

Big-Data y patentes: método para la normalización de las instituciones OCDE Top 100

María-Eugenia Espinosa-Calvo¹, Rocío Gómez-Crisóstomo², María-Antonia Ovalle-Perandones³, Félix Moya-Anegón⁴, Elena Corera-Álvarez⁵

¹<https://orcid.org/0000-0002-8955-4637> + SCImago Lab; Departamento de Información y Comunicación, Facultad de Derecho, Universidad de Extremadura, Cáceres, España. maespinosac@unex.es.

² <https://orcid.org/0000-0002-3258-6283> + Departamento de Información y Comunicación, Facultad de Ciencias de la Documentación y la Comunicación, Universidad de Extremadura, Badajoz, España. mrgomcri@unex.es.

³<https://orcid.org/0000-0002-6149-4724> + Departamento de Biblioteconomía y Documentación, Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid, Madrid, España. maovalle@ucm.es.

⁴<https://orcid.org/0000-0002-0255-8628> + SCImago Lab, felix.moya@scimago.es.

⁵ <https://orcid.org/0000-0002-2618-339X> + SCImago Research Group – IPP, Consejo Superior de Investigaciones Científicas, Madrid, España. ecorera@gmail.com.

Tipo de contribución: comunicación

Palabras clave: patentes; Big Data; normalización; instituciones; PATSTAT; Scopus

Resumen:

El incremento experimentado por el número de patentes registradas en el mundo en los últimos años (World Intellectual Property Organization, 2018), ha ocasionado que su análisis se convierta en una tarea ardua y dificultosa, propia del Big Data, principalmente, a la hora de realizar análisis patentométricos con información sobre los “applicants”, es decir, sobre las entidades generadoras de dichas patentes. La razón de ello es, por un lado, el complejo y dinámico entramado que caracteriza a estas instituciones, muchas de ellas multinacionales, instituciones con filiales, etc., y por otro, la falta de normalización presente en los datos correspondientes a las mismas, es decir, la ausencia de un proceso de depuración que trate de recoger bajo un mismo nombre o registro todas las patentes que realmente se han solicitado desde cada institución.

El presente trabajo pretende dar respuesta a ese problema, y propone un procedimiento de normalización que posibilita la realización de análisis patentométricos a nivel institucional con resultados altamente fidedignos. Para ello, se realiza una prueba piloto centrada en el listado de las instituciones Top 100 de PATSTAT de la OCDE (European Patent Office, 2018). A pesar de proceder de una fuente fiable, se parte de la hipótesis de que los “applicants” o solicitantes de dichas instituciones no están normalizados.

Para proceder a la normalización de las 100 instituciones Top, y que posteriormente se presume que podrán ser extrapolados a otros contextos, se establecen dos procedimientos estrechamente relacionados, pero claramente diferenciados entre sí, uno manual y otro automático.

El objetivo principal del procedimiento manual es la identificación de la situación actual de cada entidad (si es multinacional, si dispone de filiales, etc.), para posteriormente crear su registro o registros de manera individual o estructural, en caso de filiales, substituciones, etc., y establecer las estrategias de búsqueda pertinentes, para aglutinar bajo ese registro todas las patentes correspondientes a la entidad creada en el registro. Tras este primer procedimiento, se identifican las denominadas cadenas, aquellas estrategias de búsqueda que recuperan de forma unívoca patentes pertenecientes a una misma entidad, y que serán utilizadas en la fase automática.

En el procedimiento automático se realizan oleadas de búsquedas plenamente exitosas, pues se basan en las cadenas previamente identificadas, de manera que la mayor parte del trabajo de normalización queda cubierto por las formas más habituales de consignación del “applicant” en las solicitudes de patentes.

No obstante, después de cada procedimiento se realiza una revisión manual de los “applicants” restantes, de manera que todos aquéllos que no han sido incorporados a ninguna de las cadenas aceptadas, se asignan de manera manual a su registro correspondiente.

El resultado de este trabajo es, por un lado, el listado depurado y corregido de las cien instituciones Top establecido por la OCDE, con la eliminación de los “applicants” incorrectos y la creación de la estructura de la entidad a través de registros, y por otro, la consecuente asignación correcta de sus patentes tras el proceso de normalización.

Bibliografía:

European Patent Office. (2018). *Statistics*. Obtenido de <https://www.epo.org/about-us/annual-reports-statistics/statistics.html>

World Intellectual Property Organization. (2018). *World Intellectual Property Indicators 2017 Patents*. Ginebra (Suiza): World Intellectual Property Organization.