

# Termextraktion aus bestehenden Dokumenten

## Terminologiewerk

Terminologiewerk gehört zu den zentralen Tätigkeiten in der Technischen Kommunikation. Am Anfang steht dabei die Erfassung und Festlegung neuer Terminologie – viel Aufwand für Unternehmen, die gerade erst mit der professionellen Terminologiewerk beginnen. Große Mengen von Termen sind bereits in bestehenden Dokumenten vorhanden und müssen aufwändig extrahiert werden. Congree bietet Ihnen deshalb eine Termextraktion mit Hilfe sprachtechnologischer Verfahren. Danach werden alle relevanten Termkandidaten ausgegeben und mit wichtigen Informationen wie der Grundform oder Wortart angereichert. Als Ergebnis erhalten Sie eine Liste von Benennungen, die Sie nach Durchsicht in Ihr Terminologiesystem importieren und für den Aufbau Ihrer firmeneigenen Terminologie verwenden können. Zusätzliche linguistische und statistische Daten zu allen Termkandidaten sowie die Kontexte, in denen ein Term bisher verwendet wurde, bieten Ihnen eine sinnvolle Unterstützung bei der Festlegung Ihrer Terminologie.

## Anwendungsfälle

Ein Unternehmen mit einem elektronisch verfügbaren Dokumentbestand möchte mit der Terminologiewerk beginnen und vorhandene Dokumente als Quelle für den Terminologiewerk nutzen.

Ein Unternehmen möchte seinen bereits vorhandenen Terminologiewerkbestand substanziell erweitern, und zwar auf Basis eines noch nicht abgedeckten, elektronisch verfügbaren Dokumentbestands.

## Unsere Leistungen

### Was Sie von uns erwarten können:

- » Extraktion der Termkandidaten in ihrer Grundform
- » Ermittlung des grammatischen Genus
- » Ermittlung der Vorkommenshäufigkeit
- » Ermittlung der Belegdokumente
- » Ermittlung der Vorkommenskontexte
- » Aufbereitung der Ergebnisdaten als Excel (importierbar z. B. in Congree oder Across)
- » Ausführlicher Ergebnisbericht

## Ihr Beitrag

### Zur Extraktion der Terminologie benötigen wir von Ihnen:

- » Dokumente, aus denen Termkandidaten extrahiert werden sollen (DOC/DOCX, HTML, XML, weitere Formate auf Anfrage)
- » vorhandene Terminologie als Datenbankexport oder csv (optional)
- » Termkandidatenfilter aus Congree UMMT oder sonstige Stoppwortliste (optional)
- » Benutzerwörterbuch aus Congree UMMT (optional)