SES Fact Sheet

The Semantic Extraction Service (SES) is a stand-alone service, which ensures life-cycle automation, i.e., development, evaluation, management and maintenance, of semantic extraction application (SEA). It uses the data in the GraphDB database as its dictionary of concepts (knowledge base) to recognise mentions of entities, as well as their relevance and algorithm's confidence, on the text. It comes with an out-of-the-box default pipeline, which is configured to produce media graph basic concept look ups. The service supports multiple pools of SEA and provides an API to manage them (start/stop/reload/configure), which is very convenient in cases when more sophisticated semantic annotation is required and also when different text analysis techniques are used to span across multiple data domains. Semantic annotation is covered by special queries in the SPARQL API which retrieve the matched concepts and their features. Since the SEA are usually based on a mixture of machine learning routines and rules, SES also provides a Re-training API which allows both manual and automated re-training and evaluation of the underlying statistical models (for example, relevance and confidence criteria are updated automatically over time to adjust to new data).

Architecture features

As the Semantic Extraction Service is wired to GraphDB through its Plug-in API, its dictionaries are dynamically updated in a transactional fashion and are always in synchronization with the data in the semantic repository. It also provides high availability ensured by GraphDB's replication, load balancing and fail-over capabilities. Horizontal scaling also works in the same way as in GraphDB. Moreover, the pipelines running on a single cluster node could be pooled in the way so that multiple pipeline instances share the same dictionary and achieve more semantic annotation request throughput while optimizing memory usage. The pools are also capable of dynamic expanding with respect to the current load. Communication with the service is established over HTTP and the SPARQL protocol, which makes it easily integrable with various platforms written in different programming languages.

Interfaces, Standards, Requirements

TO DO INTRO

Licensing

TO DO INTRO

Installation and Usage

TO DO INTRO

Credits

TO DO INTRO