Journal Scholar Metrics: building an Arts, Humanities and Social Sciences journal ranking with Google Scholar data¹

Alberto Martín-Martín^{*}, Enrique Orduna-Malea^{**} and Emilio Delgado López-Cózar^{*}

^{*} albertomartin@ugr.es; edelgado@ugr.es

(Facultad de Comunicación y Documentación), Universidad de Granada, Campus de Cartuja, Granada, 18011 (Spain)

^{**} enorma@upv.es

(Instituto de Diseño y Fabricación), Universitat Politècnica de València, Camino de Vera s/n, Valencia, 46022 (Spain)

ABSTRACT

This paper describes the creation of “Journal Scholar Metrics” (JSM), a prototype web application that ranks journals in the areas of Arts, Humanities, and Social Sciences (AH&SS) on the basis of the citations their articles have received according to Google Scholar Metrics (GSM). To identify as many AH&SS journals as possible, a master list of 66,454 journals covered by various databases was developed. All AH&SS journals in that list were searched on GSM. Additionally, a series of keyword searches were carried out to identify journals covered by GSM which weren’t present in the master list. A total of 9,188 AH&SS journals with names written in Latin characters were found in the 2015 edition of GSM (which displays data about articles published between 2010 and 2014). Besides the journal-level indicators provided by GSM (H5-index and H5-median), several additional indicators were computed (H5-citations, H5-index and H5-citations without journal self-citations, and journal self-citation rate). Journals are displayed by subject categories and by country of publication. Quartiles were computed for each category, and journals in a category were further classified either as core (high affinity to the category) or related (partial affinity). A detail page for each journal is also available, displaying journal indicators, as well as a list of other databases where the journal is indexed.

1. INTRODUCTION

In April 2012 Google Scholar launched Metrics (commonly known as Google Scholar Metrics, or GSM for short), a complementary tool based on Google Scholar data and designed to become an “easy way for authors to quickly gauge the visibility and influence of recent articles in scholarly publications” (<http://googlescholar.blogspot.com.es/2012/04/google-scholar-metrics-for-publications.html>). Despite its original purpose, GSM was immediately perceived by the scientific community as a new bibliometric tool, since it computed the H-

¹ This work was supported by a four-year doctoral fellowship (FPU2013/05863) granted by the Ministerio de Educación, Cultura, y Deportes (Spain) and a postdoctoral fellowship (PAID-10-14), from the Polytechnic University of Valencia (Spain).

STI CONFERENCE, PARIS 2017

index for a wide range of scientific journals and other bibliographic sources (conferences and repository collections). In order to be included in GSM, journals or conferences must meet certain requirements: their articles must be indexed in Google Scholar; they must have published at least 100 articles over a period of five years; lastly, those articles must have received at least one citation. These criteria represent an effort (albeit arguable a crude one) to filter periodical publications from other types of documents indexed in Google Scholar.

Journal rankings in GSM are presented by languages. In the first two editions of the product there were ten different languages available (English, Chinese, Portuguese, German, Spanish, French, Korean, Japanese, Dutch, and Italian). In the 2015 edition (which displayed indicators calculated from documents published between 2010 and 2014) there were nine, because the Korean ranking was discontinued. In the 2016 edition (2011-2015) there were twelve languages, because five additional languages were added (Russian, Korean, Polish, Ukrainian & Indonesian) and two were removed: Italian, and Dutch.

For publications written in English, however, GSM also groups journals in 8 broad subject categories (Business, Economics & Management; Chemical & Material Sciences; Engineering & Computer Science, Health & Medical Sciences; Humanities, Literature & Arts; Life Sciences & Earth Sciences; Physics & Mathematics; Social Sciences), and 261 subcategories. For each journal, the list of documents with a citation count that is equal or higher than the h5-index of the journal can also be consulted. For each one of these documents, in turn, it is also possible to consult the list of citing documents.

Unfortunately, Google Scholar Metrics presents a rather restrictive visualization system. Only the top 100 sources according to their h5-index are displayed when selecting any of the language or broad subject category rankings. As for the subcategory rankings and the queries that can be made through the available search box, only the top 20 sources according to their h5-index are displayed. This effectively means that there is no straightforward method to learn how many journals are indexed in GSM, and it also means that most of the journals in GSM haven't been assigned to any subject category or subcategory (at least publicly). What's more, GSM doesn't allow grouping and ordering journals according to their country of publication.

In order to overcome these limitations, we took advantage of the various search features available in GSM's search box, and set out to collect all the Art, Humanities, and Social Science journals indexed in this product that we could find. The main goal of this project is, therefore, to gauge the extent of the journal coverage in Google Scholar Metrics, focusing our efforts in the areas of Arts, Humanities, and Social Sciences, which are also the areas that have been historically neglected by other journal rankings (JCR, SJR). The result of this work is available through a freely accessible web application which we called Journal Scholar Metrics (<http://www.journal-scholar-metrics.infoec3.es>)

Journal Scholar Metrics focus exclusively on journals belonging to the areas of Arts, Humanities, and Social Sciences, since these are the areas that have traditionally presented more difficulties in terms of bibliometric assessment, and the ones for which there is a greater lack of international, geographically and linguistically unbiased tools. In these disciplines, where research is often oriented towards local interests and where cultural peculiarities are determinant, researchers usually use national channels -and their native language- to communicate their results. This is why Google Scholar, thanks to its robots that automatically index all seemingly scientific publications without any kind of geographic or linguistic

restriction, is currently the most appropriate source of data to find evidence of scientific impact in these areas.

2. METHODS

Journal Scholar Metrics focus exclusively on Arts, Humanities, and Social Sciences journals indexed in GSM. It covers journals from all around the world and in all languages, providing that the name of the journal is displayed in Latin characters in GSM. Thus, names of journals written only in Arabic, Cyrillic, Chinese, Korean, or Japanese characters were excluded.

2.1 *Making a master list of journals*

In order to identify all journals that could potentially have been indexed in GSM, the following journal databases were consulted:

- a) Ulrichsweb: Global Serials Directory. It is considered the largest directory of periodic publications in the world. We retrieved the list of all existing scientific journals (academic/scholarly) indexed in the 162 categories («subjects») concerning Arts, Humanities, and Social Sciences. Data retrieved on September, 2013.
- b) Web of Science Master Lists: journals indexed in the 84 subject categories included in the Arts & Humanities Citation Index and the Social Science Citation Index. Data retrieved on May, 2015.
- c) SCImago Journal Rank: journals indexed in the 64 subject categories included in the areas related to the Humanities and Social Sciences. Data retrieved on December, 2015.
- d) The journal coverage lists of various international disciplinary-based databases: Anthropological Index Online, ARTBibliographies Modern, Avery Index, Communication & Mass Media, Econlit, Education Resources Information Center (ERIC), Geobase, Historical Abstracts, Index Islamicus, L'Année philologique, Library and Information Science Abstracts (LISA), MLA (Linguistics and Literature), Philosopher's Index, PsychINFO, Répertoire International de Littérature Musicale (RILM), Social Work Abstracts, Sociological Abstracts, SPORTDiscus, and Worldwide Political Science Abstracts. Data retrieved between October 2015 and February 2016.
- e) Google Scholar Metrics: All journals indexed in the categories “Humanities, Literature & Arts”, “Social Sciences”, and “Business, Economics & Management” were downloaded.

Using the information retrieved from these databases, a master list of journals was developed where titles were unified and duplicates were merged, resulting in an exhaustive database that combined the information contained in all the databases described above. The bibliographic information contained in Ulrichsweb was given precedence in case of any discrepancies in the data available from other sources.

2.2 *Finding journals in GSM*

First, we searched all journals in our master list on GSM by their names (GSM doesn't support searching by ISSN, publisher, or any other field). Secondly, in order to find journals in GSM which weren't already in our master list, we performed topic searches using meaningful keywords from each discipline. The selection of these keywords was based on the analysis of the frequency of words which appeared in the names of the journals in the master list. These keywords were also translated to several languages: English, French, Spanish, German, Italian, Portuguese, Polish, and Czech.

2.3 Data processing

Once we identified all the available journals in GSM and matched them to the journals in our master list (relying on journal title comparisons), we proceeded to download the following data:

- Journal metrics (H5-index and H5-median).
- Bibliographic information of the articles published in each of those journals, including the number of citations these articles have received. Only the articles that contribute to the H5-index of the journals are displayed in GSM (n° of citations $\geq$ H5-index), and therefore only those were downloaded.
- Bibliographic information of the documents that cite the aforementioned articles.

After downloading this information, journal self-citations were identified and computed.

2.4 Journal classification

All journals found in GSM were assigned to at least one of 13 custom subject categories in Social Sciences (Anthropology, Communication, Economics, Business, & Management, Education, Geography & Urbanism, Law, Library & Information Science, Political Science, Administration, & International Relations, Psychology, Social Work, Sociology, Sports Science, Multidisciplinary) and 9 custom subject categories in Arts & Humanities (Archaeology & Prehistory, Arts, Classical Studies, History, Literature, Language & Linguistics, Philosophy, Religion, Multidisciplinary).

A journal assigned to a subject category is considered either a core journal, or a related journal. Core journals are those publications which are considered essential in a category. Related journals are those which are linked to the category because some (but not all) of the articles they publish are relevant to the category. In order to make these distinctions we relied on the classifications made by other databases (Ulrichsweb, Web of Science, Scopus, disciplinary-based databases, etc.). As a general rule, journals were considered as core in a given category when they met at least one of the following criteria:

- a) The name of the journal contains meaningful keywords from the discipline. Word frequency lists extracted from lists of journal names were used. Polysemic words (such as “information”) were excluded.
- b) The journal is considered as a core by the specialized database in the field, and it is also classified in Ulrichsweb in the corresponding category. Only some specialized databases make the core/related distinction (or similar ones). Therefore, none the journals covered by the databases that don’t make this distinction met this specific criterion.
- c) The journal is classified in the corresponding categories in at least three of the following databases: Ulrichsweb, WoS, SJR, GSM (only journals written in English), and the specialized database for the field (as core or related).

Journals that didn’t meet any of the previous criteria but were classified in the corresponding categories in two of the aforementioned databases were considered as related.

Lastly, a manual revision was also carried out to check and fix incongruities. Changes were made on a case-by-case basis after examining the scope of the journals (articles published, relationships with other journals...)

2.5 Bibliometric indicators

Six bibliometric indicators are used:

1. *H5-index*: h-index for articles published in the last 5 complete years. It is the largest number h such that h articles published in 2010-2014 have at least h citations each.
2. *H5-median*: median number of citations for the articles that make up a journal's h5-index.
3. *H5-citations*: sum of the number of citations received by all the articles that make up the journal's h5-index.
4. *H5-index without journal self-citations*: computed in the same way as the H5-index, but excluding citations that come from articles published in the same journal.
5. *H5-citations without journal self-citations*: computed in the same way as the H5-citations, but excluding citations that come from articles published in the same journal.
6. *Journal self-citation rate*: Percentage of citations that come from articles published in the same journal.

3. RESULTS

We found 9,188 Arts, Humanities, and Social Sciences journals with titles written in Latin characters in the 2015 edition of Google Scholar Metrics (which displays data from articles published between 2010 and 2014). We are confident we found most of them, but it is quite possible that we may have missed some of them, especially if their h-index is low (5 or lower) or if they use obscure names for the title of the journal (proper nouns, words coming from Latin or ancient Greek, acronyms, etc.).

Journals are presented both by subject category and country rankings. These options can be selected directly from the Home Page of the web application (Fig.1).

Figure 1: Home Page Journal Scholar Metrics

3.1 Browsing by subject category

Subject rankings can be sorted by any of the six bibliometric indicators previously described, as well as by the country of publication, just by clicking on their respective column headers (Fig. 2).

Figure 2: Page Subject categories

Above the results table users will find three elements:

- A check box to show core or related journals. When the box is checked, core journals are highlighted using bold letters, and related journals are slightly translucent. When the box is not checked, only core journals are shown.
- A drop-down list to filter by country of publication.
- A search box from which it is possible to search any journal in the current category.

Clicking on the title of a journal in any ranking takes users to the profile page of that journal (Fig 3). That page will show basic information for that journal, like: Name of the journal and country of publication, metrics, subject categories to which the journal has been assigned and position it occupies in the category, and a list of databases where this journal is also indexed, as well as the categories these databases have assigned to the journal, according to their respective classification schemes.

Figure 3: Journal Profile Page

3.2 Browsing by country

The Home Page also presents a world map from which users can access all journals in this product published in a given country, regardless of their topic. Since there might be difficulty in clicking some countries in the world map because of their size, it is also possible to display continent maps, as well as smaller territories. The shade of blue indicates the quantity of journals in JSM published in that country, and a box with the name of the country and the exact number of journals collected so far in JSM will pop up when you hover over it with the cursor. When a country is selected, all journals published in that country will be shown.

4. CONCLUSIONS

In previous studies (Martín-Martín et al. 2014; Orduna-Malea et al. 2016), we have deeply described the underlying philosophy embedded in all Google's academic products. These tools have been created in the image and likeness of Google's general search engine: fast, simple, easy to use, and last but not least, accessible to everyone free of charge. GSM follows all these precepts, and it is a hybrid between a bibliometric tool (indicators based on citation counts), and a minimalist information product with few features, closed (it cannot be customized by the user), and simple (navigating it only takes a few clicks) (Jacsó 2012; Delgado López-Cózar & Cabezas 2012; 2013).

Journal Scholar Metrics was designed to prove that some of GSM's shortcomings can be overcome. This product uses both advanced query search procedures and journal data processing to solve GSM's limitations. Moreover, this product is focused in the Arts, Humanities and Social Sciences, precisely the areas that have been more neglected in terms of coverage by other traditional bibliographic databases, thus surfacing the wealth of data available in Google Scholar that would have otherwise stayed unknown to most users.

As a proof of this, the number of journals found in the subject categories ("Humanities, Literature & Arts", "Social Sciences", and "Business, Economics & Management") offered by GSM (2010-2014 edition) only amounts to approximately 1,800. By perusing the language rankings we can find around 300 more. However, Journal Scholar Metrics displays 9,188 journals in these categories. Moreover, the journals available in GSM are only a fraction of the journals available in Google Scholar, because GSM's inclusion criteria leave out a great number of Humanities and Social Sciences journals that don't publish at least 100 in five consecutive years.

In addition to giving more visibility to these journals, Journal Scholar Metrics has offered to the community the following methodological findings:

- A. A procedure to compute the impact of journals excluding journal self-citations using Google Scholar data, presenting two additional indicators in this regard: sum of citations to articles that contribute to the h5-index (including and excluding journal self-citations), as well as the percentage of citations that are journal self-citations. The h5-index excluding journal self-citations is also displayed.
- B. A new journal classification system: journals are classified as core or related on a given subject category depending on the level to which they are linked to said subject category. We based this scheme on the classification systems of other multidisciplinary and specialized bibliographic databases and journal directories.

REFERENCES

Delgado López-Cózar, E. & Cabezas-Clavijo, A. (2013). Ranking journals: could Google Scholar Metrics be an alternative to Journal Citation Reports and Scimago Journal Rank? *Learned Publishing*, 26, 101-113.

Delgado López-Cózar, E. & Cabezas-Clavijo, Á. (2012). Google Scholar Metrics: an unreliable tool for assessing scientific journals. *El Profesional de la Información*, 21, 419–427.

Jacsó, P. (2012). Google Scholar Metrics for Publications: the software and content features of a new open access bibliometric service. *Online Information Review*, 36, 604–619.

Martín-Martín, A., Ayllón, J.M., Orduna-Malea, E. & Delgado López-Cózar, E. (2014). Google Scholar Metrics 2014: a low cost bibliometric tool. *EC3 Working Papers*, 17. Retrieved April 28, 2017 from: <http://arxiv.org/pdf/1407.2827>

Orduna-Malea, E.; Martín-Martín, A.; Ayllón, J.M. & Delgado López-Cózar, E. (2016). *La revolución Google Scholar: destapando la caja de Pandora académica*. Granada: UNE.