# Spectra in Taxonomic Evidence in Databases III. Application in Celestial Bodies. Asteroids Families.

**Autores:**

*Angel Luis Plastino*
*PROTEM Laboratory*
*Department of Physical Sciences*
*Faculty of Sciences*
*University of La Plata*
*C.C. 727 or (115 # 48/49)*
*(1900) La Plata*
*Buenos Aires*
*Argentina*
*Phone: (54 221) 483-9061 (ext. 247)*
*E-MAIL: Plastino@venus.fisica.unlp.edu.ar*

*Rosa Beatriz Orellana*
*Mechanics Laboratory*
*Celestial Mechanics Department*
*Faculty of Astronomical and Geophysical Sciences*
*University of La Plata*
*Paseo del Bosque*
*(1900) La Plata*
*Buenos Aires*
*Argentina*
*Phone: (54 221) 421-7308*
*E-MAIL: rorellan@fcaglp.fcaglp.unlp.edu.ar*

*Gregorio Perichinsky*
*Operating System and Database Laboratory*
*Computer Science Department - South Wing - School of Engineering*
*University of Buenos Aires gperi@mara.fi.uba.ar*
*Paseo Colón Nº 850 - (1063) Buenos Aires - Argentina*
*Phone: (54 11) 4343-0891 (int. 140/145)*
*FAX: (54 11) 4331-0129*

**Abstract.**

Numerical Taxonomy aims to group in clusters, using so-called structure analysis of operational taxonomic units (OTUs or taxons or taxa) through numerical methods.
These clusters constitute families. Structural analysis, based on their phenotypic characteristics, exhibits the relationships, in terms of degrees of similarity, between two or more OTUs.
Entities formed by dynamic domains of attributes, change according to taxonomical requirements: Classification of objects to form families or clusters.
Taxonomic objects are here represented by application of the semantics of the Dynamic Relational Database Model.
Families of OTUs are obtained employing as tools i) the Euclidean distance and ii) nearest neighbor techniques. Thus taxonomic evidence is gathered so as to quantify the similarity for each pair of OTUs (pair-group method) obtained from the basic data matrix.
The main contribution of the present work is to introduce the concept of spectrum of the OTUs, based in the states of their characters. The concept of families' spectra emerges, if the superposition

principle is applied to the spectra of the OTUs, and the groups are delimited through the maximum of the Bienaymé-Tchebycheff relation, that determines Invariants (centroid, variance and radius).
Applying the integrated, independent domain technique dynamically to compute the Matrix of Similarity, and, by recourse to an iterative algorithm, families or clusters are obtained.
A new taxonomic criterion is thereby formulated.
An astronomic application is worked out. The result is a new criterion for the classification of asteroids in the hyperspace of orbital proper elements (the well-known Families of Hirayama).
Using an updated database of asteroids we ascertain the robustness of the method. Thus, a new approach to Computational Taxonomy is presented, that has been already employed with reference to Data Mining.
The Informatics (Data Mining and Computational Taxonomy), is always the original objective of our researches.

**1.Introduction.**

Classification is an abstraction technique used to collect objects with common properties.
The following hypothesis: 1) each object belongs to one (and only one) class and 2) for each class at least one object belongs to the classification, allow us to delimit the domain of objects.
The association of concepts in systematic way by recourse to numerical variables has been the source of a great variety of numerical classification techniques, that have their origin in Numerical Taxonomy.
The search of classification concepts that facilitate a robust classification structure (not modifiable by the addition of new information and not altered by the incorporation of new entities) constitutes an important endeavor. In such a line, this work develops tools based on Information Theory and, as a result, a new classification technique is found.
We discuss first principles and methods and then, as an application, we investigate the application of our ideas to the classification of asteroids (an application to celestial bodies), and the classification of plants in biological taxonomy, and the classification of objects of other disciplines.
A correspondence with ideas pertaining to the field of the Dynamic Databases is also established.
Taxonomic objects are here represented by the application of the semantics of the Dynamic Relational Database Model: Classification of objects to form clusters or families [34][35][36][37].

**2. Conclusions.**

The integrated and dynamical treatment of the domains permits an accessible normalization process, a rapid assignation of attribute - domain - values, the implementation of a model of data base Dynamic and their utilization in Numerical Taxonomy.
Our contribution here is of both a theoretical and an empirical nature. The agglomeration of objects forming classes produced by following the steps of the method (ALGORITHM) presented in this communication allows one to obtain clusters and domains with normalized values. Additionally, the density and range (in terms of the radius of the set) can be visualized as INVARIANT CHARACTERISTICS of the OTUs.
From a taxonomic viewpoint, we introduce the new and original concepts of characteristic spectra of OTUs and families' characteristic spectra.
The states of the characters of the attributes of the OTUs are considered in the spectra using principles of both superposition and interference. For the same purpose we use as well as the density and range concepts. An alternativer point of view is that of regarding the radius of the clusters as an invariant characteristic [22] [23] [27] [42].
The methodology introduced here has been applied, as an illustrative example, to the important astronomical problem of asteroids' classification. We are able to reproduce the celebrated families

of Hirayama by straightforward application of our algorithm, avoiding the intuitive and ad-hoc considerations of previous researchers [2] [3] [4] [31] [46].
Measures that allow to determine the stability of a family by its attributes' sensibility have also been studied.
The Families proposed by Hirayama have been validated by the numerical method described above.
The spectra have been used in a rather original way so as to confirm that the results obtained with our numerical methodology are congruent (consistent).
Our main contribution lies the use of dynamically independent domains to compute a CHARACTERISTIC SPECTRUM for each OTU (Object) from the Resemblance Matrix. We have computed this matrix by recourse to an iterative algorithm that allows us to obtain the clusters or families.
In the taxonomic space this method of clustering delimits taxonomics groups that can be visualized as FAMILIES' CHARACTERISTIC SPECTRA.
One should be encouraged by the present results to evaluate an Invariant relative to the Centroid :: Density :: Radius cluster for the incorporation of new members to the clusters, and also to design structure using other methods that employ our technology.

**References.**

[1]Abramson,N., "Information Theory and Coding". McGraw Hill. Paraninfo. Madrid. 1966.
[2]Arnold,J.R., "Asteroids Families and Jet Streams". The Astronomical Journal. 74: pp 1235-1242. 1969.
[3]Carusi,A., Massaro,E. "On Asteroids Classifications in Families". Astronomy and Astrophysics. Supplements 34, p 81. 1978.
[4]Carusi,A., Valsecchi,G.B. "On Asteroids Classifications in Families". Astronomy and Astrophysics. pp 327-335. 1982.
[5]Batini, C., Ceri, S., Navathe, S.B. "Conceptual Databases Design" Addison Wesley. 1998.
[6]Codd E. F. "Relational Completeness of Data Base Sublanguages". Database Systems, Courant Computer Science Symposia Series 6, Englewood Cliffs, New Jersey, Prentice-Hall. 1972.
[7]Codd E. F. "Extending the Data Base Relational Model to Capture More Meaning" ACM TODS 4, 4 pp 397-434. 1979.
[8]Codd E. F. "Relational Data Base. A Practical Foundation for Productivity" CACM 25, 2. 1982.
[9]Codd E. F. "How Relational is your Database Management System ?". Computer World. 1985.
[10]Codd E. F. "The Relational Model for Database Management: Version 2". Addison Wesley. 1990.
[11]Cramer, Harald. "Mathematics Methods in Statistics". Aguilar Edition. Madrid. Spanish. 1958.
[12]Crisci, J.V. , Lopez Armengol, M.F. "Introduction to Theory and Practice of the Numerical Taxonomy" , A.S.O. Regional Program of Science and Technology for Development. Washington D.C. Spanish. 1983.
[13]Date,C.J. "An Introduction to Datase Systems Vol. I". 6th Ed. Addison Wesley. 1995.
[14]Date,C.J. "Relational Database: Selected Writings". Addison Wesley. 1986.
[15]Date,C.J. "Relational Database: Further Misconceptions #1". Info DB, spring, 1986.
[16]Date,C.J. "A SQL Standard". Addison Wesley. 1987.
[17] Date,C.J. "Where SQL Falls Short". Datamation pp 84-86. 1987.
[18]Date,C.J. "An Introduction to Datase Systems Vol. I". 5ª Ed. Addison Wesley. 1998.
[19]Date,C.J. "Date on Databases" On proceeding of the Codd & Date Relational Database Symposium". Madrid. 1992.
[20]de Miguel, A., Piatttini, M. "Concepts and Design of Databases." Addison Wesley.1994. Spanish.
[21]Elmasri,R., Navathe,S. "Fundamentals of Database Systems".The Benjamin/Cummings Publishing Company and Addison Wesley. 1997.

[22]Feynman, R.P., Leighton, R.B. & Sands, M. “Lectures on physics, Mainly Mechanics, Radiation and Heat”. pp. 25-2 ff, 28-6 ff, 29-1 ff, 37-4. 1971.
[23]Frank,N.H. “Introduction to Mechanics and Heat”. Science Service. Washington. Editorial Atlante. Spanish. 1949.
[24]Freeman,J.A., Skapura,D.M. “Neural Networks. Algorithms, applications and techniques of programming”.Addison Wesley. Iberoamericana. Spanish. 1991.
[25]Gennari,J.H. “A Survey of Clustering Methods” (b). Technical Report 89-38. Department of Computer Science and Informatics. University of California., Irvine, CA 92717. 1989.
[26]Hamming, R.W. “Coding and information theory”. Englewood Clifs, NJ: Prentice Hall. 1980.
[27]Hetcht,E. and Zajac,A., “Optic”. Fondo Educativo Interamericano. pp. 5-11-206-207-293-297-459-534. Spanish 1977.
[28]Hirayama,K. “Groups of Asteroids Probably Common Origin”. Proceeding of Physics-Mathematics Society. Japan II:9. pp 354-351. 1918.
[29]Hirayama,K. “Groups of Asteroids Probably Common Origin”. The Astronomical Journal: 31, pp 185-188. 1918.
[30]Hirayama,K. “Present State of the Families of Asteroids”. Proceeding of Physics-Mathematics Society. Japan II:9. pp 482-485. 1933.
[31]Knêzevîc,Z., Milani,A. “Asteroids Proper Elements from an Analytical Second Order Theory”. Astronomy and Astrophysics. pp 1073. 1990.
[32]Kohonen, T. “Self - Organization and Associative Memory”. Berlin: Springer - Verlag. 1989.
[33]Perichinsky, G. “Multiple states of multiple state automata to key fast validation”. 11th. International Symposium Computer at University. Catvat. Zagreb. Yugoslavia. 1989.a.
[34]Perichinsky, G., Servetto, A., Crocco, E. “Relational Data Bases Structured on Dynamic Domains of Attributes".18th Sessions. Operative Research and Informatic’s Argentine Society.SADIO. 1989.b.
[35]Perichinsky, G., Servetto, A. "Dynamically Integrated Independent Domains on Data Bases”.19th Sessions Operations Research and Informatic’s Argentine Society.SADIO. 1990.
[36]Perichinsky, G. et Al. “Data Base Model Manager Structured on Independent Domains". Faculty of Science. National University of La Plata. Spanish. 1992.
[37]Perichinsky, G., Feldgen, M., Clúa, O. “Conceptual Contrast of Dynamic Data Bases with the Relational Model” in Proceedings International Association of Science and Technology for Development. 14th Applied Informatics Conference. Innsbruck, Austria. 1996.
[38]Perichinsky, G. , Feldgen, M. , Clúa, O. "Dynamic Data Bases and Taxonomy" in Proceedings International Association of Science and Technology for Development. 15th Applied Informatics Conference. Innsbruck. Austria. 1997.
[39]Perichinsky, G., Jimenez Rey, E.; Grossi, M. “Domain Standardization of Operational Taxonomic Units (OTUs) on Dynamic Data Bases” in Proceedings International Association of Science and Technology for Development. 16th Applied Informatics Conference. Garmisch-Partenkirchen.Germany. 1998.
[40]Perichinsky, G., Jimenez Rey, E.; Grossi, M. “Spectra of Objects of Taxonomic Evidence on the Dynamic Data Bases” in Proceedings International Association of Science and Technology for Development. 16th Applied Informatics Conference. Garmisch-Partenkirchen. Germany. 1998.
[41]Perichinsky, G., Jimenez Rey, E.; Grossi, M. “Application of Dynamic Data Bases in Astronomic Taxonomy” in Proceedings International Association of Science and Technology for Development. 17th Applied Informatics Conference. Innsbruck. Austria. 1999.
[42]Sawyer, R.A. “Experimental Spectroscopy”. Dover Publication. New York. 1963.
[43]SEI "Rationale for SQL Ada module Description language (SAMeDL)" Ver. 2.0 CMU/SEI-92-TR-16, oct 1992.
[44]Sokal, R.R., Sneath, P.H.A. "Numerical Taxonomy". W.H. Freeman and Company. 1973.
[45]Wiederhold,G. **“**Data Base Design”. McGraw-Hill Book Company.1983. Edition 1985.

[46]Williams,J.G. “Asteroids Family Identifications and Proper Elements”. Jet Propulsion Laboratory. Palomar-Leiden minor planets. Asteroids Edition. University of Arizona Press. p 1034. 1989.

[47]Fenton, N.E., Pfleeger, Sh.L. “Software Metrics”. PWS Publishing Company. 1997.

[48]Knezevic,Z., Milani,A. Farinella,P., Froechle,Ch., Froechle, Cl, “Asteroids Family Identifications and Proper Elements”. Icarus, 93, 316. 1991.

[49]Kitchenham,B., Pickard,L., Pfleeger, S.L. “Case studies for method and tool evaluation”. IEEE Software, 12(4) pp 52-62. 1995.

[50]Milani,A. “Asteroids Family Identifications and Proper Elements”.Celestial Mechanics. 57,59. 1993.

[51]Williams,J.G., “Asteroids Family Identifications and Proper Elements”.Icarus. 96,251. 1992a.

[52]Williams,J.G., “Asteroids Family Identifications and Proper Elements”.Icarus. Sbm. 1992b.

[53]Zappala, V, Cellino,A., Farinella,P., Knêzevîc,Z. "Asteroid Families. I. Identification by Hierarchical Clustering and Reliability Assessment " The Astronomical Journal, 100, 2030. 1990.

[54]Zappala, V, Cellino,A., Farinella,P., Milani,A., "Asteroid Families. II. Extension to Unnumbered Multiopposition Asteroids" The Astronomical Journal, 107, 772. 1994.

[55]Perichinsky, G., Orellana, R., Plastino, A.L., Jimenez Rey, E. y Grossi, M.D. "Spectra of Taxonomic Evidence in Databases." Proceedings of XVIII International Conference on Applied Informatics. (Paper 307-7-1 ).Innsbruck. Austria. 2000.