

Innovative Terminologieextraktion

Mithilfe von KI-gestützte Prozessen das Know-how Ihres Unternehmens fixieren.



Terminologieextraktion ist ein wesentlicher Schritt im Bereich des Terminologiemanagements und der Sprachtechnologie. Dieser Prozess befasst sich mit der automatischen Identifizierung und Extraktion spezifischer Fachbegriffe und Terminologien aus großen Textmengen oder Dokumentensammlungen. Ziel der Terminologieextraktion ist es, relevante Fachbegriffe zu erkennen, zu klassifizieren und in einer speziellen Terminologiedatenbank zu speichern.

Durch den Einsatz von statistischen Methoden, maschinellem Lernen und Natural Language Processing (NLP) können Terminologieextraktionswerkzeuge komplexe Texte analysieren und die am häufigsten verwendeten und spezifischen Begriffe identifizieren. Diese extrahierten Termini sind von besonderem Interesse für Branchen mit spezifischen Fachgebieten, wie zum Beispiel in der Medizin, Technik oder Rechtswissenschaften.

Die gewonnenen Fachterminologien aus der Terminologieextraktion dienen als Grundlage für die Erstellung von Terminologiedatenbanken, die in weiteren Schritten des Terminologiemanagements und Übersetzungsprozessen

verwendet werden. Eine präzise und effektive Terminologieextraktion trägt dazu bei, die Qualität und Konsistenz in der Kommunikation zu verbessern, Übersetzungsprozesse zu optimieren und die Genauigkeit von Texten in spezialisierten Fachgebieten zu gewährleisten.

Vorteile des Einsatzes von LLMs für Terminologieextraktion: LLMs sind besonders nützlich, da sie Kontext verstehen und semantische Beziehungen erkennen können. Sie sind in der Lage, Begriffe und ihre Übersetzungen präziser zu identifizieren und Ähnlichkeiten besser zu bewerten als herkömmliche regelbasierte Ansätze.

Was bietet die KERN AG im Bereich Terminologieextraktion an?

Die KERN AG unterstützt bei der Terminologieextraktion bzw. bei der Fixierung des Firmen-Know-hows mit Fachkenntnis, langjähriger Erfahrung im Bereich des Terminologiemanagements und technischen Extraktionsmöglichkeiten. Eine Extraktion firmenbezogener Fachbegriffe und Begrifflichkeiten bietet die Möglichkeit, die interne Sprachkonsistenz sicherzustellen und festzulegen.

Unter Zuhilfenahme von Extraktionstools kann von einem muttersprachlichen Übersetzer / einer muttersprachlichen Übersetzerin oder einem Terminologen / einer Terminologin in einem aufwendigen Verfahren eine Terminologieextraktion durchgeführt werden. Die Extraktion erfolgt anhand des Ausgangstexts und der jeweiligen Übersetzung, sie kann aber auch einem Übersetzungsprojekt vorgelagert werden. Die Extraktionstools generieren eine Liste von Terminologie-Kandidaten, die dann von Ihrem persönlichen Projektmanager selektiert und mit Ihnen abgestimmt werden.

Wie kann der Prozess für die Terminologieextraktion beschrieben werden?

Der Prozess der Terminologieextraktion und -validierung umfasst mehrere Schritte, die sowohl automatisiert als auch manuell durchgeführt werden. Dies stellt sicher, dass die extrahierten Begriffe präzise und anwendungsorientiert sind:

- Bilinguale Term-Extraktion
- Manuelle Validierung durch einen Terminologen
- Anpassung der Termbank für spezifische Anwendungszwecke

Auf Basis von welchen Parametern wird die Term-Extraktion durchgeführt?

- Source Language (Ausgangssprache)
- Target Language (Zielsprache)
- Maximum Term Length (die maximale Länge der Termini)
- Similarity (Ähnlichkeit der Begriffe zwischen Ausgangs- und Zielsprache)

Welche Eingabe- und Ausgabeformate sind für eine Term-Extraktion möglich?

Eingabeformate: Die Term-Extraktion unterstützt verschiedene bilinguale Dateiformate für die Verarbeitung mehrsprachiger Inhalte. Dazu gehören das XML-basierte XLIFF-Format, das speziell für Übersetzungsprozesse entwickelt wurde, sowie TMX für den Austausch von Übersetzungsspeichern. Auch tabellarische Formate wie XLSX und CSV können verwendet werden, um Begriffe und deren Übersetzungen zu verarbeiten.

Ausgabeformate: Die extrahierten Begriffe können in verschiedenen Formaten exportiert werden: als Excel-Tabelle (XLSX) oder CSV-Datei, die eine detaillierte Auflistung aller gefundenen Begriffe enthält. Für die professionelle Terminologieverwaltung steht das TBX-Format zur Verfügung, das speziell für terminologische Datenbanken entwickelt wurde.



Innovative Terminologieextraktion mithilfe von KI-gestützte Prozessen

Die Terminologieextraktion mit Python und LLMs bietet eine automatisierte, kontextbezogene Methode zur effizienten Identifizierung und Übersetzung mehrsprachiger Begriffe. Mit flexiblen Parametern wie „Similarity“ und „Maximum Term Length“ lässt sich der Prozess individuell anpassen. Durch die Kombination mit terminologischer Expertise wird eine präzise Termbank erstellt, deren manuelle Überprüfung die korrekte Fachsprache sicherstellt. Die strukturierte Ausgabe in Excel oder TBX-Format sowie die Anpassung an spezifische Anforderungen (CMS, TMS, SEO) ermöglichen eine nahtlose Integration in bestehende Workflows und verschiedene Systeme. Die Kern AG garantiert dabei höchste Übersetzungsqualität durch ihre erfahrenen Fachübersetzer und modernste KI-gestützte Technologie.

www.e-kern.com



KERN AG, Sprachendienste

Kurfürstenstraße 1 · 60486 Frankfurt am Main · Tel. (069) 75 60 73 - 0 · Fax (069) 74 99 98 · info@e-kern.com

Deutschland: Aachen · Augsburg · Berlin · Bielefeld · Bochum · Bonn · Braunschweig · Bremen · Darmstadt · Dortmund · Düsseldorf · Essen · Frankfurt am Main · Freiburg · Friedrichshafen · Hamburg · Hannover · Heilbronn · Karlsruhe · Kassel · Köln · Leipzig · Mainz · Mannheim · München · Münster · Nürnberg · Regensburg · Saarbrücken · Stuttgart · Ulm · Wiesbaden · Wuppertal · Würzburg

Niederlande: Amsterdam · Eindhoven · Rotterdam · Utrecht · **Frankreich:** Lyon · Marseille · Paris · **Österreich:** Graz · Innsbruck · Linz · Wien · **Polen:** Warschau · **England:** London
Vereinigte Staaten: New York · San Francisco