

Casus

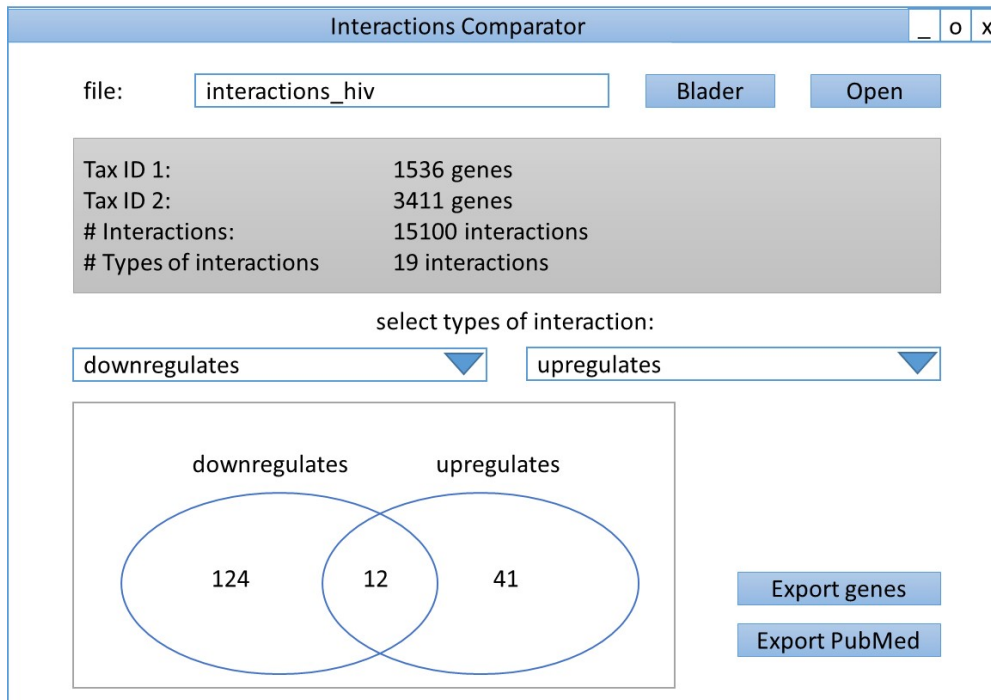
Onderzoekers hebben de interacties tussen de producten van genen van mensen en het HIV virus in kaart gebracht. Er is een bestand gegenereerd waar alle interacties in staan beschreven, zie tabel 1. Er zijn reeds ruim vijftienduizend vormen van interactie bekend, maar in de toekomst komen daar wellicht meer bekende interacties bij. Zie bestand: ftp://ftp.ncbi.nlm.nih.gov/gene/GeneRIF/hiv_interactions.gz

Tabel 1: *De regels van boven naar beneden in deze tabel staan gelijk aan de kolommen van links naar rechts in de interactions-file. Voor een volledige beschrijving, zie:*

<ftp://ftp.ncbi.nlm.nih.gov/gene/README>

Tax ID 1
Gene ID 1
Product accession.version 1
Product name 1
Interaction short phrase
Tax ID 2
Gene ID 2
Product accession.version 2
Product name 2
PubMed ID (PMID) list
Last update timestamp
GeneRIF text

Het is de bedoeling om de genen van mensen inzichtelijk te maken voor biologen. We zijn met name opzoek naar die genen waarvan de producten meer dan één type relatie hebben met de producten van genen uit het HIV virus. Bij een vergelijking op basis van typen interactie kunnen we genen van mensen selecteren die betrokken zijn bij meerdere vormen van interactie met producten van HIV genen.



Figuur 1: Voorbeeldscherf van de te ontwikkelen applicatie. De uiteindelijke applicatie dient overeenkomstige functionaliteit te bezitten, maar mag anders opgebouwd zijn. De gegeven waarden zijn fictief. ©HAN/Martijn Liebrand & Martijn van der Bruggen.

Zoals in figuur 1 valt te zien, toont de applicatie een overzicht van de gevonden interacties. Vervolgens toont de applicatie een venndiagram met overeenkomsten en verschillen tussen de verzamelingen van genen (per interactie) van mensen.

“Happy flow” van een bioloog die een interactions-bestand laadt, visualiseert en exporteert:

1. De bioloog drukt op de bladerbutton.
2. De bioloog selecteert een interactions-bestand middels een fileChooser.
3. De bioloog ziet het pad naar het interactions-bestand in een textField verschijnen en drukt op de openbutton.
4. De bioloog ziet statistieken van het ingelezen bestand (aantal genen per Tax ID, aantal interacties, aantal typen interacties) in een overzicht.
5. De bioloog kan in beide dropdownmenu's gevonden interacties selecteren. M.a.w., de dropdownmenu's zijn dynamische/afhankelijk van de gevonden typen interacties.
6. De bioloog selecteert één van de gevonden typen interacties (bv. downregulates, upregulates, cleaves, etc) per dropdown-menu.
7. De bioloog ziet in een venndiagram de aantallen genen die overlappen en verschillen tussen de twee gekozen verzamelingen van genen die horen bij de gekozen interacties.
8. De bioloog maakt een export (naar .txt file) van alle overlappende genen uit de venndiagram.
9. De bioloog maakt een export (naar .txt file) van alle gerelateerde PubMed ID's uit alle overlappende genen uit de venndiagram.

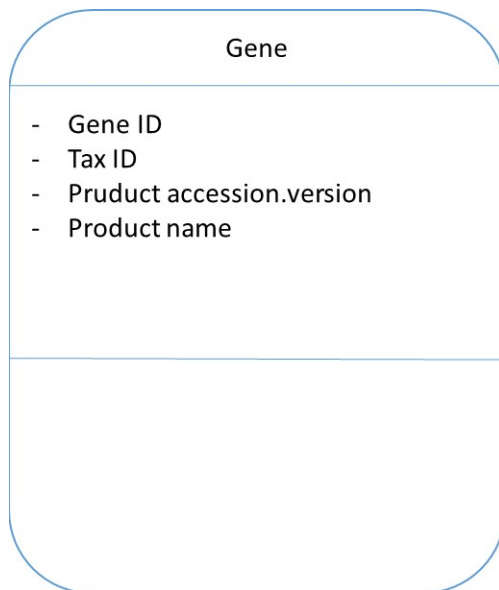
Creëer een applicatie die gegeven een interactions-bestand een verzameling van genen behorend bij een type interactie kan vergelijken met een verzameling van genen die horen bij een ander type interactie. Dit resulteert in een venndiagram met daarin de aantallen genen die overlappen, alsmede de aantallen genen die uniek zijn voor beide verzamelingen.

Technische eisen:

- De applicatie is in Java geschreven
- GUI logica moet apart worden geïmplementeerd van de applicatie logica.
- De verzamelingen bestaan uit Gene objecten.
- Big O voor de vergelijking van genen is zo klein mogelijk.
- Er wordt gebruik gemaakt van encapsulatie voor het opslaan van gegevens in Gene objecten.
- Genen kunnen middels het implementeren van de comparable interface gesorteerd worden op Gene ID.
- De applicatie dient goed gedocumenteerd (javadoc) ingeleverd te worden als een NetBeans project op GitHub.

Functionele eisen:

- De applicatie bevat een bladerbutton en een openbutton. Middels een bladerbutton selecteert een gebruiker een bestand welke verschijnt in een tekstveld. Middels een openbutton wordt het tekstveld uitgelezen en het getoonde bestand geopend.
- De applicatie laat statistieken uit het interactions-bestand zien. (aantal genen per Tax ID, aantal relaties, aantal typen relaties.)
- Middels twee dropdown menu's kan de gebruiker gevonden type interactie selecteren.
- Een venndiagram toont de hoeveelheid in verschil en overeenkomst tussen de te vergelijken verzamelingen.
- Via een exportbutton kunnen alle overlappende genen worden geëxporteerd naar een .txt bestand. Het bestand bevat per gen alle informatie uit het Gene-object.
- Via een tweede exportbutton kunnen alle PubMed ID's van alle overlappende genen (zoals getoond in de venndiagram) worden geëxporteerd naar een .txt bestand.
- Indien er geen genen of PubMed ID's zijn om te exporteren, werpt de applicatie een custom exception op en zal de gebruiker middels een pop-up geïnformeerd worden over de foutmelding.

Gene object:

Figuur 2: Het Gene object met bijbehorende attributen. De getters en setters dienen hieraan toegevoegd te worden. Let op, PubMed ID's zijn gebonden aan relaties, niet aan Gene objecten. Verder bevat het object geen informatie over relaties. ©HAN/Martijn Liebrand & Martijn van der Bruggen.

Bi6a beoordelingsformulier eindopdracht

Naam: Klas: Datum:

Competentie 2.2 (1 t/m 12)		Beoordeling door	
Onderdeel	Score	Student	Docent
1. Algemeen			
a. Code is generiek geschreven (modulaire opbouw)	10	10	
b. Exception handling: adequaat en volledig. (denk aan export van gegevens)	10	8	
2. Verwerking data			
a. Verwerking van te lezen data (efficiënt lezen en schrijven van databestanden)	10	10	
b. Generieke verwerking van data naar dataset(s).	10	8	
3. Toepassing datastructuur			
a. Gebruik van juiste geneste datastructuur.	10	7	
b. Toepassing van datastructuur is efficiënt en optimaal: juiste Big-O.	10	9	
4. Algoritme			
a. Goede analyse van biologisch probleem en juiste vertaling naar programma (functionaliteit)	10	10	
b. Opzet en werking van gevraagde algoritme (techniek)	10	8	
c. Complexiteit van algoritme (generieke methoden)	10	8	
5. GUI			
a. Flexibiliteit gevraagde GUI elementen: GUI is flexibel en reageert op wijzigingen run time.	10	10	
Eindoordeel (Het cijfer is het aantal behaalde punten delen door het maximaal aantal te behalen punten maal 10.)	Max: 100	88	
Feedback:			
Naam en paraaf beoordelaar:			

Inleveren

Geschreven code (in Netbeans project op GitHub) en een eigen beoordeling van je geschreven code aan de hand van bovenstaande beoordelingscriteria. Vul daartoe de kolom student in.

N.B. het eindoordeel van de docent is hetgeen geadministreerd wordt. Lever je eigen beoordeling en de link naar je project op GitHub in via mail bij Martijn Liebrand.

Gebruikte bronnen

1. ftp://ftp.ncbi.nlm.nih.gov/gene/GeneRIF/hiv_interactions.gz
2. <ftp://ftp.ncbi.nlm.nih.gov/gene/README>