UNIVERSITÀ DEGLI STUDI
DI TRENTO

KNOW
DIVE

DEPARTMENT OF INFORMATION ENGINEERING AND COMPUTER SCIENCE
ICT International Doctoral School

Diversity Aware Visualization

Sajan Raj Ojha

Advisor

Prof. Fausto Giunchiglia

Università degli Studi di Trento

April 2018

I hereby declare that I have created this work completely on my own and used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

Trento, April 2018
Sajan Raj Ojha

Contents

List of Figures

List of Tables

Abstract

This thesis aims to address a significant issue related with the consumption of diversified data in the field of semantics and knowledge representation by using a framework which allows the data consumption in a generic, scalable and pleasing manner. The work proposes a mixed solution by splitting the issue into four subproblems: how to preserve the richness associated with the data; how to present information about an object in a single or multiple visualization contexts; and to provide a seamless exploration of interconnected entities; and how to design a tool that offers a better user experience.

A real-world object can have various representations which lead to data diversity. However, each representation captures a view (mostly partial) of an object. To preserve the richness associated with the data, we follow an entity-centric design approach. In this approach, we represent multiple datasets related to an object as an entity with various properties. An entity is then further categorized in a group according to its similarities or differences.

Our contextual model not only considers the transformation of objects as entities but also adapts to various visualization contexts. These contexts are space, list, timeline, and network. We design a multiview visualization framework that allows simultaneous presentation of entities according to these four defined visualization contexts.

To allow seamless interaction of data with the users, we emphasized on using a multilayered architecture where: 1)datasets are aggregated and stored using an entity-centric approach, 2) visualized in various contexts and viewpoints simultaneously according to the entity types and users' need. This adaptation is capable enough to facilitate presentation and exploration of diversified data according to users need.

To prove the feasibility of our framework, we applied it to visualize diversified data in various settings. Continuous interaction with the end users produced valuable feedback and essential design suggestions. Finally, multiple prototypes were evaluated with the end users to verify their usability. The results obtained were highly favorable.

Keywords [User Experience, Open Data, Entities, eTypes, User Centered Design]

Acknowledgements

This thesis represents a milestone not only for me but my family as well. There were ups and downs throughout the period I spent at University of Trento and in the Open Space 2, but the overall experience I got was terrific. Throughout the years, I learned that end users play an integral part in any solution. If you include them in the design, they will like the tool else they will use it but not like it. Furthermore, I only thought about entities, eTypes, and knowledge. For countless nights, I just dreamt about Entity-Centric representation. Redesigning a wheel which exceeds the performance of an existing one is not a bad design. Now, I can conclude entities definitely are the future!!! :-) Lastly, throughout the period, I encountered lots of remarkable people who I wish to acknowledge.

First and foremost, I want to thank my advisor, Professor. Fausto Giunchiglia for his immense support and teaching me the importance of entities.

I want to thank my colleagues from the KnowDive group. I specifically want to thank Subhashis Das and Maria De Las Mercedes respectively. Subhashis showed me a path towards data integration and management whereas Mercedes reminded me to stay inline in the field of Human-Computer Interaction. They were the perfect check and balance for my work.

Super big hug to my entire family who never complained to me about being away but always supported me with their unconditional love and told me to stay focused and persevere.

Thank you all.

Conventions

Throughout this thesis we use the following conventions.

Text conventions

Definitions of technical terms or short excursus are italicized.

Summary or essential sentence in a paragraph is also cited as a margin note.

Margin note

Source code and implementation symbols are written in typewriter-style text.

`myClass`

The whole thesis is written in American English.

Download links are set off as footnotes.

File: myFile[a]

[a] http://disi.unitn.it/folder/file_number.file

SPARQL queries are written in typewriter-style text.

```
SPARQL
```

Listing 1: SPARQL Squery

Publications

- Ojha, Sajan Raj, Mladjan Jovanovic, and Fausto Giunchiglia. ”Entity-centric visualization of open data.” Human-Computer interaction. Springer International Publishing, 2015.

- Giunchiglia, Fausto, Sajan Raj Ojha, and Subhashis Das. ”SemUI: A Knowledge Driven Visualization of Diversified Data.” Semantic Computing (ICSC), 2017 IEEE 11th International Conference on. IEEE, 2017.

- Das, Subhashis, Sajan Raj Ojha, and Fausto Giunchiglia. ”ATOM: Ontology Aware Transportation Model.” Semantic Computing (ICSC), 2017 IEEE 11th International Conference on. IEEE, 2017.

- Zhang, Hanyu, Sajan Raj Ojha, and Fausto Giunchiglia. ”Finding Errors in a Chinese Lexico-Semantic Resource Using GWAP.” Semantic Computing (ICSC), 2017 IEEE 11th International Conference on. IEEE, 2017.

- Karanjit, Sampada, Sajan Raj Ojha, and Subhashis Das. ”Process Ontology for Confectionery SweetBot.” Semantic Computing (ICSC), 2017 IEEE 11th International Conference on. IEEE, 2017.

- Sajan Raj Ojha, Subhashis Das and Sampada Karanjit. ”A Process Ontology For A Confectionery Service Robot.” International Journal of Semantic Computing, 12(01)

Chapter 1

Introduction

Information seeking is an expensive and time-consuming process in both personal and organizational contexts [Feldman, 2004]. In our life, we spend a substantial amount of time seeking, processing, communication and disseminating information out of data generated from various sources [Robinson, 2010]. Furthermore, datasets are generated at a speed and volume making it laborious to make sense of them. Apart from the size of information, a real-world object can have multiple representations. These representations can only capture a view of the object in a specific context. These issues of representation and context in data create diversity [Giunchiglia, 2016]. Diversity is unavoidable given the fact there are multiple data, producers and users [Giunchiglia, 2006]. Big and Open Data is a typical example where heterogeneity appears intrinsically. A plethora of various datasets can be used to create plenty of services in different domains such as transportation, geospatial, land management or environment for the end users [Dadzie and Rowe, 2011b]. Although data diversity provides tremendous benefits, it also creates challenges while designing a generic User Interface [Khalili et al., 2016].

Gathering information is a costly process.

A real-world object can be represented differently, creating diversity.

Semantic Technologies offer a solid foundation for aligning knowledge across various datasets. However, the problem of effective and efficient consumption of data remains. Facilitating the 'meaningful' use of organized knowledge is a challenging task as various stakeholders require quick and

Semantic technologies deal with the knowledge representation.

easy access to the data for immediate decision making.

The main contribution is towards the development of a user-friendly interface for diversified data.

The main contribution of this thesis is in the development of an engaging diversity aware visualization framework for a variety of users. Starting from the domain analysis, requirement gathering phase to the final User Interface generation, various stakeholders were included in the design and development cycle. The next section presents the motivation for the research in more detail and sets the stage for the thesis.

1.1 Motivation

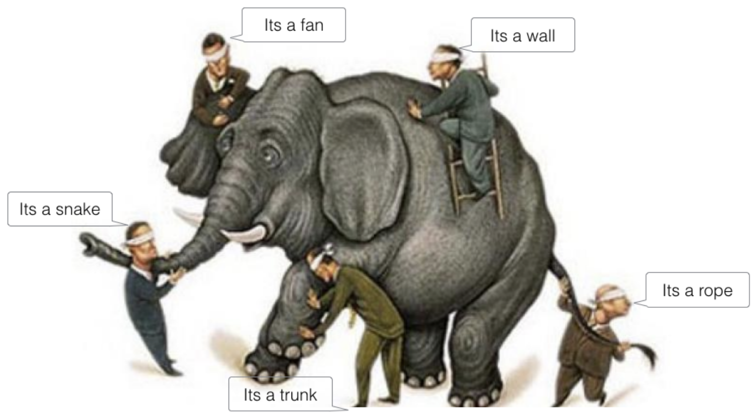


Figure 1.1: An elephant and five blindfolded men.

Diversity lies in the users' representation of the world. Depending on what she wants to see, she will perceive an object in its parts.

"*Reality is one, though wise men speak of it variously*" [Griffiths, 2007] as famously quoted by Paul J. Griffiths while explaining *universalist perspectivalism* in Rigveda (an ancient Hindu scripture). He used an example of five blindfolded men who were describing an elephant and has never seen it wholly as in Figure 1.1. One of them, touching its humongous leg, described the elephant to resemble a log while the other described it to be like a rope feeling its tail and so on. This reference can also be applied while visualizing any real-world object. There are different information associated with an object and variations are depending on the users' perspective. i.e., Diversity lies in the users' representation of the world. Depending on what she wants to see, she will perceive an object in its parts [Giunchiglia, 2016].

The example presented above is also relevant in the case of data representation and visualization. A significant increase in the generation of the data from a variety of data sources makes the knowledge organization and representation inevitable. Numerous organizations have already started integrating data from different streams like Open Data[1] showing their potential. However, the situation only reflects one side of the story. The linked data technology only allows joining of data disregarding context which leads to the information overflow. As the semantified information is intended to be finally consumed by the end user and not only by machine provides a challenge for the end user [Bizer et al., 2009a].

Various data producers generate a massive volume of data.

Semantified data should be easy to consume by various users to fulfill their need.

To cater this situation, a multitude of visualization methods and tools exists, ranging from standalone visualization systems, mashups, and browsers. However, most of the tools are either domain specific or lacks scalability.

Plethora of visualization tools and techniques exist.

A significant amount of work is done on the visualization of extensive collections of domain-specific data, and the results are impressive. However, their domain dependency results in incompatible with cross-domain datasets. There are various visual analytic tool[2]. However, the analysis is only limited to a particular domain or context.

Various domain-specific analytic tools exist, but they are not scalable.

The Semantic web community is facing a similar challenge in linked data consumption [Mazumdar et al., 2014]. Linked open data has already made a big step by adding both diversity and machine-readable semantics of data on the Web. However, the scale of the Web provides unlimited amounts of cross-domain data whose contexts impose various perspectives and interpretations from a human side. The data issue requires a general method to handle open data which is domain-independent and user-centered at the same time. Research by [Socrata, 2010] shows that people want more interactivity than the common method of down-

Research shows that common people want more interactivity than direct manipulation of data.

[1]Open Data are data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike

[2]http://bis.clients.talis.com/

loading and manipulating files. The major research challenge currently faced by the linked open data community is to figure out how to present the structured data intuitively and generically for common users [Bizer et al., 2009b].

1.2 Research Goals

This thesis explores innovative solutions to visualize and explore data defining any real-world objects following an entity-centric approach. There are two major part of the work- ***data representation*** and ***presentation for the real users***. Data has significant diversity such as multidimensional, language dependencies, nomenclature variations, variations in meaning and so on which creates hindrance while merging them. On the other hand, users are for whom the interface is designed so they can effectively comprehend and use the data which are currently hidden in various data silos and are not usable. These two factors make this thesis interdisciplinary work and to bridge the gap it requires an input from both the fields of semantics and Human-Computer Interaction.

Data are diverse making them hard to integrate across various domains. We design tools so that the final user can effectively consume the data irrespective of any technology used.

The research question this thesis attempts to answer is as follows:

How can diversified data be visualized and explored using an engaging, pleasing and contextual user interface?

We can see that the above question has some key elements underlined: **diversified data**, **effective visualization, and exploration**, **engaging, pleasing** and **contextual**. Based on this we can frame following research questions:

1. ***R1: How can the diverse data be presented clearly in the interface?***
   (a) R1.1: What are the implications of using diversified datasets from various domains in the visualization systems?
   (b) R1.2: What are the user implications?

(c) R1.3: What considerations and trade-offs are made to design both generic and a contextual interface?

2. ***R2: How can the user interface provide generic exploration and presentation of diversified data?***

   (a) R2.1: How to generalize the process of exploration of related data?

   (b) R2.2: How the relations be visualized generically?

   (c) R2.3: How the relations be visualized contextually?

   (d) R2.4: How the interaction be generalized?

3. ***R3: How can the engaging and pleasing interface be designed?***

   (a) R3.1: Can we design a pleasing interface catering the need of all the stakeholders?

   (b) R3.2: Is it possible to realize the need in practice?

   (c) R3.3: What are the compromises made during the design of such interface?

1.3 Contribution

The overall objective is to come up with a visualization framework which adapts to the various context and characteristics of the user and the diversified data. This situation required exploration of both literature and domain from various angle and context. For handling the data, we used an Entity-Centric approach and used Adaptive contextual visualization for data presentation and interaction. In this thesis, to answer the research goal we performed following steps.

We follow two step-based approach: 1.) Entity-centric data management and 2.) Contextual visualization.

- **Understanding diversity from multiple perspective.** An elaborate study of the existing tools for information visualization and semantics shows that there is a gap in the development of adaptive visualization

Contribution: Finding out impediments in the use of data in various domains and their categorization.

framework for diversified data in both contextual and generic way. This situation requires design and development of a schema that is flexible in accommodating data from various domains and also a User Interface which depicts this generic schema. Furthermore, data always provide impediments from multiple stakeholders' setting. Throughout the work, we tried to understand those various impediments in an Open Data setting in three different domains, namely, tourism, transportation, and cultural heritage. Furthermore, we also tried to understand the effectiveness of the visualization framework in other domain like education and health-care.

Contribution: Development of a framework for multiview and contextual visualization.

- **Reference model for Contextual Visualization for end users**
  Data depicting any real-world entities can be projected in any visualization that portrays them in space and time. However, an extensive literature review shows that the applications developed so far are extremely domain centric and not scalable to various incoming heterogeneous data setting. To fill the gap, we designed a framework which is both user-centric and shifts from property centric data structure to entity-centric where the visualization is guided by the context an entity is rather than the properties it possesses. Furthermore, to let the user understand the information from various angle, we devised coordinated multiview presentation and exploration of data.

Contribution: A prototype based on adaptive multi-layered architecture.

- **Multi-tiered architectural design approach**
  Designing an adaptive visualization that accommodates changes of the data at both design and runtime required a flexible architecture. For this, we adapted the software engineering principle by using a multi-tiered architecture where we segregate and encapsulates the functioning of every layer. For example, in our case, The knowledge layers only handles data and contains no business logic. Same is valid for visualization layer. The interaction between two layers is performed by a separate middle layer called mapping layer which handles all the data handling operations.

- **Use cases**

To validate our hypothesis, we designed various low fidelity prototypes by involving stakeholders in the design cycle. These prototypes were further verified and redesigned with the respective stakeholders. After the analysis phase, high fidelity prototypes were designed. We explored various domains in this work. The primary work was based on the Open Data setting where we developed a series of prototypes in various settings such as tourism and cultural heritage.

Contribution: Adapting the SemUI to various use cases.

- **Evaluation**
  Various stakeholders validated the work in the summative phase. In this phase, we asked multiple users from different domains to test and approve our application. The primary goal was to understand the effectiveness of the application. Furthermore, we also tried to find if the application is providing them a better experience and is helpful to them.

Contribution: Usability and user experience evaluation of various use cases with the end users.

1.4 Claims and Limitations

The work here was divided into two major parts: interaction and visualization of diversified data and the adaptation of the interface in many settings. Users were involved throughout the design and development phase. Furthermore, the results are then further summarized to generate a framework for visualization of diversified data.

In addition to the framework, some claims were made while answering the intended goal. They are:

1. ***C1: It is effective to visualize different types of entities in the same or different visualization context based on eTypes.***
   An entity consist of a defining class called an eType. Different eTypes exhibit different behavior visually. Facilitating users with various visualization contexts based on these eTypes will enhance the effectiveness of the visualization. This was also verified by the evaluation results (part V).

An eType guides visualization context of an entity.

2. ***C2: Seeing different properties of entities simultaneously facilitates easy exploration of related entities.***
   Different views provide different facets of the data giving the users the possibility to understand the overview of the data quickly. The evaluation carried with various stakeholders demonstrates that multiview allows an easy and intuitive mechanism for data exploration.

Simultaneously presenting various properties of an entity in different context provides a mechanism for flexible exploration.

3. ***C3: Involving users throughout the application design and development process will result in the system with the better user experience.***
   It is evident that an entity-centric approach has provided the foundation for managing large heterogeneous data. It facilitates the data handling by providing a generic and adaptive schema at design time and by providing a scalable, adaptive interface at runtime. The evaluation conducted with various users also justify that the interface was pleasing and easy to learn.

Involving users through the design and development of the interface is beneficial.

Though the framework was scalable and useful for a set of stakeholders, there were certain limitations from the beginning. They are as follows:

1. ***Real world Objects***
   The multiview framework only allows both interaction and presentation of the real world objects which has existence on the surface of the earth and can be accommodated into our generic schema. Any abstract objects which are nonphysical or out of this world are not considered in this thesis. Some examples that belong to this outcasted category are celestial objects and god.

We are not considering any abstract concepts like a god.

2. ***Number of users***
   Though we try to accommodate many stakeholders from various scenarios; the evaluation was only restricted to small population size. The assessment was done with internal stakeholders due to the limitation of time, hierarchical enterprise frameworks, language, and bureaucracies. However, we claim that given the time and authority this work can be scien-

The population size with whom the prototypes were evaluated was small.

tifically evaluated with a huge number of stakeholders.

1.5 Dissertation Roadmap

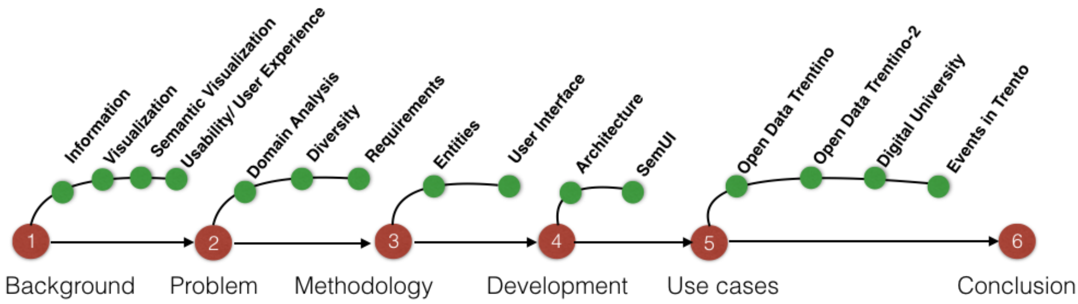


Figure 1.2: Thesis Structure

This thesis is structured into six major parts: background study, problem, solution, development of technology, use cases and conclusion as shown in Figure 1.2. Part wise description of the thesis is organized as below:

- **Part I** presents a brief review of the literature in the various fields that are connected with the work presented in this thesis. The proposed solution is highly interdisciplinary which required understanding the problem from multiple domains and facets. In regards to this, research done in the various fields is presented here. First, we explore the meaning of information in chapter 2. Topic visualization follows it in chapter 2.3 where we study various visualization tools and techniques that have been proposed. Similarly, we explore various visualization techniques in the Semantic Web and Open Data community in chapter 3. Finally, we explore the user-centric design approach in chapter 4. Here we also briefly describe usability and user experiences evaluation techniques used in our work.

Background
Information and Visualization- 2
Semantic Visualization- 3
UCD- 4.

- **Part II** motivates the problem by analyzing various domains in multiple settings to design a diversity-

Problem
Domain Analysis- 5
Diversity- 6
Requirements- 7.

aware visualization. In chapter 5, we list out various information and user needs; we explore various diversity issues in those domains in chapter 6. Finally, from domain analysis and diversity analysis, we generate a set of functional and nonfunctional requirements from stakeholders' perspective in chapter 7.

Methodology
Entity-Centric Approach- 8
User Interface- 9.

- **Part III** motivates the solution methodology towards designing a generic visualization framework for diversified data. Here we describe the notion of an entity and an entity-centric approach on the chapter 8. Chapter 9 follows the user interface design technique where we demonstrate the iterative interface design technique to caters users' need.

Development
Architecture 10.

- **Part IV** presents the technical realization of the system. Here in chapter 10, we describe a flexible visualization architecture called SemUI based on the work presented in chapter 8 and 9. Here we propose a multi-tiered flexible architecture where 1) datasets are organized as entities 2) entities are visualized according to various visualization context and 3) mapping layer that allows easy selection of entities from knowledge to visualization layer using various selection functions.

Use Cases
Trentino Entitypedia 11 and 12
Digital University 13
Events in Trento 14 .

- **Part V** illustrates the implementation of various use cases in multiple settings. We concretely present the use of entity-centric approach in various open data settings such as tourism and transportation in chapter Trentino Entitypedia 11 and Trentino Entitypedia II 12. We also present the use of SemUI as discussed in part IV in various other settings such as digital university 13 and Trentino events 13. Furthermore, we also perform usability and user experience evaluation of those systems.

Conclusion 15
Future work 16.

- **Part VI** wraps up the thesis presenting the summary of work, lesson learned and possible future works.

Part I

Background

Chapter 2

Information Visualization

In this chapter, we will try to understand information visualization as a foundation for the rest of the work. We will try to relate what information means and discuss various information exploration schemes. Finally, we will also discuss different visualization techniques aligned with our work. The goal of this study is to understand information visualization from multiple aspects.

2.1 Information

[Krikelas, 1983] defines information as "any stimulus that reduces uncertainty". [Marchionini, 1997] defines information as "anything that can change a person's knowledge". [Ackoff, 1989] defined information by contrasting with data, knowledge, and wisdom where data constitute raw observations and measurements; information represents purposeful messages, built out relationships and interconnections within data; knowledge is built by applying information and data; wisdom is evaluated understanding, created by reflecting upon knowledge.

There are various definitions of information and knowledge from multiple perspectives.

There are a few subtle difference between data, informa-

Several researchers have distinguished terms relating information, data, and knowledge.

tion and knowledge [Schreiber, 2000]. Data defines real-world facts discretely [Choo et al., 2013, Davenport and Prusak, 1998], by observations [Davenport and Prusak, 1997], and using symbols for interpretation [Quigley and Debons, 1999, Van der Spek and Spijkervet, 1997, Schreiber, 2000]. Whereas, information is associated with meaning, context and relevance to data [Van der Spek and Spijkervet, 1997, Choo et al., 2013, Davenport and Prusak, 1997, Schreiber, 2000], which varies according to the recipient [Davenport and Prusak, 1998]. It can also be defined as a flow of meaningful messages and facts that defines a condition or a situation [Nonaka, 2008, Wiig, 1994]. Knowledge, on the other hand, constitutes truths, justified beliefs, commitments, judgments, expectations, methodologies and so on [Wiig, 1994, Nonaka, 2008, Choo et al., 2013]. Knowledge comprises of valuable information from the human mind [Davenport and Prusak, 1997] and helps to answer how and why questions [Quigley and Debons, 1999].

Information can be defined as a flow of meaningful messages. Knowledge encompasses of information generated from the human mind and constitutes truth, beliefs or judgment.

2.2 Information Exploration

It is the process to satisfy an information need.

It is necessary to understand how humans explore information. There is ongoing work to understand various search patterns, strategies, and techniques to model the information seeking process [Mazumdar et al., 2014]. [Wilson, 2000] define it as the "purposive seeking for information as a consequence of a need to satisfy some goal" which involves interaction with various information sources. There is a big difference between information seeking and retrieval. Former relates to the acquiring whereas later relates to finding a system where they know where they are looking for [Marchionini, 1997].

Information seeking and information retrieval are two different processes.

Humans use various techniques for searching, browsing and monitoring information [Bates, 2002]. [Marchionini, 2006] describes exploratory search by a series of events as lookup, learn and investigate as in Figure 2.1. Lookup deals with carefully selected queries with precise results. Lookup is associated with building new knowledge and Investigate deals with the synthesis of the outcome.

Exploratory search is described as a set of activities.

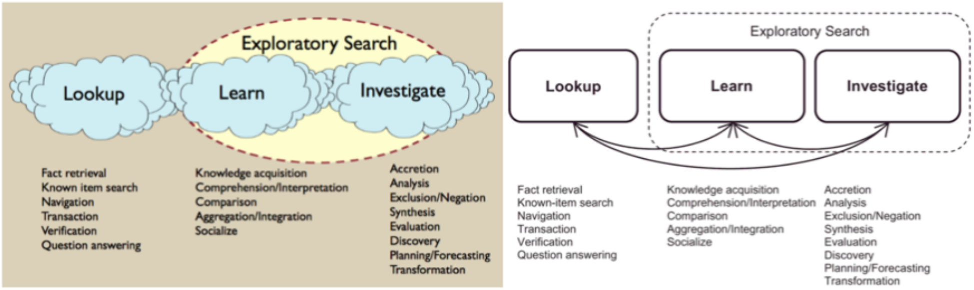


Figure 2.1: Exploratory Search and interplay of various tasks. Adapted from [Marchionini, 2006, White and Roth, 2009].

On the other hand [White and Roth, 2009] suggests that the process is not discrete but is the interplay between three tasks. Here, lookup, learning and investigating are interconnected. A user can start the search directly retrieving a specific fact (search engine). Whereas an exploratory search requires more user and system interaction.The exploratory search process might require querying and requiring the topic of interest and can last for a longer time duration depending on the task and user's motivation. This process of information exploration provides a useful pattern for us while developing our application.

Exploratory search can be an interplay of three tasks.

Exploratory Search depends on user's interest and can last longer.

2.3 Visualization

[Spence, 2001] defines visualization as the act of forming a mental model of an object in consideration. The process starts with the data which is represented using a set of valid metaphor and then presented on the interface. After that, an end user can then interact with all three states to make sense of the data.

During data visualization, data are represented and then presented using the continuous human interaction.

2.3.1 Visual Interaction Techniques

Before describing various tools, techniques, and frameworks for visualization in various contexts, it is necessary

Interaction techniques define the limitation of the visualization.

to understand what interactions are and what are the different methods associated with them. Interaction techniques reflect the usefulness and adaptivity of the solution [Shneiderman, 1996].

Here we present various interaction techniques based on the classification carried out by [Hearst, 1999]. The classification is similar to [Keim, 1997] classification but also considers user's input. Furthermore, we also adapt interaction from [Nazemi, 2016] which also aligns with our work. The techniques are:

Brushing and Linking

Allows highlighting and selection of an object.

Brushing and linking provides highlighting and selection of visual objects from various views. Various ways in which highlighting can occur are; by changing of color or size of the selected object.

Panning and Zooming

Allows changing the viewpoint of the data.

Panning and zooming provide the change of the viewpoint to the visualized data. It helps refine the part or area of interest by either moving the screen or zooming to the part of the screen in focus.

Focus plus Context

Preserves the surrounding context when an object is selected and zoomed.

When an item is zoomed, the surroundings also gets distorted. This leads to the loss of surrounding information. Focus plus context allows for zooming in the visual object by preserving the surroundings.

Semantic Zooming

Uncovers the meaning while preserving the context.

Ordinary zooming only zooms the visualized object which is not entirely informative if the object is highly multidimensional. Semantic zoom uncovers detailed information to encompass the context and meaning of a zoomed target.

Animation

It details the implication of a user's action.

It is more of an implication of user's interaction [Bartram, 1997]. For example, copying a file that shows the exact copying action. It improves user's interaction and understanding [Palmer, 1999].

Overview plus Detail

The interaction enables visualization of the object from various visual perspectives for example time and space which allows visualization of two various contexts simultaneously [Card et al., 1999].

It allows preservation of the context and detailing on the area of interest

Dynamic Queries This is a very successful approach where the user interacting will build the query adding multiple visual components satisfying the various information need. Dynamic queries benefit users as it is easy to learn and remember, flexible with more reversible actions, provide more control to the user, instant visual feedback, limits error and reduces the need of error messages [Mazumdar, 2013].

It allows visual interactive query formulation.

Direct Manipulation *Direct manipulation provides direct interaction with the user interface or visualization without the need of commands.*[Nazemi, 2016]. This technique is also somehow related to animation described above.

In this technique user's idea will be translated to an interaction command.

2.3.2 Data Classification

Here we provide various data classification that helps to define any interactions in the UI. [Card and Mackinlay, 1997, Card et al., 1999] proposed various data values by ordering as:

Data classification facilitates easy interactions.

1. Nominal: no values that can be ordered
2. Ordinal: value that can be ordered by relations between other values.
3. Quantitative: values that can naturally be order

Data can also be classified according to the dimensions. Understanding various data dimensions makes it easy for mapping the interaction with the visual information. [Card et al., 1999, Keim et al., 2006] define various data dimensions as:

Various data dimensions allow easy mapping to visualization and interaction.

1. 1-dimension: linear data;

2. 2- dimension: map data;
3. 3- dimension: real world objects;
4. Temporal- 1- dimensional data with start, end time;
5. Multidimensional- data in relational and statistical database;
6. Tree: hierarchical data;
7. Network: interlinked data;
8. Text and Hypertext: data with unknown dimensions and number;
9. Hierarchies and Graphs: data with relationships to other information entities;

2.3.3 Visualization techniques

Many classifications of visualization exist.

Various visualization technique classification has been researched which shows various criteria to provide efficient visualization. [Keim et al., 2004] classified various visualization techniques based on the display mode. They are:

1. 2D/3D: Uses plots for visualization, example 2D charts;
2. Geometric Transformation: Using statistical methods to find point of interest;
3. Icon based: map attributes of data to some representation as color or icons.
4. Dense pixel: Division of screen according to data volume.
5. Stacked: Present hierarchical data by embedding coordinate systems.

The visualization is sometimes guided by a data structure, attributes and their values.

[Ward et al., 2010] Presented a taxonomy of visualization. In this visualization, the main basis is the data structure which consists of data attributes and values.

Visualization technique is mostly dependent on data structure and data value as discussed above[Keim, 1997]. These visualization techniques defined by [Keim, 1997] allows seamless representation of data.They are:

Data structure and values guide visualization.

Geometric Visualization
This visualization technique transforms data value ranging from one dimension to multi-dimension data into a visual representation. These visual representations can be either point, line or an area. They are described as:

Transforms data into 2D or 3D visual representation.

- Points: It is the visual projection of data values as points into a graphical representation. An example of such transformation is using Cartesian coordinates for point positioning.

It is a projection of points.

- Lines: It is a visual projection of data values as lines in the graphical representation. There is a difference between point-based visualization and line as it can visualize curvatures, slopes, crossings and so on [Ward et al., 2010].

It is a projection of lines.

- Areas: It is a visual projection of data values as polygons [Ward et al., 2010]. Shape, the color of the polygons visualized can be used as an extra dimension for visualization.

It is a projection of polygons.

Graph Visualization
Data contains relationships, which are stored in different structural formats. Simply, it is the presentation of nodes and links between data. There can be various ways in which a graph be presented. [Herman et al., 2000] performed a detailed investigation of a graph-based visualization. Discussing various visual layouts on which a graph can be displayed. Some arrangements are spanning tree, 3D layout, hyperbolic layout and so on. There can be multiple graph visualization techniques such as:

Presentation of nodes and edges.

1. Hierarchical: Parent child relationship is depicted in this kind of visualization.

They are taxonomical visualization.

2. Arbitrary: They visualize information that does not contain any particular structure.

Visualize non hierarchical graphs.

Data value is represented using a pixel.

Pixel Oriented Visualization Big data might not be easy to visualize, hence his approach is more suitable where a pixel on the screen is used to represent a data value. They are categorized in two categories as:

Projection on a single 2D screen.

1. Single-window: In this situation, an entire window is used as a single projection pane. This allows mapping of multi-dimensional data into a single two-dimensional space.

A single screen is divided into many screens.

2. Multi-window: In this technique on a single screen is further divided into various other screens. This allows an enhanced view to the data point.

2.4 Summary

We motivate the definition of information.

In this chapter; we discussed what is an information. We understood information seeking behavior highly depends on the information need and environment [Marchionini, 1997]. Seeking not well-defined questions or only gathering overview of information lead to browsing. Exploration allows discovering, learning and investigating a new set of information.

We explored various visualization schemes.

In this chapter, we discussed different visualization techniques. We explored information visualization from multiple perspectives and motivate how various data types and their dimensions guide the visualization process.

We explore various semantic visualization in the next chapter.

These visualization techniques motivate and provide a platform for visualizing and exploring semantic data.

Chapter 3

Semantic Visualization

This chapter is based on the background work presented on top of the previous chapter. We will try to articulate how various visualization and interaction metaphors were developed in the field of semantics.

Tim Berners-Lee, father of World Wide Web also proposed a Semantic Web (SW) as an extension of the existing web [Berners-Lee et al., 2001]. The technology at that time was only built for humans to read. The primary focus was on the content presentation and layout which was rendered by the browser. Processing the information by the machines was hard. The invent of SW shift the paradigm, allowing computers to perform intelligent operations. To make the content understandable by machines, Resource Description Framework (RDF) which is based on Web Ontology Language (OWL) is used [McGuinness et al., 2004]. [Mizoguchi and Ikeda, 1998] listed various levels of use for an ontology as :

Web technologies like HTML was only focused on humans.

RDF uses ontologies using OWL.

1. Level1: As a common vocabulary for communication;
2. Level2: As a conceptual model;
3. Level3: As a backbone information;
4. Level4: Answering competence questions;
5. Level5: Terminology standardization by providing meanings;

6. Level6: Mean of structural organization of Databases
7. Level7: Reusing knowledge base's knowledge
8. Level8: Knowledge base reorganization

Linked data momentum is allowing various research parties to link data using RDF. The SW technologies are only limited for machine.

Then the concept of Linked Data (LD) arise where many data publishers are pushing data into more machine understandable formats [Bizer et al., 2009a, Mazumdar, 2014]. However, the SW and LD, do not offer proper support for the consumption of resources from a human perspective. They are only able to identify and relate resources [Berners-Lee et al., 2006] uniquely.

3.1 Ontology Visualization

We explore different mechanism on which ontologies can be visualized.

Before diving into the visualization of semantic data, we explore various ontology visualization tools. Various tools have been developed with the goal to visualize ontologies [Katifori et al., 2007]. Ontology visualization considers elements such as classes, instances, attributes, and relations. Ontologies are usually visualized as hierarchies using the following UI controls:

- *Intended list* - Represents an ontology as items of the hierarchical list.
- *Tree control* - Represents an ontology as nodes of a tree. Nodes can be easily expanded or retracted to adjust the information according to the user needs.
- *Scaling mechanism* - Allows for zoom-in/zoom-out functions. It is used in combination with the other UI controls.
- *Context extraction mechanism* - Extracts a particular view on the ontology usually by selecting part of its content and the visual perspective.
- *Symbologies* - This is a technique where an ontology is represented as a set of symbols on the visualization plane. Symbols follow a particular scheme in terms of shape, size and color.

The main issue with the techniques above is that they are all used to visualize data in the form of the hierarchical structures. This makes them less usable when visualizing actual objects, i.e., they do not make a clear separation between the ontology (classes) and instances. As such, they are used mainly by knowledge experts.

Most of the presentation are hierarchical without the separation of schema and data.

3.2 Consuming Semantic Data

Plethora of application tools are designed following various design approach providing various interaction mechanisms to query and visualize semantic data. We list a few of them below:

Various visualization technique exists.

1. Textual: In such kind of systems, users have to interact with the data by formalizing textual queries. Such systems are only confined within the use of technical users [Kaufmann et al., 2007, Lopez et al., 2006, Damljanovic et al., 2011, Bhagdev et al., 2008, Lei et al., 2006, Tummarello et al., 2007, Hogan et al., 2011, Cheng et al., 2008].

   A user needs to formulate textual queries.

2. Faceted: Facilitates users to build queries dynamically using hierarchical information searching approach, where they start the building the information in a generic way and drill down to until they have found the required information [Mazumdar et al., 2014, Wilson et al., 2006, Hyvönen et al., 2005, Hildebrand et al., 2006].

   Allows building of information using various facets.

3. View support: The tools in this category consist of ontology and other interactive components. Most of the view based system are graph based [Mazumdar, 2014, Athanasis et al., 2004, Catarci et al., 2004, Elbedweihy et al., 2012, Kaufmann and Bernstein, 2010].

   Consist of visual interaction components.

3.3 Visualization of Semantic Data

These days, it is easier to link up and query semantic data.

Various presentation tool to interact with Open Data exists.

A plethora of tools and applications exist to cater user needs [Mazumdar et al., 2014]. Here we present some tools and techniques that have already been developed for easy visualization of semantic information.

Consists of a tabular list.

1. Text-based: Lists out triples as a table or lists. No other interaction technique is defined [Mazumdar et al., 2014].

Consists of a tabular list of links.

2. Enriched-Text: This is an advancement over the previous design where data are presented as links and also contains icons, colored materials [Mazumdar et al., 2014, Lopez et al., 2006].

Consists of an enhanced list with data structure visualization.

3. Text and Image: In this visualization technique, a data structure can also be viewed for example images [Hildebrand et al., 2006, Wilson et al., 2006, Hyvönen et al., 2005, Petrelli et al., 2009, Glaser et al., 2004].

Supports advance visualization per domain.

4. Basic visual encoding: This visualization technique is an advancement over the previous one allowing various views based on the data features [Berners-Lee et al., 2006, Becker and Bizer, 2008].

Assists multi interactive visualization.

5. Advance visualization: Supports multiple visualization and interaction of the data [Mirizzi et al., 2010].

We compared multiple LOD application from a different perspective.

Based on the above features and the work stated in Chapter 2, we give a comparative analysis of the Open Data visualization systems found in the literature. The comparison is made against a set of requirements adapted from [Dadzie and Rowe, 2011a]. The criteria are:

If interactive components are used in the tool.

1. ***Interactive Visualization (IV)***: Refers to the use of interactive representation through different kind of widgets (such as images, buttons, and maps). Here human perception is considered in understanding the complexity of the data structure and discovery and analysis of the data [Amar and Stasko, 2005, Bizer et al., 2009b, Card et al., 1999, Hastrup et al., 2008].

A possibility of finding different relations.

2. ***Relations (R)***: Denotes different kinds of relationships within one or multiple datasets to understand

the data and find new data [Becker and Bizer, 2009, Berners-Lee, 2011, Fekete et al., 2008, Halb et al., 2008a].

3. ***Details on demand (DOD)***: Deals with exposing different level of details for the data as needed [Card et al., 1999, Hastrup et al., 2008].

   See data from various level of details.

4. ***Scalability (S)***: Denotes the ability to manage and link large amount of heterogeneous data, which are loosely coupled [Becker and Bizer, 2009, Card et al., 1999, Halb et al., 2008b].

   Able to manage large or small heterogeneous datasets.

5. ***Filtering (F)***: Refers to an ability to suppress irrelevant information and focus only on information relevant to a particular context or user session [Becker and Bizer, 2009, Card et al., 1999].

   Maintain focus on specific part of information.

6. ***History (H)***: Describes the ability to record a history of interaction allowing the user to review or retrace paths, undo/redo their actions [Bizer et al., 2009b, Heath, 2008].

   Ability to retrace back what has been done.

7. ***Faceted exploration and Navigation (FN)***: Refers to a flexible mechanism that enables setting particular context for search and exploration, and switching to another context based on relations the user explores during the session [Cao et al., 2010, Shneiderman, 1992, Farazi, 2010].

   Ability to switch from a context to another.

8. ***Domain Independency (DI)***: Means that applications are not coupled with the specific domain and can exploit a wide range of underlying datasets [Giunchiglia et al., 2012b, Berners-Lee et al., 2008, Bizer et al., 2009b, Hastrup et al., 2008].

   Able to accommodate a plethora of domains.

9. ***Target User Group (UG)***: Means we differentiate three target user groups [Dadzie and Rowe, 2011a, Mazumdar et al., 2014].

   The desired audience of the system.

   (a) Common User (C): End-user who does not have any background in ontologies.

   Generic people.

   (b) Tech User (T): End-user who has an understanding of the underlying technology and ontologies.

   Computer scientists.

Data scientists.

(c) Domain Expert (D): End-user with expertise in data of a particular domain. They might or might not know an underlying technology.

Should emphasize on the friendly presentation of data.

Table 3.1 gives a brief comparison of existing Open Data applications. When visualizing open linked data to make them accessible for common end users not familiar with Semantic Web languages, it is important not to present data as URIs or triples but in a more user-friendly way.

Most of the applications still present on the enhanced textual form.

It is hard to drill down on the data. Furthermore, they are contextual and has no filtering support.

Although some systems provide high-level interactivity and emphasize different kinds of relations in visualized datasets, they mostly use RDF to describe the data. From the usability viewpoint, it adds an extra effort to the end user to interpret and understand what is being visualized. Applications like Dipper[1], Disco[2], Marbels[3], Piggy Bank[4], Sig.ma, URI Burner[5], Zitgist[6] and IsaViz[7] employ complex notations for visualization. They mainly use knowledge graphs that quickly become cumbersome as users drill down the data. Tools like OpenLink[8],RDF Gravity[9], RelFinder[10], SIMILE/Exhibit [11]) and LESS[12] provide good visualization support, but are highly contextualized and does not support better filtering and tracking history of interaction.

Tools do not allow context switching.

Link data are automatically produced containing various errors.

Regarding cross-domain support, most of the systems allow visualizing in a predefined context, without the ability to switch to another context based on the relations user perceives during the session. This limitation also comes from the automated production of Linked Data which raises the problem of the accuracy and completeness of the datasets. In particular, incorrect or missing values, or incorrect links

[1]http://api.talis.com/stores/iand- dev1/items/dipper.html
[2]http://www4.wiwiss.fu- berlin.de/bizer/ng4j/disco
[3]http://www5.wiwiss.fu- berlin.de/marbles
[4]http://simile.mit.edu
[5]http://linkeddata.uriburner.com
[6]http://dataviewer.zitgist.com
[7]http://www.w3.org/2001/11/IsaViz
[8]http://lod.openlinksw.com/ode
[9]http://semweb.salzburgresearch.at/apps/rdf- gravity
[10]http://relfinder.dbpedia.org
[11]http://www.simile-widgets.org/exhibit/
[12]http://less.aksw.org

make it hard to correlate data [Dadzie and Rowe, 2011a].

Table 3.1: Comparing Functionality of different tools that exist in the Linked Data Community.

Application	IV	R	DOD	S	F	H	FN	DI	UG
Dipper	L	L	L	L	L	L	L	L	T
Disco	L	L	M	L	L	L	L	H	T
Marbels	L	M	M	L	L	L	L	H	T
Piggy Bank	L	M	H	L	L	H	H	H	C
Sig.ma [Tummarello et al., 2010]	L	L	M	H	H	M	L	H	TU
URI Burner	L	L	M	L	L	L	L	H	T
Zitgist	L		H	M	H	M	H	H	T
DBPedia Mobile [Becker and Bizer, 2009]	H	L	H	M	H	H	L	M	CU
Fenfire [Hastrup et al., 2008]	H	H	H	H	M	L	L	H	All
IsaViz	H	H	H	H	H	H	H	H	C;T
LESS	H	L	M	L	L	L	L	H	All
OpenLink	H	H	H	M	H	H	L	H	All
RDF Gravity	H	H	H	M	H	L	L	H	All
RelFinder	H	H	H	H	H	L	L	L	All
Tabulator [Berners-Lee et al., 2008]	H	M	H	L	L	M	L	H	TU
SIMILE/Exhibit	M	L	L	L	H	L	L	L	C
Rhizomer [García et al., 2008]	L	M	H	L	M	L	H	L	T
Sgvizler [Skjæveland, 2012]	H	M	L	L	L	L	L	L	T
LODWheel [Stuhr et al., 2011]	L	H	M	M	M	M	L	L	T
Calluna [Otjacques et al., 2012]	H	H	H	L	H	H	L	M	C

Low (L) - The system has no or low support for the category.
Medium (M) - The system supports the feature to some extent.
High (H) - The system has full support for the category.
Common User (C), Technical User (T), Domain Expert (D)

All the factors above require a certain level of expertise, either in ontologies or domain of the visualized data or even knowledge of specific UI notation. This is the reason why a majority of systems is well suited for more experienced users. Thus, LOD community is still struggling to come up

Most of the tools are designed and focused on technical experts and domain experts.

with a common visualization tool that captures open data diversity and unexpectedness in a user-friendly way.

3.4 Summary

We discuss various semantic visualization.

In this chapter, we discussed various semantic visualization tools and techniques. Different data visualization tools have been developed and used in the domain of semantics [Lanzenberger et al., 2009]. Although general guidelines for visualizing diversified data have been proposed [Shneiderman, 1996], in practice they are still not fully addressed. Our work lies at the intersection of data visualization, ontology visualization, and knowledge visualization.

We explore the user-centric design in the next chapter.

In the next chapter, we discuss various user-centric design technique and principles that are aligned with the thesis for generating user-friendly visualization technique for diversified data.

Chapter 4

User Centered Design

In this chapter, we will try to understand various user-centered design techniques and work done previously to support user engagement. This chapter will also provide the canonical foundation for the rest of the work. Here we will discuss how end users can be engaged in various design stages. Finally, we will discuss different evaluation techniques which we also implemented in our work.

Users are actively involved in the design and development of the product to understand their tasks and requirements better [Donald and Draper, 1986]. It helps to get a better understanding of their needs and goals, it grounds the product to the user's reality and also generates the feeling of ownership to them [Rogers et al., 2011].

Users always stay at the center of any product.

4.1 User Engagement

[Gould and Lewis, 1985] discuss three principles of system design as user and task, empirical measurement and iterative design. In the first principle, designers try to understand behavior user characteristic which helps then design the tool. In the second, users are engaged in using various simulation tools and prototypes. The results are then measured. The third principle focuses on redesigning con-

Various user engagement techniques in the design cycle have been proposed in the literature.

sidering the changes suggested by the users. [Rogers et al., 2011] suggest five principles as:

User's tasks and goals are the driving force behind the development.

- It is important to understand users' alignment with the technology for better support of user tasks to fulfill the goal. Here technology should not be imposed.

Understand Users' behavior.

- It is important to understand users' behavior and their intentions while they are performing their tasks.

Capture users' characteristics.

- It is important to understand the physical and cognitive ability of the users. This helps to design a better system.

Consult users throughout the design and development phase.

- Users are consulted throughout the design of the tool. Designers have to be aware of the user's sentiment and be respectful towards the users.

Consider users in all the design decisions.

- Though the users are not designing the product; the designers have to be aware that the product is being designed for them.

Users are always at the center of the design of a tool.

There are various ways in which users can be involved in the product design.

[Rubin and Chisnell, 2008] illustrates that users are always the critical part while developing a system. In the first part of the design, understanding users' context, goals, environment, and objectives are the primary concerns. Whereas the outer circle focuses on various tasks such as detail, organization, flow, and content. [Abras et al., 2004] illustrates multiple ways in which users can be included in the design and development of a solution. [Vredenburg et al., 2002] surveyed how UCD is employed by various practitioners while designing the systems.

From the survey, we can see that users are mostly involved in iterative design and evaluation. Though the approach is beneficial there is hesitance from the companies to adopt the methodology.

Users are highly considered on iteration phase followed by assessment and task analysis. Though the method is highly useful, it is not well adapted to the design and development scenario. . Considering the initial involvement cost, the methodology is not well adopted in the software development community [Nielsen, 1994] therefore they are not adequately exploited [Curtis et al., 1988]. Furthermore, lack of qualified and motivated workforce also creates a barrier for mass use [Rosenbaum et al., 2000].

Just like software design cycle, UCD can also have four stages as study, design, build and evaluate where the iterations can be made many time [Harper et al., 2008, Mazumdar, 2014]. This flexibility allows designing better products.

The design cycle can be iterated many times and can be entered at any phase.

[Harper et al., 2008] introduced another phase called *understand* which aims to perceive various human values. This is a multidisciplinary approach requires understanding from various other fields. This stage acts as a catalyst initiating discussion across various stakeholders working in multiple domains. The output of this stage is the input of the study stage.

Understand stage is multidisciplinary and acts as a catalyst for various stakeholders.

A multidisciplinary analysis has to be performed at this stage. A close study of how a user interacts with an existing solution, their work habits, practices and environment in their everyday like is studied [Mazumdar et al., 2014]. This output that deals with a user's various interaction factors acts as an input to design stage.

The study stage perceives a user's interaction practice to perform their job.

Here we design various artifacts like sketches in different environment and settings. A technical study of the existing technologies, hardware and software consideration to design better systems is also understood at this stage.

Designers and developers identify various goals.

Possible implementation solutions are designed and developed at this stage. A new system can be developed, or the existing system can be extended to cater user's need.

Various low and high fidelity solutions are designed.

It estimates the outcome of the build stage where the system performance verified. Understanding how the new solution is perceived and helping the user is the primary goal of this stage.

We evaluate the effectiveness of the system.

In this stage we perform various quantitative and qualitative evaluation of the solution. Multiple dimensions like usability, user experience, accessibility, acceptability, efficiency, effectiveness and so on are validated at this stage.

Qualitative and quantitative assessment is made.

4.2 Evaluation

In our work, we have tried to incorporate users as much as possible. Furthermore, we have followed various evaluation patterns in our work. Here we discuss different usability and User Experience evaluation done in our work.

Consideration has to be made to understand the resources and their constraints.

Software development is very time consuming and expensive process, so consideration has to be made to understand the resource and constraints (like cost, time, system representations, participants, test environment and equipment, data capture tools, analysis tools, security and so on.)while evaluating the system [Mazumdar et al., 2014].

Evaluation can be conducted in many ways.

Heuristic evaluation [Nielsen and Molich, 1990] allows coming up with positive and negative aspects of the interface. Similarly, in cognitive walk-through [Wharton et al., 1994], the solution is evaluated by a group with a predefined sequence of actions. Think aloud study [Ericsson and Simon, 1980] generates a verbal report to better understand the experimenter's interpretation. Contextual inquiry [Beyer and Holtzblatt, 1999] helps understanding users' day to day activity. CASSM (Concept-based Analysis of Surface and Structural Misfits) [Blandford et al., 2008] focuses on user's conceptual model and its fitting in the system developed.

We separately studied User Experience evaluation.

Out of many evaluation possibilities, we explored various ways to measure User Experience. In the literary work, there are many similarities between usability and user experience. Here we focus on two ways on how User Experience (UX) is defined.

4.2.1 UX Dimensions 1

[Winckler et al., 2013] surveyed various factors regarding usability and User Experience and proposed these concrete dimensions as:

How is aesthetics perceived by the user?

1. ***Visual and Aesthetic Experience***: Shows how aes-

thetics affect user perception of the system. It deals with the pleasure that users gain from the immediate perception of the system [Hekkert, 2006, Lavie and Tractinsky, 2004, Hassenzahl and Tractinsky, 2006, Alben, 1996].

2. ***Emotion***: Describes effective side of UX regarding feelings and emotions elicited as an outcome of the interaction with the system [Hekkert, 2006, Alben, 1996, Desmet and Hekkert, 2007, Mahlke and Thüring, 2007].

How user gets emotionally active seeing the system?

3. ***Identification***: Addresses the human need to express oneself through the interaction with the system [Marc, 2007, Jääskö and Mattelmäki, 2003].

How are human need expressed?

4. ***Stimulation***: Relates to the quality of the system to encourage a user to use it [Hassenzahl, 2005, Sheldon et al., 2001, Karapanos et al., 2010].

How the system encourages and discourages users?

5. ***Meaning and Value***: Denotes the quality of the system to reflect or represent values that are important to the user [Hekkert, 2006, Jääskö and Mattelmäki, 2003].

How is the system important to the user?

6. ***Social relatedness/coexperience***: Describes the pleasure that comes from the social interactions [Jääskö and Mattelmäki, 2003, Gaver and Martin, 2000].

How is social co pleasure obtained?

4.2.2 UX Dimensions 2

[Laugwitz et al., 2008] came up with more generic User Experience dimensions. They are:

1. ***Attractiveness*** relates to the overall impression of the system.

How beautiful is the system?

2. ***Pragmatic***It measures the usefulness of the system across three dimensions. Efficiency deals with how easily the system can be used. Perspicuity deals with the familiarity of the system. Finally, dependability deals with the user's feeling of the control of the system.

Usability of the system.

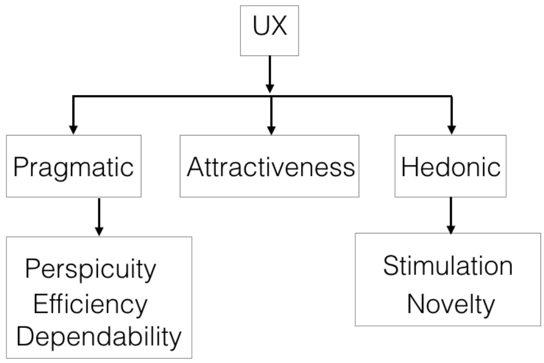


Figure 4.1: Various User Experience dimensions. Adapted from [Laugwitz et al., 2008].

User stimulation while using the system.

3. ***Hedonic*** Users stimulation is measured in this dimension. It is composed of stimulation (deals with the feeling of excitement while using the system)and Novelty (The innovativeness of the system).

4.3 Summary

Various factors associated with User Centric Design is studied.

In this chapter we investigated various techniques followed in UCD. User-centered design is a very crucial methodology that has to be implemented in the design and development of the solution. The methodology will facilitate designing and defining better products. We implemented various techniques to involve users in the design and development of the tool.

In the next part we formulate various problems aligned with the thesis.

These literary review show potential gap. The concept of designing a generic application catering to various users' need is still in its infancy. Furthermore, these work performed will guide us to formulate a goal for our thesis.

Part II

Problem

Chapter 5

Domain Analysis

The background study related to various visualization techniques illustrated studies on human visual perception, existing approaches, and existing systems and methods. Our research shows that there is still a considerable gap for visualization that accommodates data variation and adapts to various contexts in the field of semantics. In this chapter, we will try to analyze the outcome of the background study and derive our requirements for diversity aware contextual visualization which provides adequate support for our design. This chapter discusses how various requirements were outlined for addressing the research questions as mentioned in section 1.2, and how approaches were developed to address the needs. Following an entity-centric approach, the steps were continuously refined after every iteration of evaluation following several subsequent redesign stages.

The potential of contextual visualization in the field of semantics is still not yet explored.

Here in this chapter, we try to understand needs of various stakeholders.

[Bentivogli et al., 2004] defined a domain as "an area of knowledge which is recognized as a unitary. It is characterized by the name of a discipline where a particular knowledge area is developed (e.g., chemistry) or by the specific object of the knowledge area. In this phase, various roles and functions are studied to understand the workflow on how different user perceives their existing environment and how they perform their task according to their work contexts. The process of analyzing various domain continued throughout the research period. We completed series

We analyze workflow from stakeholders working in different domains and try to understand how they can achieve meaningful insight with the help of a visualization tool.

We conducted interview sessions to understand the possible stakeholders and their need.

Interviews were conducted to understand both domain and user.

of interview sessions to understand the possible outcomes. The outcome of the interview sessions was to understand how our solution provides a convenient way to perceive and explore the data providing various stakeholder with meaningful insight. The interview sessions were structured in two distinct stages: understanding information need; and stakeholder 's visualization need. The list below summarizes various questions that were asked to the stakeholders. They are:

1. What is the role of the stakeholder?
2. What existing system the stakeholder use to perform her task?
3. What kind of information does the tool provide?
4. What the stakeholder liked and disliked about the tool?
5. Where did the stakeholder learned the skills to perform the job?
6. How often she uses the system?
7. What are the problems faced while fulfilling the need?
8. How are the problems solved?
9. What are the stakeholder's primary and secondary goals?
10. How the stakeholder measured her success?
11. How to make the tool better?

We split the generic questions into two different categories.

These generic questions helped to set the stage to find both information and visualization need from the stakeholders. We split the answers into two categories to obtain the clue about data structure and information needs; and visualization needs. Along with these generic questions, stakeholders were asked domain related questions as well. These questions were asked individually to the stakeholders.

5.1 Open Data Setting

Open Data (OD) initiative is pushing forward on releasing enormous datasets to be used by the general public from various government, private and public sectors.This vast numbers of datasets could open up new opportunities for application developers and general people helping in the exploitation and generation of a new business model [Ferro and Osella, 2013, Manyika et al., 2013, Vickery, 2011]. Three most lucrative categories for application developers and the general public are tourism, transportation, and cultural heritage.

Open Data allows both transparency and innovation of data benefiting general public.

OD also allows generation of a new business model.

5.1.1 Tourism and Transportation Domain

The World Tourism Organization, a United Nations agency behind the promotion of responsible, sustainable and universally accessible tourism, has come to view tourism as a "key driver of the global economic recovery, and a vital contributor to job creation, poverty alleviation, environmental protection and multicultural peace and understanding across the globe[1]". A reason behind this assumption may be that, over the decades, tourism has become tourism has experienced continued growth and deepening diversification, in this way, representing one of the primary sources for a large number of developing countries. The increase in the growth can be the critical factor to the better economic growth. Tourism-related data includes locations data, data about service providers (organizations), buildings and so on. The data about the tourism services contains various properties and are extremely language dependent.

UNWTO claims tourism is the vital contributor on the increase of the economic growth of the country.

Tourism data is diverse and is highly contextualized to a particular language.

Similarly, transportation also has high potential to fulfill the generic stakeholders need. The report from European Commission shows that the frequency of people commuting every year using different public transport has increased significantly[2]. Out of various transportation categories, pub-

Public transportation plays a vital role in the daily life of a person.

[1] http://www2.unwto.org/
[2] https://goo.gl/ZToIpm

lic transportation data has the highest impact on a citizen's life. The data in this context majorly deals with the crucial information such as timetable, frequency and other services associated with the means of transportation. Meanwhile, in the recent days, the nature of travel pattern has become more complicated.

Currently, the data cannot support transportation multi modality suggesting the best way of transport from point A to B.

From the data perspective, publicly available data contains various issues. The data is available in the multiple formats, follow different standards, and language creating complexity while integrating them. This complexity in data also raises several issues: (1) Various public transportation services providing different transportation means, creating confusion among ordinary people to choose best travel option. (2) Currently, existing solutions mostly handle a specific mean of transportation, i.e., the user has to check multiple applications and then analysis which combination best suits her. (3) Interlinked multidimensional complexity due to overlapping of temporal (timetable) and spatial data (stops and stations). (4) The geographic information related to transportation is not interlinked with other services such as gas station or hotel.These issues create obstacle while making complex queries.

Open Data Trentino contains datasets from various categories like tourism and transportation. Six different stakeholders were interviewed.

Open Data Trentino[3] was the playground which allowed amalgamation of data provided from multiple producers in a plethora of setting like tourism, transport, recipe, provincial government and so on with the aim of providing information service to the end users. The initiative was pushed forward by the province of Trento under the provincial law to promote information society and the digital administration and for the distribution of free software and open data, formats [Bedini et al., 2014]. Six stakeholders from various areas in the province of Trento (consumers, provincial officials, and data scientists) were interviewed to understand their needs. A set of 15 open-ended questions were prepared. The stakeholders were interviewed individually. Each interview session lasted for an hour. During the interview session we tried to understand the perspective of open data, the data dissemination process and the visualization need.

[3] http://dati.trentino.it/

Information Need

The main aim of the province was the information dissemination in a scalable and human-usable way using compliant data. During the initial days, there were more than 800 datasets from a plethora of providers belonging to various categories like locations, organizations, buildings and so on. These datasets were generated using multiple formats, with or without any documentation and were unusable from a general people perspective. Data scientists primary concern about the information was the alignment of data with a generic model matching the need of both consumers and data providers. From the consumers' point of view, the situation was more complicated. They required a way in which they can be informed about a various point of interests. Currently, they used Google for all the information need. The information search pattern is shown in Figure 5.1: 1)They start the process by using a Google search. 2) They search for various POI in the area (for example a restaurant). 3) They try to see if the POI has a website. 4) They get information about the POI. 5) They come back to search. 6) Then they use another application (Google Maps) to search for the location. 7) Again go back to the search interface. 8) Then click the transportation details from her place to the POI either re-clicking on the map application or through Transportation website. 9) They take note of the mode of Transport. These to and fro operation is time-consuming. Furthermore, the way they search for information only points to specific POI and don't show the relationship between various objects in detail. Another issue was language since tourists didn't speak the native language, they wanted a multilingual tool which suggests POIs in different languages.

Province wanted a way in which open data in various categories to be easily consumed by the consumers

Data scientists' primary motivation was to find a flexible way to generate a generic schema so that various heterogeneous information can be easily be integrated.

Consumers wanted an easy way to interact with the data to obtain vital information.

Visualization Need

The context of this thesis is in the generation of generic visualization approach; the interest was to understand the basic visualizations schemes the communities distinguished. The provincial employees used their own internal tool for data management. However, they used CKAN[4] to release the data. The CKAN open data portal was only a way to

Province and data scientists wanted a generic way for discovering datasets following an integrated approach. The visualization requirement was not their primary concern.

[4]https://en.wikipedia.org/wiki/CKAN

collect resources and was not able to directly serve consumers' data need. The tool had the possibility to visualize data on a map or either as charts but it required an extra operational step to be performed by the user. Similarly, data scientist wanted a way in which they can easily visualize the interconnected data after aligning various datasets to a uniform generic schema. They also needed a faceted exploration allowing semantic jumps from one object to another. Furthermore, they also wanted a way in which the can easily find wrong links within the data. Primarily, general consumer interviewed never knew anything about Open Data; they had a problem understanding how Open Data could be beneficial for them. To make them recognize, we used a series of scenarios explaining the benefit of using Open Data. After they became aware, they suggested an integrated and interactive user interface. They expected maps and detail information about the POI they were searching. They were also unsatisfied with the tabular presentation of data. They mentioned that the chance of missing vital information is high in such presentation.

Consumers wanted an integrated interactive UI where they could quickly get information about various touristic destinations and transportation in one place.

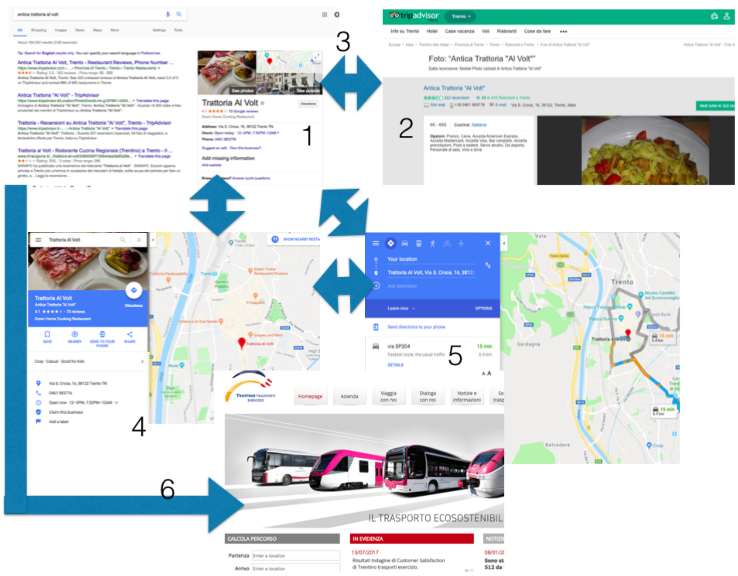


Figure 5.1: Information search process of an ordinary user, looking for a restaurant.

5.1.2 Cultural Heritage Domain

Museums around the world are releasing cultural heritage data as Linked Open Data with a mission to reach new audiences such as educators, developers and the general public by collaborating data across various museums. 14 different museums formed a collaborative to allow inter museum exploration of data [Knoblock et al., 2017]. The triple data store contained about 9.7 M triples out of which there were 153,453 objects, 20,389 artists and 18,765 related parties as shown in the table 5.1. Figure 5.2 shows the triples for an entity. Consuming information from the triple browser is hard for the general audience as she has to perform multiple back and forth operations to get complete information about an object.

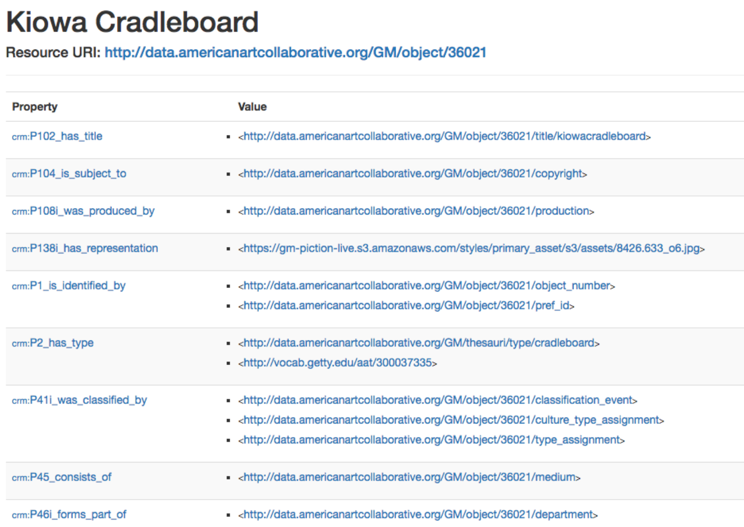
The main aim of the cultural heritage project was to reach a broader audience by integrating data across 14 museums. Triplestore stored more than nine million triples from various artists and objects. However, it was hard to browse them.

Table 5.1: List of museums with their respective artists and artifacts.

Museum	Artifacts	Actors
Indianapolis Museum of Art	22314	2077
The Walters Art Museum	801	181
Gilcrease Museum	20904	1198
Crystal Bridges Museum of American Art	1691	513
Archives of American Art	15681	6956
The Amon Carter	6421	772
Autry Museum of the American West	193	114
Dallas Museum of Art	2229	649
National Museum of Wildlife Art	2208	375
Princeton University Art Museum	13314	2866
National Portrait Gallery	16829	12552
Colby College Museum of Art	8217	2005
Smithsonian American Art Museum	42651	8896
Total	153,453	39,154

In the context of museums, an artist can have one or more than artifacts located in one or many museums. Similarly,

An artist can produce many artifacts which might be located in many different museums.



Kiowa Cradleboard

Resource URI: http://data.americanartcollaborative.org/GM/object/36021

Property	Value
crm:P102_has_title	• <http://data.americanartcollaborative.org/GM/object/36021/title/kiowacradleboard>
crm:P104_is_subject_to	• <http://data.americanartcollaborative.org/GM/object/36021/copyright>
crm:P108i_was_produced_by	• <http://data.americanartcollaborative.org/GM/object/36021/production>
crm:P138i_has_representation	• <https://gm-piction-live.s3.amazonaws.com/styles/primary_asset/s3/assets/8426.633_o6.jpg>
crm:P1_is_identified_by	• <http://data.americanartcollaborative.org/GM/object/36021/object_number> • <http://data.americanartcollaborative.org/GM/object/36021/pref_id>
crm:P2_has_type	• <http://data.americanartcollaborative.org/GM/thesauri/type/cradleboard> • <http://vocab.getty.edu/aat/300037335>
crm:P41i_was_classified_by	• <http://data.americanartcollaborative.org/GM/object/36021/classification_event> • <http://data.americanartcollaborative.org/GM/object/36021/culture_type_assignment> • <http://data.americanartcollaborative.org/GM/object/36021/type_assignment>
crm:P45_consists_of	• <http://data.americanartcollaborative.org/GM/object/36021/medium>
crm:P46i_forms_part_of	• <http://data.americanartcollaborative.org/GM/object/36021/department>

Figure 5.2: Hyperlinks represent an object on a triple store browser.

an artifact can have more than one artist associated with it. Moreover, artifacts are also related to each other. Lastly, an artist is also related to many artists as either being influenced by or was influencing. This notion will generate a dense mesh of interconnected entities. The relationship web between various entities is shown in Figure 5.3.

An artist and artifact can be connected with many different artists.

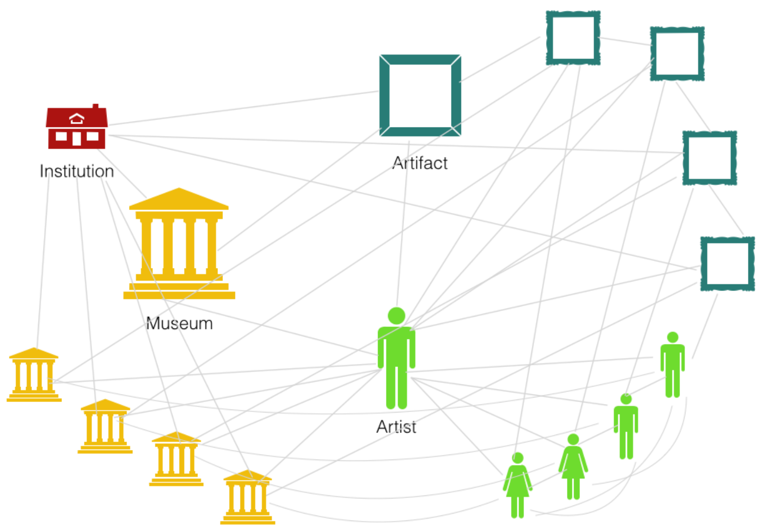


Figure 5.3: Relations between arts, artist, and museums.

The SPARQL queries to fetch an entity from the triple store were complex with many optional fields. This was due to various modeling choices from the data modelers and the complexity of CRM ontology. An example of a SPARQL query to fetch an artifact is shown below.

The structure of the SPARQL queries was complex.

```
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema\#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX schema: <http://schema.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core\#>
CONSTRUCT {
 ?entity\_uri a crm:E22\_Man\-Made\_Object ;
    rdfs:label  ?primary\_title\_text ;
    foaf:homepage   ?website\_url ;
    schema:genre  ?style\_name ;
    crm:P24i\_changed_ownership_through ?acquisition\_event ;
    crm:P102\_has\_title ?alternate\_title\_class ;
    crm:P102\_has\_title ?primary\_title\_class ; }
```

Listing 5.1: A snippet of SPARQL query to fetch an artifact.

To allow natural exploration of information, the American Art Collaborative (AAC)[5] developed a prototype application called 'Browse' (Figure 5.4 which allowed visualization of museum data to the end users. They surveyed with members of six museums to identify the goals. The people involved were curators, registrars, educators, and outside researchers who searched for alternative ways to present various museum-related information in an integrated way. The application is still on its early stage. The application used modular aggregation visualization components called toybox. The data was converted to a presentation based on the presentation profile of the toybox.

Browse application allowed users to view museum data relating person, artworks, museums and related parties.

The application was designed using modular visualization components called toybox.

Information Need

The primary information need was to make various stakeholders such as researchers, curators, and the general public aware of integrated cultural heritage data. AAC's principal requirement was to have scalable, easy-to-configure, simple to work with libraries for creating the links so that

The primary information concern for AAC was the accuracy of the information released.

[5] http://americanartcollaborative.org/

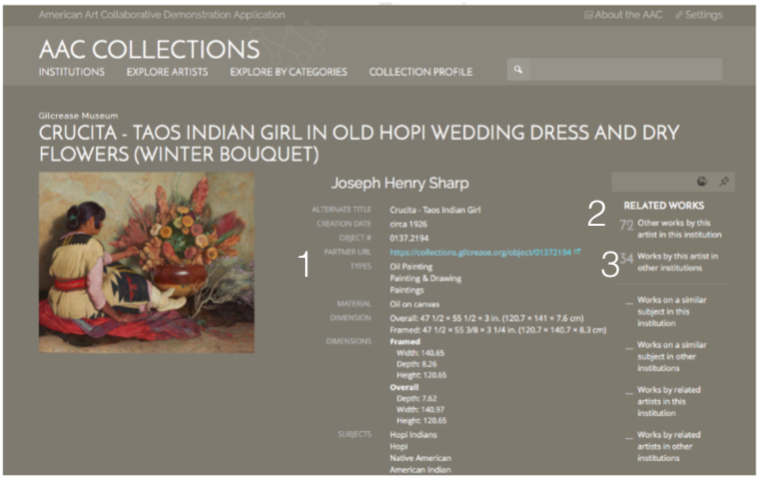


Figure 5.4: Screenshot of Browse application. Image adapted from [Knoblock et al., 2017]

they could release accurate information which was their primary concern. They used LOD technique to store the information. KARMA tool was used for the data conversion. CRM[6] ontology was used as the data model. This ontology is very complicated, and data scientist had their mapping techniques which hindered designing a reliable general query across museums. Data scientist were interested in a simple schema that executes various SPARQL queries [Knoblock et al., 2017].

Data scientist wanted a more comfortable uniform ontological model rather than CRM.

Visualization Need

AAC wanted accurate and reliable information to be presented in a human-readable way.

AAC wanted a tool where these LOD can be visualized in an understandable human form so that the managers in the museums can verify the data accurately. For easy rectification and presentation of data, browse tool was designed. As in Figure 5.4 information was presented on a list. There were other pieces of information such as 'related artworks' and another 'associated museum' for an artifact, but those were not visualized at once. They had to navigate back and forth to find related information. Browse application had a feature to show related artworks from an artist, but the visualization was limited.

[6] http://www.cidoc-crm.org/

5.2 Education Domain

Universities around the world have to maintain a massive amount of information related to people. They also host various intellectual resources such as books, papers, patents, courses, projects, budgets, and thesis. In the current scenario, all these information are scattered among various data silos. In a typical day to day scenario, these data need to be merged to fulfill the specific need. These merging processes are a particular tailored application with limited scope. Though this technique is highly practiced, it is not scalable and provides hindrance in search and analysis of data [Maltese and Giunchiglia, 2017]. There are many universities[7] in the world relying on standard Semantic Web technologies and tools to extract, convert and store data in RDF, as well as to query it using the SPARQL query language. However, the coordinated data model which facilitates explicit data interaction is seldom seen and also possesses issues ranging from localization to entity identification.

University datasets are linked using tailored software which is only good enough to achieve specific needs.

Many universities use SW techniques, but they lack centrally coordinated model.

[Maltese and Giunchiglia, 2017] defines Digital University as *"as a set of essential resources, methodologies, and tools appropriately organized to support universities' users efficiently"*. The digital university offers institutional services by providing innovative ways to present information to the end users by not deviating from standards which promote natural transformation and use of data.

The concept of Digital University is to provide interoperable institutional services to all the university joining the collation.

Various interview sessions were held to recognize how different users in the organization use University of Trento's official portal[8] to fulfill their requirements. Furthermore, we tried to understand the difficulty they faced while using the website. Finally, some suggestions about how to improve the current situation were also accessed. In this regard, six different kinds of users were chosen. They range from professors, administrative staffs, technical staff, secretary to students. A set of questions were asked to the interviewee. Furthermore, they were asked to show the problems they were facing in the existing portal. Finally,

Interview sessions were conducted with six participants to understand their needs.

[7]https://focus.library.utoronto.ca/, http://hub.hku.hk/

[8]www.unitn.it

they were shown the new interactive prototype designed [9]. They were allowed to think aloud while using the prototype. The entire interview session lasted an hour and was audio recorded. The interaction with the new prototype was also video recorded. All the users gave their consent. The motive of the interview session was explained to all the participants.

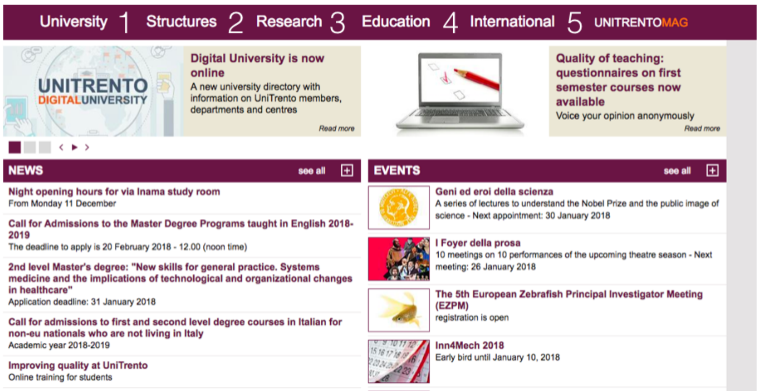


Figure 5.5: Digital University

Information Need

Technical staff search for colleagues and departments.

The university portal is shown in Figure 5.5. Various stakeholders use different parts of the portal. Technical Staffs mainly use myUnitn[10]. They have the technical expertise to interact with the system. They sometimes search for information about colleagues and departments. They mostly know who they are searching for except some rare cases they might have to look it up. They use address book feature of the phone rather than searching for people in the portal which came as a surprise. However, sometimes it takes a time to search for people on the phone (as they have to continuously push up and down key on the telephone.). So, only in these case, they search for people in the portal. Or instead, they will ask the nearby colleague.

Administrative staffs search for people and articles.

Administrative Staffs are the people who are completely used to with the system. They may or may not have the technical expertise. They fully understand the architecture and has no specific issue with the portal. They search for

[9]The interactive mock-up was developed using Balsamiq (https://balsamiq.com)

[10]www.unitn.it/myunitn

both articles (not publications) and people. They only face problems whenever the system is upgraded to a new interface. They don't usually prefer drastic changes in the way information is conveyed. Similarly, professors may or may not have technical skills but are well acquainted with information systems. Searching for colleagues and people is easy as they also have a large group of people they have to contact to. They also look for different kind of documents like publications, projects, calls etc. They feel the information provided on the websites are messy and cluttered up. The information is rather inconsistent so it has to be revised. There are many interlinked departments and the information is not completely consolidated. Secretary are helpers for professors. They have very low skills related to information systems. For many secretarial jobs, they have to use the portal. Usually, they search for people, departments and articles as asked by the respective professor. They also perform other duties like registering grades for students, manage funds, create and update new calls for students and so on. Their work mostly requires them to access the portal. Students have good knowledge about information systems. They have a range of requirements like searching for thesis, courses, choosing the most appropriate department, searching the appropriate supervisor related to their work. So, they need to query about different organizational information and didactic information at the same time.

Professors search for colleagues, lectures, publications, and projects.

Secretary search for departments, people, and articles.

Students search for professors, thesis courses, and departments for didactic purposes.

Visualization Need

All the users mentioned that the search and exploration process is not fluid. There is no uniformity on the information presented in a page. Furthermore, some pages open in new tabs some doesn't. The multilingual aspect is not maintained throughout the portal. Furthermore, it is not easy to change the language from one to another causing a major problem for the non-native stakeholders. The way in which articles are distributed is not clear. Currently, the articles presented on the portal are not associated with any people or department. So, when they are found, they will directly open as a PDF file (mostly) so a user cannot directly associate from where the file is from. Even if the documents are associated with people or department, it is not obvious. It is difficult to retrieve information effectively at a

The search and exploration process is not interactive. All of the pages are not multilingual.

Information is not segregated along time and creates hindrance while exploration.

short period and it is hard to distinguish periodic information. The organization is very messy and creates a problem when users don't know what they are searching for. Lastly, different departments in the portal use their own technique of presentation and are not standardized creating confusion for the end users.

5.3 Health care Domain

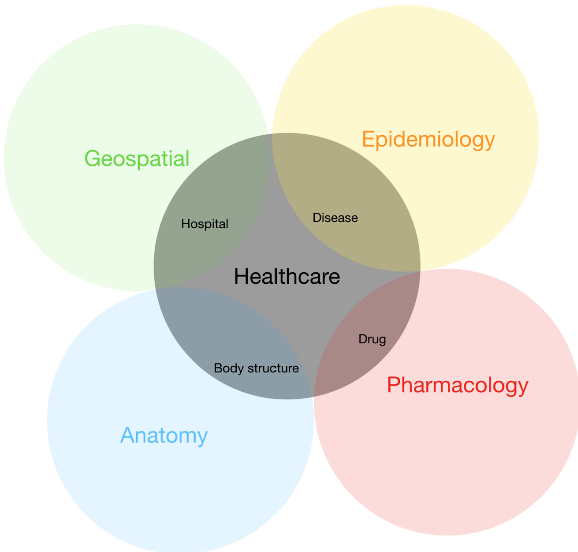


Figure 5.6: A relationship between various subdomains.

Sharing of consented health-related information do offer benefits across the jurisdiction.

Europe's aging problem requires collocation of data internationally to solve the problem associated with chronic diseases.

The health domain is a vast domain where semantic Interoperability of data is of great concern. Here the main issue is the exchange of health-related information which is explicitly shared between different stakeholders [Benson] across the international federation. Controlled sharing of health data where data can be accessed and shared across jurisdictions for the research does offers significant benefits. This is needed because Europe's population is aging[11]. This certainly means more chronic disorder. To tackle this situation to find more cures to these chronic diseases an international collation is a must. Figure 5.6 illustrates how health-care domain intersects with other domains.

[11] http://slideplayer.com/slide/7788722/

The hurdle in achieving semantic interoperability is due to variation in the terminological representation. These representations are either designed as a classification or as a corpus fulfilling different users' need. The difference in terminology used by various coding systems and also a lack of correlation among local coding systems and international coding systems limits semantic interoperability. Furthermore, Ontology-based approach along with the standard controlled vocabulary helps tackle issues where the explicit formal specification is provided to interact with different systems [Héja et al., 2008]. Making health records meaningful will only be possible if we link the Electronic Health Record (EHR) to sound clinical knowledge and then use natural language in the user interface as suggested by the IMIA working conference on clinical terminology which was held in 1984 [Benson]. This meaningfulness enables effective meaning-based retrieval. The primary stakeholders in this scenario are federation (this shares information across many jurisdictions), commercial organizations (organizations like pharmaceutical companies), and finally the data scientists.

Terminological variation and various standards of health practices create easy data collocation.

Using ontological approach may aid linking diverse set of data.

Information Need

Federation want statistically significant analyses performed rapidly and reliably on populations more significant than a single jurisdiction. Commercial organizations want to work with local authorities to provide services that support more extensive, the faster experiments based on real data. This allows them to process the information and provide an alternative result quickly. Lastly, data scientist wants to share experimental results with other two stakeholders promptly. She also has to be aware that the information shared is as accurate as possible. For this, she has to first get anonymized data from the research clients process them, integrate them by understanding the basic terminology used and analyze them.

Federation want to perform rapid analysis of the data.

Commercial organization wants faster data analysis.

Data scientist process anonymized data to produce high-quality analysis.

Visualization Need

The major need of any visualization tool is for data scientists. Currently, the anonymized datasets are scattered across many CSV files. Though the process is repetitive, this hinders an effective exploration of related files. She has to open files individually, check their data structure, under-

CSV based visualization is restrictive for a data scientist.

stand the terminology used, map them across many files and then only she will be able to use the result.

5.4 Summary

The information and visualization impediments are uniform throughout the domains. Data are scattered throughout the countries. No uniform data interaction and presentation.

Here in this chapter, we analyzed various domains and tried to figure out the common issues across them. Though multiple stakeholders are working on various tools, the information and the visualization need is somehow aligned across these domains. Firstly, datasets are scattered across various data silos and are represented in a diverse set of ways. These representation hinders generating a uniform generic schema. Similarly, all the existing system lack a consistent presentation of data. Mostly, data are presented in a tabular form. Furthermore, the exploration of the related information is limited throughout the domains. Terminological distinctions and language intolerance is also another standard issue across stakeholders. In the chapter below, we will try to elicit all the diverse set of problems that we have encountered.

Chapter 6

Diversity

The domain analysis that we performed with the stakeholders in chapter 5 proved to be beneficial. It helped us to sort out all the information and visualization needs. To authenticate what they said, we investigated various datasets from a plethora of domains to better understand different semantic and user related impediments. The investigation involved consulting various data portals[1][2][3], proprietary datasets (from digital university and medical domain), listing the datasets that we intended to use, understanding their formats, finding terms used in the datasets to denote the different kind of features that matches our need, and finally integrating them.

We extensively studied a plethora of datasets from various domains.

The work described here was motivated by the following research questions:

We try to find out both data and visualization diversity.

1. What are the various form of diversity which affects the data?
2. What are the visual impediments? and
3. How can we design a generic User Interface with a minimal effort for data whose diversity is unpredicted at design time?

[1]http://opendata.arcgis.com
[2]https://www.europeandataportal.eu
[3]http://data.americanartcollaborative.org/

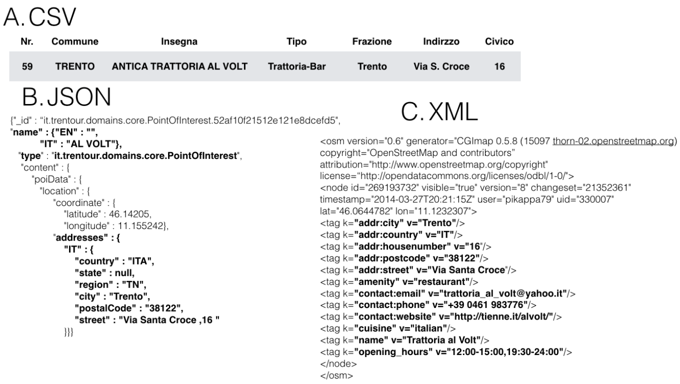


Figure 6.1: An excerpt from open data catalog that illustrates how a real-world object (Restaurant Al Volt) is represented in three different way.

Our investigation shows that there are two major diversity issues. Namely:

The same object is represented differently either because of variation in schema or properties.

These different naming might hinder data integration.

1. ***Knowledge Diversity***
   Meaning by this is the fact that the same object can be described using different schema and properties. Variations between knowledge, producers, and users, and new and old knowledge creates a barrier while establishing a certain level of connectivity between people, software agents, and IT systems to appropriately enable information exchange [Uschold and Gruninger, 2004]. For example, in Figure 6.1, the same restaurant is referred differently in various datasets, as it is referred with multiple names, has various sets of properties, used different terms to represent it as a point of interest, restaurant or building and so on. Collocating data with above mentioned multiple intricacies is hard.

The same object can be visualized using multiple ways.

2. ***Diversity in Visualization***
   Meaning by this is the fact that there can be the various ways to visualize the same entity or a set of entities, possibly visualized in different ways, where each of the various ways is the function of its local perspective. This situation becomes more prominent when we have data that are highly heterogeneous in terms of context. For example, in Figure 6.1, after the data

is converted into an entity, the same restaurant can be shown on a map (its location), list (cuisines), timeline (opening schedules) or a network (owner). Similarly, diversity in visualization occurs when the user interface fails to predict all possible combination in which the data can be presented. The vast majority of applications offers diversified visualizations at the interface level using various mash-up techniques where data are presented in a predefined domain dependent way than often ignoring other relevant facets of the data. For instance, while visualizing the information about the restaurant Al Volt in Figure 6.1, there are issues like which dataset should be prioritized and what are the properties to visualize. Similarly how to visualize contextual information such as the relation between the restaurant and its owner or menu. These issues create a barrier for easy understanding of the information [Zuiderwijk et al., 2012].

Properties that are important should be considered and then presented to the user based on her desired context.

Apart from these diversity datasets possess other problems such as:

Data Volume and formats

The size of these datasets is huge with a large number of columns. Furthermore, the datasets are stored in various formats such as XML, XLS, CSV JSON and so on. It is assumed that end user knows how to deal with multiple formats and use them [Davies, 2012]. This extra effort creates a barrier in utilizing the datasets by most of the users. In figure 6.1, record B[4] is stored in XLs format, whereas record C[5] is written as JSON or CSV.

Datasets are huge and they are produced in various commercial or open formats.

Domain Specific

Existing datasets are tailored to specific domains and aimed to be used in different contexts. They often use several ways to interact with the data. This situation requires users to use several tools differently and can increase cognitive effort.

These datasets are developed to be used in a certain context and might not be generic.

In the section below, we will try to understand these two

[4]http://goo.gl/iH0xfD
[5]http://goo.gl/77UbzF

forma of diversity in detail.

6.1 Knowledge Diversity

Name Diversity

There are many names for an object.

Name is a language unit by which an object is recognized [Giunchiglia et al., 2012a]. In the real world, an object can have many names. For example in Figure 6.1, the same restaurant is named differently across various datasets. In Figure 6.1A, it is *ANTICA TRATTORIA AL VOLT*, in Figure 6.1B, it is *AL VOLT*, whereas, in Figure 6.1C, it is named *Trattoria al volt*. Name variation exists in any real-world object and this phenomenon is unavoidable. This multitude of representation of names is easily detected by humans but hard for machines to recognize.

Property Diversity

Multidimensional data are complex with many properties of different nature.

Properties are the features that describe an object.Organizations use various properties to represent a real-world object from the different point of view. These properties only reflect a partial view of the object. For example, in Figure 6.1A, as CSV file gives information that Al Volt is a typical restaurant and bar, in Figure 6.1B, the JSON file gives more detailed attributes, including latitude and longitude, where, finally, on 6.1C, the XML file provides information also about the opening hours.

Diversity in Generality

The same object can be represented in various ways.

In most cases, the same real-world object is grouped into different categories (this is the source of many inconsistencies) [Guarino, 1998]. These classifications overlook basic fundamental ontological principles. For example, in Figure 6.1C, Al Volt is classified as a restaurant, in Figure 6.1A as a typical bar, wherein Figure 6.1A as a point of interest. End users will have confusion while searching for information in a category.

Value Diversity

The value of a property can also be diverse.

Dataset publishers use labels or tags to represent the same concept and value as a function of their local needs. For example, in Figure 6.1, the term civic number and house number are used interchangeably. Similarly, the price value of a meal in a restaurant datasets is sometimes published without mentioning the currency.

Diversity in Language

Datasets exist in many languages. This situation requires a prospective user to have knowledge of the languages to understand the data. For example, in Figure 6.1, various properties for the restaurant Al Volt are in Italian whereas in Figure 6.1B and C they are in English.

Datasets exist in multiple languages.

6.2 Diversity in Visualization

Lack of generic presentation and interaction interface

Most of the datasets are fragmented across various portals and there is no definite way to interact and visualized the content for the end users. In most of the situation, they have to download the dataset and then manipulate themselves to understand the view. This restricts the coverage of the data with the end users and can only be used by the users within a specific community. For example, scientific data are mostly consumed by scientists.

There are no generic interactive visualization interfaces for diversified data.

Inefficient search and exploration of datasets

In general, data portals lacks a proper interface that allows for an efficient interaction and exploration[Agrawala et al., 2011]. There exists no adequate way to index or adequate categorization of data. There is data duplication, obsolete or invalid or incorrect. Furthermore, finding a specific tuple in the dataset requires an extra effort for a user. The data may overlap, inaccurate, or incomplete (as shown in figure 6.1). An interface that allows for multi-faceted interaction and supports exploratory search is still missing[Hearst, 2006].

There are no standard indexing or categories hindering easy exploration of datasets.

Complex data correlation process

The process of discovering related data tuples within the datasets is rather tricky. Stakeholders correlating the data might find it cumbersome to connect the dots. A user lacking domain knowledge might also find it hard to treat the data. Looking at the record B from figure 6.1, we can notice

Correlating data can be hard as it is difficult to find relations within datasets.

that extracting different points of interest located nearby or within the address via Santa Croce is almost impossible from the dataset.

Contextual View

Dataset does not give the context in which the data can be visualized.

A real-world object exists in space-time and possesses various kinds of relationships with other objects. Based on multiple settings, data has to be visualized in space-time and also should highlight the relations with other objects. For example, the restaurant Al Volt in Figure 6.1 is located in a specific location, has properties such as opening hours and is related to its owner, and its menu. Based on the visualization need, these kinds of contextual parameters offer different insights, and they have to be well reflected in the interface. However, in open data portals, data are presented in a simple tabular view which hinders contextualization [Mazumdar et al., 2014, Agrawala et al., 2011].

Segregation of attributes and properties

Many semantic tools present both attributes and relations together.

Various semantic tools present multiple properties of an entity. For example [Mahdisoltani et al., 2014] presents all the attributes and relations together in one graph. This creates a highly dense graph. Mixing of intrinsic attributes of an entity and a related property makes the exploration process difficult.

6.3 Summary

Knowledge diversity and heterogeneity associated with visualization hinders designing a generic application to cater users' need.

Here we discussed various diversity that hinders creating a generic application. Though the research was mainly targeted toward Open Data, it can be generalized to other organizational data settings as well. Availability and access to the data source is a major hindrance. Currently, it is very hard to find the required datasets. Even if found, an expert's advice is required to make sense of the data. Lack of metadata and provenance also hinders the understanding ability and usability of data. The intrinsic diversity associated with the data hinders linking and combining various datasets to get a complete view of an object. Furthermore, the visualization limits the data usage and coverage from the end users' perspective.

Chapter 7

Requirement Analysis

After understanding the domains from various perspectives in the chapter 5 and categorizing the issue as knowledge or visualization diversity in the chapter 6, we realize a commonality between domains that satisfy our requirement of designing a generic visualization. Before starting the functional and nonfunctional requirements for generic visualization, we try to figure out the desired stakeholders.

Four significant stakeholders are handling the data from various viewpoints.

7.1 Understanding various Stakeholders

The work discusses here outlines four major categories of stakeholders.

Government
Many governments around the world participate in the release of their Data as Open Data. Their primary intention is to allow access to the data to the end users to promote transparency and innovation [Union, 2014]. Currently, the datasets are fragmented across various sources making it hard for end users to access the data. Their major motivation is to make the data visible, so the knowledge associated be strengthened and centralized. This arrangement will promote openness, good governance and provide more social benefit.

The government provide data to allow both innovation and transparency of information.

Business Organizations

Organizations prefer organized and classified information to smooth the search and retrieval of information.

In the context of the global market, where various people from multiple countries work in the multi-cultural environment. Working in such environment requires streamlined and specialized task force. These task force have a variety of roles and specific application on which they work. Each employee in the organization will have a specific information need. Information seeking process in any organization is a very expensive and time-consuming task [Feldman, 2004]. To cater this issue of providing the right information at the right time, organizations need properly classified information catalog exposed on the interface to perform day to day activity. They want an easy interaction with the information catalog.

Data Handlers

They possess domain or technical knowledge to handle the data.

These people are domain experts and computer scientists who have the understanding of the underlying technology and ontologies. However, domain experts might not have knowledge of an underlying technology. Their main responsibility is to manipulate the data. They perform various interdisciplinary research [Zuiderwijk et al., 2012]. Currently, much effort is made on the use of the data designing various applications. These experts make the tremendous effort of combining and linking data. They prefer an interactive interface which easily displays the data definitions which they use for the analysis of the data.

End Users

They have no technical expertise and require an exploratory and interactive interface.

The end user who does not have any background in handling the data. A research that was done by an independent firm[1] shows that the common people seek information directly as summary or charts, which can be easily be explored and interacted with using specific services. They seldom prefer direct data manipulations.

There is an information search pattern preferred by the end user.

In all the case, we see that stakeholders prefer finding specific data information at first and then get a detailed view of the data. Afterward, they prefer exploring various relationships by filtering and setting various contexts.

[1]http://benchmarkstudy.socrata.com/d/xkgk-r22k

7.2 Stakeholders Requirements

After a thorough investigation of various visualization techniques in part I of this thesis, we conclude that context-aware diversified data visualization catering requirements for various stakeholders do not exist. To fill the gap, several requirements for a solution to address the research questions were drawn. The requirements were identified and grouped into two main categories: Functional and Non-Functional.

Requirements were grouped as functional and nonfunctional.

7.2.1 Functional Requirements

An interactive visualization system facilitates interaction between a user and underlying data using effective visual means [Mazumdar, 2013]. Users then interact with the objects visualized to perceive required information according to their intentions. A focused technique which amplifies human cognition can be considered a better visualization [Nazemi, 2016]. To provide better visualization, a series of operations has to perform, this includes investigation on multiple layouts, various visualization paradigm, diverse data presentation and interactions. Similarly, a keen attention has to be paid to generate set of requirements that align with the user's need so they can easily obtain their required goals. The following lists the requirements that the solution must address:

The requirements should be user-centric allowing them to amplify their cognition.

FR1: Solution has to follow an Entity-Centric Approach.
A diverse set of data are designed and used in their limited contexts. Building a general purpose solution for diversified data requires the interface to handle diversity and unpredictability at both design time by using a solid knowledge representation, and runtime by using contextual and adaptive visualization technique. This technique requires moving from a property-centric data to an entity-centric representation and view of data and knowledge.

Entity-centric approach facilitates handling data diversity both at design and runtime.

FR2: Solution has to be adaptive in terms of entities and classification

New datasets should be easy to integrate.

This requirement is associated with requirement R1 which affects visualization. The newly integrated datasets should be classified at runtime so the end users can use them.

FR3: Facilitate object visualization

The tool must allow visualizing diversified data in an easy way.

The primary focus of this thesis is to allow visualization of large semantically related diverse datasets in a more human understandable manner so that the end user can easily perceive and understand the content.

FR4: Facilitate interaction with objects

The interaction mechanism should be easy and intuitive.

There should be an easy communication mechanism so the users can express their desired intention to satisfy their information need. Consideration should be made on facilitating standard interaction methods like using a keyboard and mouse along with other techniques like using a touchscreen.

FR5: Facilitate object searching

Multi-faceted search facility should be provided.

The visualization should facilitate direct searching of various objects. Various search facets should be designed for an easy query of the large-scale data.

FR6: Facilitate exploring related objects

The exploration mechanism should easy to see, and the process of exploration should be natural.

The relations between the data tuples in the dataset can be dense. Users familiar with their prime domain can easily understand the mechanism to relate this information. However, if they have to explore a diverse set of domains, then there should be a facility that allows easy exploration across multiple domains. The exploration mechanism should be easy to perceive and understand. The exploration should follow both top-down and bottom-up allowing easy refining of information.

FR7: Facilitate simultaneous presentation of Entities

The visualization should offer multiple ways of perceiving the data to the end users.

The datasets contain multiple properties making them multidimensional. These multidimensional data can also be perceived from various visual perspective. Each of these visual perspectives offers a different kind of information to the user. The solution designed should provide multiple views based on the multiple dimensions of the data.

FR8: Solution should allow easy adaptation to various context

There can be two different kinds of contexts: data and user. The first, context guides various layout, representation, and schemes required for visualization whereas the second context is based on users' role, aptitude and need. The solution should be adaptive on both data and user characteristics.

Based on the context, the solution should adapt to various contents, visual representations, and contexts.

FR9: Solution should be language independent
The datasets exist in various language making them multilingual. This issue requires a prospective user to know various languages to understand the data. The solution should be language independent so it can be expanded to include different languages.

The solution should be able to reflect the change in the language immediately.

Summary of all the functional requirements is deduced in the table 7.1.

Table 7.1: Summary of noted Functional Requirements.

	Requirements
FR1	Solution has to follow an Entity Centric Approach
FR2	Solution has to be adaptive in terms of entities and classification
FR3	Solution should facilitate Object Visualization
FR4	Solution should facilitate interaction with objects
FR5	Solution should facilitate object searching
FR6	Solution should facilitate exploring related objects
FR7	Solution should facilitate simultaneous presentation of Entities
FR8	Facilitates adaptation to various context
FR9	Solution has to be language independent

7.2.2 Non Functional Requirements

Nonfunctional requirements specify the basic characteristics rather than focusing on the specific solutions. The following lists the nonfunctional requirements of the solution:
NFR1: Domain Independent
Based on the functional requirement R1, the entity centric approach offers designing domain independent applications. In the current context, most of the existing solutions are highly domain or dataset dependent, only catering to one type of data.

Entity-Centric approach facilitates designing domain independent applications.

NFR2: Intuitive

The solution should be intuitively easy to follow.

In an open data environment, many applications are developed that uses their own metaphor making them nonintuitive when deployed to the end users. Though training session can be provided to the users, making the system self-intuitive will increase productivity and efficiency of the system.

NFR3: Generic

Users from various domain and range should be able to use the system.

The solution should be intuitive, and the visualization and interaction should cater the need of a vast number of users working in various domains. The information should be presented with least technical jargon.

NFR4: Aesthetically Usable

The tool developed should be attractive to use.

Designing an attractive and usable interface is tough. The rationale behind this requirement is that users will pay more attention and stay longer to query information in the system if the interface is aesthetically pleasing.

NFR5: Human-Friendly Representations

The objects used in the interface should be familiar to the users.

Many different visual schemes and metaphors can be used to represent an object. However, precaution has to be made so that users are always familiar with what they see. The representation should match users' visual perspective. This representation minimizes confusion among the users. An example can be using a circle to define an area range.

NFR6: Consistent representation

Objects and technique should be consistent throughout the interface.

The solution must consistently represent diversified data to reduce confusion. Objects and technique defined to describe a particular type of data should be consistent throughout the interface. For example, if a file icon is used to define a 'file', then it should remain consistent throughout the solution.

NFR7: Standard interaction mechanism

The interaction mechanism should be fluid and should be similar to other applications with similar interactions.

The solution must ensure users are familiar with the interaction mechanisms. In an open data setting, many applications are designed with various interaction mechanism. Users then have to adapt to these new interactive techniques. It will be easier for users if the interaction is based on their prior experience with tools employing similar interaction paradigms. Furthermore, the interaction should be consistent throughout the solution. For example, if double clicking on a map zooms it in one case, it should be true for all the maps throughout the interface.

NFR8: Support a wide range of devices
Various devices follow different interaction and representation technique. Achieving this nonfunctional requirement allows the system to be used by a vast number of users. However, achieving this required fulfilling nonfunctional requirement 2, 5, 6 and 7.

The solution should also be adaptive to a wide range of devices.

NFR9: Recovery from error
The solution should allow users to rectify their error and give proper, timely feedback.

Users should be able to trace back from errors.

NFR10: Support exporting data to other SW formats
Most of the application domain use Semantic Web technology as their knowledge-based solution. To be compatible with various SW technology, there should be a provision to export data to various other formats such as RDFs and JSON-LD.

The solution should be compatible with other SW data formats.

Summary of all the non functional requirements is deduced in the table 7.2.

Table 7.2: Summary of noted NonFunctional Requirements.

	Requirements
NFR1	Domain Independent
NFR2	Intuitive
NFR3	Generic
NFR4	Aesthetically Usable
NFR5	Human Friendly Representations
NFR6	Consistent representation
NFR7	Standard interaction mechanism
NFR8	Support wide range of devices
NFR9	Recovery from error
NFR10	Support exporting data to other SW formats

7.3 Summary

Stakeholders' input was vital for designing the interface.

Throughout the research, the continued involvement of various stakeholders proved to be beneficial in deriving detailed requirements. These requirements lay a foundation for using the Entity-Centric approach as a basis for our solution. During the investigation, we realized that the stakeholder's requirements are generally ignored. This results in the generation of extremely contextual, nonscalable and unusable application.

We propose solution methodology based on an entity-centric approach.

In the next part, we discuss the solution methodology based on the problem described in this part. We will explain how we will implement an entity-centric approach to solve the knowledge related issues and how the visualization metaphor was realized.

Part III

Solution Methodology

Chapter 8

An Entity-Centric Approach

Various analysis done in the chapters 5, 6 and 7 lay the foundation for a generic, scalable and a flexible solution capable of handling diversified data. This solution can be achieved by following an Entity-Centric approach. The inherent nature of entities, being highly interconnected, multi-dimensional and self-descriptive, provides an excellent framework for the solution. Using an Entity-Centric information infrastructure with a wide range of datasets provide a realistic scenario generation a generic solution. Here we try to define the various facets of an entity.

The Entity-Centric approach allows designing a generic and scalable solution.

8.1 Overview

Here we will provide a theoretical foundation for an entity. By entity we mean any real-world object which is so important to us to give it a name [Giunchiglia, 2006]. Entities are the essential elements of the knowledge. Entities can be defined either as an abstract or physical objects, can be of different types (e.g., person, location, event, artifact and so on) and are described by various properties (e.g., name, position, address, etc.) which are based on the type of the entity but also on the domain. Different domains may elicit differ-

An entity is anything that is important to us that we give it a name.

Entities can be described with a set of attributes from various domains.

ent properties for the same entity, e.g., in the education domain is enough to provide few properties to the entity type person, like name, address, faculty, while in a medical domain much more properties should be provided at a much higher level of details like blood type, among others. In an entity-centric view, data instances are described as a set of interconnected entities with their respective properties. For example, a restaurant is an entity with various properties such as coordinates, address, opening hours and so on and can be connected to other entities such as Location and owner. Based on its properties, we can easily perceive that it is a restaurant. Furthermore, we can distinguish this particular restaurant from other restaurants in the world because we align various properties of this restaurant mentally and give a unique identification to it. This identification helps us to differentiate it from any other restaurant. This is a key motivation to move from property centric representation to an entity-centric representation. While the properties we consider are goal dependent, whereas entities are independent.

Many properties can be merged to provide a unique identification to an object.

8.1.1 Property Centric to Entity Centric Transformation

We mimic the data aggregation as humans do.

Since we cannot anticipate various kinds of data that will appear, we are trying to mimic the aggregation of data in the machine in the same way as humans do. As humans adapt to diversity, our approach will also evolve and adapt based on the situation [Giunchiglia, 2016]. So that, we will put together data into entities in the same way as humans do and we believe that this approach will minimize the problem of data diversity.

Humans perceive things in space and time.

We also, simulate the visualization in a human-like adaptation [Giunchiglia, 2016] where a real-world object is perceived in space and time. This perception is the reason we have space and time. Furthermore, we perceive concepts such as relations as a network.

Entities facilitate this kind of mimicry in an open world concept.

We have to design the system to behave like humans. The entity is the only way which can facilitate the working of a machine in an open environment like Open Data where we cannot predict the future dynamics of the data. Based on the notion of an entity, we set our premise for entity-centric representation and visualization.

8.1.2 Common Sense Representation

The notion of entity is rooted in our commonsense understanding of the world. We, humans, view objects as entities holistically [Damasio, 1989, Tallon-Baudry and Bertrand, 1999]. The viewing process has the psychological perspective also termed as sense-making [Klein et al., 2006], i.e., when we perceive an object; our brain first generates a mental model [Anderson, 2013, Gentner and Stevens, 2014]. The mental model is an object's representation in a memory where various features of the object are associated with different concepts or knowledge. This association means that whenever we see an object, we combine various properties of that object (based on concepts and knowledge) and then project it as a whole object. Afterward, based on many encounters with the object our cognitive ability updates the knowledge about that object and starts recognizing, reorganizing and categorizing the properties according to their similarity and differences [Millikan, 2000, Giunchiglia and Fumagalli, 2016].

The concept of an entity is rooted in our commonsense understanding where we see things holistically.

8.2 Entity-Centric Representation

An entity-centric approach has been used for collocating the diversified data in one place using an *eType*. An *eType* provides a schema and set of rules for the creation of a conceptual representation of a real-world entity, for example, a person, a building or an organization. An entity can be expressed using following quadruple:

An eType allows creating a conceptual representation of any real-world object.

An Entity is defined using a quadruple.

$$Entity =< EID, \{Name\}, \{Property\}, eType > \quad (8.1)$$

$$Property =< \{Attribute \mid Relation\} > \quad (8.2)$$

$$Attribute =< AttributeName, Datatype > \tag{8.3}$$

$$Relation =< RelationName, eType > \tag{8.4}$$

where,

EID is a unique identifier.

EID - A unique global identifier that identifies an entity over a range of domains. It is based on the set of defining attributes an eType possesses. For example, a person can have an identifier based on his name, date of birth, place of birth or their combination.

The name is a set of strings representing the names used by the corresponding representation.

Name - An entity is always represented by a name or a set of names. It is how we recognize a real-world thing. For example, Pink city of India and Jaipur are two different names for the same entity or in the personal healthcare domain, person name, and patient's identifier can be considered the name.

Property is a non-empty set of attributes and relations that characterize the entity.

Property - A property defines the characteristic of an entity. Properties are quantitative (expressed in numbers), qualitative (expressed as adjectives), or descriptive (expressed as natural language sentences), and coexist in a name-value pair. Properties can be broadly categorized as attributes and relations (Equation 2). Attributes are intrinsic properties of an entity that describes a quality or a feature (Equation 3). For example, temporal (date of birth, date-time of an event), spatial (coordinate of a location), state (conditional attributes like open or close), and inherence (qualities like the color of a building). On the other hand, relational properties relate one entity with another entity according to an eType (equation 4). An example of relation is the *friend-of*, *located-in*, *near-by*, or *part-of*.

An eType is a reference class of an entity. There are few basic eTypes.

eType - A category of a real-world entity. The Federal Geographic Data Committee (FGDC) defines an eType as *"the definition and description of a set into which similar entity instances are classified (e.g. bridge)"* [Committee, 2015]. In other words, it is a template that defines the constraints (set of rules) for creating attributes and relations of a real-world entity (e.g., a person, a building or an organization).

8.3 eTypes and their defining attributes

To motivate the design of a generic solution for visualizing diversified data, it is necessary to understand various primary attributes an eType possesses. Clear separation of concepts of attributes, relations and entity classes leads to a flexible mechanism that combines these elements to represent knowledge. In our scenario, there are five basic types of eTypes as Location, Event, Person, Artifact, and Organization. We will discuss and present the most critical attributes that guide the visualization below:

There are five basic eTypes.

8.3.1 Location

Table 8.1: Location eType

Name	Description	DataType
Identifier	a symbol that establishes the identity of the one bearing it	STRING
Geographical name	a name by which a geographical location is known.	[] NLSTRING
Description	description of the entity	[] NLSTRING
Coordinate	a number that identifies a position relative to an axis	GEOMETRY

Location represents a point or an extent in space. The attributes are showed in Table 8.1. Notice that Geographical Name overrides Name. In our context, the most important attribute in the geographical context is Coordinate. If a data point contains coordinate, then they can be easily be presented on a map. That means the locations are tangible physical objects.

Coordinate defines the position of an object in the space.

8.3.2 Person

A person represents an individual. The defining attributes for a person eType is shown in Table 8.2. A person eType can relate to other eTypes such as Location regarding Place of Birth, Place of Death or Address. Similarly, can have a set

Person eType can have a direct relation to a location and other relationships according to their roles.

of roles such as visitor, professor, patient, doctor, founder and so on. In different context and situation, an individual can play multiple roles which will help relating to various other eTypes.

Table 8.2: Person eType

Name	Description	Data Type
Identifier	a symbol that establishes the identity of the one bearing it	STRING
Finding related to biological sex	the properties that distinguish organisms on the basis of their reproductive roles	CONCEPT
Ethnic group	an ethnic quality or affiliation resulting from racial or cultural ties	CONCEPT
Employment status	current employment status of the person	CONCEPT
Place of Birth	the place where someone was born	LOCATION
Date of birth	the date on which the person born	CONCEPT
Country of residence	the country where the person has her residency	COUNTRY
Address	written directions for finding some location	LOCATION

Person eType can also be extended as a patient. List of attributes defining a patient is shown is table 8.3.

Table 8.3: Patient eType

Name	Description	DataType
Identifier	a symbol that establishes the identity of the one bearing it	STRING
Finding related to biological sex	the properties that distinguish organisms on the basis of their reproductive roles	Concept
CHI Number	Community Health Index Number	SSTRING
Visit	person visit to the healthcare centre	Visit
Prescription	written instructions from a physician	Prescription

8.3.3 Event

An Event is an action which happens in a particular location and time. An event always involves a set of participants.The defining attributes are shown in Table 8.4. In our context, an event always constitutes at least a location and a participant. A participant can be either a person or an organization. For example, a visit to a doctor can be considered as an event which comprises of a location (can be a generic location or a building according to the availability of data) and a medical practitioner (doctor, nurse) and a patient. Using this example we can relate an event eType to location and Person eType together.

An event eType is always related to a location and involves at least a participant.

Event eType can be extended to accommodate visits.

Table 8.4: Event eType

Name	Description	DataType
Identifier	a symbol that establishes the identity of the one bearing it	STRING
LOCATION	the place where the event occurs	[] Location
Person	someone who takes part in an activity	[] Person
Organization	organization that takes part in an activity	[] Organization

Visits are classified into four categories: outpatient care, inpatient confinement, emergency room, and long-term care. Persons may transition between these settings over the course of an episode of care (for example, treatment of disease onset) [Reich et al., 2017]. Various attributes are shown in table 8.5.

There are four types of visits.

Table 8.5: Visit eType

Name	Description	DataType
Identifier	a symbol that establishes the identity of the one bearing it	STRING
Patient	a person who requires medical care	Patient
Health encounter sites	Healthcare centres	Hospital

8.3.4 Artifact

An artifact should have a creator and a location where it is located.

An artifact is a man-made object taken as a whole. The defining attributes are showed in Table 8.6. An artifact is a generic eType and can consist of other specific eTypes such as buildings, artworks and so on. To be included in an artifact eType an object should possess a creator and a location where it is located.

Table 8.6: Artifact eType

Name	Description	DataType
Identifier	a symbol that establishes the identity of the one bearing it	STRING
Name	a name by which an entity is known	[] NLSTRING
Date of making	the date on which making begins	DATE
Height	the vertical dimension of extension	FLOAT
Address	written directions for finding some location	LOCATION

8.3.5 Organization

An organization is related to a location and person.

An organization represents a group of people or a collective who work together. The attributes are shown in Table 8.7. An organization eType should be located in a particular location. Along with that, it should have founder and members. This pattern also helps to find relationships between location and person.

Table 8.7: Organization eType

Name	Description	DataType
Identifier	a symbol that establishes the identity of the one bearing it	STRING
Location	the place where event occurs	[] Location
Member	member of an organization	[] Person
Founder	founding member of an organization	[] Person

8.3.6 Prescription

Prescriptions are written instructions from a physician or dentist to a druggist concerning the form and dosage of a drug to be issued to a given patient [Miller and Fellbaum, 1998]. Snippet of a prescription eType is shown in table 8.8.

Prescriptions is a physical objects.

Table 8.8: Prescription eType

Name	Description	DataType
Name	The name by which an entity is known	[] NLSTRING
Identifier	a symbol that establishes the identity of the one bearing it	STRING
Patient	a person who requires medical care	Patient
Prescription drug	a drug that is available only with written instructions from a doctor or dentist to a pharmacist	Drug
Prescription date	the date on which a prescription is issued.	DATE
Dispense date	the date on which drugs dispense	DATE

8.4 Summary

In this chapter, we briefly described the what an entity is from cognitive principles. Similarly, we also signified a shift from property-centric approach to entity-centric approach. We then further defined the entity-centric representation where we briefly describe all the intricacies related to an entity. eTypes then lay the foundation for designing a generic solution for visualizing diversified data. The attributes an eType possess helps to generate an exploration pattern.

Entities are the basis that helps to design a generic application. In this situation, eType shades the light for developing contextual application.

Chapter 9

Interface Design

Our proposed solution uses a combination of several approaches previously discussed in part I including graph-based visualization, faceted browsing, visual mash-ups, dashboard design and so on. We try to accumulate all the positive aspects of such tools to provide better visualization and interaction support.

The interface design is influenced by various previously discussed tools and techniques from multiple communities.

The goal of this research is to come up with a visualization framework that allows presentation, interaction, and exploration of diversified data following an entity-centric approach (FR1) in a pleasing way. To fulfill this goal, three major tasks has to be complete. They are:

1. Visualize diverse data
2. Interact with the data
3. Explore the data

9.1 Proposed Architecture

To design a diversity-aware user interface following the requirements mentioned in the chapter 7, we follow the visualization technique (FR3) which is generic and has the

The solution should adapt to customized domain-specific content.

flexibility of being adaptive (FR2) to domain-specific diversified data in the architecture. The ability to present information using generic visualizations help in exploring unknown data (FR6) in a generic manner. Interaction with the visual objects will generate queries allowing users to search and explore deep in the data (FR4, FR5).

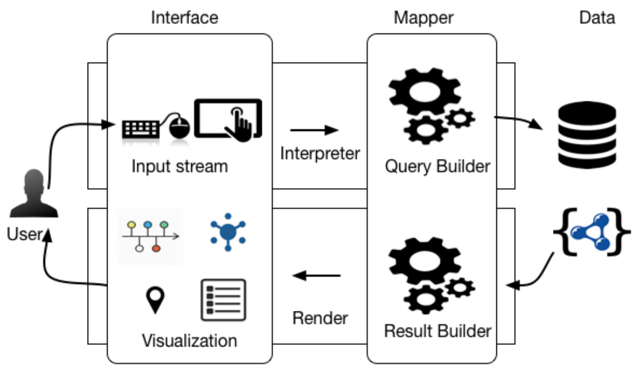


Figure 9.1: Solution Architecture

The interaction, visualization and exploration process is cyclic.

In Figure 9.1 we demonstrate the mechanism on which the goal can be achieved. Here the process of interaction, visualization, and exploration is cyclic. Firstly, a user will start the communication process by interacting with the interface using various input streams (mouse, keyboard or touchscreen). This interaction is understood by the engine and then converted to a set of queries by the mapper. The questions are then forwarded to various data stores which integrate and further build the results and then pass it to the same presentation interface as a result. The cycle can be repeated by the user by interacting with new objects on the interface. We briefly describe each component of the cycle below:

User Interface

We implement simultaneous presentation of an entity in a single interface.

It allows communication between user and data. The complexity that is involved with data management, formats is hidden from the end user. In our solution user only interacts with the visual objects on the interface. Here we implement simultaneous presentation of entities (FR7) which allows presentation and adaptation of data according to various visual contexts (FR8) providing a different kind of information by continuous aggregation of information in the interface.

Query builder
The user interface consists of various visual components (forms, map or hyperlinks), detail level (zoom), interaction(click or scroll) that allows capturing multiple user intentions. The mapper understands these intentions requests and then converted into queries which then is asked to the database.

The user intentions are understood by the mapper and converted into user queries.

Result Builder
Based on the query, information is transformed into the result by the result builder. Here various results are combined to form concrete, real-world object. The metadata associated with the object guiding the visualization is understood here. For example, coordinates require map support, data and time require a timeline.

Result builder fetches the result and then aggregates the data to design a real-world object.

Renderer
This the component then renders the updated result into different visualization context. So various coordinates are drawn on a map, data/time data is rendered on a timeline and so on.

It renders the result on the interface.

9.2 Engaging users

Another major contribution of the work is engaging users to develop a generic framework to visualize diversified data. The gathered information about domain analysis (5), diversity(6 and requirements (7 helped us to generate adaptive visualization. In the present scenario, only one factor either data or users is considered [Nazemi, 2016].

Users were involved in the design process.

We conducted a series of interviews, focus group iteratively on coming up with a better solution. During these sessions, we realize that users do show eagerness to design a better solution by involving and providing feedback on the essen-

Users gave feedbacks to design a better solution.

tial presentation and interaction components. Furthermore, they also brainstormed to give new ideas regarding the exploration of the entities. However, it was sometimes hard to find them and make them understand the situation.

We designed low fidelity mockups and then reiterate it.

Design ideas did contradict with among the users.

All the ideas regarding generic visualization were sketched as low-fidelity mockups and were further discussed with stakeholders from various domains. These discussions provided meaningful insight to implement a real solution. Discussions with various stakeholders do allow insight from data management to the user interface. Sometimes there were arguments in the focus group regarding the interface layout and implementation of technology. However, they were resolved by more experienced users. These users gave more meaningful insight on dashboard and widget-based design interface design. Furthermore, in some sessions, language was also a barrier. We used an experienced interpreter to help us during those sessions. Below we describe various iteration cycles that we performed.

9.2.1 Design Iteration I

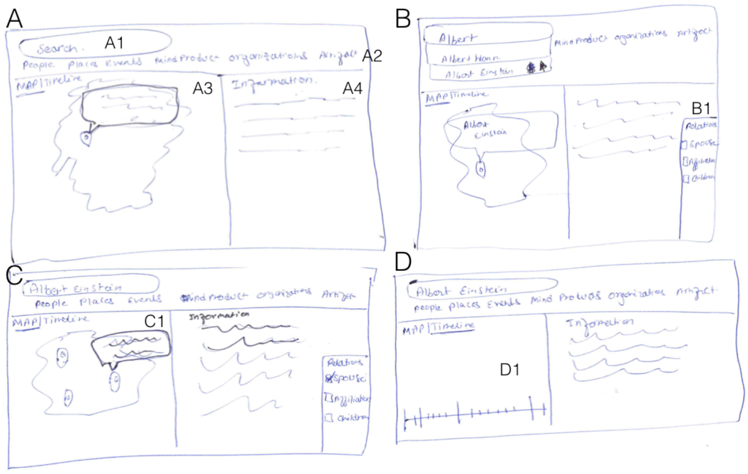


Figure 9.2: Initial Mockup version I based on tourism scenario.

The first design had search, eType, map, and timeline component.

The overall design of the solution in an Open Data setting was proposed as shown in Figure 9.2. Here Figure 9.2A shows the basic interface consisting of search component

(A1), an eType component (A2), a map Component (A3) and finally a list component (A4). Users had various ways to query and interact with data. They could search component to directly query the intended entity. They could also use the eType component to get a bulk result then filter the bulk to get the final result. The search component had the auto-fill capability (Figure 9.2B) that would help to select a particular entity. Once the entity(s) was/were selected, it/they would be displayed on the map (Figure 9.2C). Clicking the map icon would display various attributes (Figure 9.2B) and generate relations (Figure 9.2B1). The mockup also had a facility to switch to a timeline tab (Figure 9.2D) which would display the various attributes of the selected entity on a timeline and a list. We implemented two high fidelity prototypes based on this design (Chapter 11 and 12). However, the tool at that time did not have search capability.

Two different versions of high fidelity prototypes were also developed.

End users highly appreciated the proposed design. They liked the relational part where the user could click on a related entity menu and get all the desired entities. The tab that allowed switching between a timeline and map was not intuitive. It was realized that continuous switching between tabs added more cognitive load and they had to recall what they were doing to solve a task. The users suggested removing the tab for timeline and make it permanent on the same screen so that they don't have to switch using the tabs. They also mentioned that the there was no balance on the visual components and textual components. They also asked making the mockup multilingual. They also requested redesigning the eType component as it was taking a huge space and was conflicting with the search component.

This design was highly appreciated but had a cognitive overload to the end users.

9.2.2 Design Iteration II

The previous design iteration considered Open Data specific to tourism however in the second discussion session they wanted to observe the mockup on the implementation of Open Data on a cultural heritage scenario.

This iteration was more challenging as users asked for a different solution considering a different scenario.

It included refinement over previous design and also included new features of a contextual facet menu.

In the new iteration, we included the design suggestions from the users. The newly designed mockups are shown in Figure 9.3. The mockup consists of an eType drop-down menu (A1), search component (A2), language component (A3), map component (A4) and a widget that displayed *what happened in the history today?* (A5). Various entities or an entity was visualized once it was searched. Instead of a list, entities were displayed on a visual grid making them easier to choose (Figure 9.3). Once the entities are searched, a contextual facet menu that allows various filtering capability appears on the screen (B1). If the entity possesses time-related attribute, then it was also displayed below the map component (Figure 9.3.C5). We removed a horizontal eType component and used a drop-down menu to present basic eTypes (Figure 9.3.A1).

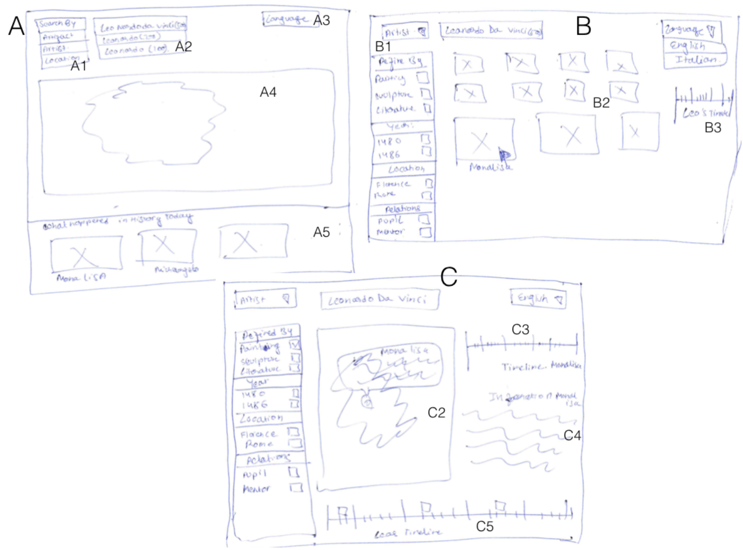


Figure 9.3: Mockup version II based on cultural heritage scenario.

The user feedback was positive for the design.

The users were very impressed with this iteration. However, they suggested that the history component cannot be used in other scenarios. They complained about the eType not being visible. They also indicated that there should be a possibility of searching entities from more than one eType at a time. Since on the mockup, it was only possible to select one eType this feature was hard to realize. All the stakeholders liked the possibility to switch language. The thumbnail representation was also considered useful. The users were still not happy that half of the screen was devoted to displaying various attributes.

9.2.3 Design Iteration III

In this cycle, we designed mockup considering scenarios related to transportation.

Design iteration II was highly successful with less feedback from the users. In the design cycle III, users only asked to add another scenario. We then adopted the mockup to the transportation setting. Users suggested shifting from two view approach to a dashboard-like view where there was one or more visual component. They were vague whether these components should be coordinated or not. i.e., change on one component would reflect the change on all the components.

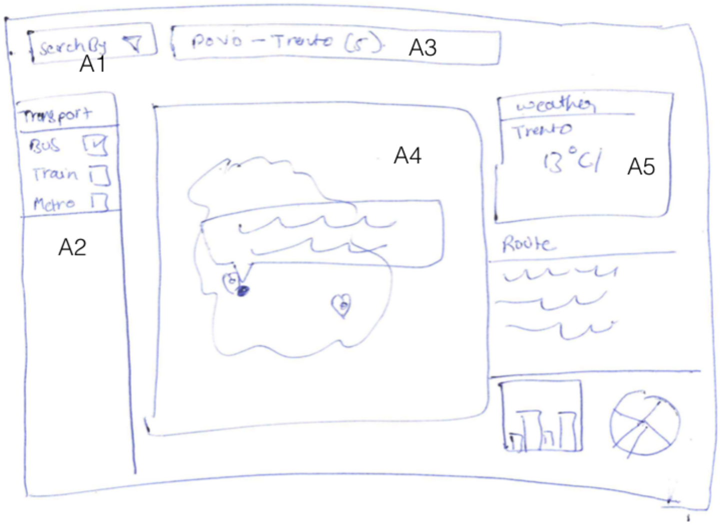


Figure 9.4: Mockup version III based on transportation scenario.

Apart from other visual components, this mockup also had an external widget.

Considering this, we iterated the mockup and removed the history component. The design this time was more modular and followed a mashup/dashboard design practice. Figure 9.4 shows the new design. Here, we created two eTypes components. One component only listed basic eType on a drop-down menu (A1) whereas another component listed various contextual facet (A2) and was always present on the screen. Search and map component was also present (A3, A4). Apart from the major component, it also consists of other widgets like (A5) feature (weather widget). We also streamlined and gave less space to present textual information as most of it was already present on the map component.

Expert users were concerned about the scalability of the design.

This design cycle was a major milestone, as users started showing concern about the scalability of the design. Here, we were designing a single mockup for various domains. For the end users, the scalability of the exploration process was a major concern which might hinder our requirement. Furthermore, they stated that the number of attributes in an entity would increase drastically.

9.2.4 Final Design

We had to go back to the design board to find a new design solution.

This information overload situation had to be tackled, so we went and restructured our design also considering the data component so that only required information in the desired context was fetched.

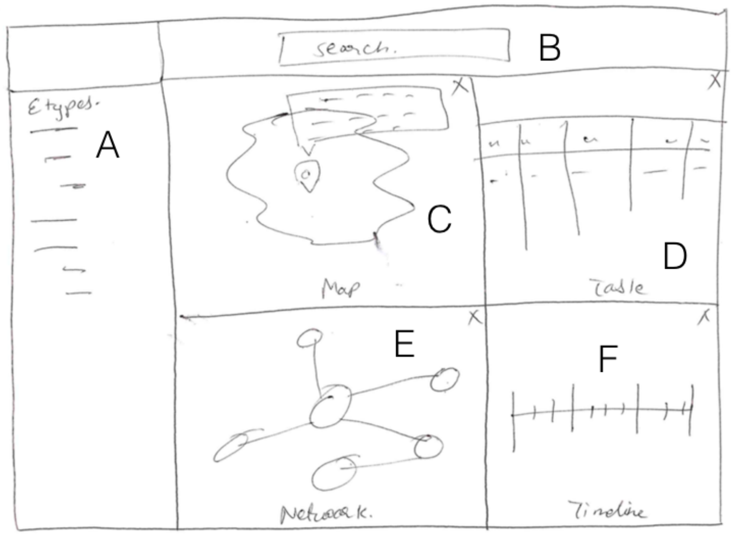


Figure 9.5: Mockup version IV

We divided the design into four tabs, one for each visualization context.

As an initial design solution, we considered designing four different views and segregating the visualization contexts based on the eTypes. There were four tabs for each basic visualization contexts: map, space, list, and Network. However, this design was not liked by the most of the users. They mentioned that the tab-based view would reduce their working and every time they had to switch tab to get a different perspective on the data.

To tackle this situation, we designed a multiview solution approach as shown in Figure 9.5. Here in this design, a single view screen view was divided into five major parts as an eType lattice (A), search (B), map (C), table (D), network (E) and timeline (f). The eType lattice contained all the generic eTypes along with required subclasses (the hierarchy was adaptable to add more domains). This arrangement allowed visualization according to various data context and roles. The search feature was adaptive to search for either concepts, eTypes or entities giving more flexibility to the users. Once an entity was searched or explored from the lattice, it was projected on all four view components. If an entity missed a defining attribute, then the view disappeared.

We designed a multiview solution with coordinated views.

The users gave very good feedback on this design. They mentioned that the ability to see various properties of an entity from different perspective simultaneously in a single screen added more excitement than previous designs. The choice of only fetching what they want as per their context was liked by most of the participants. The users had some concern about the screen size. They mentioned that thee might not be enough space to see multiple perspectives at a glance. Furthermore, they were also concern about the portability of such design approach on other small devices like tablet and phones.

The Users mentioned that the interface became more intuitive than before.
There were some concerns about the screen sizes.

9.3 Summary

Throughout the research, the continued involvement of user communities and stakeholders have been an incredibly positive and encouraging experience. Their continuous participation in the generation of the various prototypes was commendable.

Users participated in various interactions throughout the design process.

Continuous feedback from the users helped designing SemUI.

Their valuable feedback and challenging queries which were not only related to the interface shaped the way for designing a SemUI a multi-tiered solution.

Part IV

Development of Technology

Chapter 10

SemUI: Multi-layered Architecture

In this work, chapter 1, part II and III introduced the research goals, problems, requirements along with a hint of a solution. Our solution follows a multifold approach, where first the datasets are converted into meaningful entities using an entity-centric approach, after that they are visualized and explored using a contextual multi-coordinated interface. Here, multi-coordinated visualization is realized using a rearrangeable dashboard metaphor. Two major concerns were:

We implemented an entity-centric approach for data management and multi-coordinated views for presentation and exploration.

1. How multiple views could visualization and exploration of diverse entities providing information from various perspectives?
2. Technical constraint and implications when developing a generic tool for a change in data require readapting the interface from scratch.

The concerns mentioned above can be resolved if the tool is flexible from user's perspective and data perspective needs modularity on both ends. Separating the visualization side from the data side helped to design a flexible layout which allowed presentation and exploration of the entities in a generic manner.

The tool has to be adaptive regarding entity management and visualization.

10.1 Architecture

We implemented our design based on the modular multilayered architecture.

The multilayered approach is adapted from the software engineering principle which allows easy maintainability and encapsulation of data (entities) that will help in creation of high performance, generic and adaptive systems. The architecture is presented in Figure 10.1.The layers are:

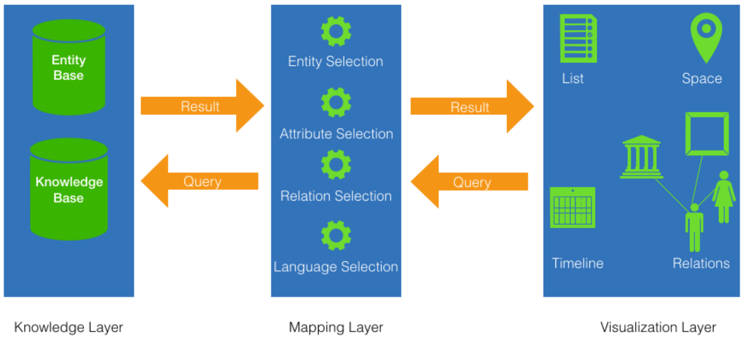


Figure 10.1: Multilayered Architecture

Comprises entities forming an entity graph.

- ***The Knowledge Layer***- It stores data as entities which are interconnected with each other forming a dense entity graph. In this graph, the nodes represent the entities, and the links describe the relations between them. The data are modeled as entities along with their attribute and relations following an entity-centric approach using an eType.

Presents entities on various visualization contexts simultaneously.

- ***The Visualization Layer***- It visualizes entities from various contextual perspectives such as timeline, space, network, and list. Furthermore, the User Interface allows visualization of an entity 1)as a whole with all of its attributes and relations and 2) also according to the users' representation of the entity. This adaptivity contributes towards developing an intuitive, natural and generic User Interface for a vast number of users.

Fetches entity from the knowledge layer, interpret them and presents them on the visualization layer.

- ***The Mapping Layer***- This layer is crucial because an entity graph is dense and the user will be overwhelmed with information overload on the User Interface. Therefore, it interacts with both *Knowledge* *layer* and *Visualization layer* and fetches specific parts of the entity graph from the *knowledge layer* according to the request issued by the *visualization layer*.

10.2 Knowledge Layer

The knowledge layer consists of two interlinked data hubs called entity base and knowledge base.

The knowledge layer comprises an entity base and a knowledge base.

- Entity Base: stores the information about the transformed diversified data from multiple datasets as a set of unique entities.
- Knowledge Base: stores the schema and vocabulary providing a template for an entity.

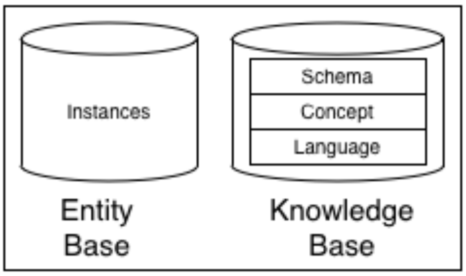


Figure 10.2: The Knowledge Layer.

The Knowledge layer maintains data as a set of interlinked entities in an entity base along with their respective eTypes in the knowledge base as an entity graph. Figure 10.2 illustrates various components in a knowledge layer. The knowledge base explicitly separates the schema and vocabulary. The existence of a generic schema which describes the data is the step towards handling diversity.

Knowledge layer maintains a network of entities and eTypes as an entity graph.

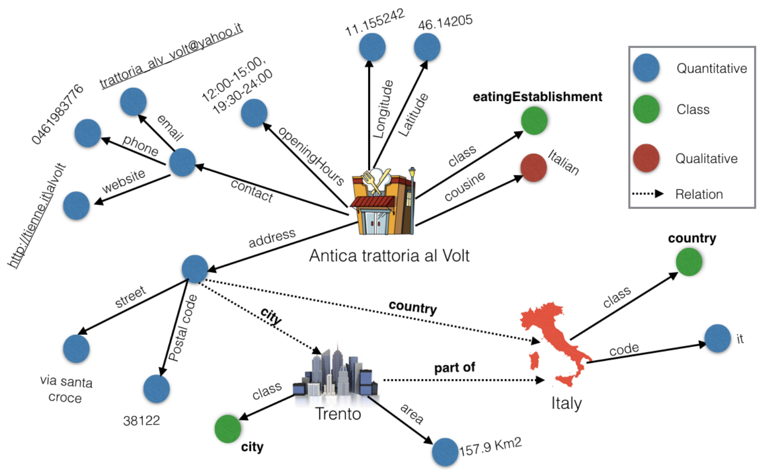


Figure 10.3: An entity graph- The legend describes different kinds of properties. The blue is a quantitative attribute; brown is a qualitative attribute. Green is a class, and dotted arrow line is a relational property.

The motivating problem in Figure 6.1 where the information about the individual restaurants is now transformed into a set of interconnected entities and relations as shown in Figure 10.3. Here, Al Volt now belongs to the class restaurant, spatial attributes like longitude and latitude, contact Phone, relational property address further linking to other entities like cities and country are merged from various datasets into a dense entity graph. Hence, instead of seeing partially created a view using separated datasets, the user can visualize the entire entity Al Volt holistically where different properties (relational and attributes) are distinguished. This arrangement further allows seamless navigation modality from one class to other classes, i.e., from a restaurant we can either navigate to a city or even further to a country or people working or in the restaurant (if data is available) and so on. In the section below, we will also discuss the three various components of a Knowledge Base.

User gets a full feature of an entity selected holistically from an entity graph.

1. Language: Language lexicalize knowledge using natural string.
2. Concept: A single specific token that connects lan-

guage and schema.

3. eType: It is a template built using concepts.

10.2.1 Language

Language codifies information from 'natural language' and further lexicalize it as the concept. The language describes an entity of a specific eType in a given domain corresponding to an entity class, relations, and the attribute names as well as the corresponding attribute values and also helps its instantiation in multiple languages. For example, Disease is defined as an impairment of health or a condition of abnormal functioning in English or *Malattia* is defined as *un indebolimento della salute o un funzionamento non normale* in Italian are the same concept in the different language.

Language codifies information from a natural language and lexicalizes it as the concept.

10.2.2 Concept

The concept is a core component of the knowledge representation which builds a connection between the *language* and the *schema* [Das and Giunchiglia, 2016]. Each concept represents a single specific meaning and is identified by a concept code. We use the concept as a formal notion denoting an element of a domain. They represent entity classes, relations, attribute names and attribute values. Concepts are used at the formal level to define an entity. For example, a concept 'building' can be attached to various label as *building* in English and *edificio* in Italian.

We use concept as a formal notion denoting an element of a domain.

A concept can be further expressed as:

$$Concept => (EC, NS, AD, QA) \tag{10.1}$$

Here, a concept has a *semantic relation* (e.g. is-a, part-of, component-of) with its parent/child concept (e.g. building is-a structure). A concept can also represent a *synset*. A *Synset* is a set of cognitive synonyms and contains terms (e.g., building, edifice) associated with the particular concept. A *lexical relation* (e.g. synonym) shows the relation

A concept has various types of relations with its parents and child concept.

between terms within a *synset*. *Semantic lexical relation* (e.g., hyponym, hypernym) is used to denote the relation between synsets. *Gloss* provides natural language description (e.g., building: *"a structure that has a roof and walls and stands more or less permanently in one place"*) of the concept. It helps to eliminate issues related to *heterogeneity in meaning*.

10.2.3 eType

We have briefly discussed what an eType and various sub eType in the section 8. An eType or schema is further defined using a quadruple as in the equation below.

$$eType = (ID, EC, NS, AD) \tag{10.2}$$

Where,
ID is a unique identifier;
EC is a concept denoting the class of the eType;
NS is a name of the eType;
AD is a non-empty set of Attribute Definitions. AD determines the set of attributes that can be associated to an instance of a certain eType. An attribute definition is a tuple, defined as:

$$AD = (ID, AN, DT) \tag{10.3}$$

Where,
ID is a unique identifier (here we are using the ID corresponding to a concept);
AN is the concept denoting the attribute name;
DT is a DataType.

Vocabularies are mapped using a concept.

The full meta-model of the eType is illustrated in Figure 10.4 where all the vocabularies are mapped using a concept. For example, eType Name, class, attribute definition and qualitative attribute are connected with the concept.

The model is also compatible with various semantic web formalism as RDF(S).

Our model presented in the Figure 10.4 is also compatible with existing semantic web 'formalism'. For instance, an

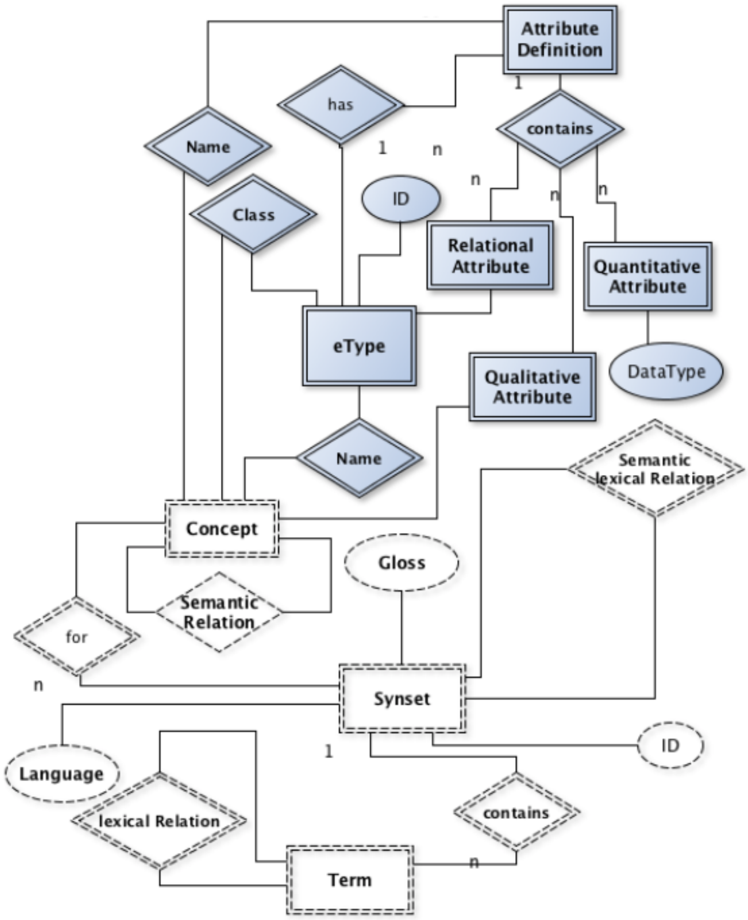


Figure 10.4: The figure describes the eType Meta Model.

eType is directly mapped with the RDF(S) as in Figure 10.5. Here, eType class such as building and event is mapped with *rdfs:Class*; attributes of an eType such as building name mapped with *rdfs:literal*; qualitative attribute which store concept (e.g. condition of a building either functional, ruin, abandoned) mapped with *skos:concept*. This functionality fulfills nonfunctional requirement NFR10.

A flexible model that addresses the functional aspects of all different scenario is key to achieving optimal results. To adjust diversity to the model we have aligned our model with DOLCE a top-level ontology [Guarino, 1998].

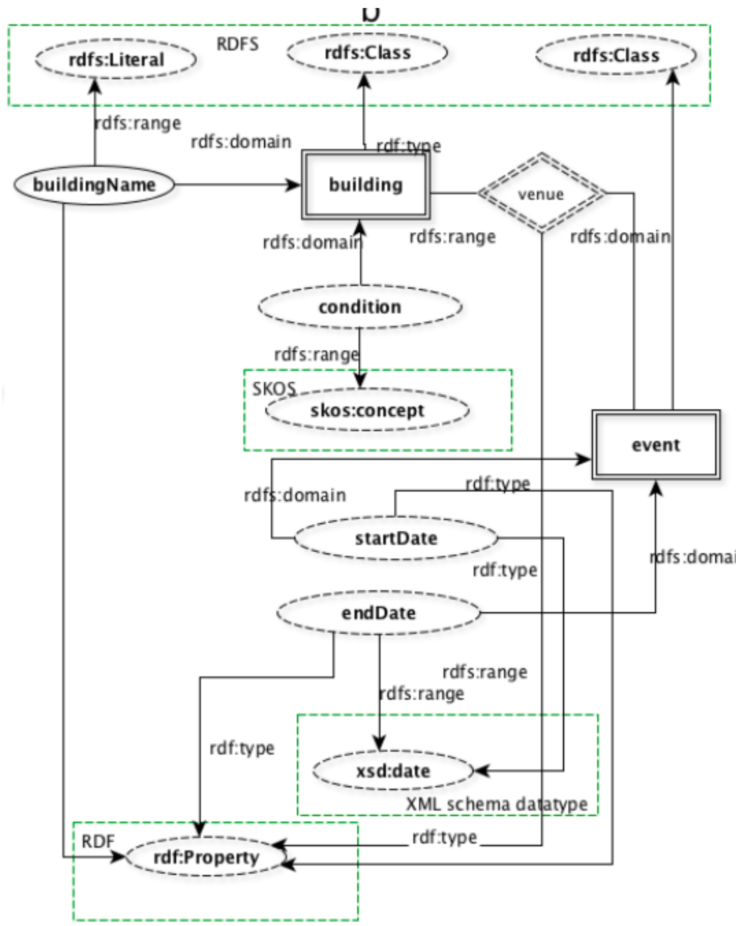


Figure 10.5: Metamodel mapped with to RDF(S).

This architecture helps us fulfilling various requirements.

Implementing an entity-centric approach in our solutions setting directly handles our functional requirements number FR1, FR2, and FR5.

10.3 Mapping Layer

Mapping layer interacts with knowledge and visualization layer.

It interacts with both *Knowledge layer* and *Visualization layer* and fetches specific parts of the entity graph from the *knowledge layer* according to the request issued from the *visualization layer* based on the user's intention. Then builds the result and present it to the interface. The major role of this layer is to select and build both queries and results. Figure 10.6 shows various components of mapping layer.

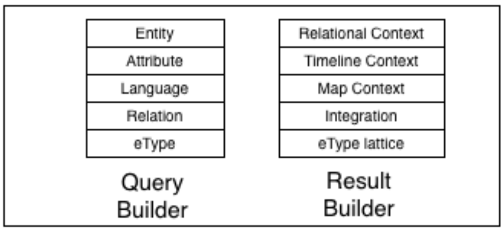


Figure 10.6: The Mapping Layer.

This layer builds the query and performs knowledge selection and filtering based on the request(s) made from the User Interface. The mapping is done according to the criteria provided in the User Interface layer. There are four ways in which selection can be made:

Selection and filtering can be done using four possible ways.

- *Entity selection* - A user can select different entities based on their types.
- *eType selection* - A user can select an eType and get all the entities as a result.
- *Attribute selection* - A user can select different attributes of the entity to visualize.
- *Relation selection* - A user can select a different kind of relations when visualizing the entities.
- *Language selection* - Based on the language that a user selects, entities can be visualized in different languages.

After receiving the result from the knowledge, layer binds the result and send it to the User Interface. This binding is possible in four different ways as:

It then builds the result and determines various visualization contexts.

- Entity integration: The entity is integrated after querying various properties for an entity from multiple entity bases.
- Timeline context: The attributes that define a timeline are then sent to the timeline engine on the User Interface.
- Map context: The attributes that define a map are then sent to a map engine on the User Interface.
- Relational context: Relational properties are then sent to Network component.
- eType lattice: Fetches an eType hierarchy and presents on the User Interface.

This technique detaches business logic from the User Interface.

This kind of arrangement detaches the knowledge from visualization and helps on writing incremental business logic in the mapping layer making the code more manageable.

10.4 Visualization Layer

The User Interface mimics the human-like perception.

Since it is possible to adapt the knowledge organization as humans, we also simulate the visualization in a human-like adaptation [Giunchiglia, 2016]. We use a single object then visualize it in space and time. As we perceive things in space and time, similarly we have abstract views like connected network views. Figure 10.7 shows various visualization contexts.

The entities are presented on the interface according to the user context.

The attributes description portrays that an object is somehow related with every other object.

The UI needs to handle and visualize variability in content. This variability needs to be captured both at design and run-time. At design time, it is addressed by representing data as an entity graph defined in the knowledge layer, whereas at run-time we need a flexible mechanism to select what we want to visualize from the entity graph and how we can visualize them. This representation implies that the presentation of the data is rather based on a different context of an eType. Description of various attributes in the section 8.3 shows a pattern that guides the contextual visualization and also helps map relation pattern for diversified

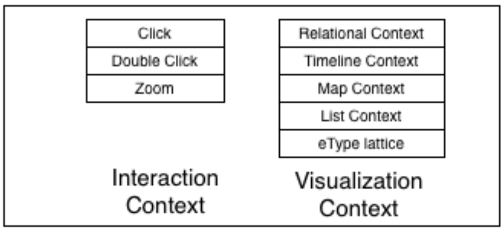


Figure 10.7: The Visualization Layer.

data exploration. We claim that there are two possible contexts as:

10.4.1 Visualization Context

The eTypes play an integral role in defining the visualization metaphor. eTypes defines the characteristics of an entity which furthermore guide the context. Since there are few eTypes, and each eType consists of a set defining attributes that guide the visualization. For example, in the context of space entities are visualized with a map and shapes. In the time context, entities are shown using time control objects (such as timeline, points, and intervals). The relational context (network) uses a graph-based presentation (with nodes and links). The list context represents a plain list that enumerates entities and their properties.

We implement multiple simultaneous views as space, timeline, network, and tabular to present data in different ways. The primary design rationale behind multiview visualization is to allow meaningful presentation (by removing all the technical jargon) and navigation of the semantic content to the end users (FR3, FR6, and FR9). Here, each view allows users to visualize entities based on their different properties. Different views provide a unique visual inter-

We use multiple simultaneous windows to present an entity.

Context / eType	Space	Time	Network	List
Location	Yes	No	Yes	Yes
Person	Yes	Yes	Yes	Yes
Organization	Yes	Yes	Yes	Yes
Artifact	Yes	Yes	Yes	Yes
Event	Yes	Yes	Yes	Yes

Figure 10.8: eTypes and their visualization contexts.

pretation of an entity. This leads to a multitude of perspectives and allows for flexibility in visualization, which further helps users to understand entities from different visual perspectives. In this work, we posit that there are at most four possible ways in which a real-world entity can be visualized. They are as follows:

- Timeline- Entities that possess time-related attributes are presented as a timeline view. For instance, date of birth of a person, construction date of a structure or an event date can be displayed in a timeline.
- Space- Entities that possess spatial (coordinate related) attributes are presented as a space view. We choose maps as a representation technique for depicting entities on space. For example, different buildings such as Pantheon or Empire State building can be shown in space.
- Network- Entity relating to another entity is pre-

sented as a network view. Here, a node represents an entity and a labeled edge depicts the relation between those entities. It is very useful for the end user to check other hidden relations of a certain entity.

- List- Entity's properties are shown as a tabular view. This table is ranked and sorted. This view is suitable to show the state (i.e. condition) and inherence (i.e. quality) type attributes.

10.4.2 Exploration Context

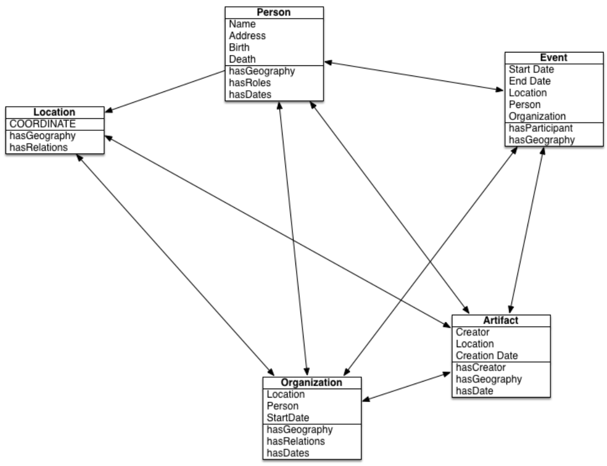


Figure 10.9: eType Mesh, showing relationship patterns in the schema.

Figure 10.9 shows the interconnection between various eTypes. This pattern makes it possible to generate a flexible framework for developing a generic exploration solution. According to this pattern, every object is related to one or more objects which creates an exploration metaphor. The inherent nature of eType, being highly interconnected, multi-dimensional and self-descriptive, provides an excellent framework as a foundation for the generating a contextual exploration solution. Apart from that, since every object is uniquely identified, facilitating disambiguation tasks

Unique identifier facilitates disambiguation of objects across multiple domains.

as well as uniquely representing distinct objects, entities, features and characteristics across domains.

10.5 Implementation

We designed various low and medium fidelity prototype.

We generated a plethora of low, medium and high fidelity prototypes before implementing the final design. Low fidelity prototypes were designed using simple drawing tools whereas medium-fidelity prototypes were designed using Balsamiq[1] and Mockplus[2]. The tool SemUI was implemented using *node js*[3] and *Angular js*[4]. The implementation can be divided into three parts as described in the implemented architecture as in Figure 10.10.

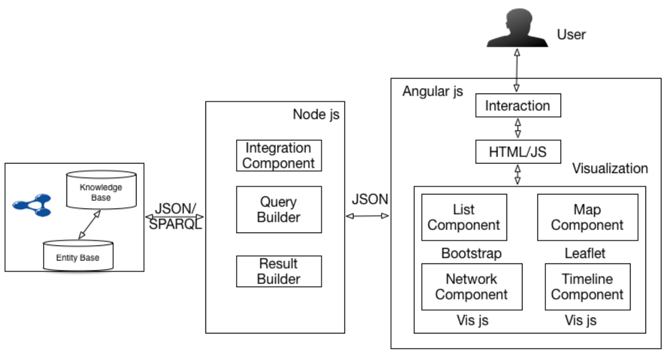


Figure 10.10: The Implementation Architecture.

The implementation supports querying either an Entity Base, a knowledge Base or a Triple Store.

In the data layer, there is an entity base or a knowledge base(FR1,2). Since our approach is adaptable to Semantic Web formalism, we also support querying triple stores(NFR10). The mapping function is implemented as a separate node js application. Finally, the interaction component and visualization component are implemented as an

[1]https://balsamiq.com/
[2]https://www.mockplus.com/
[3]https://nodejs.org/en/
[4]https://angularjs.org/

HTML/JavaScript application. The application followed a RESTFUL design approach where data were queried from the entity and knowledge base using web services.

10.6 Deploying SemUI

SemUI is deployed following a client-server technology design pattern. According to Figure 10.1, the knowledge layer is deployed as a server application. It hosts all Entity base and Knowledge Base as separate databases. We use PostgreSQL[5] for storing both entities and knowledge. Its URI uniquely identifies both Bases. The API for fetching results from both entity and knowledge bases are predefined. Based on the case studies, we import various entities and knowledge using KOS[6] UI. KOS is knowledge and an entity modeler. We use KOS to introduce and manage both entities and knowledge. For instance, on a health-care domain, 1725 entities and 338000 concepts (1.7 million terms) were imported.

We use PostgreSQL as our database system.

KOS UI tool is used to import datasets and concepts

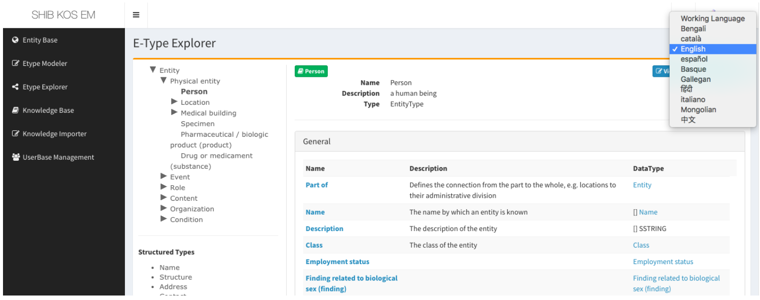


Figure 10.11: KOS UI.

The contextual layout in the UI layer stays consistent i.e., no business logic is written on the UI. UI only deals with the interaction and presentation schemes. However, the UI is separately generated, and it requires tweaking the mapping layer based on the scenario making it one separate implementation. The modular implementation scheme will allow easy tweaking of the mapping layer. In future, the

For every use case, a separate UI is implemented.

[5]https://www.postgresql.org/
[6] http://opendata-staging.disi.unitn.it:34500/shib-kos-em

Table 10.1: Summary showing how various requirements were met.

Functional Reuqirement	Solution
FR1: Solution has to follow an Entity Centric Approach	We implement the solution following an Entity-Centric approach.
FR2: Solution has to be adaptive in terms of entities and classification	We follow various eTypes based classification.
FR3: Facilitate object visualization	We implement presentation of entities by hiding all the technical jargon and visualize the entity in a human-friendly way using space, time graph and list.
FR4: Facilitate interaction with objects	We implement various interaction metaphor that allows interaction in a natural way.
FR5: Facilitate object searching	A searching functionality was implemented in the solution which allows searching for entities and types.
FR6: Facilitate exploring related objects	We separated attributes and relations and showed the relations on the graph-based view.
FR7: Facilitate simultaneous presentation of Entities	To allow users to see an entity from multiple perspectives, four different views were implemented showing specific properties of that entity.
FR8: Solution should allow easy adaptation to various context	This was possible through the eType hierarchy and the visualization context. Based on their required query, users could select a node in the hierarchy and view the results on various contexts.
FR9: Solution should be language independent	The Entity-Centric approach itself makes the solution language independent.

UI will be implemented inside KOS as a declarative application. Currently, KOS only supports the list based presentation. The list-based UI will be extended to show other visualization contexts as well.

10.7 Summary

In this chapter, we elaborate the implementation architecture of the solution. SemUI a multi-tiered solution approach makes an effort to reduce the gap between the user and the field of semantics. How the flexible architecture that handles, fetches and visualize the entities was explained. We also tried to map the solution to requirements stated in the chapter. All the requirements were fulfilled with the framework designed. Table 10.1 shows how various requirements were met.

Multi-tiered solution architecture was proposed in this chapter.

In the next chapter, we present various use cases where the designed solution architecture was implemented. Furthermore, many design solutions were evaluated with the end users following various evaluation techniques.

Various Use cases and their evaluations is presented next.

Part V

Use Cases and Evaluations

Chapter 11

Trentino Entitypedia

Based on the domain analysis we performed in chapter 5, gathered requirements in section 7 and to realize the effectiveness of an Entity-Centric approach in an Open Data setting, an experimental prototype[1] was developed using data from the province of Trento [2]. The prototype is based on the design finalized with various stakeholders (appendix C). Here, the data are stored in an entity base called Entitypedia. Aside from serving as entity repository, it provides entity-based services ranging from simpler (such as standard CRUD operations) to more advanced (such as context-dependent operations of search, matching or navigation) The contents of the entity base are described in Table 11.1.

Open Data from the province of Trento was used as an entity base.

Table 11.1: Entities classified according to various eTypes.

eType	Total
Location	20704
Building	141
Lodging Facility	442
Shopping Facility	219
Refreshment Facility	229
Transportation	3379

Developed prototype is shown in Figure 11.1. It contains four major parts: ontological menu, map, information con-

The prototype application had four main parts.

[1] http://www.opendata.entitypedia.org/spacetime

[2] dati.trentino.it: It is one of the major data catalogs with 875 datasets from different sectors.

The UI is multilingual.

tainer, and a relational menu. The ontological menu is placed on the left side and consists of different eTypes. It is created using a lightweight ontology. Currently, it supports an Italian and English. Based on the selected menu item (eType), the interface is refreshed, and entities are retrieved and visualized on the map. Participants can also

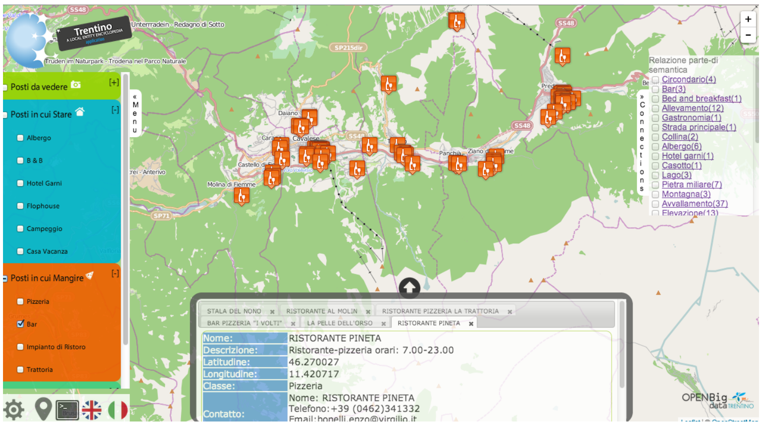


Figure 11.1: UI of the working prototype.

The objects on the interface are refreshed on any interaction made.

select multiple menu items. Once an entity is selected on the map, the information about that entity is shown on the screen and, the relational menu is created on the right side of the screen. The relational menu lists the entities that are either near or located in the same place. There are also standard setting, such as the default language, the radius to search within, and geolocational search for the entities relative to the user's position.

11.1 Evaluation

Usability and User Experience dimension of the prototype was evaluated.

The goal of the initial evaluation of the system prototype was to obtain insights on how different UX and usability dimensions are addressed. During a period of one week, nine students from different faculties in the University of Trento participated in the evaluation. The participants were students from Bachelors to Ph.D. with age ranging from 21 years to 29 years. Five participants were male, and four were female. They were given a brief overview of the system and shown how to use the prototype application called Trentino Entitypedia (TE). The participants were then asked to find and explore entities in TE. Each session lasted maximum for an hour. All the interaction of the participants and their comments were transcribed, and a video was recorded for future analysis. The participants were instructed to think aloud, and they were also asked to perform the tasks rationally without feeling any discomfort or stress. After completing the tasks, respondents were asked to fill questionnaires. The questionnaire consists of background information and questions that use 5-point Likert Scale to assess different usability and user experience criteria.

Entire participants interaction was recorded and transcribed.

Each participant was asked to spend some time on the interface to get familiar with it. Then they were asked to read the instructions about the system. Four related tasks were assigned to the participants as follows:

Participants were asked to get acquainted with the tool.

- Task 1 (T1): Find Bed and Breakfast.
- Task 2 (T2): Based on the T1, find any other Point of Interest (POI) that is in the same location (commune) as that of previously selected Bed and Breakfast.
- Task 3 (T3): Find POI within the range of 500M in Cavalese.
- Task 4 (T4): Find Sports Club in Cavalese.

In the following section, we describe both usability and User Experience dimensions for evaluation.

11.1.1 Usability Evaluation

Five usability dimensions were selected as follows: Usefulness (U), Learnability (L), Memorability (M), Satisfaction (S) and Visibility of system status (V). Evaluation study is summarized in table 11.2. Detailed evaluation dimensions and statistics can be found elsewhere[3].

Five usability dimensions were evaluated.

[3]http://goo.gl/6KtFLe

Table 11.2: Usability evaluation based on five-point Likert scale.

	U		L		M	S	V	
User	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
1	4	3	5	2	2	3	3	3
2	5	5	4	2	2	4	4	5
3	4	3	4	3	3	3	4	4
4	5	4	4	2	2	4	4	4
5	4	4	4	2	2	4	4	4
6	4	4	5	2	2	3	3	5
7	4	4	5	1	1	4	5	5
8	4	5	5	2	1	4	5	5
9	4	3	4	3	2	3	4	5
Mean	4.2	3.8	4.4	2.1	1.8	3.5	4.0	4.4
S.D	0.44	0.78	0.52	0.60	0.60	0.52	0.70	0.72

Q1 - I find the application simple to use.
Q2 - The system meets my needs.
Q3 - It is easy to learn how to use the system.
Q4 - I had to put extra effort to use the system.
Q5 - It takes some time to remember application controls.
Q6 - I didn't notice any inconsistencies while using it.
Q7 - I receive timely feedback on my actions.
Q8 - The information provided by the system is easy to understand.

11.1.2 UX Evaluation

The UX evaluation was done against five criteria.

The system was evaluated considering five UX dimensions. Aesthetics and Visual appearance (A), Emotion (E), Identification (I), Stimulation (S) and Meaning and value (M). Table 11.3 shows responses related to specific UX dimensions.

11.2 Results and Discussion

Participants were able to complete the tasks with ease.

Most of the participants were able to complete task 1 within minutes. They were also able to understand the relationship and to navigate through different relations. Task 2 also took less time to finish. The possibility to set the ra-

Table 11.3: UX evaluation based on five-point Likert scale.

	A		E		I	M	
User	Q1	Q2	Q3	Q4	Q5	Q6	Q7
1	4	4	4	3	4	3	5
2	4	5	5	4	4	4	4
3	3	4	4	3	4	4	4
4	4	4	4	4	4	4	4
5	4	4	4	4	4	4	4
6	5	5	5	4	4	4	4
7	4	5	1	3	5	5	5
8	3	5	5	5	5	5	1
9	2	4	4	4	4	4	4
Mean	3.667	4.4	4.0	3.778	4.2	4.111	3.889
S.D	0.866	0.527	1.225	0.667	0.440	0.601	1.167

Q1 - Seeing different types of entities on the same page is visually pleasant
Q2 - Icons clearly reflect the purpose of the specific entity on the map
Q3 - I like the idea that, using the system, I can discover new information connected to what I was looking for in the first place
Q4 - The fact that I can propagate through the network of connected entities makes this application somehow playful
Q5 - I see myself in using the application.
Q6 - The application gives me reliable evidence of entities in a sense that I get the information I expect
Q7 - The meaning of entities and their relations enables me to quickly navigate and extract the information I find useful

dius made the task 3 simple to be achieved, and participants were also excited to see the results. Task 4 was also easily performed by most of the participants.

On average, participants' assessments were quite high for all measured usability variables. The overall evaluation of the system was positive. Standard deviation from the mean value for each of the questions for the usability dimensions is also minimum. This result proves that system was useful for almost all of the participants. The result of usability evaluation is shown in Figure 11.2.

The assessment of the usability variables was high for most of the participants.

On average, the participants' assessments were positive for

The UX evaluation was also positive.

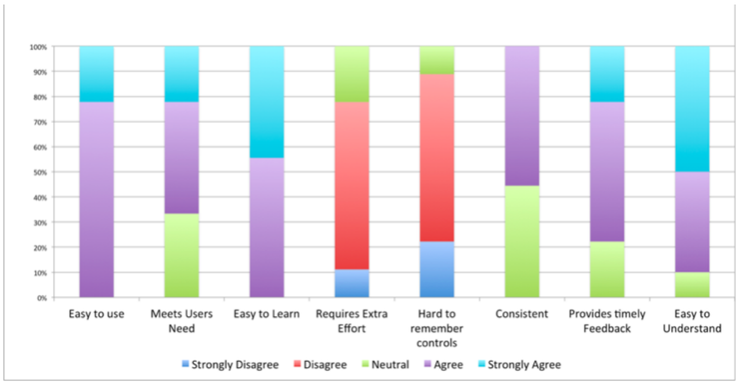


Figure 11.2: Result of the usability test.

all of the UX variables. Participant 7 and 8 gave negative remarks for Q3 and Q7, respectively. These remarks resulted in higher deviation for those questions from the mean. We explain this in the comments they have provided. User 7 noted that the system has easy traversal mechanism, but the discovery of information may not be intuitive. Similarly, user 8 mentioned that though it is easy to find entities and relations, the meaning of the words in Italian was not intuitive in some cases. For example, Cateratta was used to describe waterfall, whereas the Cascata may be an appropriate word. This issue comes from the quality of underlying entity-centric open data. The result of usability evaluation is shown in Figure 11.3.

However, some participants suggested that the concepts used were not evident.

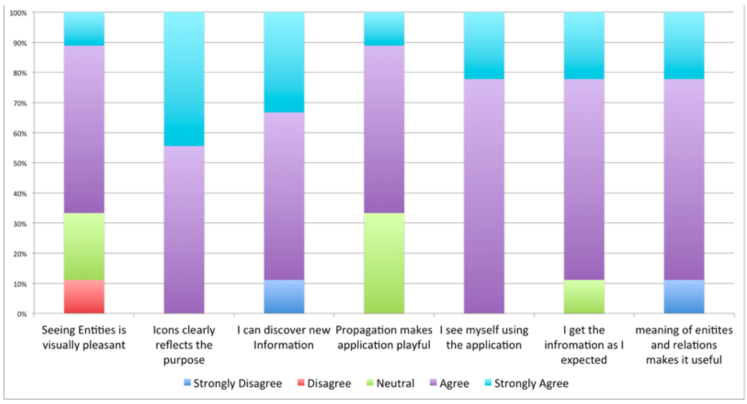


Figure 11.3: Result of the User Experience evaluation.

Both ontological and semantic menus were useful.

The ontological menu on the left and the semantic menu were both considered as user-friendly. Left menu proved to be well organized as it lets participants find specific service in less time. Participants also liked the possibility of setting entity search range on the system. Overall, the interactivity and the integrated environment was interesting for participants. Some minor bugs were also discovered during the process. The check-box on the left menu behaved with certain issues. When an item was checked on the semantic menu (right menu), the participants expected that the previously selected entity would also be shown on the map. Some suggestions about font size and type were also provided. The system was slow, and participants expected quick results.

Some minor bugs were also discovered by the participants.

The idea of seeing entities of different types on the same page excited the participants. Participants felt it is less time-consuming and can get a clear picture of what they were looking. Participant 1 mentioned that the interface was intuitive and exciting. Participants liked the idea of the consistency of menu color with an icon color which showed clarity and symmetry. For some participants, the green icon was not helpful as it also matched with the color of the map. Some icons like for street also did not make any sense for the participants. For most of the participants finding new information was an innovative thing. Participant 6 expressed that she enjoyed the way in which connected entities can be traversed. Participant 4 was surprised to see the opening and closing time for POI. All participants asked if they could start using the application immediately. They also suggested that it will be beneficial for them if there was a mobile app. Participant 3 liked the menu. The arrangement of menu items seemed logical for him. Participant 2 was doubtful whether the menu item *Things to do* was expressive. He also thought that ski and club should be placed together.

The idea of visualizing multiple entities at once was exciting for participants.
The color coding and icons were also intuitive.

The participants showed their willingness on using the fully developed application.

Table 11.4 gives the percentage of participants that indicated application features concerning specific UX dimensions. With respect to the identification, the participants were asked to choose the elements that reflect their habits in using similar systems. Regarding stimulation, the participants were asked to choose the specific features that encourage them to use the application.

The application was also considered stimulating.

Table 11.4: The application features rating concerning UX dimensions.

Application Feature	Identification	Stimulation
Relations/links between entities on the map	63%	63%
Entities details on demand	38%	38%
Entities view scalability	25%	25%
Entities view filtering	38%	50%
History of browsed entities	25%	13%
Flexible exploration and navigation through entities		50%
Flexibility/contextual switch in showing entities from different domains		38%

The table shows that the most exciting application features are relations between entities, view filtering and flexible exploration and navigation through entities. It also shows that the participants did not notice some of the features. These features require further improvement.

Participants suggested additional features as well.

Participants also suggested some features. Directional features like in Google Maps, direct search functionality, facet level service on attribute level and icons everywhere. More interactive help was also pointed out as needed. Some participants also suggested accommodating the relational menu (right menu) directly inside the information container. Participants also suggested moving the information panel which showed the details about an entity from the bottom of the page to the top.

Chapter 12

Trentino Entitypedia-2

The feedback that we received from the participants in the development of Trentino Entitypedia (chapter 11) were not ignored. Considering the changes suggested by the participants, the interface was redesigned to cater their needs. The primary issue was to include direct search functionality which allows searching for both instances and types. Two different search boxes were added. One search box on the left enabled searching for eTypes (Figure 12.13) and the second allowed searching of entities (Figure 12.1. Similarly, the information panel was moved to the top of the screen. Icons were added to the semantic relation menu as well. The instance count feature was also color coded. The redesigned prototype is shown in Figure 12.1.

The interface was redesigned for the scenario considering all the user feedback.

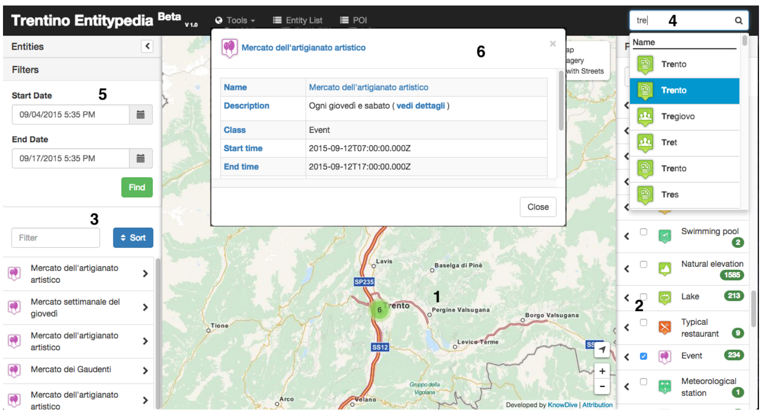


Figure 12.1: UI of the working prototype.

As more datasets were inserted, more scalable UI representation was required. To handle this situation we followed a multi-layered approach.

In addition to the user feedback, the information need also expanded. More datasets from various open data categories were now converted to entities. This situation resulted in the generation of complex entity graph. The shape of an entity graph varies in type, number and size of the entities and their relations. A user may want to visualize different kinds of content with various levels of detail or look for particular information.

The requirement also suggested that the entities can be visualized in many different ways regarding context and content.

The UI should provide flexibility in the visualization of diversified data in different ways. Different visualization contexts present data in different ways, for example in space, time or social. This implies that in our case, the presentation of the data is not driven by their nature, but rather by the context. Furthermore, UI should be able to handle and visualize variability in content. This variability needs to be captured both at design and run-time. Therefore, we had to separate what we visualize and how we do it. To cater these requirements, UI was designed following multi-tiered approach as discussed in section 10.1.

The developed UI was not scalable to visualize various other relations.

We realized that the UI designed in Figure 12.1 was not scalable. We hence designed a new interactive prototype represented in Figure 12.2. The prototype was designed using the Balsamiq[1] mockup tool. As from Figure 12.2, on the left side, the prototype contains a hierarchical menu with different eTypes, such as Person, Location, Artifact, Building, and Event. On the right side, the main view shows the different visualization contexts, in particular, space, time, graph and list. Different filtering options which allow selecting the content are available under the menu, on the left side. Only two visualization contexts are shown in the figure - the time and the graph context.

12.1 Evaluation

We evaluated against two hypothesis.

To fully evaluate our approach, we developed a proof-of-concept prototype for our solution. We experimented to evaluate the prototype. The hypotheses of the experiment

[1] https://balsamiq.com/

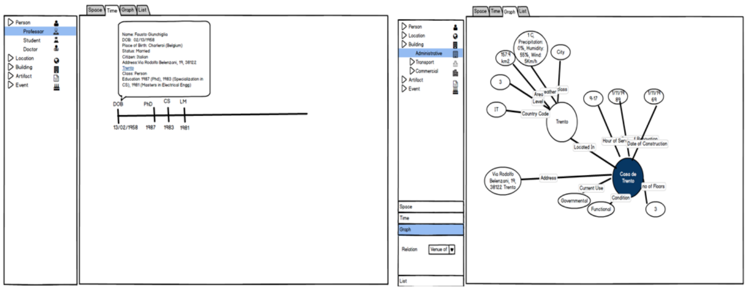


Figure 12.2: The proof-of-concept prototype. The views of two contexts are shown - time and graph. The left part of the view shows the visualization filters, whereas the right part renders the resulting content in the specific visualization context.

were:

- *H1*. It is effective to visualize different types of entities in the same visualization context.
- *H2*. It is effective to visualize the same entities in different visualization contexts.

We designed a set of six tasks to be performed using the prototype. The tasks were as follows:

1. Find a class city from the eType menu on the left.
2. Locate the specific city - Trento using the interface.
3. Explore different contexts and understand how visualization allows different properties for the chosen city.
4. Explore and traverse related entities.
5. Visualize different entities in the same context.
6. Use filters in the graph visualization context and see how the corresponding view changes.

15 different computer literate participants took part in the experiment.

The experiment was carried out with 15 participants (10 male and five female persons). The age range of the participants was as follows: 18-25 (7 participants), 26-29 (5 participants) and 30-39 (3 participants). The education level of the participants was as follows: undergraduate degree (3 participants), masters degree (9 participants), Ph.D. degree (2 participants) and 1 participant was working professional. The experiment was organized on a voluntary basis. All participants were computer literate.

A standard computer in the participants own setting was used for the experiment.

The experiment was conducted using Mac book Pro, 13-inch screen, 2.5 GHz Intel Core i5 Processor, Intel HD Graphics 4000 1024 MB graphics card, 4 GB RAM. As per participant's choice, either the Trackpad or the optical mouse was used.

Each session lasted for half an hour, and the aim was explained.

The data were collected during individual interview sessions. Each interview session lasted around 30 minutes. All the participants were described the aim of the study, the notions of entity and visualization context, and the prototype. Participants were asked to complete the tasks. While doing the tasks they were asked to think aloud. They were continuously invited to express their feelings about the interface. Upon completion of the individual task, the participants were asked to give feedback on how they perceived the task. All sessions were recorded. Upon completion of the tasks, each participant was asked to complete the online questionnaire.

12.2 Results and Discussion

The tasks were easy to understand.

The participants performed the tasks they were given, and neither of them reported any issues in understanding the tasks. They got familiar with the prototype quickly. They used the prototype without any assistance. A problem reported by three participants was locating the class city in the menu.

Location type did not have time context causing problems.

As for the time context view, the participants expressed satisfaction with the way the information was organized on the screen. When trying to visualize entities of the location type, the participants received the message that the temporal visualization was not possible. This arrangement was fine with the participants since this type of entity does not contain temporal properties.

As for the graph visualization context, the participants gave recommendations on what to pay attention when using this visualization context. Three participants pointed out that there was too much information making navigation between different entities difficult. Concerning this, all participants found that filtering was useful to reduce the information load. Three participants were skeptic about the fact that the graph context would scale if an entity had many relations. One participant suggested clustering of the nodes. Another participant expected a hyperlink in all the relational nodes. All the participants appreciated the use of colors to highlight the currently selected entity in the graph.

Participants suggested that the network was dense making it hard to understand the information.

After completing the tasks, the participants were asked to fill a questionnaire. The questionnaire used the 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The questions were the following:

Participants filled in a questionnaire.

- *Q1* - It is easy to visualize the different type of entities grouped into different categories (categories like a person, location).
- *Q2* - Discovering new entities to what I was looking for in the first place is exciting.
- *Q3* - Propagating through a list of entities is playful.
- *Q4* - Visualizing an entity in different contexts is exciting (for example, visualizing the person in space, time, graph and list).
- *Q5* - Visualizing different entities in the same context is meaningful (for example, visualizing a person and a location in the graph).
- *Q6* - Filtering information to focus on the specific portion of the entity was helpful.
- *Q7* - Getting a detailed view of an entity was helpful.

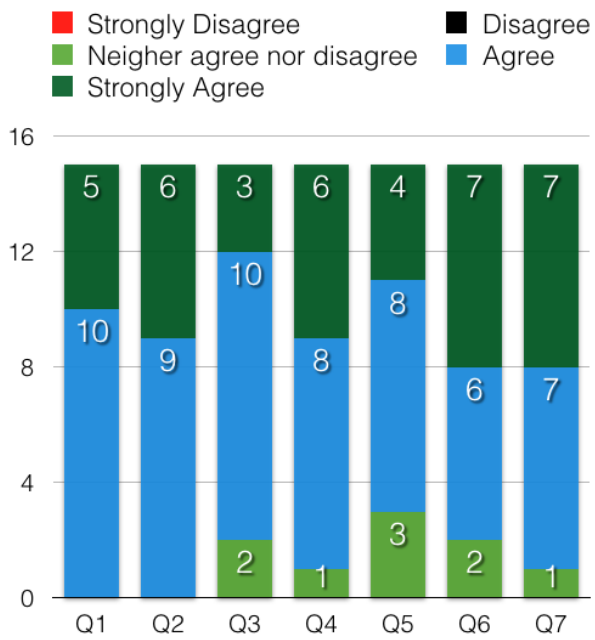


Figure 12.3: User responses on a five-point Likert scale.

The visualization was easy for users.

The results are reported in Figure 12.3. Concerning the hypotheses, we analyze the answers. We conclude that The participants found it easy to visualize entities of different types (Q1). As for the exploitative behavior (Q2, Q3), the respondents reported a positive experience, except two neutral answers. If we look at filtering behavior (Q6), we see that the respondents expressed positive attitude (with two neutral answers). The level-of-detail behavior (Q7) gives a similar result. Similarly, visualizing the entity in different settings (Q4) resulted in positive user experience, with only one neutral answer. Visualizing different types of entities in the context (Q5) was meaningful to the participants. However, this feature received three neutral answers. Given the hypotheses above, participants' assessments of the prototype's features were positive.

Various visualization contexts provided positive User Experience.

Chapter 13

Digital University

Evaluation of the hypothesis in the Trentino Entitypedia 2 as described in section 12 showed the visualization benefit following a multi-layered approach. We designed and developed robust interface called *multi viewlets* following the feedback we received. The prototype is based on the university's portal where different stakeholders such as students, staff, and outsiders (prospective students, collaborators or visiting faculty and so on) are continuously searching for different kinds of information. We generated separate random anonymized entities related to people, departments, publications, location and events happening in the university.

We learned the benefit of multilayered approach and by the feedback, we designed a multi-view tool.

Figure 13.1 shows the implemented prototype of the four multiview *viewlets* in the digital university setting. The main motivation of the designed prototype was to show different properties of an entity at a glance simultaneously. The prototype is based on the university's portal where different stakeholders such as students, staff, and outsiders (prospective students, collaborators or visiting faculty and so on) are continuously searching for different kinds of information. We generated different random anonymized entities related to people, departments, publications, location and events happening in the university.

The prototype showed multiple facets for an Entity in views called viewlets.

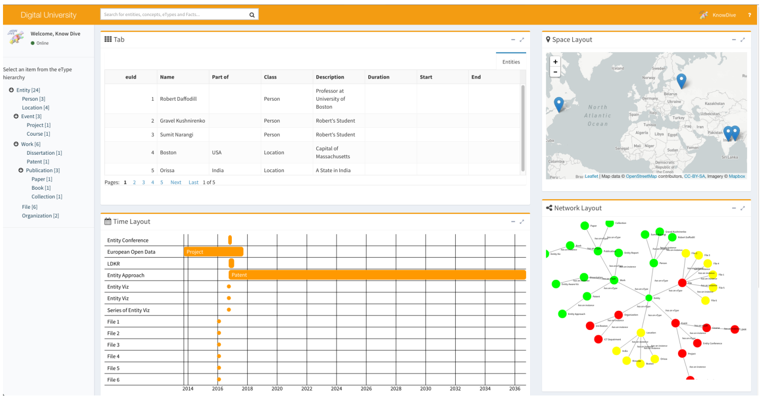


Figure 13.1: Multiview Entity Visualization. There are multiple *viewlets* called timeline (e), space (d), network (f) and tabular (c) *viewlets*. Search box (a) allows direct entity search. eType hierarchy faciliates exploration (b). History option facilitates traversing back to the previously explored entity or eType (g).

13.1 Evaluation

A preliminary user study was conducted to understand the meaningfulness of multiview viewlets.

We performed thematic analysis to validate two hypothesis.

We conducted a preliminary user study to assess the meaningfulness of multiview *viewlets*. We applied mixed method approach in our research and directly observed the way participants performed the task. An open-ended interview session was conducted at the end of the task. Data from the interview session were transcribed, analyzed and was categorized into a set of themes using thematic analysis [Braun and Clarke, 2006]. The transcribed notes were mapped to the different UX dimensions [Laugwitz et al., 2008]. The themes were both independently and collectively verified by five other researchers a snapshot of them participating in the grouping of terms is shown in Figure 13.2. Synonymous terms used in the interview notes were also verified using a lexical knowledge base called Wordnet [Miller, 1995]. Wordnet groups nouns, verbs, adjectives, and adverbs into a set of synonyms called synsets which express a distinct concept. We mainly focused on the adjectives in the transcribed notes and came to an agreeable list of codes and themes after several trials. The main argument of the analysis was to understand benefits of the multiview *viewlets* based on following hypotheses:

Figure 13.2: Independent researchers grouping attributes based on Thematic Analysis.

1. *H1*. Seeing same entity using different *viewlets* simultaneously is meaningful.
2. *H2*. Seeing different related entities facilitates easy exploration of related entities.

A total number of 10 participants (9 Male, 1 Female, age range 23-42, mean age 28.5 and S.D 5.04) took part in the user study. Five of the participants were fully aware of semantic technology and tools whereas the other five had no prior knowledge regarding semantics. Participants' highest degree of education ranged from bachelors to Ph.D. degree. All of the participants possessed a good understanding of English and volunteered for the study. The consent form was signed before the beginning of the study. We commenced the task by gathering demographic information and asked each of the participants if they knew any similar systems with the multiview layout. Only one participant gave an example of a similar system[1]. The tool mentioned do not offer multiview visualization of entities on a single layout. After that, we asked to point out advantages and disadvantages of the multiview *viewlets* to the participants. Furthermore, we also asked to list the advantages

10 participants participated in the study. All of them were computer literate.

Participants were unable to remember various tools with similar functionality.

[1] https://goo.gl/de4JTe

and disadvantages of network *viewlet* to validate both of our hypotheses.

Seven individual tasks were designed and verified by the independent researchers.

Seven individual tasks were designed and verified by two independent researchers (by a pilot study). All of these tasks were aimed at understanding presentations of entities into different *viewlets* (space, timeline, network, and table) and then explore separate related entities using one of the views (network). The tasks were:

1. (T1) Search a person named Gravel Kushnirenko;
2. (T2) Find his birth date;
3. (T3) Find his birthplace;
4. (T4) List one paper he co-authored with Sumit Narangi;
5. (T5) Find the conference where the paper was presented;
6. (T6) Where was the conference held;and
7. (T7) Starting date of that conference

The user guide was also prepared and handed to the participants.

We described the tasks thoroughly to the participants. We handed them a user guide to understand and get acquainted with the prototype. Each participant was allowed to play around with the visualization tool until they got used to it. Participants were asked to think aloud. The whole session lasted approximately for half an hour. The tasks were performed on the participants' desired setting. Four different MacBook pro with 13-inch display was used to conduct the entire experiment. All of the participants were provided with an external mouse.

13.2 Result and Discussion

In our study, we explored the meaningfulness of multiview *viewlets*. Our initial prototype shows that the age, qualification, and knowledge of semantics had no significant impact on the usage of the prototype. Out of 10 participants, eight were able to complete all the tasks successfully. Eight of the participants needed some assistance to complete task T4. Irrespective of gender, the only female participant was able to complete the task T4 without any assistance while out of nine male participants only one completed. However, this cannot prove that female won't need any support. We need more female participants to verify if the gender did affect the usage. One of the participants was biased towards full functionality rather than the visual layout and meaningfulness of the system which the participant mentioned throughout the session.

The result shows that the age, qualification and pre-knowledge semantics has no significant impact to determine the usability of the tool.

The task completion time is shown in Figure 13.3. The completion time was measured in seconds. There were no complaints on the layout from most of the participants. However, participants faced an issue with a timeline and network *viewlet*. The issue was majorly related to visibility of the text. The overlapping text created problems for users when performing tasks related to dates and relations. They were able to complete those tasks using alternative visual layouts. There were different ways in which the task could be completed. Participants chose their way, and there was no influencing pattern.

Participants complained the size of the font and an overlapping text.

Seven out of ten participants chose to search an entity (T1) through search box whereas three of them searched through the eType hierarchy. The ones who decided search box option were able to complete the task faster. There was no apparent reason why some chose the search option while the other decided to explore using the eType hierarchy. Most of the participants referred to different visual layouts simultaneously. While performing task T2 and T3 participants preferred table and space layout respectively because the overlapping dates hindered participants to complete task T2 from timeline *viewlet*. The task of finding a paper co-authored by two people was difficult for most of the participants. The result was not obvious and participants were required to think analytically using different layouts to complete the task. All of the participants chose network layout to complete this task. 3 of the participants also took significant time to complete task T7.

Participants mostly used the search feature rather than exploration. The reason for this is unclear.

All the visual layout were highly appreciated.

The task of finding papers co-authored was difficult for participants.

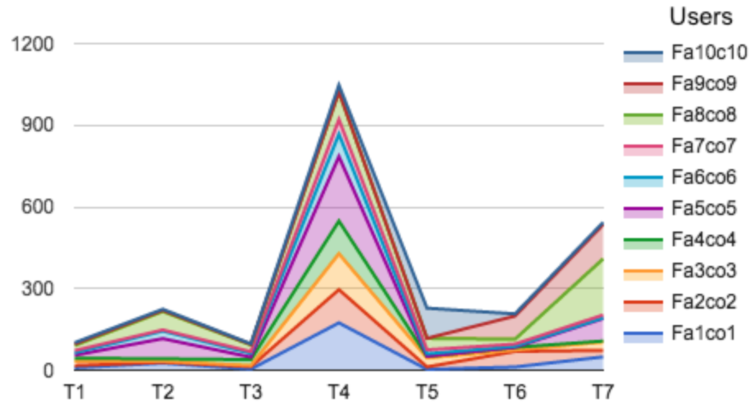


Figure 13.3: Task completion time for all the participants.

The result of the thematic analysis was meaningful.

From the thematic analysis, we found that the overall prototype was meaningful for the participants. The interface was clear, coherent, understandable and well organized. It was also valuable to them as it showed new insights into the different meaning of entities and their relations. This organization allowed participants to gain a different perspective on visualization simultaneously. The prototype was interactive, easy, friendly with a significant focus on the natural exploration of entities. Participant felt the need for such kind of tool.

The prototype was interactive.

Participants suggested more filtering options. They also mentioned it was hard to learn.

Nonetheless, the interface also caused some problems for the participants. The current issues such as overlapping of text in the network and timeline viewlet caused the central problem of confusion and obstruction. Furthermore, participants mentioned that it was difficult to learn. To solve these issues, participants suggested filtering support to remove blockage and implementation of color codes or icons to remove visual confusion.

The result supports the hypothesis.

The result from the task observation and the following interview session corroborate the proposed hypotheses of *H1*. Seeing same entity using different *viewlets* simultaneously is meaningful; and *H2*. Seeing different related entities facilitates easy exploration of related entities.

Chapter 14

Events in Trento

SemuI was finally implemented in the events scenario. The solution is designed to fetch an entity from the knowledge layer and then present various properties of that entity simultaneously on the visualization layer. In the tool, four different views are projected side by side in the same window. As a start, tab (Figure 14.1c) and network (Figure 14.1e) views are in the central part of the window. Whereas, space (Figure 14.1d) and tab (Figure 14.1f) views are on the right side of the window. Each view can be moved freely depending on user's preference. For instance, a user can drag and drop their preferred view from one position to another or they can minimize or maximize each view. The User Interface automatically understands the eType of the entities being visualized and then project them based on their defining attributes. For example, entities containing spatial attributes (e.g., longitude and latitude) are projected on a map whereas entities containing temporal attributes are projected on a timeline. The tab layout shows both attributes and relations. The network layout shows related entities along with their relation name. The user can search entities and eTypes using the search box.

SemUI with four distinct visualizations was implemented on this tool.
The visual components are simultaneously presented.

View can be dragged and dropped from a position to another position.

The eType hierarchy on the left-hand side(14.1b); where all concepts in the hierarchy are mapped to concepts defined in the knowledge layer. Furthermore, the User Interface also includes history list (Figure 14.1g) which lists previously searched eTypes and entities.

The tool also possessed an eType hierarchy and history support.

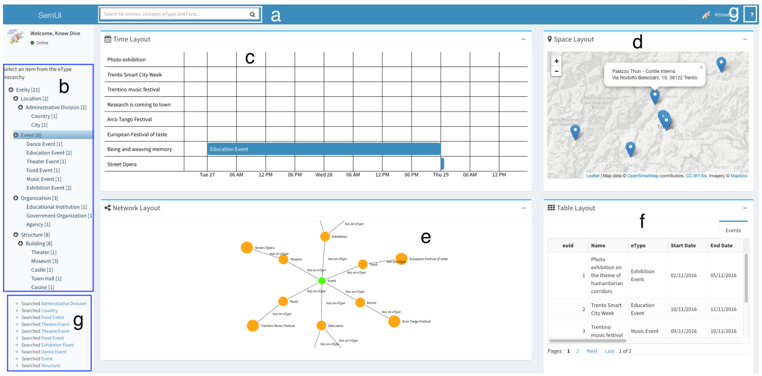


Figure 14.1: *SemUI*: semantic multiview visualization tool. It has multiple views as a timeline (c), space (d), network (e) and tabular (f). Search box (a) allows direct entity search. eType hierarchy facilitates exploration (b). History option facilitates traversing back to the previously explored entity or eType (g).

Datasets containing various events happening in Trento were imported.

The tool also incorporates different Open Data belonging to different categories related to events. We generated different entities related to eType building, event, location, and organization. The combination of sets of entities, eTypes from the *Knowledge layer* which is then shown in the *Visualization layer*. The major stakeholder in this scenario are tourists who are continuously searching for different kinds of information. The working of the tool is shown in an accompanied video[1].

14.1 Evaluation

Usability and User experience dimensions were evaluated.

We conducted a user study to assess the usability and experience of the multiple views. We asked the participants to perform a series of tasks followed by filling of an online questionnaire. An open-ended group discussion was also conducted at the end of the task. Data from the discussion session were further analyzed and mapped with the online questionnaire. The questionnaire was designed to understand different UX dimensions [Laugwitz et al., 2008] along with the specific traits of *SemUI*. These UX di-

[1] https://goo.gl/gVvZyZ

mensions perform a thorough assessment of the product using six scales with 26 terms. These scales were: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. These scales were further categorized as Attractiveness (relates to the overall impression of the system), Pragmatic (the usability of the system and basically consists of Efficiency, Perspicuity, and Dependability) and Hedonic quality(user stimulation while seeing the product). It consists of 2 scales: Stimulation and Novelty. After that, to validate the usefulness of *SemUI* we asked the participants to point out the advantages and disadvantages of the multiple views as well as to list out the advantages and disadvantages of network view.

The assessment was done against six different dimensions.

A total number of 20 participants (13 Male, 7 Female; 7 within an age range of 18-25, 12 within an age range of 26-30 and one within the range of 31-40) took part in the user study. Participants' highest degree of education ranged from undergraduate (1), postgraduate (14) to Ph.D. (4) degree. One participant was a high school graduate. The specialization of the participants ranged from computer science, economics, law, medical science and social science. Some of the participants were fully aware of the semantic technologies and tools whereas others had no prior knowledge regarding semantics. All of the participants possessed good knowledge of English and volunteered for the study. The consent form was signed before the beginning of the study.

20 participants from various fields participated in the study.

Seven individual tasks were designed and verified by two independent researchers (by conducting a pilot study). All of these tasks were aimed at understanding presentation of entities into different views (space, timeline, network, and table) and then explore different related entities using one of the views (network). The tasks were:

Seven individual related task that supported exploration of entities were designed.

1. (T1) How many educational events are there?;
2. (T2) Name one educational event;
3. (T3) When is the start and end date of that event?;
4. (T4) Who is the organizer of that event?;

5. (T5) Where is the venue of the event?;
6. (T6) What is the address of the building where it is held?;and
7. (T7) Is the venue disabled person friendly?

The objective of the evaluation and the tasks were thoroughly explained.

We described the tasks thoroughly to the participants and handed them a user guide along with an accompanying video to understand and get acquainted with *SemUI*. Each participant was allowed to play around with the visualization tool until they got used to it. Participants were asked to think aloud. The whole session lasted approximately for half an hour. The tasks were performed on the participants' desired setting. Participant's used their own computers to perform all the tasks.

14.2 Results and Discussion

16 of the participants completed all the tasks successfully.

Network, space and the tabular view was their preferred choice.

Some tasks and termed used were confusing for the participants.

The result of user evaluation[2] shows that the age, qualification, and knowledge of semantics had no significant impact on the usage of the *SemUI*. Out of 20 participants, 16 were able to complete all of the tasks successfully. The result from the task T1 shows that the participants can understand and navigate the eType hierarchy easily. All of the participants were able to complete task T1 without any problem. Only 1 participant was unable to perform task T2. Moreover, participant preferred the tabular view to complete this task. Most of the participants chose tabular view instead of timeline view to perform task T3. Task T4 required participants to explore another entity via network view. Most of the participants successfully completed the task. Similarly, participants had to use network view to complete task T5 as well. Space view was the preferred choice for the participants to complete task T6. Task T7 created confusion for almost half of the participants. We used standard controlled vocabulary designated for handicapped people. However, the term 'disabled friendly' was not clear to 8 of the participants, but after clarification, most of them were able to complete the task.

[2]https://goo.gl/bpNASM

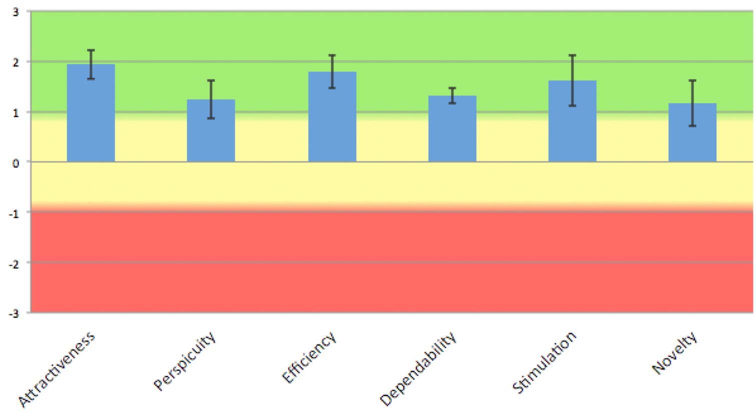


Figure 14.2: The result of different UX dimensions.

During the group discussion session, all the participants mentioned that they prefer the multiview presentation of an entity and it was the main motivating factor for them to use it again. Furthermore, they confirmed that the hierarchy is well organized and the relations between entities are shown clearly.

All the participant preferred multiview presentation.

Overall exploration and navigation of entities were flexible for them. They also agreed that the multiview presentation of entities allowed them to visualize entities as expected. Participants mentioned that the multiview presentation was not slowing them down while navigating related entities. Many participants did not use features such as details on demand, scalability of the view and history of browsed entities.

Exploration was flexible for all the participants.

The presentation did not slow them down while completing their tasks.

In the online questionnaire, we asked participants about the positive, negative and the new features they expect. As positive features, they mentioned that the UI is catchy and easy to use. The way the entities were presented was also appreciated. Navigation through network layout proved to be beneficial and more natural for most of the participants. They also mentioned that the information layout provides a complete set of knowledge from different perspectives making the system useful and practical. As negative aspects, overlapping of text in the network and timeline view created confusion and obstruction. Furthermore, participants mentioned that it was difficult to learn at the begin-

In an online questionnaire, participants stated that the User Interface was easy.

Overlapping values in various layout created some hindrance.

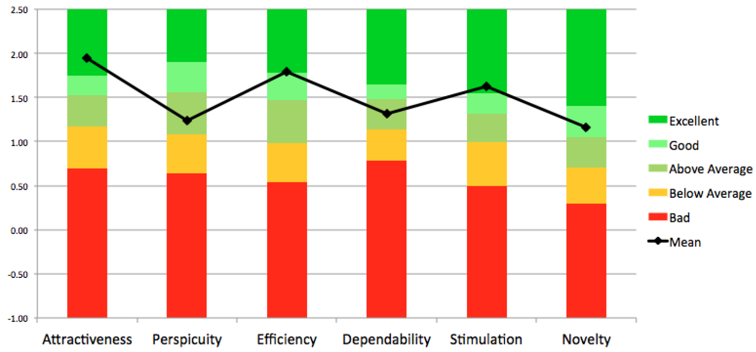


Figure 14.3: The overall UX mean of the system

ning. Some of the participants felt that the overall layout was confined making exploration difficult. One participant also expected attribute filtering on the eType lattice. People suggested some features as resizable views, full-screen mode, switch to another view on demand, pop up based labeling on network view, icons sets to represent concepts, and personalized features for selecting preferred eTypes.

Participants suggested resizable interface with more icons.

The User Experience evaluation with the participants was positive.

The UX results were overall positive and promising for all the scales: Attractiveness (1.942), Perspicuity (1.238), Efficiency (1.788), Dependability (1.313), Stimulation (1.625) and Novelty (1.163). As shown in the Figure 14.2 the overall value impression for Pragmatic (1.38) and Hedonic (0.96) features of the tool was also encouraging. Some inconsistent results were also discovered related to perspicuity and dependability. Those outliers were removed while analyzing the questionnaires.

The result verifies that the interface is engaging.

Some improvement is needed on Perspicuity and dependability of the system.

The result shows that *SemUI* is engaging and provides a satisfying result to the end users who have less knowledge about semantics. The overall mean of the system is satisfactory as shown in Figure 14.3 with a very promising result for three UX scales (Attractiveness, Efficiency, and Stimulation). However, the tool needs further improvement to increase its value on two scales (Perspicuity and Dependability). Majorly, some more work is required to reduce the learning curve of the system. The system was also able to point out issues with standard terminologies (for example the term 'disabled friendly'). This terminological issue reflects that though we can use the standard controlled vocabulary while generating an ontology, those terms cannot be directly exposed in the UI as it might create confusion to the end user. To solve this issue, we either have to use an expressive concept based on the application scenario or use the concepts along with meaningful icons.

The evaluation also pointed out that terminologies used can have a profound impact on the usability of the interface.

Part VI

Conclusion and Future work

Chapter 15

Conclusion

The increasing complexity associated with data and the introduction of open and big data offers a plethora of challenges for managing, accessing and utilizing diversified data for humans. Information visualization provides a way to investigate data, but obscure visualization schemes with a variety of interaction mechanisms and techniques will make the process more cumbersome for the end users. Furthermore, most of the application, in this case, are domain dependent and cannot cater the information need of a user from various perspectives.

Both Data management and interaction process are going to be more complicated.

The entity-centric approach offers a more meaningful, manageable and scalable way to handle diversified data as entities. The approach provides a natural classification of datasets in a more human-comprehensible way. The visualization that is aware of the diversity that exists in the data will make it easy to investigate and consume data. The results from various evaluations also seem promising.

The Entity-Centric approach will allow easy management of real-world data.

The visualization designed provides a highly interactive solution. It allows multiple simultaneous presentations of entities whose presentation is guided by the contexts. This scheme gave a fundamental notion of the contextual presentation of entities. To reach this goal, we performed multiple iterations of the design by involving the end users. Furthermore, the tool and the approach were evaluated with various users from different domains. The result ob-

We proposed interactive multiple simultaneous solutions and also involved users throughout the design.

tained was satisfactory.

15.1 Research Questions and Claims

We performed a series of iterative analysis, development, and evaluation to meet the research goals mentioned in chapter 1. The main question *How can diversified data be visualized and explored using an engaging, pleasing and contextual user interface?* was broken down into four subsequent questions. Here we reflect how various questions were handled.

R1: How can the diverse data be presented clearly in the interface?

We achieve this by using an Entity-Centric approach and multiview presentation of entities based on their eType.

In this thesis, we focus on the use of an Entity-Centric approach where data instances are redefined from a property-centric view to an entity-centric view. This transformation allows holistic representation of an entity. We visualized entities from various domains using a multiview presentation which enables visualization of various facets of an entity. The result obtained from the end users were also positive and promising. The following claim tries to answer this research question.

The evaluation result from the end users supports it.

C1: It is effective to visualize different types of entities in the same or different visualization context based on eTypes.: We see that the multiple visualizations of entities which are guided by various visualization contexts provides a flexible data consumption mechanism. For achieving this, eTypes plays a vital role. Furthermore, the evaluation of multiple use cases in section 12.1, 13.1 and 14.1 show the effectiveness of various contexts.

R2: How can the user interface provide generic exploration and presentation of diversified data?

The multi-tiered architecture and multiple view interaction provide easy navigation metaphor.

To handle this question we implemented a multi-tiered architecture in part IV. The architecture allows fetching only the users to fetch what they want from the entire knowledge base. The fetched results are then presented in a manageable graph called Network. The user study performed also support this. The claim below tries to answer this question.

C2: Seeing different properties of entities simultaneously facilitates easy exploration of related entities.: Different views provide different facets to the data giving the users the possibility to to understand the overview of the data quickly. The evaluation carried with various stakeholders in section 12.1, 13.1 and 14.1 demonstrates that multiview allows an easy and intuitive mechanism of data exploration. Similarly, the use of proper labels and icons in the manageable network view allowed the users to navigate unknown entities.

Simultaneous presentation of an entity in different context provides flexible exploration.

R3: How can the engaging and pleasing interface be designed?
End users were involved throughout the design and development of the solution. Designs were iterated many times by involving various stakeholders from multiple domains. The solution analysis done in chapter 9 motivates this idea. Furthermore, various usability and User Experience evaluation confirms our claim below.

To achieve a pleasing interface, various stakeholders were involved throughout the design and development cycle.

C3: Involving users throughout the application design and development process will result in the system with the better user experience.: Users directly interacted with the system throughout the design and development phase. They interacted with low and medium fidelity prototypes designed in chapter 9 and finally also took part in the summative evaluation of the system. The overall assessment performed in sections 12.1, 13.1 and 14.1 illustrate the effectiveness of the system designed.

Human-Computer Interaction can facilitate designing a user-centric design framework.

15.2 Lesson Learned

We handle diversified data explicitly through entities as a domain-independent(NFR1,3) and a user-centered method. From the data perspective, entities address fundamental issues of diversified data. By design, they capture context very well by encapsulating all the relevant properties (attributes) in a component. Once designed, they serve as the

The entity-centric approach is a significant data aggregator.

excellent data aggregators using eTypes. Moreover, different kinds of relations among entities make them From the usability perspective, we perceive that people intuitively think in terms of entities as objects (such as friends, events, and places) and we aim to exploit this notion in the human mental model. Our evaluation also concludes that various implementations were intuitive and supported most of the nonfunctional requirements.

In the long run, exploration of data will be a major issue.

Better management of data is the crucial issue in this data-centric community. There are many simultaneous data producers producing data in their own context. This contextual data creates diversity. An Entity-Centric solution allows better handling of the data. If we consider, Semantic Web Community, we can see that there are a plethora of design solutions to cater users need. However, various ontology and various end-user solutions will create more confusion for the users. In the long run, exploration of the data will be a significant issue.

Using end users in the design and development of any solution is still in infancy.

Though using end user in the development of any solution is still frowned upon in the research community. We feel, involving the end users is crucial. Developers of the tool might not have a clear idea for which purpose the tool will be used. Involving various users will help them generate requirements that solve their need.

Engaging and motivating end users is also an issue.

Contrary to the above statement, it is very complicated to manage end users. Making them feel that the tool is for them and explaining the requirements can be a daunting task. Another aspect can be the language barrier as users come from various domains and contexts and countries. A generic solution requires catering all their need and understand them.

Chapter 16

Future work

Performing interdisciplinary research during the Ph.D. helped us gain insight into both semantic technology and Human-Computer Interaction. We have seen several possibilities in continuing the research. Some of the areas are presented below:

There are possibilities to extend this research.

More Evaluation

We plan to evaluate the tool with more users in the future. Many use cases still need to be accommodated in an Entity-Centric system. Due to the time limitation, those domains were skipped for now. There are two use cases, ready to be evaluated. They are Cultural Heritage (AAC Project), Figure D.1 and on health care domain (SHIB), Figure D.3 . Currently, we only managed to conduct a heuristic evaluation of these tools. We have planned to perform a UX evaluation of the implemented tool. We will evaluate six different User Experience dimensions. They are attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. We plan to assess museum project with various museum workers working under AAC. The possible stakeholders in the AAC scenario are museum curators, administrators, and visitors. Similarly, we plan to evaluate SHIB with various health workers working under NHS[1]. The possible stakeholders in this scenario are data scientists, doctors, and nurses.

We need to include more domains and evaluate them.

[1]http://www.scot.nhs.uk/

Entigets

Entigets are flexible User Interface widgets

The existing architecture allows defining a generic context. However, the designed interface is not modular. We plan to develop the interface in a more modular way where various visual components can be dynamically aggregated. Entigets are individual visual components which are defined based on a context. This dynamic aggregation allows customization of the multiview interface from end users' perspective. If they feel a presentation is not useful, then they can hide the presentation and focus on another specific presentation. For example, if map presentation is irrelevant, then the user can hide it from the interface and then focus on another presentation like a timeline, network or list. The Entigets will make this possible. Right now, each visual context is present in a single application. Based on the eType the Entigets will generate a specific visualization scheme. For example, if the user selects an eType person, the interface will pull and draw four different Entigets as List, Map, Network, and Timeline. Entigets will be incorporated inside KOS. Furthermore, other visual components such as charts can also be realized.

Charts

Accommodate charts on the interface.

Charts bring new conclusive perspective to the data. In this work, we haven't focused on using charts as a visualization metaphor, however, in the future, we want to explore the possibility of visualizing charts and their implication on other views and the users.

User Exploration Framework

Understand the pattern and how Users explore the entities.

Users follow various pattern or ways to explore the information. They constantly keep on changing the track throughout the process. The pattern in which they explore and the search can be presented to them so they can learn their pattern and readjust if there are issues.

Mobile Friendly

Develop a native mobile app.

The interface currently developed can easily be used from various mobile devices. However, some interactions are not well suited. In the future, we can port it to various smaller devices as a native application.

Appendix A

Early prototypes

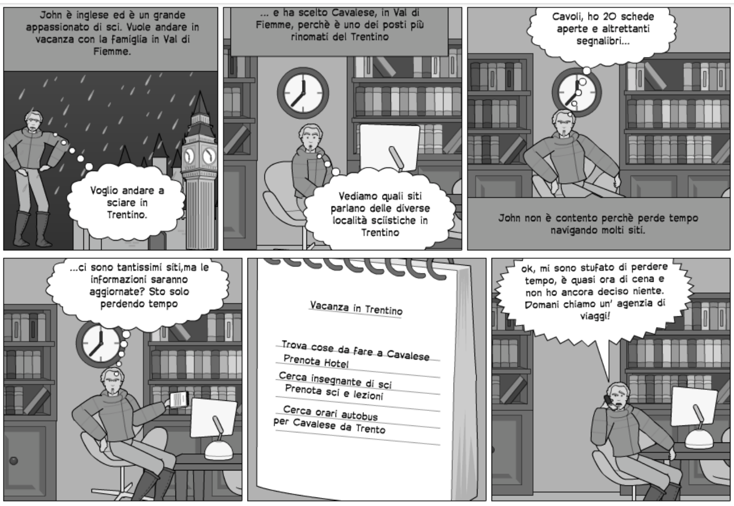


Figure A.1: Trentino Entitypedia existing scenario in Italian.

Figure A.2: Trentino Entitypedia in the new scenario in Italian.

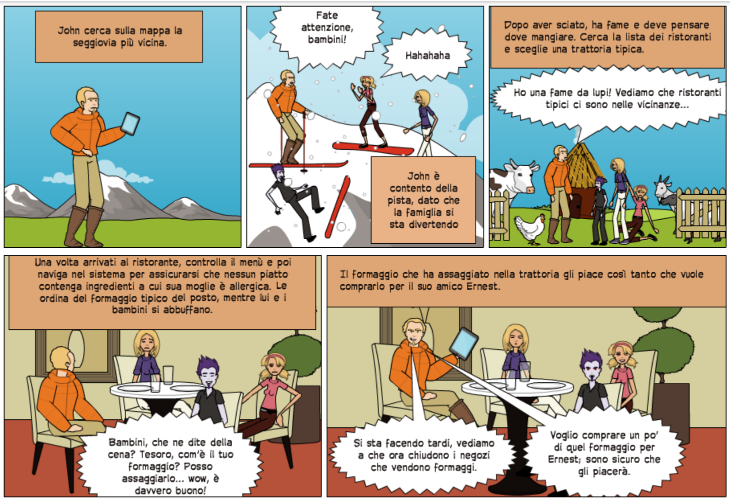


Figure A.3: Trentino Entitypedia in the new scenario in Italian.

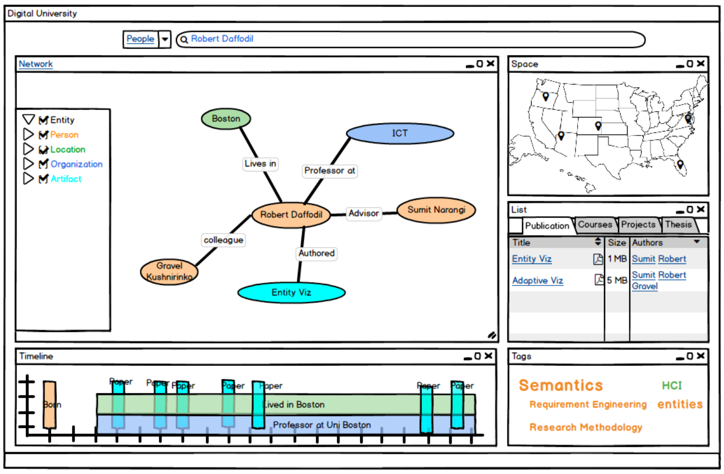


Figure A.4: Digital University Low fidelity mockup .

Figure A.5: Persona SHIB.

Appendix B

Consent

Contextual visualization of entities Consent form

I ________________________________ hereby freely and voluntarily consent to participate in the experiment conducted by researchers from the University of Trento with the purpose of understanding the effectiveness of context-based visualization.

Before beginning the session the aim and content were explained to me explicitly.
I acknowledge I have the right to question any part of the procedure and can withdraw at any time without this being held against me.
For research purposes, a video recording of the session will be made.
As long as video recording is concerned

- I allow
- I don't allow

The accompanying researcher will record my voice.
The audio and other information collected during the test will be analyzed by the researcher and will be published and presented to an audience. If this will happen, my right to privacy will be retained, i.e. personal details will not be revealed and it will not be possible to retrieve any data, which might disclose my identity. As long as analyzing and publishing is concerned.

- I allow
- I don't allow

The research team to use the data gathered during the session
I have read and understand the before:

____
Participant's Signature and Date

____
Researcher's Name, signature and Date

Figure B.1: Consent for Trentino Entitypedia.

University of Trento, KnowDive,

UNIVERSITY
OF TRENTO - Italy

Consent Form

Title of Project: DIGITAL UNIVERSITY

Name of Researcher: Sajan Raj Ojha

Our universities' official portal (www.unitn.it) is being phased out. It is currently being upgraded to a new portal called ***Digital University (DU)***. This user study aims to understand how different users (academic staffs, technicians etc.) are using the existing system to search for people and different departments within the university. We aim to understand the problems that a user faces using the current portal. Furthermore, we are asking suggestions and new User Interface functionalities that will help users to perform their search effectively and also improve the usability of the new DU system.

1. I confirm that I have read and understand the Plain Language Statement for the above study and have had the opportunity to ask questions.
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

3.
- *consent to interviews being audio-taped ,*
- *consent to screen activities being video-taped,*
- *participants to be referred to by pseudonym or identified by name in any publications arising from the research,*
- *and in instances where a dependent relationship is involved, confirmation that participation or non-participation in the research will have no effect on grades/assessment/employment)]*

4. Consent to be audio taped [ ] Yes [ ] No
5. Consent to video tape the tasks [ ] Yes [ ] No

4. I agree / do not agree (delete as applicable) to take part in the above study.

Name of Participant ______ Date ______ Signature ______

Researcher ______ Date ______ Signature ______

Figure B.2: Consent for Digital University.

University of Trento, KnowDive,

UNIVERSITY
OF TRENTO - Italy

Consent Form

Title of Project: Entity Aware Multiview data visualization

Name of Researcher: Sajan Raj Ojha

Semantic technology mainly offers restricted visualization of the content. To overcome this, we propose an entity aware approach which presents the semantic content by separating their properties as attributes and relations and presenting them simultaneously using a tool called ***SemUI***. This user study aims to understand the user's perception towards simultaneous presentation and exploration of related entities. Furthermore, we are asking suggestions and functionalities that will help users to perform their tasks effectively and also improve the usability of the prototype.

1. I confirm that I have read and understand the Plain Language Statement for the above study and have had the opportunity to ask questions.
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

3.
- *consent to interviews being audio-taped ,*
- *consent to activities being video-taped,*
- *participants to be referred to by pseudonym or identified by name in any publications arising from the research,*
- *and in instances where a dependent relationship is involved, confirmation that participation or non-participation in the research will have no effect on grades/assessment/employment)]*

4. Consent to be audio taped	[ ] Yes [ ] No
5. Consent to video tape the tasks	[ ] Yes [ ] No
6. Consent to not to share/use any information related to SemUI tool	[ ] Yes [ ] No

7. I agree / do not agree (delete as applicable) to take part in the above study.

Name of Participant	Date	Signature
Sajan Raj Ojha	05/October/2016	
Researcher	Date	Signature

Figure B.3: Consent for SemUI.

Appendix C

Questionnaires

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

Figure C.1: Dimensions for User Experience evaluation questionnaire.

Age:
Gender:
Country where you have spent most of your life
Highest degree of education
Have you noticed any other systems also having the similar layout and functionalities.

Hypothesis 1: Seeing same entity using different visual layout in a same view is meaningful

- What are the advantages/benefits of visualising an entity using different paradigms in a UI (as in Map, Timeline and Table)?
  - does it facilitate finding an entity and seeing/understanding attributes from different perspective
- What are the disadvantages/drawbacks you find with such presentation?
  - Does it hinder your objective to find an entity and understand them? Please elaborate
  - What are the suggestions you propose to solve the problems?

Hypothesis 2: Seeing different entities in a network layout facilitates easy exploration of related entities.

- What are the advantages/benefits of visualizing related entities in one network view?
  - Does it facilitate finding related entity and understanding the associated relation?
- What are the disadvantages/drawbacks you find with such view?
  - Does it hinder your objective to navigate related entities?
  - What alternative suggestions do you propose for the problems?

What type of scenarios you can imagine where this system can be used? Can you describe?

What would be added value compared to other, similar UIs

Figure C.2: System evaluation questionnaire Trentino Entitypedia 2.

Aesthetic and Visual experience

Seeing different types of entities on the same page is visually present*

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neither agree or disagree
- ◯ Agree
- ◯ Strongly agree

Icons clearly reflects the purpose of the specific entity on the map*

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neither agree or disagree
- ◯ Agree
- ◯ Strongly agree

Figure C.3: System evaluation questionnaire: Aesthetics.

Aspects and features of the application

Please list all the different aspects of the application based on your experience.

Positive Aspects *

Long-answer text

Negative Aspects *

Long-answer text

New features you like to see *

Long-answer text

Figure C.4: System evaluation questionnaire: UI aspects.

Emotions

I like the idea that, using the system, I can discover new information connected to what I was looking for in the first place.*

- Strongly disagree
- Disagree
- Neither agree or disagree
- Agree
- Strongly agree

The fact that I can propagate through the network of connected entities makes this application somehow play full.*

- Strongly disagree
- Disagree
- Neither agree or disagree
- Agree
- Strongly agree

Figure C.5: System evaluation questionnaire UI: emotions.

Please choose features of this application that reflect your habits in using similar systems.*

- ☐ Relations/ links between entities on the map
- ☐ Entities's details on demand
- ☐ Entities's view scalability
- ☐ Entities's view filtering
- ☐ History of browsed entities
- ☐ Flexible exploration and navigation through entities
- ☐ Flexibility/contextual switch in showing entities from different domains
- ☐ Other:

Please choose the specific features that encourage you to use the application*

- ☐ Relations/ links between entities on the map
- ☐ Entities's details on demand
- ☐ Entities's view scalability
- ☐ Entities's view filtering
- ☐ History of browsed entities
- ☐ Flexible exploration and navigation through entities
- ☐ Flexibility/contextual switch in showing entities from different domains
- ☐ Other:

Figure C.6: System evaluation questionnaire UI: User's own expereince.

Identification

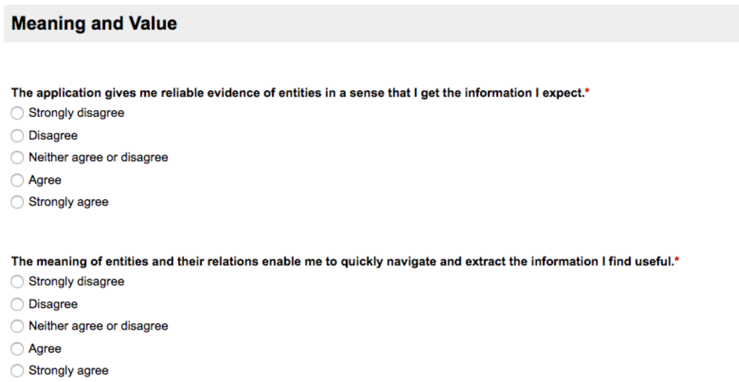
I see myself using the application*

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neither agree or disagree
- ◯ Agree
- ◯ Strongly agree

Please choose features of this application that reflect your habits in using similar systems*

- ☐ Relations/ links between entities on the map
- ☐ Entities's details on demand
- ☐ Entities's view scalability
- ☐ Entities's view filtering
- ☐ History of browsed entities
- ☐ Facet-based exploration and navigation through entities
- ☐ Flexibility-Contextual swithc in showing entities from different domains
- ☐ Other:

Figure C.7: System evaluation questionnaire UI: identification.



Meaning and Value

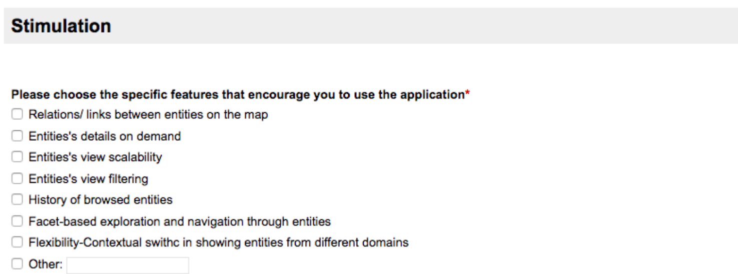
The application gives me reliable evidence of entities in a sense that I get the information I expect.*

- Strongly disagree
- Disagree
- Neither agree or disagree
- Agree
- Strongly agree

The meaning of entities and their relations enable me to quickly navigate and extract the information I find useful.*

- Strongly disagree
- Disagree
- Neither agree or disagree
- Agree
- Strongly agree

Figure C.8: System evaluation questionnaire UI: Meaning and Value.

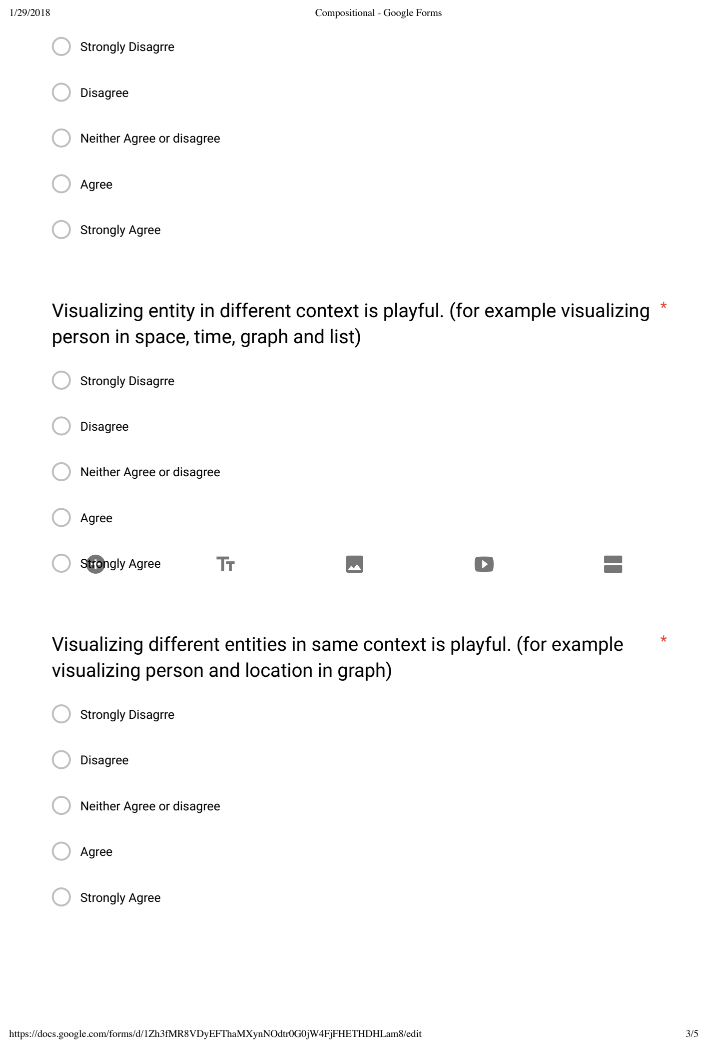


Stimulation

Please choose the specific features that encourage you to use the application*

- Relations/ links between entities on the map
- Entities's details on demand
- Entities's view scalability
- Entities's view filtering
- History of browsed entities
- Facet-based exploration and navigation through entities
- Flexibility-Contextual swithc in showing entities from different domains
- Other:

Figure C.9: System evaluation questionnaire UI: stimulation.


1/29/2018 Compositional - Google Forms

Strongly Disagrre

Disagree

Neither Agree or disagree

Agree

Strongly Agree

Visualizing entity in different context is playful. (for example visualizing person in space, time, graph and list) *

Strongly Disagrre

Disagree

Neither Agree or disagree

Agree

Strongly Agree

Visualizing different entities in same context is playful. (for example visualizing person and location in graph) *

Strongly Disagrre

Disagree

Neither Agree or disagree

Agree

Strongly Agree

https://docs.google.com/forms/d/1Zh3fMR8VDyEFThaMXynNOdtr0G0jW4FjFHETHDHLam8/edit 3/5

Figure C.10: System evaluation questionnaire 1.

1/29/2018 Compositional - Google Forms

- Strongly Disagrre
- Disagree
- Neither Agree or disagree
- Agree
- Strongly Agree

Getting detailed view about an entity was helpful *

- Strongly Disagrre
- Disagree
- Neither Agree or disagree
- Agree
- Strongly Agree

List the positive aspect(s) of the contextual visualization

Long answer text

List the negative aspect(s) of the contextual visualization

Long answer text

List the features you wanted to see

Long answer text

https://docs.google.com/forms/d/1Zh3fMR8VDyEFThaMXynNOdtr0G0jW4FjFHETHDHLam8/edit 4/5

Figure C.11: System evaluation questionnaire 2.

1/29/2018 Compositional - Google Forms

After section 1 **Continue to next section**

Section 2 of 2

Contextual visualization

Description (optional)

It is easy to visualize different type of entities grouped in different categories (categories like person, location) *

- ○ Strongly Disagrre
- ○ Disagree
- ○ Neither Agree or disagree
- ○ Agree
- ○ Strongly Agree

Discovering new entities to what I was looking for in the first place is *

- ○ Strongly Disagrre
- ○ Disagree
- ○ Neither Agree or disagree
- ○ Agree
- ○ Strongly Agree

https://docs.google.com/forms/d/1Zh3fMR8VDyEFThaMXynNOdtr0G0jW4FjFHETHDHLam8/edit 2/5

Figure C.12: System evaluation questionnaire 3.

Appendix D

SemUI: Implementations

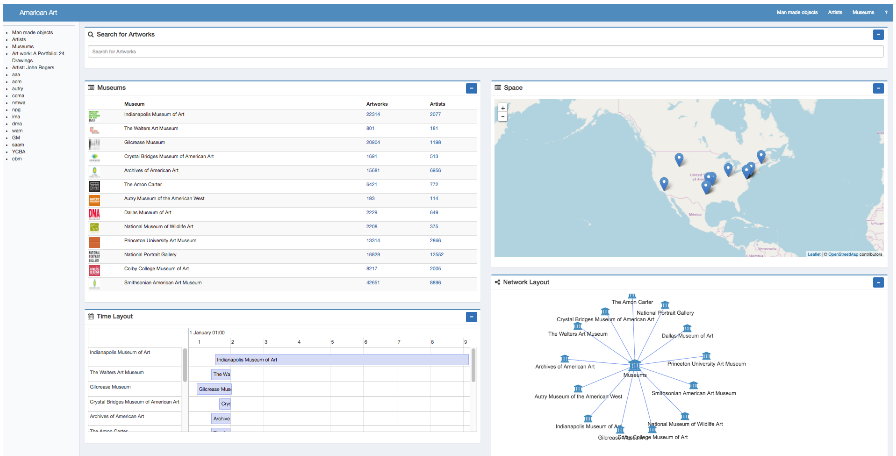


Figure D.1: *SemUI*: in a museum setting.

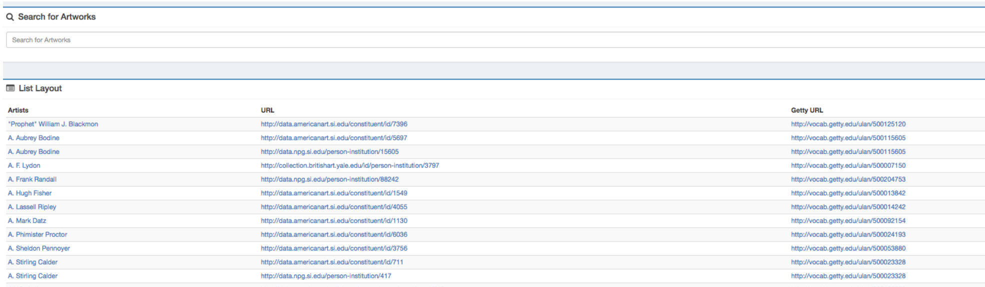


Figure D.2: *SemUI*: List view.

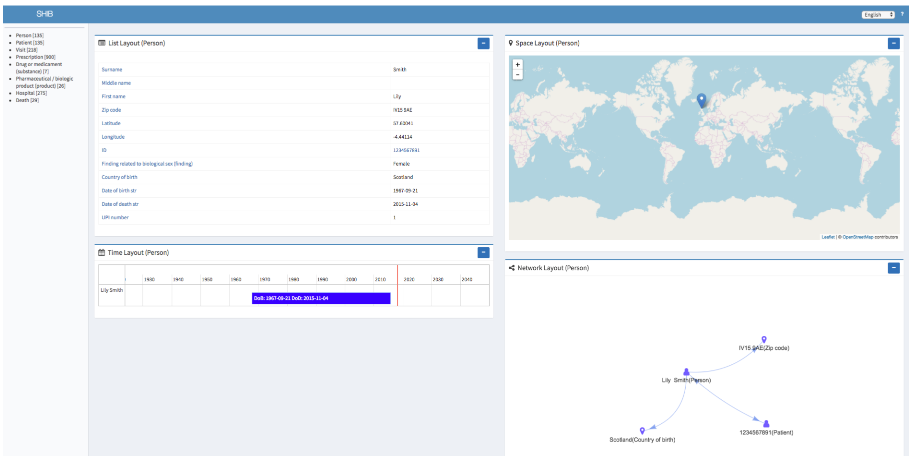


Figure D.3: SHIB implementation.

Bibliography

Chadia Abras, Diane Maloney-Krichmar, and Jenny Preece. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456, 2004.

Russell L Ackoff. From data to wisdom. *Journal of applied systems analysis*, 16(1):3–9, 1989.

Maneesh Agrawala, Wilmot Li, and Floraine Berthouzoz. Design principles for visual communication. *Communications of the ACM*, 54(4):60–69, 2011.

Lauralee Alben. Defining the criteria for effective interaction design. *Interactions*, 3(3):11–15, 1996.

Robert A. Amar and John T. Stasko. Knowledge precepts for design and evaluation of information visualizations. *IEEE Trans. Vis. Comput. Graph.*, 11(4):432–442, 2005. URL `http://dblp.uni-trier.de/db/journals/tvcg/tvcg11.html#AmarS05`.

John R Anderson. *The architecture of cognition*. Psychology Press, 2013.

Nikolaos Athanasis, Vassilis Christophides, and Dimitris Kotzinos. Generating on the fly queries for the semantic web: The ics-forth graphical rql interface (grql). In *International Semantic Web Conference*, volume 3298, pages 486–501. Springer, 2004.

Lyn Bartram. Can motion increase user interface bandwidth in complex systems? In *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on*, volume 2, pages 1686–1692. IEEE, 1997.

Marcia J Bates. Toward an integrated model of information seeking and searching. *The New Review of Information Behaviour Research*, 3:1–15, 2002.

Christian Becker and Christian Bizer. Dbpedia mobile-a location-aware semantic web client. *Proceedings of the Semantic Web Challenge*, pages 13–16, 2008.

Christian Becker and Christian Bizer. Exploring the geospatial semantic web with dbpedia mobile. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(4): 278–286, 2009.

Ivan Bedini, Feroz Farazi, David Leoni, Juan Pane, Ivan Tankoyeu, and Stefano Leucci. Open government data: Fostering innovation. *JeDEM-eJournal of eDemocracy and Open Government*, 6(1):69–79, 2014.

Tim Benson. Principles of health interoperability hl7 and snomed; 2009.

Luisa Bentivogli, Pamela Forner, Bernardo Magnini, and Emanuele Pianta. Revising the wordnet domains hierarchy: semantics, coverage and balancing. In *Proceedings of the Workshop on Multilingual Linguistic Ressources*, pages 101–108. Association for Computational Linguistics, 2004.

Tim Berners-Lee. Linked data-design issues (2006). *URL http://www. w3. org/DesignIssues/LinkedData. html*, 2011.

Tim Berners-Lee, James Hendler, Ora Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.

Tim Berners-Lee, Yuhsin Chen, Lydia Chilton, Dan Connolly, Ruth Dhanaraj, James Hollenbach, Adam Lerer, and David Sheets. Tabulator: Exploring and analyzing linked data on the semantic web. In *Proceedings of the 3rd international semantic web user interaction workshop*, volume 2006, page 159. Athens, Georgia, 2006.

Tim Berners-Lee, James Hollenbach, Kanghao Lu, Joe Presbrey, and Mc Schraefel. Tabulator redux: Browsing and writing linked data. 2008.

Hugh Beyer and Karen Holtzblatt. Contextual design. *interactions*, 6(1):32–42, 1999.

Ravish Bhagdev, Ajay Chakravarthy, Sam Chapman, Fabio Ciravegna, and Vita Lanfranchi. Creating and using organisational semantic webs in large networked organisations. *The Semantic Web-ISWC 2008*, pages 723–736, 2008.

Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pages 205–227, 2009a.

Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009b. URL `http://eprints.soton.ac.uk/271285/`.

Ann Blandford, Thomas RG Green, Dominic Furniss, and Stephann Makri. Evaluating system utility and conceptual fit using cassm. *International Journal of Human-Computer Studies*, 66(6):393–409, 2008.

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

Nan Cao, Jimeng Sun, Yu-Ru Lin, David Gotz, Shixia Liu, and Huamin Qu. Facetatlas: Multifaceted visualization for rich text corpora. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1172–1181, 2010.

Stuart K Card and Jock Mackinlay. The structure of the information visualization design space. In *Information Visualization, 1997. Proceedings., IEEE Symposium on*, pages 92–99. IEEE, 1997.

Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999. ISBN 1-55860-533-9.

Tiziana Catarci, Paolo Dongilli, Tania Di Mascio, Enrico Franconi, Giuseppe Santucci, and Sergio Tessaris. An ontology based visual tool for query formulation support. In *Proceedings of the 16th European Conference on Artificial Intelligence*, pages 308–312. IOS Press, 2004.

Gong Cheng, Weiyi Ge, and Yuzhong Qu. Falcons: searching and browsing entities on the semantic web. In *Proceedings of the 17th international conference on World Wide Web*, pages 1101–1102. ACM, 2008.

Chun Wei Choo, Brian Detlor, and Don Turnbull. *Web work: Information seeking and knowledge work on the World Wide Web*, volume 1. Springer Science & Business Media, 2013.

Federal Geographic Data Committee. Entity and attribute information, 2015. URL `https://www.fgdc.gov/metadata/csdgm-new/05.html`.

Bill Curtis, Herb Krasner, and Neil Iscoe. A field study of the software design process for large systems. *Communications of the ACM*, 31(11):1268–1287, 1988.

Aba-Sah Dadzie and Matthew Rowe. Approaches to visualising linked data: A survey. *Semant. web*, 2(2): 89–124, April 2011a. ISSN 1570-0844. doi: 10.3233/SW-2011-0037. URL `http://dx.doi.org/10.3233/SW-2011-0037`.

Aba-Sah Dadzie and Matthew Rowe. Approaches to visualising linked data: A survey. *Semantic Web*, 2(2):89–124, 2011b.

Antonio R Damasio. The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, 1(1):123–132, 1989.

Danica Damljanovic, Milan Agatonovic, and Hamish Cunningham. Freya: An interactive way of querying linked data using natural language. In *ESWC Workshops*, volume 7117, pages 125–138. Springer, 2011.

Subhashis Das and Fausto Giunchiglia. Geoetypes: Harmonizing diversity in geospatial data (short paper). In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 643–653. Springer, 2016.

Thomas H Davenport and Laurence Prusak. *Information ecology: Mastering the information and knowledge environment*. Oxford University Press on Demand, 1997.

Thomas H Davenport and Laurence Prusak. *Working knowledge: How organizations manage what they know*. Harvard Business Press, 1998.

Tim Davies. Supporting open data use through active engagement. In *Position paper: W3C Using Open Data Workshop*. W3C, 2012. doi: http://www.w3.org/2012/06/pmod/pmod2012_submission_5.pdf.

Pieter MA Desmet and Paul Hekkert. Framework of product experience. *International Journal of Design*, 1(1):57–66, 2007.

A Donald and Stephen W Norman Draper. User centered system design: New perspectives on human-computer interaction. 1986.

Khadija Elbedweihy, Stuart N Wrigley, Fabio Ciravegna, Dorothee Reinhard, and Abraham Bernstein. Evaluating semantic search systems to identify future directions of research. In *Extended Semantic Web Conference*, pages 148–162. Springer, 2012.

K Anders Ericsson and Herbert A Simon. Verbal reports as data. *Psychological review*, 87(3):215, 1980.

Mohammad Shahjahan Feroz Farazi. *Faceted Lightweight Ontologies: A Formalization and some Experiments*. PhD thesis, University of Trento, 2010.

Jean-Daniel Fekete, Jarke J Van Wijk, John T Stasko, and Chris North. The value of information visualization. In *Information visualization*, pages 1–18. Springer, 2008.

Susan Feldman. *The high cost of not finding information*. Information Today, Incorporated, 2004.

E Ferro and M Osella. Eight business model archetypes for psi re-use ("open data on the web" workshop, april 23rd-24th). google campus, shoreditch, london, 2013.

Roberto García, Juan Manuel Gimeno, Ferran Perdrix, Rosa Gil, and Marta Oliva. The rhizomer semantic content management system. In *Emerging Technologies and Information Systems for the Knowledge Society*, pages 385–394. Springer, 2008.

Bill Gaver and Heather Martin. Alternatives: exploring information appliances through conceptual design proposals. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 209–216. ACM, 2000.

Dedre Gentner and Albert L Stevens. *Mental models*. Psychology Press, 2014.

F Giunchiglia. Managing diversity in knowledge, invited talk at the european conference on artificial intelligence ecai. *Lecture Notes in Artificial Intelligence*, 2006.

Fausto Giunchiglia. Computational Humanism,. http://disi.unitn.it/ fausto/CH.pdf, 2016.

Fausto Giunchiglia and Mattia Fumagalli. Concepts as (recognition) abilities1. In *Formal Ontology in Information Systems: Proceedings of the 9th International Conference (FOIS 2016)*, volume 283, page 153. IOS Press, 2016.

Fausto Giunchiglia, Biswanath Dutta, Vincenzo Maltese, and Feroz Farazi. A facet-based methodology for the construction of a large-scale geospatial ontology. *Journal on Data Semantics*, 1(1):57–73, 2012a.

Fausto Giunchiglia, Vincenzo Maltese, and Biswanath Dutta. Domains and context: first steps towards managing diversity in knowledge. *Web Semantics: Science, Services and Agents on the World Wide Web*, 12:53–63, 2012b.

Hugh Glaser, Harith Alani, Les Carr, Sam Chapman, Fabio Ciravegna, Alexiei Dingli, Nicholas Gibbins, Stephen Harris, m schraefel, and Nigel Shadbolt. Cs aktivespace: Building a semantic web application. *The semantic web: Research and applications*, pages 417–432, 2004.

John D Gould and Clayton Lewis. Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3):300–311, 1985.

Paul J Griffiths. *An apology for apologetics: A study in the logic of interreligious dialogue*. Wipf and Stock Publishers, 2007.

Nicola Guarino. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*, volume 46. IOS press, 1998.

W. Halb, Y. Raimond, and M. Hausenblas. Building Linked Data For Both Humans and Machines. In *WWW 2008 Workshop: Linked Data on the Web (LDOW2008)*, Beijing, China, 2008a.

Wolfgang Halb, Yves Raimond, and Michael Hausenblas. Building linked data for both humans and machines. In *LDOW*, 2008b.

Edited Richard Harper, Tom Rodden, Yvonne Rogers, Abigail Sellen, Being Human, et al. Human-computer interaction in the year 2020. 2008.

Marc Hassenzahl. The thing and i: understanding the relationship between user and product. In *Funology*, pages 31–42. Springer, 2005.

Marc Hassenzahl and Noam Tractinsky. User experience-a research agenda. *Behaviour & Information Technology*, 25 (2):91–97, 2006.

Tuukka Hastrup, Richard Cyganiak, and Uldis Bojars. Browsing linked data with fenfire. 2008.

Marti Hearst. User interfaces and visualization. *Modern information retrieval*, pages 257–323, 1999.

Marti Hearst. Design recommendations for hierarchical faceted search interfaces. In *ACM SIGIR workshop on faceted search*, pages 1–5. Seattle, WA, 2006.

Tom Heath. How will we interact with the web of data? *Internet Computing, IEEE*, 12(5):88–91, 2008.

Gergely Héja, György Surján, and Péter Varga. Ontological analysis of snomed ct. *BMC medical informatics and decision making*, 8(1):S8, 2008.

Paul Hekkert. Design aesthetics: principles of pleasure in design. *Psychology science*, 48(2):157, 2006.

Ivan Herman, Guy Melançon, and M Scott Marshall. Graph visualization and navigation in information visualization: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 6(1):24–43, 2000.

Michiel Hildebrand, Jacco van Ossenbruggen, and Lynda Hardman. /facet: A browser for heterogeneous semantic web repositories. *The Semantic Web-ISWC 2006*, pages 272–285, 2006.

Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres, and Stefan Decker. Searching and browsing linked data with swse: The semantic web search engine. *Web semantics: science, services and agents on the world wide web*, 9(4):365–401, 2011.

Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Samppa Saarela, Miikka Junnila, and Suvi Kettula. Museumfinland—finnish museums on the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(2):224–241, 2005.

Vesa Jääskö and Tuuli Mattelmäki. Observing and probing. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, pages 126–131. ACM, 2003.

Evangelos Karapanos, John Zimmerman, Jodi Forlizzi, and Jean-Bernard Martens. Measuring the dynamics of remembered experience over time. *Interacting with Computers*, 22(5):328–335, 2010.

Akrivi Katifori, Constantin Halatsis, George Lepouras, Costas Vassilakis, and Eugenia Giannopoulou. Ontology visualization methods—a survey. *ACM Computing Surveys (CSUR)*, 39(4):10, 2007.

Esther Kaufmann and Abraham Bernstein. Evaluating the usability of natural language query languages and interfaces to semantic web knowledge bases. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4): 377–393, 2010.

Esther Kaufmann, Abraham Bernstein, and Lorenz Fischer. Nlp-reduce: A naive but domainindependent natural language interface for querying ontologies. In *4th European Semantic Web Conference ESWC*, pages 1–2, 2007.

Daniel A Keim. Visual techniques for exploring databases. In *Knowledge Discovery in Databases (KDD'97)*, 1997.

Daniel A Keim, Christian Panse, and Mike Sips. Information visualization: Scope, techniques and opportunities for geovisualization, 2004.

Daniel A Keim, Florian Mansmann, Jörn Schneidewind, and Hartmut Ziegler. Challenges in visual data analysis. In *Information Visualization, 2006. IV 2006. Tenth International Conference on*, pages 9–16. IEEE, 2006.

Ali Khalili, Antonis Loizou, and Frank van Harmelen. Adaptive linked data-driven web components: Building flexible and reusable semantic web interfaces. In *International Semantic Web Conference*, pages 677–692. Springer, 2016.

Gary Klein, Brian Moon, and Robert R Hoffman. Making sense of sensemaking 1: Alternative perspectives. *IEEE intelligent systems*, 21(4):70–73, 2006.

Craig A Knoblock, Pedro Szekely, Eleanor Fink, Duane Degler, David Newbury, Robert Sanderson, Kate Blanch, Sara Snyder, Nilay Chheda, Nimesh Jain, et al. Lessons learned in building linked data for the american art collaborative. In *International Semantic Web Conference*, pages 263–279. Springer, 2017.

James Krikelas. Information-seeking behavior: Patterns and concepts. *Drexel library quarterly*, 19(2):5–20, 1983.

Monika Lanzenberger, Jennifer Sampson, and Markus Rester. Visualization in ontology tools. In *Complex, Intelligent and Software Intensive Systems, 2009. CISIS'09. International Conference on*, pages 705–711. IEEE, 2009.

Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*, pages 63–76. Springer, 2008.

Talia Lavie and Noam Tractinsky. Assessing dimensions of perceived visual aesthetics of web sites. *International journal of human-computer studies*, 60(3):269–298, 2004.

Yuangui Lei, Victoria Uren, Enrico Motta, et al. Semsearch: A search engine for the semantic web. In *EKAW*, volume 4248, pages 238–245. Springer, 2006.

Vanessa Lopez, Enrico Motta, and Victoria Uren. Poweraqua: Fishing the semantic web. *The Semantic Web: research and applications*, pages 393–410, 2006.

Farzaneh Mahdisoltani, Joanna Biega, and Fabian Suchanek. Yago3: A knowledge base from multilingual wikipedias. In *7th Biennial Conference on Innovative Data Systems Research*. CIDR Conference, 2014.

Sascha Mahlke and Manfred Thüring. Studying antecedents of emotional experiences in interactive contexts. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 915–918. ACM, 2007.

Vincenzo Maltese and Fausto Giunchiglia. Foundations of digital universities. *Cataloging & Classification Quarterly*, 55(1):26–50, 2017.

James Manyika, Michael Chui, Peter Groves, Diana Farrell, Steve Van Kuiken, and Elizabeth Almasi Doshi. Open data: Unlocking innovation and performance with liquid information. *McKinsey Global Institute*, page 21, 2013.

H Marc. Aesthetics in interactive products: Correlates and consequences of beauty, 2007.

Gary Marchionini. *Information seeking in electronic environments*. Number 9. Cambridge university press, 1997.

Gary Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

Suvodeep Mazumdar. *Visualising Large Semantic Datasets*. PhD thesis, University of Sheffield, 2013.

Suvodeep Mazumdar. *Visualising Large Semantic Datasets: A Generic, Scalable and Aesthetic Solution*. PhD thesis, University of Sheffield, Department of Computer Science, 2014.

Suvodeep Mazumdar, Daniela Petrelli, and Fabio Ciravegna. Exploring user and system requirements of linked data visualization through a visual dashboard approach. *Semantic Web*, 5(3):203–220, 2014.

Deborah L McGuinness, Frank Van Harmelen, et al. Owl web ontology language overview. *W3C recommendation*, 10(10):2004, 2004.

George Miller and Christiane Fellbaum. Wordnet: An electronic lexical database, 1998.

George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

Ruth Garrett Millikan. *On clear and confused ideas: An essay about substance concepts*. Cambridge University Press, 2000.

Roberto Mirizzi, Azzurra Ragone, Tommaso Di Noia, and Eugenio Di Sciascio. Semantic wonder cloud: exploratory search in dbpedia. *Current Trends in Web Engineering*, pages 138–149, 2010.

Riichiro Mizoguchi and Mitsuru Ikeda. Towards ontology engineering. *Journal-Japanese Society for Artificial Intelligence*, 13:9–10, 1998.

Kawa Nazemi. *Adaptive semantics visualization*, volume 646. Springer, 2016.

Jakob Nielsen. *Usability engineering*. Elsevier, 1994.

Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256. ACM, 1990.

Ikujiro Nonaka. *The knowledge-creating company*. Harvard Business Review Press, 2008.

Benoît Otjacques, Mickaël Stefas, Maël Cornil, and Fernand Feltz. Open data visualization keeping traces of the exploration process. In *Proceedings of the First International Workshop on Open Data*, pages 53–60. ACM, 2012.

Stephen E Palmer. *Vision science: Photons to phenomenology*. MIT press, 1999.

Daniela Petrelli, Suvodeep Mazumdar, Aba-Sah Dadzie, and Fabio Ciravegna. Multi visualization and dynamic query for effective exploration of semantic data. *The Semantic Web-ISWC 2009*, pages 505–520, 2009.

Edward J Quigley and Anthony Debons. Interrogative theory of information and knowledge. In *Proceedings of the 1999 ACM SIGCPR conference on Computer personnel research*, pages 4–10. ACM, 1999.

Christian Reich, Patrick Ryan, Rimma Belenkaya, Karthik Natarajan, and Clair Blacketer. Omop common data model v5.1 specifications, 2017. URL `https://github.com/OHDSI/CommonDataModel/wiki`.

Mark A Robinson. An empirical analysis of engineers' information behaviors. *Journal of the Association for Information Science and Technology*, 61(4):640–658, 2010.

Yvonne Rogers, Helen Sharp, and Jenny Preece. *Interaction design: beyond human-computer interaction*. John Wiley & Sons, 2011.

Stephanie Rosenbaum, Janice Anne Rohn, and Judee Humburg. A toolkit for strategic usability: results from workshops, panels, and surveys. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 337–344. ACM, 2000.

Jeffrey Rubin and Dana Chisnell. *Handbook of usability testing: howto plan, design, and conduct effective tests*. John Wiley & Sons, 2008.

Guus Schreiber. *Knowledge engineering and management: the CommonKADS methodology*. MIT press, 2000.

Kennon M Sheldon, Andrew J Elliot, Youngmee Kim, and Tim Kasser. What is satisfying about satisfying events? testing 10 candidate psychological needs. *Journal of personality and social psychology*, 80(2):325, 2001.

Ben Shneiderman. *Designing the user interface: strategies for effective human-computer interaction*, volume 2. Addison-Wesley Reading, MA, 1992.

Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

Martin G Skjæveland. Sgvizler: A javascript wrapper for easy visualization of sparql result sets. In *Extended Semantic Web Conference*, 2012.

Socrata. Open Government Data Benchmark Study Report. Technical report, 2010.

Robert Spence. *Information visualization*, volume 1. Springer, 2001.

Magnus Stuhr, Dumitru Roman, and David Norheim. Lodwheel-javascript-based visualization of rdf data. In *Proceedings of the Second International Conference on Consuming Linked Data-Volume 782*, pages 73–84. CEUR-WS. org, 2011.

Catherine Tallon-Baudry and Olivier Bertrand. Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences*, 3(4):151–162, 1999.

Giovanni Tummarello, Renaud Delbru, and Eyal Oren. Sindice. com: Weaving the open linked data. In *The Semantic Web*, pages 552–565. Springer, 2007.

Giovanni Tummarello, Richard Cyganiak, Michele Catasta, Szymon Danielczyk, Renaud Delbru, and Stefan Decker. Sig. ma: Live views on the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4): 355–364, 2010.

Innovation Union. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*. Brussels, 2014.

Michael Uschold and Michael Gruninger. Ontologies and semantics for seamless connectivity. *ACM SIGMod Record*, 33(4):58–64, 2004.

Rob Van der Spek and Andre Spijkervet. Knowledge management: dealing intelligently with knowledge. *Knowledge management and its integrative elements*, pages 31–59, 1997.

Graham Vickery. Review of recent studies on psi re-use and related market developments. *Information Economics, Paris*, 2011.

Karel Vredenburg, Ji-Ye Mao, Paul W Smith, and Tom Carey. A survey of user-centered design practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 471–478. ACM, 2002.

Matthew O Ward, Georges Grinstein, and Daniel Keim. *Interactive data visualization: foundations, techniques, and applications*. CRC Press, 2010.

Cathleen Wharton, John Rieman, Clayton Lewis, and Peter Polson. The cognitive walkthrough method: A practitioner's guide. In *Usability inspection methods*, pages 105–140. John Wiley & Sons, Inc., 1994.

Ryen W White and Resa A Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services*, 1(1):1–98, 2009.

Karl M Wiig. *Knowledge Management Foundations: Thinking about Thinking-how People and Organizations Represent, Create, and Use Knowledge*. Schema Press, Limited, 1994.

Max Wilson, Alistair Russell, Daniel A Smith, et al. mspace: improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM*, 49(4):47–49, 2006.

Thomas D Wilson. Human information behavior. *Informing science*, 3(2):49–56, 2000.

Marco Winckler, Cedric Bach, and Regina Bernhaupt. Identifying user experience dimensions for mobile incident reporting in urban contexts. *Professional Communication, IEEE Transactions on*, 56(2):97–119, 2013.

Anneke Zuiderwijk, Marijn Janssen, Sunil Choenni, Ronald Meijer, R Sheikh Alibaks, and R Sheikh_Alibaks. Sociotechnical impediments of open data. *Electronic Journal of e-Government*, 10(2):156–172, 2012.

Index

Typeset April 10, 2018