

Análisis de las etiquetas de contenido de un banco de imágenes: *Agefotostock*

Blanca Rodríguez Bravo¹, Crispulo Travieso Rodríguez²

¹orcid.org/0000-0002-9476-7602. Universidad de León. blanca.rodriguez@unileon.es

²orcid.org/0000-0002-0774-0728. Universidad de Salamanca. ctravieso@usal.es

El objetivo fundamental de este trabajo es analizar las etiquetas asignadas a los documentos fotográficos en una agencia comercial concreta, Agefotostock, con el fin de conocer en qué aspectos inciden y cuáles son las características más comunes de estos descriptores libres. De entre las nueve categorías temáticas ofrecidas, se optó por la denominada *Gente*, pues se infirió a priori que sería la que más enfoques podría aportar sobre una misma fotografía. Finalmente se optó por seleccionar, dentro de la categoría mencionada, las primeras 100 fotos recuperadas en la última quincena de abril de 2019 ordenadas en función de su popularidad, lo que nos aseguraba contar con una muestra suficientemente amplia. El estudio de la muestra se abordó tomando los siguientes elementos de análisis: número de etiquetas asignadas por fotografía; tipo y morfología de los términos; número de términos por etiqueta; porcentaje de etiquetas de contenido frente a etiquetas de formato o técnicas; porcentaje de etiquetas denotativas y connotativas, coincidencia de etiquetas y sinonimia.

Se recopilaron un total de 6775 etiqueta, siendo la media de etiquetas/foto más elevada de lo que habíamos estimado en un principio. De estas etiquetas, 1921 son diferentes. Tras eliminar las etiquetas con errores se han analizado 1913. La media de palabras clave por documento supera las 67, lo que da muestra de la exhaustividad de la indización de este banco de imágenes. La morfología es extremadamente variada, habiéndose identificado 54 formas entre las que predominan las siguientes: sustantivo común (946 ocurrencias), adjetivo (276 ocurrencias) y la fórmula nombre+adjetivo (137 ocurrencias). En relación a la estructura de las etiquetas, la más frecuente es la de unitérminos o etiquetas simples (78,52%) frente a las etiquetas compuestas o sintagmáticas (21,48%). En el caso de las palabras clave compuestas dominan las conformadas por dos términos (11,87%) seguidas por las estructuradas en tres partes o cuatro partes (6,83%) y (2,60%) respectivamente. La presencia de etiquetas de 5 y 6 palabras es meramente testimonial. Mayoritariamente las etiquetas se formalizan en singular (89,85%). Este dato se ha obtenido tras eliminar del recuento los nombres propios y las formas verbales y adverbiales. La presencia de etiquetas que reflejan aspectos compositivos y técnicos es sorprendentemente reducida, no alcanzando el 2%. La gran mayoría se dedican a describir el contenido de la imagen por medio fundamentalmente de etiquetas denotativas (70,17%) frente a un 29,83% de palabras clave connotativas, destinadas a describir lo que la imagen sugiere o representa en nuestro contexto cultural. En cuanto a la sinonimia, se identificó un elevado índice de términos redundantes en la descripción, siendo además muy poco uniformes.