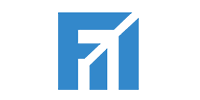
UNIVERSITATEA “ALEXANDRU IOAN CUZA” DIN IAȘI

**FACULTATEA DE INFORMATICĂ**

****

LUCRARE DE LICENȚĂ

**Învățare automată: chatbot-ul *Euterpe***

propusă de

**Alexandru Voaideș**

**Sesiunea:** *iulie, 2018*

**Coordonator științific**

**Colaborator Florin Olariu**

**UNIVERSITATEA ”ALEXANDRU IOAN CUZA” DIN IAȘI**

**FACULTATEA DE INFORMATICĂ**

**Învățare automată: chatbot-ul *Euterpe***

**Alexandru Voaideș**

**Sesiunea:** *iulie, 2018*

Coordonator științific

**Colaborator Florin Olariu**

Prin prezenta declar că Lucrarea de licenţă cu titlul „*Învățare automată: chatbot-ul Euterpe*” este scrisă de mine şi nu a mai fost prezentată niciodată la o altă facultate sau instituţie de învăţământ superior din ţară sau din străinătate. De asemenea, declar că toate sursele utilizate, inclusiv cele preluate de pe Internet, sunt indicate în lucrare, cu respectarea regulilor de evitare a plagiatului:

- toate fragmentele de text reproduse exact, chiar şi în traducere proprie din altă limbă, sunt scrise între ghilimele şi deţin referinţa precisă a sursei;

- reformularea în cuvinte proprii a textelor scrise de către alţi autori deţine referinţa precisă;

- codul sursă, imaginile etc. preluate din proiecte *open*-*source* sau alte surse sunt utilizate cu respectarea drepturilor de autor şi deţin referinţe precise;

- rezumarea ideilor altor autori precizează referinţa precisă la textul original.

Iași, *3 iulie 2018*

Absolvent *Alexandru Voaideș*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

(semnătura în original)

Prin prezenta declar că sunt de acord ca Lucrarea de licență cu titlul „*Învățare automată: chatbot-ul Euterpe*”, codul sursă al programelor şi celelalte conţinuturi (grafice, multimedia, date de test etc.) care însoţesc această lucrare să fie utilizate în cadrul Facultăţii de Informatică.

De asemenea, sunt de acord ca Facultatea de Informatică de la Universitatea „Alexandru Ioan Cuza” din Iași, să utilizeze, modifice, reproducă şi să distribuie în scopuri necomerciale programele-calculator, format executabil şi sursă, realizate de mine în cadrul prezentei lucrări de licenţă.

Iași, *3 iulie 2018*

Absolvent *Alexandru Voaideș*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(semnătura în original)

**Cuprins**

[Introducere 2](#_Toc517614347)

[**Capitolul 1.** Platforma de dezvoltare Android 3](#_Toc517614348)

[**Capitolul 2.** Tehnologii folosite 3](#_Toc517614350)

[**Capitolul 3.** Aplicația 3](#_Toc517614352)

[**Capitolul 4.** Componenta de inteligență artificială 3](#_Toc517614354)

[Bibliografie 3](#_Toc517614356)

# Introducere

La momentul actual, domeniul învățării automate se bucură de o mare popularitate și utilitate. Acesta a fost principalul motiv al alegerii temei de față. Un alt motiv a fost dorința de a retrezi interesul oamenilor pentru poezie, aflată astăzi într-un con de umbră. Pentru aplicație, am decis ca sistemul de operare să fie Android, având în vedere că oamenii se folosesc mai des de un smartphone pentru rezolvarea problemelor decât de un calculator personal. Această temă mi-a oferit și ocazia de a aprofunda domeniul învățării automate.

Lucrarea de față constituie documentația asupra unei aplicații pentru Android, bazată pe un chatbot românesc care vorbește în versuri. Există zeci de mii de chatbot-uri în lume, în aproximativ 60 de limbi (două în limba română) și pe o arie largă de specializări: de la îndrumător pentru gătit („Dinner Ideas”), până la specialist în tranzacții bancare („Neomy”) (Chatbots.org, 2018). Luând în seamă aceste statistici, aplicația nu reprezintă un lucru inedit. Însă, având în vedere numărul scăzut de chatbot-uri în limba română și abilitatea de a vorbi versificat, aceasta poate fi considerată o noutate.

Obiectivele principale ale lucrării sunt:

* Valorificarea oportunităților oferite de dezvoltarea domeniului învățării automate
* Cultivarea interesului pentru poezie
* Crearea unui instrument inteligent de divertisment

În realizarea chatbot-ului, am recurs la o serie de reprezentări a cuvintelor prin vectori de probabilități precalculați cu un algoritm Word2Vec. Algoritmul a fost aplicat pe corpusul CoRoLa, o mulțime de texte dintre care unele datează din 1989, „scopul creării acestuia fiind să ofere o imagine obiectivă a limbii române actuale scrise și vorbite” (Tufiș, et al., 2017). Numărul vectorilor preluați este de 250255, fiecare având dimensiunea 300. Am calculat cu datele din vectori utilizând similaritatea cosinus, pentru fiecare cuvânt, cele mai similare 10 cuvinte cu acesta, care au fost folosite ulterior la algoritmul de generare a versurilor.

Chatbot-ul este definit ca „an artificial living entity designed to have conversations with real human beings. This could encompass a text conversation via computers, a spoken conversation or even a non-verbal conversation” (Chatbots.org, 2018). Numele Euterpe dat chatbot-ului a fost preluat din mitologia greacă, fiind una dintre cele nouă fiice ale lui Zeus și considerată muza muzicii și a poeziei lirice.

În primul capitol se face o statistică asupra numărului de aplicații și al veniturilor dezvoltatorilor din acestea, o scurtă descriere a mediului de dezvoltare folosit, Android Studio, precum și o prezentare succintă a câtorva dintre cele mai populare chatbot-uri pe Android existente.

Capitolul doi descrie modalitățile prin care se poate realiza un chatbot și surprinde tehnologiile folosite pentru atingerea obiectivelor acestei lucrări.

Capitolul trei cuprinde o descriere completă asupra interfeței grafice și a funcționalităților aplicației.

Capitolul patru vorbește despre „creierul” chatbot-ului și algoritmul de generare a versurilor care se folosește de cunoștințele acestuia.

# Capitolul 1

## Platforma de dezvoltare Android

Aplicațiile se bucură la momentul actual de o popularitate extraordinară în rândul oamenilor, iar aceasta mulțumită facilităților și divertismentului pe care le oferă. Conform statisticilor, există 3.8 milioane de aplicații pe Google Play, 2 milioane pe Apple App Store, 669 de mii pe Windows Store, 430 de mii pe Amazon Appstore și 234 500 pe BlackBerry World (Statista, 2018).

Dezvoltatorii au profitat de nivelul de utilizare al aplicațiilor, care au devenit pentru ei o sursă de venit generoasă (YouTube, Facebook Messenger, Instagram ș.a.). Un dezvoltator Android câștigă anual 97.6 de mii de dolari, iar un dezvoltator mid-level iOS, nu cu o diferență prea mare, 96.6 mii de dolari (Charuza, 2018).

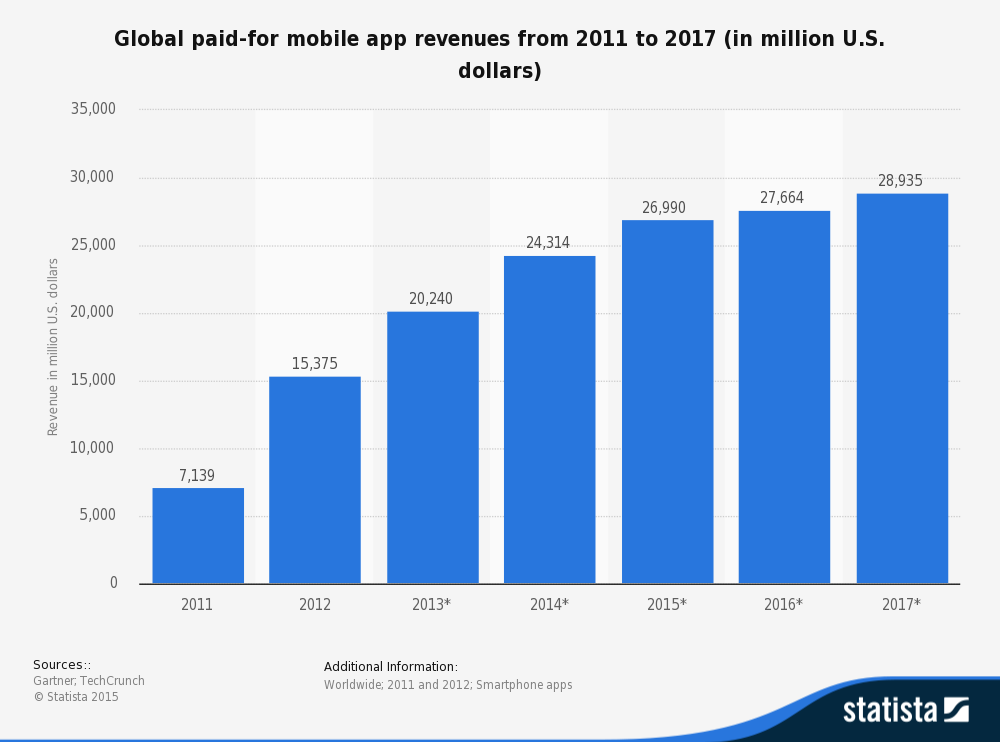


Fig.1 Creșterea veniturilor din aplicații la nivel global din anul 2011 până în 2017

În iunie 2014, Google lansează Android Studio, un IDE gratuit pentru dezvoltarea de aplicații pe sistemul de operare Android. IDE-ul vine la pachet cu un SDK și un debugger și prezintă următoarele instrumente:

* Un sistem de dezvoltare flexibil bazat pe Gradle;
* Un emulator rapid cu multe opțiuni;
* Un mediu de dezvoltare unic pentru toate dispozitivele cu Android;
* Instant Run ce permite adăugarea noilor modificări aduse aplicației fără necesitatea instalării unui nou APK;
* Șabloane de cod și integrare cu GitHub pentru a facilita construirea de caracteristici comune ale aplicațiilor și importarea codului simplu;
* Instrumente de testare extensive și framework-uri;
* Instrumente Lint pentru detecția performanței, uzabilității, versiunii compatibile și a altor probleme;
* Suport pentru C++ și NDK;
* Suport încorporat pentru Google Cloud Platform, făcând ușoară integrarea pentru Google Cloud Messaging și App Engine;

(AndroidDevelopers, 2014)

## Chatbot-uri pe Android

Așa cum s-a menționat în introducere, numărul chatbot-urilor a crescut foarte mult odată cu dezvoltarea domeniului inteligenței artificiale și al învățării automate, având o aplicabilitate foarte diversă: legislație, finanțe, sport, educație etc. Aceștia se specializează pe oferirea unei experiențe conversaționale autentice, în care interlocutorul uman nu ar trebui să poată afirma dacă poartă o discuție cu un om sau cu o mașină. Despre un astfel de chatbot se spune că a trecut testul Turing. Există companii care folosesc chatbot-uri pentru a furniza servicii mai bune clienților, dintre care amintesc 5.

### Endurance: A Companion for Dementia Patients

Dificultatea cu care se luptă pacienții cu pierderilere de memorie i-a determinat pe cei de la compania rusească Endurance să creeze un chatbot care să ajute pacienții suferinzi de astfel de probleme. Obiectivul acestuia este de a identifica deviațiile din ramurile conversaționale ce indică o problemă cu imediata amintire. În plus, chatbot-ul fiind o soluție bazată pe cloud, membri familiei pot revizui conversațiile pentru a depista potențialele degradări ale funcției de memorare și obstacolele de comunicare care pot semnala deteriorarea condiției pacientului. (Shewan, 2018)

### Casper: Helping Insomniacs Get Through the Night

În cazul în care suferi de insomnie și nu știi cum să treci peste singurătatea sufocantă, o soluție este să apelezi la Insomnobot 3000, un agent conversațional care își propune să-ți prelungească insomnia în timp ce restul lumii doarme. (Shewan, 2018)

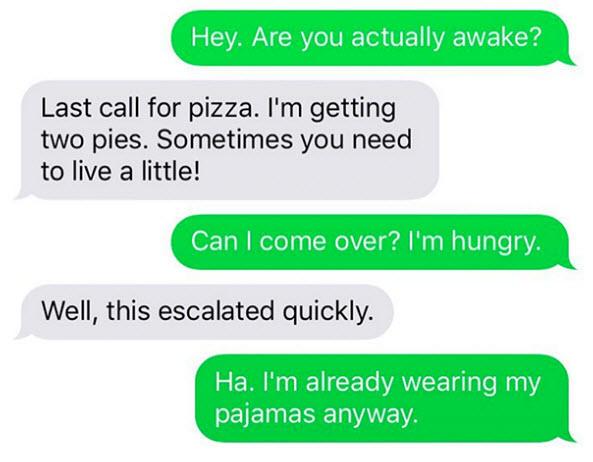


Fig.2 Un exemplu de conversație între un interlocutor uman și Insomnobot 3000

### MedWhat: Making Medical Diagnoses Faster

MedWhat are ca obiectiv să facă diagnostice medicale mai repede, mai ușor și mai precis, atât pentru pacienți, cât și pentru medici. Chatbot-ul a fost învățat prin parcurgerea a vaste volume de cercetare în medicină și lucrări științifice revizuite. (Shewan, 2018)

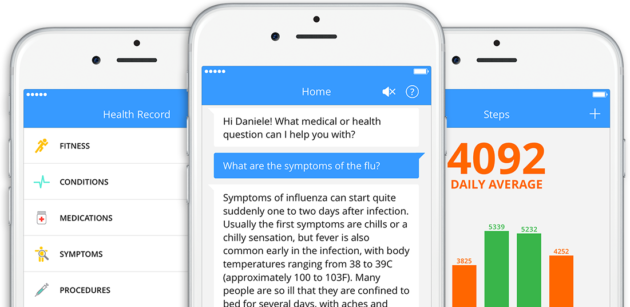


Fig.3 MedWhat fiind întrebat de simptomele gripei

### UNICEF: Helping Marginalized Communities Be Heard

UNICEF folosește chatbot-uri pentru a ajuta oamenii care locuiesc în țări în curs de dezvoltare să vorbească despre nevoile urgente cu care se confruntă comunitățile lor. Bot-ul U-Report se concentrează pe colectarea de informații de mari dimensiuni prin sondaje, deci nu este folosit pentru a purta conversații. U-Report trimite regulat sondaje pregătite pe o serie de probleme sociale urgente, iar utilizatorii răspund. UNICEF folosește apoi acest feedback ca bază pentru posibile recomandări politice. (Shewan, 2018)

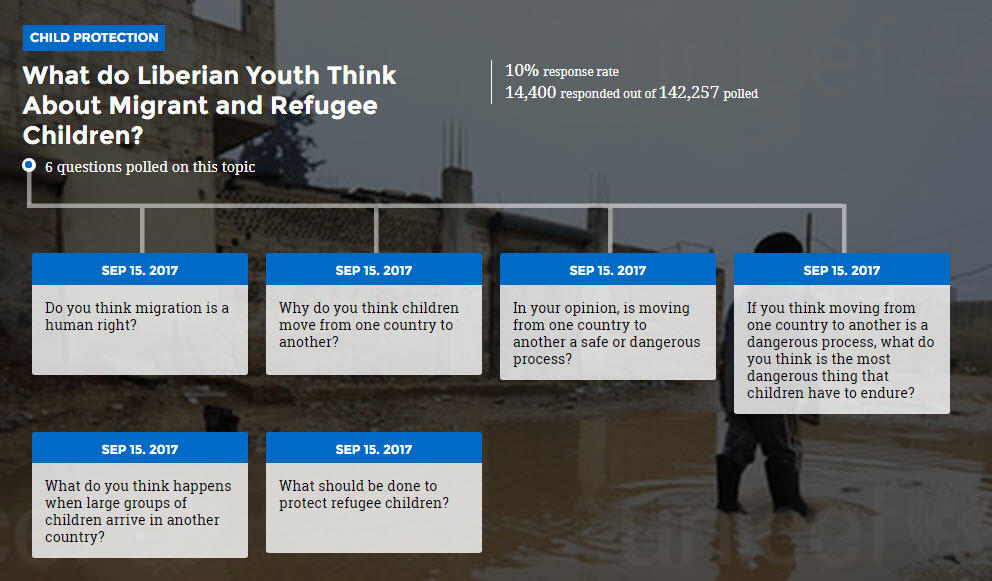


Fig.4 Exemplu de sondaj trimis de bot

### NBC: Helping Newshounds Navigate the Headlines

NBC Politics Bot permite utilizatorilor să interacționeze cu agentul conversațional prin Facebook pentru a identifica subiectele de știri care ar putea fi de interes pentru ei. După prima interacțiune, bot-ul furnizează utilizatorilor rezultate personalizate de știri (prioritizând materialele video) bazate pe preferințele lor. (Shewan, 2018)

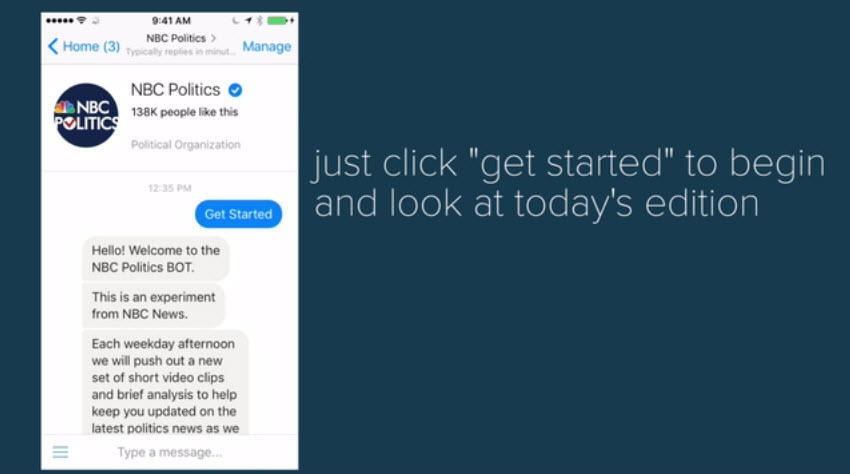


Fig.5 NBC Politics Bot la prima interacțiune

## 1.2 Obiectiv

Luând în considerare ideea de bază a unui chatbot, lucrarea de față are ca obiectiv oferirea unei experiențe conversaționale cât mai plăcute. În crearea acestuia, pe lângă vocabularul său, cea mai importantă componentă a sa va fi metoda de generare a versurilor. Algoritmul se va baza pe regulile fundamentale ale gramaticii limbii române și pe păstrarea particularităților genului liric care se va face printr-un algoritm de silabisire.

## 1.3 Concluzii

În ultimii ani, domeniul inteligenței artificiale s-a dezvoltat mult, ceea ce a dat naștere mai multor idei de soluții prin care să fie sprijinit omul (Endurance chatbot, MedWhat, UNICEF chatbot, NBC Politics Bot) sau soluții care să constituie o sursă de divertisment (Insomnobot 3000, ALICE). Datorită popularității de care se bucură sistemul de operare Android în prezent, Google a oferit dezvoltatorilor de aplicații un IDE gratuit, intuitiv și bogat în unelte, Android Studio, fapt ce m-a determinat să-l folosesc pentru construirea aplicației la baza căreia stă chatbot-ul Euterpe.

# Capitolul 2

## Tehnologii folosite

Un chatbot este o entitate artificială menită să poarte o conversație exact ca o ființă umană, fapt sugerat și de numele pe care îl poartă. Există două abordări pentru dezvoltarea de chatbot-uri: predefinirea unui set de întrebări și răspunsuri (complexitate redusă) sau prin tehnici de inteligență artificială. Ca utilizator, la prima vedere, este foarte posibil să nu recunoști metoda folosită. Totuși, dacă se întâmplă să repetați o întrebare, veți descoperi imediat cum a fost implementat chatbot-ul. Cel mai bun mod de a evalua un chatbot este prin testul Turing. Dacă omul nu poate să afirme sigur dacă discută cu un alt om sau cu o mașină, atunci bot-ul a trecut testul.

Există mai multe moduri prin care se poate realiza un chatbot bazat doar pe recunoașterea întrebărilor. Una la care probabil v-ați gândit deja e printr-o mulțime de instrucțiuni if-else. Un alt mod este prin AIML (Artificial Intelligence Markup Language), un limbaj extins din XML.

|  |
| --- |
| <category>  <pattern>WHAT IS YOUR NAME</pattern>  <template>My name is Euterpe.</template>  </category> |

Tabela 1: Exemplu de cod în AIML

A doua metodă de implementare a unui chatbot se poate face utilizând algoritmi de clasificare ce sunt aplicați pe sute sau mii de exemple și va ști să aleagă răspunsul potrivit. Astfel de algoritmi pot fi construiți folosind o librărie de învățare automată precum scikit-learn. O altă opțiune e folosirea unui cloud API: wit.ai, api.ai, Microsoft LUIS (Surmenok, 2016).

La dezvoltarea chatbot-ului, singura tehnologie ce a fost folosită este sistemul de operare Android.

## 2.1 Sistemul de operare Android

Android este o platformă software și un sistem de operare pentru dispozitive și telefoane mobile bazată pe nucleul Linux, dezvoltată inițial de compania Google, iar mai târziu de consorțiul comercial Open Handset Alliance. Android permite dezvoltatorilor să scrie cod gestionat în limbajul Java, controlând dispozitivul prin intermediul bibliotecilor Java dezvoltate de Google. Aplicațiile scrise în C și în alte limbaje pot fi compilate în cod mașină ARM și executate, dar acest model de dezvoltare nu este sprijinit oficial de către Google.

Lansarea platformei Android la 5 noiembrie 2007 a fost anunțată prin fondarea Open Handset Alliance, un consorțiu de 48 de companii de hardware, software și de telecomunicații, consacrat dezvoltării de standarde deschise pentru dispozitive mobile. Google a lansat cea mai mare parte a codului Android sub licența Apache, o licență de tip free-software și open source (Wikipedia, 2018).

Cel mai bun lucru la Android sunt multele opțiuni. Cu Android aveți sute de gadget-uri la dispoziția dumneavoastră, cea mai ieftină, cea mai scumpă și inovatoare piață. Android este, de asemenea, incredibil de personalizabil, atât în rolurile sale, cât și în aspectul său. Puteți realiza cu adevărat o experiență mobilă unică pentru dumneavoastră în acest sistem de operare (Schmidt, 2016).

Ca sistem de operare, munca Android-ului este de a acționa ca un traducător între dumneavoastră și gadget-ul dumneavoastră. Când faceți o fotografie, Android vă oferă butonul pe care îl atingeți și îi spune telefonului ce trebuie să facă atunci când îl atingeți. Când efectuați sau primiți un apel, Android vă spune telefonului cum să facă acest lucru. Când jucați un joc, Android spune jocului ce mișcări faceți și ce butoane apăsați. Este ca și Windows, dar pentru dispozitive mobile (Schmidt, 2016).

Fiecare nouă versiune de Android primește un nume de cod bazat pe literele consecutive ale alfabetului. Cea mai recentă versiune este cunoscută sub numele de Oreo, lansată în decembrie 2017.

# Capitolul 3

## Aplicația

După cum a fost menționat, aplicația a fost dezvoltată cu IDE-ul Android Studio datorită mulțimii de posibilități pe care le oferă. Versiunea minimă de Android compatibilă este 7.1.1 (Nougat). Testarea acesteia a fost făcută pe emulatoare de Android cu Nougat, de diferite rezoluții, precum și pe smartphone-ul personal.

## 3.1 Interfața grafică

Construcția interfeței s-a realizat, pentru fiecare activitate, folosind layout-uri de tip Constraint pentru a face aplicația responsive la diferite dimensiuni de ecran. De asemenea, un alt aspect ce oferă calitatea de responsive aplicației este reprezentat de folosirea pixelilor independenți de densitate.

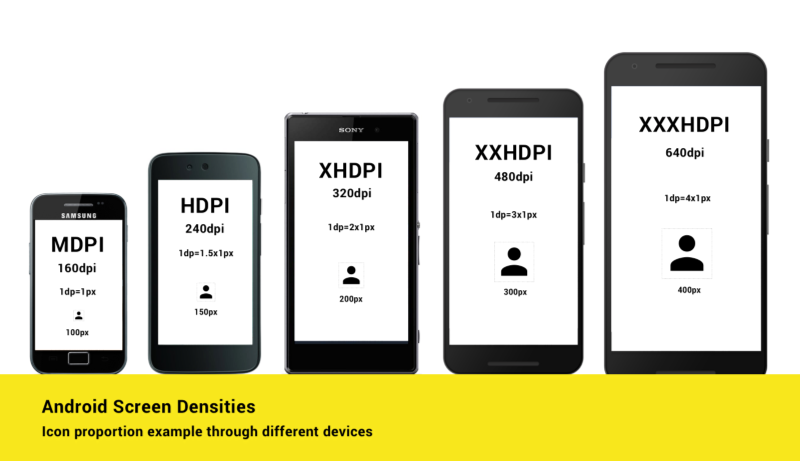


Fig.6 Densitatea pixelilor pe ecrane cu dimensiuni diferite

Interfața este alcătuită din 7 activități: ConversationActivity, MenuActivity, ModBacoviaActivity, ModEminescuActivity, ModsActivity, ModStanescuActivity și SavedVersesActivity. Fiecare activitate prezintă două layout-uri: unul pentru orientare portrait și unul pentru orientare landscape. În cadrul activităților ConversationActivity și SavedVersesActivity a mai fost folosit layout-ul de tip Relative pentru definirea aspectului unui element din obiectele ListView.

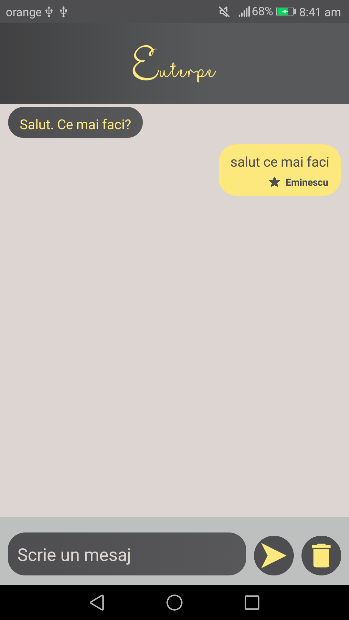
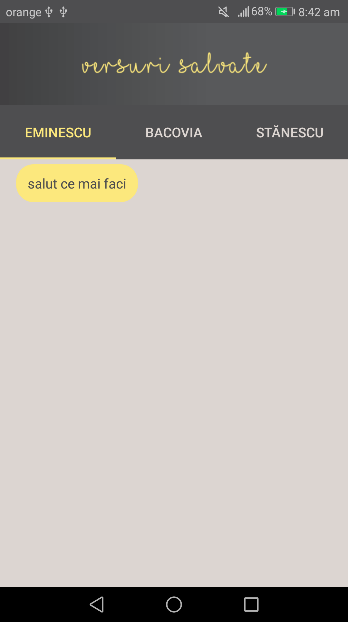
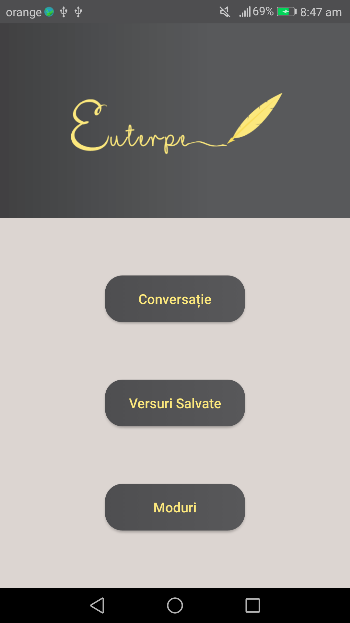
  

Fig.7 ConversationActivity, SavedVersesActivity, MenuActivity

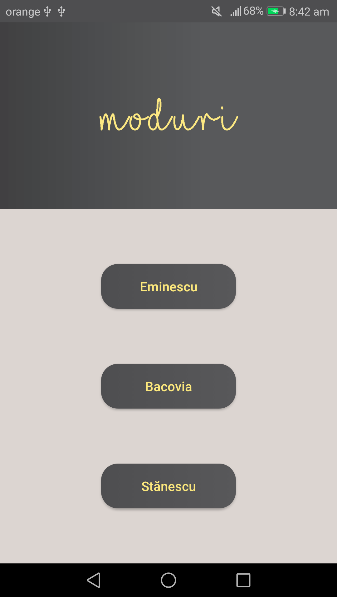
   

Fig.8 Mods Activity, ModEminescuActivity, ModBacoviaActivity, ModStanescuActivity

## 3.2 Codul

Codul operează cu 6 fișiere folosite pentru salvarea modului activ, a mesajelor din conversație, a versurilor (câte unul pentru fiecare mod) și a tab-ului selectat în SavedVersesActivity.

### MenuActivity

Clasa MenuActivity prezintă trei obiecte de tip Button ce corespund butoanelor Conversație, Versuri Salvate și Moduri din interfață. Butoanelor li se atribuie aceeași funcționalitate de apăsare: se inversează culorile și pornește activitatea corespunzătoare la apăsare, iar la ridicarea degetului culorile revin la starea inițială.

### ConversationActivity

btn.setTextColor(Color.parseColor("#ff4e4e50"));

btn.setBackground(background\_pressed);

CharSequence buttonText = btn.getText();

if(buttonText.equals(<text\_buton>))

{

startActivity(new Intent(MenuActivity.this, <activitate>.class));

}

O instanță a acestei activități conține un EditText care permite introducerea textului de la tastatură, un număr întreg pentru reținerea id-ul mesajului următor, trei ArrayList-uri pentru salvarea versurilor pentru fiecare mod, trei variabile de tip long pentru detectarea unui eventual dublu-click pe un set de versuri, un adapter personalizat pentru memorarea mesajelor și o constantă de tip char pentru marcajul de final al mesajului.

Funcționalitatea funcției principale (onCreate) se rezumă la preluarea informațiilor (versuri, conversație și mod activ) din fișiere și setarea funcționalităților butoanelor. Butonul de trimitere scrie în fișier mesajul utilizatorului și al chatbot-ului pe care le adaugă ulterior în adapter. Pentru evitarea erorilor, am stabilit un format pentru mesaje atunci când sunt adăugate în fișier:

<id>\_<modActiv>\_<esteSalvat>\_<numarMarcajeFinalInMesaj>\_<corpMesaj><marcaj>.

Butonul de ștergere a conversației elimină toate elementele din adapter, apelează funcția notifyDataSetChanged a adapter-ului pentru actualizarea ListView-ului și golește fișierul cu conversația.

Fiecare element din ListView prezintă aceeași funcționalitate la dublu-click: obiectul de tip ChatMessage setează isSaved pe true, adapter-ul este înștiințat de această modificare și actualizează lista de mesaje, scrie versurile și le adaugă în fișierul, respectiv vectorul corespunzător.

Adapter-ul funcționează pe principiul următor: dacă mesajul se află pe poziție pară, atunci aparține utilizatorului și îi setez parametri corespunzători, iar dacă se află pe poziție impară, atunci aparține chatbot-ului.

La fiecare mesaj de chatbot adăugat în conversație, ListView-ul selectează ultimul său element.

### ModsActivity

Este identică la funcționalitate cu MenuActivity.

### Cele trei activități pentru moduri

ModEminescuActivity, ModBacoviaActivity și ModStănescuActivity sunt activități cu un singur buton care activează modul respectiv.

### SavedVersesActivity

Conține un TabHost cu trei taburi în care se regăsesc trei liste (câte una pentru fiecare mod) cu versurile salvate. Eliminarea unui set de versuri se face tot cu dublu-click.

### Alte clase

Mai sunt de menționat clasa Euterpe (extinde clasa Application) a cărei instanță creează în memoria internă cele 6 fișiere menționate mai sus și le inițializează și clasa Bot în care se păstrează toată logica generării versurilor.

# Capitolul 4

## Componenta de inteligență artificială

# Bibliografie

AndroidDevelopers, 2014. *Android Developers: Intro.* [Interactiv]   
Available at: https://developer.android.com/studio/intro/  
[Accesat 30 May 2018].

Charuza, P., 2018. *Much money can earn app.* [Interactiv]   
Available at: https://fueled.com/blog/much-money-can-earn-app/  
[Accesat 29 May 2018].

Chatbots.org, 2018. *About Us: Chatbots.org - Virtual assistants, virtual agents, chat bots.* [Interactiv]   
Available at: https://www.chatbots.org/community/static/about\_us  
[Accesat 1 5 2018].

Chatbots.org, 2018. *Chatbots.org - Virtual assistants, virtual agents, chat bots.* [Online]   
Available at: https://www.chatbots.org  
[Accessed 1 5 2018].

Gibson, A., Nicholson, C. & Patterson, J., 2017. *Deeplearning4j: Open-source, Distributed Deep Learning for the JVM.* [Interactiv]   
Available at: https://deeplearning4j.org/overview  
[Accesat 20 5 2018].

Schmidt, C., 2016. *What is Android: AndroidPIT.* [Interactiv]   
Available at: https://www.androidpit.com/what-is-android  
[Accesat 31 May 2018].

Shewan, D., 2018. *ChatBots: Wordstream.* [Interactiv]   
Available at: https://www.wordstream.com/blog/ws/2017/10/04/chatbots  
[Accesat 30 May 2018].

Statista, 2018. *Number of apps available in leading app stores: Statista.* [Interactiv]   
Available at: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/  
[Accesat 29 May 2018].

Surmenok, P., 2016. *Chatbot architecture: Medium.* [Interactiv]   
Available at: https://medium.com/@surmenok/chatbot-architecture-496f5bf820ed  
[Accesat 31 May 2018].

Tufiș, D. și alții, 2017. *CoRoLa.* [Interactiv]   
Available at: http://corola.racai.ro/  
[Accesat 29 May 2018].

Wikipedia, 2018. *Android (sistem de operare).* [Interactiv]   
Available at: https://ro.wikipedia.org/wiki/Android\_(sistem\_de\_operare)  
[Accesat 31 May 2018].