**downloadDataset.py**

Per lanciare questo script bisogna specificare come parametro l’url del portale da cui si vuole scaricare la lista di dataset ai quali è associato un endpoint SPARQL. I possibili portali sono:

* <https://www.europeandataportal.eu/sparql> 🡪 url dell’European Data Portal in cui sono elencati gli open data pubblicati sui portali di tutti i paesi europei
* <http://data.europa.eu/euodp/sparqlep> 🡪 url dell’EU Open Data Portal, che elenca gli open data prodotti dalle istituzioni e dagli organi dell’UE
* <https://io.datascience-paris-saclay.fr/sparql> 🡪 url del portale IO Data Science

Quindi, lo script scaricherà sul database, verificando che non siano ivi già presenti, l’elenco dei dataset trovati su tali portali.

**automaticExtraction.py**

Lo script deve essere lanciato settando una procedura automatica in grado di avviarne l’esecuzione quotidiana. Dunque, l’effetto sortito sarà che giornalmente verrà intrapreso il tentativo di estrarre gli indici dei dataset e di generarne lo Schema Summary.

**createTable.py**

Lanciare tale script produce come risultato una tabella excel (formato .xls) da consultare per monitorare lo stato di accessibilità dei set di dati con cui lavora H-Bold. Per ogni giornata in cui è stata invocata l’esecuzione, si tiene traccia:

* del numero di url di endpoint SPARQL
* del numero di dataset per cui è stata correttamente completato il processo di estrazione degli indici in quel mese
* del numero di dataset per cui è stata correttamente completato il processo di estrazione degli indici
* del numero di dataset per cui si è riusciti a generare con successo lo SS in quel mese
* del numero di dataset per cui si è riusciti a generare con successo lo SS