Quaderns d'Intel·ligència Competitiva, Vigilància Estratègica, Científica i Tecnològica
QUIC&VECT 2 (2009), Edició 1, 72 - 73

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 per M. Welling Flensborg)
e-mail: m.welling-flensborg@gmail.com

L'article pretén explorar el camp de les Ciències de la informació i la biblioteconomia (CIB), i per a això presenta sis clústers de sectors o grups d'estudi, resultants de processar més de 1000 articles, notes i publicacions en general, fetes sobre la matèria a partir de cinc diaris científics en el camp de les CIB. L'article prossegueix amb l'anàlisi de les representacions gràfiques fetes a partir dels clústers de sectors, que permeten visualitzar informació valuosa sobre peculiaritats, tendències o incongruències aparents.

Per començar, l'article discuteix l'aplicació de la lingüística quantitativa (anàlisi estadística de text) en el context de l'informetría i la bibliometria. Actualment les tècniques més freqüents (agrupació de conceptes per criteris com a paraula, títol o autor) són basades en l'anàlisi de paraules claus recurrents, termes extrets del text, abstractes introductoris, cossos del text, títols o autors citats.

Diversos autors i les seves contribucions al camp s’esmenten, per concloure que el camp de les CIB és molt heterogeni, per incloure subdisciplines tals com a Bibliometria, Cienciometría, Informetría, anàlisi de patents i la més recent Webmetría. En aquest context, l'objectiu dels acadèmics del camp no ha de ser aportar més teoria a la lingüística quantitativa, sinó estendre l'ús de metodologies d'enfocament més ampli que puguin incloure i processar un input cada vegada més heterogeni, és a dir, processar bases de dades que contenen documents heterogenis, diferents entre si. La veritable fita a aconseguir no és pas la creixent quantitat d'articles, sinó la seva varietat en un camp cada vegada més híbrid.

Les principals tècniques discutides pels autors en aquesta publicació estan basades en l'estadística aplicada a la lingüística, i no en el processament profund del llenguatge. D'aquesta manera, l'estudi es basa en les següents preguntes en clau de tesi:

1. Es pot caracteritzar la presumpta heterogeneïtat abans esmentada mitjançant la lingüística quantitativa?
2. Quins són els principals temes d'investigació actual en el camp de les Ciències de la Informació?
3. Han creat ja la seva pròpia terminologia els nous temes d'estudi sorgits recentment?
4. Es pot visualitzar i representar l'estructura cognitiva utilitzant tècniques multivariable?
5. Com són representats a les principals publicacions científiques els temes i subdisciplines del camp?

Per poder contestar a aquestes preguntes, després de processar 938 articles i notes extretes de 5 diaris seleccionats, es presenta el vocabulari propi de les diferents subdisciplines del camp. L'article també compara diversos mètodes d'agrupació de conceptes (clustering) i mètodes de representació (mapping) amb l'objectiu d'aconseguir l'òptima representació de l'estructura cognitiva d'aquest camp (representació de text, preprocessament de text, extracció de texts, escalament multidimensional, clustering).

Els resultats es presenten de la següent manera: En primer lloc es presenta un índex general de 11151 termes o frases, que han estat identificades i categoritzades. En segon lloc s'ensenya una

representació gràfica del "set de dades", per exemple un mapa MDS (multidimensional scaling) dels 938 articles o notes en dos i tres dimensions, il·lustrant les relacions, la proximitat, o les semblances entre els articles, per exemple. Al tercer lloc, es donen a conèixer els errors, les ambigüitats o les redundàncies de dades que resulten del procés. Finalment els autors analitzen la relació que existeix entre els clústers de sectors trobats i els diaris dels quals van ser obtinguts. Respecte a això es conclou (observant els resultats gràfics) que les distàncies observades entre diferents clústers de sectors i entre les publicacions i els clústers de sectors, evidencien que el diari "Cienciometría" inclou bona part de l'espectre de la bibliometria, Tecnometría, i l'Infometría. L'esmentada postura coincideix amb els estudis i conclusions de diferents autors que afirmen que Cienciometría inclou subdisciplines com a teoria infomètrica, estudis empírics, desenvolupament d'índexs, estudis metodològics, enfocaments sociològics, i polítiques d'educació i ciència.

L'article conclou que, després del processament de l'|input de 938 articles i notes, sis diversos clústers de sectors són ben definits: dos en bibliometria, (dels quals el més gran és del camp de la bibliometria - investigació - avaluació i un altre més petit sobre temes de metodologia i teoria del camp); dos grans clústers, el primer sobre recuperació d'informació i el segon sobre temes generals i miscel·lània; i finalment van ser identificats dos petits clústers, un d'ells sobre webmetría i patents, i l'altre sobre estudis tecnològics. Dins del clúster de recuperació d'informació, es va identificar un petit subgrup sobre identificació de música que segurament ha estat detectat perquè JASIST havia publicat un temps abans una edició especial sobre música. A més de la identificació de clústers de sectors es fa menció al fet que interessants patrons van ser trobats amb l'ajut de les representacions gràfiques.