

Terminologie extrahieren - Die Spreu vom Weizen trennen

27.04.2017 07:26 von Julia Bischoff (Kommentare: 0)

In den letzten Wochen haben wir uns damit befasst, wie man Fachwörter und andere Termini aus vorhandenen Quellen im Unternehmen extrahieren kann.

Was man jetzt jedoch hat ist eine lange (oder nicht so lange) Liste an Wörtern die weiterverarbeitet werden muss, kann oder sollte. Klingt ominös? Ist es aber eigentlich gar nicht.

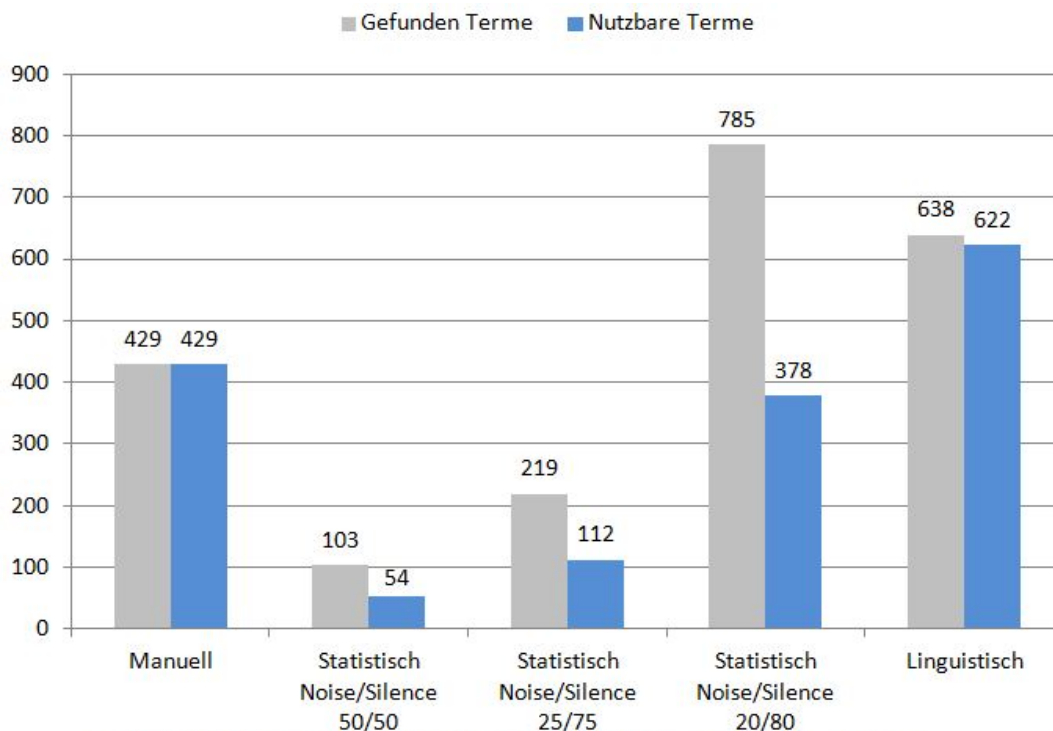
Es gibt noch eine Hürde zu überwinden, um die Spreu vom Weizen zu trennen:

Die Ergebnislisten unserer Extraktionen müssen nachbearbeitet werden um überhaupt brauchbare Terme zu erhalten.

Im letzten Blogbeitrag haben wir gezeigt, dass die linguistische Extraktion sehr gute Ergebnisse geliefert hat und das in Null-Komma-Nichts.

Es hat sich aber auch gezeigt, dass man die Ergebnisliste der statistische Extraktion noch stark nachbearbeiten muss, um eine brauchbare Termliste zu generieren.

In unserem Test ergab sich, dass die linguistische Extraktion die meisten Terme gefunden hat und von diesen auch die meisten nutzbar waren. Wenn auch eine der statistischen Extraktionen mehr Ergebnisse lieferte, so waren von diesen jedoch nur rund 50% nutzbare Terme.



Egal welche Art von Termextraktion man durchführt, die jeweiligen Ergebnislisten sind fast nie sofort benutzbar. Sie alle müssen mit mehr oder auch weniger Zeitaufwand nachbearbeitet werden, bevor überhaupt eine Bereinigung gestartet werden kann.

Wenn man Terme manuell oder statistisch extrahiert, muss man eine zeitintensive Nachbearbeitung durchführen: zum Beispiel Mehrwortkombinationen auflösen, Mehrzahl in Einzahl oder Verben in ihren Infinitiv umwandeln.

Bei der linguistischen Extraktion entfällt die zeitintensive Nachbearbeitung fast ganz. Hier muss hauptsächlich kontrolliert werden, ob unbekannte Wörter richtig rückflektiert wurden.

Erst wenn diese Nachbearbeitung durchgeführt wurde, kann man sich an die eigentliche Bereinigung der Terminologie heranwagen.

Eine umfangreiche Liste an Programmen zur Termextraktion und ein Glossar mit Fachwörtern finden sie in unserem [Basiswissen](#).

Einen Kommentar schreiben