A Three-stage Neuro-symbolic Recommendation Pipeline for Cultural Heritage Knowledge Graphs

Krzysztof Kutt[0000−0001−5453−9763], Elżbieta Sroka[0000−0001−8952−2187], Oleksandra Ishchuk, and Luiz do Valle Miranda[0000−0003−1838−5693]

Department of Human-Centered Artificial Intelligence, Institute of Applied Computer Science, Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, prof. Stanisława Łojasiewicza 11, 30-348 Kraków, Poland
{krzysztof.kutt,elzbieta.sroka,luiz.miranda}@uj.edu.pl

Abstract. The growing volume of digital cultural heritage resources highlights the need for advanced recommendation methods capable of interpreting semantic relationships between heterogeneous data entities. This paper presents a complete methodology for implementing a hybrid recommendation pipeline integrating knowledge graph embeddings, approximate nearest-neighbour search, and SPARQL-driven semantic filtering. The work is evaluated on the JUHMP (Jagiellonian University Heritage Metadata Portal) knowledge graph developed within the CHExRISH project, which at the time of experimentation contained ≈3.2M RDF triples describing people, events, objects, and historical relations affiliated with the Jagiellonian University (Kraków, PL). Despite sparse and heterogeneous metadata, the approach produces useful and explainable recommendations, which were also proven with expert evaluation.

Keywords: Cultural heritage · Recommendations · Hybrid recommender · Knowledge graphs · Knowledge graph embedding · RDF · HNSW

1 Introduction and Motivation

Cultural heritage (CH), including museum, library, and archival collections as well as the achievements of cultural, scientific, and political figures, is an invaluable source of knowledge about a community's art, history, and identity. Mass digitization has rapidly expanded the volume of digital CH resources, shifting research from scarcity to an era of "abundance," where inconsistent cataloging and metadata reduce search quality and overload researchers [22]. In practice, scholars must navigate multiple, disconnected search interfaces and iteratively refine queries, often obtaining large sets of irrelevant or low-quality results [12] that cannot be filtered due to missing or coarse metadata [9].

These challenges stem from the distinctive nature of CH metadata: (1) *VISU complexity*—vagueness, incompleteness, subjectivity, and uncertainty [13]—makes digital representation of historical narratives inherently difficult [11], (2) *Fragmented and low-quality metadata*, often limited to basic catalog descriptions [4], (3) *Heterogeneous metadata schemas* across institutions, particularly visible in large aggregators like Europeana [7], and (4) *Lack of user history*.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_37

Addressing these issues requires multiple perspectives. First, semantic data modeling—using technologies such as RDF and SPARQL [16]—enables representing CH information as knowledge graphs (KGs) that capture complex relationships among objects, creators, dates, functions, and places [4,18]. This approach is increasingly adopted by cultural institutions [16,18]. Second, automated and semi-automated metadata enrichment leverages machine learning to extract semantic features and map them to structured ontologies [6]. Expert validation complements these methods, correcting errors and adding contextual interpretation that automated tools cannot yet provide [12]. However, these efforts primarily expand and refine metadata rather than support researchers in navigating vast CH collections [10]. Although KGs have enabled semantic search [3] and query/document expansion [1], they still extend traditional keyword-based search without fundamentally improving user interaction.

We argue that recommendation systems (RS) offer a promising direction. Although widely used in domains like entertainment and e-commerce [14], RSs remain uncommon in CH and typically provide only simple thematic suggestions [2,10,15]. This work evaluates the feasibility of applying modern RS techniques to CH data. We introduce a three-stage recommendation pipeline (Sect. 2) and test it on a heterogeneous KG representing the cultural heritage of Jagiellonian University, developed within the CHExRISH project (Sect. 3). The method is generic and applicable to diverse CH datasets.

2 Methods

2.1 Three-stage recommendation pipeline

Guided by the challenges outlined in Sect. 1, we designed a recommendation pipeline grounded in four assumptions:

1. RDF KGs as data backbone, ensuring interoperability and reusability [18].
2. Graph embedding-based recommendations are required [23], as traditional content-based or collaborative filtering is unsuitable in CH due to VISU properties, metadata heterogeneity, and lack of user history.
3. An approximate nearest-neighbor (ANN) search is necessary for scalable candidate generation; HNSW or FAISS can be chosen depending on scale-latency-memory constraints [17].
4. Symbolic filtering must complement pipeline to enforce logical consistency, remove low-quality candidates, and incorporate domain constraints, following evidence that rule-based refinement improves accuracy and novelty [19].

Following these assumptions, we implemented a three-stage neuro-symbolic pipeline tailored to CH data summarized in Fig. 1 [8].

2.2 CHExRISH use case

The pipeline was developed within the CHExRISH project (https://chexrish.id.uj.edu.pl/), which creates the JUHMP (Jagiellonian University Heritage Metadata Portal). It integrates heterogeneous archival (AUJ), museum (MUJ), and

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_37

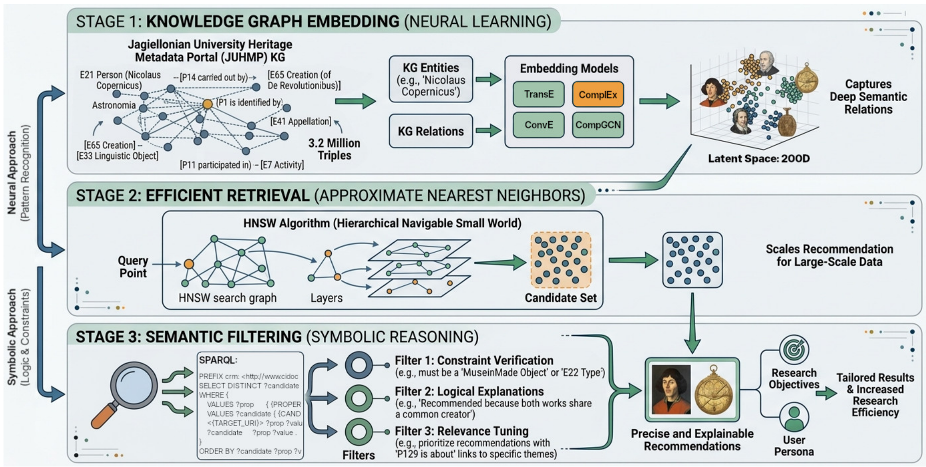

Fig. 1. Three-stage recommendation pipeline overview.

library (JL) datasets into a CIDOC-CRM-based KG capturing people, events, places, and objects associated with Jagiellonian University since the 14th century [20]. Cross-institution alignment in the second JUHMP prototype is based on a semi-automatically generated authority file with 824 verified matches [21].

Because the graph is person-centric, the primary scenario was defined as: *given a historical individual, identify other people who are similar or meaningfully connected*. Similarity deliberately remains open-ended to support exploratory research, i.e., scenarios not well served by keyword-based search.

With this in mind, two versions of the graph were used as the basis for the reported work: *Prototype2* with only cross-institution matches (402,948 triples) and *FullCAC_260128* extended with all AUJ persons (3,202,711 triples; the largest possible graph at the time of writing).

2.3 Pipeline implementation for CHExRISH use case

A preliminary benchmark on *Prototype2* compared four embedding models[1]: TransE, ComplEx, ConvE, and CompGCN. ComplEx achieved the best balance of quality, runtime (≈46 min for 10 epochs), and memory use (+64 MB). Although CompGCN achieved slightly higher accuracy, its training time (≈13.3h) made it impractical. ConvE showed consistently poor results. Hyperparameter tuning showed that $lr = 0.01$ (learning rate) and $dim = 200$ (embedding dimensionality) offered the best trade-off between predictive quality (*MRR*, *Hits@k*) and resource consumption for ComplEx. For ANN search, ComplEx embeddings were indexed with HNSWLib Python package. A grid search yielded the following configuration: $M = 16$ (graph connectivity), $efConstruction = 400$ (build-time

[1] For detailed results and a description of the experimental setup, see *ICCS2026_Supplement.pdf* in the repository linked in *Code and Data Availability*.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_37

```
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>

SELECT DISTINCT ?candidate ?prop ?value
WHERE {
    VALUES ?prop      { {PROPERTIES_URIS} }
    VALUES ?candidate { {CANDIDATES_URIS} }
    <{TARGET_URI}> ?prop ?value .
    ?candidate     ?prop ?value .
}
ORDER BY ?candidate ?prop ?value
```

Listing 1.1. SPARQL query used to check whether the target and candidates share the same value in given set of properties.

search breadth), $efSearch = 50$ (query-time search breadth). Because indexing is a one-time operation, selection criteria focused on retrieval quality (*Recall@k*) and query latency rather than build-time cost.

At query time, based on the target node, the pipeline first retrieves the nearest neighbors from the HNSW index. These candidates are then refined using a set of SPARQL queries that ensure semantic validity and contextual relevance. The semantic filtering stage is a general concept but is highly dependent on the KG used as input and covered scenarios. In evaluated scenario, candidates are restricted to instances of *crm:E39_Actor* or *crm:E21_Person*, after which a predefined set of SPARQL tests checks for shared relations with the target node. Allowed connection types include same objects connection, same events, close death dates, same place of birth, etc. For example, the "same events" filter identifies candidates linked to the same event via *crm:P11i_participated_in*, *crm:P12i_was_present_at*, or *crm:P14i_performed* (see Listing 1.1). A candidate is retained only if there is at least one such exact link. The pipeline also records the specific shared predicates, enabling for explainability.

2.4 Evaluation approach

After selecting the best model configuration, the full pipeline was applied to both KGs (Sect. 2.2). The evaluation focused on nine historically significant figures from Jagiellonian University (e.g., Nicolaus Copernicus; see Tab. 1). This approach, focusing on well-known individuals, facilitated the expert evaluation.

Quantitative evaluation focused on verifying the basic performance of the embedding and ANN stages. The experts then assessed whether the top 10 recommendations for each target were meaningful and historically grounded. They received an online questionnaire and rated each candidate on a scale: *2 – close connection*, a user reading information about X will be interested in person Y; *1 – distant connection*, the user may be interested in such a suggestion; *0 – very distant connection (incorrect)*, the user will not be interested in such a suggestion; *-1 – I do not know this person*. Optional comments could also be provided.

The expert study group included people from AUJ, MUJ and JL. These are people with extensive experience in the field of history and heritage of Jagiellonian University. Four experts were involved in the evaluation of *Prototype2*. The larger model (*FullCAC_260128*) was evaluated by 10 experts.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_37

Table 1. Number of close and distant recommendations as indicated by experts for both models. Target labels are abbreviated for clarity.

Target	*Prototype2*		*FullCAC_260128*	
	Close	*Distant*	*Close*	*Distant*
Benedykt z Koźmina . . .	1	5	2	1
Sebastian Sierakowski (hrabia)	0	7	1	9
Stanisław Reszka z Buku . . .	0	2	2	1
Jan III Sobieski (król Polski ; . . .)	0	1	0	6
Maciej Karpiga z Miechowa . . .	0	5	2	10
Mikołaj Kopernik (Copernicus) . . .	0	3	0	6
Jakub Górski (młodszy) . . .	0	2	3	6
Andrzej z Buku (starszy)	0	1	2	6
Jan Brożek (Broscius) . . .	0	2	2	6

3 Results and Discussion

3.1 Quantitative evaluation

For *Prototype2*, the final 120-epoch ComplEx model achieved $MRR = 0.2207$, indicating that the correct entity typically appeared within the top 4-5 ranked positions, although not consistently at rank 1. $Hits@1 = 0.1659$ confirms limited ability to place the correct entity first, while $Hits@3 = 0.2421$ and $Hits@10 = 0.3288$ show that the model retrieved the correct entity within the top 10 in roughly one-third of the queries. Overall, the model reliably surfaced relevant candidates but struggled to prioritize the correct one at the very top. Training required ≈22 hours on a CPU-only node.

For *FullCAC_260128*, training was completed in 108 hours (GPU node) with $MRR = 0.1393$, $Hits@1 = 0.1057$, $Hits@3 = 0.1572$ and $Hits@10 = 0.2098$. Lower metrics were expected due to the graph's characteristics: it is both sparser (many individuals with few connections) and denser in population (contains all persons linked to the University between the 14th and 18th centuries), making the embedding task more difficult. Despite this, the larger model produced semantically closer recommendations. The similarity of the top recommendation ranged from 0.40 to 0.85, compared to $[-0.41; 0.21]$ for *Prototype2* (cosine similarity ranging from -1 to 1). Its greater historical coverage also enabled the pipeline to recommend a wider and more accurate set of relevant individuals. Therefore, we expected stronger expert evaluation results for the larger graph despite the lower quantitative metrics.

3.2 Expert evaluation

Experts assessed the top 10 recommendations for each of the nine targets (historical figures), rating connections as close or distant (Tab. 1). Despite the *Prototype2* graph's limited size, experts validated at least one meaningful connection per target. When relevant links existed in the graph, the model typically surfaced them, and experts confirmed their correctness. However, the small graph

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_37

inherently restricted the number of plausible connections, limiting the frequency of close matches. The larger model (*FullCAC_ 260128*) produced more close and distant connections, as expected from its higher population density and broader temporal coverage. Experts confirmed that the pipeline successfully identified historically grounded relationships, reflecting the advantages of training on a more complete graph.

In the free comment fields, experts noted several factors that affected their ability to assess the recommendations. In some cases, individuals shared identical names and places of origin, making it difficult to determine whether a suggested connection was historically valid. The experts therefore suggested including identifiers from the AUJ database to avoid ambiguity. They also emphasized the need for greater explainability, expressing interest in seeing why the algorithm selected each recommendation, such as shared events or locations. Finally, they observed that the distinction between "close" and "distant" connections remained subjective, even with provided definitions, since evaluating the strength of historical relationships often depends on contextual knowledge and interpretation.

4 Conclusions

We presented an end-to-end three stage (embedding, HNSW candidate generation, SPARQL-based semantic filtering) recommendation pipeline for cultural heritage data. It exemplifies a neuro-symbolic recommender: vector neighborhoods provide recall, while ontological/SPARQL constraints supply correctness and explanations. This approach is aligned with current practice in heritage KGs and with sector needs for transparency [19].

The evaluation demonstrates both the feasibility and usefulness of the approach, while also highlighting the importance of ongoing collaboration with domain experts. Future development must include a dedicated user interface closely coupled with the KG exploration environment, allowing users to access stable identifiers and clear explanations for each recommendation. Experts also noted that the definitions of "close" and "distant" connections require refinement. Therefore, we plan to conduct workshops to co-design appropriate recommendation scenarios, evaluation criteria, and corresponding filters.

This hybrid framework addresses directly the overload of irrelevant or low-quality results that motivated the work by combining the predictive capabilities of neural models with the precision of symbolic constraints. With continued data curation and hybrid modeling, the method can generalize beyond persons to places, events, and artworks. Finally, text-to-SPARQL techniques [5] offer a promising path toward personalized recommendations without requiring users to manually define semantic filters.

Code and Data Availability. All code is publicly available in the repository: https://gitlab.geist.re/pro/chex-recs. Input KGs and intermediate files (e.g., embeddings) are available upon request.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_37

Acknowledgments. We would like to express our sincere gratitude to the experts from the JU Archives, the JU Museum, and the Jagiellonian Library for their time spent evaluating the pipeline.

This publication was funded by a flagship project "CHExRISH: Cultural Heritage Exploration and Retrieval with Intelligent Systems at Jagiellonian University" under the Strategic Programme Excellence Initiative at Jagiellonian University. The research for this publication has been supported by a grant from the Priority Research Area DigiWorld under the Strategic Programme Excellence Initiative at Jagiellonian University. We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2025/018037.

During the preparation of this work the authors used MS Copilot and Writefully in order to improve the readability and language of the manuscript. The initial version of the pipeline diagram (Fig. 1) was generated with Google Gemini. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alma'aitah, W.Z., Talib, A.Z., Osman, M.A.: Opportunities and challenges in enhancing access to metadata of cultural heritage collections: a survey. Artificial Intelligence Review **53**(5), 3621–3646 (2020). https://doi.org/10.1007/s10462-019-09773-w
2. Casillo, M., Colace, F., Conte, D., Lombardi, M., Santaniello, D., Valentino, C.: Context-aware recommender systems and cultural heritage: a survey. Journal of Ambient Intelligence and Humanized Computing **14**, 3109–3127 (4 2023). https://doi.org/10.1007/s12652-021-03438-9
3. Chansanam, W., Chaichuay, V., Manorom, P., Kanyacome, S., Sugimoto, S., Huang, Y.J., Zhang, G.l., Li, C.: Cultural heritage preservation through ontology-based semantic search systems. Conservation Science in Cultural Heritage **23**(1), 183–196 (2023). https://doi.org/10.6092/issn.1973-9494/20048
4. Dangerfield, M.C., Kalshoven, L.: Report and recommendations from the task force on metadata quality. Tech. rep., Europeana (May 2015), https://pro.europeana.eu/files/Europeana_Professional/Europeana_Network/metadata-quality-report.pdf
5. Hansali, A., Hocine, K.B., Zemmouchi-Ghomari, L., Ghomari, A.R.: Translation from natural language questions to sparql: A survey. In: 4th Edition of the National Study Days on Research on Computer Sciences, JERI'2020 (2020)
6. Ignatowicz, J., Kutt, K., Nalepa, G.J.: Position paper: Metadata enrichment model: Integrating neural networks and semantic knowledge graphs for cultural heritage applications. In: 2025 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2025). https://doi.org/10.1109/IJCNN64981.2025.11228881
7. Isaac, A.: Case study: Enriching and sharing cultural heritage data in europeana. Tech. rep., W3C (June 2012), https://www.w3.org/2001/sw/sweo/public/UseCases/Europeana/
8. Ishchuk, O.: Search and recommendation methods in graph knowledge bases describing cultural heritage. Master's thesis, Jagiellonian University (2025), supervisor: K. Kutt

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_37

9. Kamal, A.M., Golub, K.: Subject matters: Metadata standards and subject access for library and museum catalogues. In: Hanssen, J.M., Furuseth, S. (eds.) The Hermeneutics of Bibliographic Data and Cultural Metadata, pp. 204–239. National Library of Norway (2025)
10. Li, J., Bikakis, A.: Towards a semantics-based recommendation system for cultural heritage collections. Applied Sciences **13**, 8907 (2023). https://doi.org/10.3390/app13158907
11. Lian, Y., Xie, J.: The evolution of digital cultural heritage research: Identifying key trends, hotspots, and challenges through bibliometric analysis. Sustainability **16**(16), 7125 (2024). https://doi.org/10.3390/su16167125
12. Lorenzini, M., Rospocher, M., Tonelli, S.: On assessing metadata completeness in digital cultural heritage repositories. Digital Scholarship in the Humanities **36**(Supplement_2), ii182–ii188 (2021). https://doi.org/10.1093/llc/fqab036
13. Mariani, F.: Introducing VISU: vagueness, incompleteness, subjectivity, and uncertainty in art provenance data. In: Rochat, Y., Métrailler, C., Piotrowski, M. (eds.) Proceedings of COMHUM 2022. pp. 63–84. CEUR Workshop Proceedings, CEUR-WS.org (2022), https://ceur-ws.org/Vol-3602/paper5.pdf
14. Raza, S., Rahman, M., Kamawal, S., Toroghi, A., Raval, A., Navah, F., Kazemeini, A.: A comprehensive review of recommender systems: Transitioning from theory to practice. Computer Science Review **59**, 100849 (2026). https://doi.org/10.1016/j.cosrev.2025.100849
15. Ruotsalo, T., Haav, K., Stoyanov, A., Roche, S., Fani, E., Deliai, R., Mäkelä, E., Kauppinen, T., Hyvönen, E.: SMARTMUSEUM: A mobile recommender system for the web of data. J. Web Semant. **20**, 50–67 (2013). https://doi.org/10.1016/J.WEBSEM.2013.03.001
16. Schreur, P.E.: The use of linked data and artificial intelligence as key elements in the transformation of technical services. Cataloging & Classification Quarterly **58**, 473–485 (2020). https://doi.org/10.1080/01639374.2020.1772434
17. Shah, H.: A comparative study of hnsw implementations for scalable approximate nearest neighbor search (2025). https://doi.org/10.36227/techrxiv.175321947.71782908/v1
18. Silva, A.L., Terra, A.L.: Cultural heritage on the semantic web: The europeana data model. IFLA Journal **50**, 93–107 (2024). https://doi.org/10.1177/03400352231202506
19. Spillo, G., Musto, C., de Gemmis, M., Lops, P., Semeraro, G.: Recommender systems based on neuro-symbolic knowledge graph embeddings encoding first-order logic rules. User Model. User Adapt. Interact. **34**(5), 2039–2083 (2024). https://doi.org/10.1007/S11257-024-09417-X
20. do Valle Miranda, L., Kutt, K., Nalepa, G.J.: CIDOC-CRM and the first prototype of a semantic portal for the chexrish project. In: Bruns, O., Graciotti, A., Sartini, B., Tietz, T. (eds.) Proceedings of SemDH 2025. CEUR Workshop Proceedings, vol. 4009. CEUR-WS.org (2025), https://ceur-ws.org/Vol-4009/paper_14.pdf
21. do Valle Miranda, L., Mozolewski, M., Kutt, K., Nalepa, G.J.: A wikidata-based workflow for entity reconciliation strategies evaluation: A study on early modern polish personal names. In: Proceedings of ESWC 2026 (2026), accepted for publication (to appear)
22. Zaagsma, G.: Digital history and the politics of digitization. Digital Scholarship in the Humanities **38**(2), 830–851 (2022). https://doi.org/10.1093/llc/fqac050
23. Zhao, X., Wang, M., Zhao, X., Li, J., Zhou, S., Yin, D., Li, Q., Tang, J., Guo, R.: Embedding in recommender systems: A survey. CoRR **abs/2310.18608** (2023). https://doi.org/10.48550/ARXIV.2310.18608

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_37