**Technische documentatie**



# Inleiding

De champignonteelt maakt het grootste deel uit van de paddenstoel productie industrie. Een groot probleem binnen de champignonteelt is dat de oogst van champignons sterk kan variëren. De in de compost aanwezige micro-organismen spelen daarbij een grote rol. Tijdens de compost bereiding wisselt deze microbiële samenstelling voortdurend en heeft invloed op de oogst. In opdracht van BioCentre is een applicatie ontwikkelt om sequencing data te analyseren, een database ontworpen en gecreëerd om de gegevens in op te slaan, en een webapplicatie opgebouwd om deze resultaten te visualiseren. Het ERD is in Vertabelo ontworpen en gecreëerd in MySQL (versie 5.1.46), en deze bestaat uit 5 tabellen. De applicatie is ontwikkelt in Python (versie 3.6) en maakt gebruik van 3 verschillende scripts. Het eerste script leest een Excel bestand met daarin de sequencing data, dit gebeurt m.b.v. de openpyxl en TkInter modules. De headers worden aangepast om als bestandsnaam gebruikt te worden, en de sequenties worden gecontroleerd of ze DNA sequenties zijn. Vervolgens worden de sequenties aligned door BLASTx m.b.v. de Biopython module, en opgeslagen als XML bestanden onder de naam v.d. bijbehorende header. Het tweede script gebruikt de eerder gemaakte bestanden en leest deze wederom m.b.v. Biopython uit. Deze data wordt gesorteerd en gebruiksklaar gemaakt om via Python automatisch de database te vullen m.b.v. de mysql-connector-python module. Het derde script haalt de gegevens op van de micro-organismen en de eiwit accessiecodes, en gebruikt dit resultaat om zowel de taxonomie als eiwit functie op te halen.

Dit de data wordt opgehaald m.b.v. de Biopython module, dit gebeurt over een connectie aan de NCBI servers. Wegens het hoge aantal requests en tijdsduur van dit proces wordt gebruik gemaakt van de urllib modules om mogelijke connectieproblemen af te vangen. De webapplicatie draait op het Flask framework (versie: 1.0.2) en bestaat uit 1 Python script, 1 CSS (versie 4.13) bestand en 5 HTML5 templates. De site is online gezet op een Microsoft Soft Azure server van de Hogeschool Arnhem en Nijmegen.

De site kan benadert worden via dit URL: **< vul website url hier in >**

# Werking van de applicatie

# Software Architectuur

**Software**

Op de Front tier draait Windows 10 met daarop een installatie van Firefox 67.0.2 (64 bits). De webapplicatie maakt gebruik van HTML5 ( inclusief JavaScript embedded in HTML) en CSS 4.13.

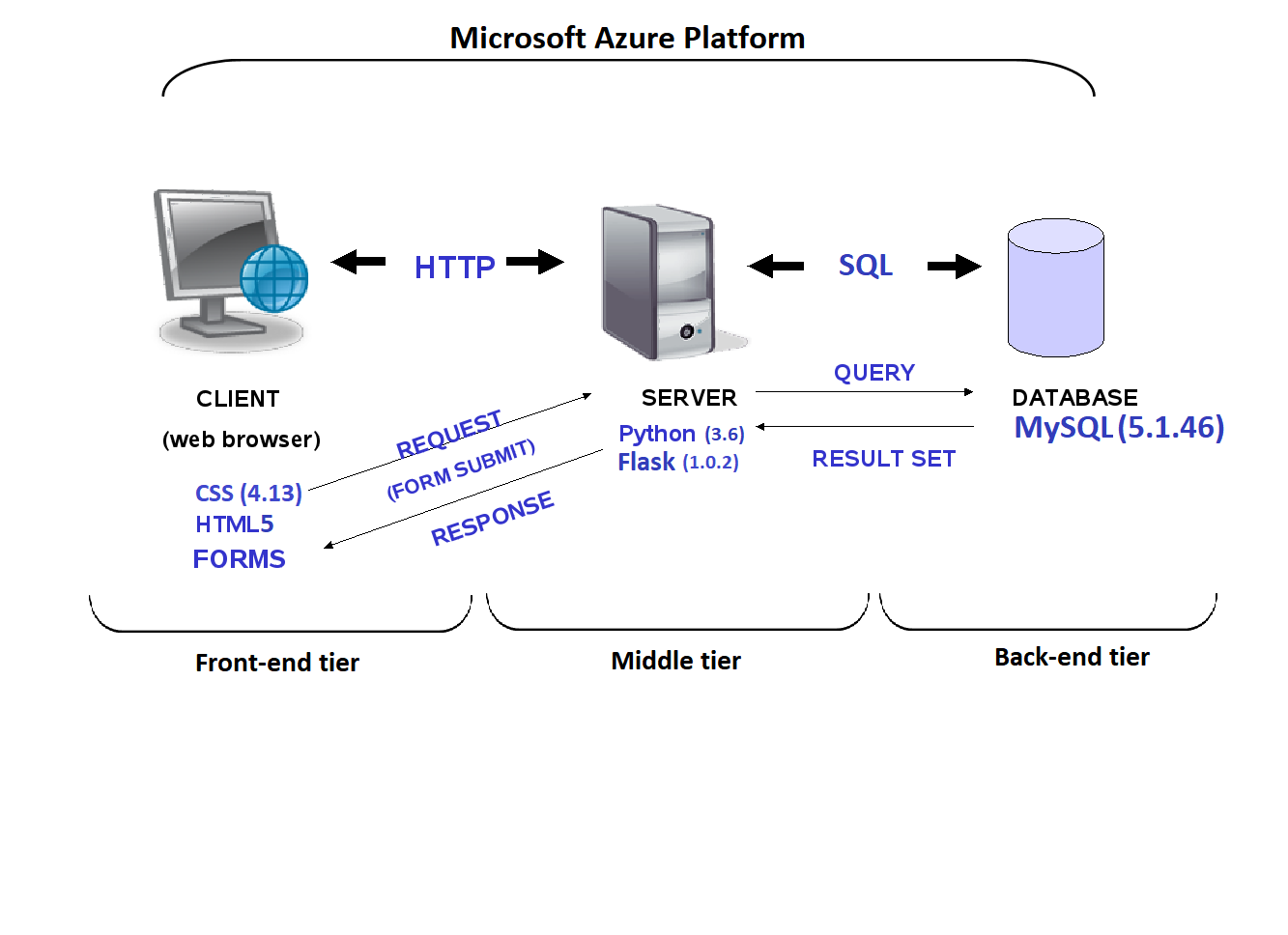
Op de Middle tier draait Windows 10 met daarop een installatie van Python 3.6 op een Flask 1.0.2 framework. Op Python 3.6 zijn de volgende modules geïnstalleerd:

-biopython 1.73  
-openpyxl 2.6.2   
-mysql-connector-python 8.0.16

Op de Back-end tier draait Windows 10 met daarop een installatie van MySQL 5.1.46 met de InnoDB engine.

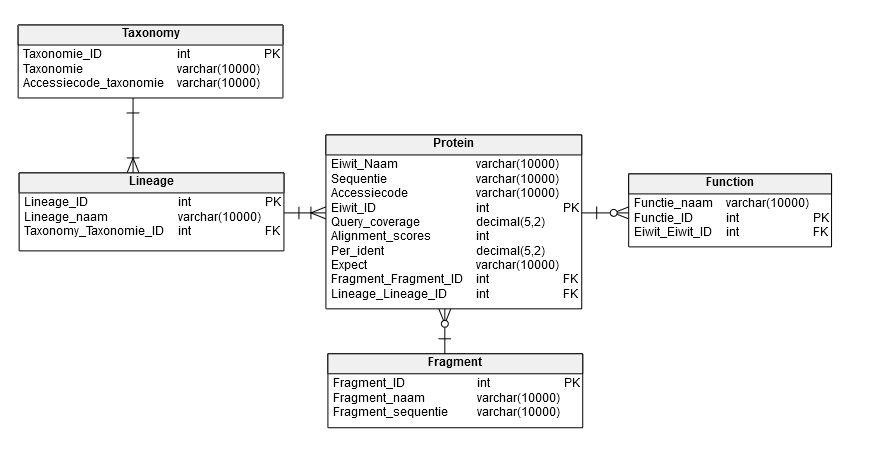
De applicatie draait online op een Microsoft Azure Platform van de HAN.

**Schematische weergave van architectuur**

****

**Database**

De database is gecreëerd in MySQL 5.1.46 m.b.v. een gegenereerd SQL script uit Vertabelo. Na het genereren van dit script is handmatig een auto-increment toegevoegd aan elke primary key. In de opzet van het ERD is er gekozen om 5 tabellen te maken. De tabel Fragment bevat de gebruikte header van het bestand en de bijbehorende DNA sequentie. Elke sequentie is aligned via BLASTx en levert verscheidene hits of juist geen hits op, dit wordt uitgedrukt als 1 op veel relatie. Deze hits bevatten een eiwit, organisme en score. De eiwitten worden in tabel Protein opgeslagen, samen met bijbehorende scores. In tabel Lineage worden de organismen opgeslagen waarvan elk eiwit 1 organisme heeft, maar een organisme kan meerder eiwitten bevatten, dit wordt uitgedrukt als 1 op veel relatie. Via een ander script wordt de eiwit accessiecode gebruikt om extra informatie op te vragen over de functie, en dit wordt opgeslagen in tabel Function. Elk eiwit kan geen of meer functies hebben, maar elke functie zit aan een eiwit, uitgedrukt als 1 op veel relatie .Via de organismen wordt een taxonomie opgehaald en in de tabel Taxonomy geplaatst. Elk organisme heeft 1 taxonomie, en omdat de taxonomie niet inclusief species is kunnen meerdere species dezelfde taxonomie bevatten, dit wordt uitgedrukt met een 1 op veel relatie.



# Test scripts

**Python script voor webapplicatie: Application.py**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Testscript: zoekbalk filtert de database gegevens op het ingevoerde zoekwoord | | | | |
| Stap # | **Instructie** | **Verwacht resultaat** | **Feitelijk resultaat** | **Opmerking** |
| 1 | Opent de website | Website opent op Home page | Werkt | Firefox als webbrowser gebruikt |
| 2 | Klikt op tabblad Results | Results pagina opent | Werkt |  |
| 3 | Vul niks in / vink niks aan en klik direct op submit | Er verschijnt geen tabel | Werkt |  |
| 4 | Vul 'proteobacteria' in / vink niks aan en klik op submit | Er verschijnt geen tabel | Werkt |  |
| 5 | Vul 'proteobacteria' in / vink alles aan en klik op submit | Tabel met resultaten verschijnt. | Werkt | Limiet niet ingevuld, wordt limit 100 gebruikt |
| 6 | Vul 'proteobacteria' in / vink alles aan en vul limiet 50 in. Klik op submit. | Tabel met resultaten verschijnt. | Werkt |  |
| 7 | Vul 'proteobacteria' in / vink niks aan en vul limiet 50 in. Klik op submit. | Er verschijnt geen tabel | Werkt |  |
| 8 | Vul '54543704085' in / vink alles aan en vul limiet 50 in. Klik op submit. | Alleen de aangevinkte kolomnamen verschijnen. | Werkt |  |
| 9 | Vul '%.;''\*``~~' in / vink alles aan en vul limiet 50 in. Klik op submit. | Alleen de aangevinkte kolomnamen verschijnen. | Werkt |  |
| 10 | Vul 'NULL' in / vink alles aan en vul limiet 50 in. Klik op submit. | Alleen de aangevinkte kolomnamen verschijnen. | Werkt |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Testscript: limietbalk filtert de database gegevens op het ingevoerde zoekwoord | | | | |
| Stap # | **Instructie** | **Verwacht resultaat** | **Feitelijk resultaat** | **Opmerking** |
| 1 | Opent de website | Website opent op Home page | Werkt | Firefox als webbrowser gebruikt |
| 2 | Klikt op tabblad Results | Results pagina opent | Werkt |  |
| 3 | Vul niks in / vink niks aan en klik direct op submit | Er verschijnt geen tabel | Werkt |  |
| 4 | Vul '50' in / vink niks aan en klik op submit | Er verschijnt geen tabel | Werkt |  |
| 5 | Vul '50' in / vink alles aan en klik op submit | Alleen de aangevinkte kolomnamen verschijnen. | Werkt |  |
| 6 | Vul bij zoekbalk 'proteobacteria' in / vink niks aan en vul in limietbalk 'woord' in. Klik op submit. | Tabel met resultaten verschijnt. | Werkt | Indien er geen input is, of input niet bestaande uit getallen, wordt limit 100 gebruikt |
| 7 | Vul bij zoekbalk 'proteobacteria' in / vink niks aan en vul in limietbalk '99999999' in. Klik op submit. | Tabel met resultaten verschijnt. | Werkt |  |
| 8 | Vul bij zoekbalk 'proteobacteria' in / vink niks aan en vul in limietbalk ''%.;''\*``~~'' in. Klik op submit. | Tabel met resultaten verschijnt. | Werkt |  |
| 9 | Vul bij zoekbalk 'proteobacteria' in / vink niks aan en vul in limietbalk 'NULL' in. Klik op submit. | Tabel met resultaten verschijnt. | Werkt |  |

**Python script voor BLASTx en maken XML bestanden: Final\_version\_Blast\_get.py**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Testscript: | | | | |
| Stap # | **Instructie** | **Verwacht resultaat** | **Feitelijk resultaat** | **Opmerking** |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |
| 6 |  |  |  |  |