Received: 25 March 2024 | Revised: 9 July 2024 | Accepted: 18 August 2024 | Published online: 2 September 2024

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2025, Vol. 4(1) 64–82
DOI: 10.47852/bonviewJCCE42022924

BON VIEW PUBLISHING

HILDEGARD: Human-in-the-Loop Data Extraction and Graphically Augmented Relation Discovery

Cosimo Palma[1,2,*]

[1] *Department of Informatics, University of Pisa, Italy*

[2] *Department of Literary, Linguistic and Comparative Studies, University of Naples "L'Orientale", Italy*

Abstract: This research paper presents HILDEGARD, an application conceived to guide a semi-expert user in the domain of cultural heritage data management toward the creation of a lightweight knowledge graph tailored for supporting Automatic Story Generation (ASG). For this purpose, a subset of CIDOC-CRM classes and properties is preliminarily selected to fit the domain of interest. The input is constituted by one or more seed-heritage objects selected from a knowledge base. In our case study, they are SPARQL-queried from a Linked Open Database for Italian Cultural Heritage. The shortest path algorithm is then run online on all couplets obtained by a combination of the Wikipedia entities from the selected entry-seeds descriptions. The retrieved entities are subsequently linked to their related DBpedia- or YAGO-entry in the chosen language, and the relationships among them are automatically retrieved. The proposed tool addresses different knowledge gaps and societal needs simultaneously, such as the lack of solutions tailored for narrative purposes in the cultural heritage domain, that is, to be used in a scenario where objects belonging to the same room must be linked through a narrative, which shall not only be coherent and informative but also engaging and interesting. The prototype, already able to generate the triples required for the following step of the proposed general ASG pipeline, is intended to be graphically enhanced so that the end user may guide the graph expansion interactively.

Keywords: knowledge graphs, digital heritage, digital storytelling, data management, ontology harmonization, human-in-the-loop

1. Introduction

The domain of cultural heritage (CH) has been transformed by digitalization more significantly than other ones bearing a broader societal impact [1]. Current research is exploring various strategies for creating a gamified connection with users, which has shown encouraging outcomes [2–5]; for instance, there has been a surge in systems that interact with users through natural language, whether via mobile apps [6] or social robots [7]. This shift is driven by an abundance of informational resources and a growing necessity for new ways to engage the public [8], whose attention's quality has shifted considerably during the last decades [9, 10]. The diminishing attention span due to an exponential increase in entertainment options, coupled with the audience's growing disconnection from traditional engagement with CH, has underscored the need to counter this trend through various strategies, ranging from the *transmedial narratives* [11, 12] to gamification [13], from virtual reality [14, 15] to interactive story-telling [2, 13, 16, 17], eventually able to leverage the visitor's physical real-time position in the museum as well as their personalized interests [10, 18] and emotions [19].

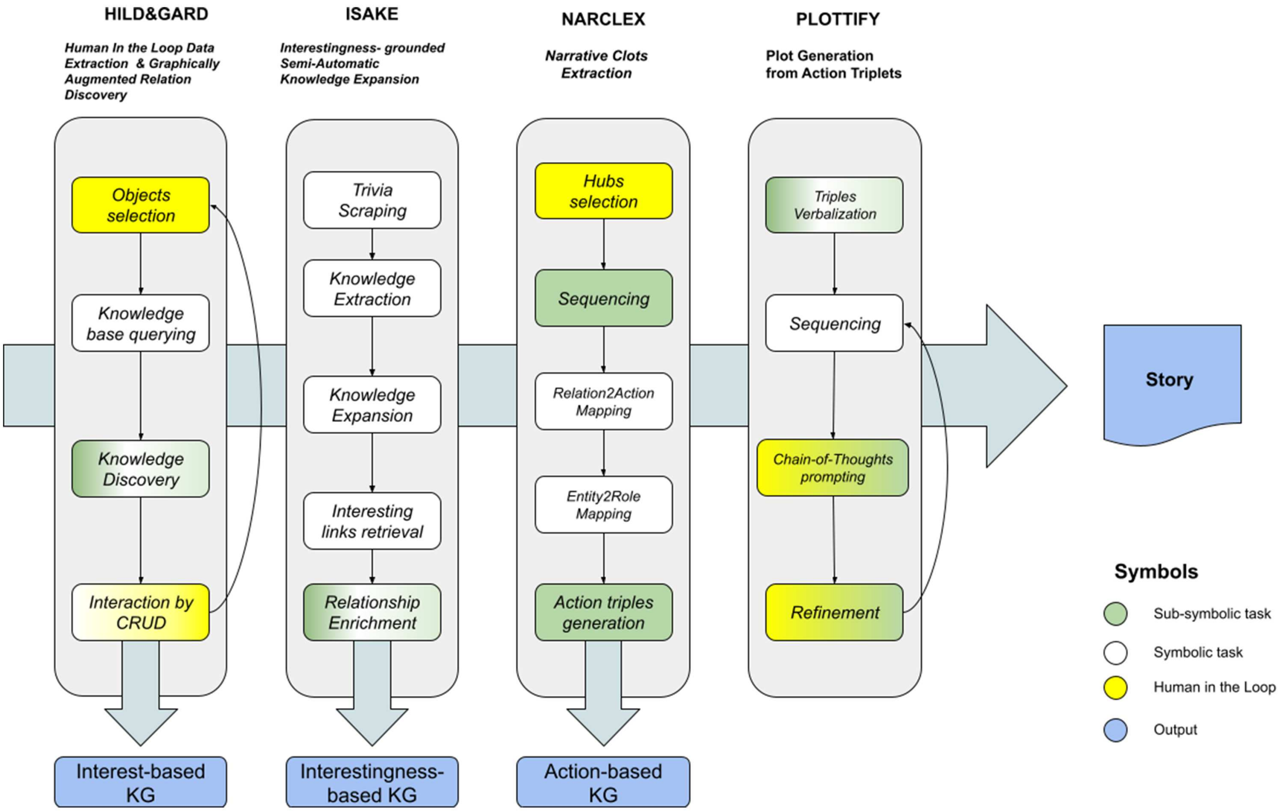
The present contribution is intended as the first module of a larger neural-symbolic [20, 21] (semi-)automated narrative generation pipeline for digital heritage (DH) drawing from Linked Open Data (LOD) [22]. "Semi-automated" here means "modular"; that is, all intermediate steps leading to the final outcome are already designed to be self-sufficient, easily interpretable, and ready to be used independently from one another at least as an aid to particular sub-tasks of story creation, delivering actionable sub-results ranging from the *mind map* to the *story text* across the *story map* and the *actions set*. The choice of keeping their eventual integration analogical also follows the principle that it is indeed the direct intervention of the user to foster his own engagement, which represents, in our case, the most important by-product of the text generation.

Being the concept of LOD immediately related to that of a *network* (or graph) [23], the data structure that better allows to visualize and understand relations between entities, it is worth to canalize efforts to fully exploit the extensive amount of standardized information in the *Semantic Web* (SW).

Thus, the costly but desirable transformation of undigitized and fragmented information currently fossilized in galleries, libraries, archives, and museums (GLAM) [24, 25] will be catalyzed by consequentially. An increase in demand for LOD, furthermore, would arguably influence positively their quality as well, which is currently spotted by many mistakes and inaccuracies. It is for this reason, that in the current study, we have assumed that LOD is intrinsically

***Corresponding author:** Cosimo Palma, Department of Informatics, University of Pisa and Department of Literary, Linguistic and Comparative Studies, University of Naples "L'Orientale", Italy. Email: cosimo.palma@phd.unipi.it

Figure 1
An overview of the neural-symbolic storytelling pipeline starting from input heritage objects



more trustworthy than synthetic data, although nowadays, neural models can even outperform humans in numerous annotation tasks [26]. Simultaneously, the prevailing view supports the notion that unearthing and revitalizing existing, yet hidden narratives hold more sway in capturing user interest than creating entirely new ones. The knowledge graph (KG) operates as a directed graph where the connections are semantically enhanced and marked with textual labels. Crafting a narrative from a KG, rather than just a textual depiction, necessitates therefore forming a knowledge base designed for storytelling from the very beginning. This process involves selecting entities and relationships that are likely to appear in genuine narratives as well as useful for constructing them (see Table 1). Moving from this consideration, in Palma [27], the following points are individuated as necessary steps to achieve constraint-based Automatic Story Generation (ASG):

1) Interest-based semantic hub construction;
2) Interestingness-based knowledge discovery;
3) Event sequences extraction from KG's narrative clots;
4) Link-aware recursive event generation.

Beyond the abovementioned features the KG will also serve as the principal interface between the user and the database [28].

For *DH*, we aim at creating a lightweight tool that fully exploits the SW in order to retrieve relationships among elements, providing therefore a bridge between domain-specific and cross-domain KGs; for the ASG, HILDEGARD (the Interest-based semantic hub construction) is the first important step toward producing simple subject-predicate-object sentences that will be afterward exploited as input for the task of outline-conditioned story generation [29, 30]. The application is open-source, which has been implemented in Python 3.7, and is fully compatible also with Python 3.9. The implementation required 2277 lines of code. The processor used is Intel(R) Core(TM) i7 –6600U CPU @ 2.60 GHz.

Despite significant advancements in generating stories with the help of machine learning (ML) techniques, ASG still faces challenges in creating outputs that are qualitatively deep and informatively rich [31]. To overcome this issue, a KG – especially its most *interesting paths* identified by the method described in the work of Palma [32] – is expected to enhance the large language model. This integration will equip the final output with user-selected contextual information throughout the entire workflow. This approach is especially beneficial in educational contexts, where the stories produced often have pedagogical applications.

Furthermore, in a curated exhibition, it often occurs that a selection of heritage objects shall be displayed in a particular sequence, to support an underlying narrative able to optimally engage the visitor [3, 33]. In general, the exploration of links among given elements, without defining a clear machine-readable question, poses itself as a relevant task, with promising developments for the field of computational (co-)creativity. The proposed tool places itself therefore at the crossroads of data management and computational creativity and tackles the problem of narrative generation connecting adjacent heritage objects in an exhibition, resorting to the harvesting of *serendipitous* connections between given entities [34]. It is observed that, generally, museum rooms revolve around the same main theme, which leads to the display of very similar objects together. Through this work, it is argued that the focal point according to which the disposition of objects in a room shall be designed is the *narrative*, which best awakens the visitor's interest, curiosity, and memorization capabilities. Nevertheless, the presented application offers the possibility to connect also very similar objects in a captivating way, by exploiting the entities featured in the object's description as retrieved from the knowledge base.

Enhancing the creation of narratives by repositioning heritage objects belonging to the same institution represents a cost-effective strategy to enhance museal revenues and visitor's engagement and information retention.

With respect to ASG, however, the Knowledge Graph built through this task constitutes already a necessary first step toward the story map [35]; from the other side, a reliable knowledge base is an indispensable requirement for texts that have not only entertainment as a goal but also and foremost the education, which requires a reliable and proof-checked ground knowledge [36]. In this context, the term "narrative" is used interchangeably with "story," which, unlike a significant segment of the existing scholarship [37], refers here to its traditional definition as an imaginative tale involving real or fictional characters, settings, and events, crafted to entertain, inform, and educate. The pivotal role of stories in learning new information [38] underscores the rationale behind integrating ASG and computational narratology with DH. Additionally, this research addresses the challenge of constructing a KG from keywords based on the premise that the user may not be aware of potential intriguing links between elements that must be utilized as thematic constraints. In Figure A4 (see Appendix), individual paths leading from "Alexander the Great" to "Anubis" show, as an example, the "hinge" entities bridging between the first two ones in the Wikipedia KG. Among them, for the story generation, only the most interesting ones will be chosen. The most adopted heuristics used in literature to abstract the broader and more complex concept of "interest" has been the concept of "similarity," which can be modeled by a formal concept analysis (FCA) approach [39] or by linguistic one, either by means of distributional semantics [40] or *synsets* as well harvested from Linked Data [41]. In the domain of CH, the linking of fragmented data may happen automatically by means of deep learning algorithms for knowledge expansion [42]. In our scenario, with the focus resting on the cognitive stimulation of the user, may it be the end user or the domain expert, we prefer to employ user intervention in the assessment and forwarding of the pipeline's tasks, instead of manually annotating a large dataset *a priori* to prepare a supervised ML classification algorithm. Therefore, our first identified task is the KG population with *human in the loop* (HIL). This approach is described in Schröder et al. [43], where human interaction is mostly intended as simple validation in a process based on ML. Nevertheless, the definition of HIL embraces more complex inputs from the user side, as well as symbolic procedures. In this paper, the term has been preferred to "semi-automatic,' or other similar expressions, because the task of automatic relation extraction and labeling according to the selected schema is indeed reserved for future implementation, featuring a pure HIL component. This would allow a continuous refinement of the predicting model without the necessity of resorting to external crowdsourcing. Finally, the task of expanding the knowledge base by link prediction [44] is hence reserved for the second step of the pipeline (see Figure 1).

In the following sections, I will first present an overview of all currently available tools for DH related to HILDEGARD, and then after showing the workflow as presented in the current version of the application, an experiment will be performed to assess the computational effectiveness of our KG-building procedure compared to the other working examples existing in the literature.

2. Related Work

Many tools have been developed aiming at technically supporting inquiries in the DH and digital history domains (see Table 2 for a comprehensive overview), such as the project *Palladio*. Developed four years ago at *Humanities + Design* (Stanford University), this project features functions for tabular data visualization, such as the *TimeLine, TimeSpan*, and *GalleryView*. Stemming from the same lab, the project *Data-Pen* (Stanford University) represents a remarkable attempt at harnessing the process of data manipulation in Digital Humanities with a *User-Experience* (UX) component. The possibility to add, remove, and edit nodes as the research shifts, thus creating graphs that are meaningful to the researcher, is one of the main points of inspiration for HILDEGARD.

Other tools not strictly designed for DH data management, such as *Infranodus* [45], focus on insights generation leveraging *text network analysis*. Among the reviewed systems, the only example that takes into account this *desideratum* is represented by *DeXTER* (DeepTextMiner), a deep learning, critical workflow to contextually enrich digital collections and interactively visualize them. It is task-oriented (as opposed to result-oriented), and it is designed to be generalizable and interoperable (i.e., it is user-experience independent) [46]. Since then, these projects have not been further expanded and maintained, unlike the project *ResearchSpace* [47], an open-source system that embraces the complexities of knowledge creation, integration, sharing, and preservation from relational and process-oriented perspectives, facilitating transparent synthesis, rejection, reconciliation, and adoption of propositions and arguments. *ResearchSpace* generates interconnected data narratives at various levels of detail.

Altogether, the above mentioned tools can be used by CH experts, researchers, or in general communities with a higher degree of expertise regarding LOD best practices. As an example, the project *Recogito*, a partner of the *Pelagios Network,* attempts to reduce technicality (thus improving usability) by proposing a collaborative annotation system, whereby people and places' labeling is powered by controlled vocabularies.

In the work of Hyvönen and Rantala [48], a knowledge-based approach for finding serendipitous semantic relations between resources in a KG[1] is presented. The developed system takes two elements as input and deploys the semantically shortest path between them.

SMBVT, a semi-automatic workflow to produce story maps from textual documents is assembled in Bartalesi et al. [35] and Metilli et al. [49], whereby natural language processing and *Wikidata* services are leveraged to extract key concepts, to assemble a logically ordered sequence of enriched dataset events, producing an interoperable LOD semantic knowledge base for event exploration and inter-story correlation analyses. The transition from a synchronic to a diachronic visualization, while fostering to some extent the "storification" of the displayed content, yet is not explicitly designed with the perspective of approaching the final generation of a narrative, which is the main aim of the present work. However, this topological interpretation of "story" matches the graph-based visualization, which in our case is used as a bridge toward the final full-textual realization. At the same time, it represents an independent pipeline block, already fully exploitable for educational and creative purposes, a topic deepened, among others, in *EduKG* [50]. Leveraging heterogeneous educational data,

[1] http://www.kulttuurisampo.fi/ff.shtml

an interdisciplinary and fine-grained ontology for uniformly modeling knowledge and resources in K-12 education is designed. Guided by this ontology, the authors propose a flexible methodology for interactively extracting factual knowledge from textbooks. In Schreiber et al. [51], the potential of SW technologies is combined with advanced presentation methods to drastically improve indexing and search functionalities within large-scale virtual collections of CH resources. The architecture of the *MultimediaN E-Culture* project is strictly built upon open web standards such as XML, SVG, RDF/OWL, and SPARQL, ensuring compatibility and widespread accessibility. Nowadays, in the CH domain, single institutions provide their open data to the national aggregator, which can forward them to the European database for CH known as *Europeana* [52, 53]. The Europeana Data Model (EDM) [54], the schema underlying the chosen dataset, is widely compliant with CIDOC-CRM, although the former lacks some equivalent classes of the latter, whose more fine-grained ontology would better allow it to capture interesting nuances for story building. The CIDOC conceptual reference model [55] schema constitutes an ideal starting point for story generation because of the rich logic specification including temporal tenses. Currently, the effort is being performed to find strategies, including ML approaches, to align already existing knowledge silos with this schema [56].

In the encoding example featured in Figure 2 [55], the event "Winkelmann is seeing Laokoon" is modeled according to the ontology, thus also instantiating a structured building block of a potential narrative.

The proposed pipeline fully relies on already existing techniques and tools. However, to the best of our knowledge, the formulated task/use-case scenario has never been addressed as such. Another relevant example for this case is *museum-digital:deutschland*, the online database of German museums. Despite the richness of the database and the user-friendliness of the query-results graph visualization, which much resembles the one envisioned for HILDEGARD, there are no connections to external general-domain knowledge.

Similarly, the beautiful *BabelNet*'s explorable network [57] is based on a single node hub. The related nodes, when clicked, shift the visualization to its own semantic hub, not showing the connection(s) between it and the previously selected node.

Other useful tools for exploring relationships among Wikipedia entities are *Lodmilla* and *LodLive*.

Conversely, in Hyvönen and Rantala [48], *Kulttuurisampo* successfully addresses the need for finding semantic relations between resources in a KG, but without providing either the KG visualization or the possibility for the user to interact with it. Therein, the shortest path algorithm is run offline, a factor that prevents the connection of the story-map knowledge initially harvested to the cross-domain knowledge needed in sight of the story generation.

The most relevant work dealing with the search for serendipitous relations is *DBpedia Relationship Finder* [58]. As of today, it is not available. In the proposed experiment, we exploit a custom-built adaptation of its key algorithmic process to assess the validity of our approach.

2.1. Problem statement

In the related work, a series of features emerge that the ideal Digital Heritage Data Management tool shall possess: a *KG visualization,* as a device to immediately grasp connections in a visual way, as well as an ideal domain-specific for representing, sharing, maintaining, and discovering information; *alternative visualizations,* such as gallery view or map view; a *user interaction paradigm* allowing to fully manipulate data; a function to retrieve and visualize relations among objects, such as the retrieval of *shortest path* between two nodes; the possibility to *expand nodes* to fetch finer-grained information about entities; the capability to *generate information* (as both entities and relationships) by means of technologies such as web scraping or generative AI; *generation of insights* using network and text analysis; interoperability, as the capability of the system to use non-native ontologies and export the result in a format that can be easily imported in

Figure 2
A glimpse of the CIDOC-CRM schema, the most established ontology for cultural heritage

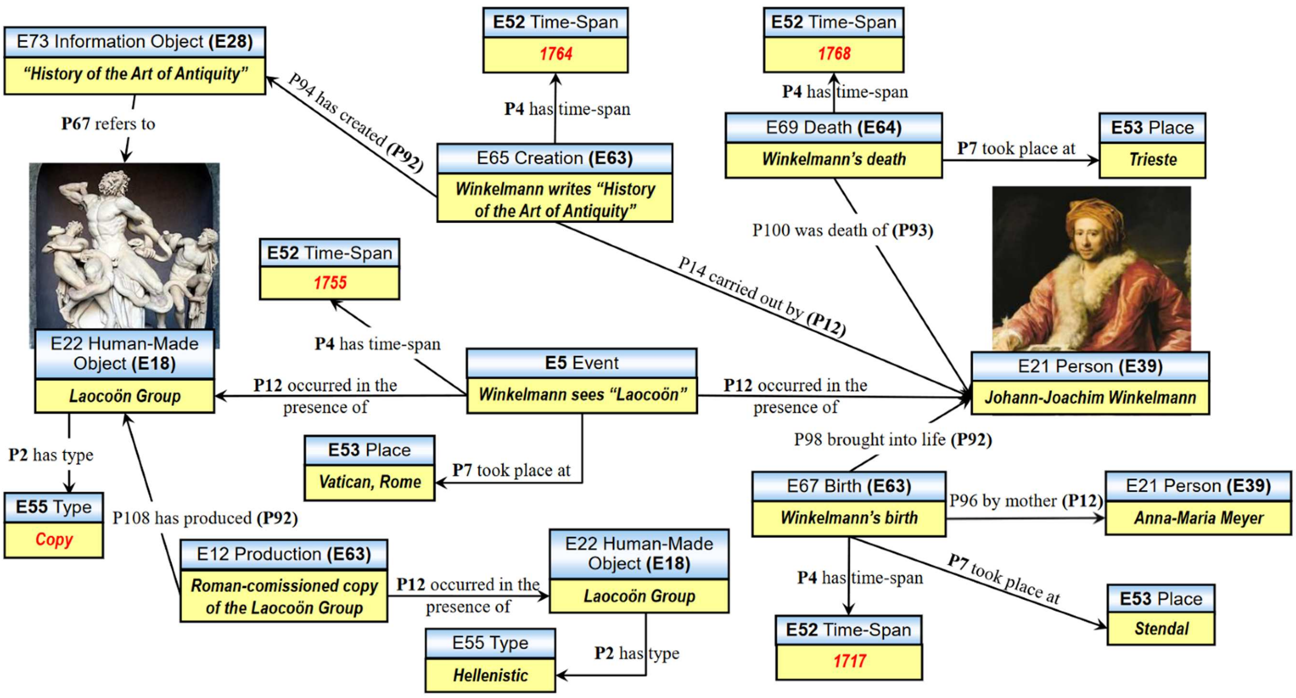


Table 1
A comparative table of the reviewed cultural heritage data management tools

Tool	1	2	3	4	5	6	7	8	9	10	11
HILDEGARD	Available	Under development/possible	Partially available	Available	Partially available	Available	Available	Available	Available	Under development/possible	Under development/possible
ResearchSpace	Available	Available	Available	Unavailable	Available	Unavailable	Unavailable	Unavailable	Unavailable	Available	Unavailable
Palladio	Available	Available	Available	Unavailable	Available	Unavailable	Unavailable	Unavailable	Unavailable	Available	Unavailable
Datapen	Available	Available	Available	Unavailable	Available	Unavailable	Unavailable	Unavailable	Unavailable	Available	Unavailable
Dexter	Partially available	Available	Unavailable	Unavailable	Available	Under development/possible	Unavailable	Unavailable	Under development/possible	Unavailable	Available
Kulttuurisampo	Unavailable	Available	Unavailable	Available	Unavailable	Unavailable	Unavailable	Available	Partially available	Available	Unavailable
SMBVT	Under development/possible	Available	Available	Unavailable	Available	Partially available	Unavailable	Unavailable	Under development/possible	Available	Unavailable
LODmilla	Available	Unavailable	Available	Partially available	Available	Unavailable	Available	Unavailable	Unavailable	Unavailable	Unavailable
LODlive	Available	Unavailable	Unavailable	Unavailable	Available	Partially available	Unavailable	Partially available	Under development/possible	Under development/possible	Unavailable
Babelnet	Available	Unavailable	Unavailable	Unavailable	Available	Available	Unavailable	Partially available	Available	Partially available	Unavailable
Museum-Digital	Available	Available	Unavailable	Unavailable	Available	Unavailable	Unavailable	Under development/possible	Available	Available	Unavailable
Infranodus	Available	Unavailable	Available	Available	Available	Available	Available	Under development/possible	Partially available	Unavailable	Available

Note:
1) KG visualization;
2) Alternative visualizations;
3) CRUD;
4) Shortest path retrieval;
5) Expandable nodes;
6) Nodes generation;
7) Relationships generation;
8) Interoperability;
9) Multilinguality;
10) Multimodality;
11) Insights generation.

Color Legend:

Available	(green)
Partially available	(light green)
Under development/possible	(yellow)
Unavailable	(red)

other system; *multilinguality*; and, finally, *multimodality,* that is, the capability of handle more data types in different languages (see Table 1).

On this basis, the objective of this paper is to create a tool that, while addressing the societal need and the ASG- knowledge gap pointed out in the introduction, tries at the same time to include as many features as possible from the ones listed here above, considering also that some of them are not currently maintained or freely available.

3. Workflow Overview: A Case Study on the National Archeological Museum of Naples

In our case study, we exploit a compact dataset, the cultural data provided by the Italian Ministry of Cultural Heritage and Activities (MiBAC), modeled according to the ArCo ontology [59, 60], which also partially reuses the established CIDOC-CRM schema. In specific, the selected museum is the National Archeological Museum of Naples.

Every heritage object entry is usually related to a conservation site, such as a museum, a description, an art technique, and so on. These metadata constitute the tails of facts (or triple sets), whereby the relation is defined by some property specified in the referred ontology, and the head is the core entity, defined by a class and a unique identifier.

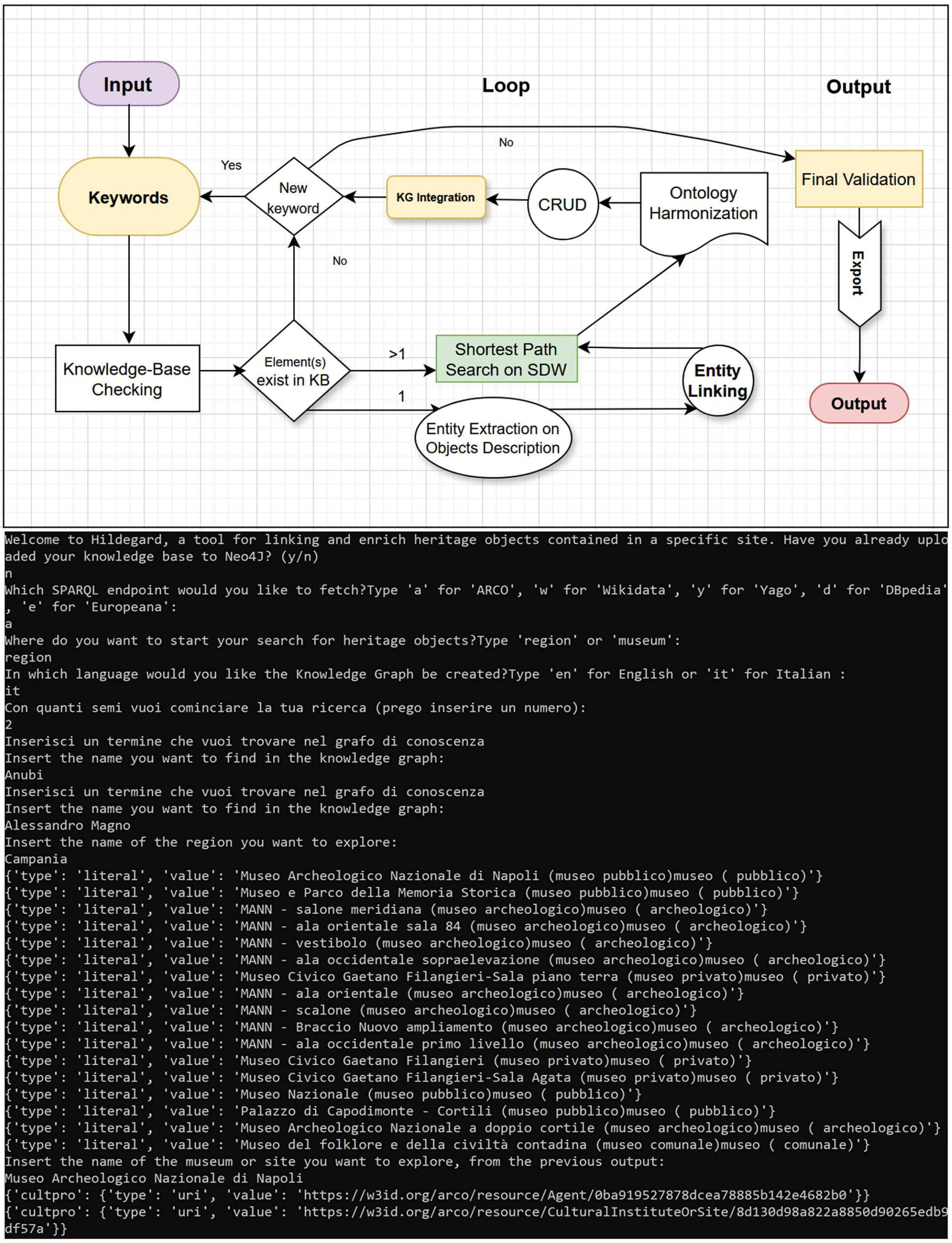
This choice reflects the initial goal to work with Italian data, as well as the practicality of dealing with a cleaner and smaller dataset, which simplifies the initial technical setup. Nevertheless, HILDEGARD can fetch multiple knowledge bases in all available languages (to be inserted in input as the related code in the *ISO 639-1* nomenclature). At the current stage of development, *Europeana* is the only CH knowledge base available in the system besides *Arco*.

As specified at the beginning, only a subset of the CIDOC-CRM schema has been chosen for manual labeling of entities and relationships, with the aim of moving from the domain of CH toward the Open World. This operation is the first encountered step for tailoring our KG for ASG. Among all possibilities, CIDOC-CRM has been preferred to (narratologically speaking) more adequate Upper Ontologies, such as DUL[2], because it is at the same time strictly related to our domain of origin, namely, the CH, and almost as rich and comprehensive as an Upper Ontology. Moreover, properties such as "refers to," "is about," or "represents" have been preferred to equivalent yet better-known alternatives because of the principle of *homogeneity*, which is regarded as higher in priority in comparison to the other common principle of LOD management, the *interoperability*.

An HIL approach allows the user to steer the construction of a semantic hub embracing and expanding the input concepts.

[2]http://www.ontologydesignpatterns.org/ont/dul/DUL.owl

Figure 3
HILDEGARD - Interest-based knowledge graph software architecture and user interaction command prompt




```
Welcome to Hildegard, a tool for linking and enrich heritage objects contained in a specific site. Have you already uplo
aded your knowledge base to Neo4J? (y/n)
n
Which SPARQL endpoint would you like to fetch?Type 'a' for 'ARCO', 'w' for 'Wikidata', 'y' for 'Yago', 'd' for 'DBpedia'
, 'e' for 'Europeana':
a
Where do you want to start your search for heritage objects?Type 'region' or 'museum':
region
In which language would you like the Knowledge Graph be created?Type 'en' for English or 'it' for Italian :
it
Con quanti semi vuoi cominciare la tua ricerca (prego inserire un numero):
2
Inserisci un termine che vuoi trovare nel grafo di conoscenza
Insert the name you want to find in the knowledge graph:
Anubi
Inserisci un termine che vuoi trovare nel grafo di conoscenza
Insert the name you want to find in the knowledge graph:
Alessandro Magno
Insert the name of the region you want to explore:
Campania
{'type': 'literal', 'value': 'Museo Archeologico Nazionale di Napoli (museo pubblico)museo ( pubblico)'}
{'type': 'literal', 'value': 'Museo e Parco della Memoria Storica (museo pubblico)museo ( pubblico)'}
{'type': 'literal', 'value': 'MANN - salone meridiana (museo archeologico)museo ( archeologico)'}
{'type': 'literal', 'value': 'MANN - ala orientale sala 84 (museo archeologico)museo ( archeologico)'}
{'type': 'literal', 'value': 'MANN - vestibolo (museo archeologico)museo ( archeologico)'}
{'type': 'literal', 'value': 'MANN - ala occidentale sopraelevazione (museo archeologico)museo ( archeologico)'}
{'type': 'literal', 'value': 'Museo Civico Gaetano Filangieri-Sala piano terra (museo privato)museo ( privato)'}
{'type': 'literal', 'value': 'MANN - ala orientale (museo archeologico)museo ( archeologico)'}
{'type': 'literal', 'value': 'MANN - scalone (museo archeologico)museo ( archeologico)'}
{'type': 'literal', 'value': 'MANN - Braccio Nuovo ampliamento (museo archeologico)museo ( archeologico)'}
{'type': 'literal', 'value': 'MANN - ala occidentale primo livello (museo archeologico)museo ( archeologico)'}
{'type': 'literal', 'value': 'Museo Civico Gaetano Filangieri (museo privato)museo ( privato)'}
{'type': 'literal', 'value': 'Museo Civico Gaetano Filangieri-Sala Agata (museo privato)museo ( privato)'}
{'type': 'literal', 'value': 'Museo Nazionale (museo pubblico)museo ( pubblico)'}
{'type': 'literal', 'value': 'Palazzo di Capodimonte - Cortili (museo pubblico)museo ( pubblico)'}
{'type': 'literal', 'value': 'Museo Archeologico Nazionale a doppio cortile (museo archeologico)museo ( archeologico)'}
{'type': 'literal', 'value': 'Museo del folklore e della civiltà contadina (museo comunale)museo ( comunale)'}
Insert the name of the museum or site you want to explore, from the previous output:
Museo Archeologico Nazionale di Napoli
{'cultpro': {'type': 'uri', 'value': 'https://w3id.org/arco/resource/Agent/0ba919527878dcea78885b142e4682b0'}}
{'cultpro': {'type': 'uri', 'value': 'https://w3id.org/arco/resource/CulturalInstituteOrSite/8d130d98a822a8850d90265edb9
df57a'}}
```

The creation, reading, updating, and deleting (CRUD) [61] as an interaction paradigm instantiated on the application's front end is perfectly mapped in the back end as MERGE (equating to a CREATE command, which also avoids duplicates), SELECT, UPDATE, and DELETE queries. Given the experimental focus of the setup, the prototype momentarily allows to select the initial elements only by the input of strings in the Python terminal[3] (see Figure 3 [27]). In the current prototype, the UI guides the user toward the creation of the KG by means of questions. Through them, the system downstreams the user interest for a specific geographical region and then for a museum or institution, whose DH database is hence fetched (see Figures A1 and A2 in Appendix). The knowledge base, which is queried by means of CIPHER on Neo4, is connected to the Python development environment. The most intuitive way to solve the abovementioned problem, in case only two start-seeds are provided, is running the shortest path algorithm (as Dijkstra's [62]). However, if the elements are several, the problem may become very complex (a realization of the NP-hard traveling salesman problem). In a real case scenario, as well as for practical reasons, the user would choose between two and a maximum of ten items, which configures the

[3]To allow a dynamic interaction with the user, this module will be implemented as a Flask web application, as well as the other pipeline's modules requiring a similar feature. Refer to paragraph 7.1 for further details.

Table 2
CIDOC-CRM classes and properties selected for HILDEGARD's output KG

CIDOC-CRM Classes	CIDOC-CRM Properties
E4 *Period*	P2 *has type*
E5 *Event*	P4 *has time-span*
E18 *Physical Thing*	P5 *consists of*
E19 *Physical Object*	P15 *was influenced by*
E21 *Person*	P20 *has specific purpose*
E53 *Place*	P53 *has (former) location*
E29 *Design or Procedure*	P67 *refers to*
E39 *Actor*	P94 *has created*
E57 *Material*	P96 *by mother*
E90 *Symbolic Objects*	P97 *from father*
CIDOC-CRM Properties	P129 *is about*
P103 *was intended for*	P130 *shows features of*
P104 *is subject to*	P132 *spatiotemporally overlaps with*
P106 *is composed of*	P138 *represents*
P124 *transformed*	P172 *contains*
P148 has component	P126 *employed*
P196 *defines*	P102 *has title*

Table 3
Comparison table of entity linking and typing performance of Wikifier and DBpedia Spotlight. *Threshold* set for both tools at 0.6.

Metric	Value
Total Wikifier	6524
Total DBpedia	3940
Total Common	650
Total Only Wikifier	5874
Total Only DBpedia	2997
Overall Wikifier Types	166,332
Overall DBpedia Types	3069
Avg Wikifier per Text	3.632516704
Avg DBpedia per Text	2.19376392
Avg Common per Text	0.361915367
Number of Texts	1796
Text Max Length	100 chars

problem as possible to be solved with doable computational costs by applying the Dijkstra's algorithm iteratively on combinations of two entities or as an alternative by means of the Floyd–Warshall algorithm [63]. Since the algorithm is run online with Wikipedia entries inserted by a web scraping module in "Six Degrees of Wikipedia," it is important to bring the results back into the selected language, in case it differs from English. The *CIPHER*-query mapping titles to descriptions is shown in Figure 4. The entities are not directly linked, although the matching metadata allows to fathom an underlying degree of similarity among them. For this reason, it is necessary to build a larger KG starting from the KB at our disposal, linking it to large cross-domain KGs such as *Wikidata, DBpedia*, or *YAGO* [64].

DBpedia would be the ideal target because it provides direct links to the other mentioned large knowledge bases. However, it currently delivers reliable results only for English, German, and Spanish. Other languages, such as Italian, if selected, redirect to non-existing web pages. For this reason, the algorithm has been programmed in such a way that the linking is performed on DBpedia only if the language-related link correctly works; otherwise, the retrieved entity must be linked to the YAGO resources, which contains everything that also exists in Wikidata, together with the enrichment from the Schema ontology.

The CIDOC-CRM properties and classes listed in Table 2 have been selected according to the axioms of *simplicity* and *generality*. All properties strictly regarding the domain of CH have been left out. This selection is thought to be redefined through usage and be further integrated with DBpedia relations harvested by SPARQL queries.

Although at this stage we still find ourselves in a synchronic dimension, a principle of *storification* is already being introduced, that is, the *contextualization*. For simplicity's sake and for better mapping with the 'title-caption-href'-structure, we query the *ArCo* knowledge base to obtain only *titles, descriptions,* and *URIs* of the entities. The retrieved triplets (tuples of three elements) are transformed into RDF triples according to the CIDOC-CRM schema following the preliminary mapping shown in Algorithm 2.

In cases of too high specificity, such as, for instance, archeological artifacts, the pointed entity may not be found even in Wikidata, an issue which could be solved by finding the entity with the highest semantic similarity to the input object. If only one entity is selected, an *Entity Extraction* may be performed on its related *Description*. To this extent, we have chosen *Wikifier* [65], showing a higher accuracy in entity detection, linking, and typing compared to the better-known *DBPedia Spotlight*[4] (see Table 3 for a detailed overview of their quantitative comparison). In case of more starting seeds, this step may be considered as optional since the *Entity Alignment* with the above mentioned KGs may suffice. Once the hub embedding the connections among the selected elements is displayed, the process of pruning, rearranging, and expanding will follow.

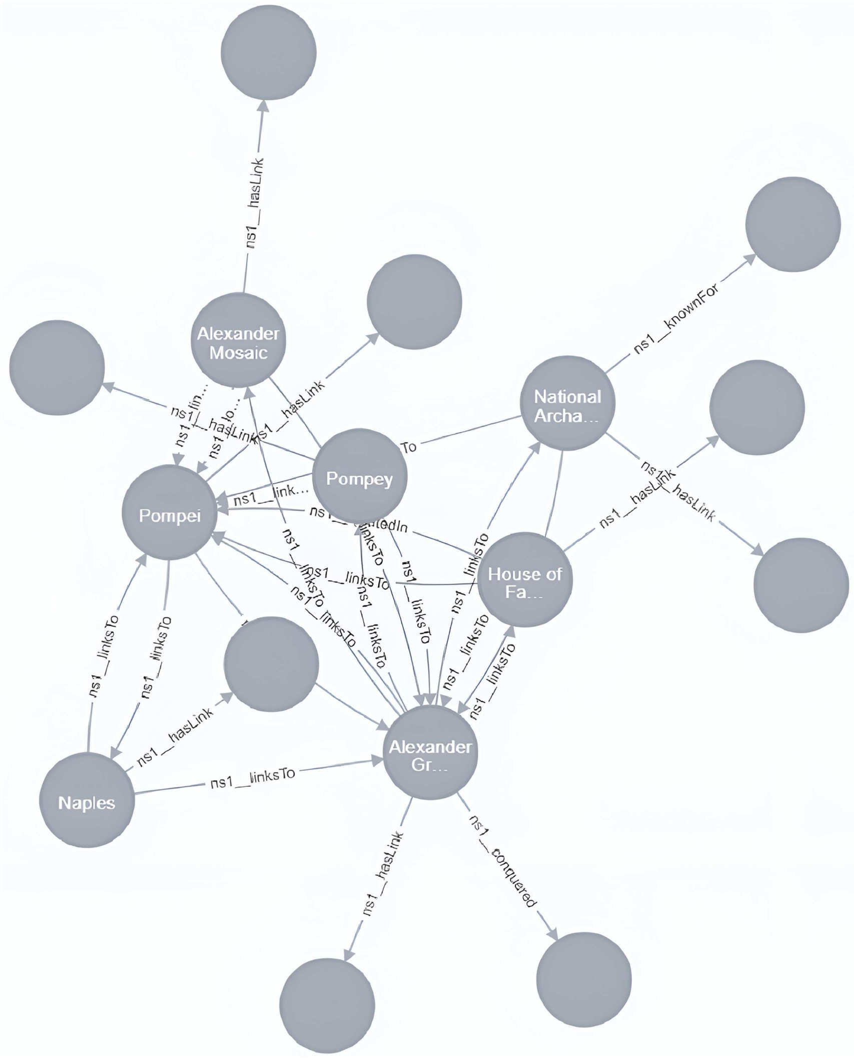
Figure 4
A setup for triangle completion

```
MATCH (d) where d.ns0__variable="descr"
MATCH (t) where t.ns0__variable="title"
MATCH (d)←[:ns0__binding]-(commonNode)-[:ns0__binding]→(t)
CREATE (d)-[:describes]→(t);
```

The workflow can be repeated *ad libitum* by selecting nodes from the newly generated ones and proceeding with a new path extraction. In the envisioned final application, this operation occurs by simply selecting the nodes, whereas in the current prototype, node titles still must be inserted analogically. The cycle will end only when the user is satisfied with the semantic

[4]https://www.dbpedia-spotlight.org/api

Figure 5
A Neo4J visualization of the Wikipedia entities extracted from the heritage objects descriptions through the shortest path algorithm



hub he cooperated to create. It produces textual files containing *wikimedia commons* objects related to the entities, the DBpedia resources related to them, and above all a JSON file containing all heritage objects queried from the selected museum integrated with the new entities linked from Wikipedia, which in turn will constitute the input of the following step of the pipeline: *ISAAKx* (Interestingness-grounded Semi-AutomAtic Knowledge eXpansion, as in Figure 1) [32].

In Figure 5, the seed entity couples are "Battle of Issus-Trier" and "Alexander the Great-Pompei," featured in the National Archeological Museum of Naples. The following method for KG building has not considered finding the shortest path between entities according to simple relationships. All connections shown above correspond to the Dbpedia property *dbo:wikiPageWikiLink*, meaning that the start entity is contained in the Wikipedia page of the end entity. This is a very general connection that in the following will be called, together with *cidoc-crm:P67_refersTo*, a *trivial relationship.* In the following section, two approaches for shortest path retrieval are compared, and a methodology for retrieving *non-trivial* relationships is provided.

4. In-Depth Analysis of Adopted Algorithms

The algorithms developed in HILDEGARD for harvesting entities and connecting them resort to a combination of web scraping and ontology harmonization. The first procedure (see Algorithm 1) leverages a website for executing the shortest path since there is no direct way to implement it using the DBpedia SPARQL endpoint. The *triple* mentioned in the algorithm defines sets of three elements that do not have to be confused with the RDF triples of the final KG.

Algorithm 1

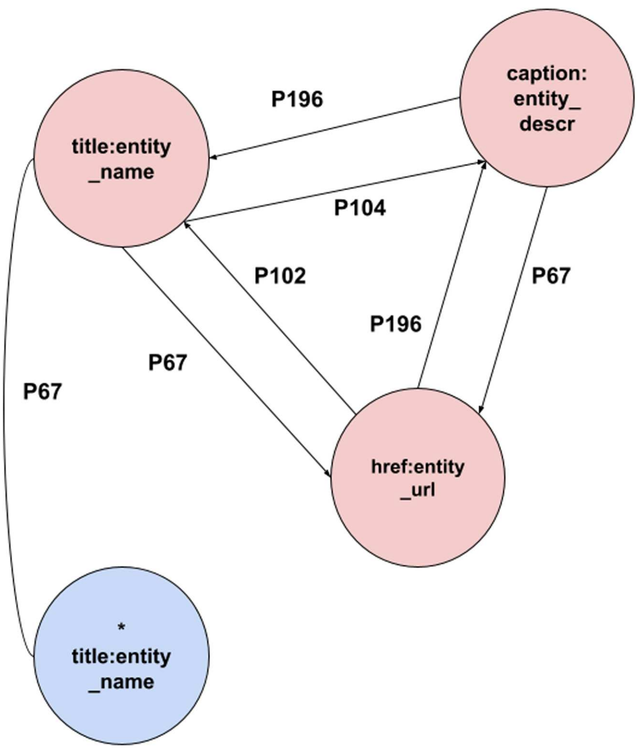
Shortest path web scraping between Wikipedia entities

```
Input  : start_entity, end_entity, file
Output: triples
browser ← openBrowser();
browser.get("https://www.sixdegreesofwikipedia.com/?source="
  + start_entity + "&target=" + end_entity);
webtext ← findLocationOfIndividualPaths();
hrefsList ← [];
titlesList ← [];
captionsList ← [];
ropeslist ← webtext.tolist();
for line in ropeslist do
    title ← findTitle(line);
    titlesList.append(title) ;
    caption ← findCaption(line);
    captionsList.append(caption);
    href ← findHref(line);
    hrefsList.append(href);
triples ← getTriplesFromLists(titlesList,
  captionsList, hrefsList);
if file T̄rue then
    // Write triples into file
    ;
return triples
```

Hereby entities and related descriptions and *urls* are harvested to mimic the same kind of information obtained by querying the museum-related LOD. Afterward, the relationships are created among *titles, URLs,* and *descriptions* using the CIDOC-CRM properties P67:refersTo, P102:hasTitle, P104:isSubjectTo, and P196:defines, as shown in Algorithm 2.

Algorithm 2

A graphic representation of the algorithm developed for ontology alignment according to our subset of the CIDOC-CRM schema (refer to Table 1)



According to this approach, *dbo:wikiPageWikiLink* has been rendered as *P67:refersTo*. In alternative to this method, only four possible options have been found:

1) Importing DBpedia locally and performing a shortest path algorithm in the Data Management System (not feasible).
2) Leveraging DBpedia KG embedding to solve the problem as an ML one.
3) Exploiting the Wikipedia API to perform a Clickstream Walk from one entity to another.
4) Querying DBpedia to find relationships between two entities, in the following fashion [58]:

Query 1

PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>

SELECT ?entity1 ?pf1 ?mid1 ?pf2 ?mid2 ?pf3 ?mid3 ?pf4 ?entity2

WHERE {
VALUES (?entity1 ?entity2) {(dbr:Anubis dbr:Alexander_the_Great)}
?entity1 ?pf1 ?mid1.
?mid1 ?pf2 ?mid2.
?mid2 ?pf3 ?mid3.
?mid3 ?pf4 ?entity2 ...

The limit of this approach is that it is unknown in advance how many degrees of similarity there are between the given entities (which means, how many hops separate them). For this reason, it results in heavy computational costs and requires a significant involvement of the user. Although in this paper we have advocated for a Human-In-The-Loop approach, it is assumed that the human interaction must be relevant to the user primarily, then only in second instance also functional to keep the workflow flowing. A partial solution to this problem would be to run first the query with the minimal number of hops, and gradually increasing their number until the query is delivering results. From another perspective, one may need to find the explicit relationships in the Dbpedia KGs, namely relationships that are more specific than *cidoc-crm:P67_refersTo* or *dbo:wikiPageWikiLink.*

This can be achieved by means of filters, such as:

Filters
...
FILTER(?entity1 != ?mid1 && ?entity2 != ?mid2)
FILTER (?pf2 NOT IN (dbo:Person, owl:Thing))
FILTER (?pf1 NOT IN (dbo:Person, dbo:wikiPageWikiLink, owl:Thing))
FILTER (?pf3 NOT IN (dbo:Person, owl:Thing))
FILTER (?pf4 NOT IN (dbo:Person, dbo:wikiPageWikiLink, owl:Thing))}
LIMIT 50

The proposed filters individuate other sorts of trivial classes, such as *dbo:Person* and *owl:Thing.*

The previous query produces results of the sort of:

http://dbpedia.org/resource/Anubis
http://dbpedia.org/property/cultCenter

http://dbpedia.org/resource/Cynopolis
http://dbpedia.org/ontology/wikiPageWikiLink
http://dbpedia.org/resource/Plutarch
http://dbpedia.org/ontology/wikiPageWikiLink
http://dbpedia.org/resource/Demetrius_I_of_Macedon
http://dbpedia.org/property/predecessor
http://dbpedia.org/resource/Alexander_the_Great

In the previous query, removing all *dbo:wikiPageWikiLink*. would cause an empty query output. Nevertheless, the finding process of non-trivial DBpedia relationships has been proven as successful mostly between entities with one degree of similarity (in other words, directly connected by *dbo:wikiPageWikiLink* or *P67:refersTo*). For example, the following output describes the relationships between "Anubis" and "Anput," as retrievable running Query 1 and filtering out all trivial relationships:

Solution

http://dbpedia.org/resource/Anubis
http://dbpedia.org/property/offspring
http://dbpedia.org/resource/Kebechet
http://dbpedia.org/property/parents
http://dbpedia.org/resource/Anput

5. Evaluation of Alternative Approaches

The hybrid nature of this research prevents a thorough comparative and quantitative analysis of the tools mentioned in Section 2. Moreover, a multifaceted UX-based experiment would require a fully functioning front end, as well as the availability of resources to proceed with human evaluation. For all these reasons, for now, we only cover the evaluation of the two other available strategies for shortest path retrieval in a KG. An explorative solution featuring KGs embeddings [66] has been considered for the entity pair *Alexander the Great-Anubis* training random walks from the start to the end entities. The result achieved through this method is:

KG- Embeddings

Alexander the Great –[wikiPageWikiLink]–> Priene Priene –[wikiPageWikiLink]–> Anubis

Despite its correctness, it represents only one of all possible shortest paths. Furthermore, it presents the trivial relationships we aim to replace with richer and more informative ones, such as those achieved in *Solution*.

The generation on random walks based on user-clickstreams provides more and more interesting paths that present trivial or even empty relationships:

Alexander the Great –[unknown_relation]–> Achaemenid dynasty
Achaemenid dynasty –[wikiPageWikiLink]–> Arses of Persia
Arses of Persia –[unknown_relation]–> Neferneferuaten
Neferneferuaten –[unknown_relation]–>Karomama II
Karomama II –[unknown_relation]–>
Nineteenth Dynasty of Egypt family tree
Nineteenth Dynasty of Egypt family tree –[wikiPageWikiLink]–> Ramesses II
Ramesses II –[wikiPageWikiLink]–> Anubis

In Table A1 (see Appendix), a sample of entity pairs obtained by this procedure is provided, for which *Query 1* has been performed iteratively, starting with one middle entity and adding more until the result is delivered. The trivial relationship has been added only in case of failure of all previous queries with an incremental number of middle entities and relationships.

The implementation of this *Query Relaxation* procedure has been necessary to test the previously stated assumption that *entities connected by one trivial relationship are likely to be linked by many non-trivial relationships as well*. Preliminary results shown in Table A1 (see Appendix) confirm it only partially.

5.1. Discussion of results

The performed experiment shows that the most efficient procedure to solve the task of *shortest path* and *relationships retrieval* is the combination of Algorithms 1 and 2, together with *Query 1* for finding non-trivial relationships. The latter can be exploited to integrate the last proposed solution as well, which is showing promising results. Since they reflect patterns occurring in real web navigation, they are considered less suitable to match one of the core rationale of the present paper, which is the unearthing of latent knowledge in databases for narrative and educational purposes.

Running the relationships finder without admitting trivial relationships has proven to be ineffective. The insertion of at least one *dbo:wikiPageWikiLink* is shown to be sufficient in most cases to achieve the awaited results.

The experiment is expected to be enriched by evaluating more models for training of the KG-Embeddings.

6. Paper's Value Proposition

This pipeline is intended to expand the knowledge of the user and create links between siloed yet related entities.

From a cognitive perspective, the difference between this kind of user experience and usual web surfing is that in HILDEGARD, previously visited nodes remain in a graph, which visually blossoms in front of the user, who is therefore encouraged to travel the already generated nodes and relationships, to progressively fill the gaps between not connected nodes. The web surfer, instead, does not try to fill a gap, but a void, with the result of not retaining the acquired information, which is more prone to fade away [38].

A novel semi-automatic workflow for generating narrative-oriented knowledge graphs for CH is established, featuring heritage objects stored in user-selected cultural sites along with entities, places, and events of broader interest harvested by integration of LOD and web scraping techniques. The workflow culminates in a triples-based dataset and develops through the following tasks:

1) Data Harvesting by ARCO SPARQL endpoint Query (title, description, URI) accessed by Python Application using user input as parameters (Italian region, site/museum. See Figures A1 and A2 in Appendix);
2) Data Import by Neo4J through integrated Python library;
3) Data Visualization through Neo4J;
4) Data scoping by user selection;
5) Extraction of all nodes containing the input entities, including their relationships;
6) Preliminary triangle completion linking descriptions and URIs to related titles;
7) Entity extraction and linking on titles and descriptions by *Wikifier*;

8) Shortest path algorithm performed on entities by web scraping "Six degrees of Wikipedia";
   For two nodes: algorithm performed bidirectionally;
   For more than two nodes: algorithm performed on all couplets obtained by their combination;
9) Extraction of titles, descriptions, and URIs;
10) Mapping triples to selected language through *DBpedia* integration;
11) Relationship finding;
12) Integration of the obtained KG in the original one;
13) Ontology harmonization.

This tool addresses different knowledge gaps and needs simultaneously, such as the lack of solutions tailored for narrative purposes in the cultural heritage domain, that is, to be used in a scenario where objects belonging to the same room must be linked through a narrative, which shall not only be coherent and informative but also engaging and interesting. For the first time, a clear solution to the problem of non-trivial relationship retrieval has been proposed. HILDEGARD fosters interoperability, featuring entities deriving from heterogeneous sources connected through a generalized version of a domain-specific ontology. Even without using the function of domain-specific knowledge, HILDEGARD can represent a friendly choice for creating relationships among the entities according to the CIDOC-CRM schema [33]. Conversely, the integration with CH knowledge bases can be skipped, to exploit the "Six-Degrees-of-Wikipedia" wrapper and the relation labeling tool, to quickly create personalized, interest-based, interactive KGs (see the Colab notebook contained in the project's Github repository). The so-created KG (see Figure 6) can be then exported according to the LOD best practices in JSON (see Figure A3 in Appendix) or RDF/Turtle format, to contribute to the digital heritage data cloud expansion.

HILDEGARD distinguishes itself through its ability to balance relevance with unexpectedness, thereby fostering heightened user engagement. This unique approach not only enhances the user experience but also opens new avenues for discovery within CH collections. Moreover, HILDEGARD can be marketed as a lightweight aggregator of functionalities that are already present in other tools but have drifted into obsolescence due to lack of maintenance. By revitalizing and integrating these existing features into a streamlined, user-friendly interface, HILDEGARD breathes new life into valuable but neglected DH tools. This strategy not only maximizes the utility of pre-existing resources but also offers a cost-effective solution for institutions seeking to modernize their digital infrastructure without the need for extensive new development. In essence, HILDEGARD serves as a bridge between legacy systems and contemporary needs, ensuring that valuable functionalities are not lost to time but are instead repurposed and optimized for current and future use in the DH sector.

7. Limitations and Future Work

The ambition and the broadness of the presented project leave much room for improvement and still present challenges with no immediate solution.

In the following, a more detailed roadmap for the next integrations is provided.

7.1. User experience and visualization

The prototype, already able to generate the graph required for the following step of the proposed ASG pipeline, is yet to be graphically enhanced so that the final user may interact directly with its nodes and links rather than by typing the input via the terminal. Preparatory to this step is the enhancement of flexibility by introducing more input parameters (for instance, specifying what to harvest from the knowledge bases) or by prototyping also the user interface by means of off-the-shelf tools like *Streamlit*, before actualizing the definitive web implementation by Flask, thus allowing to collect feedback from the real end users, which in turn might be used as a training dataset for the neural generation of relations labels.

At the current state, the implementation through a combination of *Python* with *Neo4J* may seem like an unnecessary sophistication. Although it is not required for the final user to have a proficient command of the CIPHER query language, he is expected to initialize the database at the very beginning, which may result in being overwhelming. The reason behind this choice relies on the possibility to exploit Neo4J's pleasant graph visualization in a web application to enhance UX (as well as cutting costs for front-end development) and to be able to exploit all useful Neo4J native functions for KG analysis and manipulation (thus reducing back-end development costs). In alternative to the leverage of users' interaction logs, their overall experience can be assessed through human evaluation. An attempt to exploit museal online 3D, as showcased in Google Arts & Culture[5], has also been unsuccessfully performed, as the description of heritage objects resulted illegible because of photography resolution.

7.1.1. Experimental setting for UX

To this extent, expert and non-/semi-expert users shall be distinguished. For both, the application experience shall be evaluated according to the standards of user experience and usefulness. In turn, UX is branched down in practicality and Aesthetics, while usefulness in relevance of interaction and output.

Since the new relationships have extremely general labels, assessing the correctness of the harvested triples has been required as superfluous for the moment.

Experts, as active users of DH tools reported in the literature review, will be contacted and asked to score HILDEGARD according to the above mentioned metrics. Specific questions will address the comparison between HILDEGARD and the tool they principally use to work with, in case of overlapping goals.

On the other hand, non-expert users will be asked to evaluate HILDEGARD against other interactive KGs, such as BabelNet, LodMilla, and Deutsches Museum (see Table 2).

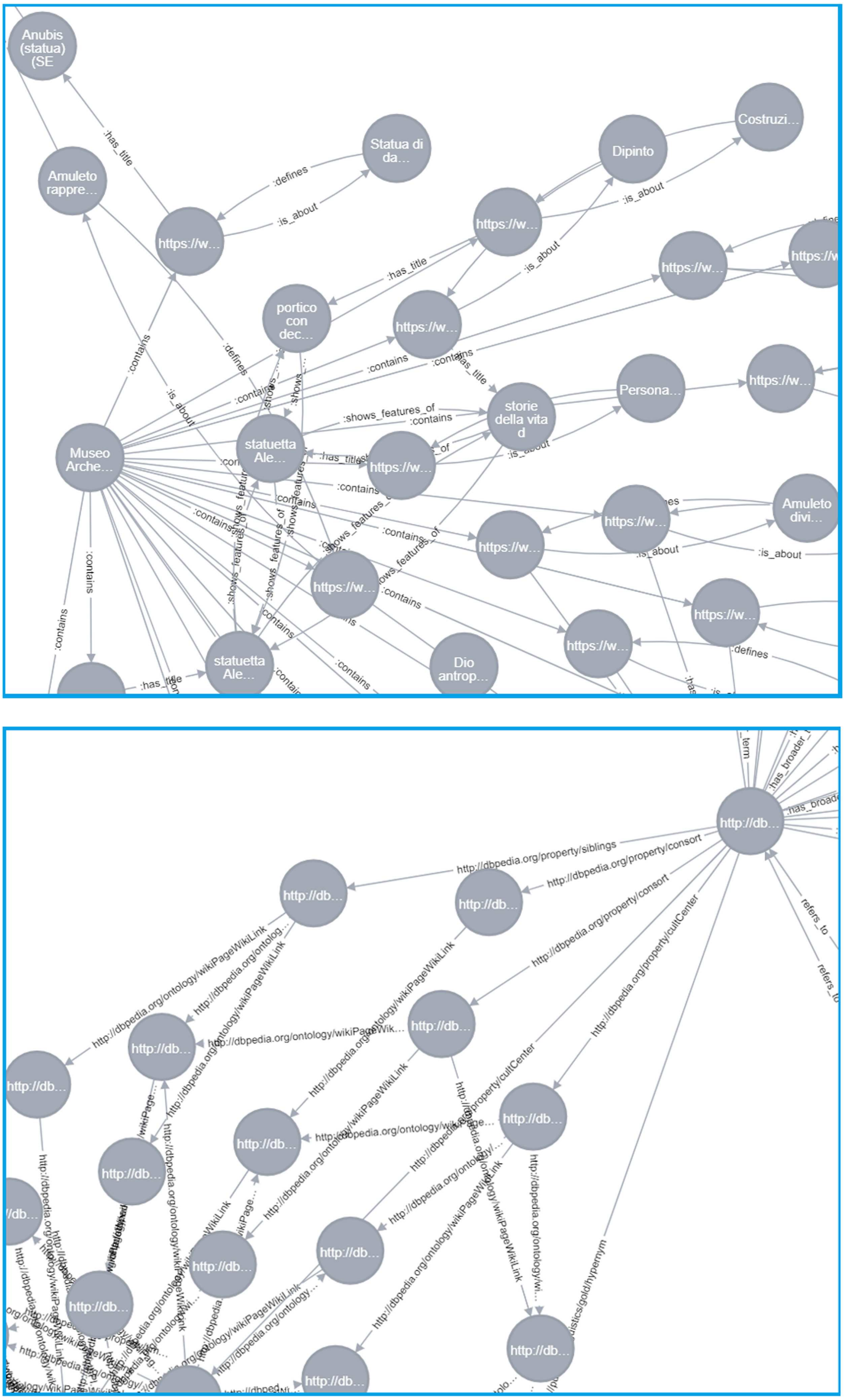
7.2. Optimization

The rationale of reusing available resources and tools does not only represent a choice of convenience but also reflects the LOD principles. Nevertheless, the final version of this tool is envisioned as a self-standing application, requiring no external dependencies. For instance, recently developed solutions such as *kuzú*[6] allow CIPHER-querying on uploaded knowledge bases without the necessity of installing Neo4j. Furthermore, the shortest back end needs to be run offline on Wikipedia- or DBpedia-dumps, which are regularly maintained and updated, because relying on an external website such as "Six Degrees of Wikipedia" (refer to Figure A4 in Appendix) cannot ensure the same continuity, stability, and performance of a native, locally stored tool. This principle is valid also for

[5] https://artsandculture.google.com/partner/national-archaeological-museum-of-naples

[6] https://kuzudb.com/

Figure 6
A partial view from the Neo4J graph visualizer of the database selected by the user through the Python terminal



more stable and fully exploitable online resources, such as the path algorithm of the above mentioned KGs.

7.3. Knowledge expansion

The selected CIDOC-CRM classes and properties will be further expanded in the upcoming modules of the general pipeline (see Figure 1). The related XML-formatted schema lists logic inference rules for every class and property: this feature can be exploited to further expand the KG. To this extent, a meta-implementation of the CIDOC Conceptual Reference Model in Neo4j may be necessary as well.

Finally, it is still necessary to harness this application with as many national and international knowledge bases for CH as possible since the full potential of the presented tool relies on the multilinguality and the heterogeneity of the sources.

In this case, knowledge silos, diverging conceptualizations, and incorrect annotations posit a pressing stumbling block for harmonization, which can nowadays be tackled by means of neural methods, such as the integration with controlled synthetic data [67].

It shall not be underestimated that currently digitized heritage objects and exhibitions very often lack important data. As an example of the gravity of this issue, it suffices to mention that the most important heritage objects of the museum used in the case study were not enlisted, whereas they could be easily spotted by a natural language query executed in the Claude 3.5 large language model. They include:

1) The Farnese Hercules
2) The Farnese Atlas
3) The Alexander Mosaic from the House of the Faun in Pompeii
4) The Toro Farnese (Farnese Bull) sculpture group
5) The Venus Callipyge statue
6) The Artemis of Ephesus statue
7) The Harmodius and Aristogeiton sculpture and plenty of other important artifacts.

Even the task of insights generation can be further enhanced by introducing an element of artificial intelligence, with the capability to fetch textual information stored in the entity-related Wikipedia page. This would allow, among others, the establishment of more precise relationship labels by means of relation extraction [68], thus contributing to the bridging of the Web of Documents to the Web of Data.

The neuro-symbolic integration of image analysis and structured data querying opens new possibilities for advanced information retrieval. When feature vectors are extracted from images, the visual data can be combined with associated metadata like geographical coordinates, categories, or related Wikipedia pages. This combination allows for sophisticated queries that merge traditional database searches with image-based similarity assessments [69], creating a hybrid system that leverages both KG embeddings and formal query languages like SPARQL.

7.4. Further application scenarios

Although this step of world building is propaedeutic to the next module of the pipeline, which will culminate in the generation of a text, it is argued that it already represents an immediately exploitable educational tool, allowing the user to explore culture interactively, hence more efficiently.

Moreover, the locally generated Linked Data can be stored, upon user's permission, in an RDF management system such as Virtuoso and Apache Jena Fuseki, becoming freely accessible to other users by means of a SPARQL-query endpoint. In this way, the produced data will respect all FAIR principles [70]: findability, accessibility, interoperability, and reusability. Filling gaps between entities represents also a powerful exercise that boosts memorization of information [71], a factor that is independent from the specific domain of the co-created KG. Therefore, the application scenarios in education result countless and fruitful.

The HILDEGARD system offers data managers a powerful tool for exploring potential enhancements to digitized heritage objects [72]. In the future, the system is expected to autonomously suggest serendipitous connections within a museum, eliminating the need for manual keyword input from operators. This capability could facilitate the creation of innovative exhibitions that draw objects from different rooms within the same museum [73], fostering unexpected connections between seemingly disparate artifacts. Such intelligent curation could not only enhance the educational value of exhibitions but also potentially increase visitor engagement and museum revenue.

Acknowledgements

This work could not be carried out without the support of several scholars and institutions. Therefore, I thank Dr. Maria Pia Di Buono, Dr. Raffaele Manna, and Prof. Dr. Johanna Monti for their criticism and suggestions, especially concerning the pipeline design and the SPARQL endpoint setting. I also thank Prof. Dr. Alessandro Codello, Prof. Dr. Carlo Strapparava, Prof. Dr. Marco Lippi and Prof. Dr. Armando Stellato for their mentoring. I acknowledge the constructive remarks from Prof. Dr. Frank van Harmelen and Prof. Dr. Wendy Mackay, particularly considering the user experience and human–machine interaction aspect of this project, and Dr. Victor de Boer, regarding practical issues in LOD management. The 5th Summer Datathon on Linguistic Linked Open Data (SD-LLOD-23), organized by NexusLinguarum and the European Network for Web-centred linguistic data science (COST-action 18209), has been fundamental for introducing me to some of the necessary tools to technically implement the application. Finally, I thank the C2DH department of the University of Luxembourg for letting me use time from my research visit there for writing down this paper, as well for presenting me the possibility to integrate the hereby presented tool in the DHARPA's data harmonization suite for Digital Humanities kiara, thus fruitfully paving the way for future adjustments and refactoring. Special thanks to this extent go to Markus Birnstein, Dr. Muriel van Ruymbeke, Florentina Armaselu, and Prof. Dr. Sean Takats.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by the author.

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in GitHub at https://github.com/Glottocrisio/HILDEGARD.

Author Contribution Statement

Cosimo Palma: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration.

References

[1] O'Neill, B., & Stapleton, L. (2022). Digital cultural heritage standards: From silo to semantic web. *AI & Society*, *37*(3), 891–903. https://doi.org/10.1007/s00146-021-01371-1
[2] Mulholland, P., & Collins, T. (2002). Using digital narratives to support the collaborative learning and exploration of cultural heritage. In *13th International Workshop on Database and Expert Systems Applications*, 527–531. https://doi.org/10.1109/DEXA.2002.1045951
[3] Mulholland, P., Wolff, A., & Collins, T. (2012). Curate and storyspace: An ontology and web-based environment for describing curatorial narratives. In *The Semantic Web: Research and Applications: 9th Extended Semantic Web Conference,* 748–762. https://doi.org/10.1007/978-3-642-30284-8_57
[4] Collins, T., Mulholland, P., Bradbury, D., & Zdrahal, Z. (2003). Methodology and tools to support storytelling in cultural heritage forums. In *14th International Workshop on Database and Expert Systems Applications*, 105–109. https://doi.org/10.1109/DEXA.2003.1232006
[5] Kaldeli, E., Menis-Mastromichalakis, O., Bekiaris, S., Ralli, M., Tzouvaras, V., & Stamou, G. (2021). CrowdHeritage: Crowdsourcing for improving the quality of cultural heritage metadata. *Information*, *12*(2), 64. https://doi.org/10.3390/info12020064
[6] Shaikh, A., & KulkarnilI, S. B. (2020). Natural language processing applications for tourism sector. *Journal of Computer Engineering*, *22*, 27–35.
[7] Hellou, M., Lim, J., Gasteiger, N., Jang, M., & Ahn, H. S. (2022). Technical methods for social robots in museum settings: An overview of the literature. *International Journal of Social Robotics*, *14*(8), 1767–1786. https://doi.org/10.1007/s12369-022-00904-y
[8] Mucha, F. (2022). *Co-creative events for engagement with digital cultural heritage collections*. PhD Thesis, University of Glasgow. https://doi.org/10.5525/gla.thesis.82754
[9] Subramanian, K. R. (2018). Myth and mystery of shrinking attention span. *International Journal of Trend in Research and Development*, *5*(3), 1–6.
[10] van Hage, W. R., Stash, N., Wang, Y., & Aroyo, L. (2010). Finding your way through the Rijksmuseum with an adaptive mobile museum guide. In *The Semantic Web: Research and Applications: 7th Extended Semantic Web Conference,* 46–59. https://doi.org/10.1007/978-3-642-13486-9_4
[11] Vrettakis, E., Kourtis, V., Katifori, A., Karvounis, M., Lougiakis, C., & Ioannidis, Y. (2019). Narralive – Creating and experiencing mobile digital storytelling in cultural heritage. *Digital Applications in Archaeology and Cultural Heritage*, *15*, e00114. https://doi.org/10.1016/j.daach.2019.e00114
[12] Valtolina, S. (2016). A storytelling-driven framework for cultural heritage dissemination. *Data Science and Engineering*, *1*(2), 114–123. https://doi.org/10.1007/s41019-016-0007-z
[13] Malegiannaki, I. A., Daradoumis, T., & Retalis, S. (2020). Teaching cultural heritage through a narrative-based game. *Journal on Computing and Cultural Heritage*, *13*(4), 1–28. https://doi.org/10.1145/3414833
[14] Bran, E., Bautu, E., Popovici, D. M., Braga, V., & Cojuhari, I. (2019). Cultural heritage interactive dissemination through natural interaction. In *Proceedings of RoCHI*.
[15] Maietti, F., Medici, M., Ferrari, F., Ziri, A. E., & Bonsma, P. (2018). Digital cultural heritage: Semantic enrichment and modelling in BIM environment. In *Digital Cultural Heritage: Final Conference of the Marie Skłodowska-Curie Initial Training Network for Digital Cultural Heritage*, 104–118. https://doi.org/10.1007/978-3-319-75826-8_9
[16] Katifori, A., Karvouni, S. M., Kourtis, V., Perry, S., Roussou, M., & Ioanidis, Y. (2018). Applying interactive storytelling in cultural heritage: Opportunities, challenges and lessons learned. In *Interactive Storytelling: 11th International Conference on Interactive Digital Storytelling*, 603–612. https://doi.org/10.1007/978-3-030-04028-4_70
[17] Da Milano, C., Falchetti, E., Migone, P., & Nisi, V. (2023). Digital storytelling, cultural heritage, and social inclusion: The MEMEX project. In D. Giglitto, L. Ciolfi, E. Lockley & E. Kaldeli (Eds.), *Digital approaches to inclusion and participation in cultural heritage: Insights from research and practice in Europe* (pp. 19–24). Routledge.
[18] Dahroug, A., Vlachidis, A., Liapis, A., Bikakis, A., López-Nores, M., Sacco, O., & Pazos-Arias, J. J. (2021). Using dates as contextual information for personalised cultural heritage experiences. *Journal of Information Science*, *47*(1), 82–100. https://doi.org/10.1177/0165551519871823
[19] Bolioli, A., Bosca, A., Damiano, R., Lieto, A., & Striani, M. (2022). A complementary account to emotion extraction and classification in cultural heritage based on Plutchik's theory. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, 374–382. https://doi.org/10.1145/3511047.3537659
[20] Hu, Y., Chapman, A., Wen, G., & Hall, D. W. (2022). What can knowledge bring to machine learning? A survey of low-shot learning for structured data. *ACM Transactions on Intelligent Systems and Technology*, *13*(3), 1–45. https://doi.org/10.1145/3510030
[21] Diaz-Rodriguez, N., Lamas, A., Sanchez, J., Franchi, G., Donadello, I., Tabik, S., ..., & Herrera, F. (2022). EXplainable Neural-symbolic learning (X-NeSyL) methodology to fuse deep learning representations with expert knowledge graphs: The MonuMAI cultural heritage use case. *Information Fusion*, *79*, 58–83. https://doi.org/10.1016/j.inffus.2021.09.022
[22] LaBouve, E., Miller, E., & Khosmood, F. (2019). Enhancing story generation with the semantic web. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, 1–11. https://doi.org/10.1145/3337722.3337742
[23] Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., de Melo, G., Gutierrez, C., . . . , & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*, *54*(4), 71. https://doi.org/10.1145/3447772
[24] Silva, A. L., & Terra, A. L. (2024). Cultural heritage on the semantic web: The Europeana data model. *IFLA Journal*, *50*(1), 93–107. https://doi.org/10.1177/03400352231202506
[25] Bikakis, A., Hyvonen, E., Jean, S., Markhoff, B., & Mosca, A. (2021). Editorial: Special issue on semantic web for cultural heritage. *Semantic Web*, *12*(2), 163–167. https://doi.org/10.3233/SW-210425
[26] Divya Venkatesh, J., Jaiswal, A., & Nanda, G. (2024). Comparing human text classification performance and explainability with large language and machine learning models using

eye-tracking. *Scientific Reports*, *14*(1), 14295. https://doi.org/10.1038/s41598-024-65080-7

[27] Palma, C. (2023). Neurosymbolic narrative generation for cultural heritage. In *Proceedings of the Second International Conference on Hybrid Human-Artificial Intelligence,* 509–523. https://doi.org/10.3233/FAIA230129

[28] Geurts, J., Bocconi, S., van Ossenbruggen, J., & Hardman, L. (2003). Towards ontology-driven discourse: From semantic graphs to multimedia presentations. In *The Semantic Web – ISWC 2003: Second International Semantic Web Conference,* 597–612. https://doi.org/10.1007/978-3-540-39718-2_38

[29] Rashkin, H., Celikyilmaz, A., Choi, Y., & Gao, J. (2020). Plotmachines: Outline-conditioned generation with dynamic plot state tracking. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP),* 4274–4295. https://doi.org/10.18653/v1/2020.emnlp-main.349

[30] Wang, Z., Zhang, X., & Du, H. (2021). Building the directed semantic graph for coherent long text generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing,* 2563–2572. https://doi.org/10.18653/v1/2021.emnlp-main.200

[31] Alabdulkarim, A., Li, S., & Peng, X. (2021). Automatic story generation: Challenges and attempts. In *Proceedings of the Third Workshop on Narrative Understanding*, 72–83. https://doi.org/10.18653/v1/2021.nuse-1.8

[32] Palma, C. (2024). Modelling interestingness: A workflow for surprisal-based knowledge mining in narrative semantic networks. In *Semantic Methods for Events and Stories Co-located with 21st Extended Semantic Web Conference,* 1–9.

[33] El-Hajj, H., & Valleriani, M. (2021). CIDOC2VEC: Extracting information from atomized CIDOC-CRM humanities knowledge graphs. *Information*, *12*(12), 503. https://doi.org/10.3390/info12120503

[34] Kotkov, D., Medlar, A., Kask, T., & Głowacka, D. (2024). The dark matter of serendipity in recommender systems. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval,* 108–118. https://doi.org/10.1145/3627508.3638342

[35] Bartalesi, V., Coro, G., Lenzi, E., Pagano, P., & Pratelli, N. (2023). From unstructured texts to semantic story maps. *International Journal of Digital Earth*, *16*(1), 234–250. https://doi.org/10.1080/17538947.2023.2168774

[36] Porzel, R., Pomarlan, M., Spillner, L., Bateman, J., Mildner, T., & Santagiustina, C. (2022). Narrativizing knowledge graphs. In *International Workshop on Knowledge Graph Summarization,* 100–111.

[37] Salas, E. L. (2021). A collection of narrative practices on cultural heritage with innovative technologies and creative strategies. *Open Research Europe*, *1*, 130. https://doi.org/10.12688%2Fopenreseurope.14178.1

[38] Robin, B. R. (2016). The power of digital storytelling to support teaching and learning. *Digital Education Review*, *30*, 17–29.

[39] Eklund, P., Wray, T., & Ducrou, J. (2011). Linking objects and their stories: An API for exploring cultural heritage using formal concept analysis. *Journal of Emerging Technologies in Web Intelligence*, *3*(3), 239–252. https://doi.org/10.4304/jetwi.3.3.239-252

[40] Lenci, A., & Sahlgren, M. (2023). *Distributional semantics*. UK: Cambridge University Press.

[41] Chatzakis, M., Mountantonakis, M., & Tzitzikas, Y. (2021). RDFsim: Similarity-based browsing over DBpedia using embeddings. *Information*, *12*(11), 440. https://doi.org/10.3390/info12110440

[42] Huang, Y. Y., Yu, S. S., Chu, J. J., Fan, H. H., & Du, B. B. (2023). Using knowledge graphs and deep learning algorithms to enhance digital cultural heritage management. *Heritage Science*, *11*(1), 204. https://doi.org/10.1186/s40494-023-01042-y

[43] Schröder, M., Jilek, C., & Dengel, A. (2022). A Human-in-the-loop approach for personal knowledge graph construction from file names. In *Proceedings of the 3rd International Workshop on Knowledge Graph Construction*, 1–15.

[44] Wang, M., Qiu, L., & Wang, X. (2021). A survey on knowledge graph embeddings for link prediction. *Symmetry*, *13*(3), 485. https://doi.org/10.3390/sym13030485

[45] Paranyushkin, D. (2019). InfraNodus: Generating insight using text network analysis. In *The World Wide Web Conference,* 3584–3589. https://doi.org/10.1145/3308558.3314123

[46] Viola, L. (2022). DeXTER: A post-authentic approach to heritage visualisation. In *Digital Humanities 2022 Conference Abstracts,* 547–554.

[47] Oldman, D. (2021). Digital research, the legacy of form and structure and the ResearchSpace system. In K. Golub & Y. H. Liu (Eds.), *Information and knowledge organisation in digital humanities* (pp. 131–153). Routledge.

[48] Hyvönen, E., & Rantala, H. (2019). Knowledge-based relation discovery in cultural heritage knowledge graphs. In *Proceedings of the Digital Humanities in the Nordic Countries: 4th Conference,* 230–239.

[49] Metilli, D., Bartalesi, V., & Meghini, C. (2019). A Wikidata-based tool for building and visualising narratives. *International Journal on Digital Libraries*, *20*(4), 417–432. https://doi.org/10.1007/s00799-019-00266-3

[50] Sun, J., Zhang, Z., & He, X. (2024). LLM4EduKG: LLM for automatic construction of educational knowledge graph. In *2024 International Conference on Networking and Network Applications (NaNA)*, 269–275. https://doi.org/10.1109/NaNA63151.2024.00051

[51] Schreiber, G., Amin, A., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M, . . . , & Wielinga, B. (2006). Multimedian E-culture demonstrator. In *The Semantic Web – ISWC 2006: 5th International Semantic Web Conference,* 951–958. https://doi.org/10.1007/11926078_70

[52] Dimoulas, C. A. (2022). Cultural heritage storytelling, engagement and management in the era of big data and the semantic web. *Sustainability*, *14*(2), 812. https://doi.org/10.3390/su14020812

[53] Candela, G., Escobar, P., Carrasco, R. C., & Marco-Such, M. (2019). A linked open data framework to enhance the discoverability and impact of culture heritage. *Journal of Information Science*, *45*(6), 756–766. https://doi.org/10.1177/0165551518812658

[54] Europeana. (2017). *Definition of the Europeana Data Model v5.2.8*. Retrieved from: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf

[55] Doerr, M., Bruseker, G., Bekiari, C., Ore, C. E., Velios, T., & Stead, S. (2020). *Volume A: Definition of the CIDOC Conceptual Reference Model*. Retrieved from: https://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM_v6.2.9%2030-4-2020%20.pdf

[56] Tzitzikas, Y., Mountantonakis, M., Fafalios, P., & Marketakis, Y. (2022). CIDOC-CRM and machine learning: A survey and

future research. *Heritage*, *5*(3), 1612–1636. https://doi.org/10.3390/heritage5030084
[57] Navigli, R., Bevilacqua, M., Conia, S., Montagnini, D., & Cecconi, F. (2021). Ten years of BabelNet: A survey. In *International Joint Conference on Artificial Intelligence,* 4559–4567.
[58] Heim, P., Hellmann, S., Lehmann, J., Lohmann, S., & Stegemann, T. (2009). RelFinder: Revealing relationships in RDF knowledge bases. In *Semantic Multimedia: 4th International Conference on Semantic and Digital Media Technologies,* 182–187. https://doi.org/10.1007/978-3-642-10543-2_21
[59] Carriero, V. A., Gangemi, A., Mancinelli, M. L., Marinucci, L., Nuzzolese, A. G., Presutti, V., & Veninata, C. (2019). ArCo ontology network and LOD on Italian cultural heritage. In *Proceedings of the CEUR Workshop,* 97–102.
[60] Carriero, V. A., Gangemi, A., Mancinelli, M. L., Nuzzolese, A. G., Presutti, V., & Veninata, C. (2021). Pattern-based design applied to cultural heritage knowledge graphs. *Semantic Web*, *12*(2), 313–357. https://doi.org/10.3233/SW-200422
[61] Martin, J. (1983). *Managing the data-base environment*. USA: Prentice-Hall.
[62] Dijkstra, E. W. (2022). A note on two problems in connexion with graphs. In K. R. Apt & T. Hoare (Eds.), *Edsger Wybe Dijkstra: His life, work, and legacy* (pp. 287–290). Association for Computing Machinery. https://doi.org/10.1145/3544585.3544600
[63] Floyd, R. W. (1962). Algorithm 97: Shortest path. *Communications of the ACM*, *5*(6), 345–345.
[64] Färber, M., Bartscherer, F., Menne, C., & Rettinger, A. (2017). Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web*, *9*(1), 77–129. https://doi.org/10.3233/SW-170275
[65] Brank, J., Leban, G., & Grobelnik, M. (2017). Annotating documents with relevant Wikipedia concepts. In *Proceedings of SiKDD*, *472*.
[66] Biswas, R., Kaffee, L. A., Cochez, M., Dumbrava, S., Jendal, T. E., Lissandrini, M., . . . , & de Melo, G. (2023). Knowledge graph embeddings: Open challenges and opportunities. *Transactions on Graph Data and Knowledge*, *1*(1), 4:1–4:32. https://dx.doi.org/10.4230/TGDK.1.1.4
[67] Rei, L., Mladenic, D., Dorozynski, M., Rottensteiner, F., Schleider, T., Troncy, R., . . . , & Salvatella, M. G. (2023). Multimodal metadata assignment for cultural heritage artifacts. *Multimedia Systems*, *29*(2), 847–869. https://doi.org/10.1007/s00530-022-01025-2
[68] Cabot, P. L. H., & Navigli, R. (2021). REBEL: Relation extraction by end-to-end language generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021,* 2370–2381. https://doi.org/10.18653/v1/2021.findings-emnlp.204
[69] Kulmanov, M., Kafkas, S., Karwath, A., Malic, A., Gkoutos, G. V., Dumontier, M., & Hoehndorf, R. (2018). Vec2SPARQL: Integrating SPARQL queries and knowledge graph embeddings. *CEUR Workshop Proceedings*.
[70] GO FAIR. (n.d.). *FAIR principles*, Retrieved from: https://www.go-fair.org/fair-principles/
[71] Cioca, L. I., & Nerişanu, R. A. (2020). Enhancing creativity: Using visual mnemonic devices in the teaching process in order to develop creativity in students. *Sustainability*, *12*(5), 1985. https://doi.org/10.3390/su12051985
[72] Freire, N., Meijers, E., de Valk, S., Voorburg, R., Isaac, A., & Cornelissen, R. (2018). Aggregation of linked data: A case study in the cultural heritage domain. In *IEEE International Conference on Big Data,* 522–527. https://doi.org/10.1109/BigData.2018.8622348
[73] Wolff, A., Mulholland, P., & Collins, T. (2012). Storyspace: A story-driven approach for creating museum narratives. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media,* 89–98. https://doi.org/10.1145/2309996.2310012

How to Cite: Palma, C. (2025). HILDEGARD: Human-in-the-Loop Data Extraction and Graphically Augmented Relation Discovery. *Journal of Computational and Cognitive Engineering*, *4*(1), 64–82. https://doi.org/10.47852/bonviewJCCE42022924

Appendix

Figure A1
A snippet of the main SPARQL-query, used to retrieve heritage objects, represented by their URIs, titles, and descriptions

```
query = """
PREFIX arco-arco: <https://w3id.org/arco/ontology/arco/>
PREFIX arco-arco: <https://w3id.org/arco/ontology/arco/>
PREFIX arco-location: <https://w3id.org/arco/ontology/location/>
PREFIX CLV: <https://w3id.org/italia/onto/CLV/>
PREFIX arco-dd: <https://w3id.org/arco/ontology/denotative-description/>
PREFIX cov: <https://w3id.org/italia/onto/COV/>
PREFIX arco-core: <https://w3id.org/arco/ontology/core/>
PREFIX datigov: <http://dati.gov.it/onto/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX io:  <https://w3id.org/italia/onto/l0/>

SELECT DISTINCT ?cultobj ?title ?museo ?descr
WHERE {
{?cultobj arco-arco:hasCataloguingAgency  <""" + str(catalogagency) + """>.
        <""" + str(catalogagency) + """>  rdfs:label  ?museo.
    ?cultobj   dc:description ?descr.
    ?cultobj  rdfs:label ?title}
    union
    {?cultobj  arco-location:hasCulturalInstituteOrSite <""" + str(cultinst) + """>.
    <""" + str(cultinst) + """> rdfs:label  ?museo.
    ?cultobj   dc:description ?descr.
    ?cultobj  rdfs:label ?title }
FILTER(LANG(?title) = '"""+ lang +"""')
FILTER("""

i = 0
while i < len(obj_list):
    query = query + f"""CONTAINS(?title, "{obj_list[i]}") || CONTAINS(?descr, "{obj_list[i]}")"""
    if i + 1 != len(obj_list):
        query = query + """ || """
    i = i + 1
query = query + """)
}"""
sparql.setQuery(query)
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
```

Figure A2
A snippet of the result from the previous query HTML format, with “National Archeological Museum of Naples” as input parameter

https://w3id.org/arco/resource/HistoricOrArtisticProperty/1500572901	"pavimento in "opus sectile" della Casa del Fauno a Pompei (disegno) di Veneri Pasquale Maria (secondo quarto sec. XIX)"@it	"Museo Archeologico Nazionale di Napoli"	"L'acquerello raffigura le decorazioni pavimentali del tablino e dell'impluvio della Casa del Fauno a Pompei. A sinistra è raffigurata la decorazione dell'impluvio, composta da rombi verdi, i cui spazi di risulta, anch'essi romboidali, sono riempiti da nove piccoli rombi policromi; sotto la decorazione sono riportati i colori dei materiali utilizzati. A destra è raffigurata la fascia mosaico a quadrati concentrici e meandri doppi prospettici, sormontata da una porzione angolare di decorazione con cubi prospettici, bordata da sette listelli policromi"
https://w3id.org/arco/resource/HistoricOrArtisticProperty/1500572906	"battaglia di Alessandro, battaglia di Alessandro contro Dario (disegno) - ambito napoletano (secondo quarto sec. XIX)"@it	"Museo Archeologico Nazionale di Napoli"	"Il disegno raffgura la battaglia di Alessandro Magno contro Dario del noto mosaico proveniente dalla Casa del Fauno a Pompei. In primo piano sulla sinistra è raffigurato Alessandro Magno a cavallo, munito di corazza con "gorgoneion" che, nell'intento di raggiungere Dario che si dà alla fuga, trafigge con la lancia un cavaliere che alza il braccio sinistro sopra la testa mentre con la destra afferra l'asta che lo ha colpito. Sullo sfondo si scoge un platano senza fronde ed uno stuolo di soldati macedoni all'attacco. In basso a destra un altro persiano, caduto da cavallo, sta per essere schiacciato sotto la ruota del carro di Dario. Questi, con un copricapo di stoffa sul capo, compare in primo piano, quasi al centro della scena, in piedi sul suo carro; dietro di lui, sullo sfondo, si spiega l'esercito persiano colto in un movimento a vortice, enfatizzato dall'auriga in primo piano, che ha invertito la rotta del carro. La scena è inquadrata da una cornice a dentelli. Il disegnatore ha integrato le parti lacunose del mosaico"
https://w3id.org/arco/resource/HistoricOrArtisticProperty/1500572911	"vasellame potorio (disegno) di Abbate Giuseppe (secondo quarto sec. XIX)"@it	"Museo Archeologico Nazionale di Napoli"	"Il disegno raffigura un gruppo di vasi potori, probabilmente in argento a giudicare dalle forme, disposti in maniera ordinata su una mensa intorno ad una statuetta di Atena, con elmo e civetta nella destra. Tra i recipienti si riconoscono "skyphoi", "kântharoi", "simpula". In basso, verso destra, è raffigurato un volatile"
https://w3id.org/arco/resource/HistoricOrArtisticProperty/1500572916	"coppia con serpente (disegno) di Abbate Giuseppe (secondo quarto sec. XIX)"@it	"Museo Archeologico Nazionale di Napoli"	"Il disegno raffigura in una cornice circolare un giovane, stante, nudo, appoggiato ad un pilastro, con manto avvolto intorno al braccio sinistro che si volge verso una fanciulla, nuda nella parte superiore del corpo, che siede in atteggiamento pensoso. Tra i due si inserisce un serpente"

Figure A3
A snippet of the output knowledge graph in JSON format


```
[{"title": "Anthropomorphism", "cidoc-relation": "P104", "descr": "Attribution of human traits, emotions, or intentions to non-human entities"}, {"descr": "Attribution of human traits, emotions, or intentions to non-human entities", "cidoc-relation": "P196", "uri": "https://en.wikipedia.org/wiki/Anthropomorphism"}, {"uri": "https://en.wikipedia.org/wiki/Anthropomorphism", "cidoc-relation": "P102", "title": "Anthropomorphism"}, {"descr": "Attribution of human traits, emotions, or intentions to non-human entities", "cidoc-relation": "P196", "title": "Anthropomorphism"}, {"title": "Anthropomorphism", "cidoc-relation": "P104", "descr": "Attribution of human traits, emotions, or intentions to non-human entities"}, {"uri": "https://en.wikipedia.org/wiki/Anthropomorphism", "cidoc-relation": "P67", "descr": "Attribution of human traits, emotions, or intentions to non-human entities"}, {"title": "Anthropomorphism", "cidoc-relation": "P67", "uri": "https://en.wikipedia.org/wiki/Anthropomorphism"}, {"title": "Human-animal hybrid", "cidoc-relation": "P104", "descr": "Entity that incorporates elements from both humans and non-human animals"}, {"descr": "Entity that incorporates elements from both humans and non-human animals", "cidoc-relation": "P196", "uri": "https://en.wikipedia.org/wiki/Human-animal hybrid"}, {"uri": "https://en.wikipedia.org/wiki/Human-animal hybrid", "cidoc-relation": "P102", "title": "Human-animal hybrid"}, {"descr": "Entity that incorporates elements from both humans and non-human animals", "cidoc-relation": "P196", "title": "Human-animal hybrid"}, {"title": "Human-animal hybrid", "cidoc-relation": "P104", "descr": "Entity that incorporates elements from both humans and non-human animals"}, {"uri": "https://en.wikipedia.org/wiki/Human-animal hybrid", "cidoc-relation": "P67", "descr": "Entity that incorporates elements from both humans and non-human animals"}, {"title": "Human-animal hybrid", "cidoc-relation": "P67", "uri": "https://en.wikipedia.org/wiki/Human-animal hybrid"}, {"title": "Anubis", "cidoc-relation": "P104", "descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head"}, {"descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head", "cidoc-relation": "P196", "uri": "https://en.wikipedia.org/wiki/Anubis"}, {"uri": "https://en.wikipedia.org/wiki/Anubis", "cidoc-relation": "P102", "title": "Anubis"}, {"descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head", "cidoc-relation": "P196", "title": "Anubis"}, {"title": "Anubis", "cidoc-relation": "P104", "descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head"}, {"uri": "https://en.wikipedia.org/wiki/Anubis", "cidoc-relation": "P67", "descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head"}, {"title": "Anubis", "cidoc-relation": "P67", "uri": "https://en.wikipedia.org/wiki/Anubis"}, {"prev_title": "Human-animal hybrid", "cidoc-relation": "P67", "title": "Anubis"}, {"title": "Anubis", "cidoc-relation": "P104", "descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head"}, {"descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head", "cidoc-relation": "P196", "uri": "https://en.wikipedia.org/wiki/Anubis"}, {"uri": "https://en.wikipedia.org/wiki/Anubis", "cidoc-relation": "P102", "title": "Anubis"}, {"descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head", "cidoc-relation": "P196", "title": "Anubis"}, {"title": "Anubis", "cidoc-relation": "P104", "descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head"}, {"uri": "https://en.wikipedia.org/wiki/Anubis", "cidoc-relation": "P67", "descr": "Egyptian deity of mummification and the afterlife, usually depicted as a man with a canine head"}, {"title": "Anubis", "cidoc-relation": "P67", "uri": "https://en.wikipedia.org/wiki/Anubis"}, {"title": "Anthropomorphism", "cidoc-relation": "P104", "descr": "Attribution of human traits, emotions, or intentions to non-human entities"}, {"descr": "Attribution of human traits, emotions, or intentions to non-human entities", "cidoc-relation": "P196", "uri": "https://en.wikipedia.org/wiki/Anthropomorphism"}, {"uri": "https://en.wikipedia.org/wiki/Anthropomorphism", "cidoc-relation": "P102", "title": "Anthropomorphism"}, {"descr": "Attribution of human traits, emotions, or intentions to non-human entities", "cidoc-relation": "P196", "title": "Anthropomorphism"}, {"title": "Anthropomorphism", "cidoc-relation": "P104", "descr": "Attribution of human traits, emotions, or intentions to non-human entities"}, {"uri": "https://en.wikipedia.org/wiki/Anthropomorphism", "cidoc-relation": "P67",
```

Figure A4
A snapshot from "Six Degrees of Wikipedia" displaying the individual paths from "Alexander the Great" to "Anubis." Heritage objects related to these entities are conserved in the National Archaeological case-study Naples as well

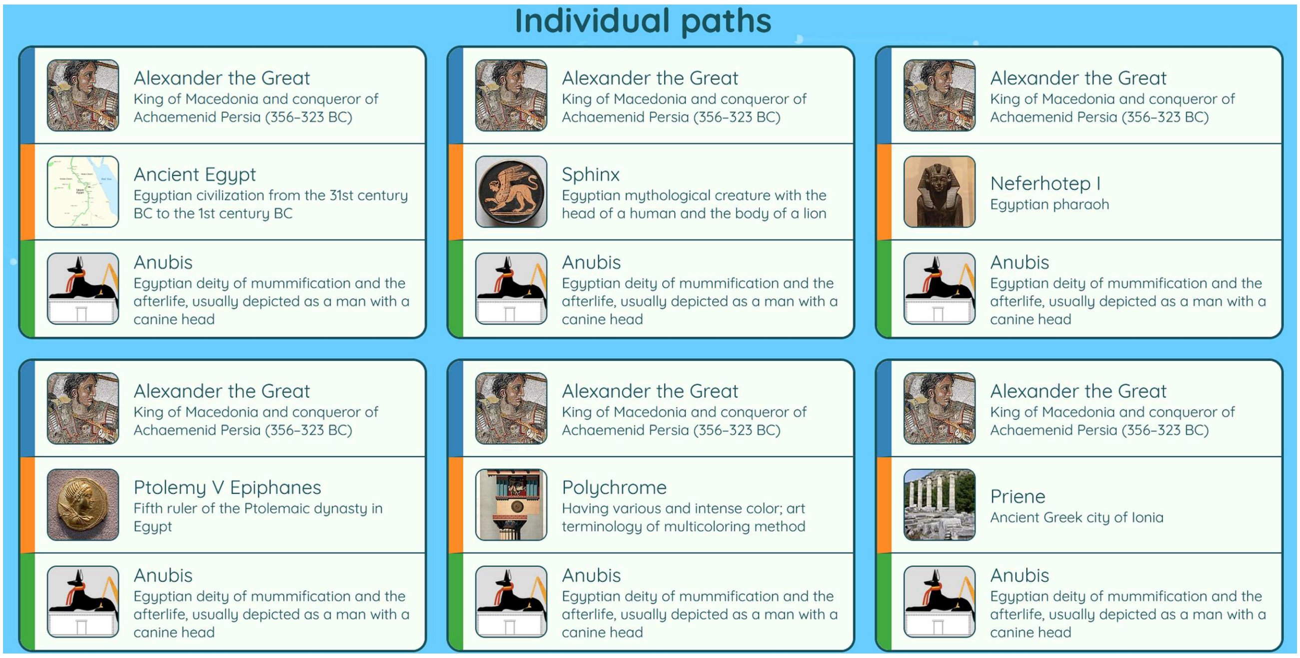


Table A1
A sample of entity pairs with number of results (relationships paths) and iterations (with Query Relaxation). The last column indicates whether trivial relationships are included

Pair	Results	Iteration	Wikipagewikilink
Agathocles of Bactria –[unknown_relation]–> Kamasarye Philoteknos	2	1	No
Alexander the Great –[unknown_relation]–> Agathocles of Bactria	2	1	No
Alexander the Great –[unknown_relation]–> Antimachus II	2	1	No
Antimachus II –[unknown_relation]–> Eumelos of Bosporus	0	5	No
Operating system –[wikiPageWikiLink]–> Dave Cutler	0	5	No
Saint_Cyril - Cyril_of_Alexandria	2	1	Yes
El_Mahalla_El_Kubra - Cyril_of_Alexandria	0	5	No
Robin_Scherbatsky - Barney_Stinson	4	1	No
Eye_of_a_needle - Cyril_of_Alexandria	0	5	No
Filioque - Cyril_of_Alexandria	0	5	No
Claus_Luthe - NSU_Spider	0	5	No
List_of_breakout_characters - Barney_Stinson	0	5	No
Patristics - Cyril_of_Alexandria	0	5	No
Saint_Cyril - Cyril_of_Alexandria	2	1	Yes
Ralph_Macchio - Barney_Stinson	0	5	No
List_of_Copts - Cyril_of_Alexandria	15	4	Yes
Tomb_of_Alexander_the_Great - Cyril_of_Alexandria	0	5	No
Alexandrian_liturgical_rites - Cyril_of_Alexandria	0	5	No
Athanasius_of_Alexandria - Cyril_of_Alexandria	15	4	Yes
Pelagius - Cyril_of_Alexandria	3	4	Yes