

Dipl.-Bibl. Thomas Müller, FH Köln

Lingo

– ein System zur automatischen Indexierung –
Anwendung und Einsatzmöglichkeiten

GfKI 2010 – Karlsruhe

Objektbeschreibung

ID: 4798

HAUPTGRUPPE: (Innen-)Architektur

UNTERGRUPPE: Einrichtung

BEZEICHNUNG: Tisch: Verfassungskonvent Herrenchiemsee

PROVENIENZ: Besatzungszonen

DATIERUNG: 1948

MATERIAL: Holz

WISSENSCHAFTLICHE OBJEKTBESCHREIBUNG: Grundgesetz, Parlamentarischer Rat, Verfassungskonvent, Herrenchiemsee, Tisch.

AUSTELLUNGSTEXT: Grundlage für das Grundgesetz

Einen ausgeprägt föderalen Staat wollen die westdeutschen Ministerpräsidenten. In ihrem Auftrag erarbeitet ein Expertengremium in Schloß Herrenchiemsee 1948 Richtlinien für ein Grundgesetz: Arbeitsgrundlage für den Parlamentarischen Rat. Die föderalistischen Grundsätze werden umgesetzt. Föderalismus, Grundgesetz, Verfassung, Parlamentarischer Rat Herrenchiemsee; Bonn Der Tisch stammt aus dem Schloss Herrenchiemsee, wo im August 1948 eine Expertenkonferenz im Auftrag der westdeutschen Ministerpräsidenten Richtlinien für ein Grundgesetz erarbeitete. Diese Richtlinien bildeten die Grundlage für den Parlamentarischen Rat, der am 1.9.1948 in Bonn seine Arbeit aufnahm. Das Grundgesetz trat am 24.5.1949 in Kraft.

Volltextsuche

<i>Indexeinträge</i>	<i>Treffermenge</i>
Anschlag	643
Anschlags	59
Anschläge	308
Anschlägen	87
Kalter Krieg	5.166
Kalten Krieges	481
Kalten Kriegs	36
Kaltem Krieg	6
Sozialdemokratische Partei Deutschlands	573
SPD	14.985
Ministerium für Staatssicherheit	1.151
Stasi	2.810
MfS	3.915

Verschiedenartigkeit von Dokument- und Suchsprache

Aufschwung der Wirtschaft

(Wortfolge von Substantiven)

Wirtschaftsaufschwung

(Kompositum)

wirtschaftlicher Aufschwung

**(Adjektiv-Substantiv-
Verbindung)**

Unregelmäßige Pluralformen

Sprache - Sprachen

Haus - Häuser

Maus - Mäuse

Museum - Museen

Wortbildung durch Wortableitung

wirtschaftlich → Wirtschaft

Indexierungsfunktionalitäten

Erzeugung grammatikalischer Grundformen

Informationen → Information

Zerlegung von Komposita

Wirtschaftspolitik → Wirtschaft, Politik

Bildung von Wortableitungen

technisch → Technik

Erkennen von Mehrwortgruppen

'soziale Marktwirtschaft'

'Konrad Adenauer'

Relationierung von Synonymen

elektronische Datenverarbeitung → EDV

lingo-Module

- Tokenizer:*** Identifizierung sogenannter Token, d.h. Zeichenketten, die Wort sein können
- Wordsearcher:*** linguistische Analyse, Grundformreduzierung
- Decomposer:*** linguistische Analyse, Kompositumzerlegung
- Multiworder:*** erkennt lexikalisierte Mehrwortgruppen
- Sequencer:*** algorithmische Erkennung von Mehrwortgruppen anhand definierter Abfolgen von Wortklassen
- Synonymer:*** generiert lexikalisierte Synonyme
- vector_filter*** und ***textwriter:*** steuern die Ausgabe der Indexierungsergebnisse

Wörterbuchstruktur

Auszug Systemwörterbuch:

überschuss=überschuss #s
überschuß=überschuss #s
kongreß=kongress #s
kongress=kongress #s
ideologisch=ideologisch #a ideologie #s
militärisch=militärisch #a militär #s
moralisch=moralisch #a moral #s
staatlich=staatlich #a staat #s

Auszug Synonymwörterbuch:

amerikanische zone deutschlands=amerikanische
besatzungszone=amerikanische besatzungszone deutschlands
britische zone deutschlands=britische
besatzungszone=britische besatzungszone deutschlands
russische zone deutschlands=sowjetzone=sbz=sowjetische
besatzungszone deutschlands=sowjetische
besatzungszone=sowjetische zone deutschlands
gagarin, juri=gagarin, juri alexejewitsch=gagarin,
yuri=juri gagarin=juri alexejewitsch gagarin=yuri gagarin

Grundformerkennung

hdg:) <Anschlag = [(anschlag/s)]>

hdg:) <Anschlags = [(anschlag/s)]>

hdg:) <Anschläge = [(anschlag/s)]>

hdg:) <Anschlägen = [(anschlag/s)]>

hdg:) <kalter Krieg|MUL = [(kalter krieg/m), (ost-west-konflikt/y)]>

hdg:) <kalten Krieges|MUL = [(kalter krieg/m), (ost-west-konflikt/y)]>

hdg:) <kalten Kriegs|MUL = [(kalter krieg/m), (ost-west-konflikt/y)]>

hdg:) <kaltem Krieg|MUL = [(kalter krieg/m), (ost-west-konflikt/y)]>

Kompositumzerlegung / Wortableitung

hdg:) <Aufschwung = [(aufschwung/s)]>

hdg:) <der = [(der/t)]>

hdg:) <Wirtschaft = [(wirtschaft/s),
(wirtschaftlich/a), (ökonomie/y)]>

hdg:) <Wirtschaftsaufschwung|KOM =
[(wirtschaftsaufschwung/k), (aufschwung/s),
(wirtschaft/s)]>

hdg:) <wirtschaftlicher Aufschwung|MUL =
[(wirtschaftlicher aufschwung/m)]>

hdg:) <wirtschaftlicher = [(wirtschaft/s),
(wirtschaftlich/a), (ökonomie/y)]>

hdg:) <Aufschwung = [(aufschwung/s)]>

Relationierung von Synonymen (1)

- hdg:) <Sozialdemokratische Partei Deutschlands|MUL =
[(sozialdemokratische partei deutschland/m),
(spd/y)]>
- hdg:) <SPD = [(spd/s), (sozialdemokratische partei
deutschlands/y)]>
- hdg:) <Ministerium für Staatssicherheit|MUL =
[(ministerium für staatssicherheit/m), (mfs/y),
(ssd/y), (staatssicherheitsdienst der ddr/y),
(stasi/y)]>
- hdg:) <Stasi = [(stasi/s), (mfs/y), (ministerium für
staatssicherheit/y), (ssd/y),
(staatssicherheitsdienst der ddr/y)]>
- hdg:) <MFS = [(mfs/s), (ministerium für
staatssicherheit/y), (ssd/y),
(staatssicherheitsdienst der ddr/y), (stasi/y)]>

Relationierung von Synonymen (2)

hdg:) <gagarin , juri|MUL = [(gagarin, juri/m),
(gagarin, juri alexejewitsch/y), (gagarin,
yuri/y), (juri alexejewitsch gagarin/y), (juri
gagarin/y), (yuri gagarin/y)]>

hdg:) <yuri gagarin|MUL = [(gagarin, yuri/m), (yuri
gagarin/m), (gagarin, juri/y), (gagarin, juri
alexejewitsch/y), (gagarin, yuri/y), (juri
alexejewitsch gagarin/y), (juri gagarin/y)]>

unregelmäßige Pluralformen

hdg:) <Sprachen = [(sprache/s), (language/y)]>

hdg:) <Häuser = [(haus/s), (building/y),
(gebäude/y)]>

hdg:) <Mäuse = [(maus/s)]>

hdg:) <Museen = [(kunstmuseum/y), (museum/s)]>

Fazit (1)

sprachliche und semantische Vereinheitlichung der Indexterme

Bereitstellung von Indextermen in der bevorzugten grammatikalischen Form (für das Retrieval i.d.R. Substantiv Singular)

automatische Erweiterung des Suchindex durch Synonyme

Fazit (2)

deutliche Verbesserung der Retrievalleistung

signifikante Erhöhung des Recalls, d.h. der Zahl der gefundenen relevanten Dokumente in der Suche

Informationsverlust wird gesenkt

Download & weiterführende Informationen unter:

<http://lex-lingo.de/>

lingo online testen:

<http://blackwinter.de/lingoweb/>

Thomas Müller, Dipl.-Bibl. (FH)
Fachhochschule Köln
Institut für Informationsmanagement
Ubierring 40, D - 50678 Köln
Tel: +49 (0)221 8275-3569, E-Mail: thomas.mueller@fh-koeln.de