ORF Voorspeller

Analyse

|  |  |
| --- | --- |
| Projectgroep 14 | |
| Document | Analyse.doc | |
| Datum | 01-01-01 | |
| Versie | 1.0 | |
| Groepsleden | Heleen Severin | |
|  | Jordi de Bruin | |
|  | Jeffrey Hiraki  Mark Verschuuren | |
|  |  | |

# Versiebeheer

|  |  |  |  |
| --- | --- | --- | --- |
| Versie | Datum | Auteur | Beschrijving |
| 0.1 | 09-02-17 | J. Hiraki | Functionele eisen toegevoegd |
| 0.1.1 | 09-02-17 | J. de Bruin | Review van functionele eisen |
| 0.2 | 09-02-17 | J. de Bruin | Het toevoegen van de introductie |
| 0.2.1 | 09-02-17 | H. Severin | Bijwerken van de begrippenlijst |
| 0.2.2 | 09-02-17 | J. Hiraki | Bepalen van de Use Case’s |
| 0.2.3 | 09-02-17 | H. Severin  J. Hiraki | Review van de introductie |
| 0.3 | 09-02-17 | J. Hiraki | Eerste versie van Use Case Diagram ingevoegd |
| 0.3.1 | 09-02-17 | H. Severin | Inleiding en begrippenlijst aangepast |
| 0.3.2 | 13-02-17 | J. Hiraki | Tweede versie van Use Case’s bepaald |
| 0.3.3 | 13-02-17 | M. Verschuuren | UC\_01\_openBestand gemaakt. |
| 0.3.3 | 13-02-17 | J. de Bruin | Eerste versie use case diagram gemaakt. |
| 0.3.4 | 14-02-17 | M. Verschuuren | Begin UC\_02\_RF\_bepaling gemaakt. |
| 0.3.5 | 14-02-17 | H. Severin | Use Cases: ORF visualisatie en BLASTen uitgewerkt |
| 0.3.6 | 15-02-17 | H. Severin | Requirements aanvullingen en aanpassingen |
| 0.4.0 | 15-02-17 | M. Verschuuren | UC\_02\_RF\_bepaling afronding en toevoeging UC\_05\_RF |
| 0.4.1 | 17-02-17 | J. de Bruin | Tweede versie use case diagram gemaakt na feedback Martijn en groepje. |
| 0.4.1 | 17-02-17 | H. Severin | Aanpassingen UC\_03\_ORF\_bepaling en UC\_BLASTen |
| 0.4.1 | 17-02-17 | M. Verschuuren | Aanpassen tekst van systeemarchitectuur |
| 0.4.1 | 17-02-17 | J. Hiraki | Aanpassen van systeemarchitectuur afbeelding. |
| 0.5.0 | 17-02-17 | M. Verschuuren | Samenvoegen van alle gemaakte bestanden. |

Contents

[Versiebeheer 2](#_Toc473787993)

[1. Introductie 4](#_Toc473787994)

[Aanleiding 4](#_Toc473787995)

[Doel van dit document 4](#_Toc473787996)

[Doel van de applicatie 4](#_Toc473787997)

[Doelgroep 4](#_Toc473787998)

[2. Requirements 5](#_Toc473787999)

[Doel en focus 5](#_Toc473788000)

[Functionele eisen 5](#_Toc473788001)

[Niet-functionele eisen 5](#_Toc473788002)

[3. Use cases 6](#_Toc473788003)

[Doel van use cases 6](#_Toc473788004)

[Actoren 6](#_Toc473788005)

[Use Case 01 - Titel 6](#_Toc473788006)

[4. Use case diagram 7](#_Toc473788007)

[Doel van de use case diagram 7](#_Toc473788008)

[Actoren 7](#_Toc473788009)

[Use case diagram 7](#_Toc473788010)

[5. Architectuur 8](#_Toc473788011)

[Doel van de architectuur 8](#_Toc473788012)

[Modules 8](#_Toc473788013)

[<<Module één >> 8](#_Toc473788014)

[<<Module twee>> 8](#_Toc473788015)

[Overzicht van modules 8](#_Toc473788016)

[6. Begrippenlijst 9](#_Toc473788017)

[Bronvermelding 10](#_Toc473788018)

# Introductie

## Aanleiding

DNA sequenties bestaan uit nucleotiden, deze zijn verdeeld in codons van drie nucleotiden per stuk. Welke nucleotiden bij welke codons horen is afhankelijk van het reading frame. Dit is dan weer afhankelijk van waar een frame begint en eindigt. Wanneer je één nucleotide opschuift, heeft dit invloed op de samenstelling van alle codons, het reading frame verandert. In totaal zijn er zes mogelijke reading frames. Wanneer een stuk van het frame een startcodon en een stopcodon is het een open reading frame (ORF). Deze ORFs kunnen in kaart worden gebracht om er vervolgens analyses (zoals BLAST) mee te kunnen doen.

## Doel van dit document

Dit document is geschreven om de functionaliteit van de applicatie vast te leggen. Er wordt beschreven wat het systeem wel en niet moet kunnen. Het document wordt gemaakt op basis van de wensen van de klant. Dit document wordt vervolgens gebruikt om een ontwerp te maken van het toekomstige systeem.

## Doel van de applicatie

Ten eerste opent de gebruiker een DNA sequentie in Fasta formaat waarna alle zes de reading frames worden geretourneerd en gevisualiseerd. Voor de zes reading frames van het DNA wordt de aminozuursequentie bepaald. Door analyse van de sequentie op start- en stopcodons worden ORF’s bepaald. Deze elementen worden gevisualiseerd in een scrollbaar paneel. De gebruiker heeft de mogelijkheid één ORF te kiezen om deze direct te BLASTen. Deze resultaten moet kunnen worden opgeslagen in de database en weer kunnen worden opgehaald om deze te visualiseren in de GUI. De BLAST resultaten moeten daarnaast kunnen worden opgeslagen in een .txt bestand. Het uiteindelijke doel is dus om van een DNA sequentie verschillende reading frames met ORF’s te visualiseren en op deze ORF’s een BLAST uit te voeren.

## Doelgroep

De doelgroep van deze applicatie zijn biologen die open reading frames willen voorspellen voor een bepaalde DNA sequentie en deze ORF’s snel willen analyseren door middel van BLAST.

# Requirements

## Doel en focus

Beschrijf kort het doel van dit hoofdstuk en waar de focus in dit project op ligt.

## Functionele eisen

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Use case koppeling | Beschrijving | Prioriteit | Moscow |
| F1 | UC\_01 | De gebruiker kan een DNA sequentie in Fasta formaat openen in een GUI. | Hoog | M |
| F2 | UC\_05 | De gebruiker kan de DNA-sequentie op slaan in een database. | Hoog | M |
| F3 | UC\_03 | De gebruiker kan een geselecteerde ORF opslaan in de database. | Hoog | M |
| F4 | NA | De GUI kan geopend worden en gesloten worden aan de hand van een toolbar. | Hoog | M |
| F5 | NA | De gebruiker ziet een visualisatie van de template- en complementaire DNA streng na het openen van een Fasta bestand. | Gemiddeld | S |
| F6 | UC\_02 | De gebruiker kan een visualisatie zien van 6 Reading Frames na het openen van een Fasta bestand. | Gemiddeld | S |
| F7 | UC\_04 | De gebruiker kan 1 of meerdere ORF’s selecteren voor het uitvoeren van een BLAST search. | Gemiddeld | S |
| F8 | UC\_05 | De gebruiker kan de resultaten uit BLAST opslaan in de database. | Gemiddeld | S |
| F9 | NA | De gebruiker ziet een visualisatie van de aminozuursequenties van 6 reading frames na het openen van het Fasta bestand. | Laag | C |
| F10 | UC\_05 | De gebruiker kan een ORF ophalen uit de database om deze opnieuw te visualiseren in de GUI. | Laag | C |

<<Indien nodig toelichten>>

## 

## Niet-functionele eisen

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Use case koppeling | Beschrijving | Prioriteit | Moscow |
| T1 | UC\_01 | De applicatie moet te gebruiken zijn op Windows en Unix. | Hoog | M |
| T2 | NA | De applicatie maakt gebruik van een externe database | Hoog | M |
| T3 | UC\_03 | De gebruiker kan het geopende bestand analyseren waarna er alle 6 mogelijke ORF’s berekend worden. | Hoog | M |
| T4 | UC\_04 | Het BLASTen mag maximaal 10 minuten duren | Gemiddeld | S |
| T4 | NA | De applicatie mag maximaal 200 mb in beslag nemen. | Gemiddeld | S |
| T5 | NA | Geheugengebruik mag maximaal 3gb ram in gebruik nemen | Gemiddeld | S |
| T6 | UC\_04 | Het BLASTen mag maximaal 10 minuten duren | Gemiddeld | S |
| T7 | NA | De applicatie werkt ook zonder internet, op het opslaan in de database en het BLASTen na. | Laag | C |

<<Indien nodig toelichten>>

# Use cases

## Doel van use cases

Use cases worden gebruikt om het systeem te beschrijven, hierin worden de actoren opgenomen. Ook de interacties worden hierin aangetoond, dus hoe een systeem reageert op een actie van de actor.

## Actoren

De actoren die betrokken zijn binnen de Use Cases zijn de gebruiker en de systeembeheerder.

## Use Case 01 – openBestand

|  |  |
| --- | --- |
| Naam | UC\_01\_openBestand |
| Versie | 1 |
| Auteur | Mark Verschuuren |
| Laatste wijziging | 13-2-2017 |
| Actoren | De gebruiker |
| Korte beschrijving | Het ophalen van de path van het door de gebruiker gekozen bestand, dit bestand moet een Fasta bestand zijn. |
| Aannames | Er wordt een path van het gekozen bestand weergegeven in een vakje.  Het bestand is een Fasta bestand. |
| Beschrijving | 1a. Gebruiker klikt op de "blader" knop.  1b. De applicatie start een interface waarbij de gebruiker kan zoeken naar een gewenst bestand.    2a. De gebruiker kiest een Fasta bestand die hij wilt openen.  2b. De applicatie zet de pathway van het Fasta bestand in een bijbehorend vakje.  3a. De gebruiker klikt op de “open” knop.  3b. Het bestand wordt geopend en gelezen |
| Alternatieve route | 2a. De gebruiker typt zelf een pathway in het vakje en drukt op “open”.  2b. Dezelfde procedure van 3 zal worden afgespeeld. |
| Excepties | 3a.   * Het door de gebruiker gekozen bestand is corrupt en zorgt voor een IOError. * Het door de gebruiker gekozen bestand  kan niet gevonden worden. * Het door de gebruiker gekozen bestand is leeg. * Het door de gebruiker gekozen bestand bevat geen DNA.   3b.   * De applicatie zal een IOError exceptie oproepen die een pop up geeft met : "IOError, kijk na of het bestand niet fouten heeft". * De applicatie zal een FileNotFound exceptie oproepen die een pop up geeft met : "FileNotFound error, de bestandsnaam klopt niet.". * De applicatie geeft een error aan met: “Dit bestand is leeg, gebruik een ander bestand!” * De applicatie geeft een error aan met : “Het bestand bevat geen DNA, kies een DNA bestand. |
| Resultaat | De applicatie heeft nu van het Fasta bestand, gekozen door de gebruiker, weergegeven in het bijbehorende vakje. Ook is het bestand al gelezen. |

## Use Case 02 – RF visualisatie

|  |  |
| --- | --- |
| Naam | UC\_02\_RF bepaling |
| Versie | 1 |
| Auteur | Mark Verschuuren |
| Laatste wijziging | 15-2-2017 |
| Actoren | De gebruiker |
| Korte beschrijving | Het visualiseren van de RF’s van de template en complementaire string op de bijbehorende panel. |
| Aannames | De readingframes komen worden in de panel gevisualiseerd, niet erbuiten. |
| Beschrijving | 1a. De gebruiker klikt op de knop “visualiseer”.  1b.   * De 3 RF’s van de template string worden boven elkaar gevisualiseerd boven de template. * De 3 RF’s van de complementaire string worden boven elkaar gevisualiseerd onder de complementaire string. |
| Alternatieve route | N.V.T |
| Excepties | 1a. De genomen index is groter dan lengte de readingframe.  2b. De applicatie zal een IndexOutOfRangeError die een pop-up geeft met : “De genomen index is te groot, er is iets mis met de readingframe!” |
| Resultaat | De readingframes van de template string staan boven elkaar gevisualiseerd boven de template string en de readingframes van de complementaire string staan boven elkaar onder de complementaire string. |

## Use Case 03 – ORF bepaling

|  |  |
| --- | --- |
| Naam | UC\_03\_ORF\_bepaling |
| Versie | 1 |
| Auteur | Heleen Severin |
| Laatste wijziging | 14-02-2017 |
| Actoren | De gebruiker |
| Korte beschrijving | Een bepaling van de ORF’s door analyse van een ingeladen sequentie op start- en stopcodons in 6 reading frames (UC\_02\_RF\_bepaling). |
| Aannames | Het ingeladen Fasta bestand dat een DNA sequentie bevat. |
| Beschrijving | 1a. De gebruiker laadt een Fasta bestand dat ORF’s bevat.  2a. De gebruiker klikt op de knop “Open”.  1b. De applicatie laat rechthoekjes zien op de berekende plek van de ORF boven of onder de lijn die de DNA sequentie representeert. |
| Alternatieve route | 1a. De gebruiker laadt een bestand van Fasta formaat in dat geen ORF’s bevat  2a. De gebruiker klikt op de knop “Open”.  2b. De applicatie geeft in een melding in een popup: “Geen ORF’s gevonden”. |
| Excepties | N.V.T |
| Resultaat | De applicatie laat rechthoekjes zien op de berekende plek van de ORF boven of onder de lijn die de DNA sequentie representeert. |

## Use Case 04 – BLASTen

|  |  |
| --- | --- |
| Naam | UC\_04\_BLASTen |
| Versie | 1 |
| Auteur | Heleen Severin |
| Laatste wijziging | 14-02-2017 |
| Actoren | De gebruiker |
| Korte beschrijving | De gebruiker kan een ORF (bepaald door het programma: UC\_01\_ORF\_bepaling) selecteren en vervolgens BLASTen. De resultaten worden in de GUI getoond. |
| Aannames | De gebruiker heeft verbinding met internet |
| Beschrijving | 1a. De gebruiker selecteert een ORF.  2a. De gebruiker klikt op de knop “BLAST ORF”.  2b. De applicatie retourneert na enkele minuten de BLAST resultaten in de GUI. |
| Alternatieve route | 1a. De gebruiker selecteert een ORF.  2a. De gebruiker klikt op de knop “BLAST ORF”.  2b. De applicatie retourneert de melding: “geen significanten overeenkomsten gevonden!”. |
| Excepties | 1a. De gebruiker selecteert een ORF, de sequentie die hier aan gekoppeld is heeft een onjuist formaat voor BLAST.  2a. De gebruiker klikt op de knop “BLAST ORF”.  2b. De applicatie geeft een melding in een pop up: “Het formaat van de sequentie is verkeerd!”. |
| Resultaat | De applicatie retourneert na enkele minuten de BLAST resultaten in de GUI. |

## 

## Use Case 05 - Ophalen uit Database

|  |  |
| --- | --- |
| Naam | UC\_05\_Ophalen uit database |
| Versie | 1 |
| Auteur | Mark Verschuuren |
| Laatste wijziging | 15-02-2017 |
| Actoren | De gebruiker |
| Korte beschrijving | Het ophalen van al ORF’s uit de database die al zijn opgeslagen in vorige visualisaties. |
| Aannames | Er is een connectie met de database.  De database is bereikbaar. |
| Beschrijving | 1a. De gebruiker klikt op de knop “Haal ORF’s op”.  1b. Een pop up zal komen met de vraag welke ORF’s hij zou willen ophalen. Deze ORF’s zijn gekoppeld aan de een ID van de naam van het Fasta bestand.  2a. De gebruiker klikt op de gewenst ID.  2b. De ORF’s van de ID zullen worden gevisualiseerd in een blauwe kleur op dezelfde positie als zij eerst stonden. |
| Alternatieve route | N.V.T |
| Excepties | 1a. De database heeft geen opgeslagen ORF’s of ID’s.  1b. De applicatie geeft een melding dat er geen ORF’s opgeslagen zijn.  2a. Er zijn al ORF’s in de applicatie gevisualiseerd.  2b. De gebruiker krijgt een melding met : “Er zijn al ORF’s gevisualiseerd, start de applicatie opnieuw op om uw gekozen ORF’s te visualiseren. |
| Resultaat | Door de gebruiker gekozen ORF’s worden gevisualiseerd in de bijbehorende panel. |

# Use case diagram

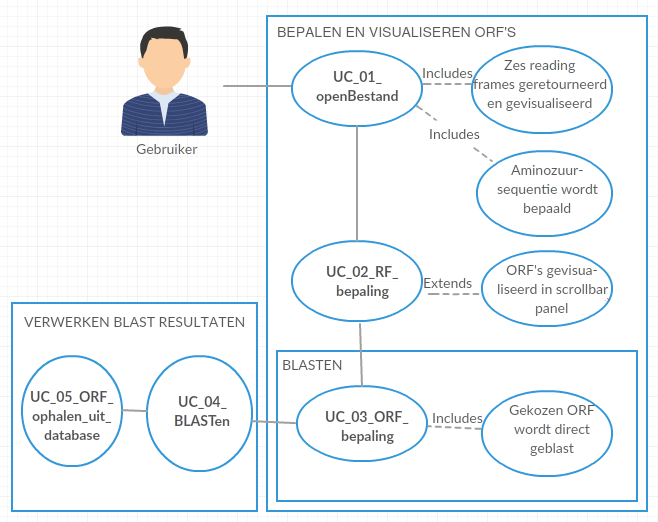
## Doel

Het doel van de een use case diagram is om te laten zien hoe de gebruiker interacties kan hebben met verschillende use cases en hoe deze use cases interacties hebben met elkaar.

## Actoren

Gebruiker

## Use case diagram



# Systeemarchitectuur

## Doel van de systeemarchitectuur

Om een goede samenloop te hebben binnen een programma en alles goed gescheiden te houden, is er een goede systeemarchitectuur nodig. Hierin staan de verschillende tiers in opgeschreven en ook de manieren hoe de tiers met elkaar communiceren.

## Tiers

### Application tier

Een belangrijk visueel onderdeel, dit is waarmee de eindgebruiker van het programma direct mee in aanraking komt. De application Tier bevat de visualisatie van de applicatie en bevat de logica van de applicatie. Zoals het openen van bestanden of het visualiseren van ORF’s. De berekeningen worden gemaakt in de achtergrond van de User tier, hier heeft de gebruiker zelf geen invloed op.

### Data tier

Data die wordt gecreëerd in de application tier kan worden opgeslagen in de data tier. In de data tier wordt vooral informatie opgeslagen over ORF’s en hun bijbehorende ID, ook BLAST data wordt hier opgeslagen. Alles wat binnen deze tier valt, heeft dus te maken met het opslaan en het bewaren van data. De data tier kan alleen worde benaderd via de application tier en heeft geen directe verbinding met BLAST.

### BLAST

Externe tier binnen de architectuur, de applicatie is in staat om de BLAST server te vragen om gegevens en te gebruiken om data te analyseren. Dit gehele proces gebeurt in de server van de BLAST en niet in de application tier.



Afbeelding 1. In de boven afgebeelde afbeelding zijn de verschillende tiers te zien. Ook is er een gescheiden segment te zien waarin BLAST staat. De Application Tier communiceert met de Data Tier. De Data Tier is vooral bedoeld voor het opslaan van informatie die de gebruiker verkrijgt. Binnen de Application Tier is er de mogelijkheid om te BLASTen, dit gebeurd niet binnen de Application Tier, maar bij het BLAST-data centrum zelf.

# Begrippenlijst

|  |  |
| --- | --- |
| Begrip | Betekenis |
| ORF | Open reading frame: potentieel gen |
| Startcodon | Codon dat codeert voor een startsignaal van translatie |
| Stopcodon | Codon dat codeert voor een stopsignaal van translatie |
| Fasta formaat | Tekst bestand voor nucleotide of proteïne sequenties |
| Database | Collectie van data in tabellen die met relaties aan elkaar gelinkt zijn |
| Codons | 3 letter sequentie van DNA of RNA dat codeert voor 1 aminzozuur |
| BLAST | Basic Local Alignment Tool: algoritme voor het matchen van een DNA of proteïne sequentie met een bestaande sequentie van een database |
| .txt | Tekst based bestand |
| Path | Bestandspad: naam van de bestandslocatie |
| Toolbar | Balk met knoppen voor bijvoorbeeld het afsluiten van een applicatie |
| GUI | Graphical User Interface: grafische en interactieve weergave van een applicatie voor de eindgebruiker. |
| Gebruiker | Eindgebruiker van de applicatie |
| Actor | Gebruiker van de applicatie: ontwikkelaar, tester, eindgebruiker |
| Applicatie | Het programma (de software) |
| Template sequentie | Sequentie waarvan het gen wordt getranscribeerd |
| Complementaire sequentie | Sequentie gebonden aan de template sequentie in de dubbele streng configuratie |
| Gb, Mb, kb | Maat voor grootte van bestand in veelvoud van bytes |

# Bronvermelding

J.Warmer (2011) Praktisch UML 3de editie. Pearson Benelux bv.