Clustering and Semantics Preservation in Cultural Heritage Information Spaces

Javier Pereira
Universidad Diego Portales
Informatics Engineering School
Avenida Ejército 441
Santiago, Chile
javier.pereira@udp.cl

Felipe Schmidt
Universidad Diego Portales
Informatics Engineering School
Avenida Ejército 441
Santiago, Chile
fschmidt@gawab.com

Pedro Contreras
Royal Holloway
University of London
Dpt. of Computer Science
Egham Hill, Surrey
TW20 0ER. England
pedro@cs.rhul.ac.uk

Fionn Murtagh
Royal Holloway
University of London
Dpt. of Computer Science
Egham Hill, Surrey
TW20 0ER. England
fionn@cs.rhul.ac.uk

Hernan Astudillo
Universidad Técnica Federico Santa María
Departamento de Informática
Avenida España 1680
Valparaíso, Chile
hernan@inf.utfsm.cl

ABSTRACT

In this paper, we analyze the preservation of original semantic similarity among objects when dimensional reduction is applied on the original data source and a further clustering process is performed on dimensionally reduced data. An experiment is designed to test Baire, or longest common prefix ultrametric, and K-Means when prior random projection is applied. A data matrix extracted from a cultural heritage database has been prepared for the experiment. Given that the random projection produces a vector with components ranging on the interval $[0, 1]$, clusters are obtained at different precision levels. Next, the mean semantic similarity of clusters is calculated using a modified version of the Jaccard index. Our findings show that semantics is difficult to preserve by these methods. However, a Student's hypothesis test on mean similarity indicates that Baire clusters objects are semantically better than K-Means when we increase the digit precision, but paying an increasing cost for orphan clustered objects. Despite this cost, it is argued that the ultrametric technique provides an efficient process to detect semantic homogeneity on the original data space.

1. INTRODUCTION

In the digital cultural heritage domain, objects (indistinctly, items or artifacts) belong to different contexts (e.g. historical, social, geographical, etc.). Thus, the ultimate purpose of the digital platforms is to help users to discover these contexts, and learn about cultural heritage, through the accessibility and exploration of artifacts, independently from the place, technology or format. In this direction, ontologies have proven to be an extraordinary tool aiding to index and retrieve items recorded in large databases. Nevertheless, it is not unusual that diversity of cultural objects induce potentially big ontologies and/or vocabularies in some domains. One single object may be ontologically described, searched and retrieved by a very small subset of concepts, as compared to the vocabulary's size. In such cases, searching for similar items would suppose processing of huge sparse ($object \times concept$) data structures. Classical techniques to cluster objects by similarity on these data spaces are not suited because of their low computer efficiency [24, 25]. On the contrary, dimensional reduction methods present special opportunities in the facing of these structures. In particular, random projection has been shown to hugely improve the computation performance when very large sparse databases are processed [7, 12] while simultaneously preserving characteristics of the original data space. However, specific complementary methods must be used for clustering purposes, which in turn helps to carry out data-set matching, and to support fast proximity searching.

Massive and high dimensional data spaces often have hidden hierarchical regularity. Following early studies [21], we seek ultrametricity in a cultural data-set, but also the semantic preservation inside clusters when allowed. An ultrametric is a distance that is defined strictly on a tree. An ultrametric induces a hierarchical structure on data. In previous work, we have compared the Baire-ultrametric and the K-Means algorithms as downstream clustering methods to random projection, finding that the former is faster when grouping objects in the context of chemical structures [21] and astronomical redshifts [5]. Nevertheless, very little is known about the quality of clustering in the context of digital cultural heritage, where semantic preservation inside clusters is relevant. By *semantic preservation* we mean the original conceptual similarity between two objects in a cluster. Regarding comparison of clustering methods, this is usually focused on evaluation of validity of clusters and al-

RIAO'10, 2010, Paris, France.

gorithmic efficiency. Several validity criteria have been developed in the literature which may be classified as external, internal or relative criteria [8]. In the external approach, groups assembled by a clustering algorithm are compared to a previously accepted partition on the testing data set. In an internal approach, clustering validity is evaluated using data and features contained in the data-set. The relative approach searches for the best clustering result from an algorithm and compare it with a series of predefined clustering schemes. In all cases, validity indexes are constructed to evaluate proximity among objects in a cluster or proximity among resulting clusters.

In our case, prior to clustering, a dimensional reduction is applied. Thus, an interesting question for us is the preservation of original semantic similarity among objects when the dimensional reduction is carried out on the data-set, and a further clustering process is performed on reduced data. In this work, an experiment is designed to test Baire and K-Means when prior random projection is applied. A data matrix extracted from an ancient folk-music archive containing information about 5000 titles and 9000 inherent characteristics has been prepared for the experiment. As random projection reduces the data matrix into a vector with components in the interval $[0, 1)$, different precision levels for clustering purposes are tested. Next, for each cluster produced by the Baire or K-Means, semantic similarity of individuals is calculated using the Jaccard index. However, since usually this index measures similarity between two sets (vectors), without considerations for semantic inclusion, we use a modified version. The mean similarity of clusters is calculated in order to compare the clustering methods. Our findings show that semantics is difficult to preserve by these methods, but a Student t hypothesis tested on mean similarity indicates that Baire is more robust than K-Means.

In which follows, we explain the research conducted for this analysis and results obtained. In section 2 the Baire and K-Means cluster methods are described, focusing on the precision issue for the former case. In section 3 the experiment process is presented and transformations applied on data-sets are justified. The section 4 presents the experiment applied on the cultural data-set and results obtained from the clustering algorithms.

2. CLUSTERING ALGORITHMS

Our purpose consists of mapping data into an ultrametric space, searching for an ultrametric embedding, or ultrametrization [23]. Actually, inherent ultrametricity leads to an identical result with most commonly used agglomerative criteria [16]. Furthermore, data coding can help greatly finding how inherently ultrametric data is. In following sections we introduce the analyzed clustering algorithms and the dimensional reduction method to be applied to the original data-set.

2.1 Ultrametric Baire space and distance

A metric space (X, d) consists of a set X on which is defined a *distance function* d which assigns to each pair of points of X a distance between them, and satisfies the following four axioms for any triplet of points x, y, z:

1. $\forall x, y \in X, d(x, y) \geq 0$ (positiveness);
2. $\forall x, y \in X, d(x, y) = 0$ *iff* $x = y$ (reflexivity);
3. $\forall x, y \in X, d(x, y) = d(y, x)$ (symmetry);
4. $\forall x, y, z \in X, d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

When talking about an *ultrametric space* we need to consider the "strong triangular inequality" or *ultrametric inequality* defined as $d(x, z) \leq max\ \{d(x, y) + d(y, z)\}$, this in addition to the positivity, reflexivity and symmetry properties for any triple of point $x, y, z \in X$.

An ultrametric space implies respect for a range of stringent properties. For example, the triangle formed by any triplet is necessarily isosceles, with the two large sides equal; or is equilateral. An ultrametric is a distance that is defined strictly on a tree.

A Baire space consists of countably infinite sequences with a metric defined in terms of the longest common prefix: the longer the common prefix, the closer a pair of sequences. What is of interest to us here is this longest common prefix metric, which we called the Baire distance [21, 4]. The longest common prefixes at issue here are those of precision of any value. Consider two such values, x_{ij} and y_{ij}, which, when the context easily allows it, we will call x and y. Each are of some precision, and we take the integer $|K|$ (where $|.|$ denotes set cardinality) to be the maximum precision. Finally, we will assume for convenience that each value is in the interval [0, 1) and this can be arranged by normalization.

Thus we consider ordered sets x_k and y_k for $k \in K$. In line with our notation, we can write x_k and y_k for these numbers, with the set K now ordered. (So, $k = 1$ is the first decimal place of precision; $k = 2$ is the second decimal place; . . . ; $k = |K|$ is the $|K|\,th$ decimal place.) The cardinality of the set K is the precision with which a number, x_k , is measured.

Consider as examples $x_k = 0.478$; and $y_k = 0.472$. In these cases, $|K| = 3$. For $k = 1$, we find $x_k = y_k = 4$. For $k = 2$, $x_k = y_k$. But for $k = 3$, $x_k \neq y_k$.

We now introduce the following distance (case of vectors x and y, with 1 attribute):

$$d_B(x_K, y_K) = \begin{cases} 1 & \text{if } x_1 \neq y_1 \\ \inf\ 2^{-n} & x_n = y_n \quad 1 \leq n \leq |K| \end{cases} \tag{1}$$

We call this d_B value Baire distance, which can be shown to be an ultrametric [17, 19, 18, 20, 21].

2.2 K-Means

K-Means is a major clustering algorithm technique that's present in various forms, first introduced by MacQueen in 1967 [14] and further developed by Hartigan and Wong [9, 10]. This algorithm groups data by minimizing the sum of the squares of distances between the data points and the cluster centroid. Suppose we have the data-set $X = \{x_1, x_2, x_3, .., x_N\}$ consisting of N observations of a d-dimensional variable X, where x_1 represents the first observation. The goal of this algorithm is to partition the set X into a number of $K > 1$ of non-overlapping clusters, where at the moment we assume the value of K is given. The algorithm has two main iterative steps; first is to update clusters according to the minimum distance rule, second is to update centroids as the centers of gravity of the clusters. This notion can be formalized by introducing a set of d-dimensional vectors μ_k, where $k = 1, ..., K$, in which μ_k is a candidate associated with the k^{th} cluster. We can now define an objective function given by:

$$D = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||X_n - \mu_k||^2, \qquad (2)$$

which represents the sum of the squares of the distances of each data point to its assigned vector μ_k [3].

2.3 Dimensionality reduction by random projection

As mentioned above it is a well known fact that traditional clustering methods do not scale well in very high dimensional spaces. A standard and widely used approach when dealing with high dimensionality is to first apply a dimensionality reduction method. For example, Principal Component Analysis (PCA) is a very popular choice to deal with this problem. It uses a linear transformation to form a simplified data set retaining the characteristics of the original data. PCA does this by means of choosing the attributes that best preserve the variance of the data. This is a good solution when the data allows these calculations, but PCA as well as other dimensionality reduction techniques remain expensive computationally speaking.

In order to apply the Baire ultrametric or K-Means algorithm our first step is to reduce the dimensionality of the original data, we choose to use random projection [11, 2, 6] not only because of performance but also because of some nice properties of this methodology. Random projection is the finding of a low dimensional embedding of a point set, such that the distortion of any pair of points is bounded by a function of the lower dimensionality.

In fact random projection here works as a class of hashing function. Hashing is much faster than alternative methods because it avoids the pair-wise comparisons required for partitioning and classification. If two points (p, q) are close, they will have a very small $\|p-q\|$ (Euclidean metric) value; and they will hash to the same value with high probability; if they are distant, they should collide with small probability.

3. EXPERIMENT DESIGN

The experiment process used in this work is depicted in Figure 1.

First, the input raw data is selected. Usually this consists of a data matrix where rows represent items in some specific domain, and columns represent characteristics or features associated with these items. In general, we may assume that the data-set is in first normal form in the database sense, that is each cell in the matrix contains just one value. A number of items, and their characteristics, are selected for the next phase. In the data comprehension step, data completeness, ambiguity and semantic quality of data are evaluated. A pre-process activity is undertaken for analysis purposes. In our case, the main process consists of the identification of semantic data associated with items. Items are annotated following the semantic characteristics, which implies coding the presence or absence of a characteristic for every item in the data structure. The final result in this step is a $\{0, 1\}$ matrix where m items are represented by n attributes. The third step consists of the random projection of the data-set. In this case, a $(n \times 1)$ normalized random vector is used for projection.

A projected data set is used for clustering purposes in the fourth step. Two clustering algorithms are applied, obtaining a given number of clusters on each case. In the fifth step, intrinsic semantics of clusters, issued from each algorithm, is calculated using a modified Jaccard index. Finally, a hypothesis test is applied to know how significant the difference between the mean similarity of two clustering algorithms is.

4. ON CLUSTERING EXPERIMENTS

4.1 Sample data

In this work, data issued from a digital cultural heritage platform, called Contexta [1], was used to carry out the classification. Contexta aims to integrate and contextualize disperse, autonomous and heterogeneous cultural information. To achieve this, it uses a middleware to integrate distributed data sources with different policies of use, providing uniform access despite multiple data types and formats. Additionally it allows semantic handling of these data and contextualization based on user and items profiles. One of the main purposes of Contexta when helping users to find cultural items is *situation awareness*. Indeed, in the cultural heritage domain, people using this kind of system are not necessarily interested in single artifacts or lists of ranked items. User needs are more oriented to general objectives, searching for elements aiding to compare, interpret, aggregate, analyze, synthesize and discover knowledge [15]. Users receiving recommendations also request explanations to sense-making processes and learning in a contextualized manner. In this setting, our study aims to determine preservation of semantics among items in clusters generated by a specific algorithm, when dimensional reduction is applied.

Let S be the sample data and $s_i \in S, (i = 1, \ldots, n)$, an item in this set. Originally twenty one fields are available as item descriptors, but seven characteristics are kept: uri-identifier, author, content description, title, associated collection, and source. Other attributes are discarded because of redundancy, incompleteness or being uninformative. For semantic purposes a term dictionary is created based on the *content description* attribute, allowing an extended feature richness of items. Keywords are produced using the Apache Lucene [13] library, filtering out words (longer than two characters) contained in a stop-list file. Processing the data source we obtain a term-based dictionary with 8674 keywords. The final set of attributes (8680) are considered semantic features. Finally, a matrix is created where the (i, j) element represents the absence or presence (0 or 1) of the j-th feature in the i-th object. The result is a binary *information matrix* I prepared to carry out the experiments.

4.2 Clustering Process

In order to assemble clusters from data, random projection is performed first on I. Normalization by column sum is performed in this data-set and the outcome is prepared for the clustering process. We know that clustering results with the Baire ultrametric are sensitive to the precision level of the projected data-set (see section 2.1), so seven levels of digit precision are selected for our experiments: two, three, four, five, six, eight and twelve digits of precision. For each precision level, ten random projection vectors are generated, which implies seventy projection outcomes in total. Interestingly, given a precision level, we have found that the resulting projection is practically the same in all cases.

Starting with the Baire methodology, we defined a tree-like structure to store the data's object identifier. Each node has ten children and each value in the projected vector is

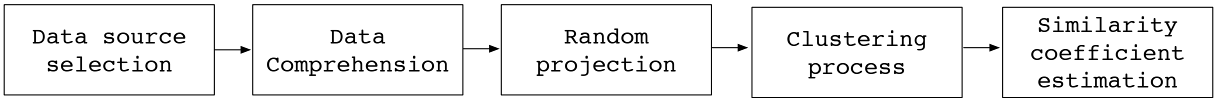

Figure 1: Experiment process design

separated by digits of precision and associated with the tree. The maximum tree depth is given by the digit precision used. For instance, when using four digits of precision we have a maximum of 10^4 possible branches, and a tree height of four. Once all the data in the projected vector is processed we check each leaf in the tree to identify the clusters. Note that only leafs with more than one element are considered. Thus, for pairwise comparison inside clusters, the 1-item leafs are excluded from the analysis.

Regarding K-Means, each experiment uses the number of clusters generated by the Baire method as initialization parameter. For example, when comparing the Baire clusters produced by four digits of precision, we use the number of clusters with more than two elements as k for K-Means (i.e. we exclude the empty and 1-items groups).

Once the clustering process is applied, we evaluate the pairwise semantic similarity within clusters. At this point, similarity between two items may be assessed by the Jaccard index, measuring the number of common items' features, divided by the total number of features present in the respective vectors in I. Given two vector rows $x, y \in I$ (representing the respective items in S), let $F(.)$ be the set of features satisfied by an object in the information matrix I, i.e. where the matrix value is 1. Then, the Jaccard coefficient of objects x and y is expressed as:

$$J(x,y) = \frac{|F(x) \cap F(y)|}{|F(x) \cup F(y)|} \tag{3}$$

Notice however that this coefficient strongly penalizes an item x such that $F(x) \subset F(y)$ and $\|F(x)\| \ll \|F(y)\|$. We propose that two items are semantically identical when the inclusion occurs because they become instances of the same class and/or subclass in an ontological system [22]. In this case, the object y may be considered a specialization of x. In consequence, a modified Jaccard coefficient is introduced here, which measures the semantic similarity between two objects, and takes into account the inclusion:

$$sim(x,y) = \frac{|F(x) \cap F(y)|}{\min\{|F(x)|, |F(y)|\}} \tag{4}$$

4.3 Results

The process has been performed ten times for each precision level and clustering algorithm. On each running, the following measures have been calculated over the number of clusters available for analysis:

1. R : average mean similarity among items within clusters;

2. R^*: average maximum similarity among items within clusters;

3. R_*: average minimum similarity among items within clusters.

Given a running (with a specific random projection vector at a determined precision level), the pairwise similarity among all items inside a cluster is averaged to obtain the mean similarity. Then R is the average of values calculated on clusters generated in such running. Equally, R^* and R_* are calculated as the average of maximum and minimum cluster similarity, respectively, among all clusters in a running.

Table 1 shows the different average values for every precision level. Also the number of clusters with more than one single element are presented for Baire and K-Means, where semantic similarity makes sense. In order to know whether differences on results obtained for Baire and K-Means are significant, a Student's test has been performed, at the 99% confidence level, for each precision level and R, R_*, and R^*. Differences are highly significant, which is explained by data in Table 2, where we notice that the standard deviation does not reach 1% of the respective average similarity.

The following results may be observed from Table 1 and Figure 2 when the precision level increases:

– for the Baire algorithm, R and R_* increase, but a sinusoidal form is observed for their respective curves;

– for the K-Means algorithm, R and R_* remain almost unchanged;

– Baire is persistently better than K-Means in R or R_*;

– the R value is not high for K-Means, indicating that semantic similarity is not consistently preserved, but it improves with higher precision in the Baire case.

Despite the fact that semantic similarity is difficult to preserve, we observe that the longer a common prefix within a resulting cluster is, the better the semantic of the original data space is preserved. Thus, objects that are semantically closer in the original data matrix I, hash closer in the 1-dimensional random projected vector, therefore more common prefixes exist among these groups. One could also hypothesize that a heterogeneous original data matrix, in the sense that there is a low semantic similarity among objects, could induce to more orphan objects (1-item clusters) in the Baire clustering method. On the contrary, it could be expected that homogeneity implies less orphan elements, but a reasonable good semantic similarity level on the resulting clusters. In fact, the Baire method provides an efficient process to detect semantic homogeneity in an information matrix. Our experiment indicates that the ratio of the number of orphan objects to the total number of objects, at a high-level precision, could measure how semantically near are rows in I. These statements will be studied in our future research.

Table 1: Average Semantic Similarity

	2 Digit		3 Digit		4 Digit		5 Digits		6 Digits		8 Digits		12 Digits	
	K-Means	Baire	K-Means	Baire	K-Means	Baire	K-Means	Baire	K-Means	Baire	K-Means	Baire	K-Means	Baire
R	0.0924	0.0924	0.1105	0.1071	0.1475	0.2302	0.1215	0.3954	0.1197	0.4349	0.1208	0.4398	0.1191	0.4398
R^*	0.6584	0.6584	0.4340	0.3596	0.4074	0.3567	0.4350	0.4806	0.4459	0.5131	0.4487	0.5179	0.4454	0.5178
R_*	0.0027	0.0027	0.0194	0.0311	0.0456	0.1580	0.0248	0.3363	0.0207	0.3792	0.0204	0.3841	0.0216	0.3841
No clusters	99	100	561	740	868	896	602	638	566	594	576	586	558	586
1-item clusters	1	1	60	173	118	1950	30	2923	25	3062	24	3079	23	3079

Table 2: Metrics Standard Deviation

	2 Digit		3 Digit		4 Digit		5 Digits		6 Digits		8 Digits		12 Digits	
	K-Means	Baire	K-Means	Baire	K-Means	Baire	K-Means	Baire	K-Means	Baire	K-Means	Baire	K-Means	Baire
R	0.0037	0.0061	0.0041	0.3463	0.0033	0.0052	0.0029	0.0043	0.0032	0.0016	0.0038	0.0001	0.0037	0.0001
R^*	0.0190	0.0782	0.0098	0.3464	0.0049	0.0037	0.0077	0.0056	0.0088	0.0017	0.0587	0.0000	0.0102	0.0001
R_*	0.0025	0	0.0043	0.3463	0.0035	0.0072	0.0025	0.0042	0.0039	0.0018	0.0032	0.0001	0.0031	0.0001
No clusters	1	0	12	11	8	9	5	11	4	2	4	1	7	1
1-item clusters	1	0	12	8	8	29	5	16	4	5	4	0	7	1

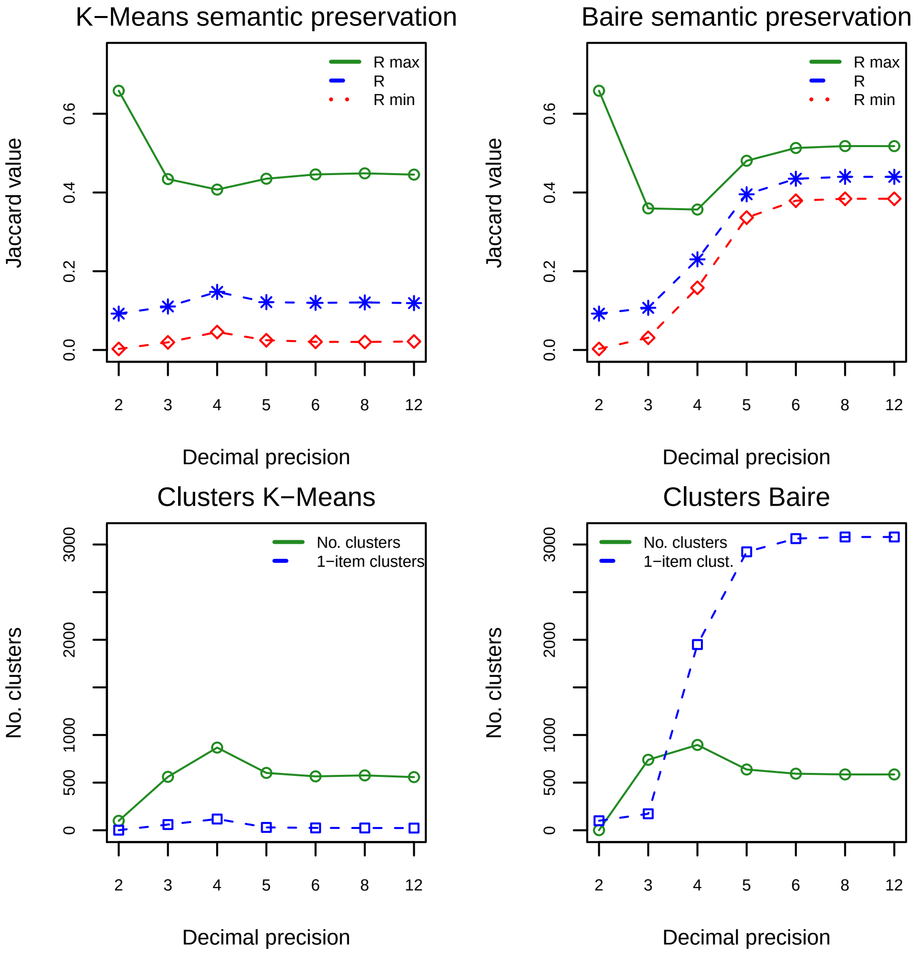

Figure 2: Jaccard values and No. of clusters

5. CONCLUSIONS

We have analyzed semantic similarity among objects when dimensional reduction is applied on a data-set and a further clustering process is performed on dimensionally reduced data.

An experiment was designed to test Baire, or longest common prefix ultrametric, and K-Means when prior random projection is applied. A data matrix extracted from an ancient folk-music archive was prepared for the experiment. Different precision levels for clustering purposes were tested and semantic similarity among group elements was calculated using a modified version of the Jaccard index. A Student's hypothesis test was performed on the mean similarity, which indicates that Baire is more robust than K-Means. However, our findings show that semantics are difficult to preserve by these methods, because the calculated similarity coefficient achieves moderate values. On one hand we can say that the Baire method in this case works as a filtering method for vectors that are semantically similar. On the other hand, once the number of centroids are chosen, K-Means works pulling the data points towards the centroids without considering if these point are closer in the original data space.

It was observed that both methods produce an important number of 1-item groups. This is a problem if clusters were to be used for data matching or for extraction purposes (i.e. when we consider groups produced by a large digit precision in relation to the data-set). Nevertheless, our results show that taking advantage of inherent data ultrametricity is a possible strategy for detecting semantic homogeneity in the original data-set. Therefore, our future research will consider alternative data sources, always in the area of semantic analysis, to prove this.

6. ACKNOWLEDGMENTS

This work was partially funded by projects CONTEXTA (FONDEF DD05I10286), UTFSM DGIP-24.08.58, and STIC-Amsud "Semantic Based Support for Collaborative Design."

7. REFERENCES

[1] H. Astudillo *et al.* Contexta: Semantic and contextualized management of distributed heterogeneous collections, 2007-2009. FONDEF Chilean Grant D05I10286, URL:www.contexta.cl.

[2] E. Bingham and H. Mannila. Random projection in dimensionality reduction: Applications to image and text data. In *KDD 2001: Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, 2001. ACM.

[3] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[4] P. Contreras and F. Murtagh. Evaluation of hierarchies based on the longest common prefix, or Baire, metric, 2007. Classification Society of North America (CSNA) meeting, University of Illinois. Urbana-Champaign. IL, USA.

[5] P. Contreras and F. Murtagh. Fast hierarchical clustering from the Baire distance. In *Classification as a Tool for Research. Proceedings of the 11th IFCS Biennial Conference and 33rd Annual Conference of the Gesellschaft für Klassifikation*, Dresden, 2009. Springer. In press.

[6] I. K. Fodor. A survey of dimension reduction techniques. In *LLNL technical report.* UCRL-ID-148494, June 2002.

[7] D. Fradkin and D. Madigan. Experiments with random projections for machine learning. In *KDD 2003: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 517–522, New York, NY, USA, 2003. ACM.

[8] G. Gan, C. Ma, and J. Wu. *Data Clustering Theory, Algorithms, and Applications*. Society for Industrial and Applied Mathematics. SIAM, 2007.

[9] J. A. Hartigan. *Clustering algorithms*. Wiley, New York, 1975.

[10] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

[11] S. Kaski. Dimensionality reduction by random mapping: fast similarity computation for clustering. In *IEEE International Joint Conference on Neural Networks*. IEEE, May 1998.

[12] P. Li, T. Hastie, and K. Church. Very sparse random projections. In *KDD 2006: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, volume 1, pages 287–296, New York, NY, USA, 2006. ACM.

[13] Lucene. Apache project, 2009. http://lucene.apache.org.

[14] J. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. LeCam and J. Neyman, editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

[15] E. Mäkelä, O. Suominen, and E. Hyvnen. Automatic exhibition generation based on semantic cultural content. In *Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007)*, November 12 2007.

[16] F. Murtagh. *Multidimensional Clustering Algorithms*. Physica-Verlag, 1985.

[17] F. Murtagh. On ultrametricity, data coding, and computation. *Journal of Classification*, 21:167–184, 2004.

[18] F. Murtagh. Quantifying ultrametricity. In J. Antoch, editor, *Proceedings in Computational Statistics, Compstat 2004*, pages 1561–1568. Springer, 2004.

[19] F. Murtagh. Thinking ultrametrically. In D. Banks, L. House, F. McMorris, P. Arabie, and W. Gaul, editors, *Classification, Clustering, and Data Mining Applications*, pages 3–14. Springer, July 2004.

[20] F. Murtagh. Identifying the ultrametricity of time series. *European Physical Journal B*, 43:573–579, 2005.

[21] F. Murtagh, G. Downs, and P. Contreras. Hierarchical clustering of massive, high dimensional data sets by exploiting ultrametric embedding. *SIAM Journal on Scientific Computing*, 30(2):707–730, February 2008.

[22] T. Ruotsalo and E. Hyvönen. A method for determining ontology-based semantic relevance. In *Database and Expert Systems Applications*, volume LNCS 4653/2007, pages 680–688, 2007.

[23] A. van Rooij. *Non-Archimedean Functional Analysis*. Marcel Dekker, January 1978.

[24] R. Xu and D. W. II. Survey of clustering algorithms. *Transactions on Neural Networks*, 16(3):645–678, May 2005.

[25] R. Xu and D. W. II. *Clustering*. IEEE Computer Society Pres, 2008.