

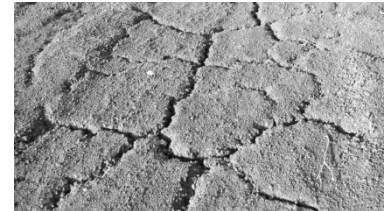
Terminologie extrahieren mit extraTerm

Ihre Terminologie auf den Weg gebracht

Sie haben noch keine firmeneigene Terminologie? Oder Sie wollen Ihre Terminologie auf der Grundlage Ihres Dokumentbestands erweitern?

Sie haben Terminologie bisher manuell extrahiert? Oder Sie setzen bereits Tools für die Termextraktion ein?

Unsere automatische Termextraktion bietet Ihnen die ideale Unterstützung für grundlegende Arbeitsschritte beim Terminologieaufbau. Vergleichen Sie auch die Ergebnisse Ihrer bisherigen Termextraktionsmethoden mit den Ergebnissen von extraTerm und überzeugen Sie sich von den Vorzügen. Wir scheuen keinen Vergleich!



Linguistische Analyse

Bei der Analyse Ihrer Textdaten setzen wir die ausgereiften sprachtechnologischen Verfahren von LCCore ein. Dabei ermittelt die morphologische Analyse für jedes einzelne Wort die Grundform, die Wortart und zugehörige grammatische Information wie z.B. das Genus von Substantiven. Durch die automatische Grundformenberechnung werden Vorkommen unterschiedlich gebeugter Formen zusammengefasst (z.B. *Dieselmotor*, *Dieselmotoren*, *Dieselmotors*). Die grammatische Analyse erkennt Wortgruppen und Satzbaumuster. Auf dieser Basis werden Termkandidaten unterschiedlicher Bildungsmuster ermittelt:

- Nominalphrasen bestehend aus einem Adjektiv und einem Substantiv wie z.B. *dieselhydraulischer Antrieb*
- zusammengesetzte Substantive wie z.B. *Dieselmotor*
- einfache Substantive wie z.B. *Differenzial*

Statistische Auswertung

Statistische Auswertungsverfahren ermitteln die Vorkommenshäufigkeit der Termkandidaten und die Anzahl der Belegdokumente pro Term. Diese Zahlen unterstützen Sie bei dem Entscheidungsprozess für die Festlegung Ihrer Terminologie.

Features

- lexikografische Grundform
- grammatisches Geschlecht
- unterschiedliche Bildungsmuster
- Vorkommenshäufigkeit
- Anzahl der Belegdokumente
- Liste der Belegdokumente
- Vorkommenskontexte
- Aufbereitung als Excel-Tabelle

Sprachen

- Deutsch
- Englisch

lfd Nr	Grundform	Originalwort	Wortart	Genus	Kriterium	Anzahl Belegstellen	Anzahl Belegdokumente	Belege
197	dieselhydraulischer Antrieb	Dieselhydraulischer Antrieb dieselhydraulischen Antriebs	adj noun	m	np	2	1	Dieselektrischer_Antrieb.html
198	Diesekraftstoff	Diesekraftstoff	noun	m	compound	2	1	Hybridantrieb.html
199	Diesellokomotive	Diesellokomotiven	noun	f	compound	1	1	Dieselektrischer_Antrieb.html
200	dieselmechanischer Antrieb	Dieselmechanischer Antrieb	adj noun	m	np	1	1	Dieselektrischer_Antrieb.html
201	Dieselmotor	Dieselmotor Dieselmotoren Dieselmotors	noun	m	compound	9	2	Dieselektrischer_Antrieb.html Hybridantrieb.html
202	Differenzial	Differenzialen	noun	n	simplex	1	1	Planetenantriebe.html

Kontextermittlung

Die Ermittlung der Vorkommenskontexte von Termkandidaten basiert auf der Kombination von statistischen und linguistischen Verfahren. Hierbei werden die Kontexte, in denen ein Termkandidat verwendet wird, auf linguistische Plausibilität überprüft. So können Sie sich bei der Sichtung der Kontexte schnell und umfassend ein Bild von der sprachlichen und fachlichen Verwendung eines Termkandidaten machen.

Zuordnung von Metadaten

Durch die Zuordnung von Dokument-Metadaten gemäß Ihren speziellen Anforderungen (z.B. Dokumentart, Baugruppe, Abteilung) können wir Ihnen spezifische Sichten auf die Extraktionsergebnisse ermöglichen. So erfahren Sie beispielsweise, welche Terme abteilungsübergreifend verwendet werden und welche nicht oder welche Terme für bestimmte Dokumentarten oder Baugruppen spezifisch sind.

Berücksichtigung der Bestandsterminologie

Bei Durchführung einer Termextraktion können wir auch Ihre Bestandsterminologie berücksichtigen. Terme, die schon in Ihrer Terminologie erfasst sind, werden gesondert ausgezeichnet. So können Sie sich ein umfassendes Bild davon machen, wie häufig und in welchen Dokumenten Bestandsterme überhaupt vorkommen. Sie wissen abschließend, welche Terme dringend in Ihrer Terminologie zu ergänzen sind und welche Bestandsterme offensichtlich irrelevant sind, da Sie in keinem oder kaum einem Dokument verwendet werden.

Evaluierung der extrahierten Terme

Kombinieren Sie die Termextraktion mit einer Evaluierung auf der Grundlage von **evalTerm**. So erfahren Sie beispielsweise, welche Termkandidaten nur in falscher Schreibung belegt sind, welche Termkandidaten unter Termbildungsgesichtspunkten problematisch sind (z.B. sehr komplexe Wortzusammensetzungen ohne Bindestrich wie *Umleitdampfeinführungsversteifungsrohr*) oder welche Termkandidaten möglicherweise Varianten voneinander sind (z.B. *Abdichtband* - *Dichtband* - *Dichtungsband*).

Weitere Informationen erwünscht?

Kontaktieren Sie uns mit Ihrem Anliegen zur Extraktion von Terminologie aus Ihren Dokumentbeständen. Wenn Sie spezielle Anforderungen an die Termextraktion haben, können wir Ihnen sicherlich weiterhelfen.

IAI Linguistic Content AG

Martin-Luther-Str. 14
66111 Saarbrücken
0681 38951-0

www.iailc.de
kontakt@iailc.de