

Suchen vereinfacht: Wie findet man die Nadel im wissenschaftlichen Heuhaufen?

Wir haben uns in diesen Blogbeiträgen bereits mit diversen Analysen von Publikationen und Patenten beschäftigt. Dabei gilt immer: Die Ergebnisse sind nur so gut, wie die Daten, auf denen sie basieren. Deshalb widmen wir uns heute dem vorgelagerten Schritt der Analyse: der Suche nach Daten. In KATI haben wir mit dem Web of Science bereits eine große Menge an Daten, aktuell fast 80 Millionen wissenschaftliche Publikationen. Die Kunst besteht somit darin, ähnlich der Nadel im Heuhaufen, die richtigen Daten zu finden. Nur weiß man noch nicht genau, wie diese Nadel überhaupt aussieht. Fest steht, die optimale Datenmenge umfasst genau die Publikationen, die zu dem Thema passen, nicht mehr und nicht weniger. In KATI geschieht die Suche über einen booleschen Suchstring, welcher iterativ verfeinert wird.

Um dies zu veranschaulichen, führen wir eine Suche zum Thema *Blutersatzprodukte* durch. Unter Blutersatzprodukten versteht man in der Medizin Substanzen, die verwendet werden, um den Blutverlust zu kompensieren und die Sauerstofftransportfunktion des Blutes nachzuahmen. Um eine Suchanfrage zu konstruieren, kann man nach dem *Schneeballprinzip* vorgehen. Das bedeutet, man startet mit einem Begriff und erweitert die Suchanfrage dann Stück für Stück basierend auf den gefundenen Ergebnissen. Wir starten zunächst mit dem Begriff *blood substitutes* und suchen in allen Titeln, Abstracts und Keywords. In KATI finden wir so fast 2.500 Publikationen. Die Manual Keywords, also die von den Autor*innen vergeben Schlagworte, weisen uns darauf hin, noch *hemoglobin base oxygen carrier* und *perfluorocarbon base oxygen carrier* zu der Suche hinzuzufügen. Iterativ ergänzen wir auch noch mit *artificial blood* und *artificial oxygen carrier*. Jetzt fällt uns auf, dass Publikationen gefunden werden, welche nicht zum Thema passen. Diese schließen wir aus, indem wir *artificial blood vessels* exkludieren. Damit sieht die Suchanfrage jetzt so aus:

```
("blood substitute*" OR "artificial blood*" OR "artificial oxygen carr*" OR ("hemoglobin base*" OR "perfluorocarbon base") AND "oxygen carr*") NOT ("artificial blood vessel*")
```

Dieses Beispiel zeigt, dass die Erstellung von Suchanfragen recht komplex werden kann. Obwohl wir noch am Anfang der Konstruktion einer guten Suchanfrage stehen, wird sie schon bei wenigen Begriffen unübersichtlich, geschweige denn, wenn noch Filter hinzukommen. Das ist so, als würden wir uns händisch durch den Heuhaufen wühlen. Klingt ineffizient? Ist es auch. Deshalb haben wir für KATI zusätzlich zu der manuellen Suchzeile einen sogenannten *Query Editor* entwickelt (siehe Abbildung 1). Dieser funktioniert nach einem Baukastenprinzip und hilft dabei, bei komplexen Suchanfragen nicht die Übersicht zu verlieren. Dazu werden die einzelnen Bestandteile der Suchanfrage übersichtlich dargestellt und können ohne tiefe Kenntnis der Syntax einfach editiert werden. Die visuelle Abgrenzung der Einzelteile der Suchanfrage durch Klammern ermöglicht es, die visuelle Darstellung so anzupassen, dass sie zu dem Inhalt der Suche passt. Außerdem gibt es verschiedene Darstellungsweisen (siehe Abbildung 2). Die Art, wie nach den Begriffen gesucht wird (bspw. *Phrasensuche* oder Suche mit *Proximity Operator*), lässt sich einfach durch Anklicken ändern, ebenso die booleschen Operatoren (siehe Abbildung 3). Es wird auf mögliche Syntaxfehler, beispielsweise in der Klammersetzung hingewiesen und die vielen Filtermöglichkeiten sind einfach auffindbar. Ein weiterer großer Vorteil: die Anzahl der Ergebnisse für die einzelnen Abschnitte der Suchanfrage wird angezeigt. Die bereits bekannten Filter in KATI werden außerdem ergänzt um den sogenannten *Condition Filter*. Dieser verknüpft einen Filter mit einer Zahl. So kann man beispielsweise die Bedingung stellen, dass die Anzahl der Autor*innen zwei sein soll, oder dass nur Publikationen mit mindestens fünf Zitationen gefunden werden sollen. Weiterhin gibt es für jeden Suchbegriff ein Kontextmenü, das es unter anderem erlaubt, den Begriff zu versetzen, zu duplizieren, zu negieren oder als *Blueprint* für

später zu speichern. Wir haben vielleicht noch keinen Magnet für die Nadel, aber der Query Editor hilft immerhin dabei, den Heuhaufen systematisch in kleinere Teile zu zerlegen.

Insgesamt hilft der Query Editor unseren Nutzer*innen, noch schneller und effizienter in KATI zu suchen und spart ihnen daneben auch noch Schreibarbeit. Da oft nur eine begrenzte Zeit zur Verfügung steht, um eine Suchanfrage aufzubauen, verbessert er indirekt so auch die Ergebnisse der Analyse. Dies ist aber noch nicht das Ende vom Lied, die Generierung einer optimalen Suchanfrage ist weiterhin eine offene Forschungsfrage, mit der auch wir uns befassen. Bis dahin, bleiben sie neugierig und datengetrieben.

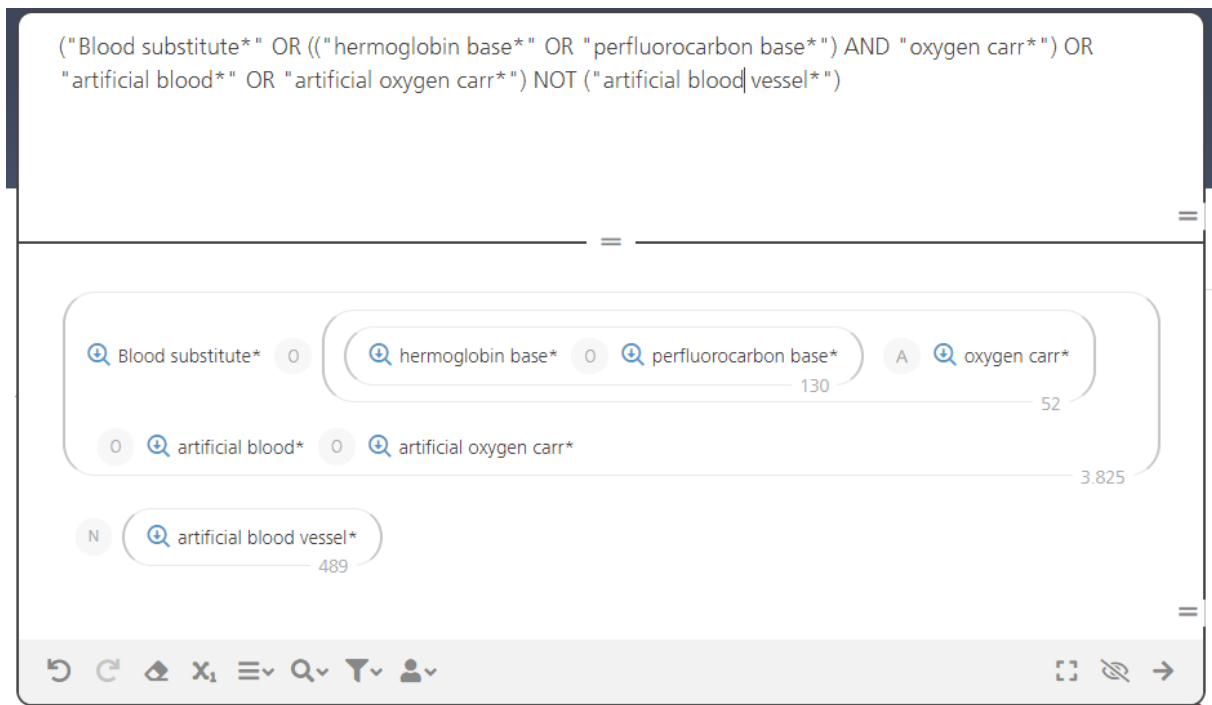


Abbildung 1: Search Query in manueller Eingabe und Query Editor

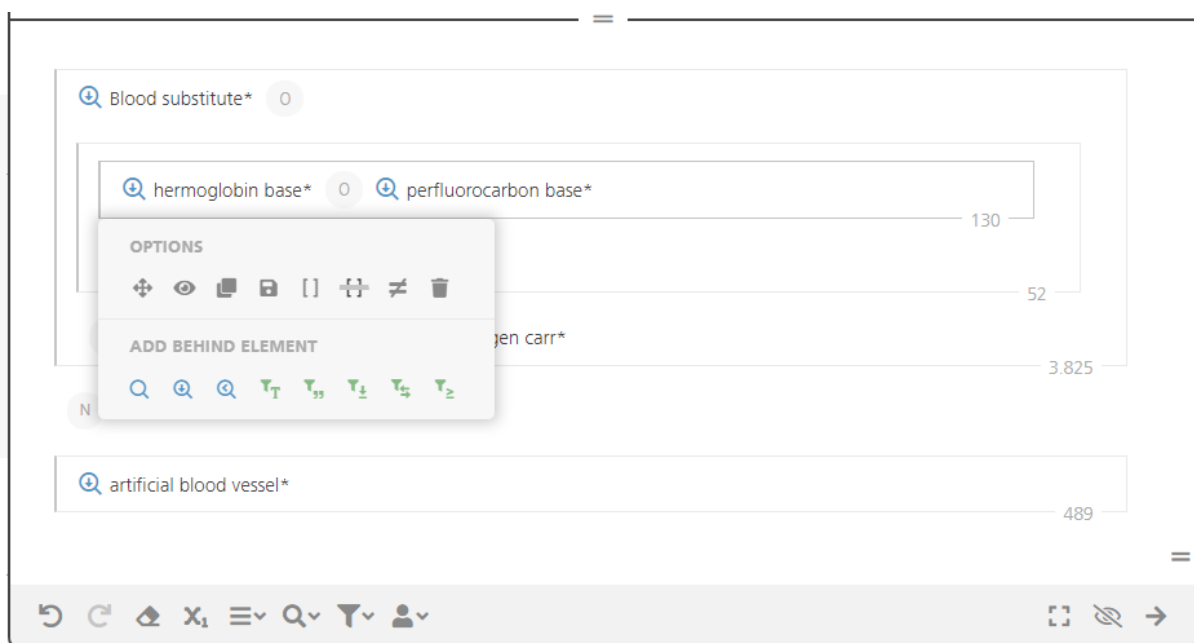


Abbildung 2: Query Editor mit anderem Layout und Optionen für Klammern

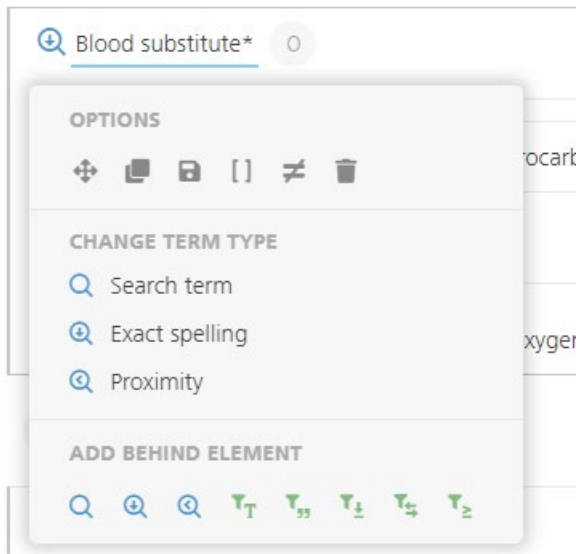


Abbildung 3: Query Editor Optionen für Begriffe