

REVIEW: “Towards mapping library and information science”

Information Processing and Management: an International Journal Special issue: Informetrics

Volume 42, Issue 6 Pages: 1614 - 1642

Frizo Janssens, Jacqueline Leta, Wolfgang Glänzel, Bart De Moor (2006)

(Review por M. Welling Flensburg)

e-mail: m.welling-flensburg@gmail.com

El artículo pretende explorar el campo de las Ciencias de la información y la biblioteconomía (CIB), y para ello presenta seis *clusters* o grupos de estudio, resultantes de procesar más de 1000 artículos, notas y publicaciones en general hechas sobre la materia a partir de cinco diarios científicos en el campo de las CIB. El artículo prosigue con el análisis de las representaciones gráficas hechas a partir de los *clusters*, que permiten visualizar información valiosa sobre peculiaridades, tendencias o incongruencias aparentes.

Para empezar, el artículo discute la aplicación de la lingüística cuantitativa (análisis estadístico de texto) en el contexto de la informetría y la bibliometría. Actualmente las técnicas más frecuentes (agrupación de conceptos por criterios como palabra, título o autor) están basadas en el análisis de palabras claves recurrentes, términos extraídos del texto, abstractos introductorios, cuerpos del texto, títulos o autores citados.

Diversos autores y sus contribuciones al campo se citan, para concluir que el campo de las CIB es muy heterogéneo, por incluir sub-disciplinas tales como Bibliometría, Cienciometría, Informetría, análisis de patentes y la más reciente *Webmetría*. En este contexto, el objetivo de los académicos del campo no debe ser aportar más teoría a la lingüística cuantitativa, sino extender el uso de metodologías de enfoque más amplio que puedan abarcar y procesar un *input* cada vez más heterogéneo, es decir, procesar bases de datos que contienen documentos heterogéneos, diferentes entre sí. Así pues el verdadero desafío no es la creciente cantidad de artículos, sino su variedad en un campo cada vez más híbrido.

Las principales técnicas discutidas por los autores en esta publicación están basadas en la estadística aplicada a la lingüística, y no en el procesamiento profundo del lenguaje. De esta manera, el estudio se basa en las siguientes preguntas en clave de tesis: 1. ¿Se puede caracterizar la supuesta heterogeneidad antes mencionada mediante la lingüística cuantitativa? 2. ¿Cuáles son los principales temas de investigación actual en el campo de las Ciencias de la Información? 3. ¿Han creado ya su propia terminología los nuevos temas de estudio surgidos recientemente? 4. ¿Se puede visualizar y representar la estructura cognitiva utilizando técnicas multi-variable? 5. ¿Cómo son representados en las principales publicaciones científicas los temas y sub-disciplinas del campo?

Para poder contestar a estas preguntas, después de procesar 938 artículos y notas extraídos de 5 diarios seleccionados, se presenta el vocabulario propio de las diferentes subdisciplinas del campo. El artículo también compara diversos métodos de agrupación de conceptos (*clustering*) y métodos de representación (*mapping*) con el objetivo de conseguir la óptima representación de la estructura cognitiva de este campo (representación de texto, pre-procesamiento de texto, extracción de textos, escalamiento multidimensional, *clustering*).

Los resultados se presentan de la siguiente manera: En primer lugar se presenta un índice general de 11151 términos o frases, que han sido identificados y categorizados. En segundo lugar se enseña una representación gráfica del “*set* de datos”, por ejemplo un mapa MDS (multidimensional scaling) de los 938 artículos o notas en dos y tres dimensiones, ilustrando las relaciones, la proximidad, o las semejanzas entre los artículos, por ejemplo. En el tercer lugar,

se dan a conocer los errores, las ambigüedades o las redundancias de datos que resultan del proceso. Finalmente los autores analizan la relación que existe entre los *clusters* encontrados y los diarios de los cuales fueron obtenidos. Con respecto a esto se concluye (observando los resultados gráficos) que las distancias observadas entre diferentes *clusters* y entre las publicaciones y los *clusters*, evidencian que el diario “Cienciometría” abarca buena parte del espectro de la bibliometría, Tecnometría, y la Infometría. Dicha postura coincide con los estudios y conclusiones de diferentes autores que afirman que Cienciometría incluye subdisciplinas como teoría infométrica, estudios empíricos, desarrollo de índices, estudios metodológicos, enfoques sociológicos, y políticas de educación y ciencia.

El artículo concluye que, después del procesamiento del *input* de 938 artículos y notas, seis diversos *clusters* son bien definidos: dos en bibliometría, (de los cuales uno de gran tamaño es del campo bibliometría – investigación – evaluación y otro más pequeño sobre temas de metodología y teoría del campo); dos grandes *clusters*, el primero sobre recuperación de información y el segundo sobre temas generales y miscelánea; y por último fueron identificados dos pequeños *clusters*, uno de ellos sobre *webmetría* y patentes, y el otro sobre estudios tecnológicos. Dentro del *cluster* de recuperación de información, se identificó un pequeño subgrupo sobre identificación de música que puede haber sido detectado por que la revista JASIST había publicado un tiempo antes una edición especial sobre música. Además de la identificación de *clusters* se hace mención al hecho de que interesantes patrones fueron encontrados con la ayuda de las representaciones gráficas.