

Ana Pérez López
Mercedes de la Moneda Corrochano
Universidad de Granada
Ángel Moros Ramírez
Universidad Central de Venezuela

Application of the Cantor Set Theory in Making Decisions about the Collections Development

ABSTRACT: The procedures by which library collections can be evaluated are quite diverse, and some are rather complex. The Cantor Set Theory is applied with a structuralist approach as a methodological aid to decision-making about the collections development. The methodology suggested here makes use of local holdings information based on an evaluative study of the Spanish university library collections.

KEYWORDS: Knowledge representation; Collections Development; Sets Theory;

1. Introduction

In view of the advances in technology and communication, it is increasingly important to think in global terms about the organization of knowledge while acting locally in collections development. Libraries today are inter-linked, in a variety of ways, on a huge chain of access. Many of our libraries, however, face budgetary limitations. Therefore, when making decisions about the collections development, information service professionals should rely on systems that can represent knowledge and help evaluate the utility of resources.

Knowledge representation can be described as the group of processes of notational or conceptual symbolization of human knowledge in the context of any thematic area. The representation of knowledge in library collections includes classification, indexing, and all those linguistic and informational operations involved in the symbolic transfer of knowledge (Barité, 1997).

The mathematical theory of sets put forth by Cantor can prove very useful in providing a graphic representation of knowledge to be used as an aid in the processes tied to collection development. The set theory describes collections of cases or objects that constitute entities per se (sets). The set is defined, then, as an entity containing other entities, a definition well-suited to library collections. Yet the set theory is not limited to describing the relationships of the sets with their elements, but also the relationships among the elements and subsets themselves. In dealing with the notion of set, we must point out two basic types of relations: those of inclusion or membership (of an element in a set, or of a subset in a greater set) and those of intersection of single elements that belong to different sets (Dauben, 1990).

The structuralist approach entails interpreting the behavior and the properties of thematic areas. Behavior is understood in terms of temporal determinants, whereas the properties are the organizational principles of classification and order. The

structuralist framework serves to identify a pattern of relationships, as it is assumed that the efficient and effective the collection development depends on the identification of an underlying structure, which comprises the coincidental relations between demand, knowledge and the patterns of publication (Baughman, 1977). Once this structure is uncovered, a collection development policy can quickly be formulated.

To facilitate the comprehension of the mathematical model, we represent the behavior and the properties of a specific thematic area, in this case LIS, as described in an evaluative study of the library collections of the Universities of Salamanca and Granada, and Carlos III University of Madrid (Pérez López, 2002). The references used in the departmental scientific production and the current state of the collections were analyzed in terms of accessibility, localization and availability, and suitability of the collections insofar as subject headings, document type and language.

2. Application of Cantor's Mathematical Theory to a Thematic Area

We shall define a universe set: $\Omega = \{\text{thematic area of Library and Information Science in Spanish university libraries}\}$.

As subsets we have:

$\beta = \{\text{Scientific production}\}$ which, in turn, contains the subsets Carlos III {subset X}, Salamanca {subset Y} and Granada {subset Z}.

Subset Δ is defined as $\Delta = \{\text{the collections of the three university libraries studied}\}$, which comprises subsets Carlos III {subset A}, Salamanca {subset B} and Granada {subset C}. From the standpoint of collection development, each one of these subsets will contain other subsets that coincide with the variables and indicators evaluated. All the sets and subsets defined can present intersections. They are not disjointive.

3.1 Representation by Extension and Comprehension

Once established which sets are to be represented, it is necessary to consider a form of notation and representation that will adequately identify the component elements. The most commonly used methods for this purpose are comprehension, when the elements are characterized by a certain property \mathbf{P} (e.g. having been cited more than three times) cited more; and notation by extension when each one of the elements that belongs to the set is indicated. This is especially appropriate for the study of collections as it allows us to take note of the elements title by title.

We may define subset \mathbf{M} by comprehension as $\mathbf{M} = \{\text{periodical publications referenced more than three times by the Departments of Library and Information Science of the Universities studied}\}$.

Given our results, we may define subset \mathbf{M} by extension as $\mathbf{M} = \{\textit{Journal of the American Society for Information Science (JASIS); Scientometrics; Information Processing and Management; Information....}\}$.

3.2 Relationships of Membership and Inclusion

Having defined the elements of the universal set $\{X\}, \{Y\}, \{Z\}, \{A\}, \{B\}, \{C\}$, the following relationship of membership is generated: $\{X\} \in \Omega, \{Y\} \in \Omega, \dots$

Given the set $\Omega = \{\text{Thematic area of Library and Information Science in Spanish University Libraries}\}$, as subsets we have, among others, $\beta = \{\text{Scientific production}\}$ which, in turn, contains the subsets Carlos III $\{\text{subset X}\}$, Salamanca $\{\text{subset Y}\}$ and Granada $\{\text{subset Z}\}$.

Subset Δ is defined as $\Delta = \{\text{the collections of the three university libraries studied}\}$, which comprises subsets Carlos III $\{\text{subset A}\}$, Salamanca $\{\text{subset B}\}$ and Granada $\{\text{subset C}\}$.

We have: $\Omega \subset (A \cup B \cup C) \cup (X \cup Y \cup Z) \dots$ Taking into account the causal relations and interrelations of inclusion, we are able to represent the extension and scope of the thematic area studied.

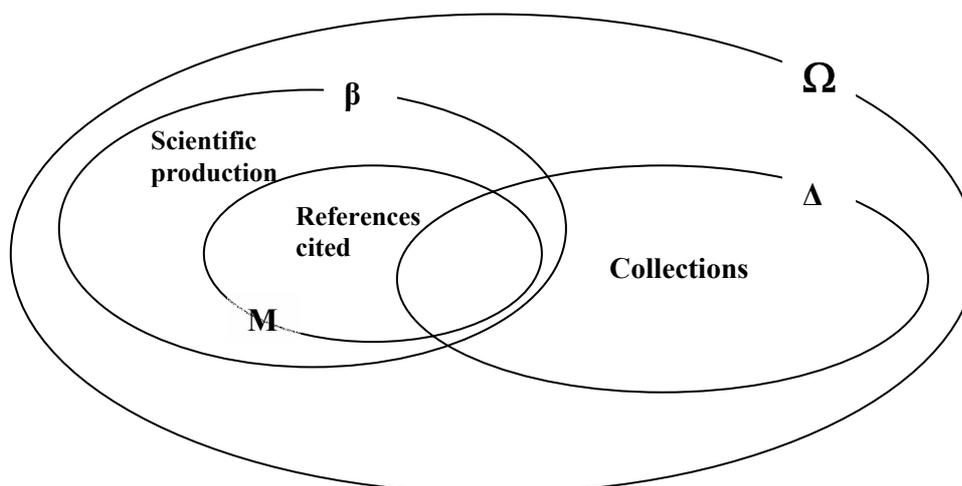


Fig. 1: The property of inclusion allows us to represent each and every one of the elements that make up the area of knowledge studied and the relationships of membership among them, as well as the degree of pertinence of the collection.

3.3. Intersection

In the cases of the Universities of Granada, Salamanca and Carlos III of Madrid, we have that:

$A = \{\text{Collection of the Department of Library and Information Science of the University Carlos III of Madrid}\};$

$B = \{\text{Collection of the Department of Library and Information Science of the University of Salamanca}\};$

And $C = \{\text{Collection of the Department of Library and Information Science of the University of Granada}\}$.

The intersection of A, B and C is expressed as $A \cap B \cap C$, and indicates the set comprising all the referenced works in the possession of the three university libraries studied, which we will denote as the basic core of the collections. This implies that those titles found in all three collections are equal elements: $x \in A; x \in B; x \in C$.

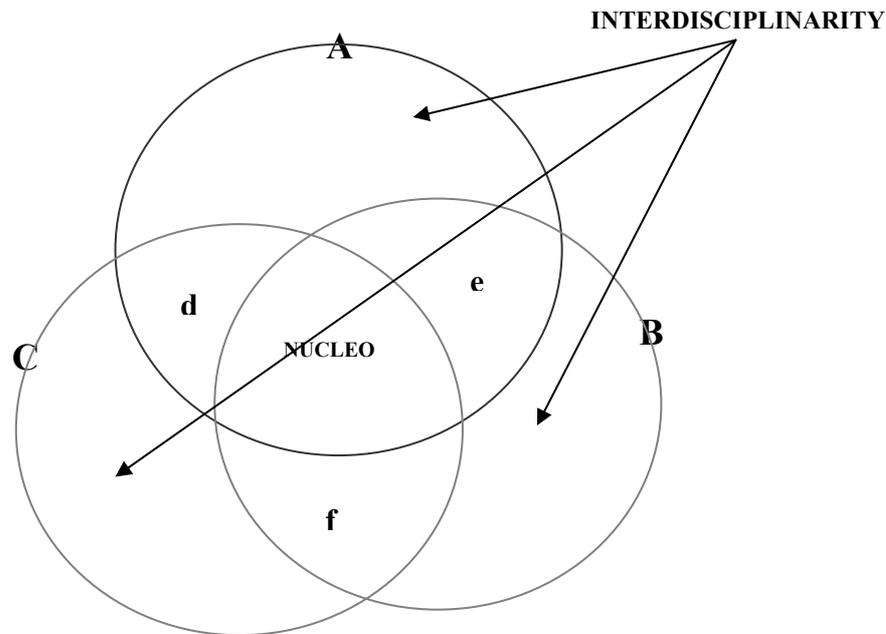


Fig. 2: An example of intersection of sets in which the element common to all three is the basic core of the collections in this thematic area.

Segment $e = A \cap B$ represents the holdings information and the lines of research commons to Carlos III and Salamanca; $f = B \cap C$ those common to Salamanca and Granada; and $d = A \cap C$ those common to Carlos III and Granada. The non-intersecting segments represent particular lines of research or interdisciplinary research connected with other thematic areas.

4. Conclusions

Cantor's theory proves to be a very useful mathematical tool for producing graphic representations of the current state of university library collections in a specific thematic area. The quality and limitations of end results will depend on the number of sets and subsets, the P properties applied, and on the indicators used, such as accessibility, organization, localization and availability.

As in our case there was an intersection of data on the scientific production of the departments with the present situation of the collection in itself, it is possible to

identify and represent the core of knowledge of a particular thematic area, the peripheral subjects of knowledge, and the degree of interdisciplinarity.

With this visualization of the state of collections, information science professionals can make well-informed decisions about acquisitions and withdrawals. This should be done with a consideration of individual library needs as well as with a cooperative sense of the needs of the university library network as a whole.

REFERENCES

- ABBOT, R. (1997). Information transfer and cognitive mismatch: a Popperian model for studies of public understanding. *Journal of Information Science*, 23 (2), 129-137.
- BARITÉ, M. (1997). *Glosario sobre organización y representación del conocimiento, clasificación, indización, terminología*. Montevideo: Universidad de la República.
- BAUGHMAN, J. C. (1977) *Toward a Structural Approach to Collection Development*. *College and Research Libraries* 38(3): 241-248.
- DAUBEN, J. W. *Georg Cantor: His Mathematics and Philosophy of the Infinite*. Princeton, NJ: Princeton University Press, 1990.
- JAENECKE, P. (1997) Knowledge Organization due to theory formation. In: *Organización del Conocimiento en Sistemas de Información y Documentación* (2) 39-55.
- MOYA ANEGÓN, F.; JIMÉNEZ CONTRERAS, E.; MONEDA CORROCHANO, M. (1998) Research fronts in library and information science in Spain (1985-1994). *Scientometrics* 42(2): 229-246
- PÉREZ LÓPEZ, A. (2002) *El Desarrollo de las Colecciones: Un estudio de Evaluación en las Bibliotecas Universitarias Españolas*. Diss. University of Granada. (aperez@ugr.es)