

# Bibliographie

- Abt Helmut A. . “Is the Astronomical Literature Still Expanding Exponentially?”. *The Publications of the Astronomical Society of the Pacific*, 110 :210–213, February 1998.
- Accomazzi A., Eichhorn G., Kurtz M. J., Grant C. S., et Murray S. S. . “Astronomical Information Discovery and Access : Design and Implementation of the ADS Bibliographic Services”. *ASP Conf. Ser. 125 : Astronomical Data Analysis Software and Systems VI*, 6 :357+, 1997.
- Atukorale A.S. et Suganthan P.N. . “An Efficient Neural Gas Network for Classification”. Dans : *International Conference on Control, Automation, Robotics and Vision (ICARCV-98)*, pages 1152–1156, Singapore, December 1998a.
- Atukorale A.S. et Suganthan P.N. . “Hierarchical Overlapped Neural-Gas Network with Application to Pattern Classification”. *Neurocomputing*, 1998b. submitted to *Neurocomputing* journal in November 1998.
- Baeza-Yates R. . *Information Retrieval : Data Structures and Algorithms*, chapter 10 : String Searching Algorithms, pages 219–240. W. B. Frakes and R Baeza-Yates, prentice hall edition, 1992.
- Bates M.J. . “Subject access in online catalogs : a design model”. *Journal of the American Society for Information Science*, 37 :357–376, 1986.
- Belew R.K. . “Adaptive information retrieval : Using a connectionist representation to retrieve and learn about documents”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 11–20, 1989.
- Belew R.K. . *Finding Out About : Search Engine Technology from a Cognitive Perspective*. Cambridge Univ., 1999. <http://www-cse.ucsd.edu/rik/foa/index.html>.
- Bhandarkar S.M., Koh J., et Suk M. . “Multiscale image segmentation using a hierarchical self-organizing map”. *Neurocomputing*, 14(3) :241–272, 1997.
- Blackmore J. et Miikkulainen R. . “Incremental grid growing : encoding high-dimensional structure into two-dimensional feature map”. Dans : *Proceedings of ICNN'93, IEEE International Conference on Neural Networks*, pages 450–455, Piscataway, NJ, 1993. IEEE Service Center.
- Blosseville M.J., Hébrail G., Monteil M.G., et Pénot N. . “Automatic classification : natural language processing, statistical analysis and expert techniques used together”. Dans : *Conference on Research and Development in Information Retrieval*, pages 51–58, 1992.
- Boughanem M., Chrisment C., et Soulé-Dupuy C. . “Query modification based on relevance back-propagation in an ad hoc environment”. *Information Processing and Management*, 35 :121–139, 1999.

- Can F. et Ozkarahan E.A. . “Concepts and effectiveness of the cover-coefficient-based clustering methodology for text databases”. Dans : *ACM Transactions on Database Systems*, volume 15, pages 482–517, 1990.
- Ceeseman P. et al . “AutoClass : a Bayesian classification system”. Dans : *Fifth International Conference on Machine Learning*, pages 54–64. Morgan Kaufmann Publishers, Inc., 1988.
- Chen H., Houston A.L., Sewell R.R., et Schatz B.R. . “Internet browsing and searching : user evaluation of category map and concept space techniques”. *Journal of the American Society for Information Science*, 49(7) :582–603., May 1998.
- Chen H., Schuffels C., et Orwig R. . “Internet categorization and search : a self-organizing approach”. *Journal of Visual Communication and Image Representation*, pages 88–102, 1996.
- Cleverdon . “Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems”. Technical report, College of Aeronautics, Cranfield, U.K., 1962.
- Darken C., Chang J., et Moody J. . “Learning rate schedules for faster stochastic gradient search”. Dans : *Neural Networks for signal processing 2 - Proceedings of the 1992 IEEE Workshop*, 445 Hoes Lane, Piscataway, NJ 08854, 1992.
- Darken C. et Moody J. . “Note on Learning Rate Schedules for Stochastic Optimization”. Dans : *Advances in Neural Information Processing Systems*, Palo Alto, 1991.
- Dreyfus G. . “Les réseaux de neurones”. *Mécanique industrielle et matériaux*, 51, septembre 1998.
- Dreyfus G. et Ploix J.L. . “Early fault detection in a distillation column, an industrial application of knowledge-based neural modelling”. *Best Neural Network Practise in Europe*, *World Scientific*, 1998.
- Egret D., Creze M., Bonnarel F., Dubois P., Genova F., Heck A., Jasniewicz G., Lesteven S., Ochsenbein F., et Wenger M. . “A global perspective on astronomical data and information : the Strasbourg astronomical Data Center (CDS)”. *Astrophysics and Space Science Library*, 203 :p. 163, 1995.
- Fritzke B. . “A growing neural gas network learns topologies”. Dans : Tesauro G., Touretzky D.S., et Leen T.K., editors, *Advances in neural information processing systems 7*, pages 625–632. MIT Press, Cambridge, MA, 1995.
- Hamey L.G.C., Yeh J.C.H., Westcott T., et Sung S.K.Y. . “Pre-processing colour images with a self-organising map : Baking curve identification and bake image segmentation”. Dans : *Proceedings of the 14th International Conference on Pattern Recognition (Brisbane, Australia)*, pages 1771–1775, Piscataway, NJ : IEEE, 1998.
- Harman D. . “Relevance Feedback Revisited”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 1–10, 1992.
- Harman D., Fox E., Baeza-Yates R., et Lee W. . *Information Retrieval : Data structures and algorithms*, chapter 3 : Inverted files, pages 28–43. W. B. Frakes and R Baeza-Yates, prentice hall edition, 1992.
- Harman D.K. . “The first Text REtrieval Conference (TREC-1)”. *NIST Special publication*, 1993.
- Haskin R.L. . “Special-purpose processors for text retrieval”. *Database Engineering*, 4(1) : 16–29, 1981.

- Heck A. et Murtagh F. . *Intelligent information retrieval : The case of astronomy and related space sciences*. Astrophysics and Space Science Library, Dordrecht : Kluwer Academic Publishers, |c1993, edited by Heck, Andre ; Murtagh, Fionn, 1993.
- Hornik K., Stinchcombe M., et White H. . “Multilayer feedforward networks are universal approximators”. *Neural Networks*, 2(5) :359–366, 1989.
- Kaski Samuel . “Fast winner search for SOM-based monitoring and retrieval of high-dimensional data”. Dans : *Proceedings of ICANN99, Ninth International Conference on Artificial Neural Networks*, volume 2, pages 940–945. IEE, London, 1999.
- Kilpatrick D. et Williams R. . “Unsupervised classification of Antarctic Satellite Imagery using Kohonen’s Self-Organizing Feature Map”. Dans : *IEEE International Conference on Neural Networks, Perth*, volume 1, pages 32–36, 1995.
- Knerr S., Anisimov V., Baret O., Gorski N., Price D., et Simon J.C. . “The A2iA Recognition system for Handwritten Checks”. *International Journal of Pattern Recognition and Artificial Intelligence*, à paraître.
- Kohonen T. . “Self-organised formation of topologically correct feature maps”. *Biological Cybernetics*, 43 :59–69, 1982.
- Kohonen T. . *Self Organizing Maps*. Springer, 1995.
- Kohonen T. . “Exploration of very large databases by self-organizing maps”. Dans : *Proceedings of ICNN’97, International Conference on Neural Networks*,, pages PL1–PL6, IEEE Service Center, Piscataway, NJ., 1997.
- Kohonen T. . “The self-organizing map”. *Neurocomputing*, 21 :1–6, 1998.
- Kohonen T., Kaski S., Lagus K., et Honkela T. . “Self-organizing maps of document collections : A new approach to interactive exploration.”. Dans : *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 238–243, 1996.
- Kurtz M. et al. . “The ADS Abstract Service”. *Bull A.A.S.*, 25, 788., 25 :788+, 1993.
- Kwok K.J. . “A neural network for probabilistic information retrieval”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 21–30, 1989.
- Lagergren E. . “Comparing Interactive Information Retrieval Systems Across Sites : the TREC-6 Interactive Matrix Experiment”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 164–162, Sydney, Australia, 1998.
- Lancaster F.W. . *Information Retrieval Systems : Characteristics, Testing and Evaluation*. Wiley, 1968.
- Lelu A. et François C. . “Information retrieval based on a neural unsupervised extraction of thematic fuzzy clusters”. Dans : *Les Réseaux Neuromimétiques et leurs applications (NeuroNîmes)*, pages 93–104, 1992.
- Lesteven S., Murtagh F., Poinçot P., Creze M., Egret D., et Murtagh F. . “Neural Networks and Information Extraction : New developments in astronomical information retrieval for electronic publications”. Dans : *American Astronomical Society Meeting*, volume 189, pages 0609+, December 1996a.
- Lesteven S., Poinçot P., et Murtagh F. . “Neural networks and information extraction in astronomical information retrieval”. *Vistas in Astronomy*, 40 :395–400, 1996b.
- Lin X. . “Map displays for information retrieval”. *Journal of the American Society for Information Science*, 48(1) :40–54, 1997.

- Lin X., Soergel D., et Marchionini G. . “A Self-organizing semantic map for information retrieval”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 13–16, 1991.
- Linde Y., Buso A., et Gray R.M. . “An algorithm for vector quantizer design”. *IEEE Transactions on communication*, COM-28 :84–95, 1980.
- Lloyd S.P. . “Least squares quantization in pcm (1957 Bell Laboratories, unpublished)”. *IEEE Transactions on communication*, IT-28 :129–137, 1982.
- Luttrell S.P. . “Image compression using a multilayer neural network”. *Pattern Recognition Letters*, 10 :1–7, 1989.
- MacQueen J. . “Some methods for classification and analysis of multivariate observations”. Dans : *the Fifth Berkeley Symposium on Mathematical statistics and probability*, volume 1, pages 281–297, Berkeley, University of California Press, 1967.
- Maehoenen P.H. et Hakala P. J. . “Automated Source Classification Using a Kohonen Network”. *The Astrophysical Journal*, 452 :L77–+, October 1995.
- Malagnini M. L. . “A classification algorithm for star-galaxy counts”. Dans : *Statistical Methods in Astronomy*, pages 69–72, November 1983.
- Maron M.E. . *Mechanized documentation : the logic behind the probabilistic interpretation*, pages 9–13. Stevens et al., National Bureau of Standards, Washington, 1965.
- Martinetz T.M. et Schulten K.J. . *A neural gas network learns topologies*, pages 397–402. 1991.
- McNeill D.K. et Card H.C. . “Data Compression at Low Power using Soft Competitive Learning”. Dans : *Proceedings of the IEEE Pacific Rim Conference on Communications, Computers, Visualization, and Signal Processing*, pages 509–512, Victoria, B.C., 1995.
- McNeill D.K. et Card H.C. . “Soft Competitive Learning in Autonomous LEGO Robots”. Dans : *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering*, volume 1, pages 181–184, St. John’s, Newfoundland, 1997.
- McNeill D.K. et Card H.C. . “Competitive Learning for Extraction of Visual Representations of Motion”. Dans : *Proceedings of the IEEE International Joint Conference on Neural Networks*, Washington, D. C., 1999. Awarded best paper of session (oral).
- Minsky M. et Papert S. . “Perceptrons”. The MIT Press, Cambridge, 1969.
- Mothe J. . *Modèle connexionniste pour la recherche d’informations*. PhD thesis, Université Paul Sabatier, Toulouse, octobre 1994.
- Mothe J. et Dkaki T. . “Interactive multidimensionnal document visualization”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 363–364, Sydney, Australia, 1998.
- Mozer M.C. . “Inductive information retrieval using parallel distributed computation”. Technical report, Cognitive science, UCSD, La Jolla, CA, May 1984.
- Poinçot P. . “Création d’une carte bibliographique par l’utilisation des cartes auto-organisatrices de Kohonen”. Dans : *Actes des conférences INFORSID-97*, pages 625–641, 1997. Cet article a reçu un prix jeune chercheur.
- Poinçot P., Lesteven S., et Murtagh F. . “Comparison of Two Document Similarity Search Engines”. Dans : *ASP Conf. Ser. 153 : Library and Information Services in Astronomy III*, pages 85+, 1998.

- Poinçot P., Lesteven S., et Murtagh F. . “A spatial user interface to the astronomical literature”. *Astronomy and Astrophysics Supplement*, 130 :183–191, May 1998.
- Poinçot P., Lesteven S., et Murtagh F. . “Maps for Information Spaces : Assesments from Astronomy”. *Journal of the American Society for Information Science*, Soumis en 1999.
- Porter M. F. . “An algorithm for suffix stripping”. *Program*, 14(3) :130–137, 1980.
- Qiu Y. et Frei H.P. . “Concept based query expansion”. Dans : *Conference on Research and Devloppment in Information Retrieval (SIGIR)*, pages 160–169, 1993.
- Rasmussen Edie . *Information Retrieval : Data structures and algorithms*, chapter 16 : Clustering Algorithms, pages 419, 442. W. B. Frakes and R Baeza-Yates, prentince hall edition, 1992.
- Ritter H.J., Martinetz T.M., et Schulten K.J. . *Neuronale Netze*. Addison-Wesley, München, 1991.
- Robertson S.E. et Sparck Jones K. . “Relevance weighting of search terms”. *Journal of the American Society for Information Science*, 27 :129–146, 1976.
- Rosenblatt F. . “The Perceptron : a probabilistic model for information storage and organization in the brain”. *Psychological Review*, 65 :386–408, 1958.
- Rumelhart D.E., Hinton G.E., et Williams R.J. . “Learning internal representation by error propagation”. Dans : *Parallel Distributed Processing, explorations in the micro-structures of cognition*, volume 1, pages 318–362. Mitt Press, 1986.
- Salton G. . *The SMART Retrieval System - Experiments in Automatic Document Processing*. Prentice Hall, 1971.
- Salton G. et Buckley C. . “Improving retrieval performance by relevance feedback”. *Journal of the American Society for Information Science*, 41(4) :288–297, 1990.
- Silverstein C., Henzinger M., Marais H., et Moricz M. . “Analysis of a Very Large Alta Vista Query Log”. Technical report, DIGITAL Systems Research Center, 1998.
- Snowdon D., Fahlen L., et Stenius M. . “WWW3D : A 3D multi-user Web browser”. Dans : *WebNet96 Proceedings Online*, San Francisco, California, USA, October 1996.
- Sparck Jones K. . “A Statistical Interpretation of Term Specificity and its Application in Retrieval”. *Journal of Documentation*, 28(1) :11–20, 1972.
- Sparck Jones K. . “Search term relevance weighting given a little Relevance Information”. *Journal of Documentation*, 35(1) :30–48, 1979.
- Tombros A. et Sanderson M. . “Advantages of Query Biased Summaries in Information Retrieval”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 2–10, Sydney, Australia, 1998.
- Urbani D., Roussel-Ragot P., Personnaz L., et Dreyfus G. . “The selection of neural models of nonlinear dynamical systems by statistical tests”. Dans : *Neural Networks for Signal Processing*, volume I, pages 229–237. IEEE Press, 1994.
- Utans J. et Moody J. . “Selecting neural network architectures via the prediction risk : application to corporate bond rating prediction”. Dans : *Proceedings of the First International Conference on Artificial Intelligence Applications on Wall Street.*, Los Alamitos, CA, 1991. IEEE Computer Society Press.
- Van Rijsbergen C.J. . *Information Retrieval*. Butterworths, Londres (UK), deuxième édition edition, 1979.

- Vesanto J. . “SOM-based visualisation methods”. *Intelligent Data Analysis*, 3 :111–126, 1999.
- Voorhees E. . “Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 315–323, Sydney, Australia, 1998.
- Wartik S. . *Information Retrieval : Data Structures and Algorithms*, chapter 12 : Boolean Operations, pages 264–292. W. B. Frakes and R Baeza-Yates, prentice hall edition, 1992.
- Widrow G. et Hoff M.E. . “Adaptative switching circuits”. Dans : *1960 IRE WESCON Convention Record, New York IRE*, pages 96–104, 1960.
- Wilkinson R. et Hingston P. . “Using the Cosine Measure in a Neural Network for Document Retrieval”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 202–210, 1991.
- Williams J.H. . *Results of classifying with multiple discriminant functions*, pages 217–224. National Bureau of Standards, Washington, 1965.
- Wong S.K.M., Cai Y.J., et Yao Y.Y. . “Computation of term associations by a neuronal network”. Dans : *Conference on Research and Development in information retrieval (SIGIR)*, pages 107–115, 1993.
- Zipf G.K. . *Human behavior and the Principle of Least Effort*. Cambridge, Mass. ; Adison-Wesley, 1949.

## Résumé :

La quantité toujours croissante des données textuelles nécessite un effort constant pour la mise au point de méthodes de stockage et de consultation, afin que la totalité des informations conservées reste accessible simplement. C'est sur ce dernier point qu'ont porté les travaux effectués au cours de cette thèse.

Le CDS, Centre de Données astronomiques de Strasbourg, est un laboratoire dont la vocation est d'organiser, regrouper et diffuser les différents types d'information en astronomie. Outre les données observationnelles sur les objets astronomiques, une quantité importante de publications (plus de 100.000) sont enregistrées au CDS. C'est une partie de ces données bibliographiques que nous avons utilisées pour la mise au point de la **carte bibliographique**.

Notre système de recherche d'information, la carte bibliographique, est fondé sur les cartes auto-organisatrices (Self Organizing Maps, SOM) : des réseaux de neurones qui permettent une classification spatiale bidimensionnelle d'un ensemble de données (une sorte de cartographie). Appliquées à un ensemble d'articles, les SOM permettent une organisation des documents telle que des articles de caractéristiques voisines (de sujets voisins) sont classés dans des zones proches et les thèmes généraux apparaissent dans des Régions bien définies. Après avoir recherché les paramètres optimaux pour l'apprentissage des SOM, nous avons développé une interface de consultation qui permet de visualiser la répartition des documents et de localiser les zones relatives à certains thèmes (requêtes par mots-clés). La carte bibliographique est accessible aux adresses suivantes : <http://simbad.u-strasbg.fr/A+A/map.pl>, <http://simbad.u-strasbg.fr/ApJ/map.pl>.

## Abstract :

The constantly growing amount of textual information needs a continuous effort for the development of storage and consultation methods so that the totality of the stored information remains simply accessible. The work carried out during this thesis addressed on this last point.

The CDS (Strasbourg astronomical Data Centre) is a laboratory dedicated to organizing, gathering and diffusing the various types of information in astronomy. In addition to observational data on astronomical objects, a significant number of bibliographical references (more than 100,000) are stored at CDS. We used a part of these bibliographical data for the development of the **bibliographical map**.

Our information retrieval system, the bibliographical map, is based on the self organizing maps (SOM) : neuronal networks which allow a two-dimensional classification (a kind of cartography) of a data set. The SOM organizes the documents in such a way that closely related articles are classified in neighbouring zones and that general topics appear in well defined areas. After the determination of optimal parameters for the SOM's learning process, we have developed a graphical interface which allows the visualization of the documents distribution and the localization of documents related to given topics (keyword queries). The bibliographical map is accessible at the following addresses : <http://simbad.u-strasbg.fr/A+A/map.pl>, <http://simbad.u-strasbg.fr/ApJ/map.pl>.