

TÉCNICAS DE WEB SCRAPING APLICADAS A LAS ESTADÍSTICAS DE INSERCIÓN LABORAL

Víctor Montañés Cobo

Instituto de Estadística y Cartografía de Andalucía
victor.montanes@juntadeandalucia.es

José E. Molina Trapero

Instituto de Estadística y Cartografía de Andalucía
josee.molina@juntadeandalucia.es

Rosa Cánovas Balboa

Instituto de Estadística y Cartografía de Andalucía
rosa.canovas@juntadeandalucia.es

Juan De Dios Ramos Poyatos

Instituto de Estadística y Cartografía de Andalucía
juand.ramos@juntadeandalucia.es

PALABRAS CLAVE

Web Scraping, Nuevas técnicas, Base de Datos Longitudinal de Población de Andalucía, Inserción laboral, Egresados universitarios

RESUMEN

La estadística de “Inserción laboral de los egresados en universidades públicas de Andalucía” ofrece, a partir de la explotación estadística de los ficheros de matrícula y rendimiento académico de los estudiantes de las universidades públicas de Andalucía, y tras el enlace con sus datos de empleo y desempleo disponibles en la *Base de Datos Longitudinal de Población de Andalucía*, información relativa al proceso de inserción laboral de los egresados universitarios que posibilita el análisis de los diferentes aspectos que caracterizan y determinan el acceso de dicho colectivo a los mercados de trabajo.

Para cada promoción de egresados se proporciona un periodo de seguimiento de cinco años, con información detallada sobre la transición de la educación universitaria al mercado laboral, la empleabilidad y tasas de inserción laboral.

Las fuentes administrativas disponibles para este análisis son insuficientes para cubrir exhaustivamente todas las posibles situaciones de relación con la actividad de este colectivo. Así, por ejemplo, no se dispone de información sobre los egresados universitarios que trabajan en el extranjero, puesto que no cotizan a la Seguridad Social española, ni sobre los mutualistas pertenecientes a MUGEJU o ISFAS y otros colectivos de profesionales colegiados que pueden ser mutualistas de mutualidades de previsión social alternativas. Estos déficits pueden dar una medición sesgada del grado de inserción laboral de determinadas titulaciones.

Durante la producción de esta publicación se hizo una primera prueba de uso de técnicas de webscraping para suplir este déficit. Entre las acciones acometidas, se usó la información publicada

en BOE de nombramientos de miembros de MUGEJU e ISFAS como indicadores de posibles situaciones de alta laboral y también las webs de perfiles profesionales *linkedIN* y *Researchgate* como indicadores de residencia o trabajo en el extranjero o incluso de estudios no recogidos en la estadística de egresados. Así mismo, se usó la información de la base de datos de Tesis Doctorales (TESEO) para completar la información de los doctorados incluidos en el fichero de explotación, enlazando cada doctorando con su tesis para recuperar las variables rama de enseñanza y ámbito/campo de estudio a partir del título de la tesis y el departamento al que estaba adscrita.

Esta nueva técnica abre una vía de posibilidades para aprovechar nuevas fuentes de información que completen lo reportado en los registros administrativos y que no se habían planteado debido a la dificultad para obtenerlas.