**Modulul 3 – Baza de cunostinte**

**Membri**: Patras Ionela, Martac Monica, Nastasiu Andrei, Savoaia Madalin, Petcu Raluca, Suciu Ana

Acest modul – avand in componenta sa 6 persoane – are ca si scop principal construirea Bazei de cunostinte care va fi utilizata de catre chatbot in conversatiile purtate cu utilizatorii.

In urma discutiilor cu modulul ce se ocupa cu implementarea efectiva a chatbot-ului am decis de comun accord ca datele sa fie stocate in format AIML – format de date derivate din XML , utilizat in mod frecvent la ora actuala in construirea chat-boturilor. ( spre exemplu:, ChaBot-ul ALICE utilizeaza AIML-uri pentru baza de cunostinte) .

Ca punct de plecare pentru realizarea bazei de cunostinte am avut colectia de fisiere oferita de catre dezvoltatorii proiectului ALICE.

De asemenea, am luat in considerare si posibilitatea folosirii altor alternative pentru imbogatirea cunostintelor& abilitatllor bot-ului, precum utilizarea unei ontologii , preluarea de date de pe retelele sociale (Twiter) sau interogarea Wikipedia, insa , din motive bine intemeiate am ajuns la concluzia ca utilizarea exclusiva a fisierelor AIML va conduce la rezultate mult mai bune din toate punctele de vedere. (Filtrarea datelor obtinute de pe retelele sociale ar fi un procedeu anevoios- care nu ar garanta obtinerea unor raspunsuri adecvate si nici nu ar putea conferi bot-ului o personalitate bine conturata. Cat despre utilizarea datelor disponibile pe Wikipedia, aceasta abordare ar face ca dialogul cu chatbot-ul sa para nenatural – robotelul ar sti mult prea multe lucruri). Tot din acest motiv, am renuntat sa folosim si o serie de fisiere AIML continand o multitudine de informatii din diferite domenii , preferand in schimb sa ne axam pe o singura directie, si anume Tehnologia. .

Per ansamblu**, activitatea modului 3** consta in :

Documentarea asupra domeniului , prelucrarea resurselor gasite( filtrarea intrebarilor nepotrivite, adaugarea raspunsurilor acolo unde lipseau/erau nepotrivite in cadrul documentelor AIML luate de la ALICE), crearea de noi fisiere AIML pentru a specializa chatbot-ul in anumite domenii, dialogarea cu modulul 1, cu care am stabilit atat formatul de fisiere folosit, cat si domeniul de specializare al bot-ului si diverse detalii de structurare ale datelor.

Din punctul de vedere **al distribuirii sarcinilor pe saptamani/membrii echipa**, activitatea desfasurata de modulul 3 este urmatoarea:

**Saptamana 1:**

**all:**

**-** documentare, cautare informatii(baza date AIML, modele de baze de date utilizate de catre alte chatbot-uri), impartire taskuri, stabilire format cu modul 1

- parcurgere a fisierelor din baza de date preluata de la ALICE( + filtrarea intrebarilor nepotrivite, adaugare raspunsuri acolo unde lipseau/erau nepotrivite)

**Ionela:**

* Java.aiml

**Madalin:**

* descarcat&convertit fisier 3gb cu date de pe retele sociale

**Monica:**

* program python care gaseste informatii pe wikipedia( modul wikipedia python)

**Andrei&Madalin:**

* documentare asupra folosirii api-ului Twitter pentru a extrage fragmente de conversatii.

**Saptamana 2:**

**Monica&Ionela :**

**-** analiza continutului bazei de cunostintei & realizarea unei scheme a acesteia ( informatii succinte asupra continutului fiecarui fisier aiml)

- adaugarea unor noi fisiere aiml (utile pentru: convertirea numerelor de la 0-255 in binar, calculul varstei unei persoane, afisarea datelor calendaristice, cunoasterea monedelor utilizate in diferite tari, afisarea de previziuni (Horoscop), curiozitati matematice, afisarea evenimentelor importante care au avut loc intr-o anumita zi a anului in trecut.)

**Ana:**

* medicine.aiml (specializare)

**Andrei:**

* apps.aiml

**Raluca:**

* animals.aiml

**Saptamana 3:**

**Monica&Ionela:**

* modificare fisiere din baza de cunostinte (referitoare la ASKRANDOMQUESTION si ConfusionAnswers)

**Ana:**

* python.aiml

**Raluca:**

* computerNetworks.aiml

**Madalin:**

* programming.aiml

**Ionela:**

* phones.aiml

**Andrei :**

* softwares.aiml

**Monica:**

* android.aiml

**Saptamana 4:**

**All:**

-completarea bazei de cunostinte cu informatii( in fisierele std-pickup.aiml, technology.aiml, technologyResponses.aiml, std-profile.aiml)