# Linking & Traversal

##### Session 1.3 (SEMANTiCS)

#### Time: Thursday, September 21, 2023 - 10:45 to 13:00

#### Chair: Kossi Amouzouvi, Postdoctoral fellow, ScaDS.AI Excellence Center

## **Talks**

### Cobalt: A Content-Based Similarity Approach for Link Discovery over Geospatial Knowledge Graphs

Purpose: Data integration and applications across knowledge graphs (KGs) rely heavily on the discovery of links between resources within these KGs. Geospatial link discovery algorithms have to deal with millions of point sets containing billions of points.

Methodology: To speed up the discovery of geospatial links, we propose Cobalt. Cobalt combines the content measures with R-Tree indexing. The content measures work based on the area, diagonal and distance of polygons which speeds up the process but is not perfectly accurate. We thus propose a polygon splitting approach for improving the accuracy of Cobalt.

Findings: Our experiments on real-world datasets show that Cobalt is able to speed up the topological relation discovery over geospatial KGs by up to 1.47\*10^4 times over other linking algorithms while maintaining an F-Measure between 70% and 90% depending on the relation. Furthermore, we were able to achieve an F-Measure of up to 99% by splitting the polygons before applying the content measures.

Value: The process of discovering links between geospatial resources can be significantly faster by sacrificing the optimality of the results. This is especially important for real-time data-driven applications such as emergency response, location-based services and real-time traffic management. In the future additional measures, like the location of polygons or the name of the entity a polygon represents, could be integrated to further improve the accuracy of the results.

| TBA |  |
| --- | --- |

### Open-world generative conversational assistants on knowledge graphs in tourism and museum projects [SP]

GNOSS is developing generative conversational virtual assistants that provide the best experience for each user and meet their needs and motivations before, during and after a visit. They offer an Augmented Visit, giving it a context through conversational capabilities based on Semantic AI, Deep Learning and the exploitation of a knowledge graph. This is possible because the technology can develop the necessary Contextual Ambient Intelligence services to provide the user with a suitable interpretation environment at any given moment, capable of whispering in our ear what we want to know at any given moment. A universal and personal Virtual Assistant, that reasons and learns.

| Ricardo A. MaturanaGNOSS founder and CEO[Affiliation page](https://www.gnoss.com/en) | Ricardo Alonso Maturana (Bilbao, 1962), graduated in Pure Philosophy at the University of Deusto and PhD in Sociology at the UNED, is the Founder and CEO of GNOSS, a technology company dedicated to the construction and exploitation of knowledge graphs, a way of integrating and representing information that allows to discover and investigate any topic in a deeper and more intuitive way and enjoy a more semantically aware web. Ricardo has been responsible for the direction, design and conceptualization of the GNOSS project and its services.  He is also the promoter of Didactalia, a social ecosystem for learning and knowledge on the web, which has more than 100,000 useful resources for teaching and learning open to the entire educational community. |
| --- | --- |

### 

### Classification of Linking Problem Types for linking semantic data

As the number of RDF datasets published on the semantic web continues to grow, it becomes increasingly important to efficiently link similar entities between these datasets. However, the performance of existing data linking tools, often developed for general purposes, seems to have reached a plateau, suggesting the need for more modular and efficient solutions. In this paper, we propose --and formalize in OWL-- a classification of the different Linking Problem Types (LPTs) to help the linked data community identify upstream the problems and develop more efficient solutions. Our classification is based on the description of heterogeneity reported in the literature --especially five articles-- and identifies five main types of linking problems: predicate value problems, predicate problems, class problems, subgraph problems, and graph problems. By classifying LPTs, we provide a framework for understanding and addressing the challenges associated with semantic data linking. It can be used to develop new solutions based on existing modularized tools addressing specific LPTs, thus improving the overall efficiency of data linking.

| TBA |  |
| --- | --- |

### 

### BiPaSs: Further Investigation of Fast Pathfinding in Wikidata

Purpose:

A previous paper proposed a bidirectional A\* search algorithm for quickly finding meaningful paths in Wikidata that leverages semantic distances between entities as part of the search heuristics. However, the work lacks an optimization of the algorithm’s hyperparameters and an evaluation on a large dataset among others. The purpose of the present paper is to address these open points.

Methodology:

Approaches aimed at enhancing the accuracy of the semantic distances are discussed. Furthermore, different options for constructing a dataset of dual-entity queries for pathfinding in Wikidata are explored. 20% of the compiled dataset are utilized to fine-tune the algorithm’s hyperparameters using the Simple optimizer. The optimized configuration is subsequently evaluated against alternative configurations, including a baseline, using the remaining 80% of the dataset.

Findings:

The additional consideration of entity descriptions increases the accuracy of the semantic distances. A dual-entity query dataset with 1,196 entity pairs is derived from the TREC 2007 Million Query Track dataset. The optimization yields the values 0.699/0.109/0.823 for the hyperparameters. This configuration achieves a higher coverage of the test set (79.2%) with few entity visits (24.7 on average) and moderate path lengths (4.4 on average). For reproducibility, the implementation called BiPaSs, the query dataset, and the benchmark results are provided.

Value:

Web search engines reliably generate knowledge panels with summarizing information only in response to queries mentioning a single entity. This paper shows that quickly finding paths between unseen entities in Wikidata is feasible. Based on these paths, knowledge panels for dual-entity queries can be generated that provide an explanation of the mentioned entities’ relationship, potentially satisfying the users’ information need.

| Leon MartinDoctoral Student at the Media Informatics Group at the University of Bamberg |
| --- |

### 

### Oxford Semantic Technologies Limited [OST]

| Peter Crocker, CEO & Co-founder Oxford Semantic Technologies Limited | Peter Crocker is the co-founder and CEO of Oxford Semantic Technologies (OST), developers of the industry-leading knowledge graph and reasoner—RDFox. In his tenure, Peter has grown OST into the established business it is today, implementing solutions across the globe. With a practical approach to clients’ needs, Peter has overseen the application of RDFox to tackle financial fraud and regulatory compliance, modernize data integration and clinical decision support, and transform configuration management and personalised recommendations.  Prior to OST, Peter served as Program Director of Product Management for the IBM Watson IoT Platform Product. He has led product development, architecture and consulted in enterprise integration. A graduate from the University of  Oxford, Peter attained a Master of Computer Science, Distinction, and a Master of Engineering Science, First class. |
| --- | --- |

### 

### 