

Het snel vinden van citaten in de Sai-Baba-literatuur

(ontsluiting van digitale documenten door middel van concordantie-software)

22 mei 2018

Inleiding

Bij het zoeken naar citaten van Swami, of naar de context daarvan, kan de grote omvang van de Sai-literatuur (Vahini serie, Summer Showers serie, Sathya Sai Speaks serie etc.) soms een belemmering zijn. Moderne software kan helpen om dit probleem op te lossen door te zoeken in digitale documenten. Zulke software is gratis verkrijgbaar via internet, evenals de (Engelstalige) Sai-Baba-literatuur. In het onderstaande wordt uitgelegd waar de concordantiesoftware en de literatuur zijn te vinden, hoe de software moet worden geïnstalleerd en hoe deze kan worden gebruikt. Daarbij is er een eenvoudige zoekmethode en een wat uitgebreidere methode. De laatste is iets complexer maar geeft de beste resultaten.

Sai-Baba-literatuur

Deze (Engelstalige) literatuur is te downloaden van de volgende websites:

<http://www.sathyasai.org/media/ebooks> (alleen Vahini serie) en

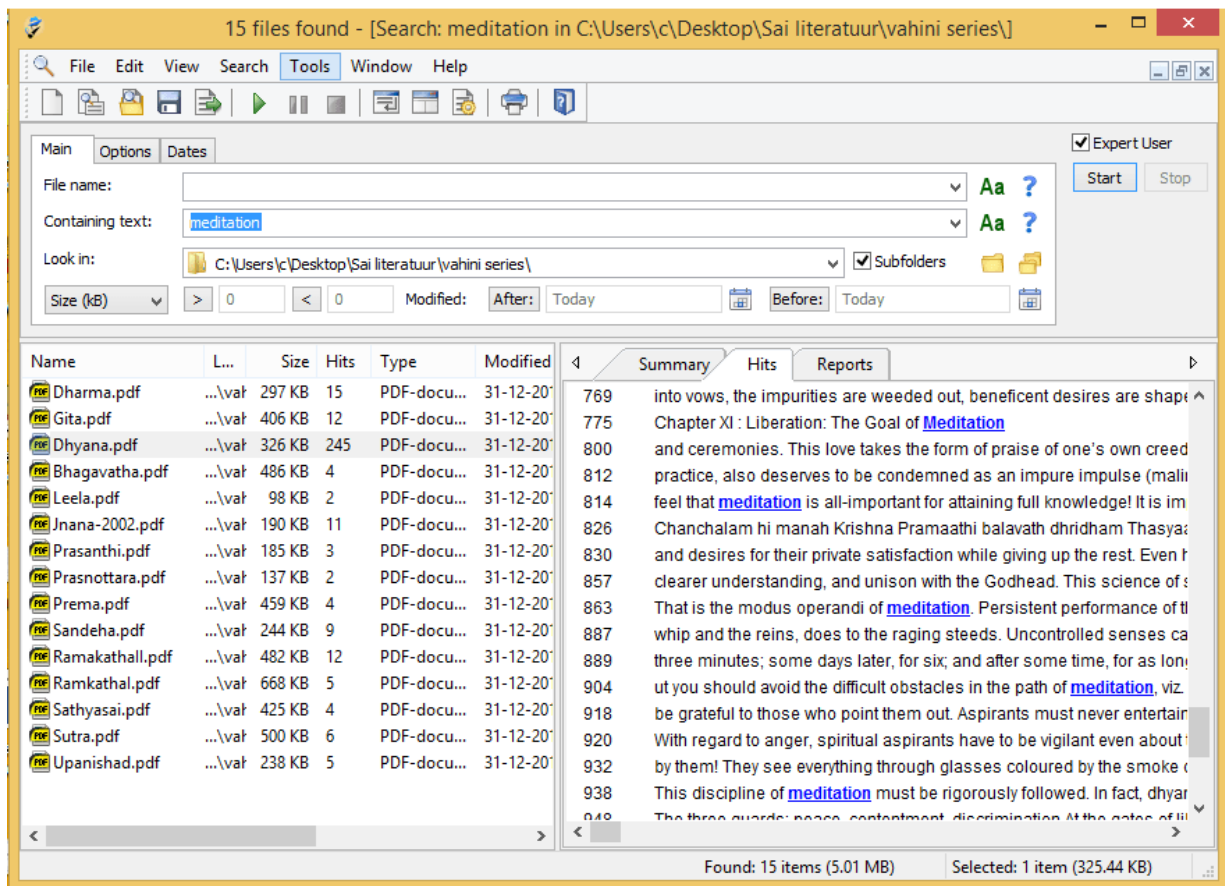
<http://www.ssbpt.info/ENGLISH/INDEX.HTML> (ook Summer Showers serie en de Sathya Sai Speaks reeks). Voor ondergenoemde toepassingen kan men het beste de pdf-versies downloaden.

Software, eenvoudige methoden

A. De tool 'Agent Ransack' geeft de mogelijkheid om in een document, in een reeks van documenten of in een hele directory (eventueel met subdirectories) naar een bepaalde term of een samenstelling van woorden te zoeken. Daarbij is ook intelligent zoeken mogelijk op woordcombinaties of woordpatronen (Boolean en regular expressions). Deze software is gratis te downloaden van : <https://www.mythicsoft.com/agentransack/> . De installatie wijst zich eigenlijk vanzelf. Volg eventueel de instructie die is aangegeven op de website.

AgentRansack is alleen verkrijgbaar voor Windowscomputers. Gebruikers van het besturingssysteem Linux of MacOS wordt aangeraden om Virtualbox (<https://www.virtualbox.org/>) te gebruiken in combinatie met Windows (voor Windows XP kun je het beste zoeken naar een oude versie van Agent Ransack van 2014 of eerder).

Agent Ransack geeft als resultaat per document het zoekwoord-in-context weer en het aantal hits, d.w.z. hoe vaak het woord in het document voorkomt. Er kan snel worden gezien in welke documenten het zoekwoord of de woordcombinatie staat, op welke regel(s) in het document het voorkomt en wat de nabije context is. Er kan in het linkerdeel van het scherm worden doorgedrukt naar het document zelf, waar de ruimere context kan worden gelezen.



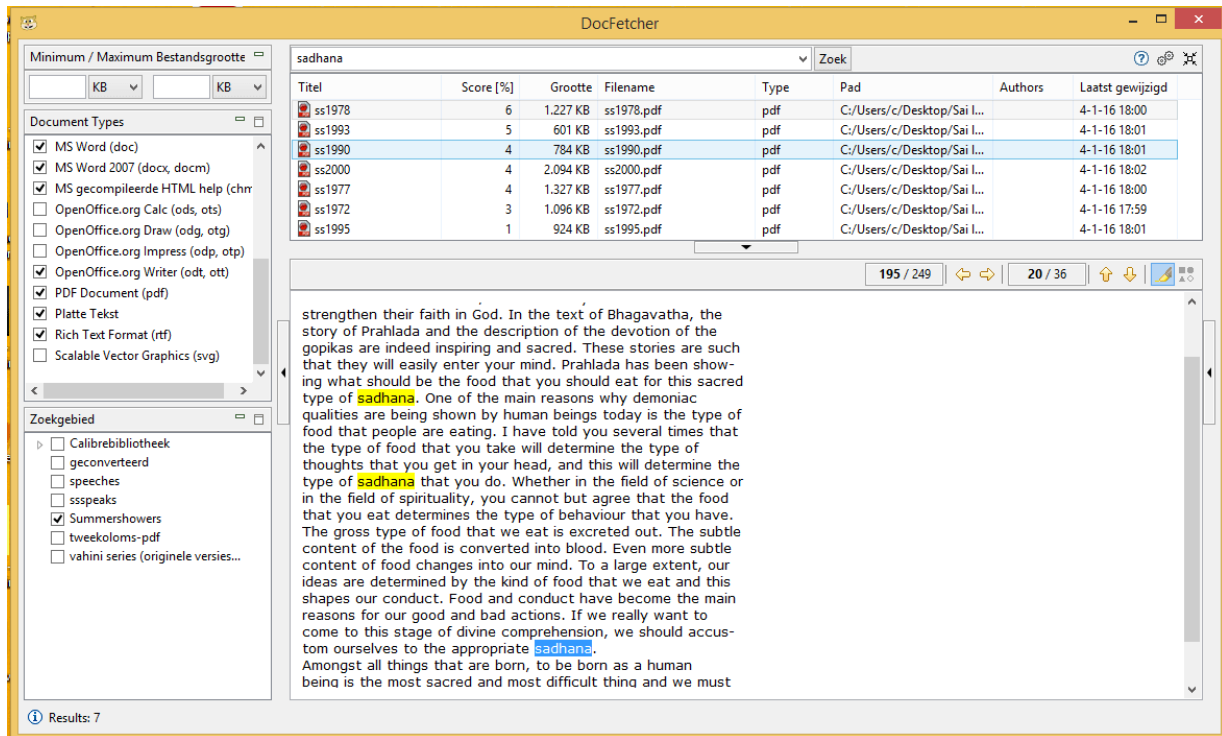
AgentRansack werkt met documenten in pdf-formaat, txt-formaat en Word-formaat.

B. De tool 'Docfetcher' geeft eveneens de mogelijkheid om in een document, in een reeks van documenten of in een hele directory (eventueel met subdirectories) naar een bepaalde term of een samenstelling van woorden te zoeken. Daarbij is ook intelligent zoeken mogelijk op woordcombinaties of woordpatronen (Boolean, wild cards, regular expressions). Deze software is gratis te downloaden van : <http://docfetcher.sourceforge.net/en/index.html> . Deze tool vereist de installatie van Java (zie de instructie op de genoemde website). De installatie is verder relatief eenvoudig.

Docfetcher werkt op Windows-, Linux- en MacOS-computers.

Docfetcher heeft het voordeel dat het met vele bestandsformaten overweg kan, waaronder pdf, txt, epub, html, doc en docx. Bij een zoekactie wordt als zoekresultaat een lijst met documenten weergegeven waarin de zoekterm (of woordcombinatie) veel voorkomt. Door op de gevonden documentnaam te klikken wordt de zoekterm in context gehighlight weergegeven. Met pijltjestoetsen kan men vervolgens in de tekst van het document bladeren (pijltjes links en rechts) of van zoekterm naar zoekterm springen (pijltjes naar boven en beneden). Gevonden tekst kan eenvoudig worden gekopieerd door het gewenste fragment met de muis te selecteren, te kopiëren en vervolgens elders te plakken.

Het zoekgebied van Docfetcher kan worden bepaald door in het deelscherm linksonder (zie bijgaande figuur) rechts te klikken en de map(pen) aan te geven waar de te doorzoeken documenten staan. De help-functie (symbooltje met vraagteken) geeft nadere uitleg over de zoekmogelijkheden.



Software, uitgebreidere methode

De tool 'AntConc' is een applicatie die voor Windows, MacOS en Linux beschikbaar is. Deze tool heeft een uitgebreidere zoekmogelijkheid, geeft de zoekresultaten grafisch weer en maakt het doorklikken naar de desbetreffende tekstlocatie mogelijk (net als Docfetcher, maar in tegenstelling tot AgentRansack). Deze concordantiesoftware heeft het nadeel dat alleen zoeken mogelijk is in documenten in platte-tekstformaat (.txt) maar heeft zoals gezegd het voordeel van grafische weergave van de locatie van het zoekresultaat. Met de extra-tool 'AntFileConverter' die is te vinden op dezelfde website kunnen documenten groepsgewijs worden omgezet van .pdf-formaat in .txt-formaat¹.

Wil men slechts één of een klein aantal pdf-documenten converteren dan kan dat simpelweg ook door de tekst de tekst van het pdf-document (rechtsclick, alles selecteren) te kopiëren en in een leeg

¹ Dit gaat prima met pdf-documenten die de tekst in één kolom hebben staan. Sommige documenten van de Sai-Baba-literatuur hebben echter een twee-koloms-opmaak. In dat geval wordt aangeraden de pdf-naar-txt-conversie te doen met de gratis PDF-reader 'NitroReader' die een uitstekende conversiemogelijkheid heeft : <https://www.gonitro.com/pdf-reader> (aanmelden is nodig). Ook is conversie met een OCR-tool mogelijk (zie bijv. <http://www.freeocr.net/>).

kladblok/notepad-document te plakken en vervolgens op te slaan. Alternatief is om het pdf-document te openen in een pdf-reader (zoals Nitroreader, zie bijgaande link) die van pdf- naar txt-formaat kan converteren. Voor groepsgewijze conversie van een groot aantal documenten is de AntFileConverter praktischer in het gebruik.

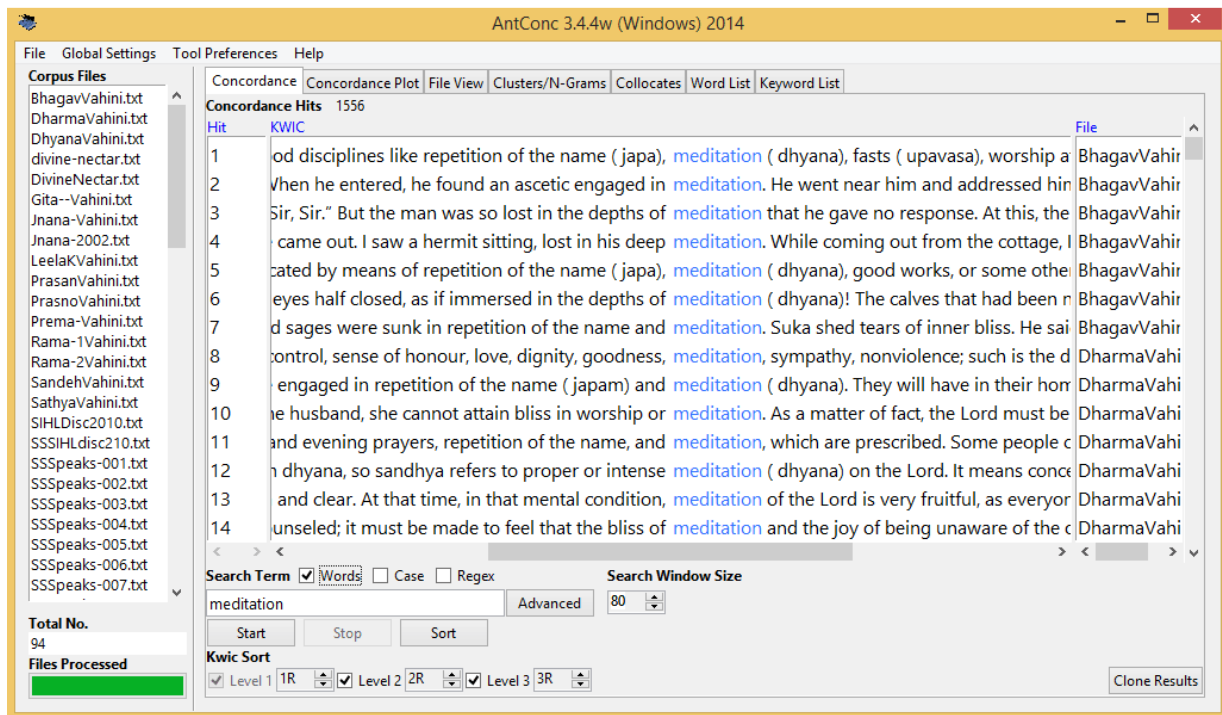
AntConc en AntFileConverter kunnen worden gedownload van de website :

<http://www.laurenceanthony.net/software.html>². De software hoeft niet te worden geïnstalleerd, hij is kant-en-klaar.

Ook met deze tool is intelligent zoeken mogelijk met woordcombinaties, woordpatronen en dergelijke.

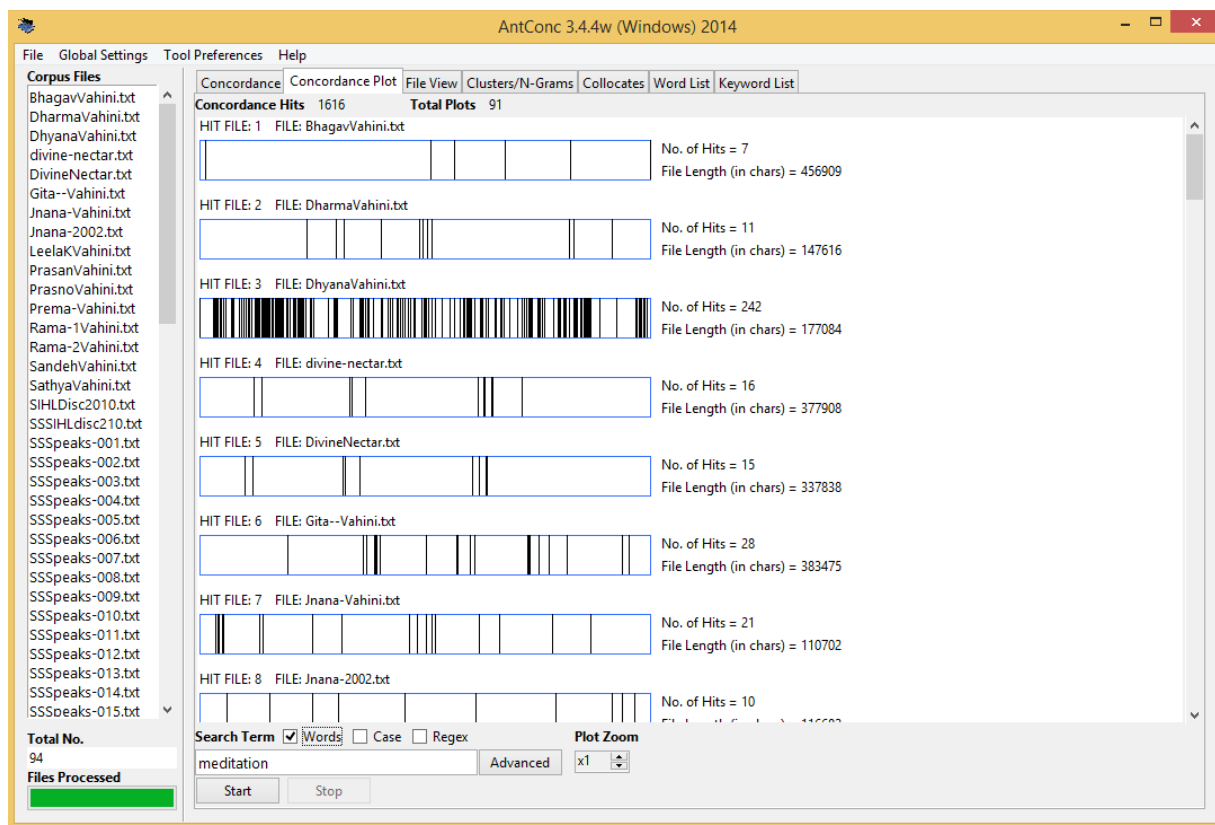
Zie onderstaand een voorbeeld waarbij is gezocht op het trefwoord 'meditation' in de complete Sai-Baba-literatuur.

Via de knop 'File' moet eerst worden aangegeven in welke (platte)tekstdocumenten moet worden gezocht. Men kan losse documenten aangeven (eventueel shift- of control-knop vasthouden bij het selecteren) of ook een hele directory. Na het ingeven van het trefwoord in het desbetreffende vak onderaan en het drukken op de knop 'start' wordt een lijst met 'hits' gegenereerd en weergegeven. Met 'Search Windows Size' kan worden ingesteld hoeveel tekens links en recht van het zoekwoord moeten worden weergegeven. Hoofdlettergevoelig zoeken is mogelijk door het zetten van een vinkje bij 'Case'.



² Noot: omdat AntConc en AntFileConverter kant-en-klare programma's zijn, kan het gebeuren dat je beveiligingsprogramma een waarschuwing geeft. Deze kan bij het gebruiken van de correcte link echter worden genegeerd.

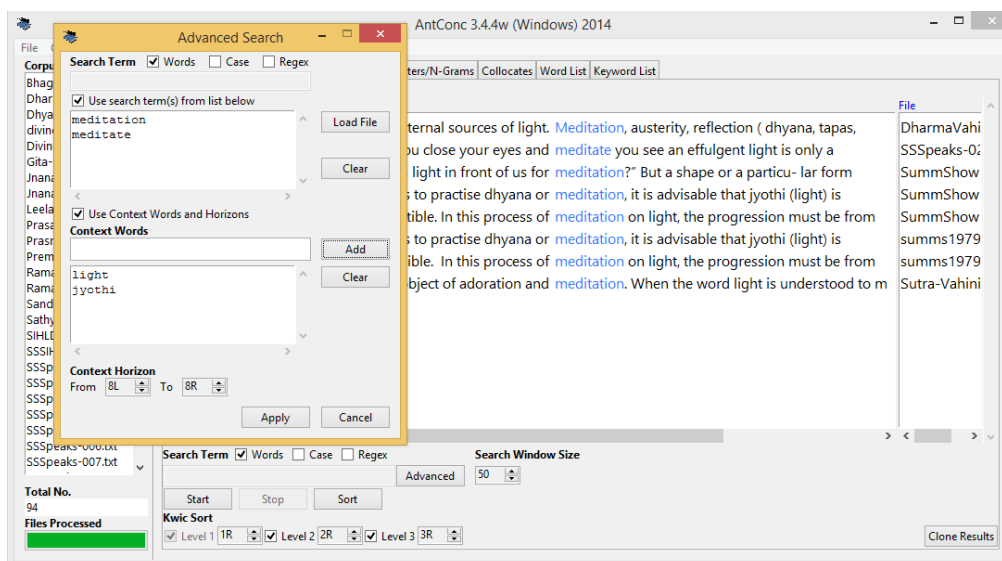
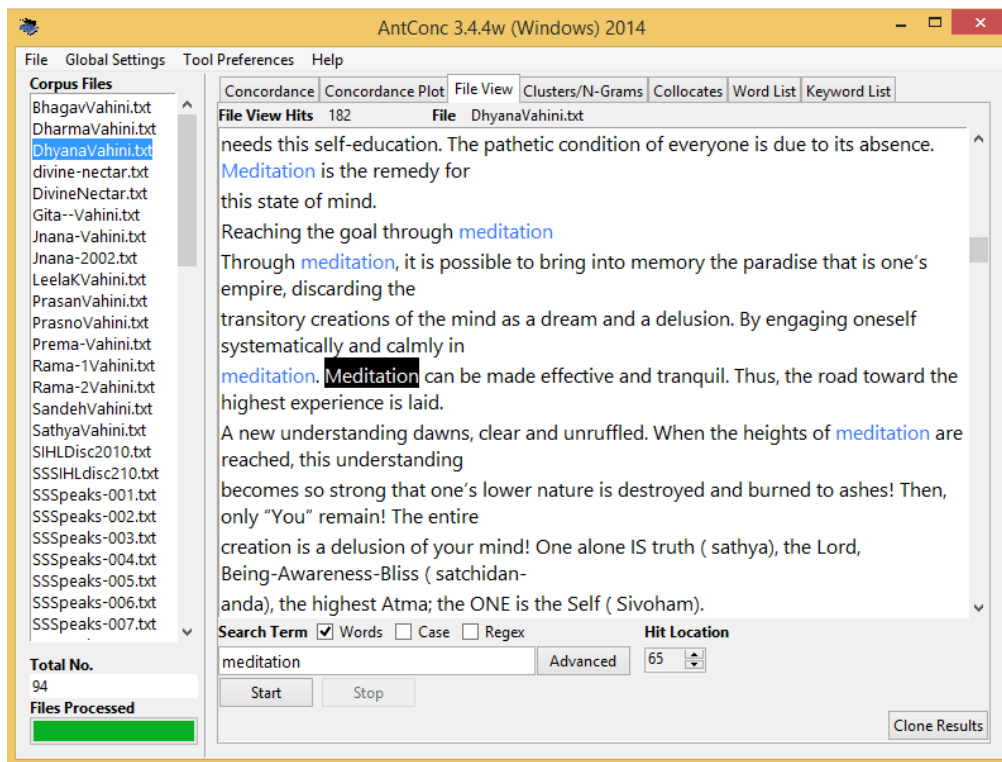
Bij aanklikken van de tab 'Concordance plot' wordt grafisch weergegeven waar het woord 'meditation' in de aangegeven documenten voorkomt. Het zal voor de kenner niet verrassend zijn dat 'Dhyana Vahini' de meeste hits geeft.



Zowel via het tabblad 'File View' als via het tabblad 'Concordance' kun je het trefwoord terugvinden in de brontekst door resp. te klikken in de streepjesgrafiek of door te klikken op het gekleurde woord. Waar in de grafiek streepjes dicht op elkaar staan kan men in het algemeen aannemen dat het desbetreffende onderwerp specifiekere wordt besproken door Swami.

Wanneer je de tekst niet alleen wilt nalezen, maar bijv. ook de desbetreffende gevonden passage wilt gebruiken voor een eigen document kun je die tekst kopiëren en plakken door middel van selecteren met de muis, kopiëren (Ctrl-C) en plakken (Ctrl-V) in het eigen document.

Voor de liefhebbers: Via de knop 'Advanced' kan ook op woordpatronen worden gezocht, bijvoorbeeld een situatie waarbij woorden binnen een bepaalde afstand van elkaar voorkomen. Zie hieronder de afbeelding van een situatie waarbij is gezocht naar de vermelding van de woorden 'meditation' of 'meditate' waarbij de woorden 'light' of 'jyothi' binnen een afstand van 8 woorden (links of rechts) van het trefwoord voorkomen.



Tot slot: Een alternatieve gratis tool, vergelijkbaar met AntConc, is te vinden op <https://www.adelaide.edu.au/carst/resources-tools/adtat/>. Toelichting: na downloaden uitpakken (rechtsklik) en als administrator installeren (rechtsklik). Deze tool heeft Java nodig, dat, na gegeven permissie, vanzelf wordt geïnstalleerd. Ook deze tool werkt uitsluitend met documenten in het .txt-formaat.