

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Ergim Ramadan

Preprost prevajalnik besedil za platformo android

DIPLOMSKO DELO

VISOKOŠOLSKI STROKOVNI ŠTUDIJSKI PROGRAM PRVE
STOPNJE RAČUNALNIŠTVO IN INFORMATIKA

Ljubljana, 2014

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Ergim Ramadan

Preprost prevajalnik besedil za platformo android

DIPLOMSKO DELO

VISOKOŠOLSKI STROKOVNI ŠTUDIJSKI PROGRAM PRVE
STOPNJE RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: viš. pred. dr. Igor Rožanc

Ljubljana, 2014

To delo je ponujeno pod licenco *Creative Commons Priznanje avtorstva - Deljenje pod enakimi pogoji 2.5 Slovenija* (ali novejšo različico). To pomeni, da se tako besedilo, slike, grafi in druge sestavine dela, kot tudi rezultati diplomskega dela lahko prosto distribuirajo, reproducirajo, uporabljajo, priobčujejo javnosti in predelujejo, pod pogojem, da se jasno in vidno navede avtorja in naslov tega dela in da se v primeru spremembe, preoblikovanja ali uporabe tega dela v svojem delu, lahko distribuira predelava le pod licenco, ki je enaka tej. Podrobnosti licence so dostopne na spletni strani creativecommons.si ali na Inštitutu za intelektualno lastnino, Streliška 1, 1000 Ljubljana.



Izvorna koda diplomskega dela, njeni rezultati in v ta namen razvita programska oprema je ponujena pod licenco *GNU General Public License*, različica 3 (ali novejša). To pomeni, da se lahko prosto distribuira in/ali predeluje pod njenimi pogoji. Podrobnosti licence so dostopne na spletni strani <http://www.gnu.org/licenses>.¹

¹V dogovoru z mentorjem lahko kandidat diplomsko delo s pripadajočo izvorno kodo izda pod katero izmed alternativnih licenc, ki ponuja določen del pravic vsem: npr. *Creative Commons* **Error! Reference source not found.** in *GNU GPL* **Error! Reference source not found.**. Zgornje besedilo je opis licence, ki ga po potrebi lahko tudi prilagodite. Če se kandidat odloči, da diplomskega dela ne bo izdal pod omenjenimi licencami, je potrebno zgornje besedilo spremeniti v naslednje: »Rezultati diplomskega dela so intelektualna lastnina avtorja. Za objavljanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.« **V obeh primerih pa iz končnega besedila odstranite to opombo.**

Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo: Preprost prevajalnik besedil za platformo android.

Tematika naloge:

V diplomski nalogi predstavite razvoj enostavne mobilne aplikacije za platformo android, ki omogoča učinkovito prevajanje besedil med izbranimi jeziki.

Cilj naloge je omogočiti hitro prevajanje krajših besedil brez uporabe svetovnega spleta. V ta namen preučite obstoječe spletne prevajalnike, izberite najprimernejšega ter njegove prevode uporabite v svojem slovarju. Aplikacija naj omogoča dopolnjevanje slovarja besed, njihovo poenostavljeno prevajanje ter uporabniku prijazen izpis. Na koncu aplikacijo primerjajte z obstoječimi spletnimi rešitvami.

IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Spodaj podpisani Ergim Ramadan, z vpisno številko 63050109, sem avtor diplomskega dela z naslovom:

Preprost prevajalnik besedil za platformo android

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom viš. pred. dr. Igorja Rožanca,
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela,
- soglašam z javno objavo elektronske oblike diplomskega dela na svetovnem spletu preko univerzitetnega spletnega arhiva.

V Ljubljani, 27. november 2014

Podpis avtorja:

Zahvaljujem se mentorju viš. pred. dr. Igorju Rožancu za mentorstvo, nasvete in popravljanje napak pri pisanju diplomskega delu. Zahvaljujem se tudi svoji družini za podporo in financiranje pri študiranju.

Kazalo

Povzetek

Abstract

Poglavje 1	Uvod	1
Poglavje 2	Uporabljena tehnologija in orodje	3
2.1	Strojno prevajanje besedil	3
2.1.1	Google prevajalnik	5
2.2	Prevajalnik Bing Translator	6
2.3	Podatkovna baza SQLite	7
2.4	Android	8
2.5	Razširljiv označevalni jezik XML	9
2.6	Razvojno okolje Eclipse	10
Poglavje 3	Izdelava aplikacije	11
3.1	Okvirne zahteve in načrt izdelave	11
3.2	Priprava besed za polnjenje slovarja	13
3.3	Tehnični opis rešitve	15
3.3.1	Lokalna podatkovna baza	15
3.3.2	Datoteka AndroidManifest.xml	17
3.3.3	Razred AsyncTask	18
3.3.4	Namestitev Bing API-ja in uporaba	19
3.3.5	Dodajanje jezika in novih besed	21
3.3.6	Prevod neznanih besed	23
3.4	Zgled in uporaba rešitev	24
3.4.1	Izdelava uporabniškega vmesnika	24
3.4.2	Lastnosti aplikacije	26

Poglavje 4	Analiza.....	29
4.1	Primerjava prevajalnikov besedil	29
4.2	Izboljšava in mnenje uporabnika.....	32
4.3	Testiranje	33
Poglavje 5	Sklepne ugotovitve	35

Seznam uporabljenih kratic

kratica	angleško	slovensko
OHA	Open Handset Alliance	poslovno združenje številnih podjetij
XML	eXtensible Markup Language	razširljiv označevalni jezik
HTML	Hypertext Markup Language	jezik za označevanje nadbesedila
IDE	Integrated Development Environment	integrirano razvojno okolje
OS	operating system	operacijski sistem
IOS	iPhone operating system	iPhone operacijski sistem
SMS	Short Message Service	kratka sporočila
SMT	Statistical Machine Translation	statistično strojno prevajanje
ASCII	American Standard Code for Information Interchange	amerški standardni nabor za izmenjavo informacij
UTF-8	Universal Character set + transformation Format-(8 bit)	splošni nabor znakov (8-bitni)
API	Application Programming Interface	vmesnik za programiranje aplikacij
SQL	Structured Query Language	strukturiran povpraševalni jezik
SDK	Software Development Kit	programski razvojni paket
RBMT	Rule-Based Machine Translation	sistemi strojnega prevajanja, temelječi na pravilih

Povzetek

Namen diplomske naloge je razviti aplikacijo za pametne telefone s platformo android, ki bo uporabniku omogočila prevajanje besedil iz izbranega jezika v drugi jezik.

Osnovna ideja naloge je uporabiti znanje obstoječega prevajalnika besedil (Bing Translator) v učinkoviti aplikaciji za pametne telefone. V podatkovni bazi je shranjenih 12000 angleških besed, ki jih najpogosteje uporabljamo v vsakodnevem pogovoru. Pri namestitvi aplikacije se uporabnik omeji na manjše število besed, nato pa izbere dva jezika prevajanja (izmed dvajset možnih). Javanska aplikacija uporabi podatkovno bazo in vnaprej zgenerira prevode v obeh jezikih. Omogočeno je tudi dodajanje novih besed iz vnešene spletne strani. Pri prevajanju je beseda lahko napisana v različnih sklonih, spolih, časih in številih, zato prevajalnik besedil išče tudi skrajšane besede. Za boljše razumevanje načina prevajanja prevajalnik besede v prevodu različno obarva.

Aplikacija je bila razvita s ciljem omogočiti čim enostavnejšo, učinkovito in hitro rešitev brez uporabe spleta. Omogočena je tudi uporaba na starejših pametnih telefonih s platformo android. V zadnjem delu diplome smo aplikacijo primerjali z znanima prevajalnikoma besedil.

Ključne besede: prevajalnik besedil, elektronski slovar, pametni telefon, android, Bing prevajalnik

Abstract

The aim of the graduation thesis is to develop an application for smartphones with the Android operating system, which will enable translation of text from a selected language into a target language.

The basic idea of the thesis is to transform the experience of the existing language translator (Bing Translator) into a functional smartphone application. The database contains 12,000 most commonly used words in everyday English. On application installation the users are limited to smaller number of words and they select the two languages for translation (out of 20 languages possible). The Java application uses the database and generates the translations in both languages in advance; the users can append new words from an arbitrary web site as well. The word can be spelt in different declinations, genders, tenses and numbers so the translator searches for the shortened forms of words. For greater clarity the application uses different colours.

The application was developed in order to provide a simple, effective and fast solution without using the internet, supporting the use on the older smartphones with the Android operating system as well. In the last part of the thesis the application was compared with two well-known text translators.

Keywords: language translator application, electronic dictionary, smartphone, Android, Bing translator app

Poglavje 1

Uvod

V današnjem času je zelo pomembno znati več jezikov: v času študija zaradi izmenjave študentov, branja literature v različnih jezikih, potovanj, lažje zaposlitve doma in v tujini, komuniciranja s sodelavci in prijatelji. Naštejemo lahko še več razlogov, zakaj je pomembno znanje jezikov. Namen te diplomske naloge je razviti aplikacijo - prevajalnik besedil, ki bo uporabniku omogočil prevajanje v več jezikov. Lastnost te aplikacije je, da omogoča uporabniku ustvarjanje lastnega prevajalnika besedil po meri. Sam si izbere jezike in koliko besed naj prevajalnik v besedilu uporabi. Vedno je omogočeno dodajanje novih jezikov in dodajanje različnih besed s spletnih strani. Zaradi naraščanja mobilne industrije in širjenja uporabe pametnih telefonov je aplikacija razvita za platformo android, ki je eden najbolj uporabnih operacijskih sistemov pri pametnih mobilnih telefonih.

Glavni cilj diplomske naloge je razviti prevajalnik besedil za platformo android, ki bo uporabniku služil kot pripomoček pri prevajanju besed v različnih jezikih na čim bolj enostaven način. Ena od možnosti je dodajanje različnih jezikov in povečanje števila besed v podatkovni bazi slovarja. Razvoj aplikacij lahko razdelimo na štiri korake. Prvi korak je nabrati dovolj besed (s spleta), ki jih največkrat uporabljamo pri vsakodnevnem pogovoru. Drugi korak je nastaviti bazo znotraj aplikacije, kjer bodo shranjene besede, in dodatne metode za dodajanje in brisanje besed znotraj baze. Sledi korak z dodajanjem Bing prevajalnika oziroma njegovega API-ja z aplikacijo, ki bo v ozadju služila za prevajanje v izbranem jeziku. Zadnji korak je izdelava vmesnika za čim lažjo in enostavnejšo uporabo aplikacije.

Na začetku sta predstavljena uporabljena tehnologija in orodje. Za tem je opisana izdelava aplikacije in razlaga rešitev, ki so bile izbrane. Sledi analiza aplikacije s primerjanjem s podobnimi aplikacijami. Na koncu je zaključek.

Poglavje 2

Uporabljena tehnologija in orodje

2.1 Strojno prevajanje besedil

Strojno prevajanje (ang. machine translation) je postopek, pri katerem računalniški program analizira besedilo in brez posredovanja človeka proizvede ciljno besedilo [1]. Prevajanje je zahtevno in ustvarjalno dejanje, zato strojno prevajanje lahko v nekaterih primerih delo olajša ali pa ga celo popolnoma nadomesti. Besedilo, ki je bilo prevedeno strojno, moramo nujno preveriti in popraviti. S strojnim prevajanjem se ukvarja predvsem računalniško jezikoslovje.

Prvi sistem za strojno prevajanje so razvili leta 1954 na univerzi v Georgetownu in prvič javno prikazali strojno prevajanje [2]. Z IBM-ovim računalnikom so prevedli več kot šestdeset povedi iz ruščine v angleščino.

Prvi spletni prevajalnik se je pojavil leta 1997 pod imenom Babel Fis, ki je bil brezplačen in je znal prevajati v 10 različnih jezikov [2]. Leta 2006 ga je prevzel Yahoo, ki ga je pod imenom Yahoo Babel Fish raširil na kar 38 jezikov. Danes na spletu najdemo dosti prevajalnikov, ki znajo prevajati v več jezikov. Na spletu so v kombinaciji s slovenščino na voljo naslednjih strojni prevajalniki: Amebis Presis [3], Google Prevajalnik [4], Bing Translator [5] in Prevajalnik.net [6].

Področje strojnega prevajanja besedil se deli na dve glavni skupini: prevajanje na osnovi pravil (ang. Rule-Based Machine Translation - RBMT) ter statistično strojno prevajanje (ang. Statistical Machine Translation - SMT).

- **Sistem strojnega prevajanja** temelji na pravilih [1]. Način zapisa pravil se razlikuje med sistemi, veže pa jih skupno dejstvo, da je postavitve takšnega sistema dolgotrajno opravilo. V to skupino sodi večina današnjih komercialnih prevajalnih sistemov, čeprav se pri gradnji poslužujejo nekaterih manj standardnih prijemov. Sistemi te paradigme izvorno besedilo najprej morfološko ter skladenjsko analizirajo ter izdelajo predstavitev vhodnega besedila, ponavadi v obliki skladenjskega drevesa izpeljave. Proces transferja prevede abstraktno predstavitev vhodnega besedila v izvornem jeziku v podobno predstavitev v ciljnim jeziku, to predstavitev sistem uporabi kot osnovo za

generacijo besedila v ciljnem jeziku, kje bistvu uporabi inverzne metode prvega dela na ciljnem jeziku. Primer: Amebis Presis [3].

- **Sistem statističnega strojnega prevajanja.** Takšni sistemi so osnovani na parametričnih statističnih modelih, ki so naučeni na poravnanih dvojezičnih učnih primerih [7]. Namesto razdeljevanja stavkov po slovničnih pravilih iščemo splošne vzorce, ki se porajajo pri uporabi jezika. Glavna prednost statističnega pristopa je, da so metode neodvisne od jezika. Čeprav so ti uporabnejši za določene jezike, med temi ni slovenščine). Glavna pomanjkljivost sistemov strojnega prevajanja na osnovi pravil je slab pregled nad delovanjem sistema, sistematske napake je zelo težko odpraviti, uvajanje lingvističnega znanja je le delno mogoče oziroma celo nemogoče. Takšni prevajalci besedi so: Google prevajalnik [4] ter Microsoft BING prevajalnik [5].

Vsaka od opisanih skupin prevajalnih sistemov ima prednosti in slabosti, vendar se moramo osredotočiti na kakovost prevajalnega sistema ter prilagodljivost in enostavnost pri naknadnem dodajanju vsebine oziroma odpravljanju napak. V tem primeru pridejo do izraza prevajalni sistemi na osnovi pravil, saj vsebujejo kar nekaj prednosti, kot sta natančnost, sledljivost prevajalnih postopkov ter enostavno dopolnjevanje [1].

Kljub vsemu napredku pa prevajalska orodja po našem mnenju še ne ogrožajo prevajalskega poklica.

2.1.1 Google prevajalnik

Opisali bomo Google prevajalnik, ki spada med sisteme statističnega strojnega prevajanja. Navedli bomo njegove lastnosti.

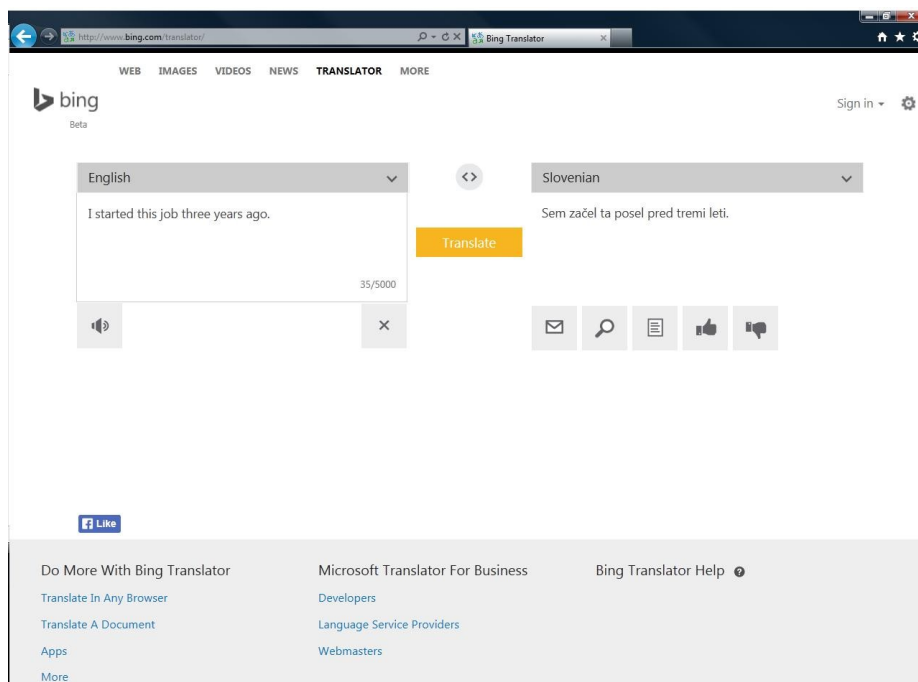


Slika 1: Google prevajalnik za operacijski sistem android.

Google prevajalnik je brezplačna prevajalska storitev za takojšnje prevajanje v številne jezike [4]. Prevaja lahko besede in stavke, v kateri koli kombinaciji podprtih jezikov. Je največji in najbolj poznan prevajalnik. Trenutno lahko prevaja v 80 različnih jezikov. Na voljo je za Android OS (slika 1) in za IOS. Za oba dva operacijska sistema je na voljo brezplačno. Lahko go uporabimo kot API za lastno razvijanje aplikacij, kjer je bila prva verzija brezplačna, ampak na žalost prevaja eno besedo naenkrat in ima manj jezikov na razpolago. Verzija 2 je plačljiva. Omogoča prevajanje več stavkov naenkrat in v 80 različnih jezikih. Google prevajalnik ima nekaj funkcij, ki olajšujejo delo z njim. Ena od njih je izgovorjanje besede. Prevajalnik zna v štiridesetih jezikih prepoznati besede in jih izgovoriti. Trenutno ta funkcija še ni pripravljena za slovenski jezik. Funkcija prepoznavanja besed je mogoča s pomočjo kamere. Ta funkcija omogoča prevajanje besedila, ki je napisano na listu. Na izboru imamo zgodovino, kjer je shranjeno, kaj smo prevajali. Zanimiva lastnost prevajalnika je prevajanje SMS sporočil, ki omogoča direktno

prevajanje v izbranem jeziku. Za nekaj jezikov je na voljo tudi prenos v pametni telefon, ki omogoča prevajanje brez internetne povezave.

2.2 Prevajalnik Bing Translator



Slika 2: Bing Translator.

Bing Translator (slika 2) je storitev, ki jo ponuja podjetje Microsoft v sklopu storitev Bing in omogoča prevajanje dokumentov ali spletnih strani v različne jezike [5]. Storitev se je najprej imenovala Live Search Translator, nato Windows Live Translator. Prevajanje je bilo osnovano na programu Systran, sedaj pa deluje na osnovi programa Microsoft Translation, ki ga razvija raziskovalna ekipa Microsoft Research. Pri prevajanju računalniških besedil Bing uporablja statistično metodo, ki jo je razvilo podjetje Microsoft. Trenutno podpira prevajanje med 31 jeziki, med njimi je tudi slovenščina.

Bing Translator je hibridni sistem za strojno prevajanje naravnih jezikov. Sistem temelji na statističnem strojnem prevajalniku. Ta uporablja tudi pravila, ki so odvisna od jezika, ter določeno mero analize izvornega besedila. Microsoft ta sistem imenuje jezikovno

obveščeno statistično strojno prevajanje besedil. Sistem je v osnovi statistični sistem za strojno prevajanje na osnovi besednih zvez, ki vključuje jezikovno odvisno analizo besedila, drevesa odvisnosti (ang. dependency trees) ter drevesa skladišne razčlemba (ang. parse trees) in pravila za poravnavo besed za generalizacijo naučenih fraz [5].

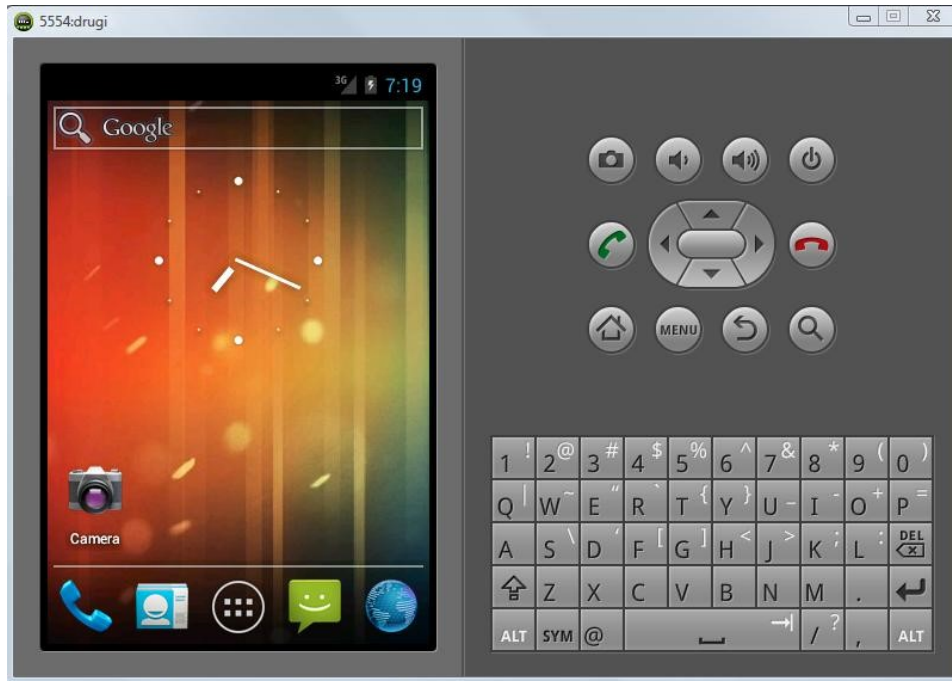
Bing Translator je na voljo tudi kot API za Javo, ki se ga da uporabiti v aplikacijah android. Imenovan je kot Microsoftov Translator API in je na voljo prek Windows Azure Marketplace. Razdeljen je na pakete, najmanjši prvi paket je brezplačen. Vsebuje pa dva milijona znakov mesečno. Večji paketi so plačljivi.

2.3 Podatkovna baza SQLite

SQLite je podatkovna knjižnica, ki nima ločenega strežniškega procesa in je odprtokodna [8]. Vse tabele, ki so kreirane v podatkovni bazi SQL, so shranjene v eni datoteki na disku. Ena od najboljših lastnosti SQLite je, da ne vzame preveč prostora. Za to se uporablja tudi pri pametnih telefonih. V vsakem sistemu android je že nameščena, tako da ni potrebe po nastavitvi. Namen aplikacije je prevajalnik, ki bo znal prevajati brez dostopa do spleta. SQLite se hitro odziva in je na samem sistemu, tako da sploh ne potrebujemo internetne povezave za dostop do baze.

Sqlite3 je vmesnik za upravljanje s SQLite relacijskimi bazami [8]. Omogoča kreiranje nove baze, dodajanje v bazo, kreiranje tabele, pisanje in brisanje v tabeli. Služi tudi kot zgled za pisanje aplikacij, ki uporabljajo SQLite knjižnico. Uporabili smo ga za kreiranje prve baze, kreiranje tabele in za napolnitev tabele z besedami.

2.4 Android



Slika 3: Android emulator.

Android je programska platforma in operacijski sistem za pametne mobilne telefone, ki temelji na Linuxovem jedru [9]. Razvija ga Google v sodelovanju s podjetji združenja Open Handset Alliance (OHA). Android je v prvi vrsti namenjen mobilnim telefonom, vendar ga najdemo tudi na drugih napravah, kot recimo tabličnih računalnikih, ročnih urah in podobno. Ime platforme oziroma operacijskega sistema izhaja iz angleške besede »android«, kar pomeni robota, ki zglada in se obnaša kot človek. Prednosti androida so, da je brezplačen in odprtokoden, omogoča razvijalcem več možnosti pri razvoju programske opreme, se ponaša z najboljšim uporabniškim vmesnikom, naprave z operacijskim sistemom android se samodejno sinhronizirajo z Googlovimi storitvami, uporabnik lahko enostavno in brez pomoči osebnega računalnika nalaga aplikacije. Danes je najbolj razširjen operacijski sistem za pametne mobilne telefone in tablične računalnike. Na sliki 3 je prikazan emulator android, ki posnema android vmesnik na osebem računalniku. Odpravlja potrebo po mobilni napravi pri razvoju aplikacij android, zato smo ga uporabili pri razvoju aplikacije.

2.5 Razširljiv označevalni jezik XML

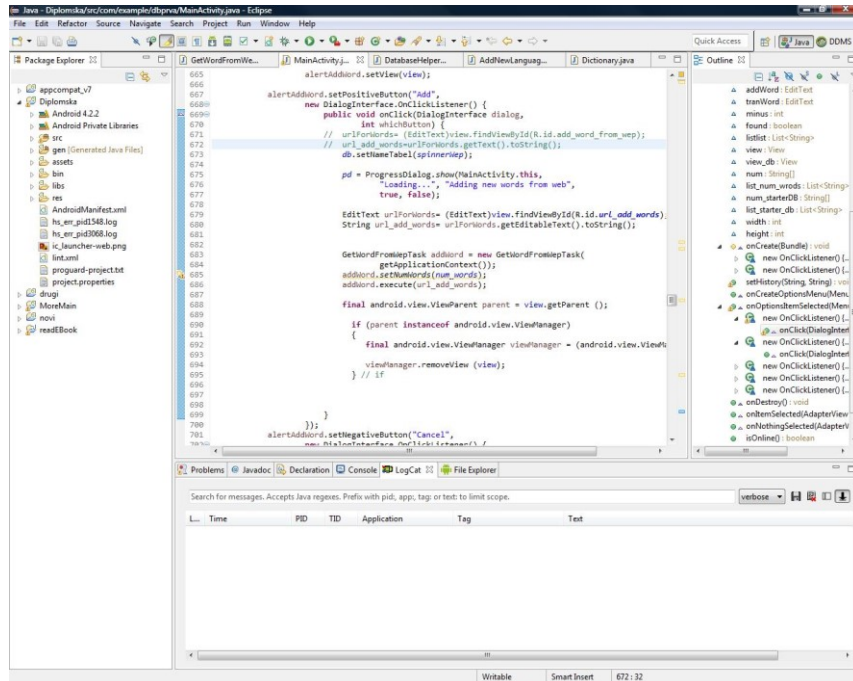
XML je tričrkovna okrajšava za angleški izraz Extensible Markup Language, razširljiv označevalni jezik [10]. XML je preprost računalniški jezik, podoben HTML-ju, ki nam omogoča format za opisovanje strukturiranih podatkov ali arhitekturo za prenos podatkov in njihovo izmenjavo med več omrežji. XML spreminja mnoge aspekte računalništva, še posebej na področju komuniciranja aplikacij in strežnikov. Da pa se ga tudi razširijo, saj ima namreč to možnost, da si lahko sami izmislimo imena etiket (ang. tag). Zelo je uporaben za komunikacije, saj ima zelo preprosto in pregledno zgradbo.

```
<?xml version="1.0" encoding="UTF-8"?>
<knjige>
  <knjiga kategorija="OTROCI">
    <naslov>Harry Potter</naslov>
    <avtor>J K. Rowling</avtor>
    <leto>2005</leto>
    <cena>29,99</cena>
  </knjiga>
  <knjiga kategorija="PROGRAMIRANJE">
    <naslov>Java SE 7 Programming Essentials</naslov>
    <avtor>Michael Ernest</avtor>
    <leto>2012</leto>
    <cena>29,39</cena>
  </knjiga>
</knjige>
```

Slika 4: Primer XML dokumenta.

Na sliki 4 je primer XML dokumenta. XML ima svoja sintaksna pravila. Etikete so občutljive na velike in male črke, vsaka etiketa mora biti zaključena. Dokument mora imeti en element, ki bo korenska etiketa, in še več sintaksnih pravil. Bralca sintaksna pravila ne zanimajo in za to rečemo, da je XML enostaven za branje.

2.6 Razvojno okolje Eclipse



Slika 5: Razvojno okolje Eclipse.

Eclipse (slika 5) je odprtokodno okolje za razvoj programov (IDE) in temelji na javi [11]. Z Eclipse je mogoče razvijati programe v različnih jezikih. Prednost uporabe Eclipse se pokaže pri razvoju zahtevnih programov, ki uporabljajo kompleksne knjižnice, za katere je potrebno podrobno poznavanje delovanja in hierarhije razredov. Večina knjižnic, tudi komercialnih, ima dodatke, ki omogočajo lažjo uporabo. Tudi pri običajnih knjižnicah se pokažejo prednosti, saj Eclipse prebere zaglavne datoteke (ang. header) in zna sproti razširjati in pripraviti pomoč za argumente funkcij, ki jih uporabljamo. Vgrajene ima še različne brskalnike kode in podporo za razhroščevalnike. Z različnimi dodatki se da doseči tudi izboljšano integracijo s servisi za nadzor kode (subversion [12], git [13], ...) ali spletnimi servisi (gForge [14] ...).

Poglavje 3

Izdelava aplikacije

V tem poglavju bomo opisali, kako smo razvili celo aplikacijo oziroma kako smo prišli od ideje do delujoče aplikacije. Najprej bomo predstavili zahteve, ki smo si jih postavili, nato celoten načrt aplikacije in nazadnje še uporabniški vmesnik in lastnosti.

3.1 Okvirne zahteve in načrt izdelave

Razvijanje prevajalnika besedil, ki zna prevajati med različnimi jeziki, je zelo zahtevno delo. S strojnim prevajanjem besedil se ukvarja računalniško jezikoslovje, ki ima več usposobljenih ljudi za to področje, ki učijo jezike, razvijajo algoritme in podobno.

Naša zahteva je bila razviti preprosto aplikacijo, ki zna prevajati besede med različnimi jeziki. Zahteve, ki jih bo vsebovala aplikacija, so:

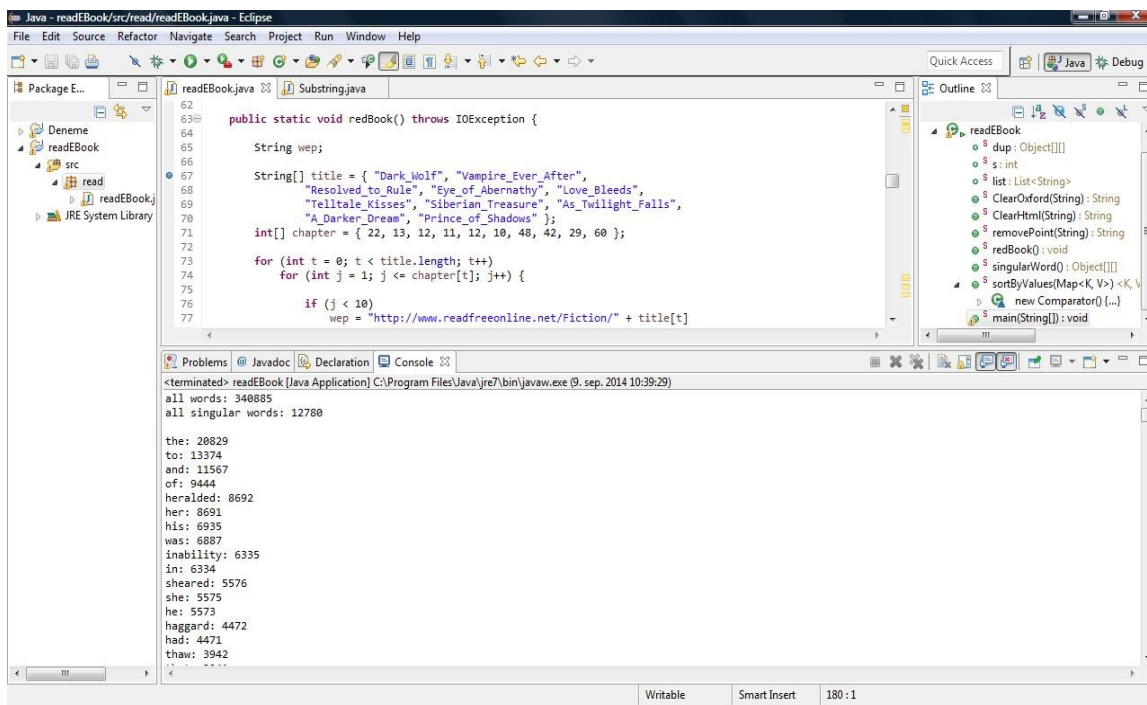
- uporaba besed iz vnaprej pripravljenega slovarja
- dodajanje novih jezikov,
- dodajanje novih besed s spletnih strani,
- iskanje tudi skrajšanih besed za prevod,
- obarvanje besed, ki niso natančno prevedene,
- pregledovanje zgodovine prevedenih besed,
- hitro delovanje,
- enostavna uporaba,
- pregled velikosti slovarja,
- uporaba brez internetne povezave,
- uporaba čim manj spomina v pametnih telefonih,
- delovanje tudi na starejših pametnih telefonih.

Zaradi lažjega in hitrejšega razvijanja aplikacije je bilo potrebno narediti celoten načrt. Aplikacijo smo razdelili na več manjših delov za lažje upravljanje. Okvirni koraki so naslednji:

- določi besede v slovarju,
- pripravi prvi vmesnik za začetno delo,
- nastavi bazo in napiši metode za upravljanje z bazami,
- nastavi Bing prevajalnik, ki bo prevajal v ozadju,
- najdi rešitev za povečanje števila besed v bazi,
- določi postopek prevajalniku besedil, ko ne najde besede,
- naredi končni vmesnik,
- testiraj,
- analiziraj: preveri, kako dobro deluje!

Po izdelavi načrta je bilo potrebno programirati vsak korak posebej in se odločiti za pravo rešitev. V nadaljevanju bomo opisali vsak korak in naše rešitve pri izdelavi celotne aplikacije.

3.2 Priprava besed za polnjenje slovarja



Slika 6: Besede za polnjenje tabel.

Prva in najpomembnejša stvar je izbira primernih besed za polnjenje baze. Katere besede največkrat uporabljamo in katere so najpomembnejše pri vsakodnevem razgovoru, je težko odločiti. Raziskava je pokazala, da izgovorimo približno 15 000 besed dnevno. Različnih besed pa je med 500 do 700 dnevno. Za dobro razumevanje in govorjenje moramo znati približno 25 000 različnih besed [15].

Napisali smo program v javi (slika 6), ki je iz interneta prebral deset različnih elektronskih knjig. Elektronske knjige, ki smo jih prebrali, in njihovi avtorji so:

1. Christine Feehan: Dark Wolf [16].
2. RaShelle Workman: Vampire Ever After? [16].
3. Darren Shan: Lady of the Shades [16].
4. Richelle Mead: Last Sacrifice [16].
5. Gillian Flynn: Gone Girl [16].

6. Chloe Neill: Some Girls Bite [16].
7. Colleen Gleason: Siberian Treasure [16].
8. Amanda Ashley: As Twilight Falls [16].
9. Suzanne Wright: From Rags [16].
10. Rachel Caine: Prince of Shadows [16].

Elektronske knjige so napisali različni pisatelji in vsaka elektronska knjiga ima različno temo. S tem smo nabrali dovolj besed z različnih področij. Vse elektronske knjige so bile napisane s 340 885 besedami, uporabljenih je bilo 12 780 različnih besed (slika 6). V samo štetjo besed nismo vključili posebnih imen. Lahko zaključimo, da se s slabimi trinajst tisoč besedami lahko pogovarjamo vsaj na deset različnih tem.

Če pogledamo samo delovanje Bing prevajalnika, vsak jezik prevede najprej v angleščino, potem v izbrani jezik. Poglejmo konkreten primer: če hočemo prevajati iz slovenščine v srbsščino, bo prevajalnik slovensko besedo najprej prevedel v angleščino, potem pa iz angleščine v srbsščino. Zaradi takšnega delovanja prevajalnika so bile izbrane knjige v angleškem jeziku. S tem smo omogočili hitrejše delovanje aplikacije.

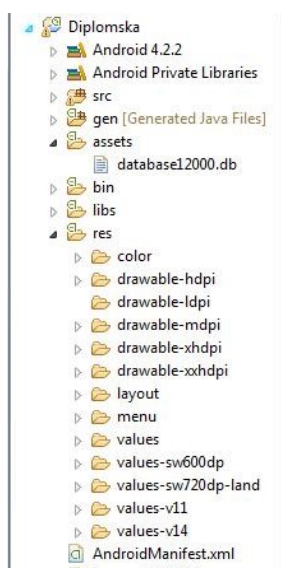
Dobljenih slabih trinajst tisoč besed smo uredili po pogostosti uporabe. Prvih 12 000 besed smo dodali v bazo. Za kreiranje baze, kreiranje tabele in dodajanje besed v tabelo smo si pomagali z vmesnikom Sqlite3.

Razlog za pisanje programov v Javi, kreiranje baze s Sqlite3 in dodajanje v androidu je hitrejše delovanje same Jave. Za branje desetih elektronskih knjig s spleta, urejanje besed, čiščenje html znakov in štetje števila pojavitev besed je bilo potrebno v androidu čakati več kot štiri minute, v Javi pa manj kot minuto. Drugi razlog za takšno odločitev je bil, da ne moramo ugotoviti, koliko časa bo še dostopna spletna stran z elektronskimi knjigami, ki smo jih izbrali za začetno bazo.

3.3 Tehnični opis rešitve

3.3.1 Lokalna podatkovna baza

Ideja je bila razviti aplikacijo, ki bo delala tudi brez internetne povezave, zato je bila potrebna lokalna baza. Zaradi hitrejšega delovanja aplikacije smo se odločili na SQLite, ki je vgrajen v samem androidu. Razvili smo razred *DatabaseHelper*, ki skrbi za odpiranje, posodabljanje in izdelavo podatkovne baze. Znotraj razreda so napisane funkcije, ki služijo za opravljanje vseh zahtevnih nalog aplikacije: prepis vsebine, kreiranje nove tabele, kreiranje tabele zgodovina, dodajanje besed, dobivanje besed, število tabel in njihova velikost, iskanje besede v tabeli.



Slika 7: V datoteki *assets* je shranjena začetna baza, ki se imenuje *database12000.db*.

Pri namestitvi aplikacije na pametni telefon se prepíše vsebina, ki smo jo kreirali s pomočjo SQLite3 in napolnili z besedami. Začetna baza *database12000.db*, ki vsebuje eno tabelo, je shranjena v datoteko *assets* znotraj androida (slika 7). Metoda kreiranja tabele zgodovina se tudi kreira ob namestitvi aplikacije na telefonu, ki služi za shranjevanje zgodovine, katere besede v katerem jeziku smo prevajali.

Po namestitvi aplikacije na pametnem telefonu ta vsebuje dve tabeli. Prva tabela je *english*, kjer so shranjene najbolj uporabljene besede. Vsaka beseda je označena z enolično številko. Ta številka bo služila za iskanje besede v drugih jezikih, ki jih bo kreiral uporabnik. Druga tabela je *zgodovina*, ki ima eno enolično število in dva teksta.

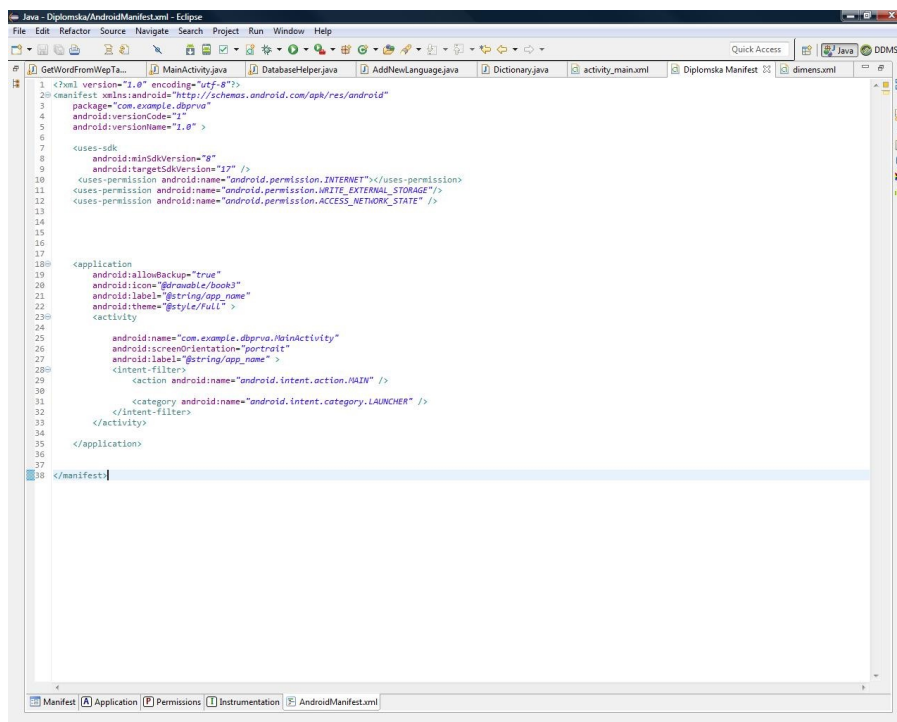
```
public void addIfNotInDatabase(List<String> list) {
    SQLiteDatabase db = this.getWritableDatabase();
    Iterator<String> it = list.iterator();

    while (it.hasNext()) {
        String word = it.next().toString();
        if (getWordFromTabel(word) == -1) {
            Dictionary di = new Dictionary(word);
            createDictionary(di);
        }
    }
}
```

Slika 8: Metoda dodajanje novih besed.

Ena najpomembnejših metod v razredu *DatabaseHelper* je *addIfNotInDatabase()* (slika 8), ki se uporablja pri dodajanju novih besed s spletnih strani. Ta metoda omogoča dodajanje novih besed v bazo z vnešenih spletnih strani, ki jih nimamo v naši bazi. Metoda sprejema en list, ki je napolnjen z besedami s spletnih strani. Vsaka beseda posebej se vzame z lista in s pomočjo metode *getWordFromTabel(String word)*, ki je tudi napisana v razredu *DatabaseHelber*, preveri, ali beseda obstaja v naši tabeli. V slučaju, da metoda *getWordFromTabel(String word)* vrne minus ena, pomeni, da besede ni v tabeli. Besede, ki niso v tabeli, s pomočjo metode *createDictionary(Dictionary di)* dodamo v tabelo.

3.3.2 Datoteka *AndroidManifest.xml*

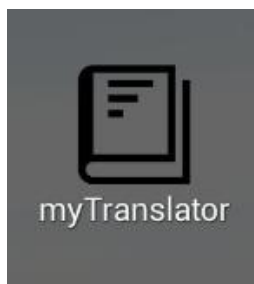


```
1 <?xml version="1.0" encoding="utf-8"?>
2 <manifest xmlns:android="http://schemas.android.com/apk/res/android"
3     package="com.example.dbrva"
4     android:versionCode="1"
5     android:versionName="1.0" >
6
7     <uses-sdk
8         android:minSdkVersion="8"
9         android:targetSdkVersion="17" />
10    <uses-permission android:name="android.permission.INTERNET"></uses-permission>
11    <uses-permission android:name="android.permission.WRITE_EXTERNAL_STORAGE"></uses-permission>
12    <uses-permission android:name="android.permission.ACCESS_NETWORK_STATE" />
13
14
15
16
17
18
19    <application
20        android:allowBackup="true"
21        android:icon="@drawable/book3"
22        android:label="@string/app_name"
23        android:theme="@style/Full" >
24        <activity
25            android:name="com.example.dbrva.MainActivity"
26            android:screenOrientation="portrait"
27            android:label="@string/app_name" >
28            <intent-filter>
29                <action android:name="android.intent.action.MAIN" />
30
31                <category android:name="android.intent.category.LAUNCHER" />
32            </intent-filter>
33        </activity>
34    </application>
35
36
37
38 </manifest>
```

Slika 9: Datoteka *AndroidManifest.xml*.

Najpomembnejša datoteka android aplikacije, ki mora biti v korenskem imeniku, je *AndroidManifest.xml* (slika 9). Operacijski sistem android jo uporablja, preden se začne izvajati aplikacija.

V datoteki je na začetku napisana verzija aplikacije. Najmanjša in ciljna verzija SDK pomeni, za katere operacijske sisteme android naj se izvaja aplikacija. Pri naših aplikacijah v *AndroidManifest.xml* je najmanjši SDK 8, ki pomeni OS Android 2.2 in ciljna verzija SDK je 17, ki pomeni OS Android 4.2.2. Za dostop do interneta in pisanja v pomnilniku je bilo potrebno dodati dovoljenje uporabnika za datoteko *AndroidManifest.xml*. Potem sledijo še ikona, ime (slika 10) in tema aplikacije.

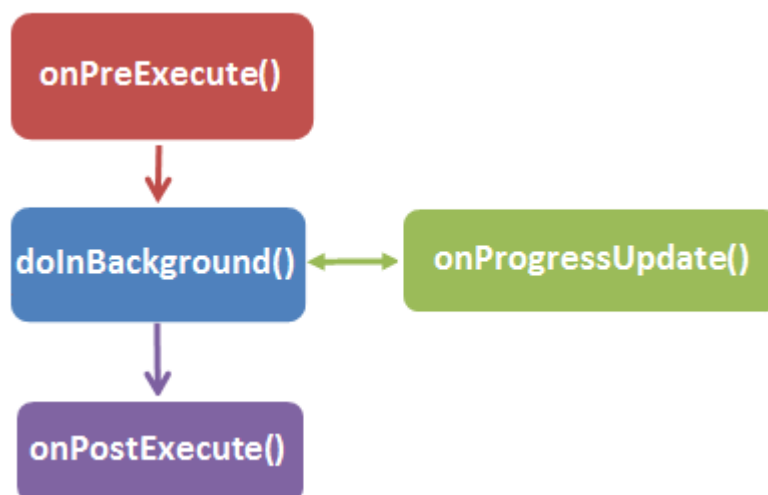


Slika 10: Ikona in ime aplikacije.

3.3.3 Razred *AsyncTask*

AsyncTask je abstraktni razred androida. Uporablja se za izvajanje dolgotrajnih operacij, kot so komuniciranje s strežnikom na spletu, prenos datotek, prenos slik in drugo. V aplikaciji je bil uporabljen pri dodajanju novega jezika in dodajanju novih besed s spleta. Namen uporabe tega razreda je uporabniku omogočiti uporabo ostale aplikacije, dokler naša aplikacija pošilja besedila strežniku za prevajanje in jih dobiva nazaj, ker to traja nekaj časa. Razred *AsyncTask* to izvaja tako, da ne zaposli glavne niti, kar bi dalo uporabniku občutek, da se je aplikacija ustavila. Razred *AsyncTask* izvaja operacije v ozadju v štirih korakih (slika 11) preko metode: `onPreExecute`, `doInBackground`, `onProgressUpdate` in `onPostExecute` [17].

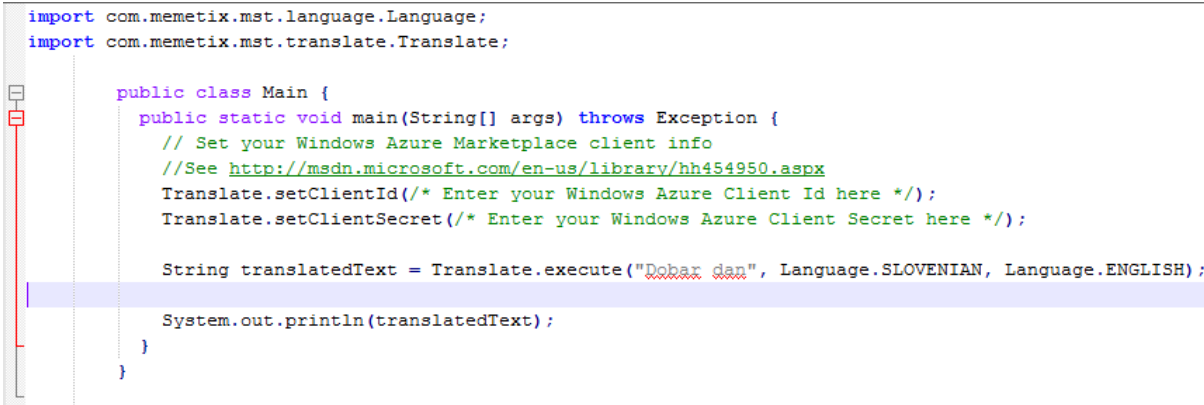
- *onPreExecute*: Ta metoda se običajno uporablja za namestitev osnovnih parametrov opravila in se izvedejo takoj, ko pokličemo razred.
- *doInBackground*: Vsa opravila, ki tečejo dalj časa, se dodajajo s to metodo. Namenjena je za izvajanje kode v ozadju.
- *onProgressUpdate*: Ta metoda se lahko uporablja za prikaz kakršnegakoli napredka, ki se dogaja v ozadju, in bo prikazan v uporabniškem vmesniku.
- *onPostExecute*: Uporabniški vmesnik prikaže končne rezultate opravila.



Slika 11: *AsyncTask* v štirih korakih.

3.3.4 Namestitev Bing API-ja in uporaba

Namestitev API-ja v Eclipsu je enostavno. Vse, kar moramo narediti, je, da dobljen API dodamo v projektno knjižnico. Za uporabo API-ja se moramo registrirati pri Microsoft Azure, od koder bomo dobili ID stranke (ang. ClientId) in njen skriti ključ (ang. setClientSecret), ki ga bomo uporabili v API-ju za povezavo do strežnika (slika 12).



```
import com.memetix.mst.language.Language;
import com.memetix.mst.translate.Translate;

public class Main {
    public static void main(String[] args) throws Exception {
        // Set your Windows Azure Marketplace client info
        //See http://msdn.microsoft.com/en-us/library/hh454950.aspx
        Translate.setClientId(/* Enter your Windows Azure Client Id here */);
        Translate.setClientSecret(/* Enter your Windows Azure Client Secret here */);

        String translatedText = Translate.execute("Dobax dan", Language.SLOVENIAN, Language.ENGLISH);

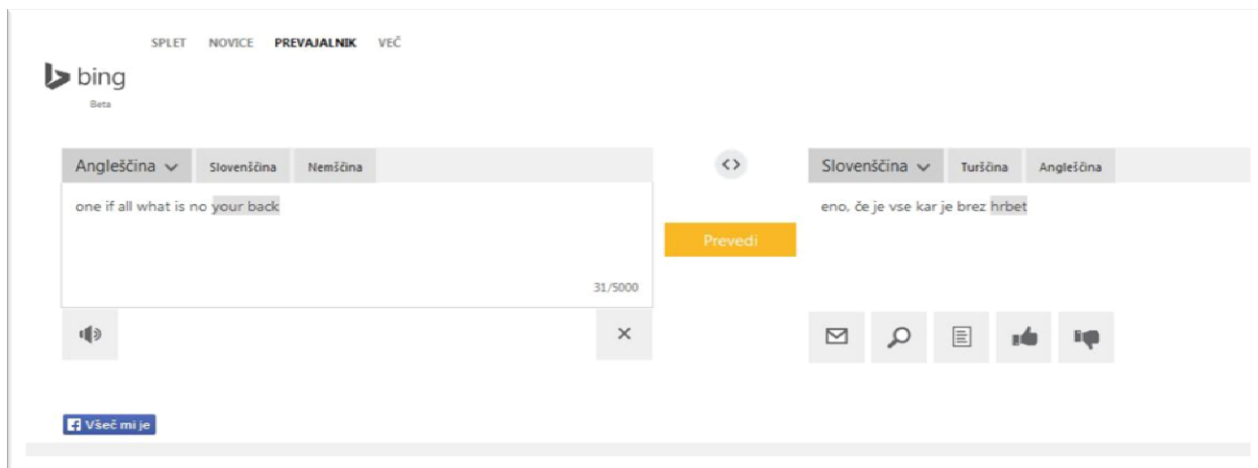
        System.out.println(translatedText);
    }
}
```

Slika 12: Primer uporabe API-ja.

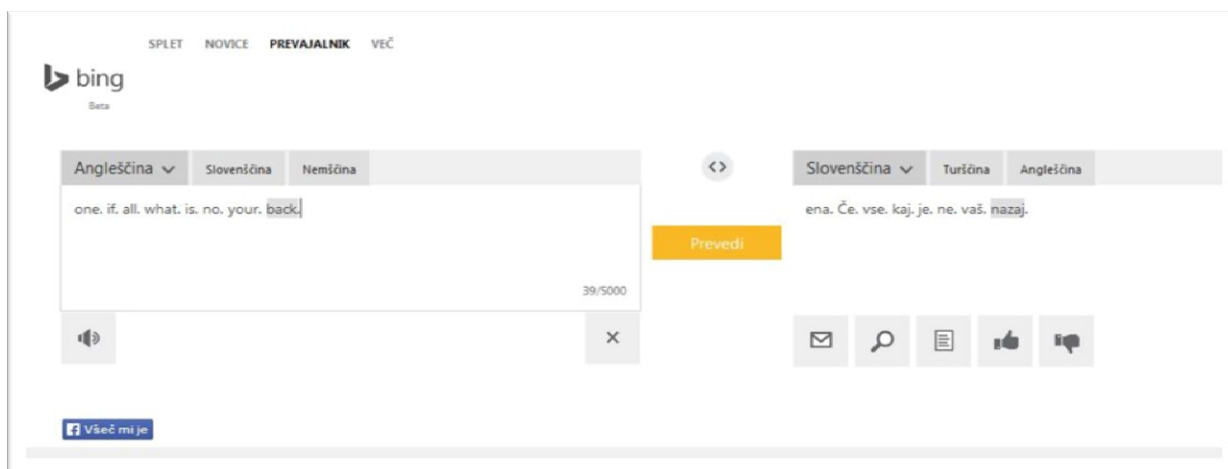
Po namestitvi obeh ključev, ki sta potrebna za povezavo do strežnika, je potrebno samo dodati tekst, ki ga bomo prevajali iz izbranega jezika v izbrani jezik. Za zamenjavo teksta, ki se bo prevajal, in za izbiranje jezika je bilo potrebno napisati razred, ki sem ga imenoval *Dictionary*. Razred *Dictionary* omogoča nastavljanje teksta, iz katerega jezika v katerega naj se prevaja, nastavljanje tabele in definiranje konstruktorjev.

Prva odločitev je bila prevajanje vsake besede posebej, vendar se je ta izkazala za neprimerno. Pošiljanje vsake besede posebej do strežnika je trajalo predolgo. Za vsako prevajanje je bilo potrebno dobiti besede iz lokalne baze, postaviti povezave s strežnikom, čakati na odgovor strežnika, prevedeno besedo shraniti v bazo. Pri več kot tri tisoč besedah je ta postopek trajal več kot šest minut; to bi bilo za uporabnika predolgo. Rešitev tega problema je bila, poslati vse besede naenkrat, s tem bi potrebovali le eno branje iz lokalne baze in (najpomembneje) enkrat vzpostavili povezavo do strežnika.

Pri pošiljanju vseh besed naenkrat smo naleteli na tri težave:



Slika 13: Nepravilno prevajanje



Slika 14: Pravilno prevajanje

Možno je, da ima ena beseda več pomenov. V svojih aplikacijah smo uporabili prvi pomen vsake besede. Zaradi več pomenov ene besede in sklanjanja prevajalnik pri različnih jezikih uporablja svoje metode za prevajanje. Kot vidimo na sliki 13, v primeru več besed skupaj za prevajanje Bing prevajalnik uporablja svoje metode tako, da včasih združi dve besedi skupaj. Primer: "your back" je preveden v "hrbet", "no" je preveden v "brez" in "one" je preveden v "eno", "what" je preveden v "kar". Tako dobljen rezultat nikakor ne bi mogli uporabiti v

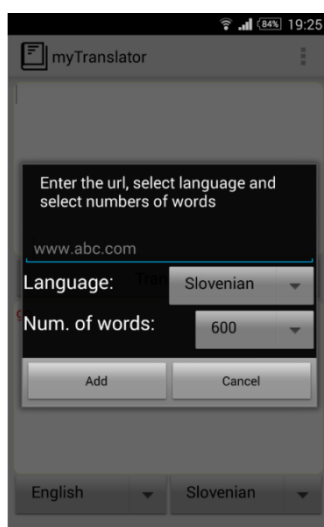
aplikaciji. Za pravilno prevajanje, ki bi bilo uporabno za aplikacijo, je bila rešitev dodajanje pike in presledka po vsaki besedi (slika 14). Po ugotovitvi takega delovanja Bing prevajalnika je bilo potrebno pred pošiljanjem vseh besed dodajati piko in presledek ter dobljenemu rezultatu, prevedenemu tekstu odstraniti pike in presledek.

Druga težava je ta, da Bing prevajalnik sprejme naenkrat največ 10 240 bajtov. Java pri ASCII črki uporablja 1 bajt za vsako črko, pri UTF-8 pa 2 bajta. Če bi naredili mali izračun in vzeli poprečje, ena beseda ima 8 črk in lahko vsebuje šumnike, bi to pomenilo manj kot 800 besed. Ker aplikacija omogoča neomejeno dodajanje novih besed, smo morali razviti dodatno kodo, ki bi celotno besedilo znala razdeliti na več delov, manjših kot 10 240 bajtov, in jih poslati strežniku za prevajanje.

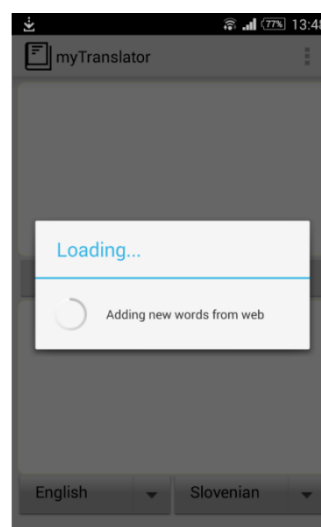
Tretja težava je v tem, da je branje iz lokalne baze, povezava do strežnika, čakanje na strežnikov odgovor in pisanje v lokalni bazi trajalo kar dvajset sekund. Uporabnik aplikacije bi se lahko počutil nerodno ali pa ujet v aplikacijo, zato je bil uporabljen razred *AsyncTask*, ki je razložen v prejšnjem podpoglavju.

3.3.5 Dodajanje jezika in novih besed

Najpomembnejša lastnost aplikacije je dodajanje novih jezikov s seznama in dodajanje novih besed s spletnih strani. Takšna funkcionalnost aplikacije omogoča uporabniku sestavljanje svojega prevajalnika besedil.



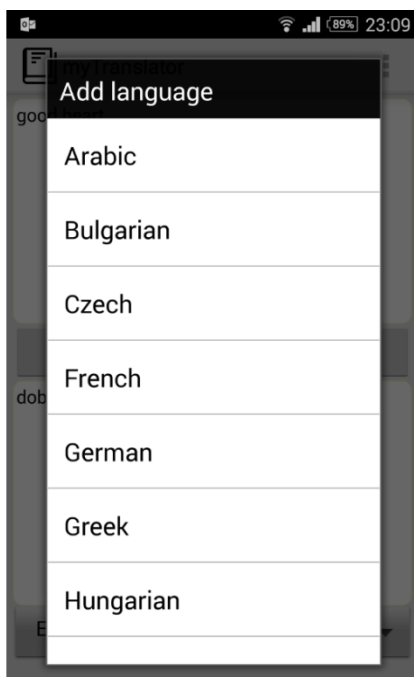
Slika 15: Dodajanje novih besede z interneta



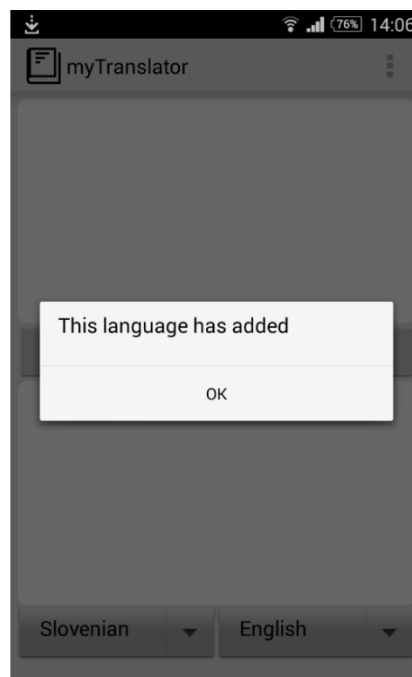
Slika 16: Obvestilo uporabniku

- Dodajanje novih besed z interneta je ena od najpomembnejših lastnosti aplikacije. Lastnost dodajanja novih besed v aplikacijo uporabniku omogoča povečanje besed v

bazi pri izbranem področju. Tako baza napolnimo z bolj uporabnimi besedami, ki uporabniku omogočajo varčevanje pri spominu pametnega telefona. Samo dodajanje besede je enostaven postopek. Ko izberemo to opcijo, je treba napisati pravilen spletni naslov, izbrati jezik in število besed (slika 15). Aplikacija se poveže na napisano spletno stran, prebere celotno besedilo, odstrani duplikate, naključno izbere vnešeno število besed, pogleda, če so izbrane besede v tabeli izbranega jezika. Besede, ki jih nimamo v tabeli, dodamo. Vsaka dodana beseda se prevede tudi v ostale jezike, ki so aktivni v aplikaciji, in se shrani vsaka v svoji tabeli. Dogodek dodajanja traja od deset do petnajst sekund. Za uporabnika, da ne bo mislil, da se je aplikacija ustavila, se na zaslonu prikaže obvestilno okno, ki sporoča uporabniku, da zadeva poteka (slika 16) in omogočeno mu je upravljanje ostalih lastnosti pametnega telefona. Zakaj izbiramo besede naključno in ne dodajamo le najpogosteje uporabljenih besed? Na začetku je bila napisana koda, da izbere najpogosteje uporabljene besede z vnešene spletne strani. Ker osnovna tabela že vsebuje najbolj uporabljene besede, se je pokazalo, da je skoraj polovica besed že v tabeli. Druga težava bi nastala, če bi uporabnik poskušal še enkrat s spletne strani povečati število besed, saj bi koda izbrala spet najbolj uporabljene besede, te pa so že v tabeli in število besed v tabeli bi ostalo nespremenjeno. Po ugotovitvi takšnega rezultata je bila koda spremenjena, naj izbere naključno izbrano število besed. S tem smo omogočili večkratno dodajanje z iste spletne strani.



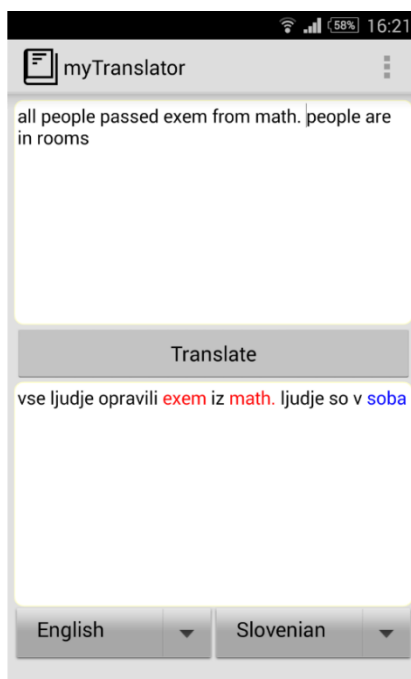
Slika 17: Dodajanje novega jezika



Slika 18: Jezik obstaja

- Dodajanje novih jezikov. Aplikacija po nastavitvi na pametnem telefonu vsebuje samo angleški jezik. Če z menija izberemo dodajanje novega jezika, se bodo pokazali na zaslonu vsi možni jeziki (slika 17), ki jih lahko dodamo. Bingov prevajalnik vsebuje 31 jezikov. Zaradi nerazumevanja uporabljenih črk in besed pri različnih jezikih nismo mogli uporabiti vseh jezikov, ki jih uporablja Bingov prevajalnik. Trenutno naš prevajalnik vsebuje 20 jezikov. Poglejmo, kaj se dogaja v ozadju, ko se uporabnik dotakne opcije dodaj nov jezik. Program najprej preveri, ali izbrani jezik že obstaja v bazi, če je, opozori uporabnika, da izbrani jezik že obstaja (slika 18), če ga ni, ga doda. Dodajanje novega jezika poteka v nekaj korakih. Najprej se kreira nova tabela v bazi z istim imenom kot izbrani jezik. Beremo vse besede, ki so shranjene v tabeli pod imenom english. Nastavimo Bing API, da prevaja iz angleškega v izbrani jezik. Prebrane besede iz baze s pomočjo API-ja prevedemo in jih shranimo v novo kreirani tabeli po vrsti tako, da imamo isti id za besede v english tabeli in v novo kreirani tabeli. V komponenti *Spinner* je dodan nov jezik, ki je pripravljen za uporabo.

3.3.6 Prevod neznanih besed



Slika 19: Uporabljanje različnih barv, ko ne najde besede

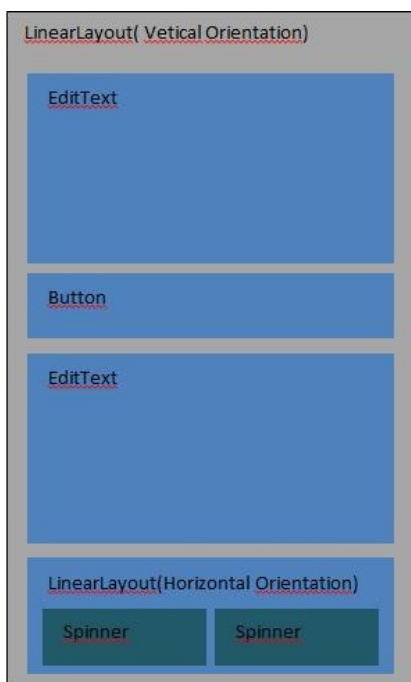
Aplikacija vsebuje tabelo, kjer so shranjene najpogosteje uporabljene besede. To so besede, ki se uporabljajo v vsakodnevem govoru v angleškem jeziku. Brez večanja števila besed smo

razvili logiko, kako bi dobili dvakrat več besed. Ideja je bila izbrisati zadnjo črko besede in poskusiti s prevodi te krajše besede. Postopek lahko ponavljamo največ trikrat pri besedah, ki so dolge več kot dve črki. Takšna logika omogoča tudi prevajanje besed, ki so napisane v drugem času, spolu, sklonu ali številu. Pri takšnem prevajanju dobljen prevod ne bo natančen, bo pa uporaben.

Na sliki 19 vidimo, ko vnesemo niz za prevajanje, program vzame vsako besedo posebej in preveri, ali je beseda v tabeli. Če je ni, izbriše zadnjo črko in ponovno preveri, ali beseda obstaja v tabeli. Najdeno besedo izpiše na zaslon, vedno uporablja modro barvo, tako da povemo uporabniku, da beseda ni prevedena popolno. Besede, ki jih sploh nimamo v slovarju, oziroma zanje nimamo prevoda, na zaslonu izpisujemo s pomočjo rdeče barve in tako sporočimo uporabniku, da beseda ni prevedena.

3.4 Zgled in uporaba rešitev

3.4.1 Izdelava uporabniškega vmesnika



Slika 20: Layout dizajn

Izdelava in načrtovanje vmesnika je poteklo v nekaj korakih. Ker je bila aplikacija namenjena vsem, je moral biti vmesnik enostaven in razumljiv, vseboval je gumb, ki izvrši funkcijo

prevajanja, dva izbirna gumba za izbiranje jezikov, med katerimi bo poteklo prevajanje, dve okni, eno za vnos teksta, drugo za prikaz prevedenega teksta, in meni.

V meniju izbiramo:

- zgodovino, ki prikazuje, katere besede so bile prevedene,
- dodajanje nove besede z interneta,
- dodajanje novega jezika,
- velikost tabele, ki prikazuje za vsak jezik, koliko besed vsebuje in
- priročnik, kako se aplikacija uporablja.

Vmesnik (slika 20) vsebuje pet komponent. Vse te komponente so razvrščene v *LinearLayout(Vertical Orientation)*, ki je vertikalno orientiran. Prva komponenta je *EditText*, ki služi za vnos teksta, druga komponenta je *Button*, ki kliče funkcijo za prevajanje, tretja komponenta je *EditText*, ki prikazuje prevedeno besedilo. Dodan je še *LinearLayout(Horizontal Orientation)*, ki je horizontalno orientiran, kjer sta dodani dve komponenti. Prva komponenta je *Spinner* za izbiro jezika, iz katerega bomo prevajali, druga komponenta pa je *Spinner*, v kateri jezik bomo prevajali.

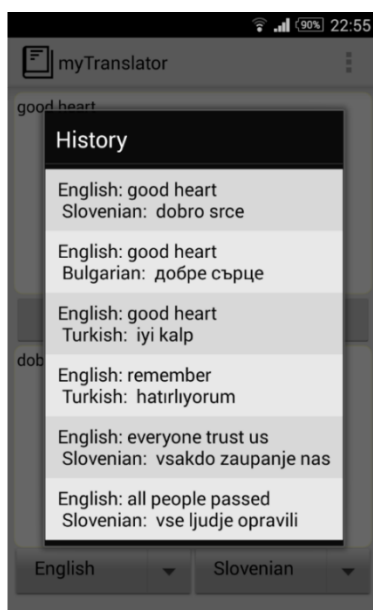
Lastnosti in postavitev komponent so zapisani v XML datoteki, ki smo jo na začetku pojasnili. Za vsako komponento je definirana velikost, barva, id in ostale lastnosti.

3.4.2 Lastnosti aplikacije



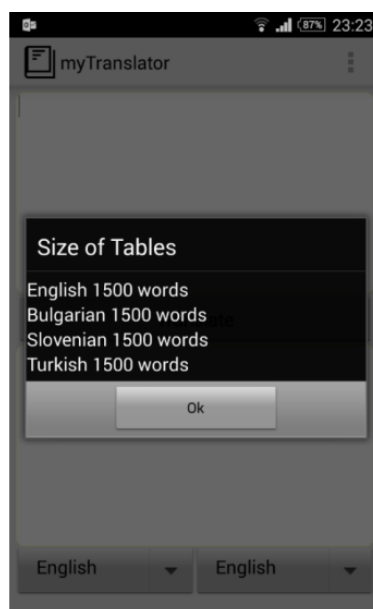
Slika 21: Meni

V tem razdelku bomo opisali lastnosti aplikacije in njeno uporabnost. Na zaslonu je gumb za prevod, dva izbirna gumba za izbiranje jezika, polje za vnos teksta, polje za prikaz prevedenega teksta in meni (slika 21). V meniju je dodatnih pet lastnosti aplikacije. Vsako posebej bomo opisali, čemu je namenjena, kako deluje in kje nastajajo težave. Lastnost dodajanja novega jezika in dodajanja novih besed smo opisali v prejšnjem poglavju.



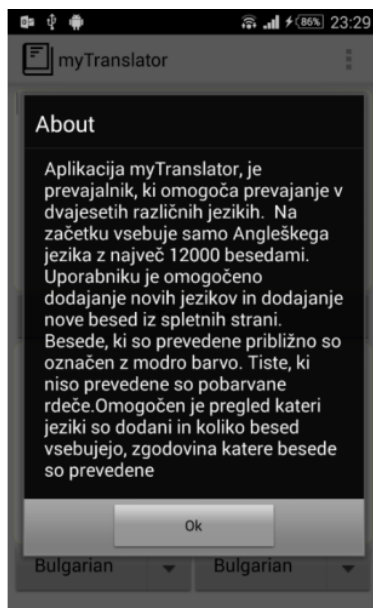
Slika 22: Zgodovina uporabniku omogoča pregled, katere besede so bile prevedene.

- Prva izbira v meniju je zgodovina. Uporabniku omogoča pregled, katere besede je prevajal in iz katerega jezika v kateri jezik (slika 22). Za prikaz zgodovine je uporabljen razred *AlertDialog*, kjer je bila dodana komponenta *RelativeLayout*. Znotraj komponente *RelativeLayout* je bila dodana komponenta *ScrollView*, ki omogoča drsanje po zaslonu, in komponenta *LinearLayout*, ki je nastavljena vertikalno in je uporabna za prikaz teksta. Tabela *History* se kreira ob prvi nastavitvi aplikacije na telefonu. Vsebuje en id in dva teksta. V id se shrani zaporedna številka ob vsakem prevajanju. V prvi tekst se shrani podatek, iz katerega jezika smo prevajali in katere besede, v drugem tekstu pa, v kateri jezik smo prevajali in prevod besede.



Slika 23: Dodani jeziki in njihova velikost

- Pregled, katere jezike vsebuje aplikacija, omogoča Tabel Size iz menija. Slika 23 prikazuje, katere tabele so kreirane in koliko besed vsebuje vsak jezik posebej. Za prikaz takšnih podatkov se povežemo z bazami, pogledamo, koliko tabel je kreiranih, iz vsake tabele preberemo njeno velikost in s pomočjo razreda *AlertDialog* prikažemo uporabniku.



Slika 24: Priročnik

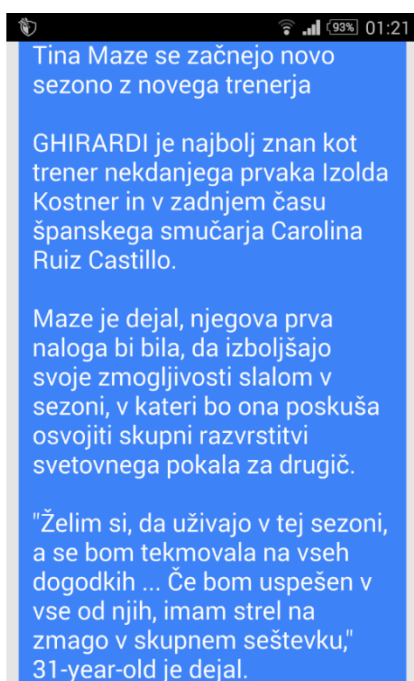
- O aplikaciji, čeprav je čisto enostavna za uporabo, je napisan priročnik (slika 24). V priročniku je opisana vsaka lastnost posebej in je razložena njena uporaba. Tako je uporabnik v celoti informiran in mu je omogočena popolna uporaba aplikacije.

Poglavje 4

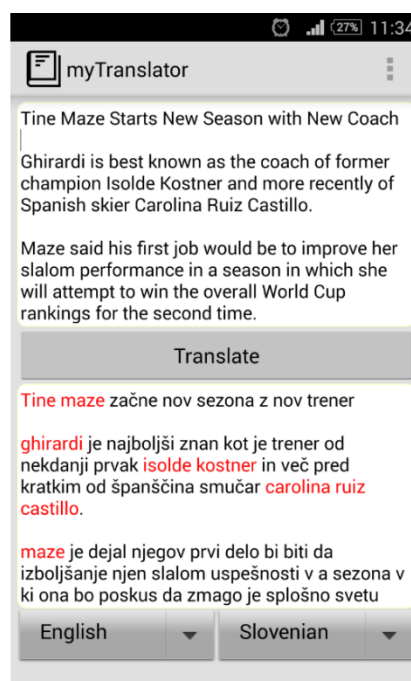
Analiza

4.1 Primerjava prevajalnikov besedil

Opisali in primerjali bomo, v čem je boljša ali slabša naša aplikacija od Google in od Bing prevajalnika.



Slika 25: Google prevajalnik



Slika 26: myTranslator prevajalnik

Prevajalnik besedil myTranslator je imel na voljo 12 000 besed. Jeziki, ki jih vseboval, so angleščina, slovenščina in turščina. Uporabili smo ga brez internetne povezave. Google prevajalnik in Bing Translator sta pri primerjavi uporabljala internetne povezave, kar pomeni, da sta imela več besed na voljo. Za boljšo primerjavo smo s spleta vzeli tri naključno izbrane članke z različnimi temami in v različnih jezikih. Pri primerjavi bomo upoštevali čas, neprevedene, približno prevedene, napačno prevedene, pravilno prevedene in v odstotkih pravih prevodov besed. Na sliki 25 je prikazan Google prevajalnik, ki je prevedel prvi članek iz slovenščine v angleščino. Na sliki 26 je prevajalnik besedil myTranslator, ki smo ga uporabili za prevajanje prvega članka iz slovenščine v angleščino.

	Poraba pomnilnika	Čas v sekundi	Neprevedene	Približno prevedene	Napačno prevedene	Pravilno prevedene	V odstotkih pravih prevod
myTranslator	1,79 MB	3	0	16	0	75	82%
Google pre.	9,59 MB	4	0	4	6	81	89%
Bing pre.	8 MB	6	0	6	6	79	86%

Tabela 1: Primerjava prevoda prvega članka z myTranslator-jem, Google prevajalnikom in Bing prevajalnikom.

Prvi članek je napisan v slovenščini. Govori o Ljubljani in Bledu [18]. Ima skupno 91 besed. Prevedli smo ga iz slovenščine v angleščino s tremi prevajalniki besedil. Primerjava pri prevajanju besedil je narejena v tabeli 1.

	Poraba pomnilnika	Čas v sekundi	Neprevedene	Približno prevedene	Napačno prevedene	Pravilno prevedene	V odstotkih pravih prevod
myTranslator	1,79 MB	3	0	19	0	71	78%
Google pre.	9,59 MB	5	2	11	0	77	85%
Bing pre.	8 MB	7	2	12	2	74	82%

Tabela 2: Primerjava prevoda drugega članka z myTranslator-jem, Google prevajalnikom in Bing prevajalnikom.

Drugi izbrani članek je bil napisan v angleščini s skupno 90 besedami in smo ga prevedli v slovenščino s tremi prevajalniki besedil. Tema članka je bila Tina Maze in njen novi trener [19]. Tudi drugi članek je bil približno preveden, vseboval je napake pri vseh treh prevajalnikih. Pri prevajanju drugega članka se je pokazala slabost Bingovega prevajalnika, ki je prevedel osebno ime Maze v Labirint. Ostala dva prevajalnika besedil sta prepoznala osebno ime in ga nista prevedla. Natančna primerjava je vidna v tabeli 2.

	Poraba pomnilnika	Čas v sekundi	Neprevedene	Približno prevedene	Napačno prevedene	Pravilno prevedene	V odstotkih pravilen prevod
myTranslator	1,79 MB	3	0	18	0	53	74
Google pre.	9,59 MB	6	0	8	3	60	85%
Bing pre.	8 MB	7	2	6	4	59	83%

Tabela 3: Primerjava prevoda tretjega članka z myTranslator-jem, Google prevajalnikom in Bing prevajalnikom.

Zadnji članek govori o turističnem mestu Antalija [20], ki je napisan v turškem jeziku s skupno 71 besedami. Tudi zadnji članek je bil preveden s tremi prevajalniki. Dobljen prevod je bil pri vseh treh prevajalnikih besedil približno prav, opazne so bile manjše napake. Čas in analiza besed prevajanja je vidna v tabeli 3.

	Poraba pomnilnika	Čas v sekundi	Neprevedene	Približno prevedene	Napačno prevedene	Pravilno prevedene	V odstotkih pravilen prevod
myTranslator	1,79 MB	9	0	53	0	199	78%
Google pre.	9,59 MB	15	2	23	9	218	86%
Bing pre.	8 MB	20	4	24	12	212	84%

Tabela 4: Primerjava prevodov vseh treh člankov z myTranslator-jem, Google prevajalnikom in Bing prevajalnikom.

V tabeli 4 je narejena primerjava vseh treh člankov. Iz primerjave je vidno, da je myTranslator hitrejši od ostalih dveh prevajalnikov besedil. Hitrejše delovanje mu omogoča lokalna baza, ki ne potrebuje povezave z internetom. Naša aplikacija je vsebovala tri jezike po 12 000 besed in je bila velika samo 1.79 MB, kar pomeni, da je bila do sedemkrat manjša od ostalih prevajalnikov besedil. Boljši rezultati so se pokazali pri neprevedenih in napačno prevedenih besedah. Slabost, ki se je pokazala pri našem prevajalniku besedil, je, da so prevedene besede prevedene le približno. Nismo pa upoštevali sklonov, časa, števila in spolov.

Pametnih telefonov ne uporabljamo samo za pogovor, ampak tudi kot fotoaparata, mp3 prevajalnik, kot napravo za igranje iger, za branje elektronske pošte in ostalo, kar uporablja spomin pri telefonu. Ena od največjih težav pri pametnih telefonih je spomin. Na trgu so starejši pametni telefoni, ki imajo manjši spomin kot novejši, ki pa so tudi spominsko omejeni. Aplikacija, ki omogoča dodajanje novih jezikov in dodajanje besed z interneta, omogoča varčevanje s spominom, ki je omejen. Pri primerjavi se je pokazalo, da Google prevajalnik, če vsebuje dva jezika za uporabo brez internetne povezave, zasede 238 MB. Naša aplikacija pri dveh jezikih z 12 000 besedami potrebuje samo 1,61 MB prostora. Zavedamo se, da ima Google prevajalnik petdesetkrat več besed, ampak dve tretjini nikoli ne bomo uporabili.

Google prevajalnik je eden od najbogatejših z besedami in najbolj popularen. Tudi tako velik prevajalnik ne vsebuje vseh besed, ki obstajajo, ali pa pri prevajanju uporablja napačne metode. Na takšne napake uporabnik sploh ni opozorjen, da nekatere besede sploh niso prevedene. Takšno slabost smo ugotovili in jo pri naši aplikaciji popravili. Za rešitev smo uporabili barvanje besed. Besede, ki ne obstajajo v bazi, ki niso prevedene, smo pobarvali z rdečo barvo, besede, ki so bile približno prevedene, smo označili z modro barvo, ostali tekst, ki je bil preveden, pa s črno barvo. Rešitev z barvanjem besed uporabniku omogoča lažje opazovanje, katera beseda ni bila prevedena ali katera beseda je bila približno prevedena.

4.2 Izboljšava in mnenje uporabnika

Aplikacija je še v razvoju. Da bo konkurenčna podobnim aplikacijam, bi bilo smiselno dodati še nekaj funkcij, ki bodo olajšale delo s prevajanjem. Opisali bomo, katere lastnosti bi bilo treba še dodati v prihodnosti in kakšno je bilo mnenje uporabnika.

V današnjem času so za komuniciranje najpogosteje uporabljeni elektronska pošta in kratka sporočila (ang. SMS). Uporabno bi bilo, če bi aplikacija omogočala povezavo do naše elektronske pošte ali do kratkih sporočil in bi jih lahko prevedla iz napisanega jezika v izbrani

jezik. Uporabniku takšna aplikacija olajša delo z nerazumljivo elektronsko pošto in kratkim sporočilom. Ne bi bilo treba kopirati besed in jih prenašati v prevajalnik.

Pravilo, ki pravi: »Govori, kakor pišeš, in kakor pišeš, tako beri!« ne drži za vse jezike. Preveč jezikov je, ki se izgovarjajo drugače, kot so napisani, zato je dobro, da aplikacija podpira govor. Prepoznavanje govora in izgovarjava besed namesto nas bi bila uporabna, ko se pogovarjamo z osebo, katere jezika ne razumemo. Takšna lastnost bi omogočila natančno izgovarjanje besed in boljšo komunikacijo v različnih jezikih.

Aplikacija trenutno vsebuje 20 različnih jezikov, kar pomeni, da manjka še več kot petdeset jezikov, da bi vsebovala vse svetovne jezike. Pri dodajanju novih jezikov moramo te vsaj malo poznati, kar pomeni, da trenutno nismo za to usposobljeni.

Najbolj uporabna lastnost, ki jo moramo dodati, je branje teksta s pomočjo kamere in prevajanje. Tekst, ki je napisan na papirju, preko kamere lahko preberemo in ga dodamo znotraj aplikacije, da se bo prevajal v izbranem jeziku. Branje teksta s kamero bo olajšalo delo, da ne bo treba več pisovati strani, ki jih moram prevajati.

Mnenje uporabnika o aplikaciji je bilo pozitivno. Za preverjanje mnenja uporabnika je bil narejen api in poslan desetim uporabnikom, ki so opisali svoje mnenje kot pozitivno in izrazili zadovoljstvo z aplikacijo. Eni uporabniki so uporabili aplikacijo na potovanju, drugi za študij kot pripomoček za prevajanje besedil. Komentirali so, da bi bilo bolje, če bi imela aplikacija več besed na začetku. Lepo bi bilo, če bi bili na začetku dodani tudi nemščina in italijanščina. Pri vsakem jeziku je lepo, če je dodana še zastava vsake države. To so bile opombe, ki se jih da upoštevati. Aplikacija jim je bila všeč, ker je majhna in ne zasede preveč prostora, ima možnost dodajanja novega jezika in novih besed.

4.3 Testiranje

Pri razvijanju aplikacije je pomembno testiranje. Vsaka napaka pred končnim razvitjem aplikacije lahko vzame veliko časa in denarja, kar na koncu ne bi bilo dobro ne za razvijalca ne za uporabnika. Testiranje aplikacije v našem primeru pomeni predvsem testiranje kode. Testiranje kode moramo opravljati vedno pri dodajanju nove funkcije aplikaciji. S testiranjem ne bomo odkrili vseh napak, ki so nastale v času razvoja, lahko pa jih zmanjšamo. Po uspešnem testiranju lahko rečemo, da aplikacija dela to, kar smo pričakovali.

Pri razvoju naše aplikacije smo po zaključku vsake funkcije testirali ali ta pravilno deluje. Med razvijanjem aplikacije smo testiranje izvajali sami. Vnašali smo različne besede, uporabljali smo več tabel, dodajali več jezikov itd. Med razvojem smo jo testirali na

emulatorju, ko smo jo dokončali, smo jo testirali na Sony Xperia Z, ki uporablja operacijski sistem android različice 4.4.2 in pri Samsungu SII Plus, ki uporablja operacijski sistem android različice 4.1.2. Izkazalo se je, da odlično deluje.

Končno aplikacijo smo poslali desetim uporabnikom, da bi jo testirali in povedali svoje mnenje. Uporabniki niso bili vključeni v testiranja v času razvoja, tako da niso nič vedli o sami aplikaciji, obveščeni so bili le o namenu uporabe. Testirali so vse funkcije aplikacije in izkazalo se je, da aplikacija deluje pravilno. Toda tudi po takšnem testiranju ne moremo zagotoviti, da aplikacija ne vsebuje nobene napake.

Poglavje 5

Sklepne ugotovitve

Cilj diplomskega dela je bil narediti mobilno aplikacijo za platformo android, ki bo prevajala besede v več jezikov. Na začetku smo na kratko opisali tehnologijo in orodja, ki smo jih uporabili pri razvoju aplikacije. Potem smo opisali, kako smo načrtovali razvoj, kako se je aplikacija razvijala, na katere težave smo naleteli med razvojem in katere rešitve smo uporabili.

Na koncu smo dobili preprosto in hitro delujočo aplikacijo za prevajanje med različnimi jeziki. Aplikacija, ki smo jo uspešno zaključili, je karakteristična po tem, da uporabniku omogoča dodajanje novih jezikov in dodajanje novih besed s spletnih strani. Po dodajanju jezika in novih besed aplikacijo lahko uporabljamo brez internetne povezave, kar pomeni, da jo lahko uporabljamo kjer koli in kadar koli. Takšne lastnosti uporabniku omogočajo bolj uporaben in hitrejši prevajalnik besedil, ki ne vzame preveč prostora na pametnih telefonih. Na trgu je kar nekaj podobnih aplikacij, ki imajo različne funkcionalnosti. Da bo naša aplikacija konkurenčna ostalim, jo moramo še naprej razvijati. Opisali smo, kakšne funkcionalnosti ima naša aplikacija in v čem se razlikuje od ostalih podobnih aplikacij. Aplikacija je bila testirana med razvojem in na koncu in se je izkazala za delujočo. Z delovanjem aplikacije in njeno uporabnostjo smo zadovoljni. Tudi uporabniki, ki so jo testirali, so izrazili pozitivno mnenje. Predvideli smo in opisali tudi izboljšave naše aplikacije, ki smo jih kot pripombe dobili od uporabnikov. Dokončana aplikacija lahko uporabniku služi kot pripomoček za prevajanje besed med različnimi jeziki v vsakodnevem življenju.

Literatura in viri

- [1] Strojno prevajanje, *Wikipedija*, (2013). [Online]. Dosegljivo: http://sl.wikipedia.org/wiki/Strojno_prevajanje.
- [2] History of MT in a nutshell, (2005). [Online]. Dosegljivo: <http://www.hutchinsweb.me.uk/Nutshell-2005.pdf>.
- [3] Amebis Presis, *AMEBIS*, (2012). [Online]. Dosegljivo: <http://presis.amebis.si/>.
- [4] O Google Prevajalniku, *Google*, (2014). [Online]. Dosegljivo: http://translate.google.com/about/intl/sl_ALL/.
- [5] Bing Translator, *Microsoft*, (2014). [Online]. Dosegljivo: <http://www.microsoft.com/translator/products.aspx>.
- [6] Prevajalnik.net, Prevajalnik, 2014. [Online]. Dosegljivo: <http://www.prevajalnik.net/>.
- [7] Statistično strojno prevajanje, 2014. [Online]. Dosegljivo: http://sl.wikipedia.org/wiki/Statisti%C4%8Dno_strojno_prevajanje.
- [8] About SQLite, SQLite, 2014. [Online]. Dosegljivo: <http://www.sqlite.org/about.html>.
- [9] M. Podlesnik, Kaj je Android, 2011. [Online]. Dosegljivo: <http://slo-android.si/prispevki/kaj-je-android.html>.
- [10] M. Rouse, What is XML, TechTarget, 2013. [Online]. Dosegljivo: <http://searchsoa.techtarget.com/definition/XML>.
- [11] Razvoj programov v okolju Eclipse, [Online]. Available: <http://hpc.fs.uni-lj.si/eclipse>.
- [12] Eclipse Subversion, eclipse, 2014. [Online]. Dosegljivo: <http://www.eclipse.org/subversive/>.
- [13] Working with Eclipse and Git, Magnolia community, 2013. [Online]. Dosegljivo: <https://wiki.magnolia-cms.com/display/DEV/Working+with+Eclipse+and+Git>.
- [14] GForge, 2014. [Online]. Dosegljivo: <https://gforge.com/gf/#inside>.
- [15] Günde kaç kelimeyle konuşuyorsunuz?, KIGEM, 2014. [Online]. Dosegljivo: <http://www.kigem.com/gunde-kac-kelimeyle-konusuyorsunuz.html>.

- [16] Read books online, 2011. [Online]. Dosegljivo: <http://readfreeonline.net/>.
- [17] G. Android, What is AysncTask in Android?, 2013. [Online]. Dosegljivo: <http://programnerguru.com/android-tutorial/what-is-async-task-in-android>.
- [18] Ljubljana in Bled - mesti za zaljubljene, RTV SLO, 2014. [Online]. Dosegljivo: <http://www.rtv slo.si/tureavanture/evropa/ljubljana-in-bled-mesti-za-zaljubljene/349686>.
- [19] Tina Maze Starts New Season with New Coach, The Slovenia Times, 2014. [Online]. Dosegljivo: <http://www.sloveniatimes.com/tina-maze-starts-new-season-with-new-coach>.
- [20] Antalya Hakinda Kisa Bilgi, 2014. [Online]. Dosegljivo: <http://www.izafet.net/threads/antalya-hakkinda-kisa-bilgi.695191/>.