

DYNAMIC CLUSTERING ALGORITHM MEDIA REPORTS INTERNET IN STORYLINES

Alexey Potemkin

PhD in The Academy of the Federal Guard Service of the Russian Federation

Orel, Russia

Anton Borodaschenco

PhD in The Academy of the Federal Guard Service of the Russian Federation

Orel, Russia

Abstract. It is proposed to message text clustering algorithm Internet media information, which is a modification of known approaches based on key words. The proposed algorithm can be used for dynamic clustering of news reports on storylines online. Storylines are a set of messages for one event or phenomenon. This is achieved as described in the article modifications of existing approaches. We consider the implementation of this algorithm is described in detail features of the implementation, greatly reducing resource consumption when monitoring thousands of Internet media information. Using this algorithm can significantly reduce the dimension of the solution to the problem of information-analytical processing of media materials. These problems include: the definition of an actual «agenda» at a predetermined time interval, analysis of case information flows, people search for references or brands in news texts.

Keywords: text; message; mass media; news; topic; subject; Internet.

References

1. The Internet in Russia: the dynamics of penetration. Autumn 2015' 2016, Public Opinion Foundation, retrieved 27 June 2016, <http://fom.ru/SMI-i-internet/11889>.
2. Lande D.V. Modeling the dynamics of information flows // Modern problems of science and education. – 2012. – № 6 (part 3) – P. 652-654
3. Kushnarev D.A. Classification algorithms for clustering text documents // Karpovskiy scientific readings. – 2011. – № 5(part 1) – P.179-183.
4. Henry Anaya-Sánchez, Aurora Pons-Porrata, and Rafael Berlanga-Llavori A document clustering algorithm for discovering topics, Pattern Recognition Letters/ Vol: 31, No: 6, pp: 502-510, April 2010.

5. Dodonov A.G., Lande D.V. Methods of an analytical study of the dynamics of events based on the monitoring of the Internet Web Resources // Information technology and security: the basics of information security. – Kiev: IPRI NAN Ukraine. – 2014. – P. 3-17.
6. Dodonov A.G., Lande D.V. Modeling and analysis case information flows // Information counteraction to the terrorism threats. – 2013. – № 20. – P. 52-59.
7. Romanenko A.A. Feature selection and stepwise logistic regression for credit scoring // Machine Learning and Data Analysis. 2012. T. 1. № 3. P. 305-310.
8. Kreines M.G., Afonin A.A. Clustering of text collections: help with a content search and analysis tool // Internet portals: content and technology. P. 4. 2007. P. 510-537.
9. Zhigalkina N.F. A dynamic approach to the problem of clustering // Mathematical structure and modeling. – 2000. – №. 1 (5).
10. Aduenko A.A., Kuzmin A.A., Strijov V.V. Feature selection and metrics' optimisation when clustering documents collection // News of the Tula State University: natural sciences. – 2012. – №.3.
11. Potemkin A. V. Monitor the flow of dissemination messages on global information networks Internet // Herald of computer and information technologies.–2015. – № 10. – P. 44–49.
12. Borodaschenco A.Y., Potemkin A.V., Sazonova E.A., Shekshuev S.V. The mass media similar publication finding algorithm // Internet magazine «Naukovedenie». – 2015, №4. <http://naukovedenie.ru/PDF/74TVN415.pdf>

Contact

Alexey Potemkin

The Academy of the Federal Guard Service of the Russian Federation
302034, Russia, Orel, Priborostroitel'naya St., 35
alex.potemkin85@mail.ru

Anton Borodaschenco

The Academy of the Federal Guard Service of the Russian Federation
302034, Russia, Orel, Priborostroitel'naya St., 35
bay55@mail.ru