

Наслов рада

УДК 811.163.41'322.2

САЖЕТАК: Овде сажетак текста...

КЉУЧНЕ РЕЧИ: прва кључна реч,
друга кључна реч, трећа кључна реч.

РАД ПРИМЉЕН: 22. мај 2015.

РАД ПРИХВАЋЕН: 30. октобар 2015.

Име Презиме

moj@mejl.bg.rs

Подаци о установи

још података

*нови ред ако не стаје у
претходни*

Име Презиме

imeprezime@gmail.com

*Катедра за рачунарство
и информатику*

1. Увод

У аутоматској или полуаутоматској класификацији текста углавном се користе два различита приступа: приступ заснован на лексичко-семантичким језичким ресурсима и приступ заснован на машинском (De Heer, 1974; Manning и Schütze, 1999) учењу. Системи прве врсте користе лексичко-семантичке мреже, као што су WordNet уз ресурсе и ...

...

Затим ће бити приказана у основним цртама коришћена (PCJ, 2011) методологија (део 3.): метода МПВ са структурним излазом и њена прилагођавања за примену на вишекласну и хијерархијску класификацију (тачка 3.1) ...

Главни резултат рада – резултат класификације текстова – биће приказан у делу 4.. Биће приказани и резултати евалуације и однос према сродним (упоредивим) резултатима (Drugačiji citat za Manning, 1999). Најзад, у делу 5. интерпретираћемо и дискутовати добијене резултате, утицај примењених метода и могућности усавршавања.

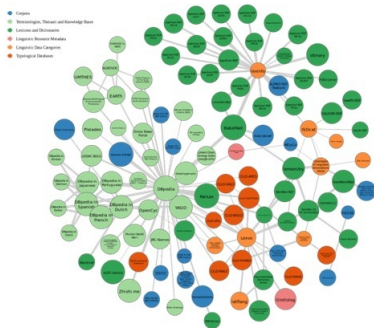
Још цитата: (Pettigrew, 2015), (Eyheramendy, 2003; Graovac, 2012; Gross, 1997), (Joachims, 2009)

2. Подаци (Dataset)

Ебарт (www.arhiv.rs/novinska-arhiva/) представља највећи архив новинских текстова савременог српског језика у дигиталном облику.

...

Овако добијени корпус се карактерише дрволиком структуром приказаном на Слици 1.



Слика 1. Опис слике

3. Методологија

3.1 Метод структуралних подржавајућих вектора

3.2 Н-грами

3.3 Репрезентација података

Постоје разне варијанте за рачунање вредности tf и idf , а у овом раду су коришћене следеће мере:

1. *classic tf-idf*: $tf \cdot \log\left(\frac{n}{n_k+1}\right)$
2. *log tf-idf*: $1 + \log(tf) \cdot \log\left(\frac{n}{n_k+1}\right)$
3. *boolean1 tf-idf*: $\log\left(\frac{n}{n_k+1}\right)$
4. *boolean2 tf-idf*: $\log\left(1 + \frac{n}{n_k}\right)$

...

3.4 Опис имплементације

Оба класификатора су настала прилагођавањем јавно доступног и бесплатног оквира за МСПВ методу SVM^{struct}¹.

Подаци су припремани на следећи начин:

- за сваки текст ...
- имплементиран је скрипт...

3.5 Опис експеримента

Расподела текстова по класама такође прати ову размеру и приказана је у Табели 1.

Класа	Скуп за подучавање	Скуп за тестирање	Цео корпус
1	10681	5340	16021
2	7173	3586	10759
3	12290	6145	18435
4	772	386	1158
5	9172	4586	13758
6	140	70	210
7	32	15	47
8	166	83	249
Укупно	40426	20211	60637

Табела 1. Број текстова у скупу за подучавање и скупу за тестирање по класама

¹ http://www.cs.cornell.edu/people/tj/svm_light/svm_struct.html

3.6 Евалуација

4. Резултати

5. Закључак

Литература

- De Heer, T., “Experiments with syntactic traces in information retrieval”, *Information Storage and Retrieval* Vol. 10 (1974), no. 3: 133–144, accessed September 1, 2015. <http://www.sciencedirect.com/science/article/pii/0020027174900151>
- Eyheramendy, Susana, David D. Lewis и David Madigan, “On the Naive Bayes model for text categorization”, *In Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics Conference* (2003): 332–339
- Graovac, Jelena, “Serbian text categorization using byte level n-grams”, *In Proceedings of CLoBL 2012: Workshop on Computational Linguistics and Natural Language, 5th Balkan Conference in Informatics, 2012*, 93–97, accessed September 1, 2015. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.416.6155&rep=rep1&type=pdf>
- Gross, Maurice, “The construction of local grammars”, *In Finite State Language Processing eds. Emmanuel Roche and Yves Schabbs* (1997): 329–354, accessed September 1, 2015. <https://halshs.archives-ouvertes.fr/halshs-00278316/document>
- Joachims, Thorsten, Thomas Finely и Chun-Nam John Yu, “Cutting Plane Training of Structural SVMs”, *Machine Learning Journal* Vol. 77 (2009), no. 1: 27–59, accessed September 1, 2015. <http://link.springer.com/article/10.1007/s10994-009-5108-8>
- Manning, Christopher и Hinnrich Schütze, *Foundations of Statistical Natural Language Processing*, Vol. 999, MIT Press, 1999
- Pettigrew, Sarah Ellen, “The Future of the Book and the Slow Book Movement”, *У History of the book: Disrupting Society from Tablet to Tablet*, Western Oregon University, 2015, 139–159, http://digitalcommons.wou.edu/cgi/viewcontent.cgi?article=1014&context=history_of_book
- PCJ, “Речник српског језика [Dictionary of Serbian Language]”, Матица српска, 2011