

Wie klappt die Termextraktion?

Mit der Termextraktion lässt sich der Schatz eines Unternehmens heben. Am Ende steht die Terminologie – einsprachig oder sogar in mehreren Sprachen. Worauf zu achten ist, damit die Extraktion ihre Aufgabe erfüllt, darüber haben sich die Teilnehmer im Termcafé ausgetauscht.

TEXT *David Bodenstein, Beate Früh und Markus Nickl*

Wer hätte das nicht gerne: Man packt seine Dokumente in einen Trichter, wirft die Extraktionsmaschine an und unten purzeln die Termkandidaten heraus. Idealerweise sind sie schon vorqualifiziert und mit Definitionen versehen. Ganz so einfach funktioniert es in der Realität natürlich nicht. Wenn man aber einige Faktoren beachtet, dann lässt sich die Termextraktion sinnvoll und ergebnisorientiert nutzen.

Grundsätzliche Fragen stellen

Wie bei so vielem gilt auch bei der Termextraktion: Gute Planung ersetzt eine Menge Nacharbeit. Das geht schon bei den Grundlagen los, etwa bei der Überlegung, ob die Termextraktion einsprachig oder mehrsprachig abläuft. Beide Richtungen haben ihre Berechtigung; mehrsprachige Extraktion bedeutet aber – logischerweise – einigen Zusatzaufwand und setzt Kenntnisse der Fremdsprachen voraus, die in die Extraktion mit einfließen.

Überhaupt spielt es eine große Rolle, wer die Extraktion vornimmt und welche Dokumente als Basis dienen sollen, denn daran orientiert sich oft die Wahl des richtigen Tools. So haben Terminologinnen und Terminologen zum Aufbau des Erstbestandes einer Terminologiedatenbank oder zumindest Terminologiesammlung ein hohes Interesse daran, möglichst viele (relevante) Dokumente zu analysieren und die Termkandidaten daraus zu entnehmen.

Unterschiedliche Arbeitsweisen

Fachleute für Terminologie nutzen im Allgemeinen sowie je nach Verfügbarkeit und Budget statistische und linguistische Extraktionstools oder Konkordanzwerkzeuge. In der Übersetzung orientiert man sich dagegen meistens am Einzeldokument oder den Dokumenten eines Auftrages und arbeitet eher punktuell. Ansatzpunkt ist häu-

fig das Translation Memory. Generell kommen auch bei der übersetzungsorientierten Termextraktion oft Tools der Übersetzungsumgebungen zum Einsatz.

In der Technischen Redaktion wiederum zielen wir eher auf die Dokumentqualität und damit auf eine mittlere Ebene; wir orientieren uns an Modulen oder Topics. So kommen hier Controlled-Language-Werkzeuge mit linguistischen Regelwerken zum Einsatz. Diese weisen potenzielle Termkandidaten für die Terminologieex-

traktion aus, im Idealfall verbunden mit Angaben zur Häufigkeit des Vorkommens sowie zum Kontext, außerdem unter Berücksichtigung vorhandener Terminologiebestände. Allerdings kann es zu Problemen kommen, etwa wenn linguistische Tools nur Substantive als Termkandidaten vorschlagen, wir aber auch die Bezeichnungen für Tätigkeiten normieren wollen.

Der Kern von Termextraktion

Eine große Rolle spielt auch der Zweck der Termextraktion. Geht es zum Beispiel darum, einen Erstbestand aufzubauen? Oder will man Lücken in einem Terminologiebestand identifizieren? Sollen Regeln für die Benennungsbildung im Unternehmen ermittelt, Übersetzungen vorbereitet oder Übersetzungsfehler verhindert werden? Oder will man einen Überblick über die Synonyme im eigenen Content sowohl in der Ausgangssprache als auch in einer oder mehreren Zielsprachen bekommen? Jeder Zweck wirkt sich unterschiedlich aus. Gilt es, Lücken zu schließen, dann möchte man möglichst wenige Termkandidaten angeboten bekommen, die schon im Termbestand erfasst sind (Deltaextraktion). Bei der Synonymsuche interessiert der Bestand schon eher. Bei der Benennungsbildung kann auch nur einsprachig gearbeitet werden, bei übersetzungsunterstützender Terminologie eher nicht.

Grob gesprochen lässt sich Termextraktion auf die Frage herunterbrechen: „Wer extrahiert was zu welchem Zweck?“ Jede Antwort macht sich unterschiedlich bei Budget, Tools, Laufzeit und Datenqualität bemerkbar. Eine perfekte Maschine zur Terminologieextraktion, wie eingangs beschrieben, gibt es leider nicht. Aber mit einer klugen Vorbereitung lassen sich Wege vermeiden, die man ansonsten unnötig gegangen wäre. ☺

Terminologie für Anfänger und Experten

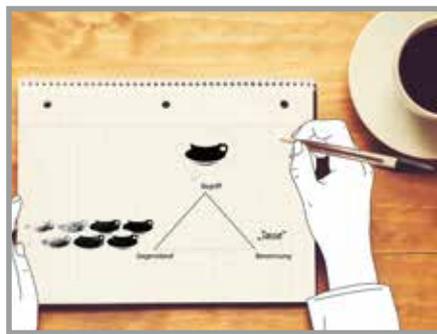


ILLUSTRATION CSH / istock

Das Termcafé ist eine kostenfreie Veranstaltungsreihe für Terminologie-Interessierte aller Richtungen, insbesondere auch für Technische Redakteure und Redakteurinnen. Sie wurde im September 2020 von Beate Früh (Büro b3), Markus Nickl (doctima GmbH) und David Bodensohn (itl AG) ins Leben gerufen. In lockerer Kaffeehausatmosphäre wird gemeinsam über Fragen aus der Terminologiepraxis gesprochen. Derzeit finden die Veranstaltungen digital statt. [Mehr unter https://www.termcafe.de/](https://www.termcafe.de/)