

Универзитет у Нишу
Електронски факултет Ниш

Интелигентни системи
Пети домаћи задатак – Обрада природних језика

Студент:
Димитрије Јовић, 928

Ниш, фебруар 2020. год.

С А Д Р Ж А Ј

1. Формулација проблема	3
2. Имплементација <i>Swarm</i> алгоритма.....	4
3. Референце.....	Error! Bookmark not defined.

1. ФОРМУЛАЦИЈА ПРОБЛЕМА

Као велики заљубљеник у криптографију, наш јунак Мирко, је читајући и истраживајући на интернету наишао на веома занимљив задатак из криптографије. Задатак је поставио један угледни професор. Наиме, цео смисао задатка се састојао у томе, да је професор желео да покаже свим заљубљеницима у криптографију, да је могуће уз познавање неког природног језика, односно познавања фреквенци појављивања слова у том језику, тј. која се слова у том језику највише употребљавају, могуће кодирати неки текст, и да је притом тешко декодирати га. С тога је у оквиру самог задатка поставио неколико фајлова са кодираним текстовима, као и фајлове са фреквенцама појављивања одређених слова. Циљ задатка је био да се дати фајлови дешифрију. Након неког времена, и уз малу помоћ својих другова, Мирко је успео да дешифрије дати текст.

Iš učđžr csl Tđ ělurćr pđSđjl, sđcbl ěr tN uđ ěrdr šrhNudNsN nprěr udNbr jrcđž Tzjrbr. Jđp lěr fzžr, lěl ěztłbl zštzađjTđ ězAlějl błTđ uđ npđsěrpr lěhrA z sđđdujN nlbpđs (jNTđ jN nlspđtjl jrnłhNjTrsN ěr Nhr dTzěN błTN hlžž tNsN jr ěpAzjvz lSrTrjTr NdN prělusN, r ěr Nh jN dNvd, jN sđdl, jđ lěr jN jrThrjTN sprž; sN dTzěN uz brł TrTđ Nš blžr jNTđ npłbdTzěNdł nNdd: unldTr uđ jđ ěNěN eNěłs błTN zjzspr bdTzSr) slđNbl Tđ pđsbl z lěłh ěěłdNSjłh ěđbz, ěr uđ Slěđb nNsr brbl Tđ hlžžBđ ěr srbrě hđrě Slěđb zlncsđ nlusłTN. Oěžlěp Tđ, ěr uhl hN šdhđTr žěđ uz plhrjsNSjN dTzěN zěđb eNěđđN N žěđ Bđ ěrdTěr zěđb eNěđsN.

opđtr urhl nredTNěł nprsnN ěłžrarTđ N sNnlěđ z jru: zěđb Tđ blě jru nlusłTrł Tđěrj sNn dTzěN (nlpđě lJNA łtNSJNA, "spđšěđjNA", "pđrdjNA") N TđěNjl csl uđ hđjTrđł, sl Tđ łtdNb z blhđ uđ srT sNn nłTrěđTNěrl. kđbrěr Tđ Nžprł zdlžž ěrTpljr, NdN Grhtđsđ, Ncrł brł blhNsr z Jzejz mptNTz, nludđ slžr šr brfrjubNł usldłh zdlžđ jđuAěrBđjNA žđjNTr... mpđBjl ěłtr! iNdł Tđ srěr šr jTđžr zšłpr, Něłdr, łłžłěr, Něđrdr ur ěđđNbNł I. LNb z blhđ uđ ěrjru srT dNb TrěđTr jđupđBjNTN Tđ Nrbl Bđ jđbl sđcbl z ěrjrcjTđh jTđžłěłh łtdNSTz nlšjrsN jđbrěrcjTđž unjTrđNvz N ěrTpljr, ur zšěruNhr N nlžđđěłh znpsNł z ěrdTNjz: ěrjru uđ lJ nłTrěđTzTđ brł Sdrj šđrsjđ hđrěđeN, cNplbl N rdTbrěł lěđđł šrhđjTđjl Tđ jrThłěđpJNTNłh bplTđh, jđbrěrcjTr ěđđNbr blur Tđ tpNedTNěł šrSđcdTrjr, jrc urěrcjTN plhrjsNb Něđ nł ězpděNhr, Nžpr, ěłfNpr, luěrTr, jđhr jNbrěěNA jršłpr, jđ sđeN jNSđhz uđh hłěr zundAz z ěpzcšž NdN jršNěž crhnłljr, zSđusěžTđ z ěđđNbNł unłpsubNł ěłžrarTNhr, dđsN, nprěN nlšjrnjusěr N šrtłprěđTr NA, žłěłpN parěł ěNěđ TđšNbr, jđ SNSr jNcsr, šěprě Tđ, tzTrj, jrněrS, uTrTrj, Nžpr nł jđbN nzs brpsđ šr jlěrv, jđ nlšjrtđ npNTrsđđTusěł, z dTztrěN ěNěN urhl fNšNSbz usprjz, jđ ludBr Nšžđđěr jNcsr šr šdhđTz, z jrTěNěđ udzSrTđěr jđ prěN jNcsr rbl uđ jđ trěN "nludłěNhr". gđN Nrbl jr npěN nlžđđě pršđNSrj, nłsnzjl ěpzžłTrSN, lJ Tđ z lułłěN uěłłTłt tprs pladjN łjlhđ ur ěžłłh blulh: brł N łjrT lě TzSđ N lěrT lě ěrjru Nhr uěłTđ sđlpNTđ błTđ uđ uzbltdTrěrTz ur eNěłslh; brł N łjrT npđ jTđžr N lěrT uđ uěrbN Sru uzěrpr ur łtNSrTNhr, ur usěrpjlcBz, ur SNjTđjNvrhr, ur dTzěNhr. mrhl, ěłb Tđ łjrT ur ěžłłh blulh sđeNł, trp z uđtN, Tđějłh zšłpz z dNvz jđblž ěđđNblž ndujNbr, luěrTrSr NdN Tzjrbr Nš plhrjr błTN hz Tđ nłhržrł ěr shzpjł nlějluN zěrpđ, ěłsđđ lěrT lě ěrjru Něđ nł uěđsz jđ npNšjrtzBN jNblžr, jđ zžđđěrTzBN uđ jN jr blžr, jlěđj hrsNvlh NdN uěłžr uđrěłdTztdTr, NdN uěłTđ bpěN, NdN uěłTđ łbldNjđ, nršđBN urhl ěr hz vNndđ zěđb tzěz uTrTjđ, r hrcjr łshđjl ěđšrjr. I bl Bđ nłpđBN ěr jNTđ ěrjru lěrT unłpsNusr N tđštpNejr dznrSr jruddějNb jđbrěrcjTđž jđuAěrBđjlž žđjNtr brěr uđ ěłtpl złSN jTđžłě nldlerT Nšěrj bljěđjvNtr N łtNSjłž slbr eNěłsr? Esł, nł slh uěłłh nldlerTz, nł slT jđupđadjlusN, nł eđđTN ěr uđ jrhdšjđ ěpzcšž ur uěłTNł uAěrsrjTNhr, lJ Tđ plhrjsNb, "zuNTrjr ždrěr", zěđb uđrěrj ur uěłTNA jrusrpjłusN N nlpđě uěłTNA łshđjNA hrjNpr, tpNedTNěł jlěđjlž lěđdr, hljěđjublž eNěłsr. gđN ěłb Tđ u Tđějđ usprjđ lěrT jlěN uNhnrsNSjNTN, N jr Nšžđđě nlšNsNějNTN, "pđrdjNTN" (ěrdTěr štlž Trbl pršěNTđjlž sđđdujlž N Szđjlž eNěłsr!), u ěpzžđ usprjđ Tđ usprcjNTN, npršjNTN, usł nzs rěđBN jNANDNusr jđžł hr błTN ěrTplj. Jđp jrcN ěrTpljN uz šjrdN hljłł, SNsrdN hljłł, ludBrđN hljłł, eNěđđN TđějNł lSrTjl lězAlěđTđjNł eNěłslh, nlpđě uěNA jrusrpjłusN rdbłAldr, ěłb lěN ěrjru jđ ěđpzTz jN z csr, jđ šrnjNhrTz uđ jN šr csr, jđ šjrTz jNcsr, jđ eđđđ jNcsr Nšěrj "ěrjrcjTđž" N "lěłšđhrđTublž"; vđł Nh uđ eNěłs, vđđł šjrjTđ, npđudđNdł jr blez, jr srT unldTjTN łłłsrS jrcđ łultđ.

Слика 1: Део шифрованог текста

2. ИМПЛЕМЕНТАЦИЈА ОБРАДЕ ПРИРОДНИХ ЈЕЗИКА

Обрада природног језика (*NLP*) је подобласт лингвистике, рачунарства и вештачке интелигенције која се бави интеракцијама између рачунара и људских (природних) језика, посебно како програмирати рачунаре да обрађују и анализирају велике количине података о природном језику. Изазови у обради природног језика често укључују препознавање говора, разумевање природног језика и генерисање природног језика.

Што се тиче самог домаћег, одабран је четврти задатак, код кога је било потребно дешифровати изабрани текст. Да би се у дешифровао изабрани текст, потребно је било обратити пажњу на неколико ствари. На почетку било је потребно уклонити из текста, све бројеве, специјалне карактере, знакове интерпункције, јер у самом тексту задатка стоји да се вероватноћа појављивања рачуна без тих карактера. Након тога, било је потребно одредити вероватноћу појављивања сваког слова у тексту, одређивање вероватноће је веома просто, одреди се број понављања слова у тексту, и подели се са укупним бројем слова. Након тога, било је потребно из тако одређених вероватноћа, отклонити све вероватноће за које се у оквиру фајла, који садржи вероватноће појављивања слова у нешифрованом тексту, не појављују одређена слова.

Када се изврши уклањање слова, потребно је упарити слова по вероватноћи појављивања, односно, упарити слова по најсличнијим вероватноћама из шифрованог текста и из нешифрованог текста. Након тога, могућа је замена слова, након чега се добија дешифровани текст. Међутим, јавио се проблем приликом замене слова, где се већ промењени карактер, поново мења неким другим. С тога је било потребно пре саме замене слова, одредити индексе где се слова која се мењају налазе, како не би дошло до дуплог мењања.

Iz svega što je dosada rečeno, teško da bi se dala zamisliti prava slika našeg junaka. Jer ova fuga, ovo duboko uzbuđenje duhovno koje se pretvara odmah u telesni pokret (nije ni potrebno napominjati da ima ljudi koji mogu biti na vrhuncu očajanja ili radosti, a da im ni lice, ni telo, ne oda ni najmanji trag; ti ljudi su kao jaje iz koga nije proključivalo pile: spolja se ne vidi život koji unutra ključa) toliko je retko u ovom dvoiličnom veku, da se čovek pita kako je moguće da takav mlad čovek uopšte postoji. Odgovor je, da smo mi zemlja gde su romantični ljudi uvek živeli i gde će valjda uvek živeti.

Treba samo pažljivo pratiti događaje i tipove u nas: uvek je kod nas postojao jedan tip ljudi (pored onih običnih, "trezvenih", "realnih") i jedino što se menjalo, to je oblik u kome se taj tip pojavljivao. Nekada je igrao ulogu Bajrona, ili Gambete, išao kao komita u Južnu Srbiju, posle toga za kafanskim stolom uloge neshvaćenih genija... Srećno doba! Bilo je tada za njega uzora, idola, bogova, ideala sa velikim I. Lik u kome se danas taj lik javlja nesrećniji je iako će neko teško u današnjem njegovom obličju poznati nekadašnjeg sanjalicu i Bajrona, sa uzdasima i pogledom uprtim u daljinu: danas se on pojavljuje kao član zlatne mladeži, široko i aljkavo odelo zamenjeno je najmodernijim krojem, nekadašnja velika kosa je brižljivo začesljana, naš sadašnji romantik ide po žurevima, igra, šofira, osvaja, nema nikakvih nazora, ne teži ničemu sem možda uspehu u društvu ili nazivu šampiona, učestvuje u velikim sportskim događajima, leti, pravi poznanstva i zaboravlja ih, govori rdavo više jezika, ne čita ništa, zdrav je, bujan, napadač, sjajan, igra po neki put karte za novac, ne poznaje prijateljstvo, u ljubavi vidi samo fizičku stranu, ne oseća izgleda ništa za zemlju, u najviše slučajeva ne radi ništa ako se ne bavi "poslovima". Ali iako na prvi pogled različan, potpuno drugojači, on je u osnovi svojoj brat rođeni onome sa dugom kosom: kao i onaj od juče i ovaj od danas ima svoje teorije koje se sukobljavaju sa životom; kao i onaj pre njega i ovaj se svaki čas sudara sa običajima, sa stvarnošću, sa činjenicama, sa ljudima. Samo, dok je onaj sa dugom kosom težio, bar u sebi, jednom uzoru u licu nekog velikog pesnika, osvajača ili junaka iz romana koji mu je pomagao da tmurno podnosi udare, dotle ovaj od danas ide po svetu ne priznajući nikoga, ne ugledajući se ni na koga, nošen maticom ili svoga slavloljublja, ili svoje krvi, ili svoje okoline, pazeći samo da mu cipele uvek budu sjajne, a mašina otmeno vezana. I ko će poreći da nije danas ovaj sportista i bezbrižna lupča naslednik nekadašnjeg neshvaćenog genija kada se dobro uoči njegov položaj izvan konvencija i običnog toka života? Eto, po tom svom položaju, po toj nesređenosti, po želji da se nametne društvu sa svojim shvatanjima, on je romantik, "usijana glava", uvek slavan sa svojih nastranosti i pored svojih otmenih manira, brižljivo nošenog odela, mondenskog života. Ali dok je s jedne strane ovaj novi simpatičniji, i na izgled pozitivniji, "realniji" (valjda zbog jako razvijenog telesnog i čulnog života!), s druge strane je strašniji, prazniji, sto puta veći nihilista nego ma koji Bajron. Jer naši Bajroni su znali mnogo, čitali mnogo, osećali mnogo, živeli jednim očajno oduhovljenim životom, pored svih nastranosti alkohola, dok ovi danas ne veruju ni u šta, ne zanimaju se ni za šta, ne znaju ništa, ne žele ništa izvan "današnjeg" i "ovozemaljskog"; ceo im se život, celo znanje, preselilo na kožu, na taj spoljnji omotač naše osobe.

Слика 3: Део дешифрованог текста

3. РЕФЕРЕНЦЕ

- [1] <https://numpy.org/doc/1.18/reference/>
- [2] <https://docs.python.org/2/library/collections.html>
- [3] <https://docs.python.org/3/library/re.html>