Predlog projekta iz SIAP-a

Ovaj dokument sadrži opis teme i definicije projekta, kao i motivaciju za odabranu temu. Pored toga, u ovom dokumentu dat je i kratak pregled radova na sličnu temu, kao i skup podataka i softver koji će biti upotrebljen za izradu projekta.

Tema projekta je generisanje kuhinjskog recepta na osnovu datih sastojaka. Ulaz u sistem predstavlja niz sastojaka (koje korisnik ima na raspologanju), od kojih se bira podskup koji će biti upotrebljen za recept. Zatim se na osnovu odabranih sastojaka generiše niz instrukcija za pripremu recepta, a na kraju se na osnovu sastojaka i teksta generiše slika koja prikazuje kako bi gotovo jelo trebalo da izgleda.

# Definicija projekta

Generisanje recepta na osnovu datih sastojaka. Potrebno je izabrati podskup sastojaka tražeći koji sastojci najbolje idu jedni uz druge na osnovu recepata iz skupa podataka. Sledeći zadatak je kreirati model koji će na osnovu odabranih sastojaka generisati set instrukcija za pripremu jela u obliku teksta. Poslednji zadatak je napraviti model koji na osnovu sastojaka i instukcija generiše sliku gotovog jela. Ideja projekta je da se na osnovu proizvoljnog skupa sastojaka generiše kompletan recept za pripremu.

# Motivacija

Prosečan čovek uglavnom zna da pripremi nekolicinu različitih jela, što dovodi do izuzetno monotone, ponekad i dosadne, ishrane. Da bi neka osoba pripremila za sebe nepoznato jelo, pored toga što je potrebno da istraži način njegove pripreme, potrebno je i da nabavi odgovarajuće sastojke, što oduzima vreme i motivaciju za takvim poduhvatom. Ciljano rešenje predviđeno je da se koristi otvaranjem frižidera i plakara, i unošenjem namirnica koje se nalaze unutra kako bi se dobila ideja, i time donele novosti u kuvanje prilikom svakodnevnog života. Nada je da će ovaj sistem dati motivaciju za kuvanjem raznovrsnim predlozima i novim idejama, kao i probuditi inspiraciju za istraživanje drevne umetnosti pripreme hrane, možda čak i stvaranjem nečeg što do sada nije viđeno, a sve to bez napuštanja sigurnosti doma.

# Pregled radova na sličnu temu

|  |  |
| --- | --- |
|  |  |
| [1] Salvador, A., Hynes, N., Aytar, Y., Marin, J., Ofli, F., Weber, I. and Torralba, A., 2017. Learning cross-modal embeddings for cooking recipes and food images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3020-3028). |  |
| Vancouver |  |

**Zadatak rada**: Pronaći *embedding* koji kombinuje sastojke, instrukcije i sliku recepta u jedinstvenu reprezentaciju. Zatim za sliku pronaći odgovarajuci recept iz skupa postojecih recepata (*im2recipe*).

**Metodologije**: U ovom radu sastojci su predstavljeni pomoću *word2vec-a*. Reprezentacija recepta predstavlja kombinaciju *embedding-a:* Od skupa sastojaka napravljen je *embedding* pomoću bidirekcionog LSTM-a, od instrukcija pomoću dvostruke LSTM mreže, a od slika upotrebnom vgg-16 i resnet-50 modela.

**Skup podataka**: Recipe1m

**Evaluacija**: Evaluacija im2recipe tako što daju čoveku sliku i 10 recepata, i provere da li on bolje izabere recept nego model. Model uglavnom radi slično ili bolje.

**Relevantno za ovaj rad**: Efikasan način za reprezentaciju instrukcija i recepata.

**Razlike**: U našem radu slika neće biti korišćena za reprezentaciju recepta. Pored toga, ovaj rad se bavi pronalaženjem recepta na osnovu slike, dok se u našem radu slika generiše na osnovu recepta, i recept na osnovu sastojaka.

[2] H. Lee, H., Shu, K., Achananuparp, P., Prasetyo, P. K., Liu, Y., Lim, E. P., & Varshney, L. R. (2020, April). RecipeGPT: Generative Pre-training Based Cooking Recipe Generation and Evaluation System. In *Companion Proceedings of the Web Conference 2020* (pp. 181-184).

**Zadatak rada**: Generisati recepte kao listu instrukcija iz liste sastojaka i generisati listu sastojaka iz instrukcija i naslova recepta.

**Metodologije**: Generisanje liste instrukcija rađeno je prilagođavanjem [OpenAI GPT-2](https://openai.com/blog/gpt-2-1-5b-release/) neuronske mreže, koja za ulaz ima spoj naslova i liste sastojaka.

**Skup podataka**: Recipe1m

**Evaluacija**: korišćene su BLEU i ROUGE metode evaluacije (standard n-gram overlap based metrics), kao i NTED (normalized tree edit distance) za proveru kvaliteta generisanih instrukcija u odnosu na *ground truth.*

**Relevantno za ovaj rad**: Metodologija genersanja instrukcija iz sastojaka i metodologija razdvajanja sastojaka na količinu, meru i sam sastojak. Pored toga, rezultati ovog rada su korisni kao referenca za poređenje.

**Razlike**: Arhitektura neuronske mreže za generisanje liste instrukcija.

[3] El, O. B., Licht, O., & Yosephian, N. (2019). GILT: Generating images from long text. *arXiv preprint arXiv:1901.02404*.

**Zadatak rada**: Generisati sliku gotovog obroka iz liste sastojaka i liste instrukcija recepta

**Metodologije**: Za generisanje slike u ovom radu se koristi *stack* GAN, koji se sastoji iz nekoliko generatora i diskriminatora organizovanih kao stablo, sa ulazom koji pored šuma ima i listu sastojaka spojenu sa listom instrukcija.

**skup podataka**: Recipe1m

**Evaluacija**: IS - Inspection-Score, u radu opisan kao suboptimalan. Generisana slika se klasifikuje (u klasu *ImageNet*-a) i očekuje se što veća vrednost IS.

Kao drugu meru ovaj rad se oslanja na ljudsku evaluaciju, gde učesnici u evaluaciji određuju snagu relacije izmedju generisane slike i recepta, generisane slike i prave slike, i u kojoj meri generisana slika predstavlja stvarnu hranu.

**Relevantno za ovaj rad**: Metodologija generisanja slike iz teksta.

**Razlike**: U našem radu biće korišćena jednostavnija cGAN arhitektura od arhitekture iz ovog rada.

# Skupovi podataka

Za ovu temu pronađeno je dva istaknuta skupa podataka. Za ovaj rad, relevantna obeležja skupa podatak su naslov recepta, skup sastojaka, instrukcije i slika.

Prvi skup podataka, [recipe1m+](http://pic2recipe.csail.mit.edu/), sadrži oko milion recepata i 13 miliona slika. Recepti su predstavljeni kao lista sastojaka i lista tekstualnih instrukcija. Ovo je najveći i najpopularniji skup podataka na ovu temu, sa preko 300 asociranih radova i preko 200 citata. Nažalost, ovaj skup nije direktno dostupan za preuzimanje sa sajta zbog tehničkih problema autora, ali su autori naveli da ga dele na zahtev putem *email*-a. Ukoliko se dobije pristup ovo skupu podataka, biće korišćen u ovom radu.

Drugi skup podataka je [Recipe box](https://eightportions.com/datasets/Recipes/?fbclid=IwAR3jYCBoyEx_76nDWt1bR8tiDduWcSKswrcSClwUbJpVryJtfo6dpoantMA), koji sadrži oko 125.000 recepata sa listom sastojaka i instrukcijama u tekstulanoj reprezentaciji, od kojih oko 70.000 imaju pridodatu i sliku. Iako znatno manji od *recipe1m+*, trebalo bi da je dovoljan za potrebe ovog projekta i biće korišćen kao rezerva.

# Metodologija

U ovom poglavlju navedene su najvažniji delovi projekta i njihova metodologija.

Razdvajanje sastojaka na količinu, meru i sastojak – U skupu podataka sastojci su predstavljeni u tekstualnoj formi (npr. ½ cup milk), koje je potrebno razdvojiti na tri navedena dela (1/2, cup, milk). U ovu svrhu biće isprobane dve metode od kojih će biti odabrana ona koja se pokaže uspešnijom: [*Stanford Stanza*](https://stanfordnlp.github.io/stanza/) (*POS tagging*) sa razdvajanjem na osnovu tipa reči, i [*ingredient-phrase tagger*](https://github.com/nytimes/ingredient-phrase-tagger)*. Stanford Stanza* je *pipeline* neuronskih mreža za NLP, dostupnih pomoću *python* alata. Stanza POS koristi bidirekcionu LSTM mrežu u osnovi, kako bi reči ulaznog teksta označio labelom uloge u rečenici i morfoloških karakteristika (jednina/množina, glagolsko lice...). Iz dobijenih oznaka i redosleda reči u sastojku može se zaključiti koja reč koji deo predstavlja. *Ingredient-phrase tagger* je *python* biblioteka koja za ulazni sastojak kao izlaz daje parove ključ-vredost koji predstavljaju različite delove sastojka (količina, merna jedinica, naziv sastojka, ostalo, komentar). Osnovni razloga ovog razdavajanja je određivanje naziva sastojka, kako naš model ne bi pravio razliku između dva ista sastojka u različitim količinama (što bi bio slučaj ako bi se posmatrao sirov tekstualni ulaz).   
 Eksplorativna analiza – cilj ovog dela je izvući relevantne osobine skupa podataka koje mogu biti korisne za izdvajanje podskupa sastojaka iz datih proizvoljnih sastojaka. Neke od stvari koje će biti posmatrani su *outlier*-i u sastojcima na osnovu frekventnosti njihovog pojavljivanja, korelacije među sastojicma kao i postojanje duplikata u skupu podataka. Svrha eksporativne analize je poboljšanje rezultata modela koristeći otkrivene osobine skupa podataka.

Izdvajanje podskupa – cilj ovog dela je za dati skup sastojaka izdvojiti podskup sastojaka koji će činiti recept, na osnovu njihovih odnosa zaključenih iz skupa podataka. Za reprezentaciju sastojaka biće korišćen word2vec za svaki sastojak. Potrebno je pronaći adekvatnu metodologiju za izdvajanje podskupa. Neke od metode koje će biti isprobane su klasterovanje pomoću K-means i DBScan algoritama, sa idejom da najveći klaster predstavlja najveći podskup sastojaka koji može da se upotrebi u jednom receptu. K-means je odabran zbog jednostavnosti implementacije i brzine izvršavanje, ali kako ne postoji sigurnost da su klasteri unutar podataka pravilnih oblika, postoji mogućnost da će DBScan davati bolje rezultate.

Generisanje instrukcija – cilj ovog dela projekta je generisati set instrukcija recepta na osnovu liste sastojaka. Lista sastojaka biće predsavljena kao prosečna vrednost vektora sastojaka koje sadrži. Za generisanje instukcija biće upotrebljena dvostruka LSTM mreža trenirana nad instukcijama i sastojcima iz skupa podataka, gde su ulaz sastojci, a izlaz set instrukcija.

Generisanje slike - generisanje slike gotovog obroka radiće se na osnovu generisanog seta instrukcija, zajedno sa listom sastojaka, reprezentovanih kao *doc2vec*, pomoću cGAN neuronske mreže, koja za ulaz ima listu sastojaka i listu instukcija a kao izlaz sliku, sa podskupom od oko 100.000 recepata za koje su date slike. Ukoliko rezultati generisanja instukcija ne budu zadovoljavajući, moguće je generisati sliku pronalaskom instukcija iz skupa podataka za recept sa najsličnijim sastojcima kao ulazni sastojci.

Evaluacija: ne postoji standardni način evaluacije ni za jedan deo projekta, tako da će se ovaj rad oslanjati na ljudsku evaluaciju, uz eventualnu pomoć nekih od metoda evaluacije iz navedene relevantne literature. Podaci će biti podeljeni na trening i test skup, a empirijska evaluacija će biti vršena nad testnim skupom. Za neke od korišćenih modela postoje standardne metode evaluacije, tako da će KNN i DBScan biti evaluirane na osnovu prosečne mere sličnosti unutar dobijenih klastera. LSTM mreža za generisanje instrukcija biće evaluriana poređenjem sa ciljanim tekstom pomoću *BLEU* algoritma. Poređenje *Stanford Stanza* i *ingredient phrase tagger* modela biće rađeno na osnovu procenta uspešno razdovjenih sastojaka u odnosu na ukupan broj stastojaka za svaki model.

# Plan rada

* Pronaći odgovarajući način za formatiranje i skladištenje podataka
* Razdvojiti sastojke na količinu, meru i sastojak
* Analizirati podatke i zabeležiti relevantna saznanja
* Razviti model za izdvajanje podskupa sastojaka
* Razviti model za generisanje instrukcija recepta
* Razviti model za generisanje slike

# Tim

R2-19/2020 Nenad MIšić R2-20/2020 Dušan Milunović

R2-21/2020 Branislav Anđelić