|  |  |
| --- | --- |
| Projectgroep 11 | |
| Document | Analyse.doc | |
| Datum | 12-05-2017 | |
| Versie | 1.0 | |
| Groepsleden | Koen van der Heide | |
|  | Rick Beeloo | |
|  | Thomas Reinders | |
|  |  | |

CoTexThen

Automatische PubMed analyse

# Versiebeheer

|  |  |  |  |
| --- | --- | --- | --- |
| Versie | Datum | Auteur | Beschrijving |
| 0.1 | 07-05-2017 | Koen van der Heide | Toevoeging eerste eisen. |
| 0.2 | 11-05-2017 | Thomas Reinders | Aanleiding + Doel v. App + UC\_diagram + namen toegevoegd |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Contents

[Versiebeheer 2](#_Toc482367670)

[1. Introductie 4](#_Toc482367671)

[Aanleiding 4](#_Toc482367672)

[Doel van dit document 4](#_Toc482367673)

[Doel van de applicatie 4](#_Toc482367674)

[Doelgroep 4](#_Toc482367675)

[1. Requirements 5](#_Toc482367676)

[Doel en focus 5](#_Toc482367677)

[Functionele eisen 5](#_Toc482367678)

[Niet-functionele eisen 6](#_Toc482367679)

[2. Use cases 7](#_Toc482367680)

[Doel van use cases 7](#_Toc482367681)

[Actoren 7](#_Toc482367682)

[Use Case 01 - Navigeren 7](#_Toc482367683)

[Use Case 02 – Visualisatie selectie 8](#_Toc482367684)

[Use Case 03 – Detail weergaven 8](#_Toc482367685)

[Use Case 04 – Detail export 9](#_Toc482367686)

[Use Case 05 – Overlap selectie 9](#_Toc482367687)

[Use Case 06 – Data invoer 10](#_Toc482367688)

[3. Use case diagram 11](#_Toc482367689)

[Doel van de use case diagram 11](#_Toc482367690)

[Actoren 11](#_Toc482367691)

[Use case diagram 11](#_Toc482367692)

[4. Systeemarchitectuur 12](#_Toc482367693)

[Doel van de systeemarchitectuur 12](#_Toc482367694)

[Het doel van de systeemarchitectuur is de samenhang aan te geven tussen de applicatie en het opslag systeem dat deze applicatie gebruikt. Hierin hebben we tier 1 ( client tier), betrokken bij direct contant met de gebruiker en tier 2(server tier), betrokken bij het genereren en verwerken van data en tier 3(database tier) betrokken bij de opslagen van de data (genen, organismen, (stress) condities, PMID’s en links naar relevante artikelen). 12](#_Toc482367695)

[Tiers 12](#_Toc482367696)

[Client tier 12](#_Toc482367697)

[Logica tier 12](#_Toc482367698)

[Database tier 13](#_Toc482367699)

[Samenhang 13](#_Toc482367700)

[Overzicht van het geheel 13](#_Toc482367701)

[5. Begrippenlijst 14](#_Toc482367702)

[Bronvermelding 15](#_Toc482367703)

# Introductie

## Aanleiding

**Het handmatig analyseren van artikelen op specifieke termen kost veel tijd en gezien de hoeveelheid beschikbare data is dit proces haast onmogelijk. Anna Czerednik ondervond dit probleem ook waarna ze ons bioinformatici heeft benaderd om een oplossing hiervoor te vinden.**

## Doel van dit document

Dit document dient ter verduidelijking van de geplande applicatie en hoe het ontwerp van deze applicatie het biologische vraagstuk omzet naar een IT vraagstuk. In dit document zijn dus de eisen, use cases (met samenvattend diagram) en de systeemarchitectuur van de applicatie samengesteld. Deze informatie is bedoeld om enerzijds de gebruiker en opdrachtgever de geplande functionaliteit van de applicatie uit te leggen en anderzijds de programmeurs de gewenste functionaliteit te tonen.  
De analyse is geschreven naar de methodes uit het boek *Praktisch UML, Jos Warmer & Anneke Kleppe, Vijfde editie*.2

## Doel van de applicatie

**Zoals in de inleiding genoemd is, kost het handmatig analyseren van artikelen veel tijd en is de hoeveelheid beschikbare data enorm. Daarom bouwen we een applicatie die in staat is met behulp van textmining in combinatie met database searches te zoeken naar artikelen over anthocyanen om zo antwoord te leveren op de vragen van Anna. De resultaten zullen vervolgens worden opgeslagen in de database die de gebruiker (bioloog) kan raadplegen via een webpagina om een overzicht te krijgen. Deze applicatie moet biologen een duidelijk beeld geven over welke planten anthocyanen produceren, welke genen daarbij betrokken zijn en welke (stress) condities hier op van invloed zijn.**

## Doelgroep

Dit document is bedoeld voor de toekomstige gebruikers (biologen) en de opdrachtgever. Ook is dit document gericht aan de ontwikkelaars die deze applicatie zullen ontwikkelen.

# Requirements

## Doel en focus

Het doel van dit onderdeel is het vastleggen van de functionele eisen van de applicatie, waarin ingegaan wordt op wat de toekomstige applicatie moet kunnen. De focus van dit project ligt bij het opzoeken en overzichtelijk presenteren van stress omstandigheden die leiden tot anthocyanen productie per plant en per anthocyanen regulerend gen. De functionele eisen zijn hier de gewenste mogelijkheden zoals aangegeven door de opdrachtgever vertaald naar testbare applicatie functies, de niet-functionele eisen geven de technische vereisten onderliggend aan deze functies.

## Functionele eisen

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Use case koppeling | Beschrijving | Prioriteit | Moscow |
| F1 | UC\_01\_Navigeren | De gebruiker kan de applicatie via het internet benaderen. | Hoog | M |
| F2 |  | De gebruiker kan de applicatie laten bepalen welke planten anthocyanen produceren onder stressomstandigheden door artikelen te doorzoeken. | Hoog | M |
| F3 |  | De gebruiker kan de applicatie laten bepalen welke genen van deze planten anthocyanen productie reguleren onder stress. | Hoog | M |
| F4 |  | De gebruiker laat de applicatie zoeken naar de exacte stressomstandigheden per gen op die anthocyanen productie kunnen verhogen. | Hoog | M |
| F5 |  | De gevonden organismes, genen en stress omstandigheden worden in een database opgeslagen. | Hoog | M |
| F6 |  | De gevonden organismes, genen en stress omstandigheden worden voor de gebruiker getoond in een sunburst diagram. | Hoog | M |
| F7 |  | De gebruiker kan links naar artikelen die relevant zijn voor de opgeslagen genen en stress omstandigheden verkrijgen. | Hoog | M |
| F8 |  | De overlap tussen de organismes, de eerder gevonden genen toebehorende aan deze organismes en de stress omstandigheden die deze genen beïnvloeden worden ook gevisualiseerd in het sunburst diagram. | Gemiddeld | S |
| F9 |  | De gebruiker kan van een gewenst item in een diagram meer details (organisme, genen, condities en de link naar relevante artikelen) in een aparte tabel opvragen. | Gemiddeld | S |
| F10 |  | De gebruiker kan zelf kiezen welke overlap getoond wordt in het diagram, er kan bijvoorbeeld voor gekozen worden om per omstandigheid alle genen te tonen of juist om per gen alle omstandigheden die dit gen beïnvloeden te tonen. | Gemiddeld | S |
| F11 |  | De gebruiker kan voor een gegeven gen de applicatie bekende orthologen in andere organismes laten tonen. | Gemiddeld | S |
| F12 |  | De gebruiker kan de applicatie de locatie in de plant (wortels, bladeren, etc.) waarop de gevonden stress omstandigheden effect hebben qua anthocyanen productie laten tonen. | Laag | C |
| F13 |  | De gebruiker kan zelf een custom omstandigheid meegeven waarna de applicatie bijbehorende organismes en genen zoekt en toont. | Laag | C |
| F14 |  | De gebruiker kan kiezen tussen meerdere soorten diagrammen voor de visualisatie. | Laag | W |
| F15 |  | De gebruiker kan de applicatie lokaal uitvoeren. | Laag | W |

## Niet-functionele eisen

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Use case koppeling | Beschrijving | Prioriteit | Moscow |
| T1 |  | De applicatie kan draaien op zowel een Windows als Linux server. | Hoog | M |
| T2 |  | De applicatie doorzoekt PubMed artikelen met een big O van O(n). | Hoog | M |
| T3 |  | De database kan niet aangepast worden zonder dat de gebruiker hiervoor als beheerder ingelogd is. | Hoog | M |
| T4 |  | De applicatie is geschreven in Python. | Gemiddeld | S |
| T5 |  | De webpagina is geschreven in HTML, CSS en Javascript. | Gemiddeld | S |
| T6 |  | Op de eerste dag van de maand wordt er een back-up van de database gemaakt. | Laag | C |
| T7 |  | De webpagina heeft een rode achtergrond. | Laag | C |
| T8 |  | De applicatie bevat een GUI wanneer deze lokaal uitgevoerd wordt. | Laag | W |
| T9 |  | De applicatie ondersteunt HTTPS verbindingen. | Laag | W |
| T10 |  | De sessie wordt voor iedere gebruiker opgeslagen in een cookie. | Laag | W |

# Use cases

## Doel van use cases

Een use-case beschrijft de reeks interacties van de actor (gebruiker of beheerder) met de applicatie in vocabulaire leesbaar voor zowel de ontwikkelaar als de gebruiker. Deze use cases kunnen hierdoor uitstekend gebruikt worden als communicatiemiddel tussen de klant en de ontwikkelaars om zo de eisen van de klant te realiseren. Verder worden er ook variaties in de reeks van stappen weergegeven. Deze variaties kunnen alternatieve interacties zijn die uiteindelijk hetzelfde gewenste doel bereiken maar ook interacties die niet tot het gewenste resultaat leiden zoals exceptioneel gedrag. De use cases specificeren dus de handelingen vanuit de gebruiker en de gewenste reacties van de applicatie hierop.

## Actoren

**De gebruiker,** deze gebruikt de applicatie om een overzichtelijke weergave te krijgen van: de genen die invloed hebben op anthocyanen productie, de organismes waarin deze genen aanwezig zijn, de (stress) condities die anthocyanen productie beïnvloeden en relevante artikelen betreft deze onderwerpen. De gebruiker kan hiervoor kiezen uit verschillende diagrammen (graaf, tree en sunburst) en de resultaten eenvoudig exporteren.

**Beheerder**, deze gebruikt de applicatie om de data dat weergegeven wordt aan te passen. Zo kan de beheerder data in de database manipuleren.

## Use Case 01 - Navigeren

|  |  |
| --- | --- |
| Naam | UC\_01\_Navigeren |
| Versie | 1.0 |
| Auteur | Rick. |
| Laatste wijziging | 08-05-17 |
| Actoren | Gebruiker. |
| Korte beschrijving | Applicatie benaderen via internet. |
| Aannames | Server is online.  Gebruiker heeft toegang tot de server. |
| Beschrijving | **1a.** Gebruiker browsed naar de URL van de applicatie.  **1b.** Applicatie retourneert de webpagina. |
| Alternatieve route | 1a. Gebruiker geeft direct IP-adres in.  1b. ga verder bij **1b.** |
| Excepties | 1b E1 De browser van de gebruiker ondersteunt geen Javascript.  1b E2 De browser van de gebruiker reageert niet op request/response van applicatie. |
| Resultaat | Gebruiker ziet de webpagina. |

## Use Case 02 – Visualisatie selectie

|  |  |
| --- | --- |
| Naam | UC\_02\_visualisatie\_selectie |
| Versie | 1.0 |
| Auteur | Rick. |
| Laatste wijziging | 08-05-17 |
| Actoren | Gebruiker. |
| Korte beschrijving | Weergeven van het gewenste diagram. |
| Aannames | Server is online.  Gebruiker heeft toegang tot de server.  Applicatie beschikt over data om een diagram te maken. |
| Beschrijving | **1a.** Gebruiker klikt op “visualisatie” dropdown menu.  **1b.** Applicatie toont de visualisatie opties: “sunburst”, “graaf” en “tree”.  **2a**. Gebruiker selecteert gewenste visualisatie optie.  **2b.** Applicatie update het diagram naar de gewenste visualisatie optie. |
| Alternatieve route | - |
| Excepties | 2b E1 De applicatie kan de data niet omzetten in het gewenste diagram. |
| Resultaat | Gebruiker ziet het gewenste diagram. |

## Use Case 03 – Detail weergaven

|  |  |
| --- | --- |
| Naam | UC\_03\_detail\_weergeven |
| Versie | 1.0 |
| Auteur | Rick. |
| Laatste wijziging | 08-05-17 |
| Actoren | Gebruiker. |
| Korte beschrijving | Details opvragen uit een deel van diagram. |
| Aannames | Server is online.  Gebruiker heeft toegang tot server.  Applicatie geeft een diagram weer. |
| Beschrijving | **1a.** De gebruiker klikt op een item van het diagram.  **1b.** De applicatie laat een tabel zien met daarin de  conditie, het organisme, de genen van het  betreffende item en de links naar relevante artikelen. |
| Alternatieve route | - |
| Excepties | 1b. E1 De gebruiker klikt op een deel van het diagram dat geen data bevat. |
| Resultaat | Gebruiker ziet een tabel met informatie over geselecteerd item. |

## Use Case 04 – Detail export

|  |  |
| --- | --- |
| Naam | UC\_04\_detail\_export |
| Versie | 1.0 |
| Auteur | Rick |
| Laatste wijziging | 08-05-17 |
| Actoren | Gebruiker |
| Korte beschrijving | Details uit het diagram exporteren |
| Aannames | Server is online.  Gebruiker heeft toegang tot server.  Applicatie geeft een diagram weer.  Gebruiker heeft detail weergaven open staan. (zie UC\_03\_detail\_weergaven)  De browser staat pop-ups toe. |
| Beschrijving | **1a.** De gebruiker klikt op de “exporteer” knop onder de detail tabel.  **1b.** De applicatie opent een bestands kiezer.  **2a.** De gebruiker selecteert een gewenste opslag map.  **2a.1** De gebruiker geeft een bestandsnaam in en klikt op “opslaan.”  **2b.** De applicatie slaat de condities, organismes en genen op (als tekst) in de opgeven map. |
| Alternatieve route | 2a.A1 De gebruiker plakt een pad in de bestandskiezer.  2b. A1 ga verder bij **2a.1**. |
| Excepties | 2a. E1 De opgegeven bestandsnaam is al in gebruik.  2a. E2 De applicatie heeft geen schrijfrechten. |
| Resultaat | Gewenste data is geëxporteerd. |

## Use Case 05 – Overlap selectie

|  |  |
| --- | --- |
| Naam | UC\_05\_overlap\_selectie |
| Versie | 1.0 |
| Auteur | Rick |
| Laatste wijziging | 08-05-17 |
| Actoren | Gebruiker. |
| Korte beschrijving | Details uit diagram exporteren. |
| Aannames | Server is online.  Gebruiker heeft toegang tot server.  Applicatie geeft een diagram weer. |
| Beschrijving | **1a.** De gebruiker klikt op “verander getoonde overlap.”  **2a.** De applicatie toont drie dropdowns menu’s: condities, organismen en genen.  **2a.** De gebruiker selecteert de gewenste overlap volgorde.  **2b.** De applicatie updatet de lay-out van het diagram. |
| Alternatieve route | - |
| Excepties | 2b E1 De gewenste lay-out is hetzelfde als de huidige lay-  out. |
| Resultaat | De gebruiker ziet de gewenste lay-out in het diagram. |

## Use Case 06 – Data invoer

|  |  |
| --- | --- |
| Naam | UC\_06\_data\_invoer |
| Versie | 1.0 |
| Auteur | Rick |
| Laatste wijziging | 08-05-17 |
| Actoren | Beheerder. |
| Korte beschrijving | Data updaten in de database. |
| Aannames | Database draait op de server.  Beheerder heeft toegang tot de database.  Beheerder heeft de rechten om data in de database te wijzigen. |
| Beschrijving | **1a.** De beheerder logt in op de database.  **1b**. Database toont de data in de database.  **2a.** De beheerder selecteert een tabel in de database.  **2b.** De database laat de tabel zien.  **3a.** De beheerder manipuleert data in de tabel en commit dit.  **3b**. De database slaat de wijzigingen op. |
| Alternatieve route | - |
| Excepties | 3a E1 De database crasht. |
| Resultaat | Beheerder heeft data in database aangepast. |

# Use case diagram

## Doel van de use case diagram

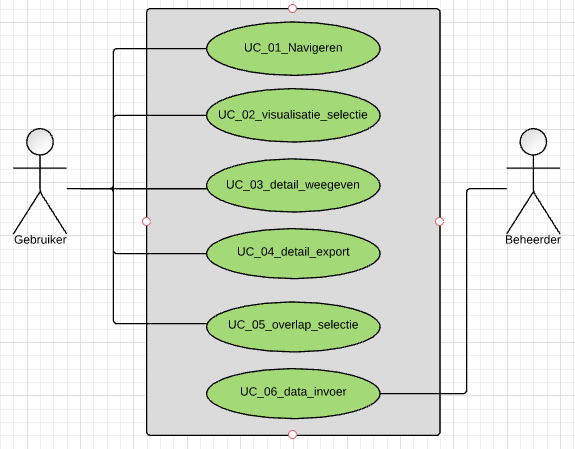
Het use case diagram geeft een grafisch overzicht van de mogelijke use cases die de actoren, in dit geval de gebruiker en de beheerder, kunnen uitvoeren om de opdrachtgever op een hoog niveau een idee te geven van welke functionaliteit per actor gepland is voor deze applicatie.

## Actoren

**De gebruiker,** deze gebruikt de applicatie om een overzichtelijke weergave te krijgen van: de genen die invloed hebben op anthocyanen productie, de organismes waarin deze genen aanwezig zijn, de (stress) condities die anthocyanen productie beïnvloeden en relevante artikelen betreft deze onderwerpen. De gebruiker kan hiervoor kiezen uit verschillende diagrammen (graaf, tree en sunburst) en de resultaten eenvoudig exporteren.

**Beheerder**, deze gebruikt de applicatie om de data dat weergegeven wordt aan te passen. Zo kan de beheerder data in de database manipuleren.

## Use case diagram



Figuur 1, **use case diagram.** In deze use case diagram zijn de verschillende acties gevisualiseerd die de gebruiker en beheerder kunnen uitvoeren. UC\_01 is het navigeren naar de webpagina door de gebruiker. UC\_02 is het veranderen van het soort diagram dat weergegeven moet worden. UC\_03 is het verkrijgen van alle informatie van een specifiek item uit het diagram. UC\_04 is het exporten van de informatie uit UC\_03. UC\_05 kan worden gebruikt voor het veranderen van de overlap volgorde bijvoorbeeld van: conditie, organisme en genen naar organisme, conditie en genen.. UC\_06 is voor de beheerder om de data in de database bij te werken.

# Systeemarchitectuur

## Doel van de systeemarchitectuur

## Het doel van de systeemarchitectuur is de samenhang aan te geven tussen de applicatie en het opslag systeem dat deze applicatie gebruikt. Hierin hebben we tier 1 ( client tier), betrokken bij direct contant met de gebruiker en tier 2(server tier), betrokken bij het genereren en verwerken van data en tier 3(database tier) betrokken bij de opslagen van de data (genen, organismen, (stress) condities, PMID’s en links naar relevante artikelen).

## Tiers

### Client tier

De client tier vorm de gebruikersinterface van het systeem, in dit geval de webpagina. Bij het initeel laden de webpagina is contact met de server vereist. Dit vanwege het feit dat de webpagina voor het genereren van het diagram de onderliggende structuur en data moet ophalen van de server. (zie “applicatie logica”). Als de website eenmaal geladen is kan de gebruiken op verschillende manieren interactie met de webpagina vertonen:

* Selectie van visualisatie mogelijkheden
* Overlap voorkeuren
* Detail weergaven
* Detail export

Al deze interacties zullen aan de client zijde worden afgehandeld. Dit heeft als grote voordeel dat de pagina niet herladen hoeft te worden, wat relatief veel tijd bespaart.

### Logica tier

Dit tier bevat twee applicaties:

* Visualisatie
* Applicatie logica

De “**applicatie logica**” is verantwoordelijk voor het ophalen van de volgende data:

* Genen die invloed hebben op anthocyanen productie.
* Organisme waarin deze genen aanwezig zijn.
* (stress) condities die anthocyanen productie beïnvloeden
* PMID’s en links naar de artikelen waar deze informatie uit afkomstig is

Deze gegevens zullen vervolgens worden opgeslagen in de database (zie “database tier”).

De “**visualisatie**” vormt de brug tussen de server en de client. Zo wordt gedurende het initieel laden van de webpagina de data uit de database opgehaald en omgezet in de onderliggende structuur voor een diagram (“visualisatie” figuur 2). Deze structuur wordt vervolgens doorgegeven aan de client.

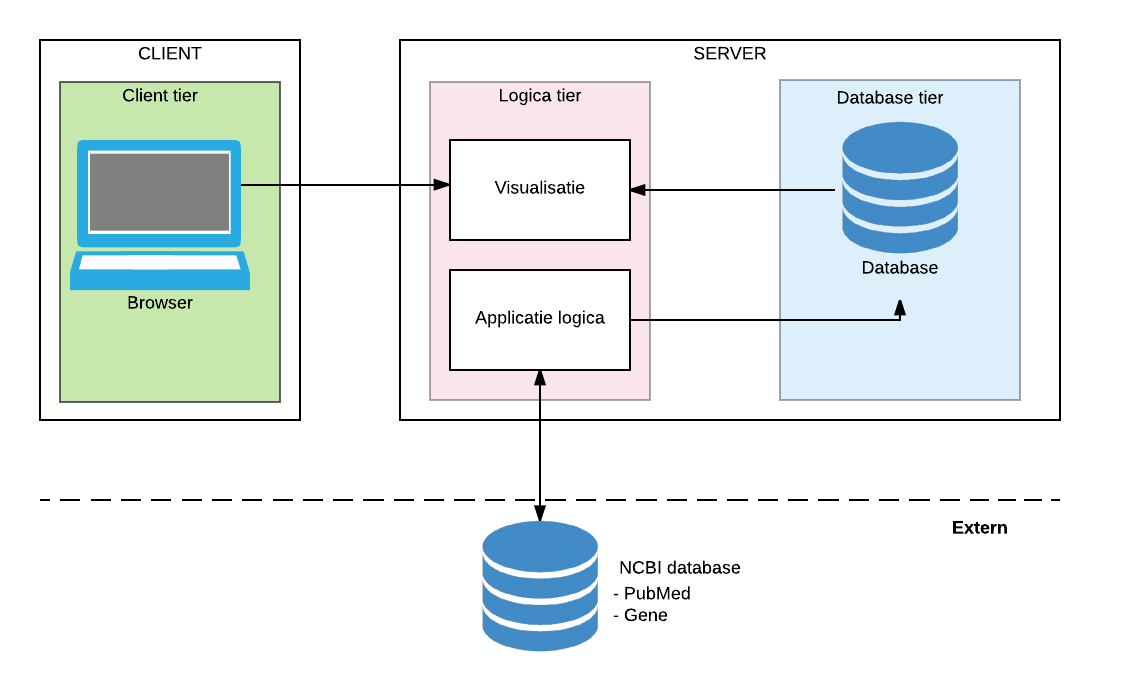
### Database tier

Dit tier is verantwoordelijk voor de opslag van de data afkomstig van de “applicatie logica” (zie ***figuur 2***).

### Samenhang

De “applicatie tier” tier is verantwoordelijk voor het genereren van de data en het opslaan hiervan in de database (onderdeel van de “datbase tier”). Verder levert de “applicatie tier” de basis van het diagram aan de “client tier” die op zijn beurt verantwoordelijk is voor het aanpassen van dit basis diagram naar de wensen van de gebruiker.

## Overzicht van het geheel



Figuur 2, **Systeemarchitectuur.**  Deze figuur laat de systeemarchitectuur zien van de applicatie. Er kunnen 3 tiers worden onderscheden: client, logica en database. De “applicatie tier” tier is verantwoordelijk voor het genereren van de data en het opslaan hiervan in de database (onderdeel van de “datbase tier”). Verder levert de “applicatie tier” de basis van het diagram aan de “client tier” die op zijn beurt verantwoordelijk is voor het aanpassen van dit basis diagram naar de wensen van de gebruiker.

# Begrippenlijst

|  |  |
| --- | --- |
| Begrip | Betekenis |
| Big O | Hiermee wordt de tijdscomplexiteit van een algoritme aangegeven. Een big O(1) betekent bijvoorbeeld dat er een constante hoeveelheid tijd nodig is onafhankelijk van de input. |
| Tier | Een tier verwijst naar de basis van een applicatie. Voorbeelden zijn client tier en server tier. |
| NCBI | National Center for Biotechnology Information, beheert onder andere een database die weer bestaat uit een verzameling van andere databases zoals Gene, Protein, Entrez, etc. |
| GUI | Graphical User Interface, de interface waarmee de gebruiker direct interactie vertoont door bijvoorbeeld het klikken op knoppen. |
| PubMed | Online database beheerd door NCBI met daarin biomedische literatuur. |
| CSS | Cascading Style Sheets, bieden opmaak mogelijkheden voor de webpagina(’s). |
| Model | Een versimpelde weergave van een concept. In dit geval worden voorbeeld zinnen gebruikt om een mathematisch model te bouwen. Dit model bevat de zin structuur en betekenis van de woorden. |
| Javascript | JavaScript is een scripttaal om webpagina's interactief te maken. In dit geval om het diagram aan de client zijde te kunne updaten. |
| HTML | HyperText Markup Language, opmaaktaal voor het wereldwijde web. |
| Graaf | Een diagram dat verbanden aangeeft. In dit geval kan een graaf de relatie aangeven tussen (stress) condities, genen en organismen. |
| Tree | Een weergave van termen in een boom structuur. In dit geval kan er dus worden gekozen om te beginnen met de condities waaraan d.m.v. “takken” de organismen zijn gekoppeld en daaraan weer de genen. |
| Sunburst | Kan worden vergeleken met de “Tree” structuur zoals hierboven beschreven, echter is deze circulair (wat interpretatie vergemakkelijkt). |
| Python | Een “hoog-niveau” programmeertaal uitermate geschikt voor tekstverwerking. Dit is de reden waarom Python de voorkeur heeft in deze applicatie. |
| HTTPS | Een beveiligd informatie overdracht protocol. |
| Cookie | Een bestandje dat door een website op je computer wordt geplaatst als je die site bezoekt. In dit geval zou het nuttig zijn om inloggegevens te onthouden. |
|  |  |

# Bronvermelding

1. *Praktisch UML, Jos Warmer & Anneke Kleppe, Vijfde editie*. (juni 2011)