An Ontology and Knowledge Graph Infrastructure for Digital Library Knowledge Representation

Stefano Ferilli[1(✉)] and Domenico Redavid[2]

[1] Department of Computer Science, University of Bari, Bari, Italy
stefano.ferilli@uniba.it
[2] R&D Department, Artificial Brain S.r.l., Bari, Italy
redavid@abrain.it

Abstract. New technologies for storing and handling knowledge provide unprecedented opportunities for enhanced fruition of digital libraries and archives. Going beyond document retrieval based on lexical content or metadata, using the context of documents, and/or of their content, may provide very new ways to put them in perspective and grasp a deeper understanding thereof, also for non-technical users.

Several components are needed to support this new perspective: suitable ontological resources to describe such variated knowledge, collaborative tools to collect the precious knowledge scattered across many scholars and practitioners spread all over the world, and to store it in a knowledge base, fruition tools to make the collected knowledge available to all interested stakeholders (scholars, researchers, but also common people).

This paper proposes the GraphBRAIN environment as a possible infrastructure. It is a general-purpose tool that allows its users to design and populate knowledge graphs, to collaboratively enrich them, and to exploit advanced fruition tools, consultation and analysis tools. Its functionality may also be provided as a set of Web services to end-user applications. An initial version of the ontology and knowledge graph for digital libraries and archives are also presented and discussed in the paper.

1 Introduction

While there has been a traditional focus on digital libraries and archives from the collection and consultation perspective, the current availability of new technologies for storing and handling knowledge provides unprecedented opportunities to handle further, high-level functionalities. One such functionality is an enhanced fruition that goes beyond 'simple' document retrieval based on lexical content or on the available metadata. For scholars and practitioners, but also for non-expert end users, a very relevant component, full of interesting information, may be the context of the document as a whole, and/or of its content, that allows to put it in perspective and grasp a deeper understanding thereof. For instance,

M. Ceci et al. (Eds.): IRCDL 2020, CCIS 1177, pp. 47–61, 2020.
https://doi.org/10.1007/978-3-030-39905-4_6

it might be interesting to know that a novel was first cited in a document that was found in a certain place of a certain country, and that a character in that novel was inspired to a real person, who was a friend of the author, who lived in a city, where an event took place that inspired him to write the novel. And, maybe, that the novel was used as the screenplay for a movie, which was shown for the first time in a certain theatre of a famous town, at the presence of some very important public persons. And so on.

Of course, collecting, storing and using such information is not trivial, for several reasons. First, the knowledge to be represented spans through a wide range that goes beyond the typical expertise of researchers and scholars, also involving amateurs, collectors, and enthusiasts. Also, the available knowledge may be scattered and spread across many persons, each perhaps knowing just part of the story, or specialized only on some aspects of it. Moreover, satisfactory usage of the available information might involve complex information patterns and aggregates, that might be domain-dependent and different for the different kinds of users involved. In short, a switch from data bases to knowledge bases is needed, so as to allow a shared understanding of the involved concepts, to improve the reuse of the available information, and to enable reasoning tasks on it that return additional higher-level, non-trivial information. Leveraging the enthusiasm of practitioners in this field, a possible solution would be to adopt a collaborative approach for the building and enrichment of the knowledge base. In a *wiki* approach, the motivation to share knowledge would be the possibility of using also the information contributed by other people.

However, a collaborative approach in which many people, with different expertise, culture, background and perspectives contribute small pieces that together make up the big picture, requires suitable schemes to represent and organize the knowledge in this field. From a traditional database perspective, these schemes are the table definitions. From the knowledge base perspective, these schemes are typically in the form of ontologies. Since this paper aims at merging both perspectives, we need a solution that may serve both as a database schema and as an ontology. Unfortunately, due to the very different traditional approach to data management in digital libraries and archives, the currently available resources, developed in the cultural heritage landscape, are unsuitable. Hence, a need for a new scheme, to be shared and reused by all the stakeholders involved in this area of interest, which is one of the contributions of this paper.

Handling the functionality described above requires an appropriate infrastructure, made up of advanced data representation and storage facilities, as well as of advanced information handling tools and algorithms. This paper proposes a solution based on the Web application **GraphBRAIN**, an on-line tool to collaboratively design, build and maintain knowledge bases. It was used to define a first version of the scheme/ontology describing the contextual information about digital libraries and archives, to serialize it in Web Ontology Language (OWL), and to build a first version of the knowledge graph.

This paper is organized as follows. The next section quickly recalls some related works. Then, after describing the features and interface of GraphBRAIN

in the Sect. 3. Section 4 provides the ontology for the 'contextual' description of digital libraries and archives, and Sect. 5 briefly overviews its current content. Section 6 reports a sample case study and, finally, Sect. 7 concludes the paper and outlines future work issues.

2 Related Work

Concerning the ontology development functionality, several tools have been proposed in the literature, each one with specific targets as regards the construction, editing, annotation and merging of ontologies [1]. Among them, the most popular and mature tool is protégé[1], based on the OWL-API, which is fully compliant with the OWL specifications by W3C[2]. GraphBRAIN adopted the same OWL-API for its ontology export functionality, so that the generated ontologies are fully compliant with the standard and may be edited using protégé. We developed a specific ontology definition and handling tool for several reasons. First, it had to be embedded into GraphBRAIN's interface, so that the administrators could seamlessly and collaboratively build and refine the ontologies. Second, while existing tools are mainly aimed at defining formal ontologies starting from an RDF knowledge base model, our motivation was in the need to define a schema for the graph DB, and the translation in standard ontology format was a consequential objective in order to enable OWL reasoning capabilities.

On the methodological side, some works exist that analyze the possibilities of cooperation between ontologies and graph DBs. In [3] the potential of applying graph DBMSs to an ontological context in order to create essentially an ontological tensor, e.g. the algebraic counterparts of the combinatorial multilayer graphs, is outlined, and its complexity is assessed. [9] discusses technical issues that might limit the impact of symbolic Knowledge Representation on the Knowledge Graph area, and summarizes some developments towards addressing them in various logics.

Another stream of related work is the development of ontologies and/or knowledge graphs. While standard ontologies used for describing resources in the library/archive domain do exist (e.g., the Dublin Core Metadata Initiative, or DCMI [8]), to the best of our knowledge, nothing exists for the specific objective of expanding its area of interest, from the strictly scholar approach to a broader, 'contextual' one that may be attractive also for non-specialized users. However, other resources are available for closely related topics. For instance, focusing on cultural heritage, and on the Italian landscape, **Cultural-ON** (Cultural ONtology) [11]. Very close to our perspective are also [10,14], concerning the development of a database relating movies to the places in which they were shot. Like GraphBRAIN, they adopt a collaborative approach, and aim at describing more than just the formal or technical aspects of filmography, also in a touristic perspective. These initiatives might be connected to GraphBRAIN to

[1] https://protege.stanford.edu.
[2] http://owlcs.github.io/owlapi.

enrich its knowledge base and provide a more effective and varied service to its users.

3 GraphBRAIN

GraphBRAIN[3] is a general-purpose knowledge base management system aimed at covering all stages and tasks in the lifecycle of a knowledge base, from knowledge acquisition, to knowledge organization, to (personalized) knowledge fruition and exploitation. The knowledge base is implemented as a graph database, using the Neo4j [13] DBMS. Nodes and arcs may have associated attribute-value maps; nodes (representing individuals) may be labelled with one or many labels (usually representing classes), while each arc (representing a relationship) may be labeled with one type only. No schema handling is provided for by Neo4j, meaning that the user is totally free to use any type and/or attribute name for any single node and arc. While ensuring great flexibility, this does not allow to associate a clear semantics to the graph items. For this reason, GraphBRAIN requires its users to work according to pre-specified data schemes, expressed in the form of ontologies. Thus, a characterizing feature of GraphBRAIN is its bringing to cooperation a database management system for efficiently handling, mining and browsing the individuals, with an ontology level that allows it to carry out formal reasoning and consistency or correctness checks on the individuals.

The administrators of the knowledge base may build and maintain the general and domain-specific ontologies by specifying the types of entities and relationships to be considered, each with its attributes and associated datatypes. The universal class is implicit, so the user must start the ontology description from the top-level classes, which are automatically considered as disjoint by the system. Each top-level class may be the root of a hierarchy of subclasses, for which no assumption about disjointness is made. Some classes and relationships may appear in different ontologies, possibly with different attributes, in order to reflect different perspectives on them. In particular, GraphBRAIN provides a top-level ontology, defining very general and highly reusable concepts (e.g., **Person**, **Place**) and relationships (e.g., **Person.wasIn.Place**). It plays a crucial role to interconnect the domain-specific ontologies, ensuring an overall connected knowledge graph. Indeed, there is a single, shared graph underlying all the domains. Thanks to the classes shared across different domains, this allows the system to reuse knowledge across domains, and thus to reach a wider range of outcomes for satisfying the user information needs. So, if an individual is used by different ontologies, it acts as a bridge among those ontologies, allowing the users of a domain to obtain additional information coming from other domains.

The ontologies are saved in an internal format, used as a schema for the graph database. The tool may also export them into standard Semantic Web formats, to make them publicly available for reuse. Currently, it can serialize them to Ontology Web Language (OWL)[4] format, with namespace prefix **lam**,

[3] A demo of the system can be found at http://193.204.187.73:8088/GraphBRAIN/.
[4] http://www.w3c.org/owl.

so that it can be published and exploited for ensuring semantic access to the knowledge base and make it interoperable with other resources. The tool models the particular case of different collection types by declaring some specific OWL classes and sub-properties. For example, object property **lam:belongsTo** has concept **lam:Collection** as the range, but several disjoint classes as domain (e.g., **lam:Person** and **lam:Document**). The tool defined one sub-property of **lam:belongsTo** for each of these domain classes. In this way, instead of having a generic property:

(**lam:Person** or **lam:Document**) **lam:belongsTo lam:Collection**

one may assert instances of either relationships:

lam:Person lam:personBelongsTo lam:PersonCollection
lam:Document lam:documentBelongsTo lam:DocumentCollection.

GraphBRAIN uses the ontologies to drive and support knowledge base creation and enrichment, plus all other functionalities, including a set of advanced tools for searching and browsing the knowledge base, and a set of mining, analysis and knowledge extraction tools that may be used interactively by end users or provided as services to other systems for obtaining selective and personalized access to the stored knowledge.

Information is fed into the knowledge base by interaction with users or by automatic knowledge extraction from documents and other kinds of resources (e.g., the Internet). The interactive interface, shown in Fig. 1, consists of two form-based tabs, one for entities (Fig. 1, bottom-left) and one for relationships (Fig. 1, bottom-right), allowing the user to insert/update/remove instances and their attribute values. The forms are automatically generated by the system from the internal format specification of the ontologies. For this reason, albeit GraphBRAIN may handle several ontologies, each specifying a different domain, the form-based interface for data management and querying requires the user to select one of the available domains in order to load the corresponding scheme/ontology to be used (Fig. 1, top-left). While the knowledge base content may be published as linked open data (LOD) [7], it is not available in its entirety as LOD. Indeed, it is accessible only through the querying and graph browsing facilities in the on-line interface, or through pre-defined tools exposed as services, that, based on their input parameters, return relevant portions of the graph serialized as RDF.

Additional functionality is also provided. First, users may manage (add, show, delete) attachments for each instance. In this way GraphBRAIN goes beyond knowledge management tools, becoming a full-fledged digital library, whose content is indirectly organized according to formal ontologies, and thus may foster interoperability with other systems. Second, users may add comments, or approve/disapprove, each entity or relationship instance, and even each single attribute value thereof. This can be used to ensure some kind of 'distributed' quality assurance on the content of the knowledge base, and to establish a trust mechanism for the users. Using the comments, the users may also provide useful

suggestions to improve and extend the ontologies. Also, users are encouraged to provide high-quality knowledge, because using a combination of their number of contributions and trust they are assigned 'points' that they may spend in using advanced features provided by GraphBRAIN.

The same form-based interfaces can be used to query the knowledge base for instances of entities and relationships. The retrieved instances may be graphically displayed in another tab, as nodes and arcs in the graph (see Fig. 1, top-right). This allows the user to continue his search in a less structured way, by directly browsing the graph (by expanding or compressing node neighbors). This is useful to explore the available knowledge without a pre-defined goal in mind, but letting the data themselves drive the search. Thus, serendipity in information retrieval is supported, and the users may find unexpected information that is relevant to their information needs.

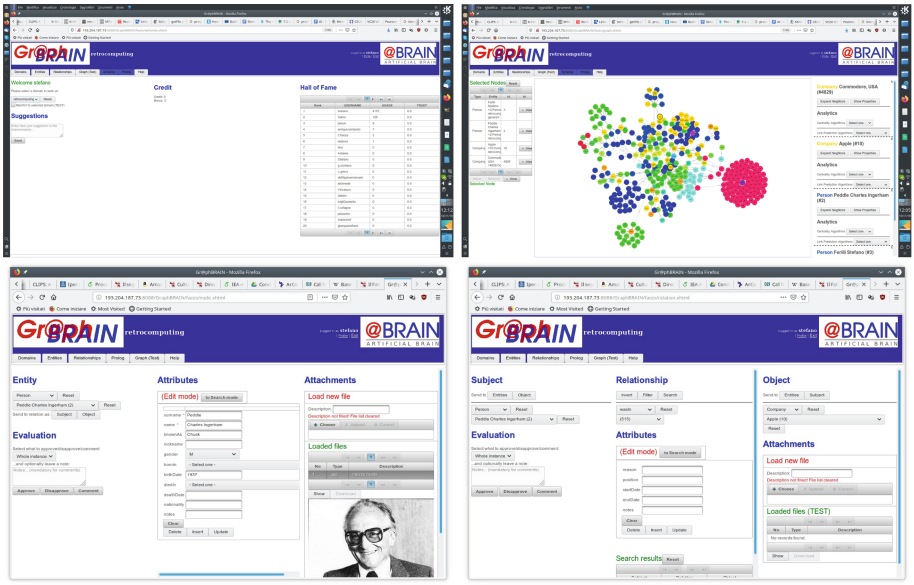

Fig. 1. GraphBRAIN interface for managing and consulting the knowledge base.

Moreover, several analysis, mining and information extraction functionalities are provided, such as:

- assess relevance of nodes and arcs in the graph, and extract the most relevant ones;
- extract a portion of the graph that is relevant to some specified starting points (nodes and/or arcs);
- extract frequent patterns and associated sub-graphs;
- predict possible links between nodes.

Some of the underlying algorithms are reused from the literature; others have been purposely extended to improve their ability to return personalized outcomes that may better satisfy the user's information needs. This would ensure that each user obtains tailored information, which is another novelty introduced by GraphBRAIN. For instance, since the graph is too large to be entirely displayed, when opening the graph tab, the neighborhood (computed by a modified version of the Spreading Activation procedure) of the most relevant nodes (based on PageRank, betweenness and harmonic centrality, etc.) is shown. If a user model is available, based on statistics collected about his previous interaction with the system, the starting nodes may be those more related to his interests, preferences, aims, background, etc. Of course, the displayed portion of the graph may also be the result of a specific user query.

4 Ontological Schema for the Knowledge Base

At the time of writing, the ontology for the 'contextual' description of digital libraries and archives includes 75 classes, 51 relationships and 75 attributes. Some are domain-specific, while some others are borrowed from other (general or domain-specific) ontologies already present in the system. The latter are crucial for allowing us to link domain-specific knowledge to contextual one and to knowledge belonging to other domains, providing information that usually would not be present in library- or archive-specific systems, but might be useful to better understand the library/archive items and to indirectly relate them to other library/archive items. In the following we will describe the main components of the ontology in an informal and intuitive style. For the sake of brevity, relationships and attributes will not be described. Of course, each class or relationship may have its own attributes, and inherits those of its super-classes (if any).

The top-level classes, and their immediate subclasses (if any), are the following (a short description is provided when not obvious).

- **Award**: any kind of recognition that can be awarded to, or record that can be marked by, persons, companies, devices, documents, or components. It has 3 subclasses:
  - **Education**: associated to (more or less formal) educational levels (e.g., B.Sc., M.Sc., PhD, etc., but also certifications, etc.).
  - **Prize**: awards formally granted (usually by some institution);
  - **Record**: the recognition of being the first or the best in doing something;
- **Collection**: any conceivable grouping of items. This ontology currently provides for 2 specific kinds of groupings, corresponding to subclasses:
  - **Persons** (e.g., families, teams, etc.).
  - **Documents** (e.g., series, archives, etc.).
- **Category**, with 2 subclasses:
  - **Concept**, useful to tag documents;
  - **Subject**, useful to categorize content.
- **Company**, currently used to represent both companies and institutions, corresponding to 2 subclasses of this class.

- **Document**, in its most general definition as "something that serves as evidence or proof". As such, it is not limited to printed documents (or documents that might in principle be printed, such as a PDF or word-processor file), but also includes audio-video recordings. It has currently 14 subclasses:
  - **Advertisement**, **AudioRecording**, **Book**, **Booklet**, **Card**, **Deed**, **Leaflet**, **Letter**, **Magazine**, **Manual**, **Movie**, **Picture**, **Postcard**, **Poster**
- **Event** (5 subclasses),
  - **Conference**: a meeting with mainly research or educational purposes.
  - **Fair**: a convention mainly oriented towards selling products and commerce.
  - **Show**: a convention mainly oriented towards showing new products.
  - **Lecture**
  - **Historical Event**: any significant event that should be recorded (e.g., the presentation of a book, etc.).
- **IntellectualWork**: the original result of an intellectual effort, relevant for methodological or practical purposes (9 subclasses)
  - **Algorithm** (e.g., Quicksort);
  - **Approach** (e.g., Step-Wise Refinement for algorithm design);
  - **Invention** (e.g., the Microprocessor);
  - **ProgrammingLanguage**
  - **Subject** (e.g., Information Theory, started by Shannon, or Graph Theory, started by Euler);
  - **Technology**
  - **Theorem**
  - **TheoreticalModel** (e.g., Turing's machine);
  - **WorkOfArt** (e.g., a novel).
- **Item**: a specific, identifiable specimen of a (mass-produced) object, in our case a **Document** (e.g., a signed or numbered copy of a book).
- **Package**: a specific packaging of documents (e.g., a set of books sold together);
- **Person**: reporting personal data about persons;
- **Place**: It is the root of a hierarchy currently made up of 27 subclasses, of which its direct subclasses are:
  - **Administrative**, **Building**, **Geographic**, **Mansion**
- **Software** with 19 subclasses, its direct ones being:
  - **Development**, **Educational**, **Embedded**, **OfficeAutomation**, **OperatingSystem**, **Videogame**

Domain-specific classes are **Award**, **Category**, **Company**, **Document**, **IntellectualWork**[5]. Classes borrowed from the general ontology are **Event**, **Person**, **Place**, **Collection**, and **Item** (the set of subclasses of **Collection** and

[5] Due to the pervasive use of documents in our lives, most elements in this ontology might be regarded as belonging to the general ontology. However, because of its specific focus, this ontology provides much more detailed and richer descriptions for them (in terms of subclasses, attributes and relationships).

Item is extended by defining additional domain-specific subclasses). Classes borrowed from other domain-specific ontologies are **Package** and **Software** (useful to represent digital media often packaged with books, magazines, etc.).

The most relevant relationships in the ontology are:

- **Person.developed.Document** (authors, editors, etc.),
- **Company.produced.Document** (publishers),
- **Document.belongsTo.Collection** (series, collections, etc.),
- **Company.produced.Collection**,
- **Company.wasIn.Place**,
- **Document.wasIn.Place**,
- **Document.concerns.Category**.

Also, other relationships were included to connect classes belonging to different partial ontologies, e.g.:

- **Software.packagedWith.Document**, **Document.requires.Software**.

5 Current Content of GraphBRAIN

A prototype of GraphBRAIN is currently in use as part of a larger ongoing project [5], in which GraphBRAIN will act as the knowledge base management platform underlying an integrated system under development, aimed at supporting all stakeholders involved in touristic activities (tourists, entrepreneurs and institutions). The library/archive domain-specific ontology perfectly fits this overall system, given the central role that documents play in the touristic perspective (e.g., books or movies describing or showing places or cultural heritage artifacts or collections, documents stored in certain places or institutions, etc.). Ontologies for the following domains are currently present in the system (ontology names are the same as the domain names):

general including very general concepts and relationships that are expected to be present in almost all domains;
tourism concerning history, cultural heritage items, points of interest, logistics and services, etc.;
food especially concerning the perspective of typical dishes and beverages from specific touristic regions;
computing concerning computing devices and their history[6] [6];
libraries&archives the ontology described in the previous section.

[6] It is included as a kind of cultural heritage, with the aim of integrating it with more traditional kinds of cultural heritage, both from a scholarly perspective and for fostering its fruition in a touristic perspective. E.g., a tourist interested in the history of computing, while in Bari, might be spotted the chance to visit the collection at the Department of Computer Science, in order to see a specimen of the Olivetti Programma 101 computer.

where 'general' may be considered as a top-level ontology, while the others are domain-specific ontologies purposely developed for the project. Albeit (partly or fully) reusing existing standard vocabularies, they also extend them for the project's specific needs. Table 1 reports overall structural statistics and a comparison among the three most populated ones.

Table 1. Statistics on some ontologies in GraphBRAIN

Ontology	Main classes	Subclasses	Attributes	Relationships	Attributes
Libraries&Archives	12	63	51	51	24
General	17	27	79	88	23
Computing	15	97	111	117	21

The available ontologies share some classes and relationships, by which knowledge items from different domains can be related to each other, extending in this way the available scope of search beyond the single perspectives. In particular, the *general* ontology acts as a hub to inter-link the other ontologies, and allow specific information from one domain to be connected to specific information from other domains. It has significant overlapping with the library/archive ontology described in the previous section, specifically as regards: **Category**, **Document**, **Person**, **Place**, **Series**.

Particularly interesting is class **Category**, aimed at hosting items from different taxonomies. Currently, it is filled with the WordNet lexical ontology [4,12] (under subclass **Concept** for its conceptual part, and under class **Word** for its lexical part) and with the standard part of the Dewey Decimal Classification (DDC) system [2] (under subclass **Subject**). The latter is fundamental for the library domain, because it provides labels to classify the documents. Also the former may play a significant role, allowing us to tag the documents with relevant concepts and words that are, in turn, related to each other, allowing to find non-trivial paths between documents. Specifically, words may be used for lexically tagging other items, while concepts may be used to semantically tag them. Note that the classes in these taxonomies are reified, becoming individuals in the knowledge graph. This allows to handle them within the graph, instead of formalizing thousands of classes in the ontology. So, the categories and words may be linked to individuals of other classes (e.g., documents, persons, places) and used as tags to express information about them (e.g., 'Alan Turing' might be linked with 'Computer Science', 'World War II', etc.).

The current content of the GraphBRAIN knowledge base is summarized in Table 2. For each ontology, the number of instances (*Inst*) and attributes (*Attr*), for both classes and relationships, is shown, along with the average number of attributes (A/C) and relationship instances (R/C) per class instance. Column $A/C+R/C$ reports the average amount of information (i.e., the sum of number of attributes and number of relationships) associated to each class instance. Obviously, the vast majority of knowledge items is in the *general* ontology, including items automatically loaded from WordNet and the DDC taxonomy, plus

Table 2. Statistics on the current content of the GraphBRAIN knowledge base

Ontology	Classes			Relationships				
	Inst	Attr	A/C	Inst	Attr	A/R	R/C	A/C + R/C
Libraries&archives	9 902	31 615	3.19	13 649	10 614	0.78	1.38	4.57
General	333 020	1 744 116	5.24	488 639	39 186	0.08	1.47	6.71
Tourism	250	1173	4.69	318	54	0.17	1.27	5.96
Food	181	405	4.01	65	0	0.00	0.36	4.37
Computing	551	2 096	3.80	739	343	0.46	1.34	5.14
Total	343 904	1 779 405	5.17	503 410	50 197	0.10	1.46	6.63
Total knowledge items	2 123 309			553 607				

other items manually entered by the users. Next comes the *libraries&archives* ontology, also mostly automatically loaded from the records of a private collection, including 4.266 different books, mostly concerning general knowledge, linguistics, literature, history, folklore, and computer science. Then, with much less items, come the ontologies whose knowledge items were manually entered using the on-line collaborative interface: the *computing* ontology, which was the first domain-specific ontology built in GraphBRAIN, and the *tourism* and *food* ontologies, that are the most recently added (and thus the less populated).

There are usually (except for the *food* ontology) less class instances than relationship instances, indicating a quite connected graph, which is important for interlinking the knowledge and providing the users with information based on graph browsing. The R/C parameter reveals that the general subgraph is the most connected, followed by the libraries, food and computing subgraphs and finally by the tourism subgraph. As expected, the average number of attributes per instance is larger for class instances than for relationship instances. Indeed, relationships are by themselves information carriers. Comparing A/C and A/R, we see that the 'information density' is different between classes and relationships for the various domains. For classes, the richest information is in the general subgraph, while the poorest is in the libraries domain, suggesting the much information was missing in the records. For relationships, the richest domain is the library one, while the poorest is food. This shows that much relevant information in the library domain is in the relationships rather than in the attributes, which makes sense considering the strict interplay among several entities (documents, authors, publishers, places, categories, series).

6 Sample Case Study

Since the system is already on-line, no specific evaluation or validation is foreseen for it, except for the exploitation in the tourism-related project, that is indeed contributing in highlighting and fixing problems, refining the ontologies and feeding knowledge graph, and identifying the aspects of the approach to be extended and improved. Also, the on-line system allows any user to provide feedback, comments and suggestions, that will be carefully taken into account in the future versions of GraphBRAIN.

Instead, we provide in the following a sample case study. Due to space limitations, we can only provide very simple examples of use of GraphBRAIN, that hopefully suggest its potential in supporting the users.

User *stefano* logs into the system, and selects the **libraries&archives** ontology/schema. Then it moves to the Entities tab, and using the search facilities selects the following instances (for the sake of readability, their distinguishing attributes will be used instead of their graph id):

- Person:(Stefano Ferilli)
- Company:(Commodore)
- Company:(Apple)

He reads the available information (attribute values) for them, adds some missing information and fixes some errors. This automatically raises his interaction score in the 'hall of fame', and decreases the trust of the users who provided the wrong information. Also, *stefano* sends these instances to the graph, and sends (Commodore) to the Relationships tab, as the object.

Then, he moves to the Relationships tab, where Company:(Commodore) is now selected as the object, and presses the Search button, that returns the list of all relation instances (triples) having (Commodore) as the object instance. Among these triples, he selects:

Person:(Chuck Peddle).wasIn.Company:(Commodore)

He reads the associated information, and decides to know more about (Chuck Peddle). He sends it to the graph as well, and then sends it to the Entity tab.

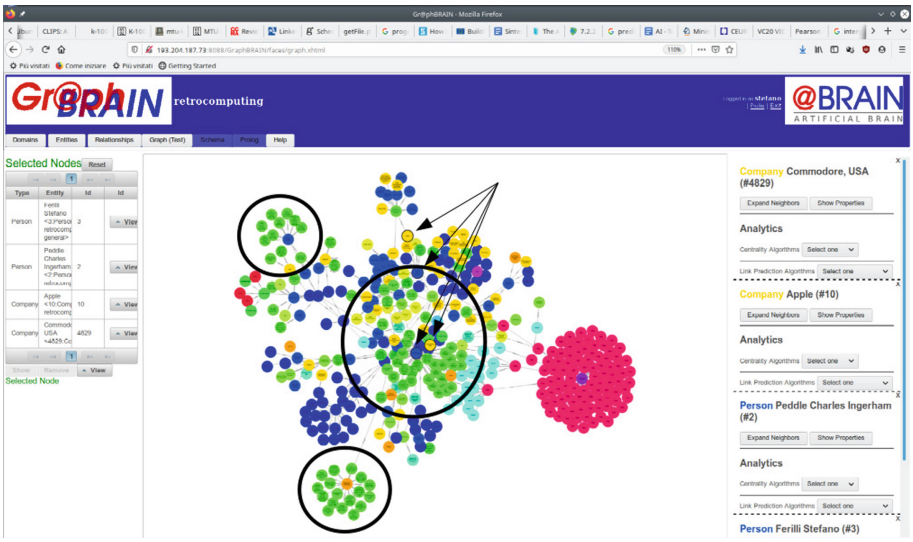

Fig. 2. Portion of GraphBRAIN's knowledge base.

He now turns to the Entity tab, where Person:(Chuck Peddle) is now selected. He reads the associated information, sees his picture in the attachments, and then moves to the Graph tab. There, he finds the list of instances he previously sent to this tab, and a portion of the graph (see Fig. 2) specifically selected starting from these instances and expanding them based on his preferences (e.g., he is more interested in entities **Company** and **Document**, and in relationships **wasIn** and **developed**), including[7]:

- Person:(Chuck Peddle).wasIn.Company:(Commodore)
- Person:(Chuck Peddle).wasIn.Company:(MOS Technology)
- Person:(Chuck Peddle).wasIn.Company:(Apple)
- Person:(Chuck Peddle).developed.Document:(6800 disclaimer)
- Person:(Stefano Ferilli).wasIn.Company:(UniBA)
- Person:(Stefano Ferilli).wasIn.Place:(Bari)
- Person:(Stefano Ferilli).wasIn.Event:(VCFI2019)
- Person:(Stefano Ferilli).developed.Document:(UAsCdB)
- Person:(Stefano Ferilli).developed.Document:(ACotE translation agreement)
- Document:(UAsCdB).wasIn.Collection:(Biblioteca F-Messito)
- Document:(ACotE translation agreement).wasIn.Collection:(Archivio Apulia Retrocomputing)
- Document:(UAsCdB).wasIn.Event:(VCFI2019)
- Document:(UAsCdB).concerns.Company:(Commodore)
- Document:(UAsCdB).concerns.Person:(Chuck Peddle)
- Company:(Apulia Retrocomputing).produced.Document:(UAsCdB)
- Company:(Apulia Retrocomputing).wasIn.Place:(Bari)
- Company:(Apulia Retrocomputing).owned.Collection:(Archivio Apulia Retrocomputing)

Note the occurrence of many instances of the entities and relationships in which *stefano* is more interested. Figure 2 shows the selected portion of the knowledge graph, where the starting nodes (instances) are indicated by arrows. Green nodes are documents, which allows *stefano* to easily spot further documents he might be interested in, obtained by indirect relationship with his initial interests. In particular, one can note potentially interesting aggregates of documents (indicated by circles in the figure).

stefano browses the graph (e.g., by looking at the owners of interesting documents and to the place where they can be found), and spots Event:(VCFI2019). He expands its neighbors, obtaining more information about it (e.g., its venue, other participants, and documents on show there). He asks for the centrality score of node Person:(Stefano Ferilli) based on the PageRank algorithm, obtaining 0.21375000000000002. He also asks for link prediction based on the Resource Allocation algorithm, obtaining 27 suggestions. Then he logs out the system.

This example shows how the proposed system can be used for describing and exploiting the contextual information of a specific digital library or archive, in ways that other traditional systems currently used by specialised users do not provide.

[7] For the sake of compactness, the book title 'Commodore - Un'azienda sulla cresta... del baratro' was reported as the acronym 'UAsCdB'.

7 Conclusions and Future Work

The current availability of new technologies for storing and handling knowledge provides unprecedented opportunities for enhanced fruition of digital libraries and archives, that goes beyond 'simple' document retrieval based on lexical content or metadata. For any kind of users, the context of the document as a whole, and/or of its content, may provide very interesting information, that allows to put it in perspective and grasp a deeper understanding thereof.

This new perspective requires, on one hand, suitable ontological resources to describe such variated knowledge, and, on the other, collaborative tools to collect the precious knowledge scattered across many scholars and practitioners spread all over the world, and to store it in a knowledge base, to make it available to all interested stakeholders (scholars, researchers, but also common people).

These solutions are provided through GraphBRAIN, a general-purpose tool developed to design and populate knowledge graphs, and to allow collaborative enrichment thereof, in addition to advanced fruition, consultation and analysis tools, that may be used as an intermediate layer to provide services to end-user applications aimed at personalized fruition of cultural heritage, also in a touristic perspective.

There are several directions for ongoing and future work. On the ontological side, we are currently extending the number and content of ontologies in GraphBRAIN, and specifically we are refining the ontology for digital libraries and archives, based on the feedback emerging from actual use of the system during the tourism-related project development or obtained by standard users of the on-line prototype. Concerning the knowledge base, we plan to contact pilot users and associations willing to contribute their knowledge. As to the platform, we are continuously improving the interface, also adding functionalities and features. The analysis and mining algorithms, in particular, will be extended and adapted for providing ever more advanced tools and services aimed at supporting researchers, scholars and other stakeholders in tailored fruition of the knowledge base.

References

1. Abburu, S., Babu, G.S.: Survey on ontology construction tools. Int. J. Sci. Eng. Res. **4**, 1748–1752 (2013)
2. Dewey, M.: A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library. Amherst, Mass (1876)
3. Drakopoulos, G., Kanavos, A., Mylonas, P., Sioutas, S., Tsolis, D.: Towards a framework for tensor ontologies over neo4j: representations and operations. In: 8th International Conference on Information, Intelligence, Systems & Applications, IISA 2017, Larnaca, Cyprus, 27–30 August 2017, pp. 1–6. IEEE (2017)
4. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. MIT Press, Cambridge (1998)

5. Ferilli, S., De Carolis, B., Buono, P., Di Mauro, N., Angelastro, S., Redavid, D.: Una piattaforma intelligente per la gestione integrata del settore turistico. In: Primo Convegno Nazionale CINI sull'Intelligenza Artificiale - Workshop on AI for Cultural Heritage, p. 2 (2019). http://www.ital-ia.it/submission/163/paper. (in Italian)
6. Ferilli, S., Redavid, D.: An ontology and a collaborative knowledge base for history of computing. In: Proceedings of the 1st International Workshop on Open Data and Ontologies for Cultural Heritage (ODOCH-2019), at the 31st International Conference on Advanced Information Systems Engineering, CAiSE 2016. Central Europe (CEUR) Workshop Proceedings, vol. 2375, pp. 49–60 (2019)
7. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web. Morgan & Claypool Publishers, Williston (2011)
8. ISO/TC 46/SC 4 Technical Committee: Information and documentation - the Dublin core metadata element set - part 1: Core elements. Technical report ISO 15836–1:2017 (2017)
9. Krötzsch, M.: Ontologies for knowledge graphs? In: Artale, A., Glimm, B., Kontchakov, R. (eds.) Proceedings of the 30th International Workshop on Description Logics, Montpellier, France, 18–21 July 2017. CEUR Workshop Proceedings, vol. 1879. CEUR-WS.org (2017). http://ceur-ws.org/Vol-1879/invited2.pdf
10. Lavarone, G., Orio, N., Polato, F., Savino, S.: Modeling the concept of movie in a software architecture for film-induced tourism. In: Calvanese, D., De Nart, D., Tasso, C. (eds.) IRCDL 2015. CCIS, vol. 612, pp. 116–125. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41938-1_13
11. Lodi, G., et al.: Semantic web for cultural heritage valorisation. In: Hai-Jew, S. (ed.) Data Analytics in Digital Humanities. MSA, pp. 3–37. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54499-1_1
12. Miller, G.A.: WordNet: a lexical database for English. Commun. ACM **38**, 39–41 (1995)
13. Robinson, I., Webber, J., Eifrem, E.: Graph Databases, 2nd edn. O'Reilly Media, Newton (2015)
14. Zilio, D., Micheletti, A., Orio, N.: Crowdsourcing for film-induced tourism: an approach to geolocation. In: Grana, C., Baraldi, L. (eds.) IRCDL 2017. CCIS, vol. 733, pp. 108–116. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68130-6_9