

## **DBpedia e Linked Open Data: il Semantic Web come chiave per l'utilizzo degli Alternative Data**

DBpedia è il tentativo di estrarre un'informazione strutturata da Wikipedia e renderla disponibile in maniera open e free nella forma di Linked Open Data accessibili a tutti tramite il Web. Nasce da una collaborazione tra la Free University of Berlin, la Leipzig University e OpenLink Software e vede la pubblicazione del primo dataset open nel 2007. Negli anni il progetto ha visto una notevole crescita ed è, ad oggi, gestito da una vera e propria associazione, la DBpedia Association, che conta sulla collaborazione di diverse società private che, in quanto membri ufficiali, forniscono il loro know-how.

L'informazione viene estratta automaticamente e periodicamente da Wikipedia, principalmente dagli "infobox" e dalle "categorie" presenti in ogni articolo, andando a costituire l'insieme di milioni di *entità* e *fatti* che costituiscono il cuore di DBpedia. Nello specifico, le *entità* sono connesse in *fatti* tramite triple RDF costituite da Soggetto(entità), Predicato (relazione), Oggetto (entità). Grazie a questa formalizzazione, è possibile accedere in maniera semantica ai contenuti di Wikipedia. Inoltre, DBpedia è connessa tramite link RDF ad altri database, rendendo possibile il suo utilizzo come hub di collegamento tra le diverse fonti di dati.

DBpedia rappresenta quindi una ricca fonte di dati cross-dominio, strutturati e dotati di collegamenti semantici il cui utilizzo è di fondamentale importanza per molte applicazioni di Intelligenza Artificiale nel campo del Natural Language Processing. Finscience, la fintech company del gruppo Datrix, è membro DBpedia e impiega ampiamente i diversi servizi per estrarre insights di interesse finanziario dagli *Alternative Data* raccolti dal web nella forma di dato testuale. Nello specifico, tramite l'utilizzo di DBpedia, è possibile identificare in maniera univoca di quali entità si parli in un determinato testo, per esempio una notizia estratta da un blog di economia, a prescindere dalla lingua in cui il testo è stato scritto tramite Named Entity Recognition. Inoltre, il nuovo modello di release mensile di DBpedia permette a Finscience il tracciamento di entità sempre aggiornate, di fondamentale importanza in un panorama altamente dinamico come quello finanziario.

## **Relatori**

**Matteo Bregonzio** Head of Research and Development a 3rdPlace S.r.l.

Nel 2011 riceve un dottorato in Computer Science con specializzazione in machine learning e data mining presso l'università Queen Mary University of London. Nello stesso anno inizia a lavorare per IHS Markit come analista specializzandosi sui prodotti derivati. Nel 2015 dopo un breve passaggio in Barclays Capital come senior analyst si sposta a BNP Paribas dove segue il trading quantitativo dei prodotti derivati. Dal 2017 lavora presso 3rdPlace come Head of Research and Development, supervisiona il team di data Scientist e Data Analyst. Matteo ha insegnato Machine Learning e data Analysis presso Queen Mary University of London. Ha organizzato la Summer School 2019 con l'università degli Studi della Tuscia - Data Science & Machine Learning. Interviene regolarmente agli workshop organizzati Talent Garden.

**Leone De Marco** Data Scientist a 3rdPlace S.r.l.

Nel 2014 si laurea in Bioinformatics presso l'Alma Mater Studiorum di Bologna. Successivamente, inizia una breve carriera accademica che conclude nel 2018 conseguendo un dottorato in Life Sciences presso l'Università di Camerino. Durante i tre anni di dottorato, svolti principalmente presso l'Università di Pavia, pubblica nove paper scientifici in riviste peer-reviewed oltre a contribuire alla didattica con lezioni frontali e supervisione di studenti durante il tirocinio di tesi. A inizio 2019 si iscrive al Master in Data Science presso il Data ScienceTech Institute a Biot, Francia. Da settembre 2019 lavora presso 3rdPlace come Data Science Intern, prima, e Data Scientist, poi, all'interno del team di Research and Development. Coordina e definisce le metodologie di analisi dati e data exploration. Analizza i requisiti di progetto e svolge le attività di data exploration e discovery. Inoltre, supervisiona la parte di visualizzazione e presentazione risultati.