbon	englisch						
5	condition	englisch						
5	coordination	englisch						
5	location	englisch						
5	injection	englisch						
5	production	englisch						
5	collision	englisch						
5	determination	englisch						
5	direction	englisch						
5	duration	englisch						
5	calculation	englisch						
5	prediction	englisch						
4	comparison	englisch						
4	identification	englisch						
4	introduction	englisch						
4	correction	englisch						
4	verification	englisch						
4	acceleration	englisch						
4	fe-simulation	-verbindung						
4	corporation	englisch						
4	reduction	englisch						
4	certification	englisch						
4	qualification	englisch						
4	computation	englisch						
3	criterion	englisch						
3	aileron	englisch						
3	combination	englisch						
3	conversion	englisch						
3	condensation	englisch						

Fortsetzung Anhang A-9-14

Token mit Zeichenfolge -on

3	competition	englisch						
3	calibration	englisch						
3	bwbkonfiguration	f						
3	acquisition	englisch						
3	attention	englisch						
3	equation	englisch						
3	classification	englisch						
3	iron	englisch						
3	reason	englisch						
3	mon	f						
3	selection	englisch						
3	solution	englisch						
3	optimization	englisch						
3	kapton	f						
3	non	englisch						
3	telecommunication							
3	upon	englisch						
3	inclination	englisch						
3	inspection	englisch						
2	orientation	englisch						
2	combustion	englisch						
2	piston	englisch						
2	utilization	englisch						
2	collection	englisch						
2	acton	f						
2	representation	englisch						
2	chevron	englisch						
2	beson	f						
2	actuation	englisch						
2	specification	englisch						
2	suppression	englisch						
2	tension	englisch						
2	ascension	englisch						
2	assumption	englisch						
2	backpropagation	englisch						
2	suspension	englisch						
2	authorisation	englisch						
2	richtfunktion					richt		
2	cooperation	englisch						
2	execution	englisch						
2	foundation	englisch						
2	hilsimulation	f						
2	d-simulation	-verbindung						
2	distortion	englisch						
2	horton	f						
2	ignition	englisch						
2	denton	englisch						

Fortsetzung Anhang A-9-14

Token mit Zeichenfolge -on

2	decomposition	englisch						
2	cross-validation	-verbindung						
2	nikon					nikon		
2	investigation	englisch						
2	elongation	englisch						
2	kon	f						
2	construction	englisch						
2	constellation	englisch						
2	consolidation	englisch						
2	m-carbon	-verbindung						
2	michelson	#e						
2	concentration	englisch						
2	motion	englisch						
2	x-position	-verbindung						
2	completion	englisch						
2	interconnection	englisch						

Anhang A-9-15

Token mit Zeichenfolge -ung

Anz.	Token		#s	#a	#v	#u	#e
21	x-richtung	-verbindung					
18	y-richtung	-verbindung					
12	z-richtung	-verbindung					
10	evaluierung		evaluierung				
10	vorauslegung		vorauslegung				
9	-richtung	-verbindung					
7	vorraussetzung		vorraussetzung				
6	schlussfolgerung	gef. Variator					
6	auftragung		auftragung				
5	fe-rechnung	-verbindung					
5	längung		längung				
4	impactbelastung	englisch					
4	klebeverbindung	-					
4	v-stellung	-verbindung					
3	bremsleistung					brems	
3	kettrichtung	f					
3	effusionskühlung		effusion				
3	-regelung	-verbindung					
3	fe-berechnung	-verbindung					
3	ölversorgung		ölversorgung				
3	auditierung		auditierung				
3	-optimierung	-verbindung					
3	-verteilung	-verbindung					
3	impactschädigung		impact				
3	tiefpassfilterung		tiefpass				
3	zentrierung		zentrierung				
3	portierung		portierung				
3	nährung	f					
3	ösung	f					
2	tung	f					
2	-überwachung	-verbindung					
2	größenordnung	v					
2	ömung	f					
2	heißgasströmung		heißgas				
2	berücksichtigung		berücksichtigung				
2	berechnung	f					
2	-verbindung	-verbindung					
2	hitzdrahtmessung		hitzdraht				
2	ung	f					
2	relaminarisierung	-					
2	sendeleistung		sendeleistung				
2	elevationsrichtung		elevation				
2	schraubverbindung					schraub	
2	schränkung		schränkung				
2	sichtbarmachung	-					
2	schlussbetrachtung	gef. Variator					

Fortsetzung Anhang A-9-15

Token mit Zeichenfolge -ung

2	schlussbemerkung	gef. Variator					
2	d-visualisierung	-verbindung					
2	-durchführung	-verbindung					
2	cp-verteilung	-verbindung					
2	fb-anstrengung	-verbindung					
2	-führung	-verbindung					
2	feberechnung	f					
2	stufung	f					
2	fe-modellierung	-verbindung					
2	femrechnung	f					
2	danksagung	-					
2	taillierung		taillierung				
2	thermalisierung		thermalisierung				
2	finitisierung		finitisierung				
2	ti-legierung	-verbindung					
2	traggasversorgung		traggas				
2	-steuerung	-verbindung					
2	mindestanforderung	-					
2	grenzschichtabsaugung	f					
2	ässigung	f					
2	vr-umgebung	-verbindung					
2	vordimensionierung		vordimensionierung				
2	ir-strahlung	-verbindung					
2	kraftstoffstufung	-					
2	young	englisch					
2	-aufbereitung	-verbindung					

Anhang A-9-16

Token mit Zeichenfolge -werk

Anz.	Token		#s	#a	#v	#u	#e
17	fahrwerk		fahrwerk				
3	bugfahrwerk	s.o.					
2	dreibeinfahrwerk	s.o.					
1	kemtriebwerk						
1	b-triebwerk						
1	e3referenztriebwerk						
1	einziehfahrwerk						
1	e-kerntriebwerk						
1	frontfahrwerk						
1	basisfahrwerk						
1	h-leitwerk						
1	zweiwellen- zweistromtriebwerk						
1	kufenfahrwerk						
1	mea-bugfahrwerk						
1	mea-fahrwerk						
1	tandemfahrwerk						
1	t-leitwerk						
1	vleitwerk						
1	v-leitwerk						
1	wankeltriebwerk						
1	zusatzfahrwerk						
1	getr1ebfantriebwerk						
			triebwerk				
			leitwerk				