Metodología de creación del grafo de conocimiento geográfico de Aragón (ICEARAGON)

Descripción técnica de las fases de desarrollo de ICEARAGON

PORTOLÉS, David; MARTÍNEZ, Rafael; LÓPEZ, Fernando

N.º del tema de las jornadas: Nº. 1 - Aplicaciones y tecnología

Resumen

En esta comunicación se presenta la metodología seguida para la creación del grafo de conocimiento geográfico (GeoKG) de Aragón (ICEARAGON).

El objetivo principal de ICEARAGON es el deseo de comprender la realidad territorial de Aragón para mejorar el proceso de toma de decisiones por parte de la Administración autonómica y de los ciudadanos.

La primera fase consiste en la normalización y abstracción de las realidades territoriales a incluir, para dar paso a continuación a la creación de un modelo lingüístico formado por taxonomías, vocabularios, listas controladas o tesauros (propios o externos, como GBA Thesaurus). Estas clasificaciones (desarrolladas mediante SKOS) conforman la base para la creación del GeoKG.

Seguidamente se ha evolucionado hacia un modelo conceptual mediante la creación de diferentes ontologías modeladas con OWL con el objetivo de dotar de mayor precisión semántica a los conceptos y relaciones anteriores.

En ICEARAGON se han desarrollado ontologías que describen elementos específicos (especialmente para elementos territoriales característicos de Aragón, como ibones, paisaje, glaciares, entre otros), reutilizando aquellos conceptos más generales de ontologías ya existentes y extendidas (BTN100, Wikidata, schema.org).

En esta fase de modelado, se han planteado algunas reglas geográficas (conforme al paradigma Evento-Condición-Acción), así como restricciones y validaciones (SHACL) si bien está previsto que esta parte se desarrolle y extienda ampliamente en el futuro, con la aplicación prevista de razonadores.

Por último se ha desarrollado una fase para incorporar datos en RDF al grafo a partir de diferentes fuentes de información geográfica. Este proceso consiste en la creación de procesos automáticos ETL para los datos estructurados (como el denominado GCG que aplica plantillas TTL a capas PostGIS), así como procesos de minería de datos para datos no estructurados como puede ser el caso del Boletín Oficial de Aragón (mediante el proceso denominado GeoBOA).

En ICEARAGON se ha realizado un esfuerzo para relacionar conceptos propios con otros elementos. Por un lado, todos los elementos geográficos (SpatialThing) ofrecen enlaces directos a aplicaciones de ICEARAGON (Visor 2D / 3D, atlas, cartoteca, descargas, etc.). Sin embargo, para dotar de un mayor contexto semántico, se han fomentado los

enlaces a iniciativas externas, para maximizar el valor de la quinta estrella de los Datos Enlazados (Linked Open Data). Concretamente se ha incluido (o está previsto incorporar) enlaces a Wikidata, DBpedia, Biblioteca Virtual de Defensa, Real Academia de la Historia, Biblioteca Nacional de España (BNE), Dirección General de Tráfico, Cross-Forest, entre otros.

Adicionalmente se ha habilitado un servicio API de búsqueda de entidades (denominado EntitiesSearchService) que realiza las tareas de reconocimiento de entidades pasando de un servicio de búsqueda de texto tradicional a uno de conceptos con relaciones. El diseño de la API está inspirado en el Knowledge Graph Search API de Google.

El resultado más visible de cara al usuario hasta la fecha es la disponibilidad de un buscador de entidades geográficas el cual muestra los datos estructurados en función de su tipo en una ficha (infobox o detallada) así como elementos relacionados.

Palabras claves

Grafo de conocimiento Geográfico, ICE, Geoknowledge graph, Aragón, Interoperabilidad, Linked Open Data, Web semántica.

Autores

David Portolés
dportoles@idearium.eu
Idearium Consultores SRL

Rafael Martínez
rmartinezceb@aragon.es
Instituto Geográfico de Aragón. Gobierno de Aragón

Fernando López
flopezm@aragon.es
Instituto Geográfico de Aragón. Gobierno de Aragón