



eTblast – It's Only Words, And Words Are All I Have...

Die wissenschaftliche Literatur ist heute so umfangreich, dass man auf spezialisierte Suchprogramme zurückgreifen muss, um die relevante Information herauszufiltern. Chemisch orientierte Suchprogramme wie SciFinder verwenden meist Formeln zur Suche, während im medizinisch-biochemischen Bereich die übliche Suche noch immer textbasiert ist. Bei der Arbeit mit Formeln ist es längst gang und gäbe, dass man nicht nach einer genau festgelegten Verbindung suchen muss, sondern bei entsprechender Auswahl der Parameter – z. B. hier ein genau festgelegter Substituent, an jener Stelle hingegen ein beliebiger – auch das Umfeld erfasst. Bei der Textsuche aber (so auch bei der weit verbrei-

teten Suche in PubMed) erfolgt ein genauer Abgleich mit dem Suchbegriff – ist er nicht in der angegebenen Form vorhanden, wird das Dokument ärgerlicherweise nicht gefunden. So geht einem der eigentlich wichtige Artikel leicht „durch die Lappen“. Hier verspricht nun eTblast Abhilfe, ein Suchprogramm, das in der Gruppe von Harold Garner am Southwestern Medical Center der University of Texas in Dallas entwickelt wurde. Die faszinierende Idee dahinter: Man gibt nicht mehr einen einzelnen Suchbegriff vor, sondern einen ganzen Textabschnitt, z. B. einen Abstract (Abbildung 1). Darüber nachzudenken, worauf es ankommt, entfällt – eTblast denkt für Sie: Das Programm extrahiert aus dem Text eigenständig die relevanten Suchbegriffe und wichtet sie, durchsucht die circa 12 Millionen Einträge der Datenbank Medline und sortiert die Ergebnisse nach Ähnlichkeit. Der Text sollte 200–500 Worte lang sein und kann als Datei (im reinen Textformat) eingelesen oder in eine Suchmaske kopiert werden. Die erste, schnelle Suche ergibt nach 2–5 min Wartezeit eine lange Ergebnisliste; darauf aufbauend kann man eine iterative Suche durchführen, bei der man nun umfangreiche Optionen zur weiteren Gestaltung hat. So kann man die Bewertungskriterien und den erfassten

Dokumententyp auswählen, bestimmte Suchbegriffe festlegen, die unbedingt vorkommen müssen, eine von mehreren vorgegebenen Stoplisten – das sind Listen mit Worten, die nicht zur Suche benutzt werden – auswählen oder eine eigene einlesen und die Suche durch Einbeziehung medizinischer Synonyme ausdehnen. Fast wird es des Guten hier zu viel: Die Vielfalt der Optionen macht die Suche recht unübersichtlich und durch die zeitliche Verzögerung, mit der man die Ergebnisse einsehen kann – sie werden einem hier mit manchmal erheblicher Verzögerung als E-Mail zugestellt – wird eine Optimierung der Suche umständlich. Der Wert der medizinischen Synonyme war bei meiner Testsuche fragwürdig, da sich nicht beeinflussen ließ, welche Synonyme verwendet werden; es ging in meinem Text um „H3“ (Histon 3), das per Synonym auch als „medizinisches Objekt“ Fumagillin erschien, eine mir bis dato peinlicherweise völlig unbekannte Verbindung. So etwas dürfte wohl öfter passieren – hier sind noch Verbesserungen nötig, um diese Option wirklich sinnvoll einsetzen zu können. Nicht zuletzt sollte der Support dringend verbessert werden; mehrere Anfragen an die unter „Kommentare, Anregungen und Beschwerden“ angegebene Kontaktadresse blieben schlachtrichtig ohne Antwort.

eTblast ist einen Versuch wert, wenn man sich in ein neues Thema einarbeiten will; für die tägliche Routine wird es bis auf Weiteres bei der altgewohnten Schlagwortsuche bleiben. Und es ist die Mühe wert zu prüfen, ob es nicht für das konkrete Arbeitsfeld bereits andere, genau angepasste Textanalyse-Werkzeuge gibt wie Textpresso^[1] für den Modellorganismus *C. elegans*.

Christoph Weise

Freie Universität Berlin (Deutschland)

[1] <http://www.textpresso.org>

 Für mehr Informationen besuchen
Sie: <http://invention.swmed.edu/eblast/>
oder nehmen Sie Kontakt auf mit
etblast@innovation.swmed.edu

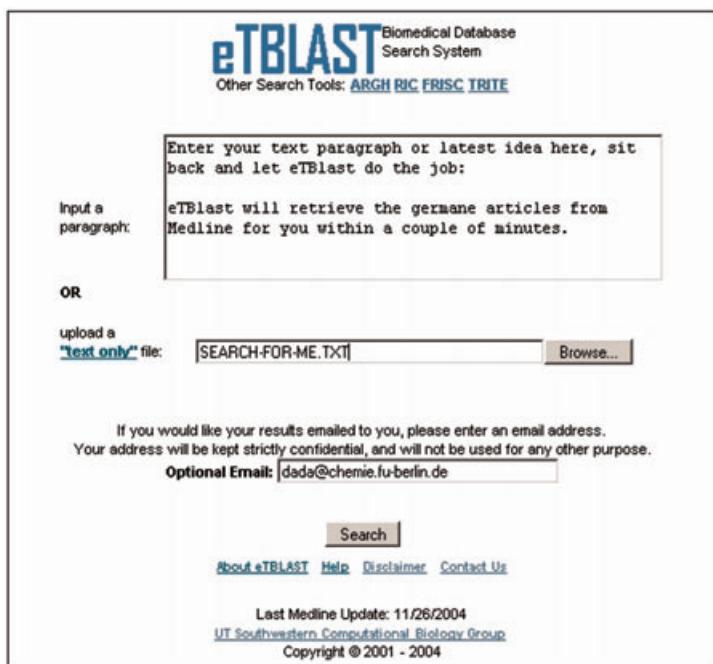


Abbildung 1. Suchmaske von eTblast.