

Un viaje a la ecología del movimiento a través de la minería de texto

Rocío Joo* Simona Picardi† Matthew E. Boone‡ Thomas A. Clay§
Samantha C. Patrick¶ Vilma S. Romero-Romero|| Mathieu Basille**

El campo de la ecología del movimiento ha experimentado un crecimiento sin precedentes en la última década. El desarrollo masivo de dispositivos que permiten seguir animales y personas (e.g. GPS, TDRs, acelerómetros, cámaras de video), y de herramientas de programación y análisis, han originado como consecuencia un mayor acceso a grandes volúmenes de datos, generación de resultados y publicación de artículos científicos. Quisimos revisar de manera cuantitativa el estado de la ecología del movimiento como ciencia, analizando diferentes aspectos como los temas investigados, los softwares y los dispositivos de seguimiento utilizados.

En esta presentación se describe la metodología utilizada para lograr este objetivo. A través de criterios de búsqueda que explicaremos brevemente, identificamos más de 8000 publicaciones científicas en inglés en ecología del movimiento realizadas en la última década. Para el análisis utilizamos varias técnicas de minería de texto. El énfasis en esta presentación será principalmente en dos enfoques: “diccionario” y modelamiento de temas. Un enfoque tipo “diccionario” fue utilizado para revisar aspectos como los softwares y los dispositivos de seguimiento. Para cada aspecto se definió un diccionario como una lista de categorías (e.g. “R”, “Matlab”), en la que, a su vez, cada categoría estuvo definida por una lista de términos (e.g. “R Software”, “R Development Core”). Para cada documento, de encontrarse uno de los términos en el texto (e.g. en la sección Materiales y Métodos o en el Resumen), el documento fue asociado a la categoría correspondiente al término. Para estudiar los temas investigados en los artículos se utilizó un enfoque distinto, puesto que ellos no fueron definidos a priori. La identificación de temas latentes y desconocidos se realizó mediante el ajuste de modelos de Asignación Latente de Dirichlet (Latent Dirichlet Allocation) a los resúmenes de las publicaciones. Estos modelos son esencialmente modelos bayesianos jerárquicos de tres niveles (temas, palabras, resumen) para documentos. El modelo define cada tema como una mezcla de palabras, donde ciertas palabras tienen mayor probabilidad de ocurrencia en ciertos temas. Como cada resumen podría estar compuesto por uno o más temas, las palabras utilizadas en el resumen reflejan los temas subyacentes.

Los modelos de Asignación Latente de Dirichlet nos permitieron caracterizar 15 temas desarrollados en ecología de movimiento. Asimismo, encontramos una creciente hegemonía de R respecto a los otros softwares y de GPS respecto a otros dispositivos de seguimiento.

Palabras clave:

minería de texto - ecología del movimiento - modelamiento de temas

*University of Florida, USA, rocio.joo@globalfishingwatch.org

†University of Florida

‡University of Florida

§University of Liverpool

¶University of Liverpool

||Universidad de Lima

**University of Florida