The 7th Open Society Conference (OSC) 2025
Faculty of Law, Social and Political Science
Universitas Terbuka
DOI: 10.33830/osc.v3i1.6969 / ISSN: 3032-2227

A Conceptual Framework for Inclusive Semantic Retrieval in Digital Heritage Archives

KianLam Tan[1], HengWei Lee[1], ChenKim Lim[2]

[1] School of Digital Technology (DiGiT), Wawasan Open University (WOU), Penang, Malaysia
[2] Institution for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia
e-mail: andrewtan@wou.edu.my; hwlee@wou.edu.my; kim@ukm.edu.my

Abstract

Open cultural heritage platforms like Europeana are widely used to store and share the rich history of Europe. Access to such large collections of linked data is still not widely distributed due to technical difficulties and the poor usability of semantic search interfaces. This paper proposes a conceptual framework for improving information retrieval on cultural heritage knowledge graphs. The model is built on the EDM and RDF Schema and introduces weighted semantic relationships between resources. It encodes weights as contextual metadata on edges and uses category-based weighting, external linked data sources (DBpedia, Freebase, etc.), and optimized graph traversal to improve the relevance and inclusivity of search results. By assigning relative relevance to each graph edge based on the similarity of shared metadata categories, the proposed method increases the accuracy of query results, while improving their coverage to include semantically similar entities, allowing non-domain expert users, educators, and underrepresented communities to browse and discover new information. This work is in line with efforts to create an inclusive digital public sphere in communication studies by building on the idea of enhancing active semantic and semiotic forms of participation rather than limiting users to passive and non-inclusive forms of access. The paper is theoretical in nature, providing a proof of concept and evaluating model characteristics; however, it also briefly describes the future evaluation strategy and use cases beyond the cultural heritage context.

Keywords: *digital cultural heritage, semantic search, weighted relationship, semantic search, linked open data*

INTRODUCTION

Digital libraries and archives are an increasingly important and active way to preserve and share cultural heritage [1, 2]. The Europeana platform is among the largest in Europe and acts as an aggregator for online museums, libraries, archives, and galleries, providing a unified interface and linked data infrastructure based on the Europeana Data Model (EDM) [3]. Although the semantic web technologies allow to semantically interlink the graph's entities using RDF [4], data accessibility and discovery is still a largely unsolved and often unaddressed problem, especially for non-specialist users.

DOI: 10.33830/osc.v3i1.6969 / ISSN: 3032-2227

This paper aims to explore how adding semantic relationships [3] to the entities in a cultural heritage graph can make retrieval more intuitive, accurate, and inclusive.

The emergence of digital transformation has enabled a leap in progress for cultural communities and their interest in their pasts. The rise of the open access cultural data has allowed information sharing on an unprecedented scale [5]. However, current search technology is unable to fully allow for inclusive participation in this form of digital communicaton due to the structural nature of search and discovery algorithms which are designed for consistency and efficiency of edge traversal over user-centered interface design. This creates a semantic divide that prevents users from fully participating in cultural exploration.

Search and discovery on RDF graphs is already an ability built into most semantic data browsers through the concept of linked data [4]. While users can follow graph edges in either direction, RDF-based queries currently treat all relationships as equal in weight and do not build upon the more or less obvious importance of some graph connections over others. This results in flattened and "inelastic" navigation of graphs with little discrimination on search results for relevance. The fact that users are not always able to find a string match for the names or metadata of the information they are looking for or possess the necessary domain knowledge to know the name of specific categories, metadata elements, or collections results in cultural information systems that prevent full user participation in access.

Although entity relationships are well-defined in cultural knowledge graphs based on RDF triples, traditional search engines do not usually weight the importance of individual RDF predicates (relationships). This research seeks to address this issue by treating edge relationships not as binary flags but as value-bearing connections.

BACKGROUND AND RELATED WORK

The Resource Description Framework (RDF) allows us to describe resources and their relationship with other resources using subject-predicate-object triples [4]. EDM based on RDF allows for the description of digital resources by creating a new resource. EDM ontologies [6] currently represent many types of cultural entities and their metadata (author, title, location, time, etc.), but these ontologies do not have weights attached to the relationships between individual types in order of semantic importance or category value for general retrieval.

A body of related work already exists on general RDF-based approaches [11] to graph traversal using semantic search, which includes both the use of weighted categories or improved graph traversal in a data model.

[7] detail the possibility of creating a semantic space dependency between nodes in a network using a frequency-similarity distribution through relative grammatical distance. This work applied the concept of semantic weighting in natural language rather than information systems or metadata, but shows the possibility of ascribing differential weight to connections in a network.

[8] discuss the Linked Open Data of cultural institutions in particular and connectedness as a proxy for relevance by providing weights for semantic connections through usage and co-occurrence statistics. [9] use the Smithsonian American Art Museum as an example of entity resolution and mapping through DBpedia in order to connect their datasets to the Linked Open Data (LOD) cloud.

The MARiMbA (Metadata to Resources for Museums and Galleries in RDF- Any) framework and tool [10] as a possibility for cultural institutions to automatically convert their metadata into

DOI: 10.33830/osc.v3i1.6969 / ISSN: 3032-2227

RDF by using a shared vocabulary for cultural heritage is already an extant piece of technology. The use of weighted edges in traversal or auto-optimized or "elastic" graph algorithms are currently not implemented, making this paper and proposed methods relevant for RDF based knowledge graphs.

While this work details relevant RDF triples and graphs on the semantic web, it does not add any computational detail or specific model to the existing approach of weighting metadata and edges for RDF triples or expanded graph search.

CONCEPTUAL FRAMEWORK

The proposed model is a category-based weighted graph approach for information retrieval in RDF-based system. In this model:

- **Vertices (nodes)** represent entities of cultural significance (paintings, sculptures, artists, historical periods, geographic locations, etc).
- **Edges (relationships)** represent RDF predicates linking the vertices (e.g., createdBy, influencedBy, belongsToMovement).
- **Weights (edge-values)** quantify the semantic relevance of the relationships between vertices using shared metadata categories.

The fundamental equation used to calculate a relationship weight (r) is:

$$r = \sum_{i=0}^{n} c_i \times w_i$$

Where:

- c_i represents the presence (1 or 0) of a shared metadata category between nodes.
- w_i is the assigned importance or relevance score of that category.

Once the relationship weight r is computed, it is compared to a threshold T. If $r \geq T$, the edge is added to the graph. Otherwise, the relationship is excluded. This threshold acts as a semantic filter to retain only meaningful and contextually significant links between entities.

The benefit of this model is that it is able to both prioritize more relevant edges to follow through graph traversal and to improve the quality and relevance of retrieved connections when edges are followed with a search query. For example, a search for "Baroque" will return not only direct matches (paintings tagged with "Baroque"), but also semantically similar concepts in higher weight edges, such as other paintings or artistic movements from the same time period, the artist's painting style, or other events related to the concept.

METHODOLOGICAL STRATEGY

While this is a conceptual paper, a high-level overview of an evaluation strategy is included for the purpose of practical application in future research. This will be completed using the following components.

DOI: 10.33830/osc.v3i1.6969 / ISSN: 3032-2227

1. **Data Acquisition and Integration:**
   Europeana's RDF data will be used as a primary dataset and will be integrated with external linked data sources such as DBpedia [12], Freebase [13], and other public cultural heritage graphs.

2. **Edge Construction and Weight Assignment:**
   Metadata will be parsed from RDF triples to discover shared metadata categories. Each category is weighted using a linear distribution [14] based on frequency, user behavior data, or by heuristic or expert-based rules. Edges are constructed between entities only if their relevancy score exceeds a threshold value.
3. **Graph Optimization:**
   Graph pruning methods [15] are used to remove excess nodes and eliminate unnecessary cyclic paths. The implementation may also use various graph optimization techniques such as shortest-path caching or batch traversal indexes to ensure system scalability.
4. **Evaluation Plan:**
   Manual validation: Domain experts are used to compare the enriched query results against a gold standard or benchmark.
   Quantitative metrics: Precision, recall, and mean average precision (MAP) are calculated based on predefined benchmark queries.
   User studies: User testing and public studies from both non-expert general users and researchers provide qualitative feedback on usability using scenario-based search tasks.

The methodology was selected based on the task of implementing the conceptual model presented in this paper in the real world and existing in an information system. It was chosen for the simplicity of using a publicly available cultural heritage dataset as well as extant graph pruning techniques and a rigorous combination of automated and human-in-the-loop evaluation methods.

IMPLICATIONS FOR INCLUSIVE DIGITAL ACCESS

A core feature of the proposed model is its focus on inclusivity. By encoding metadata relevance directly into the form of the knowledge graph, users are not required to perform complex, obscure, or technical queries. This feature is especially impactful for the following groups of users:

- Non-specialist and casual users such as students or novices that do not have domain-specific language or vocabulary.
- Diverse and underrepresented communities by allowing the model to also return content related to broader or adjacent topics instead of searching only for a string match.
- Multilingual or non-English language speakers, potentially by also being applied to multilingual linked data datasets.

This feature will support the goal of creating a more democratic and equitable public sphere of cultural communication by bridging the technical data structures and abstract concepts inherent in the data with an interface that aligns more closely with the intuitive models users have when searching and browsing in the real world. Adopting such models in the future can increase

DOI: 10.33830/osc.v3i1.6969 / ISSN: 3032-2227

accessibility to museums, especially for marginalized groups, improve educational outcomes and traffic for lesser-known works, and promote trans-cultural information exchange.

As digital libraries and cultural heritage collections continue to grow and expand with new digital objects, more scalable and context-aware solutions such as those proposed by this paper will be increasingly necessary to support open and equitable cultural access, especially for marginalized voices.

CONCLUSION AND FUTURE WORK

This paper proposes a conceptual model of weighted relationships for cultural heritage knowledge graphs in order to allow for more accurate and inclusive retrieval of linked metadata.It formalizes the concept of weighted graph traversal in the form of encoded contextual relevance within the network itself.

Future work will be required to build this model in a real environment, for example in the Europeana platform or other dataset, which will consist of several tasks. This will include the development of automated methods of weight assignment for graph edges, refining and calibrating the threshold value, and the extension of the model for culturally adaptive and multilingual metadata categories.

Long-term user studies will also be required to measure the impact of this and other similar models on user behavior, discovery mechanisms, and the perceived inclusivity of cultural information platforms. This work combines technical and user-centered design considerations in order to propose a model that has the potential to create a more inclusive digital space for cultural participation.

ETHICAL AND TECHNICAL CONSIDERATIONS

While the proposed model opens up exciting avenues for enhancing semantic retrieval in our domain, it is imperative to consider a range of ethical and technical issues that could surface during real-world deployment. Addressing these concerns from the onset will not only ensure the responsible implementation of the model but also contribute to its effectiveness and acceptance in diverse contexts.

Ethically, the assignment of weights to different metadata categories is a practice that could inadvertently perpetuate existing biases present in cultural heritage data. If certain cultures, periods, or well-documented collections are favored in the weighting scheme, there is a risk that lesser-known or underrepresented entities might be systematically excluded from search results. This calls for a carefully curated approach to ensure that diverse metadata are given appropriate weight and that the system does not replicate historical imbalances. The involvement of cultural experts in designing and reviewing the weight assignment process could mitigate such risks. Additionally, the automatic linking of entities based on shared metadata could lead to unexpected or culturally insensitive connections. It is crucial to have mechanisms in place to filter, review, or moderate such links to maintain contextually relevant and culturally respectful results.

On the technical side, the incorporation of external linked data sources (DBpedia, Freebase) and the dynamic nature of RDF datasets introduce challenges related to data consistency, version control, and system reliability. The model needs to accommodate frequent updates to RDF datasets, which may necessitate recalculating edge weights and updating the graph structure, potentially impacting system performance. Users may also require explanations or justifications

DOI: 10.33830/osc.v3i1.6969 / ISSN: 3032-2227

for the relationships suggested by the system, particularly when indirect links between entities lead to unexpected or surprising results. The system should provide a form of explainable traversal path or metadata visualization to support this requirement.

Future implementations will have to consider trade-offs between inclusivity and computational efficiency, and they need to adopt scalable, context-aware, and culturally sensitive approaches. By proactively considering these ethical and technical dimensions, the system can be designed to not only be effective but also responsible and inclusive, setting a solid foundation for broad application in digital cultural heritage access.

REFERENCES

[1] J. M. Perrin and R. G. Weaver, “Context is Key: Library and Archive Collaboration for Digital Projects,” Collaborative Librarianship, vol. 12, no. 1, p. 7, Jan. 2020

[2] X. Du, “Digital library archives information integration based on multidimensional data mining,” International Journal of Reasoning-based Intelligent Systems, vol. 14, no. 4, p. 169, Jan. 2022, doi: 10.1504/ijris.2022.126659.

[3] A. L. Silva and A. L. Terra, “Cultural heritage on the Semantic Web: The Europeana Data Model,” IFLA Journal, vol. 50, no. 1, pp. 93–107, Dec. 2023, doi: 10.1177/03400352231202506.

[4] M. R. Saeed and V. K. Prasanna, "Extracting Entity-Specific Substructures for RDF Graph Embedding," *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, Salt Lake City, UT, USA, 2018, pp. 378-385, doi: 10.1109/IRI.2018.00063.

[5] A. Shiri and E. Villanueva, “Methodological Diversity in the Evaluation of Cultural Heritage Digital Libraries and Archives: Approaches, Frameworks, and Methods,” Nov. 2020, doi: 10.29173/CAIS1177.

[6] H. Daoudi, I. Chaker and A. Zahi, "Comparative Study of Cultural Heritage Ontologies," *2023 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, Rabat, Morocco, 2023, pp. 1-7, doi: 10.1109/ICTMOD59086.2023.10438114.

[7] S. M. Carta, S. Consoli, A. Giuliani, A. S. Podda and D. R. Recupero, "CulturAI: Semantic Enrichment of Cultural Data Leveraging Artificial Intelligence," in *IEEE Access*, vol. 10, pp. 127328-127344, 2022, doi: 10.1109/ACCESS.2022.3226070.

[8] S. Oguri, M. Urata, M. Endo and T. Yasuda, "Data Model for Registered Tangible Cultural Properties Using Linked Data," *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, Kobe, Japan, 2020, pp. 651-654, doi: 10.1109/GCCE50665.2020.9291783.

[9] P. Szekely et al., “Connecting the Smithsonian American Art Museum to the Linked Data Cloud,” Springer, Berlin, Heidelberg, 2013, pp. 593–607. doi: 10.1007/978-3-642-38288-8_40.

[10] D. Vila-Suero and A. Gómez-Pérez, “datos.bne.es and MARiMbA: an insight into library linked data,” Library Hi Tech, vol. 31, no. 4, pp. 575–601, Nov. 2013, doi: 10.1108/LHT-03-2013-0031.

[11] D. Lamprecht, “RDF Knowledge Graph SemOpenAlex-SemanticWeb,” Dec. 2023, doi: 10.5281/zenodo.10299131.

[12] T. Puspa Rinjeni, S. Suci Indasari, A. Indriawan and N. Aini Rakhmawati, "Movies Analysis on DBpedia and Wikidata Using Community Detection and Centrality Algorithms," *2022 International Electronics Symposium (IES)*, Surabaya, Indonesia, 2022, pp. 380-386, doi: 10.1109/IES55876.2022.9888338.

[13] D. Chen and H. Zhao, "Research on the Construction Method of Domain Semantic Knowledge Base Based on Freebase," *2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)*, Changchun, China, 2018, pp. 222-226, doi: 10.1109/ICISCAE.2018.8666910.

[14] F. Merovci and I. Elbatal, “A New Generalization of Linear Exponential Distribution: Theory and Application,” Jan. 2015, [Online]. Available: http://www.naturalspublishing.com/files/published/64pc56a36o1g6y.pdf

[15] Z. Chen *et al.*, "RGP: Neural Network Pruning Through Regular Graph With Edges Swapping," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 10, pp. 14671-14683, Oct. 2024, doi: 10.1109/TNNLS.2023.3280899.