

CARTOBOT: BUSCANDO NOTICIAS SOBRE CAMBIOS EN LAS CARRETERAS

Diseño del sistema

IRENE MARTÍNEZ ENCINAS

ETSI en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid
irene.mencinas@alumnos.upm.es

EATRIZ ASENSIO DE LEÓN

ETSI en Sistemas Informáticos, Universidad Politécnica de Madrid
beatriz.asensio.deleon@alumnos.upm.es

FRANCISCO SERRADILLA GARCÍA

ETSI en Sistemas Informáticos, Universidad Politécnica de Madrid
francisco.serradilla@upm.es

MIGUEL ÁNGEL MANSO CALLEJO

ETSI en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid
m.manso@upm.es

ALICIA GONZÁLEZ JIMÉNEZ

Instituto Geográfico Nacional
agjimenez@fomento.es

GONZALO MORENO VERGARA

Instituto Geográfico Nacional
gmvergara@fomento.es

FRANCISCO JAVIER GARCÍA GARCÍA

Instituto Geográfico Nacional
fjgarcia@fomento.es

FRANCISCO JAVIER GONZÁLEZ MATESANZ

Instituto Geográfico Nacional
fjgmatesanz@fomento.es

RESUMEN: Desde el año 2017, la Subdirección General de Geodesia y Cartografía del IGN ha fijado el objetivo prioritario de mejorar de los procesos de mantenimiento y actualización de los productos que por mandato legal tiene encomendados. Por esa razón, a finales de ese año se lanzó el **Programa para el cambio productivo de Bases de Datos de Información Geoespacial (BDIG)**, que tiene como objetivo implantar una metodología de actualización continua de los objetos geográficos recogidos en las bases de datos de Información Geoespacial. De esta manera se consigue cambiar la antigua forma predictiva, en la que se planifica un área de actualización en plazos fijos de actuación, a otra adaptativa, en la que las operaciones de mantenimiento se realizan donde se han detectado modificaciones en el territorio.

El proyecto **CARTOBOT**, vinculado al dicho programa, probablemente sea la parte más innovadora de la gestión de las BDIG, puesto que es la encargada de obtener cambios en el territorio por técnicas no convencionales, como es la explotación de información no estructurada mediante la utilización de *bots* y técnicas de Inteligencia Artificial (*o Deep Learning/Machine Learning*).

Este proyecto se está llevando a cabo mediante la fórmula de convenio de investigación entre el IGN y el grupo MERCATOR de la UPM en el que colaboran miembros de ambas organizaciones siguiendo una metodología AGIL.

En esta comunicación se presenta el diseño preliminar de un sistema basado en *bots* (programas informáticos que efectúan automáticamente tareas repetitivas en Internet) para la detección de noticias y otras comunicaciones sobre posibles cambios producidos en infraestructuras viarias. Estos *bots* exploran periódicamente redes sociales, páginas web oficiales y otras páginas web que contienen comunicaciones relativas a obras, licitaciones y realizan un procesamiento automático de la información para detectar la ubicación geográfica de esos cambios. Este sistema se desarrolla para explorar el potencial de la información no estructurada de internet como medio para conocer los cambios en el territorio, principalmente, para actualizar la red de transporte por carretera asfaltada.

En esta comunicación se describen los trabajos previos al diseño del sistema: revisión, ampliación y priorización de enlaces web de referencia en los que se tiene constancia de la publicación de noticias relacionadas con la temática. Seguidamente se justifica la selección de un conjunto representativo de estas fuentes como objetivos de los primeros *bots* a desarrollar con la finalidad de explorar la diversidad de fuentes, formatos y estructuras de datos que permitan dar generalidad al resto de fuentes. En este primer conjunto de fuentes hay redes sociales, páginas web que publican noticias con alta frecuencia y que precisan ser filtradas cuidadosamente para identificar potenciales cambios en la red de viales, páginas web con alto contenido dinámico basado en código JavaScript, y páginas HTML bien estructuradas o en formatos XML y JSON. El diseño de este conjunto de *bots* iniciales ha permitido extraer las primeras conclusiones y diseñar el sistema que se presentará aquí.

Este consta de tres módulos: planificación de trabajos, almacenamiento estructurado y análisis de información. El módulo de planificación de trabajos permite una integración flexible de tareas de *scraping* (extracción) de información, soportando distintos modelos, fuentes de información y frecuencias de consulta. El módulo de almacenamiento estructurado consiste en una base de datos definida mediante los modelos de información creados en la fase inicial. El módulo de análisis realiza un procesamiento periódico de la información incorporada en el sistema para determinar si las noticias o eventos identificados tienen entidad suficiente para generar una comunicación que alimente al sistema de incidencias del IGN, inciGeo, o si por el contrario, se ha de comunicar al servicio técnico humano que revise la información y asista al agente para aprender o dictaminar si es o no una incidencia que tenga que ser gestionada por dicho sistema.

PALABRAS CLAVE: Cartobot, *bot*, web, *scraping*, carreteras, planificación de trabajos.