

CUMULUS

Team members:

Cretu Andra

Scortanu Raluca

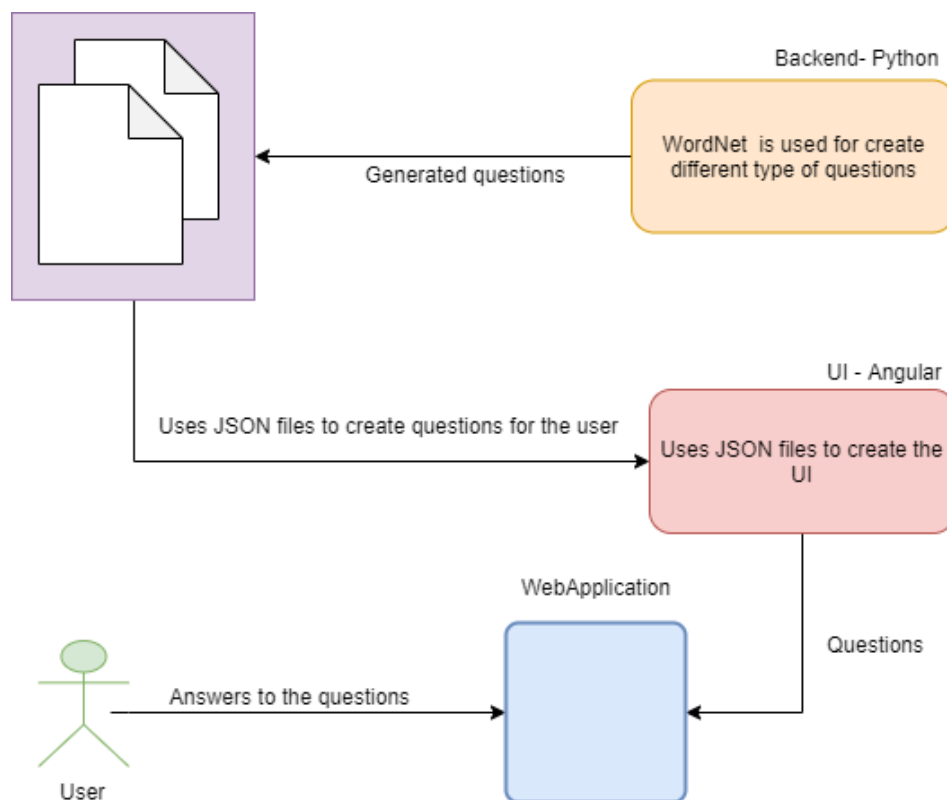
GitHub repository: https://github.com/CretuAndraElena/Proiect_NLP/tree/master/Proiect

Project's short description (max 50 words):

CUMULUS a fost gandit ca un POC(proof of concept) sau un proiect pilot, in care am intentionat sa demonstram că este posibila construirea unei sistem de tip e-learning pornind de la functionalitatile oferite de o librerie. Am avut ca punct de plecare libraria wordnet care a fost ajutata de alte cateva librarii si framework-uri pentru a dezvolta acest prototip al aplicatiei.

Aplicatia are ca scop oferirea unui mediu placut de invatare al unei limbi straine. In etapa actuala a aplicatiei utilizatorii isi vor putea testa cunostintele cu ajutorul unor seturi de intrebari cu o dificultate redusa (variante de raspuns sau input) , impartite pe diferite domenii (animale, sporturi, ocupatii, culori) care se adreseaza utilizatorilor incepatori. Utilizatorii tinta sunt copii cu varsta peste 6 ani precum si adulti care doresc sa isi imbunatateasca cunostintele sau sa invete o limba straina.

Architecture (schema):



Project's details (max 1000 words), including:

- **Motivation:**

Ideea proiectului a plecat de la dorinta de a ne familiariza cu libraria wordnet si de a pune bazele unui proiect e-learning usor de construit si imbunatatit. Prima etapa din procesul de invatare a unei limbi straine este are la baza cuvinte uzuale ale limbii si se axeaza pe asimilarea acestora cu cuvinte din limba materna sau cu cuvinte inrudite ale acestora. Astfel ca, aplicatia de fata utilizeaza functionalitatile oferite de libraria wordnet(sinonime, antonime, definitii, cuvinte din acelasi domeniu) si construiește un set de intrebari care sa antreneze utilizatorii aplicatiei in prima etapa de invatare a unei limbi straine.

- **Challenges:**

Etapele de analiza si planificare au reprezentat o adevarata provocare deoarece au necesitat luarea in calcul a multor variabile si probleme care pot aparea pe parcursul dezvoltarii. Alegerea utilizatorilor tinta care sa se plieze pe tipul intrebarilor generate precum si analiza pietei aplicatilor e-learning de tip gaming au fost unele dintre cele mai provocatoare task-uri.

O alta etapa in cadrul careia am intampinat o serie de provocari a fost gasirea unei librarii care sa ne ofere traducerea cuvintelor utilizate din wordnet. Multe dintre cele testate au avut o serie de limitari precum numarul de request-uri. Faptul ca in cadrul proiectului au fost extrase pentru a fi folosite in jur de 10000 de cuvinte ne-a impiedicat din a utiliza unele librarii mult mai cunoscute precum **google.translate**

- **State of the art:**

Nenumarate studii ne arata faptul ca invatarea este favorizata de sistemele e-learning de tip gaming. Colaborarea cu institutiile statului in dezvoltarea unor astfel de sisteme de invatare poate aduce beneficii atat sistemului educational cat si celor care vor lucra la dezvoltarea acestora. Exista nenumarate aplicatii de tip e-learning insa consideram ca introducerea lor in scoli ar fi un pas important in dezvoltarea sistemului educational in Romania.

Cumulus a fost vazuta pe parcursul dezvoltarii ca o aplicatie ce va fi utilizata in scoli. Orele de limbi straine se invart de multe ori in jurul manualelor inechite ce contin informatii neactualizate. Am intentionat ca aplicatia dezvoltata de noi sa poata fi folosita in scoli pentru a ajuta la procesul de invatare a unei limbi straine.

- ***Short description of each module in the architecture, including methods to be used, programming language, additional resources (if any), input/output formats, etc.***

Aplicatia a fost impartita in doua module back-end si front-end. Acestea comunica intre ele cu ajutorul fisierelor json generate de modulul de back-end care contin intrebari ce vor fi oferite utilizatorilor cu ajutorul modulului de front-end.

Modulul de back-end este responsabil cu generarea de fisier json care sa contina diferite tipuri de intrebari(cu variante de raspuns sau input). Pentru generarea intrebărilor a fost utilizat python-ul ca limbaj de programare si diferite librării. Libraria **Wordnet** fost utilizata ca librerie de baza, aceasta ne-a oferit o mare parte din informatiile necesare generării întrebărilor iar **yandex.Translater** a fost utilizat pentru obtinerea traducerii cuvintelor extrase din corpus-ul wordnet. Pe langa aceste librării, in generarea fisierelor json a mai fost utilizate librării precum **json** si **random**.

Modulul de front-end este responsabil cu consumarea informatiilor oferite de modulul de back-end si de a le oferi utilizatorilor intr-un format atragator. Modulul are la baza o aplicatie web construita cu ajutorul framework-ului Angular. Acest modul ofera posibilitatea de login si inregistrare a utilizatorilor precum si un mediu placut de invatare a unei limbi straine. Utilizatorul va putea alege o categorie si răsunde la întrebările asiguate acelei categorii.

- ***Description of the integration of different modules: problems foreseen, if any (incompatibility of programming languages, input/output format changes, etc.).***

Dupa cum am specificat in sectiunea anterioara, integrarea modulelor s-a dorit a fi cat mai simpla si usor de adaptat nevoilor utilizatorilor. Astfel ca am decis ca legatura dintre cele doua module sa se faca cu ajutorul fisierelor json.

Fisierele json ce contin întrebările generate in modulul de back-end vor fi consumate de modulul de front-end. Intentionam ca întrebările generate automat sa poata fi verificate si corectate de o persoana specializata(professor), astfel la utilizatorii finali sa ajunga informatii corecte.

Imbunatirea calitatii întrebărilor generate reprezinta un factor foarte important in dezvoltarea aplicatiei. Acest lucru poate aduce dupa sine necesitatea schimbării modului in care modulul de back-end va genera întrebările, posibil schimbarea limbajului de programare sau al librăriilor utilizate. Utilizarea fisierelor json ca metoda de comunicare intre module va ajuta in cazul in care modulele vor suferi modificari, acestea fiind usor de utilizat in transferul de informatii intre module.

Project management (no. max. words no. imposed):

- ***Assignment of team members for each different module (one team member per module)***

Back-end: Cretu Andra

Front-end: Scortanu Raluca

- ***Will the project have a team leader? If yes, who is it? What are its main duties?***

In etapele de analiza si planificare fiecare a venit cu idei care au fost puse cap la cap iar aplicatia de fata este rezultatul colaborarii noastre.

In cadrul etapei de dezvoltare nu s-a simtit nevoia unui lider. Fiecare dintre noi a lucrat in cadrul modulului cu care este familiarizat, ceea ce a ajutat in oraganizare. Pe parcursul dezvoltarii, fiecare dintre noi si-a adus input-ul in cadrul celui alt modul. Fiecare dintre noi a solicitat feedback in diferite etape ale dezvoltarii.

- ***Detailed deadlines and milestones (for each subtask of each module, but also for the inclusion of all modules in the final system). Deadlines are actual dates.***

La dezvoltarea modulelor s-a lucrat in paralel, acestea fiind independente.

Back-end:

- extragerea informatiilor necesare cu ajutorul libreriei **wordnet**
- gasirea unei librarii pentru obtinerea traducerii cuvintelor utilizate pentru construirea intrebarilor
- utilizarea libreriei **yandex.Translater** pentru a obtine traducerea cuvintelor
- generarea diferitelor tipuri de intrebari
- construirea fisierelor json

Front-end:

- formarea modulelor specifice unei aplicatii (login/register flow, selectarea categoriilor, crearea diferitelor tipuri de intrebari)
- crearea componentelor si a rutelor specifice fiecari parti
- alegerea designului
- includerea fisierului json obtinut in modulul de back-end

- ***Qualitative measures for the evaluation of the project (e.g. no. of annotations, no. of sentences processed, etc.). These measures will be used to assess the projects.***

In cadrul proiectului au fost utilizate din biblioteca wordnet in jur de 10000 de cuvinte(cele mai utilizate cuvinte din limba engleza). Acesta au fost impartite in 6 domenii (animale, fructi, legume, sporturi, ocupatii, culori). Din cuvintele extrase initial au fost alese doar cele din domeniile de mai sus, pentru acestea au fost extrase definitia, sinonimele si antonimele precum si traducerea in limba romana. In fisierul **list.json** se regasesc toate cuvintele utilizate precum si toate informatiile despre acestea. Intrebarile utilizate in aceasta etapa a proiectului au fost construite cu o parte din cuvinte, acestea necesitand verificarea intrebarilor generate pentru a putea fi cat mai corecte. Astfel ca in fisierul final **questions.json** se regasesc 240 de intrebari.