Some interesting web-available abstracts and papers on clustering:

An Analysis of Recent Work on Clustering Algorithms (1999), Daniel Fasulo
Abstract: This paper describes four recent papers on clustering, each of which approaches the clustering problem from a different perspective and with different goals. It analyzes the strengths and weaknesses of each approach and describes how a user could decide which algorithm to use for a given clustering application. Finally, it concludes with ideas that could make the selection and use of clustering algorithms for data analysis less difficult.
http://citeseer.nj.nec.com/did/208269

Hierarchical Model-based Clustering For Large Datasets (1999), Christian Posse
Abstract: In recent years, hierarchical model-based clustering has provided promising results in a variety of applications. However, its use with large datasets has been hindered by a time and memory complexity that are at least quadratic in the number of observations. To overcome this difficulty, we propose to start the hierarchical agglomeration from an ecient classifcation of the data in many classes rather than from the usual set of singleton clusters. This initial partition is derived from a subgraph of the minimum spanning tree associated with the data. To this end, we develop graphical tools that assess the presence of clusters in the data and uncover observations difficult to classify.... http://citeseer.nj.nec.com/posse99hierarchical.html

Model-Based Hierarchical Clustering (1999), S. Vaithyanathan and B. Dom
Abstract: We present an approach to model-based hierarchical clustering by formulating an objective function based on a Bayesian analysis. This model organizes the data into a cluster hierarchy while specifying a complex feature-set partitioning that is a key component of our model. Features can have either a unique distribution in every cluster or a common distribution over some (or even all) of the clusters. The cluster subsets over which these features have such a common distribution correspond to the nodes (clusters) of the tree representing the hierarchy. We apply this general model to the problem of document clustering for which we use a multinomial likelihood function and….http://citeseer.nj.nec.com/386534.html

Refining Initial Points for K-Means Clustering (1998) P. S. Bradley, Usama M. Fayyad
Proc. 15th International Conf. on Machine Learning
Abstract: Practical approaches to clustering use an iterative procedure (e.g. K-Means, EM) which converges to one of numerous local minima. It is known that these iterative techniques are especially sensitive to initial starting conditions. We present a procedure for computing a refined starting condition from a given initial one that is based on an efficient technique for estimating the modes of a distribution. The refined initial starting condition allows the iterative algorithm to converge to a "better" local minimum. The procedure is applicable to a wide class of clustering algorithms for both discrete and continuous data. We demonstrate the application of this method to the popular K-Means... http://citeseer.nj.nec.com/bradley98refining.html

EM algorithms for self-organizing maps (1999). T.Heskes, J. Spanjers, W. Wiegerinck
Abstract: Self-organizing maps are popular algorithms for unsupervised learning and data visualization. Exploiting the link between vector quantization and mixture modeling, we derive EM algorithms for selforganizing maps with and without missing values. We compare self-organizing maps with the elastic-net approach and explain why the former is better suited for the visualization of high-dimensional data. Several extensions and improvements are discussed. 1 Introduction Self-organizing maps are popular tools for clustering and visualization of high-dimensional data [8, 13]. To derive an error function for the self-organizing map, we will follow the vector quantization interpretation given in, among... http://citeseer.nj.nec.com/280386.html

On the use of self-organizing maps for clustering and visualization (1999). A. Flexer. Principles of Data Mining and Knowledge Discovery.
Abstract: We will show that the number of output units used in a self-organizing map (SOM) influences its applicability for either clustering or visualization. By reviewing the appropriate literature and theory as well as our own empirical results, we demonstrate that SOMs can be used for clustering or visualization separately, for simultaneous clustering and visualization, and even for clustering via visualization. For all these different kinds of application, SOM is compared to other statistical approaches. This will show SOM to be a very flexible tool which can be used for various forms of explorative data analysis but it will also be made obvious that this flexibility comes with a price in terms... http://citeseer.nj.nec.com/105424.html

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

DeGruijter, J.J., McBratney, A.B., 1988. A modified fuzzy k means for predictive classification. In: Bock,H.H.(ed) Classification and Related Methods of Data Analysis. pp. 97-104. Elsevier Science, Amsterdam.

Roubens, M., 1982. Fuzzy clustering algorithms and their cluster validity. European Journal of Operational Research
10, 294-301.

Xie,X.L., Beni,G.1991. A validity measure for fuzzy clustering. IEEE Transactions of Pattern Analysis and Machine Intelligence 13, 841-847.