**Aanpak**

Om informatie te verzamelen over planten die in staat zijn om anthocyaan te produceren zullen er twee stappen worden uitgevoerd: ophalen van bekende genen die invloed hebben op anthocyaan productie en het doorzoeken van artikelen om zo te bepalen welke omstandigheden (bijv. stress) invloed hebben op de expressie van deze genen.

Coderende genen

Voor het zoeken naar coderende genen zal er gebruikt worden gemaakt van de NCBI gene database. Deze database zal worden doorzocht door te zoeken naar de anthocyaan in combinatie met een filter voor planten. Deze stap zal resulteren in een lijst van genen die invloed hebben op anthocyaan productie. Door het toepassen van de filter zullen alleen genen worden opgehaald die aanwezig zijn in planten (meer specifiek: “Embryophyta”). Aangezien elk gen is gekoppeld aan een organisme kunnen de organisme die anthocyaan produceren gelijktijdig worden opgehaald.

Deze aanpak heeft de voorkeur boven het doorzoeken van artikelen om informatie op te halen over de genen en de organisme. Ten eerste is het doorzoeken van de NCBI gene database enorm snel omdat NCBI elk gen al heeft gekoppeld aan bepaalde sleutelwoorden (anthocyanin in dit geval). Ten tweede zullend deze resultaten veel betrouwbaarder zijn dan bij het gebruik van een text-mining techniek[[1]](#footnote-1). Text-mining wordt ten eerste beperkt doordat NCBI alleen het ophalen van abstracts ondersteund, ofwel als er genen in de rest van het artikel worden genoemd kunnen deze nooit gevonden worden (tenzij deze ook worden gedownload wat de zoektijd enorm zou maken). Verder zal text-mining altijd informatie missen omdat een algoritme nooit perfect zal werken op alle artikelen (denk hierbij aan afwijkende zinsopbouw, alternatieve namen, namen die lijken op een gen naam maar in werkelijk geen gen naam zijn etc.).

Bepalen van omstandigheden

1. [↑](#footnote-ref-1)