

GEOBIG. GESTIÓN DE GRANDES VOLÚMENES DE DATOS ABIERTOS

Un enfoque diferente a la gestión de información geográfica, preparándo datos para la inteligencia artificial

ALEJANDRO GUINEA DE SALAS
Geograma
Alejandro.guinea@geograma.com

RESUMEN: La cantidad de información geográfica disponible está creciendo a pasos agigantados. Cada día se publican nuevos datos en la red, a menudo a través de portales de datos abiertos. Estos datos implican unas inversiones no desdeñables, tanto para su creación como para su mantenimiento, que sin embargo son difíciles de reaprovechar por la industria, dada la heterogeneidad de los propios datos, pero también de las formas de acceder a ellos.

Paralelamente, la tecnología relacionada con la inteligencia artificial ha sufrido un gran avance, debido sobre todo a los servicios en la nube proporcionados por las grandes operadoras.

Sin embargo, se hace complejo aplicar estas tecnologías a grandes conjuntos de datos abiertos, por su dispersión y la dificultad de procesarlos. Por ello, es necesario preparar los datos de forma previa a cualquier proceso de análisis.

Durante la presentación se expondrá cómo se han preprocesado más de 125 000 conjuntos de datos de todo el mundo, que contienen más de 250 millones de elementos, y se han preparado e integrado en una arquitectura capaz de aplicar procesos de forma automatizada, independientemente de su origen, formato, o sistema de referencia, incluyendo procesos de aprendizaje automático.

Se expondrán aspectos como la identificación y acceso a los datos, los formatos y sistemas de referencia encontrados, y los procesos seguidos hasta obtener un sistema capaz de procesar miles de capas en paralelo. Este sistema permite obtener el máximo valor de la ingente cantidad de datos geográficos disponibles. Para ello, se están desarrollando procesos de enriquecimiento de metadatos, como la definición precisa de la cobertura de la información, o la detección automática del proveedor.

Aunque sin ninguna duda los datos enmarcados en INSPIRE facilitan enormemente la aplicación de este sistema, la arquitectura planteada y los algoritmos desarrollados tienen aplicación más allá de INSPIRE. Especialmente en los casos donde INSPIRE no se aplica por estar los datos fuera del ámbito de la directiva o por tratarse de datos, ya sean internos o abiertos, con más detalle del recogido en las especificaciones de datos definidas por la comisión europea.

PALABRAS CLAVE: *Big data*, metadatos, procesos en la nube, inteligencia artificial, *machine learning*, aprendizaje automático, información geográfica, datos abiertos, *open data*.

