



Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
Σχολή Θετικών Επιστημών
Τμήμα Βιολογίας

*«Ανάπτυξη εφαρμογής Web για την πλοήγηση,
την αναζήτηση και τη διαχείριση βιολογικών
δεδομένων»*

Επιλεγείσα Βάση Δεδομένων: DisGeNET

Αικατερίνη-Μαρία Χαλιάσου

ΠΜΣ “Βιοπληροφορική και Υπολογιστική Βιολογία”
Μάθημα: Αρχιτεκτονική Εφαρμογών Διαδικτύου και Βιοπληροφορική
Διδάσκων: Δρ. Ι. Χαμόδρακας

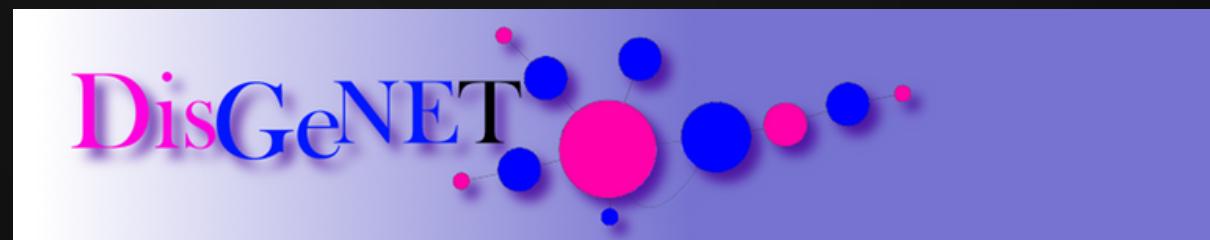
Εισαγωγικά ...

- *Περιεχόμενο*

Η Βάση Δεδομένων DisGeNET είναι μία πλατφόρμα που περιέχει μια από τις μεγαλύτερες δημόσια διαθέσιμες συλλογές γονιδίων και παραλλαγών που σχετίζονται με ανθρώπινες ασθένειες (Piñero et al., 2021- Piñero et al., 2019- Piñero et al., 2016- Piñero et al., 2015).

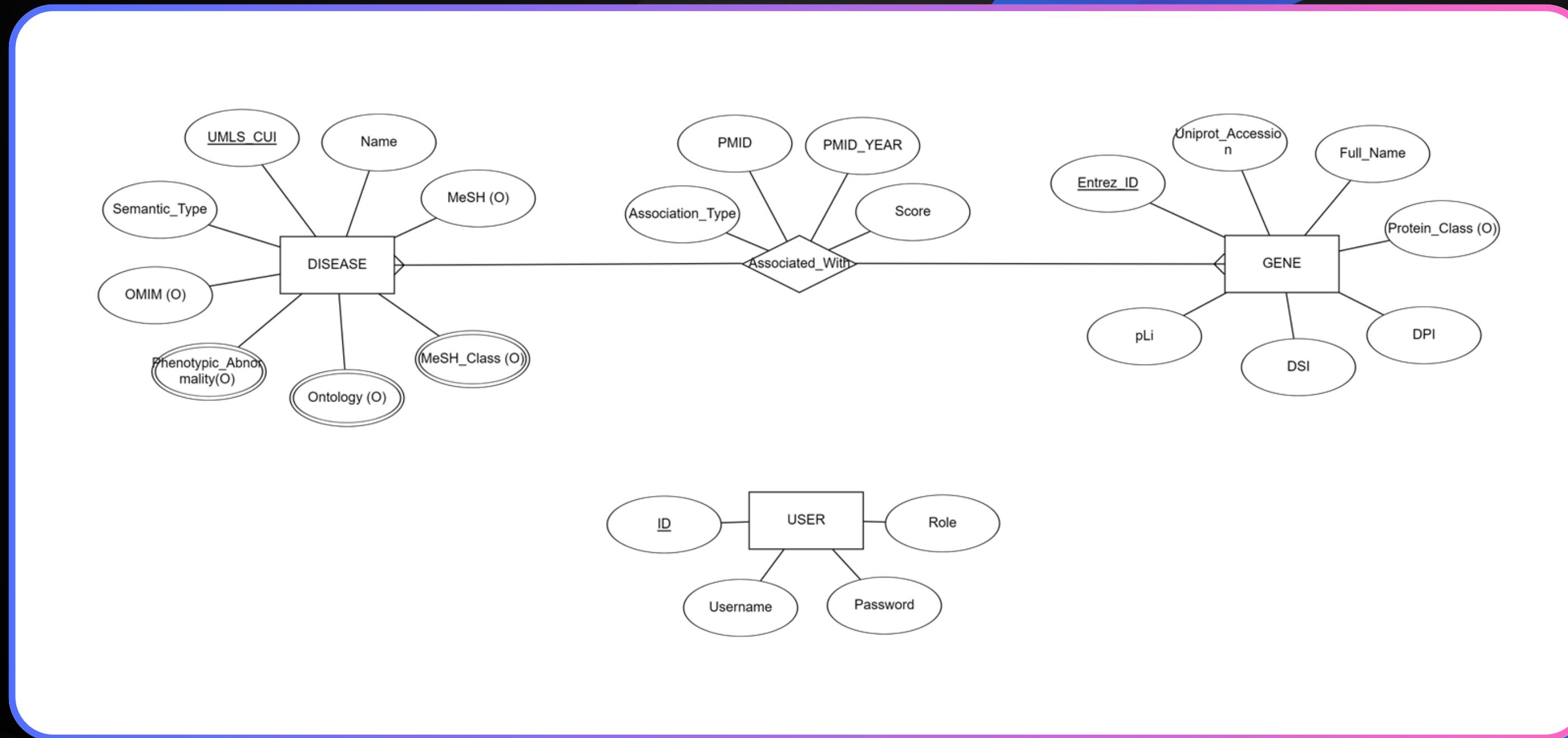
- *Όγκος δεδομένων*

- Η τρέχουσα έκδοση του DisGeNET (v7.0) περιέχει 1.134.942 συσχετίσεις γονιδίων-ασθένειας (GDAs), μεταξύ 21.671 γονιδίων και 30.170 ασθενειών, διαταραχών, χαρακτηριστικών και κλινικών ή μη φυσιολογικών ανθρώπινων φαινοτύπων.



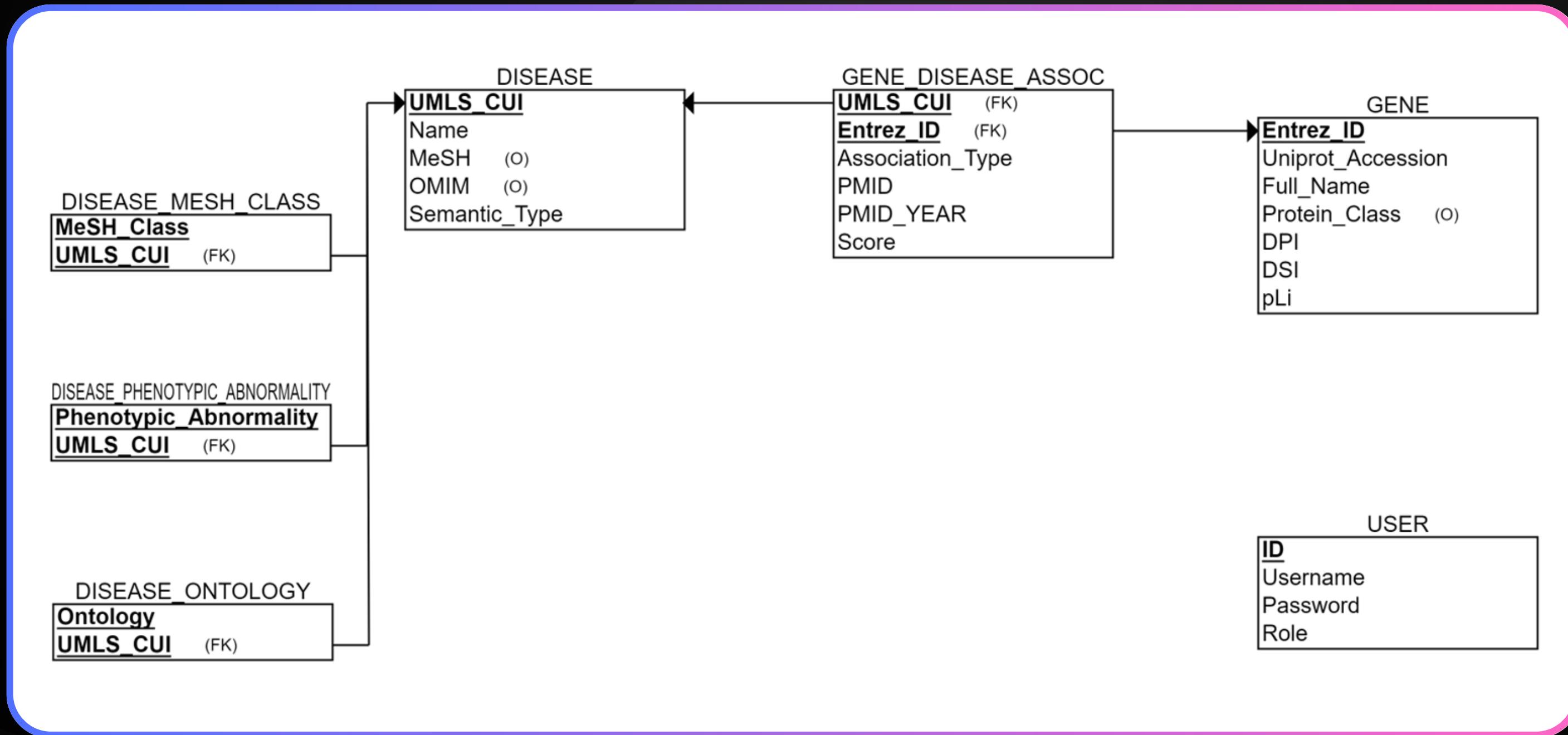
Προσέγγιση DisGeNET

Μοντέλο Οντοτήτων Συσχετίσεων



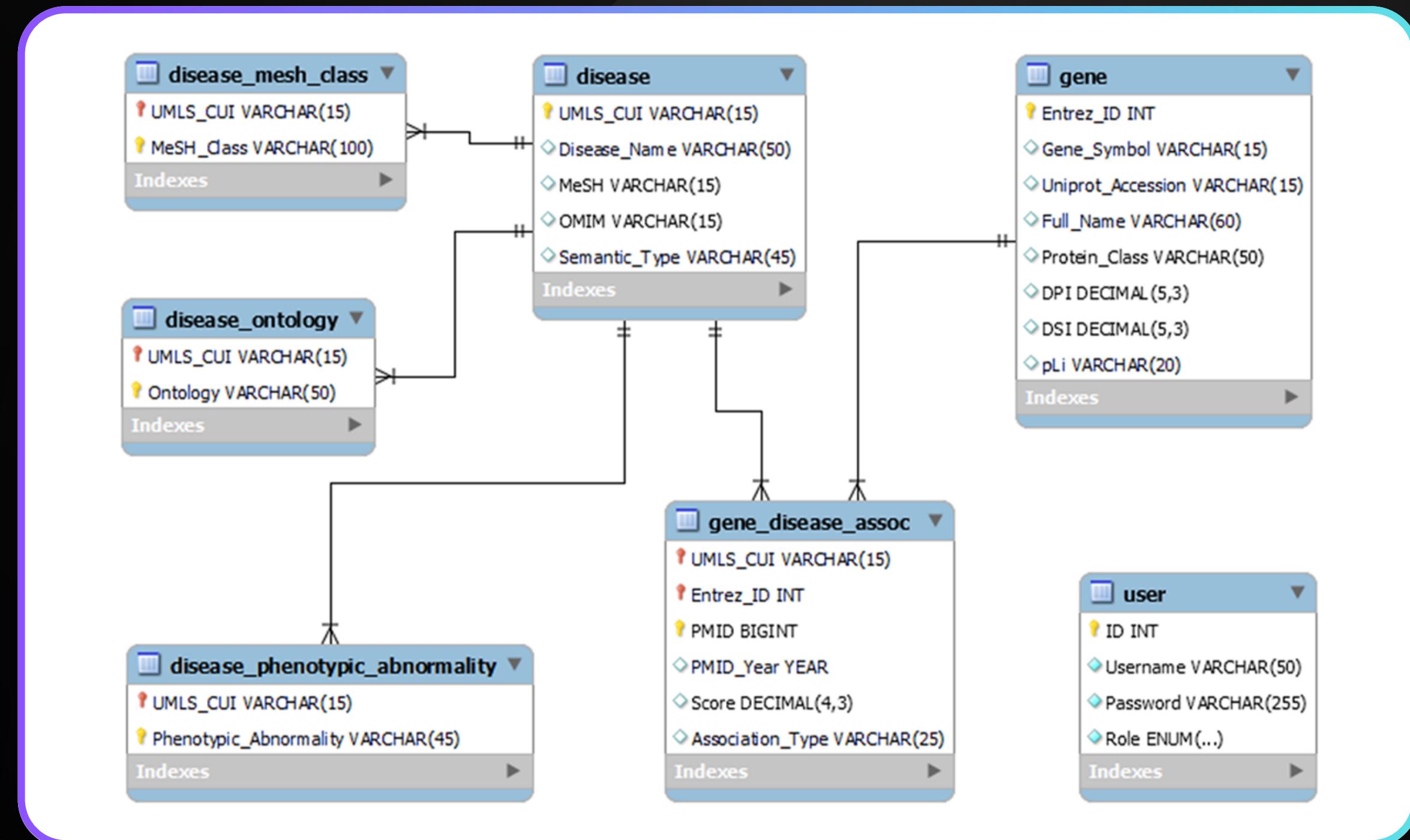
Προσέγγιση DisGeNET

Σχεσιακό Μοντέλο



Προσέγγιση DisGeNET

Σχεσιακό Μοντέλο στη MySQL



Υλοποίηση Εφαρμογής



- Υλοποίηση με ReactJS σε Typescript:
 - Components και Icons με Antd
- Υλοποίηση σε Python:
 - FastAPI web framework
 - Uvicorn: Server συμβατός με Async Server Gateway Interface (ASGI). Δρα ως το συνδετικό στοιχείο που χειρίζεται τις συνδέσεις ιστού από το πρόγραμμα περιήγησης ή τον api client και στη συνέχεια επιτρέπει στο FastAPI να εξυπηρετήσει το request.
- Υλοποίηση μέσω MySQL Workbench

Υλοποίηση Εφαρμογής

Back-end

Δημιουργία κατάλληλου environment:

```
conda create -n webApp python=3.9
```

Εγκατάσταση απαραίτητων πακέτων:

```
pip install fastapi/sqlalchemy/uvicorn/pydantic
```

Επικοινωνία με τη Βάση ορίζοντας το DATABASE_URL
="mysql://root:root@localhost:3306/mybiodb" και εγκαθίδρυση σύνδεσης μέσω sqlalchemy

```
backend > db > database.py > ...
1
2 from sqlalchemy import create_engine
3 from sqlalchemy.orm import sessionmaker
4 from config import DATABASE_URL
5
6 engine = create_engine(DATABASE_URL)
7 SessionLocal = sessionmaker(bind=engine)
```

Ρύθμιση FastAPI με την υποστήριξη CORS και συμπερίληψη routes και route handlers

Ξεκίνημα uvicorn server (cd backend):

```
uvicorn main:app --reload
```

backend > main.py > ...

```
1 from fastapi import FastAPI
2 from fastapi.middleware.cors import CORSMiddleware
3 from fastapi.responses import JSONResponse
4
5 # Import the route modules
6 from routes import disease, gene, user
7
8 # FastAPI app initialization
9 app = FastAPI(
10     title="BioInformatics",
11     description="Bioinformatics Web Application",
12     version="1.0.0",
13 )
14
15
16 origins = [
17     "http://localhost",
18     "http://localhost:3000",
19     "*",
20 ]
21
22 app.add_middleware(
23     CORSMiddleware,
24     allow_origins=origins,
25     allow_credentials=True,
26     allow_methods=["*"],
27     allow_headers=["*"],
28 )
```

Υλοποίηση Εφαρμογής

Front-end

Δημιουργία React Project:

```
npx create-react-app frontend --template typescript
```

Εγκατάσταση απαραίτητων πακέτων και ξεκίνημα εφαρμογής:

```
npm start
```

Δημιουργία MobX stores για την παρακολούθηση αλλαγών των ιδιοτήτων/μεθόδων των κλάσεων -> State Management.

Ρύθμιση μιας οντότητας axios (=http) ώστε να πραγματοποιεί αιτήσεις HTTP στο backend διακομιστή και συμπερίληψη ενός interceptor που ελέγχει την ύπαρξη authorization token στο localStorage του browser.

Δημιουργία κλάσης bioService που εμπεριέχει όλες τις μεθόδους αλληλεπίδρασης με το backend API και τα dtos.

```
class BioService {
  public async getDiseases(search: string | undefined): Promise<DISEASE[]> {
    const result = await httpService.get(
      `/api/disease/diseases${search ? `?search=${search}` : ""}`
    );
    return result.data;
  }
}

export interface DISEASE {
  Semantic_Type: string
  Disease_Name: string
  MeSH: string
  UMLS_CUI: string
  OMIM: string | null
}
```

frontend > src > services > **TS** httpService.ts > ...

```
1 import Modal from "antd/lib/modal";
2 import axios from "axios";
3
4 const http = axios.create({
5   baseURL: "http://localhost:8000",
6   headers: {
7     "Content-type": "application/json",
8   },
9 });
10
11 // Request interceptor
12 http.interceptors.request.use(
13   (config: any) => {
14     // Set your auth token from localStorage
15     const token = localStorage.getItem("token");
16     if (token) {
17       config.headers.Authorization = `Bearer ${token}`;
18     }
19
20     return config;
21   },
22   (error: any) => {
23     // Do something with request error
24     return Promise.reject(error);
25   }
26 );
```

Το αποτέλεσμα

The screenshot shows a web application interface with a dark blue background featuring a 3D DNA helix model. At the top, there is a navigation bar with a user icon (N), a "Logout" button, and a search bar containing the placeholder "Type your search terms". To the left of the search bar is a dropdown menu titled "Select Entries" with two options: "Disease" and "Gene". A large, semi-transparent white box is centered on the page, containing the following text in a light blue font:

*Please select one data category from the drop-down list
in the search bar to browse or interact with it*

At the bottom left of the main content area, there is a small footer text: "Bio web App ©2023".

Το αποτέλεσμα

The screenshot displays a web-based application interface for managing biological data, specifically focusing on disease entries. The background of the application is a detailed 3D rendering of a DNA double helix.

The interface includes the following components:

- Header:** A top navigation bar with a user icon labeled 'N' and a 'Logout' button.
- Left Sidebar:** A vertical sidebar titled 'Select Entries' containing buttons for 'Disease' and 'Gene'.
- Search Bar:** A search bar at the top of the main content area with a dropdown menu set to 'Disease' and a placeholder 'Type your search terms'.
- Table Header:** A header row for the 'DISEASE ENTRIES' table with columns: UMLS_CUI, Disease_Name, MeSH, OMIM, Semantic_Type, and Actions.
- Table Data:** A list of disease entries:

	UMLS_CUI	Disease_Name	MeSH	OMIM	Semantic_Type	Actions
▼	C0001080	Achondroplasia	D000130	100800	Congenital Abnormality	
▼	C0002395	Alzheimer's Disease	D000544	104300	Disease or Syndrome	
▼	C0003872	Arthritis, Psoriatic	D015535		Disease or Syndrome	
▼	C0003873	Rheumatoid Arthritis	D001172	180300	Disease or Syndrome	
▼	C0024141	Lupus Erythematosus, Systemic	D008180	152700	Disease or Syndrome	
▼	C0028754	Obesity	D009765	601665	Disease or Syndrome	
▼	C0030567	Parkinson Disease	D010300	516000	Disease or Syndrome	
▼	C0033860	Psoriasis	D011565		Disease or Syndrome	

Bio web App ©2023

Το αποτέλεσμα

The screenshot displays a bioinformatics web application interface. At the top, there are two tabs: "Disease" and "Gene". The "Gene" tab is selected, indicated by a blue background and white text. Below the tabs is a search bar with the placeholder "Type your search terms" and a magnifying glass icon.

On the left side, there is a sidebar titled "Select Entries" with two dropdown menus: "Disease" and "Gene". The "Gene" menu is currently active. A modal window titled "Please select entries" is open, showing a table of gene entries:

	UMLS_CUI	Disease_N
▼	C0001080	Achondroplasia
▼	C0002395	Alzheimer disease
▼	C0003872	Arthritis, Psoriatic
▼	C0003873	Rheumatoid arthritis
▼	C0024141	Lupus Erythematosus
▼	C0028754	Obesity
▼	C0030567	Parkinson disease
▼	C0033860	Psoriasis

On the right side, the main content area is titled "GENE ENTRIES" and contains a table of gene entries:

	Entrez_ID	Gene_Symbol	Uniprot_Accession	Full_Name	Protein_Class	DPI	DSI	pLi	Actions
▼	348	APOE	P02649	apolipoprotein E		0.962	0.338	0.0018685	
▼	351	APP	P05067	amyloid beta precursor protein	Enzyme modulator	0.846	0.422	0.046544	
▼	716	C1S	O60260	parkin RBR E3 ubiquitin protein ligase	Enzyme	0.846	0.431	0.94085	
▼	958	CD40	P25942	CD40 molecule		0.846	0.396	0.84603	
▼	968	CD68	P34810	CD68 molecule	Transporter	0.808	0.408	0.0000014063	
▼	1401	CRP	P02741	C-reactive protein		0.962	0.299	0.0036969	
▼	1471	CST3	P01034	cystatin C		0.808	0.436	0.0039011	

At the bottom center, the text "Bio web App ©2023" is visible.

Λειτουργικότητα Εφαρμογής

Πλοήγηση στα δεδομένα

Γίνεται με την επιλογή μίας κατηγορίας Disease / Gene από τις drop-down επιλογές στη μπάρα αναζήτησης.

Άνοιγμα περισσότερων λεπτομερειών αλλά και των συσχετίσεών μίας εγγραφής με την επιλογή της στον πίνακα.

Λειτουργικότητα Εφαρμογής

Αναζήτηση δεδομένων

- Συγκεκριμένη, με την εισαγωγή ενός μοναδικού γνωρίσματος που φέρει η εγγραφή προς αναζήτηση.
- Γενική, με την εισαγωγή μιας γενικής ομάδας στην οποία μπορεί να ανήκουν παραπάνω από μία εγγραφές.
- Βάσει ελεύθερου κειμένου

The screenshot displays a web-based bioinformatics application. At the top, there is a navigation bar with tabs for 'Gene' and 'Enzy', where 'Enzy' is highlighted with a purple border. To the right of the tabs are a search icon and a 'Logout' button. The background features a blue and white DNA helix graphic. Below the navigation bar, the title 'GENE ENTRIES' is centered above a table. The table has columns for Entrez_ID, Gene_Symbol, Uniprot_Accession, Full_Name, Protein_Class, DPI, DSI, pLi, and Actions. Each row contains a disclosure triangle, the Entrez ID, gene symbol, UniProt accession number, full name, protein class (highlighted with a purple border), and numerical values for DPI, DSI, and pLi, along with edit and delete icons. A '+' icon is located in the top right corner of the table area. At the bottom of the page, the text 'Bio web App ©2023' is visible.

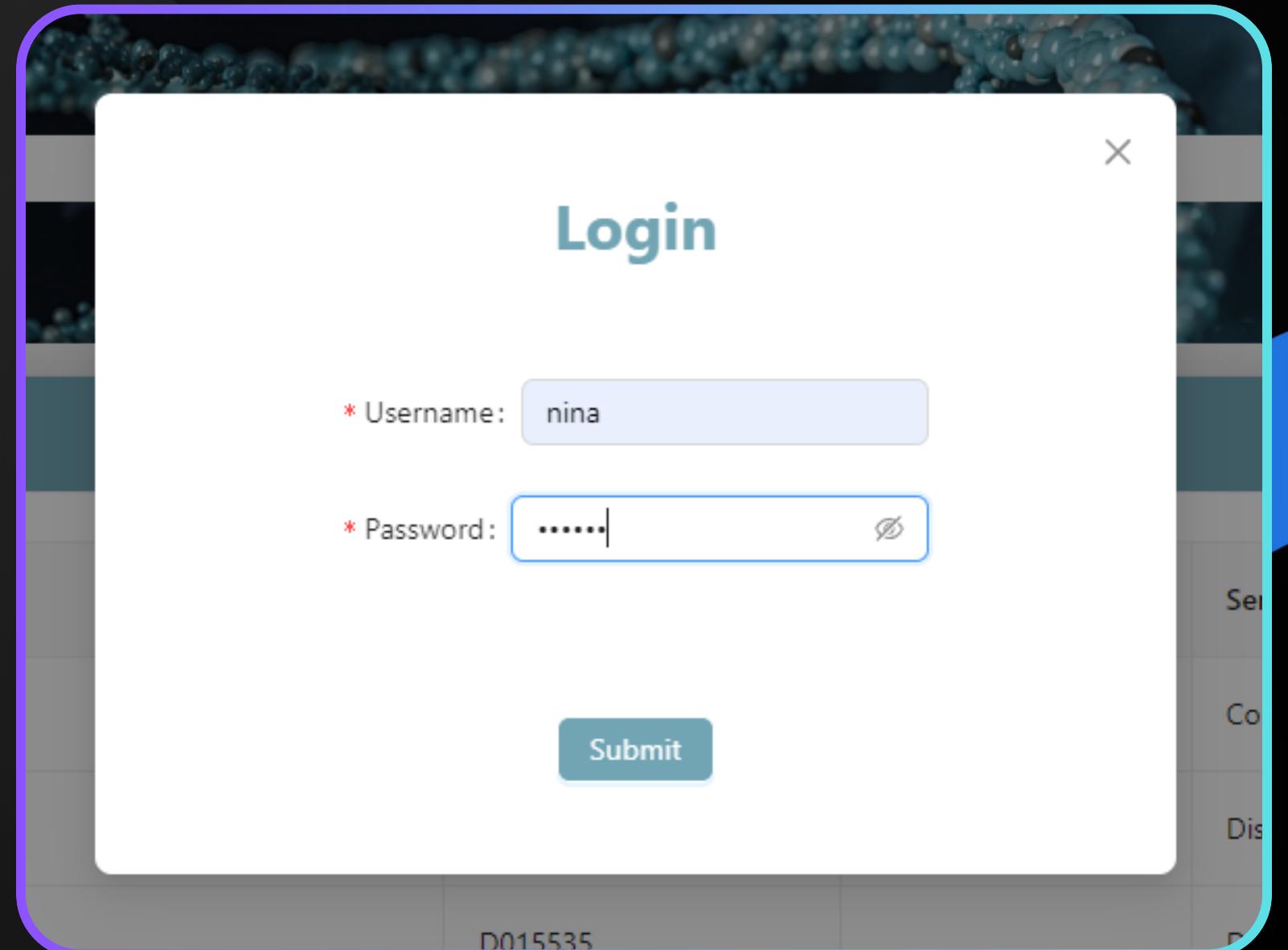
	Entrez_ID	Gene_Symbol	Uniprot_Accession	Full_Name	Protein_Class	DPI	DSI	pLi	Actions
▼	7128	TNFAIP3	P21580	TNF alpha induced protein 3	Enzyme	0.808	0.481	0.99972	
▼	7038	TG	P01266	thyroglobulin	Enzyme modulator	0.808	0.48	2.8802e-59	
▼	5071	PRKN	O60260	parkin RBR E3 ubiquitin protein ligase	Enzyme	0.846	0.431	6.9346e-7	
▼	4311	MME	P08473	membrane metalloendopeptidase	Enzyme	0.808	0.407	1.465e-18	
▼	2147	F2	P00734	coagulation factor II, thrombin	Enzyme	0.885	0.415	0.0011118	
▼	1636	ACE	P12821	angiotensin I converting enzyme	Enzyme	0.923	0.328	1.0344e-37	
▼	716	C1S	O60260	parkin RBR E3 ubiquitin protein ligase	Enzyme	0.846	0.431	0.94085	
▼	351	APP	P05067	amyloid beta precursor protein	Enzyme modulator	0.846	0.422	0.046544	

Bio web App ©2023

Λειτουργικότητα Εφαρμογής

Ρόλοι χρηστών

- Εισαγωγή χρήστη administrator στην βάση με username: nina, password: 123456 και role: amdin
- By default η εφαρμογή λειτουργεί σε anonymous mode
- Απαιτείται σύνδεση του administrator για διαχείριση των δεδομένων (εισαγωγή, τροποποίηση, διαγραφή)



Λειτουργικότητα Εφαρμογής

Εισαγωγή Δεδομένων

Insert New Disease Or Association

* UMLS_CUI

Disease Name

MeSH

OMIM

Semantic Type

Disease Mesh Class

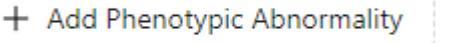
+ Add MeSH Class

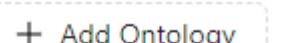
N Logout

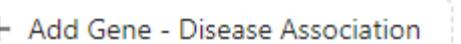
Disease Typeahead search bar: Type your search terms 

DISEASE ENTRIES

UMLS_CUI	Disease_Name	MeSH	OMIM	Semantic_Type	Actions
					

Disease Phenotypic Abnormality 

Disease Ontology 

Gene Disease Association 

Submit

Λειτουργικότητα Εφαρμογής

Τροποποίηση Δεδομένων

Edit C0001080

* UMLS_CUI
C0001080

Disease Name
Achondroplasia

MeSH
D000130

OMIM
100800

Semantic Type
Congenital Abnormality

Disease Mesh Class

* MeSH Class
Congenital, Hereditary, an

Gene Disease Assoc

* Entrez_ID	* PMID	PMID_Year	Score	Association_Type
2261	31048079	2020	1.000	GeneticVariation
2688	30706088	2019	0.090	Biomarker
3605	31184530	2019	0.010	GeneticVariation
4882	22645228	2012	0.210	GeneticVariation
5741	29104492	2017	0.030	Biomarker

+ Add GENE DISEASE ASSOC

Update Disease

Λειτουργικότητα Εφαρμογής

Διαγραφή Δεδομένων

The screenshot displays a web-based application interface for managing gene entries. The background features a blue DNA helix graphic. At the top, there is a navigation bar with tabs for "Gene" and "Enzy", a search bar, and user account information ("N" and "Logout"). Below the header, a table lists gene entries with columns for Entrez_ID, Gene_Symbol, Uniprot_Accession, Full_Name, Protein_Class, DPI, DSI, and two small icons. A modal dialog box is overlaid on the table, asking "Are you sure you want to delete this entry?" with "Cancel" and "OK" buttons.

Entrez_ID	Gene_Symbol	Uniprot_Accession	Full_Name	Protein_Class	DPI	DSI		
7128	TNFAIP3	P21580	TNF alpha induced protein 3	Enzyme	0.808	0.481	0.99972	
7038	TG	P01266	thyroglobulin	Enzyme modulator	0.808	0.48	2.8802e-59	
5071	PRKN	O60260	parkin RBR E3 ubiquitin protein ligase	Enzyme	0.846	0.431	6.9346e-7	
4311	MME	P08473	membrane metalloendopeptidase	Enzyme	0.808	0.407	1.465e-18	
2147	F2	P00734	coagulation factor II, thrombin	Enzyme	0.885	0.415	0.0011118	
1636	ACE	P12821	angiotensin I converting enzyme	Enzyme	0.923	0.328	1.0344e-37	
716	C1S	O60260	parkin RBR E3 ubiquitin protein ligase	Enzyme	0.846	0.431	0.94085	
351	APP	P05067	amyloid beta precursor protein	Enzyme modulator	0.846	0.422	0.046544	

Bio web App ©2023

Σας ευχαριστώ πολύ
για την προσοχή σας!