Unlocking the Butterfly Effect: Finding Hidden Connections in Barcant Butterfly Collection with Ontology Matching and Knowledge Graphs

Rakshitha VijayKumar, Shenghui Wang, Andreas Weber, Faiza A. Bukhsh

University of Twente, The Netherlands

Extended Abstract:

Biodiversity data is essential for conservation and ecosystem management, yet its rapid accumulation from diverse sources presents challenges for integration and analysis. This study addresses these challenges by applying ontology matching and knowledge graph techniques to the Barcant Butterfly Collection (BBC), one of the largest butterfly archives in the Caribbean. The research focuses on transforming and standardizing the BBC dataset using the Darwin Core (DwC) [1] framework to facilitate interoperability with global biodiversity information systems. Through the construction of a domain-specific ontology for the BBC, aligned with existing biodiversity ontologies such as the Biological Collections Ontology (BCO)[2] and Taxonomic Rank Ontology (TaxRank)[3], this work creates a structured, interoperable knowledge base. The ontology alignment process enhances the consistency and integration of taxonomic, ecological, and geographic data. A knowledge graph is generated to reveal hidden connections and patterns within the collection, enabling advanced querying and analysis.

The results demonstrate the potential of these methodologies to unlock valuable insights from legacy biological collections, enhancing their utility for conservation efforts and biodiversity research. By aligning traditional taxonomic data with modern computational techniques, this research contributes to the broader field of biodiversity informatics, setting a precedent for the digitization and semantic enrichment of biological collections worldwide.

This study demonstrates the potential of ontology matching and knowledge graph techniques in enhancing the analysis and utilisation of biodiversity data from historical collections as can be seen in Figure 1 (details can be seen in section 5.3 of [4]). By focusing on the Barcant Butterfly Collection, a rich yet underutilized dataset, the research successfully transformed it into a structured, machine-readable format aligned with widely recognised standards like the Darwin Core (DwC)[5]. The integration of the collection into the broader biological collections ontology (BCO) and TaxRank ontology allowed for enhanced semantic interoperability, facilitating a deeper analysis of species distribution, taxonomy, and ecological relationships.

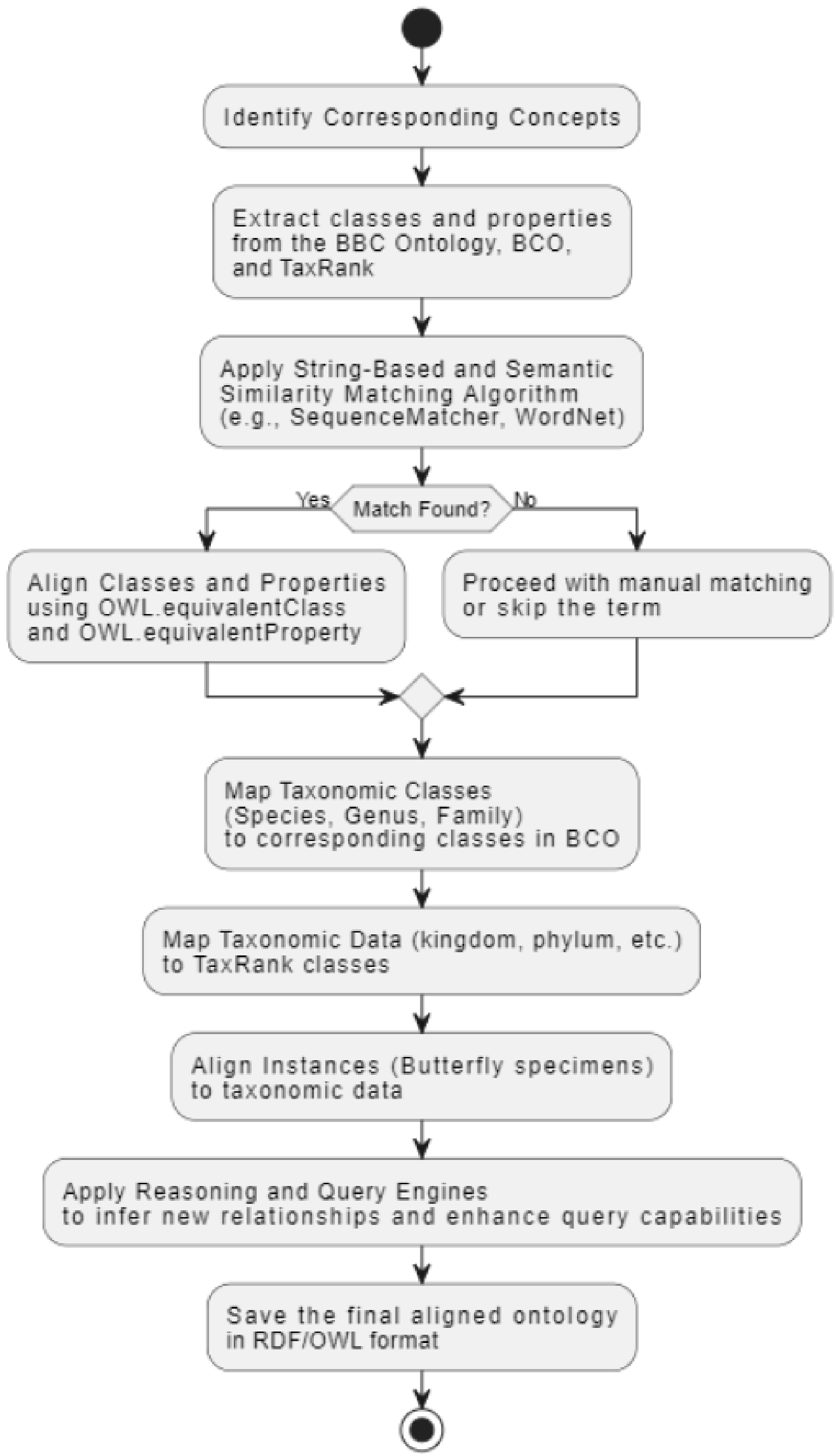

Figure 1: Flowchart of the ontology alignment process

The ontology alignment process and the subsequent construction of knowledge graphs offered significant insights by uncovering connections that were previously obscured due to the disparate nature of the data, as can be seen in the following Figure 2. (details can be seen in section 5.4 of [4])

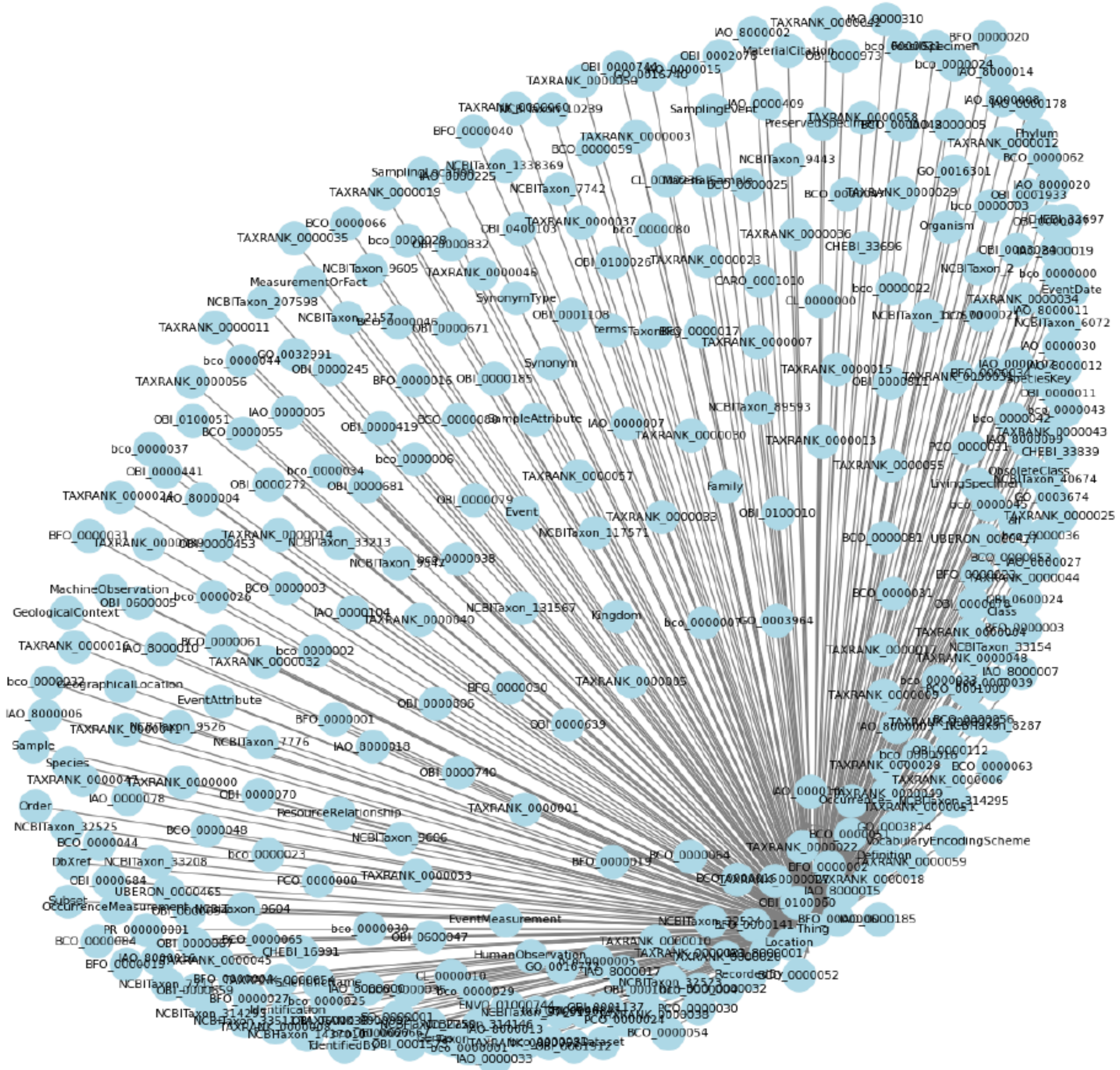

Figure 2: Resulting Knowledge Graph

The ecological range maps generated as part of this study are particularly noteworthy, offering new visual perspectives on species distributions and highlighting areas of biodiversity significance that align with protected areas data from the World Database on

Protected Areas (WDPA). Furthermore, the research underscores the importance of aligning historical collections with modern biodiversity datasets to unlock new opportunities for conservation and scientific inquiry. By bridging the gap between historical taxonomic records and modern ecological frameworks, this study contributes to a growing body of work that seeks to integrate traditional biodiversity knowledge with contemporary data science methods.

Future Research Outlook

Several promising avenues for future research have been identified as a result of this study. First, while the ontology alignment in this work focused on taxonomic and geographic data, future efforts could extend this to include more complex ecological interactions, such as species competition, symbiosis, and predation. Incorporating these ecological dimensions into the ontology would enable a more comprehensive understanding of biodiversity dynamics. Second, there is significant potential for enhancing the scalability of ontology alignment and knowledge graph generation by incorporating machine learning and artificial intelligence (AI) techniques. Using automated tools such as machine learning classifiers and natural language processing (NLP) could streamline the ontology alignment process, improving both speed and accuracy when dealing with large, heterogeneous datasets. In particular, AI could assist in identifying complex semantic relationships between entities that are not immediately apparent through traditional matching techniques.

Additionally, this research could be extended by incorporating real-time data streams from citizen science platforms, environmental monitoring networks, and remote sensing technologies. By integrating dynamic data sources, the knowledge graphs could be continuously updated, allowing for real-time analysis of biodiversity changes and enabling more timely conservation interventions. This would also facilitate predictive modelling, allowing the researchers to forecast potential changes in species distributions and ecosystem health in response to climate change and habitat loss.

Another potential extension is the application of this methodology to other historical and contemporary biological collections globally. The successful alignment and enrichment of the Barcant Butterfly Collection demonstrate that similar techniques could be applied to other under utilized biological datasets, contributing to creating a comprehensive, globally integrated biodiversity knowledge system. This would enhance scientific research and support conservation initiatives by making biodiversity data more accessible, interoperable, and actionable.

Lastly, future work should focus on enhancing the visualization tools used for biodiversity analysis. While the current study utilized ecological range maps, more sophisticated visualization techniques such as interactive knowledge graphs, geospatial heatmaps, and 3D modeling could provide deeper insights and improve stakeholder engagement. Such tools could facilitate communication between researchers, conservationists, and policymakers, thereby enhancing the impact of biodiversity data on decision-making processes.

In conclusion, this research lays the groundwork for the continued development of ontology-driven biodiversity informatics, with significant implications for global biodiversity research and conservation. By refining and expanding the methods and technologies presented here, future studies can further bridge the gap between historical, biological records and modern conservation strategies, ensuring the preservation of global biodiversity for generations to come.

References:

[1]John Deck, Robert Guralnick, Ramona Walls, Stanley Blum, Melissa Haendel, Andrea Matsunaga, and John Wieczorek. Meeting report: identifying practical applications of ontologies for biodiversity informatics. Standards in Genomic Sciences, 10:1–6, 2015.

[2] Ramona L Walls, John Deck, Robert Guralnick, Steve Baskauf, Reed Beaman, Stanley Blum, Shawn Bowers, Pier Luigi Buttigieg, Neil Davies, Dag Endresenet al. Semantics in support of biodiversity knowledge discovery: An introduction to the biological collections ontology and related ontologies. PLOS ONE 9(3):e89606, 2014.

[3] Joshua S. Madin, Shawn Bowers, Mark Schildhauer, and Matthew B. Jones. Advancing ecological research with ontologies, 03 2008.

[4] VijayKumar, Rakshitha (2024) Unlocking the Butterfly Effect: Finding Hidden Connections in Barcant Butterfly Collection with Ontology Matching and Knowledge Graphs. https://essay.utwente.nl/104562/ 2024