

МОЛБА
ЗА ОДОБРАВАЊЕ ТЕМЕ МАСТЕР РАДА

Молим да се одобри израда мастер рада под насловом:

„Класификација текста према сентименту применом метода машинског учења”

Значај теме и области:

Живимо у времену експоненцијалног раста количине доступних информација од којих је већина записана у текстуалном формату. Приликом доношења одлука у пословном свету, али и у свакодневном животу, људи се често ослањају на мишљење и искуства других људи који су своје ставове јавно изнели путем социјалних мрежа, блогова, форума и слично. У циљу ефикаснијег добијања информација из велике количине текстуалних докумената, корисно је извршити њихову класификацију према сентименту на класе докумената са исказаним позитивним, негативним или неутралним ставом.

Специфични циљ рада:

Циљ овог рада је примена различитих метода машинског учења у решавању проблема класификације докумената према сентименту на српском и енглеском језику. Биће имплементирани различити начини представљања докумената: врећа речи, вектор униграма и биграма речи и њихова комбинација, вектор врста речи, вектор позиција речи у тексту, као и Word2Vec модел. Претходно ће бити извршено препроцесирање текста које подразумева чишћење текста (уклањање стоп речи и интерпункцијских знакова) и трансформацију текста (лематизација или стемовање). Од метода машинског учења биће коришћене потпуно повезане неуронске мреже, метод потпорних вектора и наивни Бајесов класификатор. Класификација текстуалних докумената према сентименту ће бити извршена над јавно доступним корпусима филмских рецензија на српском и енглеском језику. Биће извршено поређење и анализа добијених резултата за различите методе машинског учења, различите начине представљања докумената и за различите корпусе.

Остале битне информације:

Апликација ће бити развијана у програмском језику *Python* и биће јавно доступна.

Литература:

1. Manning, Christopher D., and Hinrich Schütze., *Foundations of statistical natural language processing*, Vol. 999. Cambridge: MIT press, 1999.
2. Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques." *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, 2002.

Марија Радовић, 1065/2018, МР

(име и презиме студента, бр. индекса, смер и модул)

Сагласан ментор др Јелена Граовац, доцент

(својеручни потпис студента)

(својеручни потпис ментора)

(датум подношења молбе)

Чланови комисије

1. доц. др Младен Николић

2. доц. др Јована Ковачевић

Катедра _____

је сагласна са предложеном темом.

(шеф катедре)

(датум одобравања молбе)