Bibliographie

[Aas et Eikvil, 1999] Aas, K. et Eikvil L. (1999). Text Categorization : A Survery. Technical report. Norwegian Computing Center.

[Abbès et Hassoun, 1999] Ramzi Abbès, Mohamed Hassoun. 1999. « Conception d'un prototype de concordancier de la langue arabe : des éléments de réflexion », Colloque Génération Systématique et Traduction Automatique, Rabat, 15-17 novembre 1999.

[Abbès et Hassoun, 2002] Abbès, Ramzi et Hassoun, Mohamed. AraFreq : un outil pour le calcul de fréquences de mots arabes. In : abdelfattah braham. Colloque international sur le traitement automatique de l'arabe, 18-20 Avril 2002, Manouba-Tuni : Université la Manouba. pp 213-225.

[Abbès 2004] Abbès, Ramzi. La conception et la réalisation d'un concordancier électronique pour l'arabe. Thèse pour le doctorat de l'informatique. Institut National des Sciences Appliquées de Lyon (INSA). 2004.

[Abbès et Dichy, 2008] Abbès, Ramzi et Dichy, Joseph, « Extraction automatique de fréquences lexicales en arabe et analyse d'un corpus journalistique avec le logiciel AraConc et la base de connaissances DIINAR.1 » in : Heiden, Serge et Bénédicte Pincemain, Actes des JADT 2008, 9esjournées internationales d'analyse statistique des données textuelles (Proceedings of JADT 2008, 9th International Conference on Textual Data statistical Analysis).

[Al-Shalabi and Obeidat, 2008] Al-Shalabi R., Obeidat R., (2008). Improving KNN Arabic Text Classification with N-Grams Based Document Indexing. Proceedings of the Sixth International Conference on Informatics and Systems. March 27-29, 2008. Cairo, Egypt.

[Apté et al., 1994] Apté, C., Damerau, F. J., et Weiss, S. M. (1994). Automated learning of decision rules for text categorization. ACM Transactions on Information Systems, 12(3) : 233-251.

[Bawaneh, AlKoffash, Al Rabea, 2008] Bawaneh, M., AlKoffash, M., Al Rabea, A. Arabic Text Classification using K-NN and Naive Bayes. Journal of Computer Science 4 (7) : 600-605, 2008.

[Beesley, 1998] Beesley K., 1998, Arabic Morphological Analysis on the Internet, in A. Ubaydly (ed.), 6th International Conf. and Exhib. on Multilingual Computing (ICEMCO 98), Centre of Middle Eastern Studies, University of Cambridge, p. 3.1.1-10.

[Belkin et al. 1992] Belkin, N. J. and Croft, W. B. 1992. Information filtering and information retrieval : two sides of the same coin?. Commun. ACM 35, 12 (Dec. 1992), 29-38.

[Bhagwat et al. 2007] Bhagwat, D., Eshghi, K., and Mehra, P. 2007. Content-based document routing and index partitioning for scalable similarity-based searches in a large corpus. In Proceedings of the 13th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining (San Jose, California, USA, August 12 - 15, 2007). KDD '07. ACM Press, New York, NY, 105-112.

[Biebricher et al., 1988] Biebricher, P., Fuhr, N., Lustig, G., Schwantner, M., and Knorz, G. 1988. The automatic indexing system AIR/PHYS - from research to applications. In Proceedings of the 11th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Grenoble, France). Y. Chiaramella, Ed. SIGIR '88. ACM Press, New York, NY, 333-342.

[Blum et Mitchell, 1998] Blum, A. et Mitchell, T. Combining Labeled and Unlabeled Data with Co-training. Proceedings of the 11th Annual Conference on Computational Learning Theory, pp. 92-100, 1998.

[Borko et Bernick, 1964] Borko, H. et Bernick, M. (1964). Automatic document classification, part ii : additional experiments. Journal of the Association for Computing Machinery, 11(2) : 138-151.

[Bouckaert, 2004] Bouckaert, R. (2004), Naive Bayes Classifiers That Perform Well with Continuous Variables, Lecture Notes in Computer Science, Volume 3339, Pages 1089 – 1094.

[Buckely et al., 1994] Buckley, C. Salton, G., Allan J., et Singhal, A. Automatic Query Expansion Using SMART : TREC 3, In Proc. 3rd Text Retrieval Conference, NIST, 1994.

[Candillier, 2006] Candillier L. « Contextualisation, visualisation et évaluation en apprentissage non supervisé ». Thèse de doctorat. Université Charles de Gaulle, Lille 3. 2006.

[Cao, 2002] Cao, L.J. "Support Vector Machines Experts for Time. Series Forecasting", Neurocomputing, 2003, 51, pp. 321-339.

[Caropreso et al., 2001] Caropreso, M. F., Matwin, S., et Sebastiani, F. (2001). A learner-independent evaluation of the usefulness of statistical phrases for automated text categorization. In Chin, A. G., editor, Text Databases and Document Management : Theory and Practice, pages 78-102. Idea Group Publishing, Hershey, US.

[Carpuat et Wu 2007] Carpuat M. et Wu D. Improving Statistical Machine Translation using Word Sense Disambiguation. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 61–72, Prague, June 2007.

[Catanzaro et al., 2008] Catanzaro B., Sundaram N., Keutzer K., "Fast Support Vector Machine Training and Classification on Graphics Processors", In : International Conference on Machine Learning, 2008.

[Chai et al., 2002] Chai, K. M., Chieu, H. L., and Ng, H. T. 2002. Bayesian online classifiers for text classification and filtering. In Proceedings of the 25th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Tampere, Finland, August 11 - 15, 2002). SIGIR '02. ACM, New York, NY, 97-104.

[Chen, Leen, et Hwang, 2005] Chen, C., Lee, H., and Hwang, C. 2005. A Hierarchical Neural Network Document Classifier with Linguistic Feature Selection. Applied Intelligence 23, 3 (Dec. 2005), 277-294.

[Cohen, 1961/1970] Cohen D. (1961/70), « Essai d'une analyse automatique de l'arabe », 1961 (T.A. informations), in D. COHEN, Études de linguistique sémitique et arabe, Paris : Mouton, 1970.

[Cohen, 1970] Cohen D., 1970, Études de linguistique sémitique et arabe, Paris, Mouton.

[Cohen et Singer, 1999] Cohen, W. W. et Singer, Y. 1999. Context-sensitive learning methods for text categorization. ACM Trans. Inform. Syst. 17, 2, 141–173.

[Cornuéjols et Miclet, 2002] Cornuéjols, A. et Miclet, L. Apprentissage Artificiel : Méthodes et Algorithmes. Eyrolles 2002.

[Cornuéjols, 2002b] Cornuéjols, A. (2002). Une nouvelle méthode d'apprentissage :

Les svm. séparateurs à vastes marges. Bulletin de l'AFIA, 51 :14–23.

[Creecy et al., 1992] Creecy RH, Masand BM, Smith SJ and Waltz DL (1992). Trading mips and memory for knowledge engineering : Classifying census returns on the connection machine. Comm. ACM, 35 :48–63.

[Cristianini et Shawe-Taylor , 2000] Cristianini N., Shawe-Taylor J., An Introduction to Support Vector Machines, Cambridge University Press, 2000.

[Desclés, et al., 1983] Desclés J.-P. Abaab, H., Dichy J., Kouloughli, D.E., et Ziadah, M.S. Conception d'un synthétiseur et d'un analyseur morphologiques de l'arabe, en vue d'une utilisation en Enseignement assisté par Ordinateur, Rapport rédigé sous la direction de J.-P. DESCLÉS, à la demande du Ministère français des Affaires étrangères (sous-direction de la Politique linguistique).

[Denoyer et al. 2003] Denoyer, L., Vittaut, J., Gallinari, P., Brunessaux, S., and Brunessaux, S. 2003. Structured multimedia document classification. In Proceedings of the 2003 ACM Symposium on Document Engineering (Grenoble, France, November 20 - 22, 2003). DocEng '03. ACM Press, New York, NY, 153-160.

[Deerwester et al., 1990] Deerwester, S. Dumais, S. Landauer, T. Furnas, G. et Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American Society of Information Science, 416(6) : 391-407.

[Denis et al. 2003] F. Denis, R. Gilleron, A. Laurent, M. Tommasi, Text Classification and Co-Training from Positive and Unlabeled Examples, Proceedings of the ICML 2003 Workshop : The Continuum from Labeled to Unlabeled Data, 80-87, 2003.

[Dichy et Hassoun, 1989] Dichy J. et Hassoun M. (eds.), 1989, Simulation de modèles linguistiques et Enseignement Assisté par Ordinateur de l'arabe - Travaux SAMIA I, Paris, Conseil International de la Langue Française.

[Dichy, 1990] Dichy J., 1990, L'Écriture dans la représentation de la langue : la lettre et le mot en arabe, thèse d'État, Université Lumière-Lyon 2.

[Dichy, 1997] Dichy J. 1997. "Pour une lexicomatique de l'arabe : l'unité lexicale simple et l'inventaire fini des spécificateurs du domaine du mot". Meta 42, printemps 1997, Québec, Presses de l'Université de Montréal : 291-306.

[Dichy, 1998] Dichy J., 1998, Mémoire des racines et mémoire des mots : le lexique stratifié de l'arabe, in T. Baccouche, A. Clas et S. Mejri (eds.), La Mémoire des mots, Ve Journées du réseau "Lexicologie, Terminologie et Traduction" de l'AUF (Tunis, 25-27 sept. 1997), Revue Tunisienne de Sciences Sociales, n°117, p. 93-107.

[Dichy, Braham, Ghazali, Hassoun, 2002] Dichy J., Braham A., Ghazali S., Hassoun M., "La base de connaissances linguistiques DIINAR.1 (DIctionnaire INformatisé de l'Arabe, version 1)", paper presented at the International Symposium on The Processing of Arabic, Tunis (La Manouba), 18-20 April 2002.

[Dumais et al., 1998] Dumais, S., Platt, J., Heckerman, D., et Sahami, M. Inductive Learning Algorithms and Representations for Text Categorization, Technical Report, Microsoft Research, 1998.

[Dunning, 1993] Dunning, T. E. Accurate Methods for the Statistics of Surprise and Coincidence. In Computational Linguistics, volume 19 :1, pages 61-74, 1993.

[Duwairi, 2007] Duwairi, R. Arabic Text Categorization. The International Arab Journal of Information Technology, Vol. 4, No. 2, April 2007. pp : 125-131.

[Eibl et Pfeiffer, 2005] Eibl G. et Pfeiffer K. Journal of Machine Learning. Volume 6 (Feb) :189—210, 2005.

[El-Halees, 2007] El-Halees, A. Arabic Text Classification using K-NN and Naive Bayes.The Islamic University Journal (Series of Natural Studies andEngineering) Vol. 15, No.1, pp 157-167, 2007, ISSN 1726-6807, http//www.iugzaza.edu.ps/ara/research/

[Elkourid, Bensaid, and Rachidi, 2004] M. Elkourdi, A. Bensaid, and T. Rachidi, Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm, Proceedings of COLING 20th Workshop on Computational Approaches to Arabic Script-based Languages, Geneva, August 23rd-27th, 2004, pp. 51-58.

[Escudero et. al, 2000] Escudero, G., Màrquez, L., and Rigau, G. 2000. Boosting Applied toe Word Sense Disambiguation. In Proceedings of the 11th European

Conference on Machine Learning (May 31 - June 02, 2000). R. L. Mántaras and E. Plaza, Eds. Lecture Notes In Computer Science, vol. 1810. Springer-Verlag, London, 129-141.

[Fawcett, 2005] Fawcett, T. An Introduction to ROC Analysis. In ROC Analysis in Pattern Recognition, Vol. 27, No. 8. (June 2006), pp. 861-874.

[Feng et. al, 2005] Feng, S. L. and Manmatha, R. 2005. Classification Models for Historical Manuscript Recognition. In Proceedings of the Eighth international Conference on Document Analysis and Recognition (August 31 - September 01, 2005). ICDAR. IEEE Computer Society, Washington, DC, 528-532.

[Field, 1975] Field B. Towards automatic indexing : automatic assignment of controlled-language indexing and classification from free indexing. Journal of Documentation, 31(4) :246-265, 1975.

[Forman, 2003] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.

[Forman, 2007] Forman G. Computational Methods of Feature Selection. CRC Press/Taylor and Francis Group. 2007.

[Freund, 1995] Yoav Freund. Boosting a weak learning algorithm by majority. Information and Computation, 121(2) :256–285, 1995.

[Freund et Schapire, 1997] Freund, Y., et Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139.

[Freund et Schapire, 1999] Freund, Y., et Shapire, R. E. A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence, 14(5) :771-780, September, 1999.

[Fuhr et al., 1991] Fuhr N., Hartmann S., Knorz G., Lustig G., Schwantner M., and Tzeras K. AIR/X – a rule-based multistage indexing system for large subject fields. In Proceedings of RIAO'91, 3rd International Conference "Recherche d'Information Assistée par Ordinateur", pages 606-623, Barcelona, ES, 1991.

[Gale et. Al, 1993], Gale W. A., Church K. W., Yarowsky D. A method for disambiguating word senses in a large corpus. Computers and the Humanities, 26(5) :415-439, 1993.

[Gambäck et al. 2006] Gambäck, Björn and Sahlgren, Magnus and Argaw, Atelach Alemu and Asker, Lars (2006) Applying machine learning to Amharic text classification. In : WOCAL 5 : 5th World Congress of African Linguistics, 7-11 August 2006, Addis Ababa University, Ethiopia.

[Ghenima, 1998] Malek Ghenima. 1998. Analyse morpho-syntaxique en vue de la voyellation assistée par ordinateur des textes écrits en arabe. Thèse de doct., ENSSIB/Université Lyon 2.

[Gordillo et al. 2007] Gordillo, J. and Conde, E. 2007. An HMM for detecting spam mail. Expert Syst. Appl. 33, 3 (Oct. 2007), 667-682.

[Govindarajan, 2007], Text Mining Technique for Data Mining Application. Proceedings of world academy of science, engineering and technology volume 26 december 2007.

[Grainger, Dichy, et El-Halfaoui, 2003] In Jean-Pierre Jaffré, éd., Dynamiques de l'écriture : approches pluridisciplinaires, revue Faits de langue n° 22, Paris : Ophrys, 2003, p. 77-86.

[Guérif, 2006] Guérif, S. Réduction de Dimension en Apprentissage Numérique Non Supervisé, Thèse de doctorat, Université Paris 13, Villetaneuse, Décembre 2006.

[Guo et. Al, 2004] Guo G., Wang H., Bell D., Bi Y., et Greer K.. An kNN Model-based approach and its Application in Text Categorization. Proc. Of 5th International Conference on Intelligent Text Processing and Computational Linguistic, CICLing-2004, LNCS 2945, Springer-Verlag, pages 559-570.

[Hassoun, 1987] Hassoun M., 1987, Conception d'un dictionnaire pour le traitement automatique de l'arabe dans différents contextes d'application, thèse d'État (Sciences de l'information), Université Lyon 1.

[Huang, 1998] Huang, 1998 Y.L. Huang, A theoretic and empirical research of cluster indexing for mandarin Chinese full text document, The Journal of Library and Information Science 24 (1998), pp. 1023–2125 (in Chinese).

[Hull, 1994] Hull, D. A. 1994. Improving text retrieval for the routing problem using latent semantic indexing. In Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval (Dublin, Ireland, 1994), 282–289.

[Ide et al. 1998] Ide N. et Véronis J. (1998). Word Sense Disambiguation : The State of the Art. In Computational Linguistics, 24(1), 1-40, 1998.

[Iyer et al. 2000] Iyer, R. D., Lewis, D. D., Schapire, R. E., Singer, Y., and Singhal, A. 2000. Boosting for document routing. In Proceedings of the Ninth international Conference on information and Knowledge Management (McLean, Virginia, United States, November 06 - 11, 2000). CIKM '00. ACM Press, New York, NY, 70-77.

[Jaeger et. al, 2005] Jaeger, S., Ma, H., and Doermann, D. 2005. Identifying Script on Word-Level with Informational Confidence. In Proceedings of the Eighth international Conference on Document Analysis and Recognition (August 31 - September 01, 2005). ICDAR. IEEE Computer Society, Washington, DC, 416-420.

[Jalam, 2003] Jalam, R. (2003) "Apprentissage automatique et catégorisation de textes multilingues". Thèse de doctorat, Université Lumière Lyon 2.

[Jain et al. 1994] Jain, S. and Sharma, A. 1994. On the intrinsic complexity of language identification. In Proceedings of the Seventh Annual Conference on Computational Learning theory (New Brunswick, New Jersey, United States, July 12 - 15, 1994). COLT '94. ACM Press, New York, NY, 278-286.

[Joachims, 1997] Joachims, T. 1997. A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization. In Proceedings of ICML-97, 14th International Conference on Machine Learning (Nashville, TN, 1997), 143–151.

[Joachims, 1998] Joachims, T. (1998). Text categorization with support vector machines : learning with many relevant features. In Nédellec, C. and Rouveirol, C. editors, Proceedings of ECML-98, 10th European Conference on Machine Learning, pages 137-142, Chemnitz, DE. Springer Verlag, Heidelberg, DE.

[Joachims, 1999] Joachims, T. (1999). In Bratko, O. and Dzeroski, S., editors, Proceedings of ICML-99, 16th International Conference on Machine Learning, pages 200-209, Bled, SL. Morgan Kaufmann Publishers, San Francisco, Us.

[Joachims, 2002] Joachims, T. (2002) Learning to Classify Text Using Support Vector Machines : Methods, Theory and Algorithms. Kluwer Academic Publishers.

[Khreisat, 2006] Khreisat, L. Arabic Text Classification Using N-Gram Frequency Statistics A Comparative Study, Proceedings of the 2006 International Conference on Data Mining. Las Vegas, USA, 2006, pp. 78-82.

[Kim et al., 2000] Kim, Y., Hahn, S., and Zhang, B. 2000. Text filtering by boosting naive Bayes classifiers. In Proceedings of the 23rd Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Athens, Greece, July 24 - 28, 2000). SIGIR '00. ACM, New York, NY, 168-175.

[Kotsiantis, 2007] Kotsiantis S. Supervised Machine Learning : A Review of Classification Techniques, Informatica Journal 31 (2007) 249-268.

[Larkey et. al, 2002] Larkey, L., L. Ballesteros, and M. Connell, 2002. Improving Stemming for Arabic Information Retrieval : Light Stemming and Co-occurrence Analysis. Proceedings of the 25th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2002), Tampere, Finland, August 11-15, 2002, 275-282.

[Lee, Chen and Hwang, 2000] H.M. Lee, C.M. Chen, C.W. Hwang, A neural network document classifier with linguistic feature selection, in : Proceedings of 13th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, New Orleans, Louisiana, USA, IEA/AIE 2000, 19–22 June 2000, pp. 555–560

[Lefèvre, 2000] Lefèvre P. La Recherche d'informations, Hermès, 2000.

[Lesk, 1986] Lesk M. E. Automatic word sense disambiguation : how to tell a pine cone from an ice cream cone. In V. DeBuys, editor, Proceedings of SIGDOC-86, 5th ACM International Conference on Systems Documentation, pages 24-26, New York, US, 1986.

[Lewis, 1998] Lewis, D. D. 1998. Naive (Bayes) at forty : The independence assumption in information retrieval. In Proceedings of ECML-98, 10th European Conference on Machine Learning (Chemnitz, Germany, 1998), 4–15.

[Lewis , 1992a] Lewis, D.D. Feature Selection and Feature Extraction for Text Categorization. Proceedings of Speech and Natural Language Workshop, pp. 212-217, Morgan Kaufmann, San Francisco, US, 1992.

[Lewis, 1992b] Lewis, D. D. (1992). An evaluation of phrasal and clustered representations on a text categorization task. In Belkin, N. J., Ingwersen, P. and

Pejtersen, A. M., editors Proceedings of SIGIR-92, 15th ACM International Conference on Research and Development in Information Retrieval, pages 37-50, Kobenhavn, DK. ACM Press, New York, US.

[Lewis et Ringuette, 1994] Lewis, D. D. et Ringuette, M. (1994). A comparison of two learning algorithms for text categorization. In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Informatin Retrieval, pages 81-93, Las Vegas, US.

[Lewis et al. 1996] D. D. Lewis, R. E. Schapire, J. P. Callan, and R. Papka. Training algorithms for linear text classifiers. In Proceedings of SIGIR-96, 19th ACM International Conference on Research and Development in Information Retrieval, pages 298–306, Z¨urich, CH, 1996.

[Li et Jain, 1998] Li, Y. H. et Jain, A. K. 1998. Classification of text documents. Comput. J. 41, 8, 537–546.

[Liddy et al. 1994] Liddy, E. D., Paik, W., and Yu, E. S. 1994. Text categorization for multiple users based on semantic features from a machine-readable dictionary. ACM Trans. Inf. Syst. 12, 3 (Jul. 1994), 278-295.

[Liebowitz, 1997] Liebowitz, J. 1997 The Handbook of Applied Expert Systems. 1st. CRC Press, Inc.

[Lin et Liu, 2005] Lin, Y. and Liu, T. 2005. Robust Face Detection with Multi-Class Boosting. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cvpr'05) - Volume 1 - Volume 01 (June 20 - 26, 2005). CVPR. IEEE Computer Society, Washington, DC, 680-687

[Liu et al., 2002] Liu, Y., Yang, Y., et Carbonell, J. (2002). Boosting to correct the inductive bias for text classification. In Proceedings of CIKM-02, 11th ACM International Conference on Information and Knowledge Management, McLean, US. ACM Press, New York, US.

[Liu, et al. 2005] Liu Y., Chang F., Lin C. Language Identification of Character Images Using Machine Learning Techniques. Proceedings of the Eighth International Conference on Document Analysis and Recognition. Page 630-634, 2005.

[Maron 1961] Maron. M. Automatic indexing : an experimental inquiry. Journal of the Association for Computing Machinery. 8(3) : 404-417, 1961.

[MeSH 1993] National Library of Medicine, Bethesda, US. Medical Subject Headings (MeSH), edition 1993.

[Mesleh, 2007] Mesleh, A. Chi Square Feature Extraction Based Svms Arabic Language Text Categorization System. Journal of Computer Science 3 (6) : 430-435, 2007.

[Mitchell, 1997], Mitchell, T. M. (1997). Machine Learning Computer Science. McGraw-Hill. New York.

[Mladenič et Grobelnik, 1998] Mladenic, D. et Grobelnik, M. (1998). Word sequences as features in text learning. In Proceedings of ERK 98, the Seventh Electrotechnical and Computer Science Conference, pages 145-148, Ljubljana, SL.

[Mohamadally et Fomani, 2006] Mohamadally H. et Fomani B. SVM : Machines à Vecteurs de Support ou Séparateurs à Vastes Marges. 2006. http://georges.gardarin.free.fr/Surveys_DM/Survey_SVM.pdf. [Date de dernière visite: Mai 2010].

[Moulinier, 1996] Moulinier, I. (1996). Une approche de la catégorisation de textes par l'apprentissage symbolique. PhD thesis, Université Paris 6, Paris.

[Myers et al., 2000] Myers, K., Kearns, M., Singh, S., et Walker, M. A. 2000. A boosting approach to topic spotting on subdialogues. In Proceedings of ICML-00, 17th International Conference on Machine Learning (Stanford, CA, 2000), 655–662.

[NASA 1994] National Aeronautics and Space Administration, Washington, US. NASA thesaurus, édition 1994. Accessible depuis l'adresse suivante : http ://www.sti.nasa.gov/nasa-thesaurus.html [Date de dernière visite: Mai 2010].

[Ng et al., 1997] Ng, H. T., Goh, W. B. et Low, K. L. (1997). Feature selections, perceptron learning, and a usability case study for text categorization. In Belkin, N. J., Narasimhalu, A. D., and Willett, P. editors, Proceedings of SIGIR-97, 20th ACM International Conference on Research and Development in Information Retrieval, pages 67-73, Philadelphia, US. ACM Press, New York, US.

[Ouersighni, 2001] Riadh Ouersighni. 2001. "A major offshoot of the DIINAR-MBC project : AraParse, a morpho-syntactic analyzer of unvowelled Arabic texts". In ACL 39th Annual Meeting. Workshop on Arabic Language Processing : Status and Prospect, Toulouse, pp. 66-72.

[Peng et al., 2003] Peng, F., Schuurmans, D., and Wang, S. 2003. Language and task independent text categorization with simple language models. In Proceedings of the 2003 Conference of the North American Chapter of the Association For Computational Linguistics on Human Language Technology - Volume 1 (Edmonton, Canada, May 27 - June 01, 2003). North American Chapter Of The Association For Computational Linguistics. Association for Computational Linguistics, Morristown, NJ, 110-117.

[Peters and Sheridan, 2001] Peters, C. and Sheridan, P. (2001). Accès multilingue aux systems d'information. In 67th IFLA Council and General Conference.

[Pilászy., 2005], István Pilászy. Text Categorization and Support Vector Machines. In the Proceedings of the 6th International Symposium of Hungarian Researchers on Computational Intelligence

[Plantié, 2008] Plantié M., Roche M., Dray G., EGC 2008 : Un système de vote pour la classification de textes d\'opinion, Laboratoire LGI2P, Laboratoire LIRMM.

[Quinlan, 1986] Quinlan, J. R. Induction of decision trees. Machine Learning, 1(1) :81–106, 1986.

[Quinlan, 1993] Quinlan, J. R. (1993) C4.5 : Programs for Machine Learning. Morgan Kaufmann Publishers Inc.

[Raheel, 2009] Raheel, S. (2009). L'organisation des Connaissances et la Recherche d'Information Textuelles par l'Application des Méthodes Statistiques. 7ème colloque du chapitre français de l'ISKO. Lyon, France.

[Raheel, Dichy, and Hassoun, 2009b] S. Raheel, J. Dichy, M. Hassoun. The Automatic Categorization of Arabic Documents by Boosting Decision Trees.In the proceedings of the 5th International IEEE/ACM Conference on Signal-Image Technology and Internet-Based Systems , IEEE CS Press, Marrakech, Morocco, November, 2009.

[Raheel and Dichy, 2010a] Raheel, S., and Dichy J. Reducing Data Sparsity in a Language Dependent Automatic Classification of Arabic Documents. In the

proceedings of the 3rd. IEEE International Conference on Information Systems and Economic Intelligence, Sousse, Tunisia, 2010. Pages : 37-46.

[Raheel and Dichy, 2010b] Raheel, S., and Dichy, J. An Empirical Study on the Feature's Type Effect on the Automatic Classification of Arabic Documents. In the proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics. Iaşi, Romania. 2010. Springer LNCS 6008, Pages : 673-686.

[Rakotomalala, 1997] Rakotomalala R., « Approches constructives – Historique et évolution des graphes d'Induction », Première partie d'un exposé de synthèse, commun avec Mme D'Alché-Buc F. du LAFORIA, Université Paris 6, Rencontres Groupe de Travail de Recherche en Apprentissage (GTRA), Mai 1997.

[Rakotomalala, 2005] Rakotomalala R., "Arbres de Décision", Revue MODULAD, n°33, pp. 163-187, 2005.

[Rätsch, 2003] Rätsch G. Robust multi-Class boosting. In EuroSpeech, Geneva, September 2003. IEEE.

[Rish, 2001] Rish, Irina. (2001). An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence.

[Rudin, Daubechies, and Schapire, 2004] Rudin, C., Daubechies I., and Schapire, E. R. On the dynamics of boosting. In Advances in Neural Information Processing Systems 16, 2004.

[Sable et Hatzivassiloglou, 2000] Sable, C. L. et Hatzivassiloglou, V. (2000). Text-based approaches for non-topical image categorization. International Journal of Digital Libraries, 3(3) : 261-275.

[Salton et Buckley, 1988], Salton, G., McGill, M. J. Term weighting approaches in automatic text retrieval, Information Processing and Management, Vol. 24, No. 5, pp. 513-523, 1988.

[Salton et McGill, 1983] Salton, G. and McGill, M. J. 1986 Introduction to Modern Information Retrieval. McGraw-Hill, Inc.

[Schapire, 1990] Schapire, R. E. The strength of weak learnability. Machine Learning, 5(2) :197–227, 1990.

[Schapire, Freund, Barleet, and Lee , 1997] R. E. Schapire, Y. Freund, P. L. Bartlett, and W. S. Lee, "Boosting the margin : a new explanation for the effectiveness of voting methods," in Proc. 14th Int. Conf. Machine Learning. San Mateo, CA, 1997, pp. 322–330.

[Schapire et al., 1998] Schapire, R. E., Singer, Y., et Singhal, A. (1998). Boosting and Rocchio applied to text filtering. In Croft, W. B., Moffat, A., van Rijsbergen, C. J., Wilkinson, R., et Zobel, J. editors, Proceedings of SIGIR-98, 21st ACM International Conference of Research and Development in Information Retrieval, pages 215-233, Melbourne, AU. ACM Press, New York, US.

[Schapire et Singer, 1999] Schapire, R. E. and Singer, Y. 1999. Improved Boosting Algorithms Using Confidence-rated Predictions. Machine Learning 37, 3 (Dec. 1999), 297-336.

[Schapire et Singer, 2000] Schapire, R. E. et Singer, Y. (2000). BOOSTEXTER : a boosting-based system for text categorization. Machine Learning, 39(2/3) : 135-168.

[Schütze et al. 1995] Schütze, H., Hull, D. A., and Pedersen, J. O. 1995. A comparison of classifiers and document representations for the routing problem. In Proceedings of the 18th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Seattle, Washington, United States, July 09 - 13, 1995). E. A. Fox, P. Ingwersen, and R. Fidel, Eds. SIGIR '95. ACM Press, New York, NY, 229-237

[Schütze and Pedersen, 1995] Schütze, H. and Pedersen, J. (1995). "Information retrieval based on word senses." Proceedings of SDAIR'95. April 1995, Las Vegas, Nevada.

[Schütze, 1998] Schütze H. Automatic word sense discrimination. Computational Linguistics, 24(1) :97-124, 1998.

[Scott et Matwin, 1999] Scott, S. et Matwin, S. Feature Engineering for Text Classification. Proceedings of ICML-99, 16th International Conference on Machine Learning ,pp. 379-388, Morgan Kaufmann, San Francisco, US, 1999.

[Sebag et Gallinari, 2002] Sebag, M. et Gallinari, P. 2002. Apprentissage Artificiel: Acquis, Limites et Enjeux. In J. Le Maître, editor, *Assises 2002* : *Information - Interaction - Intelligence*. Cépaduès, 2002.

[Sebastiani, 1999] Sebastiani, F. (1999). A tutorial on automated text categorization. Proceedings of ASAI-99, 1st Argentinian Symposium on Artificial Intelligence, Buenos Aires, AR, 1999, pp. 7-35.

[Sebastiani et al, 2000] F. Sebastiani, A. Sperduti & N. Valdambrini. An Improved Boosting Algorithm and its Application to Automated Text Categorization. Proceedings of CIKM-00, 9th ACM International Conference on Information and Knowledge Management , pp. 78-85, McLean, US, 2000.

[Sebastiani, 2002] Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Surveys, 34(1) : 1.47.

[Shannon, 1948] Claude E.Shannon A mathematical theory of communication. Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

[Shannon, 1949] C.E. Shannon, The communication theory of secrecy systems, Bell System Technical Journal 28 (1949) (4), pp. 656–715.

[Simon 2005] Simon, R. 2005, Catégorisation automatique de textes et

cooccurrence de mots provenant de documents non étiquetés, Maître ès sciences

(M.Sc.), Université Laval.

[Song et al., 2006] Song Y., Zhou D., Huang J., Councill I. G., Zha H., C. Lee Giles : Boosting the Feature Space : Text Classification for Unstructured Data on the Web. ICDM 2006 : 1064-1069.

[Stricker, 2000] Stricker M. (2000) Réseaux de neurones pour le traitement automatique du langage : conception et réalisation de filtres d'informations, Thèse de Doctorat de l'Université Pierre et Marie Curie.

[Suárez et Palomar, 2002] Suárez, A. and Palomar, M. 2002. A maximum entropy-based word sense disambiguation system. In Proceedings of the 19th international Conference on Computational Linguistics - Volume 1 (Taipei, Taiwan, August 24 - September 01, 2002). International Conference On Computational Linguistics. Association for Computational Linguistics, Morristown, NJ, 1-7.

[Syiam, Fayed, et Habib, 2006] M. Syiam, Z. Fayed, and M. Habib, An Intelligent System for Arabic Text Categorization, International Journal of Intelligent Computing and Information Sciences, Vol.6, No.1, 2006, pp. 1-19.

[Tzeras et Hartman, 1993] Tzeras K. et Hartman S. (1993). Automatic indexing based on bayesian inference networks. In : Proc. 16th Ann. Int. ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'93), pp. 22–34.

[Vapnik, 1995] Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer Verlag.

[Vapnik, 1998] Vapnik, V. (1998). Statistical Learning Theory. John Wiley, NY.

[Vickrey et al. 2005] Vickrey, D., Biewald, L., Teyssier, M., and Koller, D. 2005. Word-sense disambiguation for machine translation. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (Vancouver, British Columbia, Canada, October 06 - 08, 2005). Human Language Technology Conference. Association for Computational Linguistics, Morristown, NJ, 771-778.

[Weiss et al., 1999] Weiss, S. M., Apté, C., Damereau, F. J., Johnson, D. E., Oles, F. J., Goetz, T., et Hampp, T. 1999. Maximizing text-mining performance. IEEE Intell. Syst. 14, 4, 63–69.

[Wiener, 1995] Wiener, E. D. (1995). A neural network approach to topic spotting in text. Master's thesis, Department of Computer Science, University of Colorado at Boulder, Boulder, US.

[Wiener et al., 1995] Wiener, E. D., Pederson, J. O., et Weigend, A. A. (1995). A neural network approach to topic spotting. In Proceedings of SDAIR-95, 4th Annual Symposium on Document Analysis and Information Retrieval, pages 312-332, Las Vegas, US.

[Witten et Frank, 2005] Witten, I. H. and Frank, E. (2005) Data mining : practical machine learning tools and techniques. (second ed). Morgan Kaufmann, San Francisco, CA.

[Wu et al., 2008] Wu, X., Kumar, V., Ross, Ghosh, J., Yang, Q., Motoda, H., Mclachlan, G., Ng, A., Liu, B., Yu, P., Zhou, Z.-H., Steinbach, M., Hand, D., Steinberg, D., January 2008. Top 10 algorithms in data mining. Knowledge and Information Systems 14 (1), 1-37.

[Yang, 1994] Yang Y. (1994) Expert network : Effective and efficient learning from human decisions in text categorization and retrieval. In : 17th Ann. Int. ACM

SIGIR Conference on Research and Development in Information Retrieval (SIGIR'94), pp. 13–22.

[Yang, 1999] Yang, Y. (1999). An evaluation of statistical approaches to text categorization. Information Retrieval, 1 (1/2) : 60-69.

[Yang et Chute, 1994] Yang Y and Chute CG (1994) An example-based mapping method for text categorization and retrieval. ACM Transaction on Information Systems (TOIS), pp. 253–277.

[Yang et Liu, 1999] Yang, Y. et Liu, X. (1999). A re-examination of text categorization methods. In Hearst, M. A., Grey, F., et Tong, R., editors, Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval, pages 42-49, Berkeley, US. ACM Press, New York, US.

[Yang et Pedersen, 1997] Yang Y., Pederson J., 1997. A comparative study on feature selection in text categorization. In J. D. H. Fisher, editor, The Fourteenth International Conference on Machine Learning (ICML'97), page 412-420. Morgan Kaufmann.

[Zaafrani, 1997] Riadh Zaafrani. 1997. « Morphological analysis for an Arabic Computer-aided learning system », Proceedings of DIALOGUE'97, International Conference on computational linguistics and its applications, June 10-15, 1997, Yasnaya Polyana, Russia.

[Zhou et al. 2007] Zhou, D., Burges, C. J., and Tao, T. 2007. Transductive link spam detection. In Proceedings of the 3rd international Workshop on Adversarial information Retrieval on the Web (Banff, Alberta, Canada, May 08 - 08, 2007). AIRWeb '07, vol. 215. ACM Press, New York, NY, 21-28

[Zighed, 1992] Zighed, D. A., Auray, J., and Duru, G. (1992). SIPINA : Méthode et logiciel. Editions A. Lacassagne, Lyon, France.

[Zighed et Rakotomalala, 2000] Zighed, D. A. et Rakotomalala, R. (2000). Graphes d'induction. Apprentissage et Data Mining. Hermes Science Publication, Paris.