

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. *Ganti V.* Clustering Categorical Data Using Summaries / Ganti V., Gerhke J., Ramakrishan R., In Proc KDD'99, 1999
2. *Анисимов А.В.* Система обработки текстов на естественном языке / Анисимов А.В., Марченко А.А. // Искусственный интеллект. – № 4.– 2002. –С. 157-163.
3. *Aizawa A.* An information-theoretic perspective of *TF-IDF* measures / A. Aizawa. // Information Processing and Management. – 2003. – №39. – С. 45–65
4. *Барсегян А.А.* Методы и модели анализа данных: OLAP и Data Mining / А. А.Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб: БХВ-Петербург, 2004. – 336 с.
5. *Oren Eli Zamir.* A Phrase-Based Method for Grouping Search Engine Results. University of Washington, Department of Science & Engineering.
6. *Susan T. Dumains, George W. Furnas, Thomas K.Landauer.* Indexing by Latent Semantic Analysis. Bell Communications Research 435 South St. Morristown, NJ 07960. Richard Rashman: University Of Western Ontario.
7. *R. Sibson.* SLINK: An optimally efficient algorithm for the single-link cluster method. King's College Research Center, King's College, Cambridge, and Cambridge University Statistical Laboratory.
8. *Eui-Hong (Sam) Han, George Karyris.* Centroid-Based Document Classification: Analysis & Experimental Results. University of Minnesota, Department of Computer Science / Army HPC Research Center, 4-192 EECS Bldg., 200 Union St. SE, Minneapolis, MN 55455 USA.
9. *Eui-Hong (Sam) Han, George Karyris.* Concept Indexing. A Fast Dimensionality Reduction Algorithm with Application to Document Retrieval & Categorization. University of Minnesota, Department of Computer Science / Army HPC Research Center, 4-192 EECS Bldg., 200 Union St. SE, Minneapolis, MN 55455 USA.

10. *Eui-Hong (Sam) Han, George Kapyris, Vipin Kumar.* Text Categorization Using Weight Adjusted k -Nearest Neighbor Classification. University of Minnesota, Department of Computer Science / Army HPC Research Center, 4-192 EECS Bldg., 200 Union St. SE, Minneapolis, MN 55455 USA.
11. *Douglass R.Cutting, David R.Karger, Jan O.Pedersen, John W.Turkey.* Scatter / Gather: a Cluster-based Approach to Browsing Large Document Collections.
12. *Marti A.Hearst, Jan O.Pedersen.* Reexamining the Cluster Hypothesis: Scatter / Gather on Retrieval Results. Xerox Palo Alto Research Center, 3333 Coyote Hill RD, Palo Alto, CA 94304 .
13. *Jouko Lampinen, Erkki Oja.* Clustering Properties of Self-Organizing Maps. Lappeenranta University of Technology, Department of Information Technology, PO box 20, SF-53851 Lappeenranta, Finland.
14. *LA Soshnikova, VN Tamashevich, G. Uebe, M. Sheffer.,* 1999. Multidimensional statistical analysis in economics. Uniti: Moscow.
15. Sartaj Sahni:<http://www-pub.cise.ufl.edu/~sahni/dsaaj/enrich/c16/suffix.htm>
16. Esko Ukkonen. On-line construction of suffix trees. Department of Computer Science, University of Helsinki, PO Box 26 (Teollisuuskatu 23), FIN-00014 HUT, Finland.
17. *Alan Griffiths, H. Claire Luckhurst, and Peter Willett.* Using Interdocument Similarity Information in Document Retrieval Systems. Department of Information Studies, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom. Information Retrieval 1 997 p.365-373.
18. *Voorhees EM,* 1986. Implementing agglomerative hierarchical clustering algorithms for use in document retrieval . Information Processing and Management, 22: 456-476.
19. *Bjornar Larsen, Chinatsu Aone.* Fast and effective text mining using linear-time document clustering. In the proc. Of the Fifth ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining, pages 16-22, 1999..

20. *Teuvo Kohonen*. Self-Organization of Very Large Document Collections: State of the Art. Helsinki University of Technology, Neural Networks Research Center, PO Box 2200, FIN-02015 HUT, Finland.
21. *Oren Zamin, Oren Etzioni*. Grouper: A Dynamic Clustering Interface to Web Search Results. Department of Computer science and Engineering, University of Wahington, Box 352350 Seattle, WA 98195-2350 USA.
22. *Oren Zamin, Oren Etzioni*. Web Document Clustering: A Feasibility Demonstration. Department of Computer science and Engineering, University of Wahington, Box 352350 Seattle, WA 98195-2350 USA.
23. *А.М.Дубров, В.С.Мхитарян, Л.І.Трошин.*, 1998. Багатовимірні статистичні методи. Москва «Фінанси і Статистика»
24. *Michael W. Berry, Todd A. Letsche*. Computational Methods for Intelligent Information Access. Department of Computer Science University of Tennessee Knoxville, TN 37996-13031. Susan T. Dumains: Information science Research Group. Bellcore, 445 South Street Room 2L-371, Morristown, NJ 07962-1910.
25. *SERobertson . K. Spak Jones.*, 1976. Relevance weighting of search terms. Journal of the American Society for Information Science May -June 1976.
26. *Christopher D. Manning, Hinrich Schutze*. Foundation of Statistical Language Processing. MIT Press.
27. *Dan Pelleg, Andrew Moore*. X-means: Extending K -means with Efficient Estimation of the Number of Clusters . School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA.
28. *С.А. Айвазян, В.С. Мхитарян.*, 1998. Прикладная статистика и основы эконометрики. Издательство "ЮНИТИ", Москва.
29. *Е.С.Борисов*, 2007. Классификатор текстов на естественном языке, Москва
30. *Peter J. Rousseeuw*, 1987. "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". Computational and Applied Mathematics 20: 53-65.
31. *Hall, M.J.*, 2003. The interpretation of non-homogeneous hydrometeorological time series a case study. Meteorological Applications, 10, 61–67

32. *Krus, D.J., & Blackman, H.S.*, 1988. Test reliability and homogeneity from perspective of the ordinal test theory. *Applied Measurement in Education*
33. *Nguyen Xuan Vinh, Julien Epps and James Bailey*, 2009. "Information Theoretic Measures for Clustering Comparison: Is a Correction for Chance Necessary?" *ICML'09: Proceedings of the 26th Annual International Conference on Machine Learning*. ACM. pp. 1073–1080
34. *Lawrence Hubert and Phipps Arabie*, 1985. "Comparing partitions". *Journal of Classification*. 2 (1): 193–218
35. *W. M. Rand*, 1971. "Objective criteria for the evaluation of clustering methods". *Journal of the American Statistical Association*. American Statistical Association. 66 (336): 846–850
36. *Andrew Rosenberg and Julia Hirschberg*, 2007. V-Measure: A conditional entropy-based external cluster evaluation measure
37. *Олександр Стрел і Джойден Гош*, 2002. "Кластерні ансамблі - це система повторного використання знань для об'єднання декількох розділів". *Журнал Machine Learning Research* 3: 583-617
38. *Вінь, Енс і Бейлі*, 2009. "Теоретичні заходи щодо порівняння кластеризації". *Праці 26-ї щорічної міжнародної конференції з машинного навчання - ICML'09*
39. *Fabian Pedregosa; Gaël Varoquaux; Alexandre Gramfort; Vincent Michel; Bertrand Thirion; Olivier Grisel; Mathieu Blondel; Peter Prettenhofer; Ron Weiss; Vincent Dubourg; Jake Vanderplas; Alexandre Passos; David Cournapeau; Matthieu Perrot; Édouard Duchesnay*, 2011. "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research*. 12: 2825–2830
40. *Eli Bressert*, 2012. "SciPy and NumPy: an overview for developers". O'Reilly. p. 43
41. *Lawrence Hubert and Phipps Arabie*, 1985. "Comparing partitions". *Journal of Classification*. 2 (1): 193–218
42. *R.C. de Amorim, C. Hennig*, 2015. "Recovering the number of clusters in data sets with noise features using feature rescaling factors". *Information Sciences*. 324: 126–145.

43. *Peter J. Rousseeuw*, 1987. "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". *Computational and Applied Mathematics* 20: 53–65