Knowledge Graphs: Semantics, Machine Learning, and Languages
M. Acosta et al. (Eds.)

doi:10.3233/SSW230011

Virtual Reality Based Access to Knowledge Graphs for History Research

Julia BECKER [a] Mario BOTSCH [b] Philipp CIMIANO [a,1] Melanie DERKSEN [b]
Mohammad Fazleh ELAHI [a] Angelika MAIER [a] Marius MAILE [a]
Ingo Oliver PÄTZOLD [a] Jonas PENNINGROTH [a] Bettina REGLIN [a]
Markus ROTHGÄNGER [a] Silke SCHWANDT [a]

[a] *Bielefeld University*
[b] *TU Dortmund University*

Abstract. Purpose: Knowledge graphs have so far been intensively used in the cultural heritage domain. Current interaction paradigms and interfaces however are often limited to textual representations or 2D visualizations, not taking into account the 4D nature of data. In digital history in particular, where events as well as geographical and temporal relationships play an important role, exploration paradigms that take into account the 4D nature of event-related data are important, as they have the potential to support historians in generating new knowledge and discovering new relationships. In this paper, we explore the potential of virtual reality as a paradigm allowing digital humanities researchers, historians in particular, to explore a semantic 4D space defined by knowledge graphs from an egocentric perspective. **Methodology:** We present *eTaRDiS*: a virtual reality based tool supporting immersive exploration of knowledge graphs. We evaluate the tool in the context of a task in which historians and laypersons with a history background explore DBpedia and Wikidata. We report results of a study involving 13 subjects that interacted with the data in eTaRDiS in the context of a specific task, in order to gain insights regarding the interaction patterns of users with our system. The usability of the tool was evaluated using a questionnaire including questions from the System Usability Scale (SUS) in addition to task-specific questions. **Findings:** The usability evaluation showed that our tool achieved an overall SUS score of 71.92, corresponding to a ‘satisfactory’ rating. While the mean score reached with laypersons with a history background was quite high with 76.0, corresponding to a rating of ‘excellent’, the score for historians was lower with 69.4, corresponding to a ‘sufficient to satisfactory’ rating. A qualitative analysis of the interaction data revealed that participants quickly identified the relevant information in the tasks using a variety of strategies and taking advantage of the features provided in eTaRDiS. **Value:** eTaRDiS is to our knowledge the first virtual reality based exploration tool supporting the exploration of knowledge graphs. The findings of the usability evaluation and the qualitative analysis of exploration patterns show that the system could potentially be a valuable tool for allowing digital humanities researchers to explore knowledge graphs as a way to discover new relationships between historical events and persons of interest.

Keywords. knowledge graphs, VR, digital humanities, digital history, linked data, semantic web

[1] Corresponding Author: Philipp Cimiano, cimiano@cit-ec.uni-bielefeld.de

1. Introduction

Knowledge graphs have been shown to be useful in the cultural heritage (CH) domain to connect and access large collection databases originating from multiple sources [1]. Hyvonen et al. [2], for example, have developed the SAMPO data model which provided the basis for implementing multiple data portals to access information about cultural heritage. With this model, content can be made semantically interoperable to support access, search, and discovery, thus overcoming boundaries of data silos. While interoperability is an important dimension in providing access to CH data, investigating by which access paradigms digital humanities researchers can be supported in answering their genuine research questions using knowledge graphs is also an important research avenue.

In this paper, we are in particular concerned with how to leverage knowledge graphs for the work of historians. Conceptions of history are widely shared as narratives that decisively depend on socio-cultural frames of reference and systems of representation [3]. Exploring and understanding these systems is a central task of the humanities and cultural studies, with the goal to communicate a spatio-temporal relationship, e.g. a historical moment, which equals the construction of meaning and ultimately knowledge. In history, agents, places, and events and their relationships play an important role. Events in particular can be regarded as 4D objects as they take place in a particular geographical location and have a temporal extension. Agents in turn can be located at different places at different times and places can undergo changes over time. It is thus key that tools that provide access to knowledge graphs to historians account for the 4D nature of relevant entities, allowing to explore the spatio-temporal relationships between entities.

The idea of using virtual reality technology (VR) in the CH domain is not fundamentally new. VR has been applied to the visualization of historical artifacts or full museums to enable a more interactive exploration of such artifacts or museum collections [4,5,6]. Our approach differs from the *virtual museum* paradigm in that we do not follow the paradigm of 'bringing history to life' but aim to facilitate the exploration of historical data networks. Jósza et al. [7] create 4D spatio-temporal models of large dynamic urban scenes containing various moving and static objects. Although their work has not been carried out in the CH domain, it shows that VR is an adequate medium to visualize and explore spatio-temporal 4D data.

Knowledge graphs built from Linked Open Data (LOD) such as DBpedia and Wikidata can provide a basis for the construction of different views on networks. From these knowledge graphs, the resources of historical events, persons, places, etc., and related multimodal data can be retrieved and organized in a network for analysis in historical studies. Using the three-dimensional space in immersive scenarios makes it possible to explore new spatio-temporal relations of historical events, persons, etc. in virtual space. In this way, it becomes possible to observe the actions of people in a network of space and time that makes new connections visible [8, p. 8]. According to Ayers, the visualization of time happens through the visualization of the movement of events and people in space [8]. This perception is central to the construction of events [9, p. 193] and persons [10], which can also be interpreted as specific spatio-temporal relations.

Towards the goal to leverage knowledge graphs for the work of historians, we present a new approach to the exploration of knowledge graphs that relies on VR to create an immersive space in which the spatio-temporal connections between entities can be explored. Our approach allows to explore the semantic neighborhood of entities from different perspectives: focusing on entities that are geographically close, temporally close, or semantically close, depending on the research question.

Specifically, we present eTaRDiS, a VR application to enable novel spatio-temporal access to knowledge graphs that allows exploring the neighborhood of nodes by filtering according to different aspects: time, location, and semantic distance in particular. As a VR tool, eTaRDiS offers a new perspective on spatio-temporal relationships by deconstructing existing historical narratives, allowing the user to explore, compare, and (re)order abstract knowledge fragments, i.e. data from a knowledge graph. As a consequence, eTaRDiS can be framed as a “possibility space” [11], a sort of laboratory for data and an immersive exploration environment. An initial expert review of our approach gave us valuable insights into the use cases and requirements that guided our developments [12].

The application consists of two main views: the *Hub* (see section 3.2.1) and the *Fragmentarium* (see section 3.2.2). The Hub allows a user to select a dataset to explore and a node to start the exploration. The Fragmentarium allows to explore the neighborhood of a given node; we call such neighborhoods that are defined according to some distance measure *historical fragments*. The neighborhoods are visualized in VR as labeled spheres floating around the user. Selecting one of the related data points, in turn, updates the visualization by showing those data points related to the newly selected one. This enables the exploration of the network of related agents, places, events, etc. according to one’s own individual interest.

As a main contribution of this paper, we describe the design and implementation of the eTaRDiS system. We further present the results of a user study involving 13 participants. As main result, the study showed that users generally found it easy to identify the relation between historical fragments using the features of the eTaRDiS VR application and by exploring the semantic neighborhood of fragments. A qualitative analysis of the interaction revealed that participants used different strategies to solve the tasks, tested the boundaries of the rules in our virtual environment, and liked the playful approach to exploring historical data. In the study, our tool achieved an overall System Usability Scale (SUS)[13] score of 71.92, corresponding to a ‘satisfactory’ rating. Further analysis showed that the self-assessment of users regarding their usage of digital media had an influence on the SUS score. Participants who rated themselves as ‘good’ to ‘extremely good’ at using digital media have a higher mean SUS score of 77.8 (a rating of ‘satisfactory to excellent’) compared to the participants who rated themselves as ‘very bad’ to ‘not bad’ at using digital media (mean SUS score: 58.8).

2. Related Work

Virtual reality (VR) has been shown to be a valuable tool for interaction with networks in several areas, especially in the natural sciences. For example, Pirch et al. [14] developed the ‘VRNetzer platform’ for the interaction with large network structures of genes to identify genes indicating rare diseases (see also the work of Buphamalai et al. [15]). Some applications, namely HisVA [16], VaiRoma [17], and POLIS [18], have been implemented for the exploration and analysis of data from the Semantic Web through 2D visualizations. However, no VR applications have been developed in the CH domain and the digital humanities to explore existing knowledge graphs.

According to Kidd [19], there is an immersive turn in museums and heritage contexts, one consequence of which has been a broader range of digital and other media utilized in museums. The goal of *immersive VR* is to give the impression to users that they are truly in a synthetic world that can be reached via using a Head-Mounted Display (HMD) because it allows users to focus on the display of projected VR without distraction [20]. For example, the immersive exploration of CH data in virtual and augmented reality has been proven useful in higher education in Art History [21]. Furthermore, Razum et al. [22] developed the virtual research environment TOPORAZ in which a 3D model of the main market of the city of Nuremberg is linked to a database in four time layers [22]. Although we also use immersive VR in eTaRDiS and make use of an HMD, our approach differs from the approach of Casu et al. [21] and Razum et al. [22] in that they follow the paradigm of 'bringing history to life', while in our approach VR supports the exploration of a network of historical fragments.

McIntire and Liggett discussed a bunch of data and information visualization applications to find out which kind of tasks might benefit or even suffer from a third dimension [23]. Among others, 3D proves advantageous to tasks including precise spatial localization of objects, complex imagery analysis, and manually interacting with data or virtual information. Furthermore, they found that a 3D visualization can provide performance improvements that correlated with cognitive benefits and that facilitate a better understanding of spatial and/or multidimensional data. Overall, they concluded that 3D is especially beneficial for data interpretation tasks.

Wagner Filho et al. [24] stated that exploring 3D scatter plots with an HMD leads to a smaller effort in finding information and offers a much larger subjective perception of accuracy and engagement as opposed to desktop applications, but may suffer from occlusions. Kraus et al. [25] investigated the impact of immersion on cluster identification tasks in scatterplot visualizations. Their results indicate that task performance differs between the investigated visualization design spaces in terms of accuracy, efficiency, memorability, sense of orientation, and user preference. In particular, the 2D visualization on the screen performed worse compared to the 3D visualizations with regard to the measured variables. The study shows that an increased level of immersion can be a substantial benefit in the context of 3D data and cluster detection. Overall, they state that virtual environments can indeed provide suitable design spaces for abstract visualizations such as scatterplots. Furthermore, it became apparent that getting an overview of three-dimensional data can be enhanced by means of VR due to a more natural navigation, and better orientation and memorability capabilities.

Wagner et al. [26] investigated the effect of exploration mode and frame of reference in immersive analytics. They found that egocentric exploration of space significantly reduced mental workload. Exocentric exploration, in turn, improved performance on some tasks. They concluded from their research that generally a room-scaled environment should be favored if the necessary space is available; both the egocentric and exocentric perspectives should be offered so that users can decide whether and when to switch between these perspectives based on their interests.

It can be concluded from previous work that VR seems to be a suited tool to provide access to historical data from multiple dimensions and perspectives, leading to a better understanding of the data and of the corresponding relationships inherent in it. The above-mentioned studies have substantially influenced the design of our application, in particular the choice of providing an egocentric perspective to the user at initiation in order to avoid occlusions. According to Wagner et al. [26], we let the user transform the initially room-scaled visualization to enable the switch to an exocentric perspective.

3. Design and Implementation of eTaRDiS

This section describes the design and implementation of eTaRDiS. First, we provide an overview of the types of users that the system is intended to support and how their needs differ. We then describe the actual implementation of the system, emphasizing our key design choices to meet the needs of our target users.

3.1. User Groups

Opening up the virtual space for the field of humanities and cultural studies also meant designing eTaRDiS to be intuitively usable and making it easily accessible to historians, laypersons interested in history, and people with an affinity for technology and an interest in VR. The eTaRDiS is intended for three stereotype users: Historians that are experts in a specific area of historical scholarship or that are academically active in that area, laypersons who are interested in history and cultural studies, and users who have no affinity for dealing with cultural or historical data. In our user study, we investigate the perspective of historians and laypersons.

The user group of historians knows research discourses and their contexts. In the composition of the sub-dataset, this group is able to translate their research perspectives in a very fine-granular way into the filter options offered by eTaRDiS and thus operationalize their questions. Exploration in the Fragmentarium not only provides an overview but also encourages them to change their perspective on a topic. The user group of laypersons includes not only people who have a basic affinity for history but also those who are on their way to a degree in the field of cultural and historical studies. In their use of eTaRDiS and their choice of filters, they are clearly more open and less specific than described in the first user scenario because they go into eTaRDiS with a broader view. It can be assumed that they use the insular knowledge they have acquired over the years and select a topic focus for their exploration against this background. They are concerned with the controlled reduction and expansion of complex units of information. They use their time in the Fragmentarium to gain overviews and consume the (text and image) data deposited for the information much more intensively.

3.2. Implementation

The eTaRDiS system consists of a VR application and a backend database. The VR application has been implemented using the *Unity* game engine[2], so that it supports the visualization of and interaction with knowledge graphs in a 3D immersive environment. The VR application consists of two virtual spaces: the *Hub* and the *Fragmentarium*. These spaces are clearly separated by different purposes and forms of interaction.

[2]Although we specialized our developments on the *HTC Vive*[3], the underlying concepts are applicable to other VR hardware as well.

[3]https://www.vive.com/de/product/vive-pro/

3.2.1. Hub

The Hub serves as a central retreat within the virtual environment of eTaRDiS in which basic settings such as language, type of keyboard input, etc. can be edited. In addition, the Hub is the starting point for virtual data exploration. An appropriate interface enables the user to filter the data to create a sub-dataset adequately tailored for individual (research) interests (see Figure 1), which can then be subsequently visualized and explored in the subsequent room, the Fragmentarium.

In order to allow users to get an overview of selected data, information about the filtered sub-dataset is dynamically displayed on a dashboard to allow a user to filter the dataset according to individual interests. In addition to a display of the total amount of nodes (*fragments*) in the created sub-dataset, temporal and spatial information is shown. Statistics are displayed for the entire sub-dataset, and the spatial distribution of the individual fragments becomes visible on a world map. Based on this information, initial considerations can already be made for selected filters.

In addition to filtering, the Hub interface allows loading *saved points* and *presets*. Saved points allow a user to continue a session at a later point in time and thus to start at the same point in the Fragmentarium, while presets can be used to save and load defined filters. presets can also be edited later.

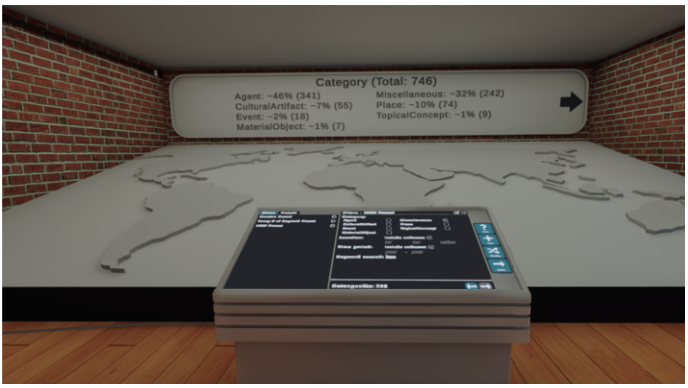

Figure 1. Filtering the dataset dynamically updates the statistical information on a dashboard in the Hub.

3.2.2. Fragmentarium

The Fragmentarium is the interface allowing a user to explore the neighborhood of a node in the knowledge graph selected in the Hub. The visualization consists of a historical main fragment and related fragments. The user is standing on the main fragment, represented as a colored circular platform, and is therefore intended to explore the scene from an immersive egocentric viewpoint. Related fragments are represented as labeled colored spheres that surround the user (see Figure 2). While the color indicates the fragment's category (e.g., agent, event), the sphere's size is related to the number of sources (i.e., `dbo:wikiPageExternalLinks`) available in the corresponding DBpedia article.

For the arrangement of the related fragments, two kinds of distance measures are used. The first one is used to position the spheres on a spherical shell around the user so that fragments with a low dissimilarity are close to each other. The second one determines their individual distances to the center according to their dissimilarity to the main fragment. The minimal and maximal distances to the center and thus to the user are 1.7 m and 3 m. A high dissimilarity results maximally in a distance of 3 m to the center.

To position the spheres on a spherical shell, we rely on the *UMAP* algorithm [27], which is a dimension reduction technique. With the dissimilarity calculation of the related fragments among themselves as its input metric and the haversine as its output metric, UMAP finds a suitable arrangement of the related fragments around the user. To prevent dislocations of the neck, in our case, the UMAP algorithm gets two additional artificial points in its optimization process that are fixed on the poles and constantly repel all the other data points, which bypasses constellations in which spheres accumulate directly above or beneath the user.

We rely on four attributes for which we calculate dissimilarities: i) temporal (t), ii) spatial (s), iii) semantic (sem), and iv) categorical (c). Users can define the neighbourhoods to explore by arranging other nodes that are closest with respect to these different dissimilarities. We first compute pairwise distances for the related fragments in one visualization for each of the above-mentioned attributes and afterwards, scale them per attribute a to an interval from 0 (which means *similar*) to 1 (which means maximal *dissimilar*) to get a normalized measure of dissimilarities for each individual attribute. If there is no data available for a fragment $n \in F$ and one of its attribute values a_n, then the value is considered invalid. More formally, for a visualization with the related fragments F, the dissimilarity *diss* for two fragments $i, j \in F$ regarding one attribute $a \in \{t, s, sem, c\}$, with a_i, a_j being their corresponding attribute values, is defined as:

$$\mathrm{diss}_a(i,j) = \begin{cases} \frac{\mathrm{dist}_a(i,j) - \min_{k,l} \mathrm{dist}_a(k,l)}{\max_{m,o} \mathrm{dist}_a(m,o) - \min_{k,l} \mathrm{dist}_a(k,l)}, \text{ with } k,l,m,o \in F, k \neq l & \text{if } a_i, a_j \text{ are valid,} \\ 0 & \text{if } a_i, a_j \text{ are invalid,} \\ 1 & \text{otherwise.} \end{cases}$$

Note that two related fragments may have different dissimilarity values after the scaling process. This is due to the fact that two fragments may appear hand in hand for various main fragments which might have distinct sets of related fragments with different minimal and maximal distances, potentially leading to different scalings. We describe the way we compute the different distances in what follows:

Temporal distance: For the temporal distance dist_t between two fragments, the minimal separating interval regarding their time periods t_i and t_j is considered. If there are time periods provided that do not overlap, then the difference in days is taken into account. If their periods overlap, or if no time period is given for both fragments, their distance is 0.

Formally, a fragment's time period t_i is an interval $\left[t_{i,start}; t_{i,end}\right]$ starting with a start date $t_{i,start}$ and ending with end date $t_{i,end}$, each of which is represented in days (the number of days that have elapsed from common era to the corresponding date associated with the fragment). Two time periods t_i and t_j overlap, if their intersection is not empty: $t_i \bigcap t_j \neq \emptyset$. Their temporal distance in days $\mathrm{dist}_t(i,j)$ can be calculated as follows:

$$\mathrm{dist}_t(i,j) = \begin{cases} \min\left\{||t_{i,start} - t_{j,end}||, ||t_{j,start} - t_{i,end}||\right\} & \text{if } t_i \bigcap t_j = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Spatial distance: The spatial distance dist_s between two fragments is defined as the shortest distances of the corresponding locations. Since locations are represented in pairs of longitudes and latitudes (ϕ, θ), the distance $dist_s$ between two locations (ϕ_i, θ_i) and (ϕ_j, θ_j) can be calculated via the great circle distance with $r_{earth} \approx 6371,009$ km being the earth's radius:

$$\text{dist}_s(i,j) = r_{earth} \arccos\left(\sin(\theta_i)\sin(\theta_j) + \cos(\theta_i)\cos(\theta_j)\cos(\phi_i - \phi_j)\right).$$

Semantic distance: We make use of the Sentence Mover's Distance by Clark et al. [28] which is an evaluation method for multi-sentence texts. As sentence vectors, we use a fine-tuned sBERT model [29] based on the pre-trained model 'all-MiniLM-L12-v2'[4]. We applied the Sentence Mover's Distance to the abstracts as textual descriptions of entities, and use uniform weights for the sentences in the calculation of the distance.

Categorical distance: The categorical distance is per definition a dissimilarity and estimates to 1 if two fragments do not share the same category. Otherwise, it is set to 0. Given the categories of two fragments, c_i and c_j, the categorical distance dist_c is computed as:

$$\text{dist}_c(i,j) = \begin{cases} 1 & \text{if } c_i \neq c_j, \\ 0 & \text{otherwise.} \end{cases}$$

The user can choose according to which dissimilarity the data should be arranged. In our scenario the distance measure that is used as the input metric for the UMAP algorithm takes the dissimilarity values of the categories of the related fragments among themselves into account while the second one makes use of the semantic dissimilarities between the related fragments and the main fragment. Thus, related fragments that share the same category appear in clusters around the user and related fragments that have a high semantic dissimilarity to the main fragment are spatially distant to the center. Therefore, the visualization gives clues regarding the relationship between the main fragment and its corresponding related fragments and regarding the interrelationships between related fragments at one glance.

3.3. Design Principles

The features in the Hub and the Fragmentarium were developed to meet the requirements for the scholarly primitives by Unsworth [30] `comparing`, `annotating`, `discovering`, and `sampling`. These primitives, among others, are common methods of humanities researchers, basic functions common to scholarly activity across disciplines, over time, and independent of theoretical orientation [30][5].

The primitive of `comparing` is addressed by the so called *relation details* in the Fragmentarium, which show pairwise connectivity information for the nodes regarding space and time (see Figure 2). The `annotation primitive` is realized by allowing users to mark nodes to find them later again more quickly. If nodes are marked, their mark appears in the color of their respective category on the compass, which is a guidance line in the upper field of view (see Figure 3). Furthermore, for each node, a *detail window* can

[4] https://www.sbert.net/docs/pretrained_models.html

[5] Pacheco [31] gives an elaborate analysis of the scholarly primitives in the digital humanities, also from other researches over the years.

be opened with collapsable additional information like its *characteristics* (e.g., locations, time period, etc.), an image, and an abstract (see Figure 3). In that window, a space is reserved for annotation by allowing users to enter notes. The primitive `discovering` is addressed with the visualization and features in the Hub and the Fragmentarium as a whole. The user can apply filters or show statistics on datasets in the Hub. At the same time, this addresses the requirements of `sampling` because the filters in the Hub can be used to select data according to the user's interests. Selecting a new node of interest with a ray pointer attached to the user's controller in the Fragmentarium makes it the new *main fragment* and updates the visualization. Thus, the user is able to move from node to node while inspecting their properties and relations. A history shows the order of the main fragments visited so far (see Figure 3). Following the recommendations of Wagner et al. [26], we offer both the egocentric and exocentric perspective to the user, who is allowed to grab and transform the whole constellation of the related fragments (which includes translating, uniform scaling, and rotating). Thereby, the user may switch from the initial egocentric to an exocentric perspective. Finally, the application offers a help menu that explains all functionalities.

3.3.1. Backend

In order to implement the data backend, we rely on Neo4j[6] as a graph database designed to manage and query large knowledge graphs. We rely on DBpedia as the main data source for several reasons. First, DBpedia, being built on top of Wikipedia, is the most widely [32] used online encyclopedia, and one of the most prominent examples of truly collaboratively created content. The Wikimedia Commons[7] is part of the Wikipedia encyclopedia family that contains over twenty-five million audio, video, and image files [33], including scanned books, historically significant photographs, illustrative figures, and maps. Accessing historical data via several media plays an important role in their exploration and interpretation. Third, DBpedia is interlinked to other related information and also connected to other datasets [34] that allow us to explore historical data through semantic relatedness. In addition to DBpedia, we use Wikidata as it represents information on the time and location of a resource by a set of well-defined properties, which is crucial for historical data exploration.

First, we extract resources from DBpedia including Wikipedia text, images, video, and audio files. For each resource, we extract its connections in the forward and backward directions. After that, we find the equivalent entry in Wikidata, provided via the property `owl:sameAs` in a DBpedia resource, and retrieve temporal and spatial information.

In order to integrate data from DBpedia and Wikidata into a knowledge graph database in the context of historical studies, we have worked with historians in the eTaRDiS project to develop an appropriate database schema. We re-structured the classes of DBpedia into the seven categories: `Agent`, `Place`, `Event`, `Topical Concept`, `Material Object`, `Cultural Artifact` and `Miscellaneous`. The reason is that DBpedia classes were not designed for the context of historical data. For example, the Motherland Calls (i.e., a resource in DBpedia `res:The_Motherland_Calls`) is a statue representing *'Heroes of the Battle of Stalingrad'* but the class of the resource is `dbo:ArtWork` in DBpedia, which is re-structured as `Cultural Artifact` in our

[6]https://Neo4j.com/
[7]https://commons.wikimedia.org/wiki/Main_Page

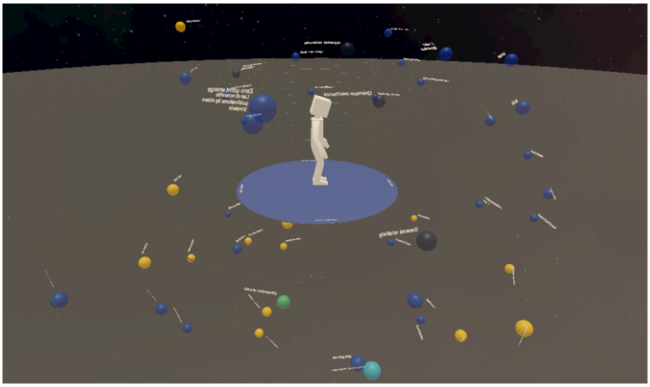

Figure 2. An external view of the visualization shows the main fragment (here Albert Einstein) represented as a colored platform the user is standing on. Related Fragments are represented as colored, labeled spheres. Their arrangement gives clues regarding the fragments' interrelationships.

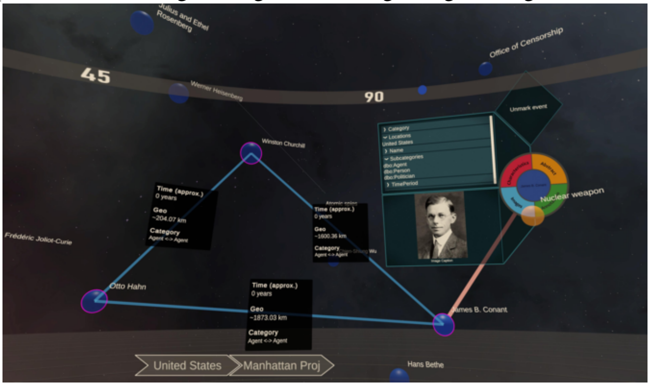

Figure 3. The user explores the scene from an egocentric perspective. For every sphere, its corresponding detail window reveals deeper information on the corresponding fragment by showing up its time period, locations, a picture (if available), and a describing text. The history localized in the lower area of the field of view shows the order of the main fragments visited so far (here United States). In the upper part of the field of view, the compass serves as a guidance line; if nodes are marked, their mark appears on the compass in the colour of their respective category. The feature referred to as 'relation details' helps to figure out how two fragments are related. It shows the gap in time, the shortest distance between their locations, and their categorical differences.

database. The resources of class `dbo:Person`, `dbo:Organization`, `dbo:Species`, `dbo:Language` and `dbo:EthnicGroup` were classified into the category `Agent`. Furthermore, the resources of class `dbo:MeanOfTransportation`, `dbo:Currency`, `dbo:Device`, `dbo:Food` and `dbo:ChemicalSubstance` were classified as `Material Object`. All the resources that could not be classified into one of the above categories are classified into `Miscellaneous`.

In another mapping process, we mapped properties from DBpedia and Wikidata to a minimal set of descriptive properties for each of our seven categories. The descriptive properties are i) `category`, ii) locations, iii) `name`, iv) `sub-categories`, v) `time period` and vi) `properties`. All properties containing geographical information of locations are listed. Respectively, all properties containing time information are collected in the descriptive property `time period`. Table 1 shows an example of six descriptive properties for category Agent (i.e., *res: Albert_ Einstein*). The descriptive property properties consists of 26 properties and some of them are shown in the Table. These sets of minimal descriptive properties of each resource were used in the detail window of a fragment in the Fragmentarium as a standardized brief description (see Figure 3).

	Descriptive properties	**Values**
1	category	Agent
2	name	Albert Einstein
3	locations	Germany, United States, Switzerland, Austria
4	subcategories	dbo:Agent, dbo:Person, dbo:Scientist
5	time period	0.1.01.1879 - 31.12.1955
6	properties/gender properties/positions properties/religion properties/abstract	male professor pantheism "Albert Einstein was a German-born theoretical physicist, widely acknowledged to be one of the greatest physicists..."

Table 1. An example of six descriptive properties for the resource *res:Albert_Einstein*. The descriptive property properties consists of 26 properties for this resource.

To access a knowledge graph database and apply adequate filter functions, an API was developed. The API allows us to query the database for specific resources that can be visualized as historical fragments in the Fragmentarium. In addition, filters could be applied to query the database for resources in certain geographical places or time frames. Additionally, the API was used to compute the central and relative semantic distances between fragments based on the abstracts from DBpedia.

4. User Study

In this section, we present a user study that helped us to gain insights into how users interact with eTaRDiS. The user study was designed with two goals: 1) to evaluate if the user (a historian or a layperson with an interest in history) could retrieve relevant information regarding a certain historical question, 2) to assess the usability of the application.

The participants started in the Hub and were asked to open a preset with the main fragment Hundred Years' War (i.e., the resource res:Hundred_Years'_War) as the starting point in the Fragmentarium and got a brief tutorial on the main functionalities. After the introduction, they were instructed to choose a preset with the fragment Albert Einstein (i.e., the resource res:Albert_Einstein) as the starting point of four successive tasks and were asked to "think aloud" (method based on [35]) in the Fragmentarium. In this section, we describe the tasks, participants, and results of the user study.

4.1. Tasks

The underlying question for the tasks was: *'What is the relation between the fragments World War II (WWII), J. Robert Oppenheimer (Oppenheimer) and the Manhattan Project?'*. In order to answer the question, the participants were first asked to find the three fragments in the Fragmentarium. As a second task, they had to retrieve information on the location and time for each fragment. As a third task, they were then asked to read the abstract of the fragments and explore the distance to other fragments as a basis to find the relation between the different fragments. Finally, as a fourth task, they were asked to assess *'how prominent the relation between the different fragments was'*. We relied on the System Usability Scale (SUS) developed by Brooke [13] as a basis for a questionnaire administered to the participants, adapting the questions to the eTaRDiS setting. The participants assigned a score on a five-point scale from *'not true at all'* to *'fully agree'* for each of the ten questions regarding aspects of the usability of eTaRDiS.

4.2. Participants

The participants were recruited through a seminar of historical studies and the digital humanities and received credits for their participation in the study. In total, 13 students participated in the user study. The age of the participants was between 24 and 42 years (average age 29.13). Six participants had a background in historical studies. Three participants even had an occupation related to history, e.g., as an archivist, or they were working in a project related to history. Five participants were interested in history and had a background in different research areas of social sciences (see Section 3.1). The interviews were transcribed and analyzed using qualitative content analysis. The total dataset comprises 2 hours 48 minutes and 41 seconds of audio and video recording.

4.3. Results

For eleven out of 13 participants, eTaRDiS was the first VR experience. All participants were able to complete the given tasks and eleven participants found the relation easy to identify. In the first task, the participants needed between 8 and 84 seconds (42.85 seconds on average) to find the three relevant fragments. One participant inferred from the names of persons on fragments coloured in blue that all blue fragments denote Agents and used this correct inference to find the (blue-coloured) fragment for Oppenheimer. Another participant used a sphere's size to find the fragment WWII, assuming the sphere for the event to be relatively large. Two other participants used the labels of the surrounding fragments to locate Oppenheimer and the Manhattan Project, assuming the fragments had to be in the proximity of the fragments of physicists. Four participants marked the corresponding spheres. The proximity of the fragments Manhattan Project and WWII represented a challenge to the exploration for some participants as the label of one of the fragments was partially occluded by the open detail window of other fragments.

In the second task, the participants used three strategies to find the location and time of the fragments. The first strategy used by six participants was to open the characteristics in the detail windows of the fragments and read the information under location and time period. For WWII, no locations were given but the participants found the relevant information in the compilation of the place property. The second strategy used by five participants was to read the abstract of each fragment. The third strategy used by two participants was to read the characteristics and the abstracts. In all strategies, the participants zoomed in or moved the detail windows for a better view of the texts.

In the third task, the participants also used three strategies to identify the topical relation of the fragments. In the first strategy, three participants read the abstracts. In particular, the participants used the names of the fragments as keywords to search for relevant information in the abstracts. In the second strategy, six participants used the relation details in combination with the abstracts, and in the third strategy, three participants additionally opened the images given for the fragments. One participant answered the question without using any of the information or functions in the Fragmentarium.

As stated before, eleven participants out of 13 found the relation of the three fragments easy to identify with the information and functions given in the Fragmentarium. In the third and fourth tasks, six participants stated that they had prior knowledge regarding the relation of the fragments. Five of these six participants were in the group of historians. The layperson with prior knowledge stated that they knew that *"Oppenheimer invented the atomic bomb [and] that it goes hand in hand with the Second World War, of course. However, the fact that the Manhattan Project is related to the atomic bomb and that Oppenheimer also belonged to the Manhattan Project, would not have come to my mind before"*. Several information and functions were seen as helpful by the participants to identify the relation. One participant referred to the images given for a fragment, and one participant referred to the proximity of the fragments. Two participants referred to the relation details, the relative distance weighted by category, or related fragments. Four participants found the abstract and other textual information helpful.

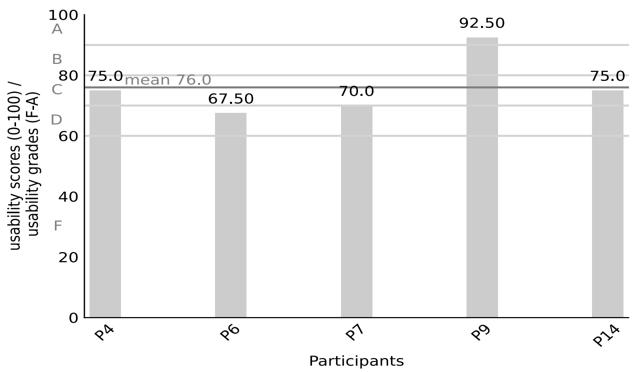

Figure 4. SUS scores of laypersons.

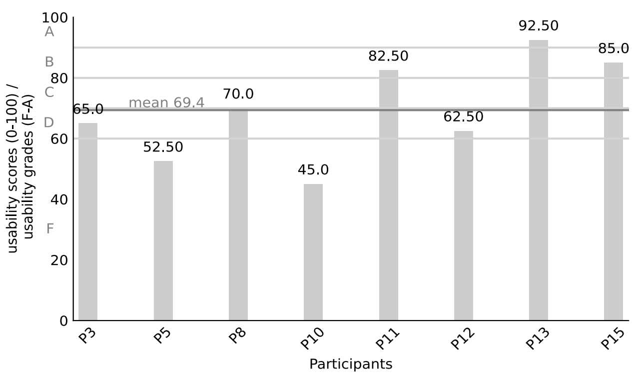

Figure 5. SUS scores of historians.

In the evaluation of the usability questionnaire, we computed the SUS score for each participant, the mean overall, and the mean for each of the two groups, laypersons (five participants) and historians (eight participants) (see Figure 4). We use the scheme of Bangor et al. [36] as a reference for the interpretation of the resulting SUS scores regarding the acceptability of the system. In addition to the usability scale, we give the corresponding school grades for the system suggested by Bangor et al. in Figure 5. The overall mean SUS score is 71.92, which corresponds to a 'satisfactory' rating in the interpretation scheme of Bangor et al. However, the mean of SUS scores is different for the two groups of participants. The mean score of laypersons interested in history is 76.0, corresponding to a 'satisfactory to excellent' rating. The mean score of historians is 69.4, corresponding to a 'sufficient to satisfactory' rating.

5. Discussion

Most of the participants were able to use the functionalities of the virtual environment without any problems. They used the existing features and tested the boundaries of the rules for the interaction with objects in the virtual environment, e.g., when they tried to use the grab button to move the fragments or other fixed elements (e.g., the windows of the relation details). Some participants who used a VR application for the first time stated that they liked the playful approach to exploring historical data. Two participants who worked in an educational context suggested using eTaRDiS for knowledge transfer. Additionally, the participants stated that they learned something new. For example, one participant said *"didn't know the topic before, but that the connection exists, you could actually find out pretty quickly by reading the abstract"*. Three participants even looked for information unrelated to the task. One participant stated regarding the information in the characteristics of the fragment WWII: *"Yes, that is also interesting. How many casualties have there been?"* (usability study 2023, P9). Another participant looked at the other fragments related to Albert Einstein and stated: *"See what else is around [the fragment WWII]. Physical review. Munich. Mozart is irritating me a bit. Why is Mozart there?"* (usability study 2023, P4). A general observation was that the participants tried to explain the proximity of fragments or why fragments were visualized in the Fragmentarium with their prior historical knowledge. Therefore, the VR environment stimulated the reflection of their prior knowledge and the potential generation of new knowledge by discovering new relationships between historical fragments.

On the other hand, some participants commented after participating in the study they felt overwhelmed by the novel visual impressions in eTaRDiS due to their lack of VR experience. However, in the usability questionnaire, four participants gave five out of five points (corresponding to the answer 'fully agree') and six participants gave four out of five points for the question if they would like to use eTaRDiS more often. The other three participants gave three out of five points, corresponding to a neutral answer.

eTaRDiS allows users to take different perspectives on a topic, and the results showed that the participants found the implemented features helpful for this. In particular, the participants explored the adjacency of individual fragments to put them into context. In addition, they combined different aspects like the arrangement of the fragments, their detail windows, and their relation details.

The SUS scores of 76.0 (mean) for the group of laypersons corresponding to a 'satisfactory to excellent' rating and the mean score of historians of 69.4 corresponding to a 'sufficient to satisfactory' rating are a promising first rating for our prototype system.

Another perspective on the SUS scores is gained by considering the participant's digital media affinity (DMA), which users were asked to indicate in a self-assessment on a six-point scale, in addition to their prior VR experience. Our DMA scale ranges from *'very bad'* to *'extremely good'*. Nine participants who stated that they had a *'good'* to *'extremely good'* DMA have a SUS score between 62.5 and 92.5 (avg. 77.8) and four participants who rated themselves as *'very bad'* to *'not bad'* at using digital media have a SUS score between 45.0 and 70.0 (avg. 58.8). An independent samples t-test between the groups yields a p-value of 0.039 and an absolute t-value of 2.70, which shows a high influence of the stated DMA on the SUS score. We observed a positive Pearson correlation between DMA and SUS scores, r = 0.75. Based on the threshold values of Cohen[37], these effect size can be considered large. Two participants stated that they are regularly using VR applications. Their SUS scores are 70.0 corresponding to a 'satisfactory' rating and 82.5 corresponding to an 'excellent' rating. Eleven participants had no VR experience and had SUS scores between 45.0 and 92.5 (avg. 71.1). The broad range of SUS scores indicates that the lack of VR experience might not be of relevance regarding perceived usability. The p-value of 0.57 in an independent samples t-test for the SUS scores in the two groups with or without VR experience shows that the difference in the two groups is not statistically significant. Therefore, the quite high results in the group with VR experience can be seen as informed trends but need to be validated in further studies.

6. Conclusion and Future Work

We presented eTaRDiS, a VR exploration tool to support access to knowledge graphs for historians. In the presented user study, our tool achieved an overall SUS score of 71.92, corresponding to a 'satisfactory' rating. While the mean score reached with laypersons interested in history was quite high with 76.0, corresponding to a rating of 'satisfactory to excellent', the score for historians was lower with 69.4, corresponding to a 'sufficient to satisfactory' rating. Further analysis showed that participants who rated themselves as *'good'* to *'extremely good'* at using digital media have a higher mean SUS score of 77.8 compared to the participants who rated themselves as *'very bad'* to *'not bad'* at using digital media (mean SUS score: 58.8). The findings of the usability evaluation and the qualitative analysis of exploration patterns show the system's potential to be a valuable tool for allowing digital humanities researchers and laypersons to explore knowledge graphs. With eTaRDiS, such users are able to reflect on their prior knowledge as well as generate new knowledge by discovering new relationships between historical fragments.

In future work, we aim to integrate other historical databases to allow more diverse data to be explored using eTaRDiS. In addition, the feedback of the participants in the user study provided insights that we aim to integrate into eTaRDiS via further features specific to research purposes in historical studies.

Acknowledgements

This research is part of the project eTaRDiS[8] (Exploration of Temporal and Spatial Data in Immersive Scenarios), which is funded by the Federal Ministry of Education and Research (BMBF). Furthermore, we are grateful to all participants attending our user study.

[8] https://digital-history.uni-bielefeld.de/etardis/

References

[1] Hyvönen E. Cultural Heritage on the Semantic Web. In: Publishing and Using Cultural Heritage Linked Data on the Semantic Web. Cham: Springer International Publishing; 2012. p. 1-11. doi:10.1007/978-3-031-79438-4_1.

[2] Hyvönen E. "Sampo" model and semantic portals for digital humanities on the Semantic Web. In: DHN 2020 Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, CEUR Workshop Proceedings. vol. 2612; 2020. p. 373-8.

[3] Nünning V, Nünning A, Neumann B, Horn M. Cultural Ways of Worldmaking: Media and Narratives. Berlin, New York: De Gruyter; 2010. doi:doi:10.1515/9783110227567.189.

[4] Karadimas D, Somakos L, Bakalbasis D, Prassas A, Adamopoulou K, Karadimas G. Current and Potential Applications of AR/VR Technologies in Cultural Heritage. "INCEPTION Virtual Museum HAMH: A Use Case on BIM and AR/VR Modelling for the Historical Archive Museum of Hydra Greece". In: Moropoulou A, Korres M, Georgopoulos A, Spyrakos C, Mouzakis C, editors. Transdisciplinary Multispectral Modeling and Cooperation for the Preservation of Cultural Heritage. Cham: Springer International Publishing; 2019. p. 372-81.

[5] Skamantzari M, Kontogianni G, Georgopoulos A, Kazanis S. Developing a virtual museum for the Stoa of Attalos. In: 2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games); 2017. p. 260-3. doi:10.1109/VS-GAMES.2017.8056611.

[6] McCaffery J, Miller A, Vermehren A, Fabola A. The Virtual Museums of Caen: A case study on modes of representation of digital historical content. In: 2015 Digital Heritage. vol. 2; 2015. p. 541-8. doi:10.1109/DigitalHeritage.2015.7419571.

[7] Józsa O, Börcs A, Benedek C. Towards 4D virtual city reconstruction from Lidar point cloud sequences. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. 2013 05;II-3/W1:15-20. doi:10.5194/isprsannals-II-3-W1-15-2013.

[8] Ayers EL. Turning Toward Place, Space, and Time. In: David J B, John C, Trevor M H, editors. The Spatial Humanities : GIS and the Future of Humanities Scholarship. Bloomington & Indianapolis: Indiana University Press; 2010. p. 1-13.

[9] Nünning A. Making Events – Making Stories – Making Worlds: Ways of Worldmaking from a Narratological Point of View. In: Nünning V, Nünning A, Neumann B, Horn M, editors. Cultural Ways of Worldmaking: Media and Narratives. Berlin, New York: De Gruyter; 2010. p. 189-214. doi:doi:10.1515/9783110227567.189.

[10] Eakin PJ. How Our Lives Become Stories: Making Selves. Cornell University Press; 1999.

[11] Sicart M. Defining Game Mechanics. Game Studies. 2008;8(2).

[12] Derksen M, Weissker T, Kuhlen T, Botsch M. Towards Discovering Meaningful Historical Relationships in Virtual Reality. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); 2023. p. 697-8. doi:10.1109/VRW58643.2023.00191.

[13] Brooke J. SUS-A quick and dirty usability scale. In: Usability evaluation in industry. 1st ed. London: CRC Press; 1996. p. 189-94. doi:10.1201/9781498710411.

[14] Pirch S, Müller F, Iofinova E, Pazmandi J, Hütter CVR, Chiettini M, et al. The VRNetzer platform enables interactive network analysis in Virtual Reality. Nature Communications. 2021;12(1):24-32. doi:10.1038/s41467-021-22570-w.

[15] Buphamalai P, Kokotovic T, Nagy V, Menche J. Network analysis reveals rare disease signatures across multiple levels of biological organization. Nature Communications. 2021;12(1):6306. doi:10.1038/s41467-021-26674-1.

[16] Han D, Parsad G, Kim H, Shim J, Kwon OS, Son KA, et al. HisVA: A Visual Analytics System for Studying History. IEEE Transactions on Visualization and Computer Graphics. 2022;28(12). doi:10.1109/TVCG.2021.3086414.

[17] Cho I, Dou W, Wang DX, Sauda E, Ribarsky W. VAiRoma: A Visual Analytics System for Making Sense of Places, Times, and Events in Roman History. IEEE Transactions on Visualization and Computer Graphics. 2016;22(1):210-9. doi:10.1109/TVCG.2015.2467971.

[18] Krishnan M, Ober J, Pyzyk M. POLIS : Designing a Visualization Tool for the Research of Complex Sociopolitical Landscapes. Parsons Journal for Information Mapping (PJIM). 2014;6(2):1-9.

[19] Kidd J. "Immersive" heritage encounters. The Museum Review. 2018;3(1).

[20] Immersive Virtual Reality. In: Furht B, editor. Encyclopedia of Multimedia. Boston, MA: Springer US; 2008. p. 345-6. doi:10.1007/978-0-387-78414-4_85.

[21] Casu A, Spano LD, Sorrentino F, Scateni R. RiftArt: Bringing Masterpieces in the Classroom through Immersive Virtual Reality. In: Giachetti A, Biasotti S, Tarini M, editors. Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference. The Eurographics Association; 2015. p. 77-84. doi:10.2312/stag.20151294.

[22] Razum M, Sack H, Tietz T, Bruns O, Göller S, Scharm T, et al. TOPORAZ: Ein digitales Raum-Zeit-Modell für vernetzte Forschung am Beispiel Nürnberg. Information - Wissenschaft & Praxis. 2020;71(4):185-94.

[23] McIntire JP, Liggett KK. The (Possible) Utility of Stereoscopic 3D Displays for Information Visualization: The Good, the Bad, and the Ugly. In: 2014 IEEE VIS International Workshop on 3DVis (3DVis); 2014. p. 1-9. doi:10.1109/3DVis.2014.7160093.

[24] Wagner Filho JA, Rey MF, Freitas CMDS, Nedel L. Immersive Visualization of Abstract Information: An Evaluation on Dimensionally-Reduced Data Scatterplots. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR); 2018. p. 483-90. doi:10.1109/VR.2018.8447558.

[25] Kraus M, Weiler N, Oelke D, Kehrer J, Keim DA, Fuchs J. The Impact of Immersion on Cluster Identification Tasks. IEEE Transactions on Visualization and Computer Graphics. 2020;26(1):525-35. doi:10.1109/TVCG.2019.2934395.

[26] Wagner J, Stuerzlinger W, Nedel L. The Effect of Exploration Mode and Frame of Reference in Immersive Analytics. IEEE Transactions on Visualization and Computer Graphics. 2022;28(9):3252-64. doi:10.1109/TVCG.2021.3060666.

[27] McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software. 2018;3(29):861. doi:10.21105/joss.00861.

[28] Clark E, Celikyilmaz A, Smith NA. Sentence Mover's Similarity: Automatic Evaluation for Multi-Sentence Texts. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics; 2019. p. 2748-60. doi:10.18653/v1/P19-1264.

[29] Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics; 2019. p. 3982-92. doi:10.18653/v1/D19-1410.

[30] Unsworth J. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this. In: Symposium on Humanities Computing: Formal Methods, Experimental Practice, King's College. London; 2000. .

[31] Pacheco A. Digital humanities or humanities in digital: revisiting scholarly primitives1. Digital Scholarship in the Humanities. 2022 03;37(4):1128-40. doi:10.1093/llc/fqac012.

[32] Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, et al. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. Semantic web. 2015;6(2):167-95.

[33] Vaidya G, Kontokostas D, Knuth M, Lehmann J, Hellmann S. DBpedia Commons: Structured Multimedia Metadata from the Wikimedia Commons. In: Proceedings of the 14th International Semantic Web Conference; 2015. .

[34] Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z. DBpedia: A Nucleus for a Web of Open Data. In: Proceedings of the 6th International Semantic Web Conference (ISWC). vol. 4825 of Lecture Notes in Computer Science. Springer; 2008. p. 722-35.

[35] Van Someren M, Barnard YF, Sandberg J. The think aloud method: a practical approach to modelling cognitive processes. London: AcademicPress. 1994;11.

[36] Bangor A, Kortum P, Miller J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies. 2009 may;4(3):114–123. doi:10.5555/2835587.2835589.

[37] Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York: Routledge; 1988. doi:10.4324/9780203771587.