FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Evaluation of an Information Visualization Technique for Large Overlapping Sets

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Information and Knowledge Management

eingereicht von

Molham Rajjo
Matrikelnummer 0735308

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: Ao.Univ.-Prof. Mag. Dr. Silvia Miksch
Mitwirkung: Dr. Bilal ALsallakh

Wien, 20.09.2014 ________________ (Unterschrift Verfasser) ________________ (Unterschrift Betreuung)

FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Evaluation of an Information Visualization Technique for Large Overlapping Sets

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Information and Knowledge Management

by

Molham Rajjo
Registration Number 0735308

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: Ao.Univ.-Prof. Mag. Dr. Silvia Miksch
Assistance: Dr. Bilal ALsallakh

Vienna, 20.09.2014

(Signature of Author) (Signature of Advisor)

Erklärung zur Verfassung der Arbeit

Molham Rajjo
Porzellangasse 30, 1090 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum) (Unterschrift Verfasser)

Acknowledgements

First of all, I want to thank my advisors Bilal ALsallakh and Silvia Miksch for their great support and help throughout the thesis. They provided good advice and guidance whenever I needed.

Furthermore, I want to thank my family for their assistance and encouragement during all the years. They provided the right circumstances to successfully accomplish my study.

Finally, many thanks to all who participated in the experiment. It would not have been possible to perform the experiment without their help.

Abstract

Sets are an essential mathematical concept that allow to treat a collection of objects as a mathematical object on its own right. They are widely used in computer science to model a variety of problems, query results, and the results of algorithms. Several problems can be modeled by defining a number of sets over a collection of elements and analyze the relations between these sets. Set-typed data are used to represent the memberships of elements in the sets. The sets that are defined over the same elements potentially overlap. The overlaps between the respective sets contain various patterns that are worth to explore and analyze.

Visualizing overlaps between sets is a challenging problem due to the exponential growth of possible overlaps with the number of sets. Some techniques for visualizing overlapping sets focused on simplifying the representation of the overlaps between sets. Other techniques can be used for large and complex sets. Radial Sets is a new InfoVis technique to analyze set memberships for a large number of elements. It is used for visualizing large overlapping sets in a more scalable and flexible way than conventional methods such as Euler diagrams.

This work presents the results of an empirical evaluation of the Radial Sets technique to explore the usefulness the tasks it was designed. For the evaluation of this technique a quantitative study has been performed. The study was conducted by means of a controlled experiment where 32 participants had to solve tasks that are instances of seven pattern finding tasks. Three hypotheses have been created covering these pattern finding tasks. Each task was assigned to one of these hypotheses. Each user has to solve 60 questions which have been divided into two groups: training questions and evaluation questions. The aim was to evaluate how well Radial Sets are performing these tasks by measuring the compilation time and errors the users made when solving these questions. Therefore, the evaluation questions were the main part of the experiment. The results of the training tasks were excluded from the evaluation results. Additionally, the experiment included a qualitative feedback to elicit usability and understandability aspects based on users opinion.

The evaluation results revealed that Radial Sets are effective at representing sets and how the elements belong to each set. The representation of overlaps as arcs was also intuitive but required detailed explanation about how the overlap sizes are computed, in particular, whether they depict the absolute number of elements, or the proportion of shared elements.

Kurzfassung

Mengen stellen ein grundlegendes mathematisches Konzept dar. dadurch ist es möglich, eine Menge von Objekten als ein individuelles mathematisches Objekt zu betrachten. Das Konzept wird häufig in der Computerwissenschaft verwendet, um verschiedene Probleme und Ergebnissen von Queries zu illustrieren. Ein Problem kann durch die Definition von mehreren Mengen über eine Sammlung von Elementen modelliert werden und die Beziehungen zwischen diesen Mengen können dann analysieren werden.

Außerdem wird die Mengendatenstruktur verwendet, um die Mitgliedschaften von Elementen in den Mengen darzustellen. Die Mengen, die über die gleichen Elemente definiert sind, können sich potentiell überlappen. Die Überschneidungen zwischen den jeweiligen Mengen sind es wert, untersucht und analysiert zu werden.

Die Visualisierung der Überschneidungen zwischen den Mengen ist ein herausforderndes Problem, da die Überscheidungen in Korrelation zur Anzahl der vorhandenen Mengen exponentiell zunehmen. Einige Techniken zur Visualisierung von überlappenden Mengen konzentrieren sich auf die Vereinfachung der Überschneidungen zwischen den Mengen. Für große und komplizierte Mengen wurden dahingehend andere Methoden entwickelt. Radial Sets ist eine visuelle Technik, um die Mengen-Mitgliedschaft für eine große Anzahl von Elementen zu analysieren. Diese Methode ermöglicht eine übersichtlichere und flexiblere Visualisierung der Überschneidungen, als man sie mit klassischen Methoden wie Euler Diagramme erzielen kann.

Diese Arbeit präsentiert die Ergebnisse einer empirischen Evaluation für Radial Sets Technik, um die Verwendbarkeit der beschriebenen Aufgaben zu erläutern. Für die Evaluation wurde eine quantitative Studie angewendet. Sie wurde von einem Experiment mit 32 Probanden geführt, die Aufgaben lösen sollten, die Instanzen von sieben Aufgaben sind. Hierzu wurden drei Hypothesen aufgestellt, die den Ablauf geregelt haben. Jeder Proband musste 60 Fragen lösen, die in zwei Gruppen aufgeteilt wurden: Übungsfragen und Fragen zur Evaluation. Ziel des Ganzen war es, herauszufinden, wie gut das Tool die Aufgaben erledigt. Dies geschieht durch Messung der Zeit und Fehler von der Probanden, als sie die Aufgaben lösten.

Dementsprechend waren die evaluierenden Fragen der Hauptteil des Experiments. Die Übungsfragen wurden dabei von den Evaluationsfragen herausgenommen. Die subjektive Fragestellung war primär als Rückmeldung der Nutzer über die Benutzerfreundlichkeit und den Entwurf bzw.

Probleme derer gedacht.

Die Auswertungsergebnisse haben gezeigt, dass Radial-Sets darin besonders effektive sind, die Sets und die Zugehörigkeit der Elemente zu den Sets zu repräsentieren. Die Darstellung der Überschneidungen als Bögen war auch intuitiv. Eine detaillierte Erklärung ist darüber nötig, wie die Überlappungsgrößen berechnet wurden, insbesondere, ob sie die absolute Anzahl der Elemente oder der Anteil der gemeinsamen Elemente wiedergeben.

Contents

CHAPTER 1

Introduction

Information Visualization (InfoVis) has been defined as ”the use of computer-supported interactive visual representations of abstract data to amplify cognition [1, p. 7]”. For example, line chart is often used to show the development of certain quantities over time. Another example, are Euler Diagrams [2] that depict the overlaps between sets.

As InfoVis research is becoming more established and the methods for evaluating InfoVis techniques are in the focus of InfoVis community being even better defined and explored. Also more research work has being performed to evaluate and compare existing techniques in their efficiency of performing the tasks they are designed for.

Visualization of large overlapping sets in InfoVis remains a relatively unexplored problem. A variety of real-world problems can be modeled by defining multiple sets over a collection of elements. Analyzing and exploring elements-set membership and overlaps between sets provides insights in the data which can be useful for solving such problems [3, 4].

Advanced tools and techniques are needed in order to solve such problems. Visualization techniques can be used to explore the whole data visually and to expose a multitude of patterns in the data. Such techniques are needed to gain insights in the data which might have been overseen by applying traditional methods. For example, a bar chart can be used to analyze the average rating of movies' genres as shown in Fig. 1.1.

As an example, a movie producer might want to extract more knowledge about the data in order to support some strategic decisions. For instance, detect if movies that have multiple genres have a high or low average rating.

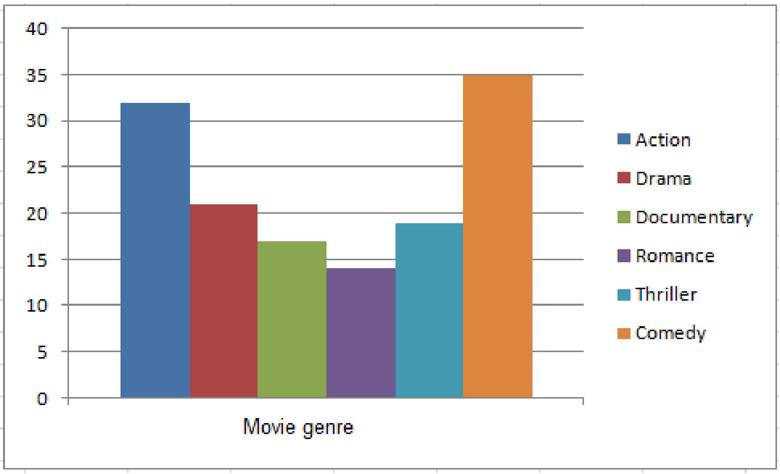


Figure 1.1: A Bar chart visualizing the average rating of movies' genres

Another example is a line chart which visualizes the number of produced movies in the movie industry over time. It can be used to compare the number of produced movies according to their genres as shown in Fig. 1.2. Again, a producer might ask for more details about the produced movies. For example, the number of the produced movies that belong to only (Drama and Action in this example) but nothing else. Such a query may require complex data processing and is time consuming using traditional tools. By using advanced visualization techniques more knowledge and insights about the data can easily be extracted.

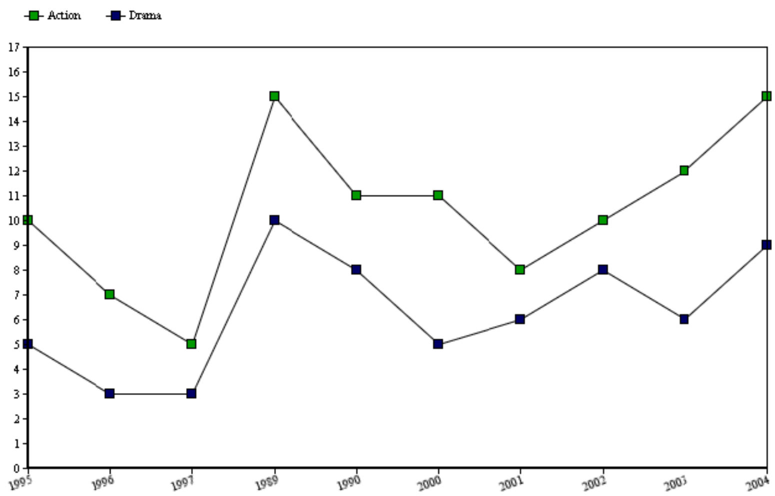


Figure 1.2: A Line chart visualizing the number of produced movies according to their genres and release data

Many InfoVis techniques have been developed for visualizing overlapping sets. Each technique was designed to support a group of tasks and to deal with different kinds and sizes of data-sets. An important question is which kind of visualization is best suited for which kind of tasks and data-sets. Evaluating a visualization technique provides an evidence of its effectiveness and defines for which kind of tasks and data sets it is best suited for.

In order for an Information Visualization technique to be adopted by the industry, a thorough evaluation needs to be conducted. The type of data the visualization is targeted at, the intended tasks, and users the visualization should support need to be clearly specified to design the evaluation accordingly [5].

1.1 Motivation

In many applications heterogeneous data tables contain multi-valued attributes that often store the memberships of the table entities to multiple sets. For example, which languages a person masters, which skills an applicant documents, or which features a product comes with. With a growing number of entities the resulting element-set membership matrix becomes very rich of information about how these sets overlap [6].

A variety of different potential visualization techniques for visualizing overlapping sets exist. It is a challenge for the visualization designers to decide which data representation should be used a new interactive visualization tool. This is because different tasks might need different data visualizations or because some visualizations are better at performing certain tasks.

The InfoVis community has integrated several human perception research results into guidelines and principles [7]. Such principles help the designers to find appropriate visual encoding and interactions for the data being visualized.

However, it is crucial to evaluate if the chosen design decisions are applicable for the possible users. It is impotent to estimate whether the design of the Information Visualization is appropriate for the data that will be represented and is the best for the tasks the users will perform [8].

1.2 Problem Statement

In many areas of science as well as economy, analyzing and evaluating complex multivariate data are necessary. It is difficult especially for large amounts of data to extract knowledge and relations. Therefore, an appropriate representation is needed.

Radial Sets (chapter. 3), is a new interactive InfoVis technique for overlapping sets. It enables quickly finding and analyzing different kinds of overlaps between the sets, and relating these overlaps to other attributes of the table entities [6].

The ineffectiveness of this new technique has not been evaluated. This work provides an empirical evaluation of Radials Sets. The evaluation will conducted by means of quantitative method, e.g., time and error will be recorded and the results will be analyzed.

It has become crucial for researchers to present evidence of measurable benefits to support the adoption of their techniques. In other words, it is necessary for a well adoption of novel visualization techniques to provide evidence that the visualizations satisfy their proposed purpose and meet the expectations and needs of users. Moreover, it is necessary to conduct an evaluation in order to move the concept to application [9].

Conducting an evaluation of a new visualization tool is a non-trivial task. The designer has to choose the right tasks to be performed and the right research questions to be answered. Additionally, it is also a challenge to pick the appropriate evaluation methodology [8].

After choosing the evaluation methodology, the evaluation has to be designed with the aim to answer the defined research questions. From the research questions one or more hypotheses have to be derived. During the execution of the evaluation, specific data need to be collected (e.g., time and error). The collected data have to be reviewed and analyzed to provide an evidence in order to accept or reject the hypothesis [9].

There is awareness that evaluation is important. It is possible to measure the effectiveness of a new visualization technique with a quantitative evaluation, while a qualitative evaluation can be used to assess the clarity of the conceptual design [8, 9].

1.3 Aim of the work

The main objective of this thesis was to conduct an evaluation of the Radial Sets technique which will be introduced in details in chapter. 4. The evaluation aimed to assess the effectiveness of the new visual technique for dealing with large overlapping sets in performing and solving the pattern-finding tasks mentioned in chapter. 3.

Evaluating how well Radial Sets are performing these tasks was measured in the study regarding time and correctness. The conclusions from the evaluation will discuss possible improvements of Radial Sets and provide useful understanding for future designers of visualization tools.

1.4 Research Questions

The thesis addresses the following research questions:

- **State of the art research:**
  - **Q1:** Which related visualization techniques for overlapping sets are described in the scientific literature?
  - **Q2:** Have these techniques been evaluated?
    If yes, which principles and scenarios were used? And what were the results?
- **Evaluation:**
  - **Q1:** Is the Radial Sets technique effective in performing tasks it is designed to support?
  - **Q2:** How can the Radial Sets technique be improved to satisfy the objectives of the design?

1.5 Methodological approach

At the beginning it was essential to define the hypothesis/hypotheses based on the research questions related to the evaluation. The next step was choosing the appropriate evaluation design based on literature of different evaluation scenarios and methodologies, which was necessary for the validity of the evaluation results and the concluded results.

Different scenarios and methods for the evaluation of visualization techniques, such as qualitative (e.g., observations or interviews) and quantitative (e.g., the laboratory experiment) evaluation techniques have been discussed [10]. For the evaluation of Radial Sets, quantitative evaluation methods have been used. This includes measuring the time and error made by the users when performing the evaluated tasks. In addition, the new technique has been qualitatively evaluated in order to assess the clarity of the conceptual design and to elicit usability issues.

Based on the hypotheses and the methodology, a list of appropriate evaluation questions had to be determined, an appropriate data-set had to be found, and the number of subjects had to be decided. The experiment interviews were based on questionnaires that involved tasks and questions that are instance of the patterns-finding tasks (mentioned in chapter. 3). The questionnaires included questions about participants' demographic data and whether they have experience with Information Visualization.

The quantitative and qualitative data were collected automatically using a software library for visualization evaluation, EvalBench [11]. It supports both quantitative and qualitative evaluation methods. It enables users to perform a list of tasks according to the evaluation design and measures time and error the users made when solving these tasks. The data were recorded in log files and xls files.

Once the evaluation design was built and the required XML files for EvalBench were created, the evaluation of Radial Sets was finally ready to be performed. The next step was inviting the participants for the experiment interviews. At the beginning users got an introduction about set-typed data and the most common representation for this type of data. The next introduction was about the new visualization technique. At the end of the introduction users received a short tutorial about the visualization tool.

While users were solving the tasks, the time and error they made were recorded automatically using EvalBench. The participants' demographic data and their experience with Information Visualization were collected at the beginning of the interview. The qualitative data and the participants' notes were collected at the end of it.

The collected quantitative and qualitative data were reviewed and analyzed. The analysis results and the conclusion will be discussed later in chapter. 5. Additionally, further improvements, recommendations, and suggestions will be discussed at the end of this work.

CHAPTER 2

Related Work

This chapter introduces a general overview on the evaluation of Information Visualization techniques. Some strategies and approaches for evaluating InfoVis tools are discussed in the first section.

The second section is a discussion of some visualization techniques for overlaps between sets. A brief description of each technique is provided followed by a discussion of the scalability of the techniques and which tasks they support. How each technique has been evaluated and which methods have been used is also presented. The techniques are categorized into two groups: those with user study and those without user study.

Set-typed data is introduced and the most common representation for this type of data (Euler Diagrams [2]) is presented. Visualizing overlaps between sets will be discussed along with the limits in terms of the number of sets that can be visualized at once.

Additionally, some other related work is discussed. Particularly, the evaluation of another technique, called Contingency Wheel [12] is studied. This is because Radial Sets use the same visual metaphor as Contingency Wheel.

In the final section of this chapter the results of the discussed techniques is presented as a summary. It includes both categories along with the scalability of each technique in terms of the number of sets and the elements they can depict.

2.1 Evaluation of Information Visualization

Information Visualization techniques have a wide variety of applications in different domains. For example, in the medical domain visualization tools are used to analyze and explore medical data to get more knowledge about patients and diseases. It could be difficult to infer such knowledge by using other analysis or statistical methods.

The variety of InfoVis techniques results in several methods and diverse approaches for evaluating these techniques. These evaluations provide an evidence of utility and effectiveness of the new Information Visualization technique.

Seven scenarios have been studied by Lam et al. to evaluate Information Visualizations [8]. These scenarios can be used to set the evaluation goals, pick the research questions, and to consider the appropriate methodology for the evaluation. This encourages the selection of the evaluation goals before considering the methods.

Lam et al. classified 361 papers that include evaluations according to 17 tags as shown in Table. 2.1. These tags have been summarized to the seven scenarios. The scenarios are categorized into scenarios for understanding data analysis processes and scenarios which evaluate the visualizations themselves [8].

The main goal of the evaluation in the first category is to understand the underlying process and the roles played by visualizations. This may requires recording the users' performance and feedback in order to analyze the user experience.

In the second category, the evaluation focuses on the visualization itself in order to test usability issues or the design concept. In this case only a part of the visualization techniques may be tested. Each scenario has an identified goal, a definition, a group of common evaluation questions, and applicable evaluation methods [8].

From these seven scenarios this work focuses on the following two scenarios, since they are most related to our tasks:

- Evaluating visual data analysis and reasoning (VDAR), which was used to measure the effectiveness of the new technique, Radial Sets, in analyzing the data, performing the tasks, and deriving knowledge about domain of the data set.
- Evaluating user performance (UP) to measure the time and error when users perform the tasks. The (UP) scenario was also used to assess the clarity of the conceptual design and to elicit usability issues.

The outputs of both scenarios are numerical values, along with confidence intervals for these values (chapter. 5).

Paper Tags	EuroVis	InfoVis	IVS	VAST	Total	Scenario
Process						
1. People's workflow, work practices	3	1	3	0	7	UWP
2. Data analysis	0	5	3	5	13	VDAR
3. Decision making	0	2	1	4	7	VDAR
4. Knowledge management	0	1	0	2	3	VDAR
5. Knowledge discovery	1	1	0	1	3	VDAR
6. Communication, learning, teaching, publishing	0	0	4	1	5	CTV
7. Casual information acquisition	0	4	0	0	4	CTV
8. Collaboration	0	3	2	4	9	CDA
Visualization						
9. Visualization-analytical operation	0	12	1	0	13	UP
10. Perception and cognition	17	24	15	3	62	UP
11. Usability/effectiveness	25	84	31	18	158	UP&UE
12. Potential usage	7	1	5	9	22	UE
13. Adoption	0	1	3	1	5	UE
14. Algorithm performance	17	37	15	0	69	VA
15. Algorithm quality	1	10	12	5	28	VA
Not included in scenarios						
16.Proposed evaluation methodologies	0	3	0	2	5	-
17.Evaluation metric development	2	6	1	1	10	-

Table 2.1: Original coding tags, the number of papers classified, and the final scenario to which they were assigned. **UWP:** Understanding environments and work practices. **VDAR:** Evaluating visual data analysis and reasoning. **CTV:** Evaluating communication through visualization. **CDA:** Evaluating collaborative data analysis. **UP:** Evaluating user performance. **UE:** Evaluating user experience. **VA:** Evaluating visualization algorithms (adapted from Lam et al. [8]).

The diversity of the existing evaluation methodologies reflects the difficulty in deriving comprehensive taxonomy for them. For example, laboratory based method can be used to summarize the effectiveness of an interface (summative) or to inform design (formative) [8, 13, 14]. Lam et al. summarize taxonomies of existing evaluation methods and their respective focus as shown in

Table. 2.2. They emphasize performing evaluations that are based on the evaluation goals and questions instead of methods and methodologies [8].

Carpendale [15] discussed the evaluation of Information Visualization in general. She discussed the importance of empirical research to encourage the adoption of visualization tools. Some challenges facing empirical research have been listed. For example, choosing the right questions, the right methodology, appropriate data analysis, and finally relating the new results to the existing research results. Possible types of empirical methodologies have been discussed [15, 16] as shown in Fig. 2.1.

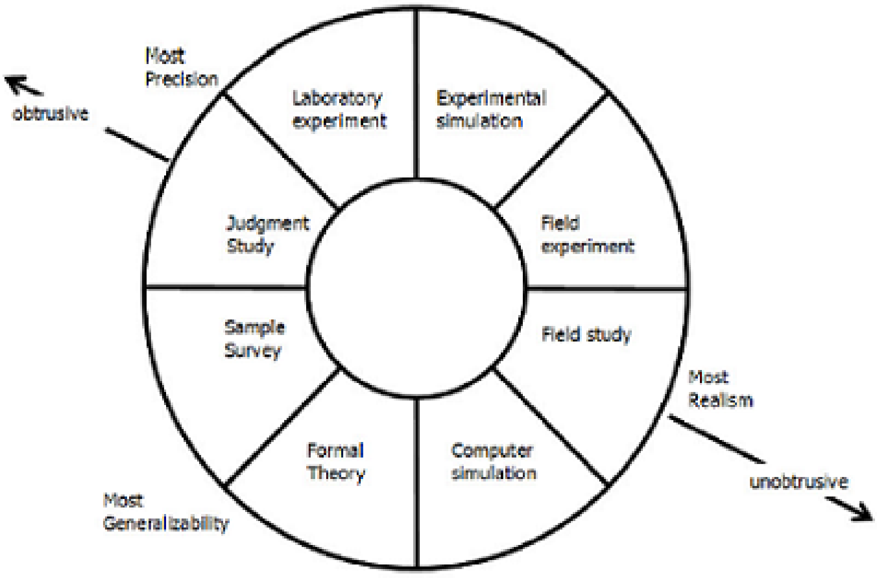


Figure 2.1: Types of methodologies organized to show relationships to precision, generalizability, and realism.(adapted from Carpendale [15, 16]).

Carpendale mentioned that all studies share some common factors. For example, they all start with questions, they all relate the research questions to the existing concepts and research results, and they all have a method [15].

Quantitative (e.g., laboratory experiment) and qualitative (e.g., observations, interviews) evaluation techniques have also been discussed. She explained the methodology and challenges of each evaluation method. Quantitative evaluation encompasses defining one or more hypothesis, determining the dependent and independent variables, and identifying the statistical methods. A simple process of a traditional experiment has also been introduced [15] as shown in Fig. 2.2.

Plaisant [9] summarized the current evaluation practices and reviewed some related challenges, for example, improving usability testing or matching the new tool with real problems. Users might need sometimes to explore the data from multiple perspectives over time and to use different tools. They also might have to answer unexpected questions before having a look to the visualization. For example, biologists might want to analyze the data set for months in order to find patterns [9].

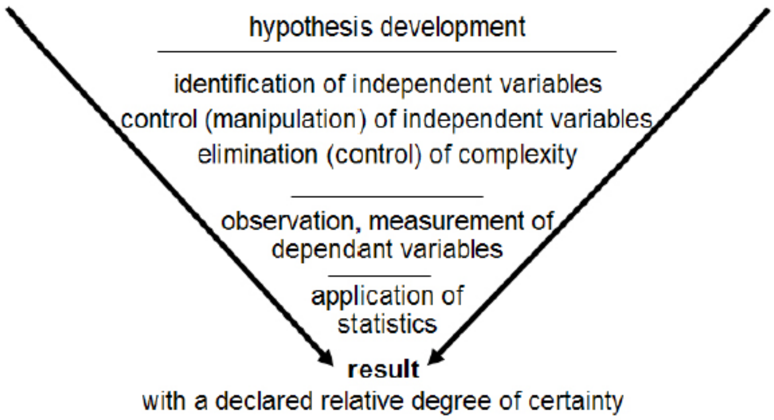


Figure 2.2: A simple schematic of the traditional experimental process (adapted from [15]

She also presented some refined evaluation methodologies and discussed possible steps to improve Information Visualization evaluation, for example, creating benchmark data sets and tasks. Case studies report on users performing real tasks. They are used to describe the entire process and the reaction of dealing and exploring the data for the first time. However, the results can not be generalizable [9].

In summary, there are various evaluation methods for evaluating Information Visualizations, summarized in Table. 2.2. Choosing the appropriate method depends on the purpose. The evaluation results might not fulfill the goal, unless the evaluators define the right questions to ask, set the right relevant instances (tasks), choose the right variables to evaluate, select the appropriate data set to test or users, and to determine the appropriate evaluation method. Such a procedure is not straightforward and can be challenging for evaluators [8].

Type	Categories	Refs
Evaluation goals	Summative (to summarize the effectiveness of an interface), formative (to inform design).	Andrews [13], Ellis and Dix [14]
Evaluation goals	Predictive (e.g., to compare design alternatives and compute usability metrics), observational (e.g., to understand user behavior and performance), participative (e.g., to understand user behaviour, performance, thoughts and experience).	Hilbert and Redmiles [17]
Evaluation challenges	Quantitative (e.g., types validity: conclusion (types I & II errors), construct, external/internal, ecological), qualitative (e.g., subjectivity, sample size, analysis approaches).	Carpendale [10, 15]
Research strategies	Axes (generalizability, precision, realism, concreteness and obtrusiveness) and research strategies (field, experimental, respondent, theoretical).	McGrath [16]
Research methods	Class (e.g., testing, intersection), type (e.g., log file analysis, guideline reviews), automation type (e.g., none, capture), effort level (e.g., minimal effort, model development).	Ivory and Hearst [18]
Design stages	Nested Process model with four stages (domain problem characterization, data/operation abstraction, encoding/interaction technique design, algorithm design), each with potential threats to validity and methods of validation.	Munzner [19]
Design stages	Design/development cycle stage associated with evaluation goals "exploratory" with "before design", "predictive" with "before implementation", "formative" with "during implementation", and "summative" with "after implementation"). methods are further classified as inspection (by usability specialists) or testing (by test users).	Andrews [13]
Design stages	Planning & feasibility (e.g., competitor analysis), requirements (e.g., user surveys), design (e.g., heuristic evaluation), implementation (e.g., style guide), test& measure (e.g., diagnostic evaluation), and post release (remote evaluation).	Usability.net [20]
Design stages	Concept design, detailed design, implementation, analysis.	Kulyk et al. [21]
Data and method	Data collected (qualitative, quantitative), collection method (empirical, analytical).	Barkhuus and Rode [22]
Data	Data collected (qualitative, quantitative, mixed-methods).	Creswell [23]
Evaluation scope	Work environment, system, components.	Thomas and Cook [24]

Table 2.2: Taxonomies of evaluation methods and methodologies based on the type of categorization, the main categories themselves, and the corresponding references (adapted from Lam et al. [8]).

.

2.2 Techniques for visualizing overlapping sets

Sets are an essential concept in mathematics. A set is a collection of unique objects, which are called elements of the set. Elements of a set are grouped together with a certain property in common, for example: the elements of the set ”Clothes” share the property ”things to wear” [6].

Sets are simple and because of their generic notion, they are widely used in computer science to illustrate real-world concepts, for example: which markers a gene contains, or which properties a product has. Sets are also used to represent query results and the results of different algorithms [6].

Set-typed data are commonly used to represent the membership of a collection of elements in different sets, for example they can represent people memberships of different clubs, or the features a product comes with [6]. In a data-set, the sets that are defined over the same elements potentially overlap. As the number of elements increase, the overlaps between the respective sets contain various patterns that are worth to explore and analyze [6].

Visualizing overlaps between sets is a challenging problem that has been approached in various ways. The major reason behind the complexity of this problem is the exponential growth of possible overlaps according to the number of sets: A set system with (m) sets can exhibit up to (2^m) distinct intersections between the sets [6, 25]. Each element lies in one of these intersections, based on its memberships of the different sets. Although a large portion of these distinct intersections is empty in practice, the number of non-empty overlaps can still be large, even with a dozen of sets. These overlaps are salient features of set data with many analysis tasks typically concerned with different kind of overlaps between the sets [6].

Some techniques for visualizing overlapping sets bypass the complexity problem by limiting the number of sets and overlaps that can be visualized at once. Other techniques avoid visualizing the overlaps explicitly and convey more abstract information about the set system instead [6].

In the following, some existing techniques for visualizing overlapping sets are presented. A brief description of each technique will be introduced and the evaluation method will be discussed. The techniques are categorized into two groups: those with user study and those without user study. Finally, as a summary, the techniques will be compared in a table according to their respective categories.

Techniques with user study

Euler diagrams [2, 26] are the most familiar and natural representation for set-typed data. Their graphical representations are widely used to provide a very effective way for depicting overlaps between sets [27].

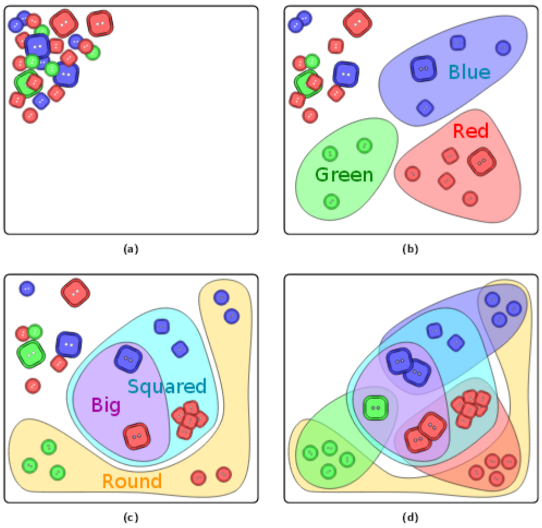


Figure 2.3: An example of an Euler diagram derived from the process of ordering a box of buttons. (a) The buttons to be ordered. (b) Displacement of the buttons according to their colour. (c) Further organization of the space according to size and shape of the buttons. (d) The Euler diagram that can be extracted from the process (adapted from Simonetto [27]).

Euler diagrams represent sets as internal region of closed curves. Elements are placed inside the region of the set which they belong to. Elements inside two or more regions represent the overlap between these sets as shown in Fig. 2.3 [27]. Euler diagrams have a wide variety of uses in many diverse areas. They are a very valuable Information Visualization technique, since they can be used to easily retrieve non-trivial patterns and knowledge from complex data [6].

Evaluation

Some user studies have been conducted for evaluating Euler diagrams [28–30]. However, these diagrams are severely limited in the number of sets they can handle, since the complexity of the diagram increases rapidly with an increasing number of sets. This is because the number of possible overlaps grows exponentially (2^m) with the number of sets (m) which exceeds the topological constraints of these diagrams. Therefore, possible overlaps can be depicted clearly only with a small number of sets $(m \leq 4)$ [6].

It has been showed that for any collection of sets up to eight sets $(m \leq 8)$, non empty overlaps between these sets can be represented by an extended Euler diagram. Such diagrams are defined by some properties such as relaxing the conditions on the contours and allowing holes in the regions [31].

There is a variety of techniques to generate Euler diagrams. They focused on different aspects of the diagrams generation. For example, some focused on the draw-ability of any Euler diagram where others focused on the readability of the generated diagrams.

Some of them are evaluated, such as ComED [32], and others are not, such as Rodgers et al. method [33]. However all techniques are restricted to a small number of sets in compare to Radial Sets.

Riche and Dwyer [32] presented two approaches to simplify the overlaps between sets resulting in a strict hierarchy that can be easily arranged and drawn. The first approach, called **ComED**, splits the overlapping sets, and used compact rectangular shapes for representing the sets. The split regions of a particular set are linked with hyperedges.

The second approach, called **DupED**, avoids depicting the overlaps between the sets explicitly. It rather represents the set regions with simple separate rectangles, and places the elements that belong to a set inside the set region. The elements that belong to multiple sets are duplicated in each set regions. The instances of the same elements are linked with hyperedges as shown in Fig. 2.4 [32].

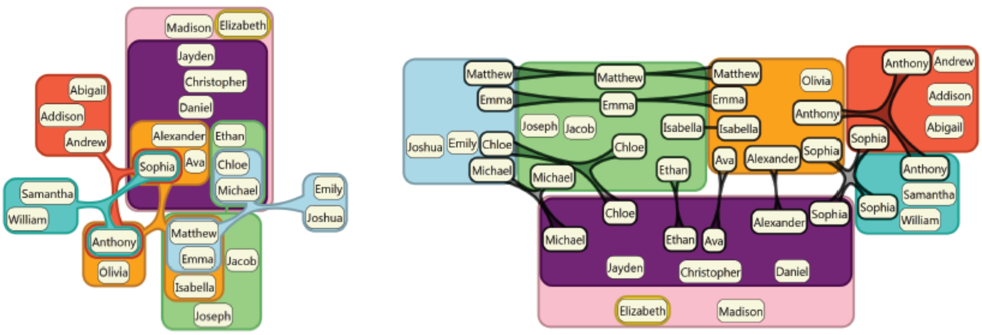


Figure 2.4: Compact Rectangular Euler Diagram (left) and Euler Diagram with Duplications (right) (adapted from Riche and Dwyer [32]).

Evaluation

Riche and Dwyer evaluated the readability of ComED, DupED and DrawnED (Hand-Draw Euler diagrams) by performing a controlled experiment. The evaluation aimed to measure how users deal with both general and detailed tasks. Two controlled experiment have been performed. They involved 5 readability tasks such as, count the number of the sets and assess the elements in a specific overlap [32].

Hypothesis: The controlled experiment has been built based on four hypotheses. The hypotheses assumed generating more readable diagrams using one of the three techniques. Some hypotheses assumed that using a particular technique would result in increasing or decreasing the performance. Comparing the effectiveness of the techniques in solving the related tasks has also been discussed. For example, the first hypothesis assumed that DupED is more effective than the other techniques for solving tasks related to sets. While ComED is more effective for solving tasks related to elements [32].

Method: The first experiment compared the performance of ComED and DupED. The second controlled experiment aimed to assess the performance of ComED in comparison with traditional Euler diagrams, due to the results of the first experiment. They conducted a comparative experimental study for ComED, DupED and Hand-Draw Euler Diagrams (DrawnED) [32].

The same procedure has been used for both experiments, a within-subject design. For example, in the first experiment, two techniques have been evaluated with four different levels of difficulty of the data set. The evaluation runs three times on different orders of the data sets with 5 tasks to be solved.

(Exp1): 2 Vis x 4 Levels x 5 Tasks x 3 repetitions.
(Exp2): 3 Vis x 4 Levels x 5 Tasks x 2 repetitions.

The users received training on each technique before the evaluation. Time and error have been recorded while users were performing the tasks. The time of each task has been limited to 40 seconds. User comments have been collected using a questionnaire. The experiment lasted 60 minutes including training and post-experimental questionnaire [32].

Tasks: The evaluation contained five readability tasks to be solved by the users. The main focus of these tasks was on sets and elements. Such as, set count or element membership. One task was introduced concerning the overlaps between the sets. Users have to answer multiple choices questions that are instances of these tasks. The order of the tasks was fixed, where the order of the datasets was randomized to avoid memorization effect. An example of the tasks and their instances is: (Elements membership) as the task, and (Which set(s) contain element 0?) as an instance of the task [32].

Users: 18 users with general computer experience have been recruited for the experiments, 9 for each one. Users have been classified according to age and gender as shown in Table. 2.3.

	Users	**Male**	**Female**	**Color-blind**	**Age range**
Experiment 1	9	6	3	0	21- 47
Experiment 2	9	2	5	0	25- 40

Table 2.3: The number of users participated in the experiment

Data set: The number of sets and elements used in the experiments has been controlled. Also the number of overlaps and discontinuous set regions has been limited. Multiple instances of the data sets have been created to avoid memorization.
Table. 2.4 shows the number of sets, elements, and overlaps used in each experiment [32].

Experiment 1	**Sets**	**Elements**	**2-set**	**3-set**	**4-set**	**disc.**
(D1) Easy min.	4-5	~10	~3	3	0	1
(D2) Easy add.	4-5	~15	~3	3	0	1
(D3) Med min.	6-7	~15	~4	2	1	1
(D4) Med add.	6-7	~25	~4	2	1	1
Experiment 2						
(D3) Med min.	6-7	~15	~4	2	1	1
(D4) Med add.	6-7	~35	~4	2	1	1
(D5) Hard min.	8-9	~25	~6	3	2	3
(D6) Hard add.	8-9	~45	~6	3	2	3

Table 2.4: Parameters used to generate Euler diagrams per difficulty level(adapted from Riche and Dwyer [32]).

Results: The results have been analyzed for each experiment using ANOVA (analysis of variance). Time and error for each task were reported. User preferences and a comparison between the techniques have been listed [32]. Table. 2.5 shows the summary of the results.

Task	**Accuracy**	**Time**	**Preference**
SetCount	DupED =ComED = DrawnED	DupED < DrawnED <ComED	DupED >ComED > DrawnED
SetComparison	DupED >ComED = DrawnED	DupED < DrawnED <ComED	DupED >ComED > DrawnED
SetIntersection	DupED >ComED = DrawnED	DupED <ComED = DrawnED	DupED >=ComED > DrawnED
EltCount	ComED = DrawnED > DupED	DrawnED <ComED < DupED	ComED > DrawnED > DupED
EltMembership	ComED = DrawnED = DupED	DupED <ComED = DrawnED	DupED >=ComED > DrawnED

Table 2.5: Summary of the results (adapted from Riche and Dwyer [32]).

However, both approaches suffer from several limitations in terms of scalability. The authors recommended further experiments with larger and complex data-sets to assess the scalability of these methods [6].

Euler-like methods have also been used to reveal set memberships over existing visualizations. Another layout and additional attributes are used to determine the positions of the visual items [6]. An examples of such methods are Bubble Sets [34], LineSets [35] and Kelp diagrams [36].

LineSets, is a set visual representation based on linking all elements of the sets with a curve. Alper et al. [35] mentioned that it can be used for large and complex sets. It improves the read-

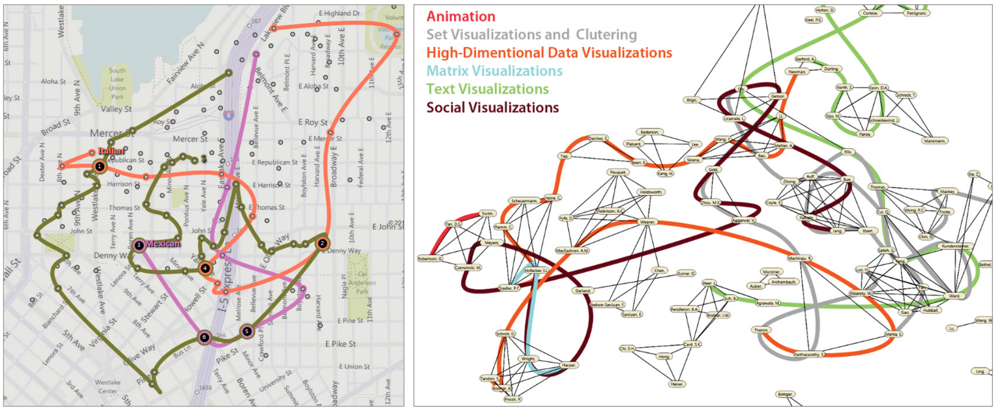


Figure 2.5: LineSets showing restaurant categories on a map (left), LineSets showing communities on a social network (right). (adapted from Alper et al. [35]).

ability of the overlaps between sets by avoiding or minimizing representing shapes overlaps. This results in supporting more readability tasks. For example, allow users to identify how two or more sets overlap with each other as shown in Fig. 2.5 [35].

Evaluation

To explore the potential of LineSets, a controlled experiment has been performed. The evaluation aimed to assess its effectiveness by means of a user study. The controlled experiment has been conducted comparing LineSets with another technique called Bubble Sets (discussed later in this section) [34, 35].

The controlled experiment measured error, time, and user preference. The procedure used in the study was a within-subjects design:

2 Visualizations (LineSets, Bubble Sets) X 2 Data type (map, social network) X 3 Difficulty levels (number of elements, sets, and intersections) X 4 Tasks of varying complexity.

Data set: The study has been conducted on two types of data-sets: hotels and social networks with different levels of difficulty. The difficulty level has been defined by limiting the number of sets, the set sizes (the number of elements), and the number of sets overlaps.

Tasks: In order to assess the readability of LineSets, four generic tasks have been chosen. The tasks cover both overview questions (e.g., “how many sets?”) and detail questions (e.g., “which sets does a particular element belong to?”). An example of the tasks and the instances of these tasks are listed in Table. 2.6 [35].

	Task Type	**Task Text**
T1	Overview: number of sets	"How many groups of hotels are shown?"
T2	Overview: size of a set	"Which one is tagged more in users profiles, Matrix or Pulp Fiction?"
T3	Membership	"Which bands do Alan and Tim both like?"
T4	Intersection	"How many hotels have free parking and breakfast?"

Table 2.6: Tasks and associated examples used in the experiment (adapted from Alper et al. [35]).

Users: 12 users have been recruited for the experiment. Each user answered 24 multiple choice questions per technique. The session lasted about 60 minutes. (RM-ANOVA) repeated measures analysis of variance has been used to analyze the collected data. Subjective users' ratings about the readability of the techniques have also been reviewed. A summary of the comparison between LineSets and Bubble sets is shown in Fig. 2.6 [35].

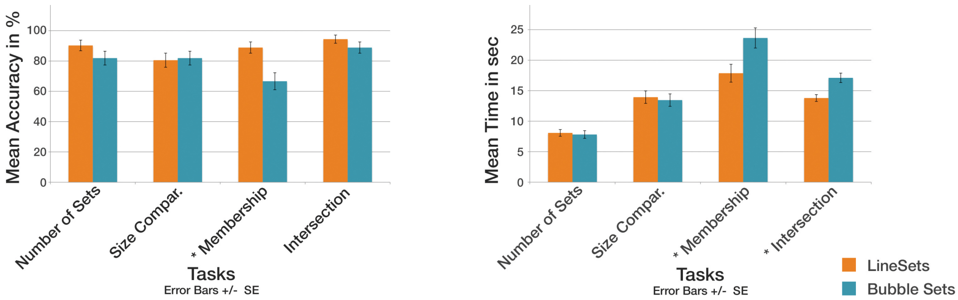


Figure 2.6: Summary of the results, mean accuracy (left) and task completion times (right). (adapted from Alper et al. [35]).

In summary, the scalability of LineSets depends on set size. Limitations include the representation of exact same sets, which are difficult to identify in LineSets. This is because the curves are superimposed. Some solutions have been proposed to solve these limitations. For example, it could be possible to offset the LineSets for same sets which makes the curves become parallel and similar sets more salient [35].

KelpFusion [37], is a method for visualizing set membership of items. It is based on a hybrid representation that uses a mix of hull techniques such as Bubble Sets [34], Euler diagrams [2] as well as line/graph-based techniques such as LineSets [35] and Kelp Diagrams [36].

KelpFusion generates fitted boundaries for groups of elements in a given arrangement. By using a fixed allocation area for each set and scaling the representations of the sets to fit within the allocation area, the readability of the set overlaps will be improved as shown in Fig. 2.7 [37].

Figure 2.7: KelpFusion applied to restaurants in Boston (left) and to cities in Europe (right). (adapted from Meulemans et al. [37]).

Evaluation

To assess the readability of KelpFusion, a controlled experiment comparing it to Bubble Sets and LineSets has been performed. The goal was to evaluate the mixed use of hulls and links compared to a single concave hull as generated by Bubble Sets.
The controlled experiment has been conducted with a within-subject design:

3 Visualization Techniques X 4 Tasks X 2 Difficulty Levels X 3 Repetitions [37].

Data set: Real data of restaurant locations in the Boston area gathered from Bing Maps has been used as a data set. The data has been grouped to cuisine, price qualification, and rating of restaurants in order to form the sets. They filtered the data to define different levels of difficulty for the set arrangements. The number of sets, the number of elements in each set, and the number of 2-set, 3-set, and 4-set intersections have been controlled as shown in Table. 2.7 [37].

	# Sets	# Elements	# 2-set	# 3-set	# 4-set.
Medium 1	4	15 to 39	22	3	0
Medium 2	4	17 to 49	17	3	0
Hard 1	5	12 to 29	17	4	2
Hard 2	5	14 to 29	16	5	2

Table 2.7: Data set statistics. (adapted from Meulemans et al. [37]).

The spatial arrangement of the sets was not controlled and real geographic data was used. The colors for the sets were based on ColorBrewer [38]. The experiment lasted 60 minutes including training for each visualization technique with two participants at a time [37].

Users: Accuracy and completion time of the tasks performed by the users has been measured. 13 users (7 males and 6 females) with general computer experience participated in the study. Each user had to answer 72 questions in a multiple-choice format. User preferences and comments have also been recorded [37].

Tasks: The evaluation involved 4 readability tasks to be solved by the users. The tasks focused on sets, for example, exploring elements contained in a set. On elements, For example, determining which sets an element belongs to and on overlaps between sets. Table. 2.8 shows the task types and an example for each type [37].

Tasks	Example
Size Overview	Are there more Thai or more French restaurants?
Size Count	How many restaurants serve Italian food?
Sets Intersection	How many Thai restaurants are rated 5?
Set Membership	What is the highlighted restaurant?

Table 2.8: Task types and associated examples. (adapted from Meulemans et al [37]).

Hypotheses, Results: Seven hypotheses have been conjectured. Five hypotheses related to accuracy and completion time and two related to participants' preferences. They covered comparing the effectiveness of the two evaluated techniques in solving the tasks, assuming a better performance of LineSets. For example, for the size overview tasks they assumed that KelpFusion will outperform LineSets or the readability of KelpFusion and LineSets is better than Bubble Sets. They used repeated-measure analysis of variance (RMANOVA) to analyze accuracy and time performance results. Fig. 2.8 shows a summary of the time and accuracy results.

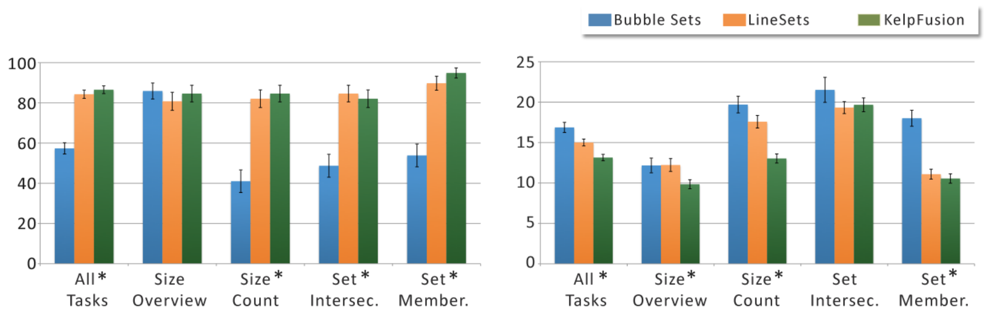


Figure 2.8: Accuracy (left) and time (right) results. (adapted from Meulemans [37]).

Finally, the limitations of the experiment have been presented. For example, the inferred results apply to a limited number of data sets and low level tasks. The scalability, the advantages, and drawbacks of the technique in comparison to other techniques has been addressed [37].

In summary, methods for generating Euler diagrams and Euler-Like diagrams often enforce several restrictions on depicting the set, the elements, and the overlaps between the sets. They are severely limited in the number of sets they can handle. They can partially cover the tasks related to sets count, elements count, and overlaps between sets [6].

Other approaches have been presented to visualize element-set memberships using different visual representations than Euler diagram. For example, some methods used node-link diagrams as visual representations. Others involve matrix-based or frequency-based representations [6].

A matrix can be used to depict the element-set membership by representing the sets as columns and the elements as rows. The ordering of the rows and columns can simplify the matrix, which improves the ability to find pattern in the matrix [39], such as finding a group of elements that tend to have similar patterns of element-set membership [40, 41].

A variety of approaches have been devised for matrix reordering to allow pattern discovery [42]. In addition many interactive tools have been presented for handling the reorderable matrix [43]. For example EnsembleMatrix [44] and MatrixExplorer support the exploration of social networks [45]. Other methods use matrices to visualize element-set membership. The simple approach of a matrix provides a flexible way for representing such relationships.

ConSet [46], is an interactive visualization tool to explore relationships among multiple sets. The sets are depicted as rows and the elements are depicted as columns. The element-set memberships are represented in the cells as shown in Fig. 2.9.

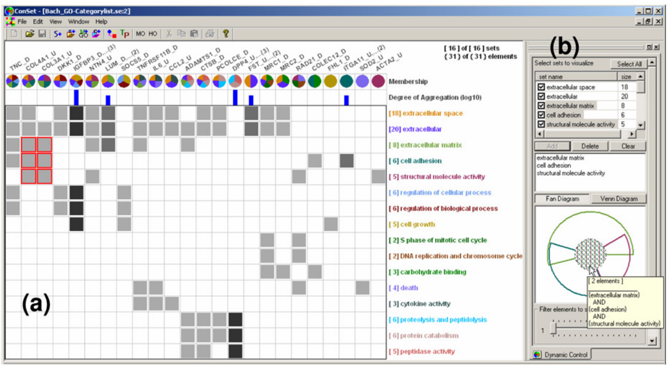


Figure 2.9: ConSet with 16 sets and 31 elements, (a) The Permutation Matrix view shows an overview of the relationships among sets and elements. (b) Dynamic Control view enables users to filter sets and elements. (adapted from Kim et al. [46])

The elements information such as element name, set membership, and degree of aggregation are summarized from top to bottom, each one in a separate row. The cells can be coloured by color-coded set membership which allows determining the sets an element belongs to [46]. Reordering methods for sets and elements have been used, such as HAC (Hierarchical Agglomerative Clustering) ordering [47]. For example, a row is moved to the top and ordered by name and cardinality. A column is moved to the right and ordered by name and number of set memberships. This results in simplifying the matrix and facilitating solving several patterns-finding tasks. For example, finding a group of elements that has the same set-membership [46].

The relationship between sets can be visualized in a dynamic control view by highlighting the elements that belong to an intersection. This allows exploring the overlaps between the sets and solving related tasks. For example, identifying all elements that belong to an intersection between two or more sets [46].

Evaluation

To evaluate how well ConSet works, a qualitative usability study has been performed. The aim of the study was to assess and to identify usability issues. In order to augment the study, the authors compared ConSet with another tool that is designed to solve the same tasks, called VennMaster [48].

During the study, time and error that users made have been measured. However, the study had some limitations to be considered as a controlled user study. The time to complete each task has been measured using a stop watch. The number of errors users made when they are solving the tasks, the number of time-outs and give-ups has been counted [46].

Users: The user study recruited 8 users (5 males and 3 females). The users were biologists. One pilot study has been performed before the evaluation [46].

Tasks: The evaluation involved 9 tasks to be conducted in 3-minute time limit for each task. Users had the possibility to give up a task at any time. The main focus of the tasks was estimating the set sizes and the elements and intersections between the sets following a group of questions. An example question is determining the three largest sets or naming the elements in the intersection of three sets. The same procedure was repeated for both techniques. Each session lasted 38 minutes on average [46].

Data set: Two similar data sets from GoMiner have been used. Each data set includes two files, the category and the gene summary file. The tool combined these files to generate the sets of genes. The number of the sets and the elements in each data set is shown in Table. 2.9.

	# Sets	# Elements
Data set 1	16	31
Data set 2	23	28

Table 2.9: Data set statistics

Users' notes and suggestions concerning usability issues were collected during the sessions and reviewed afterward. No statistical analysis on the measured variables has been performed. They argued that the number of users was too small. The results were reported as raw numbers without referring to statistical significance. A summary of the results is shown in Fig. 2.10 [46].

Finally, the limitations of the study have been discussed. It is important to mention that a usability study has been performed. The comparison with VennMaster aimed to augment the study. The study might be considered as a controlled user study, since time and error have been measured but after improving the limitations [46].

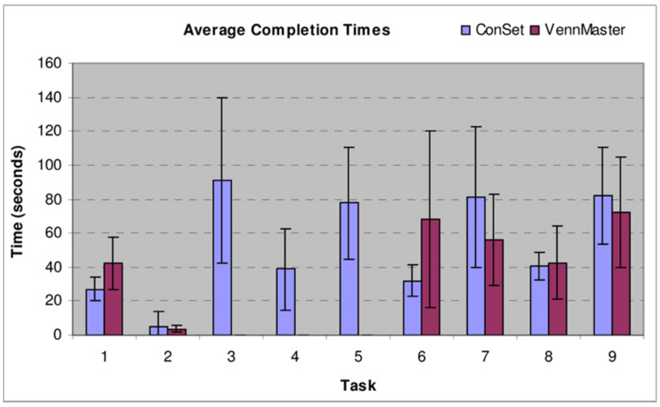


Figure 2.10: ConSet, VennMaster average completion times. (adapted from Kim [46]).

Ghoniem et al. [49] presented a comparative evaluation to assess the readability of graphs representations. It has been conducted by comparing two representations of graphs, matrix-based representations, and node-link diagrams.

The evaluation has been performed on seven generic tasks. For example, counting the number of nods in the graph and finding a link between two specified nodes. The hypothesis assumed that the representation is readable for a given task if the users can answer it quickly and correctly. If a user needs more time or answers wrong, the representation is not well-suited for that task. 36 subjects with advanced experience in computer science (postgraduate students and confirmed researchers) participated in the evaluation [49].

The used data was random graphs with different sizes and different link densities as shown in Table. 2.10. The compared graphs were not familiar to user (e.g., equally unfamiliar) [49].

Size/ Density	**0.2**	**0.4**	**0.6**
20	graph 1	graph 2	graph 3
50	graph 4	graph 5	graph 6
100	graph 7	graph 8	graph 9

Table 2.10: The types of graphs used for the experiment (adapted from Ghoniem et al. [49]).

An evaluation program has been developed to represent the graphs according to the both representation techniques. Time and error have been recorded while the users are performing the tasks. The representation technique, matrix or node-link has been selected randomly to avoid memorization. Each evaluation session consists of 126 questions with 45 seconds as a limit time for each task. The same procedure has been performed for both techniques:

2 visualization x 9 graphs x 7 tasks [49].

The time and error collected data were analyzed using a qualitative and quantitative method such as Box-Plot and non parametric test of Wilcoxon respectively. An example of the results is shown in Fig. 2.11.

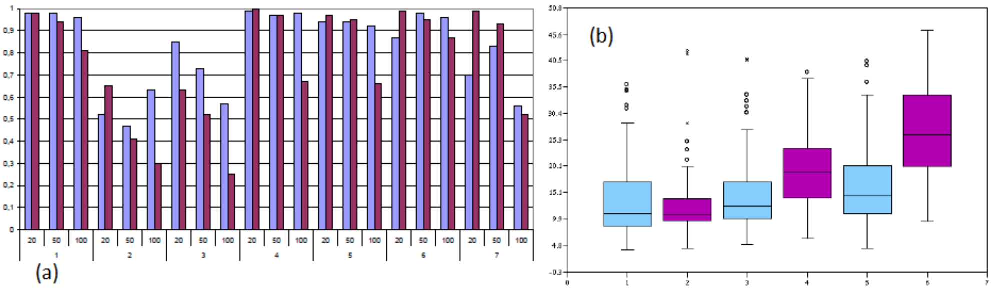


Figure 2.11: An example of results summary, (a) Percentage of correct answers. (b)Distribution of answer time. (adapted from Ghoniem et al. [49]).

Ghoniem et al. [49] showed that with a larger number of vertices $(V \geq 21)$, the matrix-based design outperforms node-link diagrams in several low-level reading tasks. Node-link diagrams perform better only on path finding. However, matrices are limited in solving some pattern-finding tasks specific to the set data. An additional separate matrix is used for exploring the intersection between two sets [6].

Wittenburg et al. [4] presented a method, BarExam, for visualizing set-valued attributes. It is an extension to bargrams [50] for depicting such attributes. The sets are depicted as rows in the bargrams. The sets are arranged from the largest set to the smallest set. The elements are represented on the horizontal dimension as bars. The elements are arranged according to their memberships to the corresponding sets, first topmost set then the second topmost set and so on. The bars are drawn according to this arrangement in each row as shown in Fig. 2.12 [4].

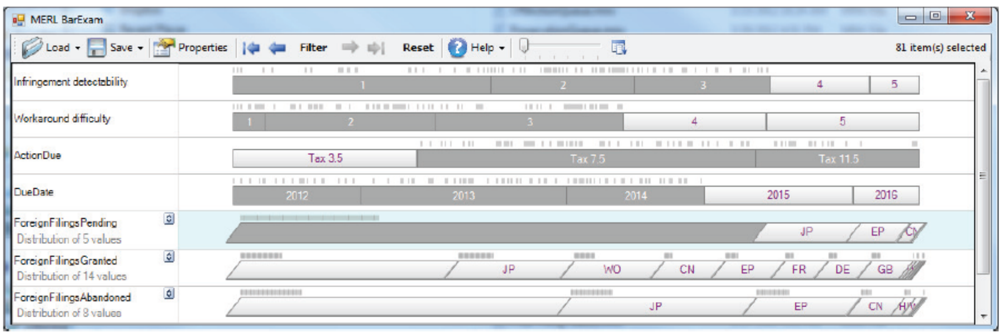


Figure 2.12: BarExam, Use case involving reducing maintenance fees in the management of a patent portfolio. (adapted from Wittenburg et al. [4]).

Such representation allows revealing various overlaps between the depicted sets. This can be used to partially solve the overlap-related tasks. The order of the elements is defined by the sets, which results in limitations with a large number of sets. For example, it becomes difficult to identify the overlap between the two bottommost sets, because the elements that belong to this overlap are scattered across different above bars [6].

Evaluation

An evaluation [4] of the new method and a general usability study of BarExam have been conducted. The evaluation aimed to infer usability and design feedback. The study included two parts, first exploration of the design regarding parallelograms vs. rectangles for set-valued attributes which included three qualitative questions. For example, one question investigates which of the two design variants is preferable and why.

The second part included two tasks to be solved using BarExam and three qualitative questions. For example, ”how likely would you be to use the BarExam tool in the future”.

Data set: A data set from a car models database has been used. The data set contained 200 items with 19 attributes, such as car model, price, warranty years, and color.

Users: 16 users with different age range and education levels participated to the evaluation. Their characteristics are shown in Table. 2.11.

Age Range	18-22 (12.5%), 23-26 (50%), 27-34(37.5%)
Sex	75% Male, 25% Female
Education	Computer Science Bachelor (37.5%), Telecommunication Bachelor (12.5%), MSc Computer Science (25%), PhD Candidate (25%)
Use of computer as main activity	More than 5 years (75%), 1 to 3 years (25%)
Data visualization experiences (courses/seminars)	No (50%), 1 course/seminar (25%), more than three courses/seminars (25%)

Table 2.11: Participants’ characteristics (adapted from Wittenburg et al. [4]).

Results: The questions have been answered by the users using a Likert scale from one to seven. The study was not expected to yield statistical significance, but rather produce usability and design feedback. The survey results have been reviewed and summarized as shown in Table. 2.12 [4].

Likert scale	Q4	Q5	Q6
Strongly agree	43.75%	50.00%	25.00%
More than agree	31.25%	43.75%	62.50%
Agree	25.00%	6.25%	6.25%
Not sure	0.00%	0.00%	6.25%
Disagree	0.00%	0.00%	0.00%
More than disagree	0.00%	0.00%	0.00%
Strongly disagree	0.00%	0.00%	0.00%

Table 2.12: Survey results for three questions (adapted from Wittenburg et al. [4]).

Techniques without user study

Several techniques have been proposed for visualizing overlapping sets and element-set memberships. Some of these techniques have been evaluated only by performing a case study. The effectiveness and user performance have not been assessed. This results in some open questions whether the proposed technique are effective or not, or whether these techniques outperform other techniques that support similar tasks or not.

The evaluation of a novel technique provides an evidence of its utility and effectiveness. In the following some techniques for visualizing overlapping sets are presented. A common factor shared between them is that no user study has been performed to evaluate the effectiveness of the technique. A brief description of each technique will be presented, followed by a general discussion concerning the evaluation.

Bubble Sets [34], is a technique used to visualize set relations over existing visualizations. It provides a continuous bounding contour, an implicit surface, for each set (Fig. 2.13a). This contour contains all elements in the respective sets. This maximizes the set membership inclusion and minimizes the inclusion of non-set members. Additionally, this can guarantee that all set members are included within one container but cannot guarantee the exclusion of non-set member. Two sets may overlap even if they do not share any elements. Such overlaps encode no information.

A case study has been presented to demonstrate the flexibility of bubble sets. The case study aimed to display set relations using isocontour surfaces over prefuse-based visualizations. The isocontour surface calculation and rendering was implemented in Java. The implementation has been used as an extension to a toolkit, called prefuse [51]. Also, Bubble set were demonstrated with a scatter plot. They have been used in a reimplementation of the GapMinder Trendalyzer [34, 52].

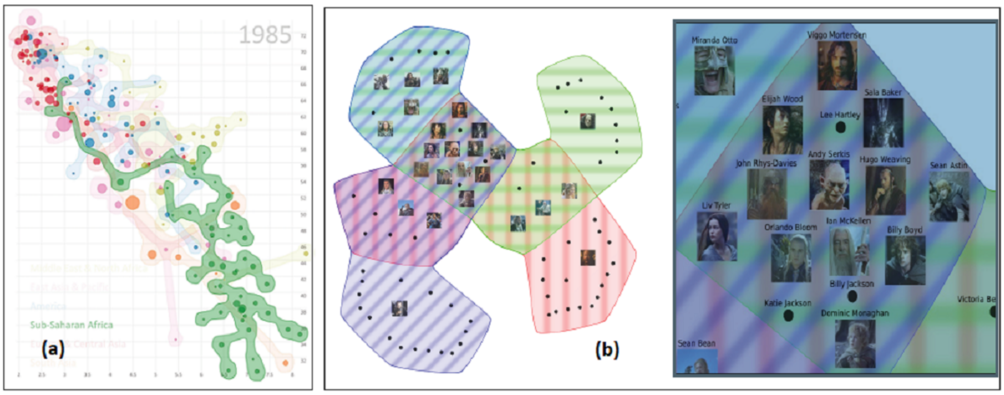


Figure 2.13: (a) Bubble Sets [34]. (b) An Euler diagram of IMDB movies [53].

However no evaluation method has been used, no pattern finding task has been proposed, and the effectiveness has not been assessed.

Flower et al. [54] proposed automated generation, in case of drawability, of any Euler diagram. They solved the problem of drawability by identifying the properties which classify a diagram as drawable or undrawable [26]. They used the concept of a (plane dual graph) of a concrete diagram. Spanning trees, the circularisation process, and addition of arcs take place resulting in all drawable diagrams with two or three contours [6, 54].

Additionally, the authors proposed an algorithm to solve the problem of drawability of any Euler diagram. They did not involve it as a part of a toolkit. A Java program has been written to implement the algorithm and sample output is generated. However, no evaluation was conducted on the readability or the effectiveness of the resulting diagrams.

Rodgers et al. [33] proposed a method that generate Euler diagram in the sense that any instance is drawable. The diagrams can be drawn by allowing disconnected regions and by minimizing some properties according to a chosen prioritization, such as permitting more than two curves to pass through a single point, permitting some curve segments to be drawn concurrently, and permitting duplication of curve labels [6, 33].

The method has not been evaluated. A software system has been used to implement the method. The authors illustrate the methodology only by generating the diagrams. The involved ideas have been demonstrated with output from the software system.

Simonetto et al. [53] presented a technique for the automatic generation of Euler-like diagrams. The algorithm generates an output even for undrawable instances of any collection of input sets. Bézier curves and transparent coloured textures have been used to improve the readability of the diagrams. They authors proposed that by using textures in addition to colour, it will be more efficient to represent the regions. They used $(c = 8)$ colours and textures to assure that no two overlapping sets will have the same colour and texture combination. To generate undrawable instances, the algorithm allows disconnected regions or allows to introduce holes in the regions as shown in Fig. 2.13b.

Simonetto et al. tested their approach on the internet movie database (IMDb), without performing any evaluation. They applied two examples on the date set by considering a subset of the films as sets. For each film, a set of actors are considered as the elements. However, no evaluation was conducted and no pattern finding tasks have been defined. Some patterns have been demonstrated, for example Katie Jackson makes cameo appearances in all three films.

Many approaches have been presented both for drawing and for visualizing bipartite graphs as node-link diagrams [6]. Misue [55] developed a technique for drawing bipartite graphs called **Anchored maps**. He assumed that the node set of a bipartite graph is divided into two sets, *anchor nodes* and *free nodes*. The anchor nodes are placed on a circle and the free nodes are placed at suitable positions according to the anchor nodes. Each free node is connected with links to the anchor nodes which it has edges with as shown in Fig. 2.14a.

Anchored Map can be used to represent a set system by depicting the sets as anchor nodes, and the elements as free nodes. Such representation enables determining which elements belong exclusively to which set, and which elements belong to multiple sets [6, 55].

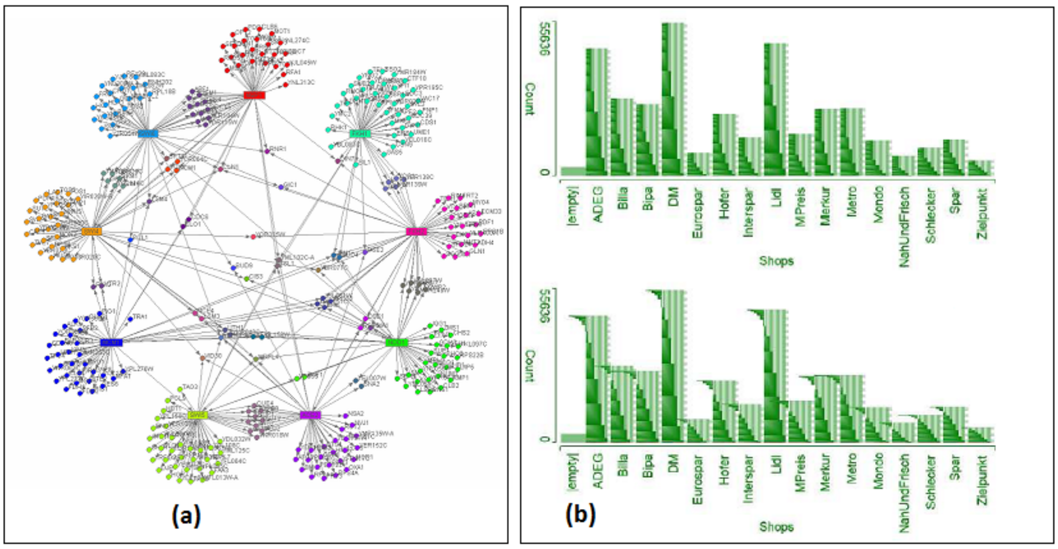


Figure 2.14: (a) Anchored maps [55]. (b) Set'o'grams [56].

The effectiveness of anchored map has been discussed with regard to the aesthetics of drawing results. Two kinds of diagrams have been generated. The two layouts have been compared in terms of aesthetics of drawing results. The algorithm has been implemented in Java. No evaluation method has been mentioned in the discussion. The scalability of the technique has not been measured.

Set'o'grams [56], have been presented by Freiler et al. as an interactive visual approach for analyzing and exploring set-typed data. They extend histograms for depicting overlaps between sets and for identifying additional relations between elements.

The sets are depicted as bars. The first bar represents the empty set. Each bar is divided into sub-bars. These sub-bars represent the degree of the elements that belong to the corresponding set. The degree of an element represents the number of sets that contain this element.

Starting from the bottom, the first sub-bar contains the elements that belong to only the respective set. The second sub-bar contains the elements that belong both to the respective set and to exactly one other set. The next sub-bar contains the elements that belong to the respective set and to two additional sets, and so on. The sub-bar width is varied in order to distinguish between consecutive sub-bars. The width of the sub-bar is reduced when the number of shared elements between the respective set and other sets increases as shown in Fig. 2.14b.

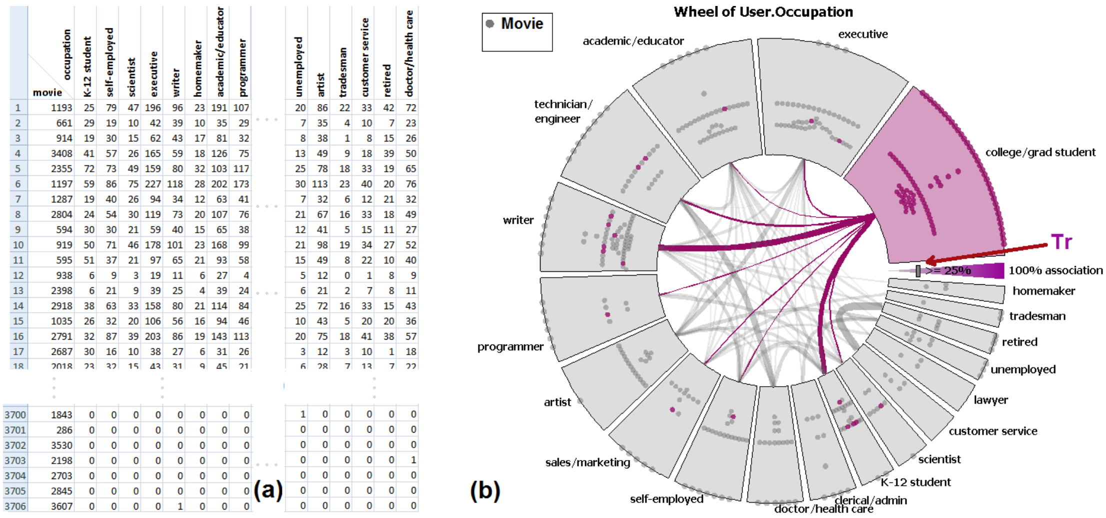
Set'o'gram has been used to demonstrate the usefulness of set-typed data without employing an evaluation methodology. It has been used to analyze a customer-relationship management (CRM) data set. A Set'o'gram has been generated followed by analyzing the group's data and discussing the patterns revealed. An example, is finding that a particular shop has the highest number of customers, but a very small amount of "exclusive" customers [56]. However, the usability issues and effectiveness have not been evaluated.

Related evaluation results

In this section, an evaluation of an interactive InfoVis method, called dot-based contingency wheel [57] will be presented. This technique has the same visual metaphor as Radial Sets.

The analysis of categorical data is usually based on contingency tables, which represent relationships between two or more categorical variables. However, when these tables become large and rich of information it might be complicated to extract information or to detect associations in the data. Therefore, visualization methods are used to provide insights in them. This makes analyzing and revealing such associations easier [12].

The dot-based contingency wheel has been designed to analyze positive associations in an asymmetrically large (n x m) contingency table [57]. The table columns are depicted as sectors forming a ring chart. The table's cells are depicted as dots. If a cell's row and column are positively associated, a dot is created in its column sector. To reduce the overlapping, the dots are distributed along the angular dimension. The radial positions of the dots are based on the associations. The dot is placed closer to the outer boundary if the association is high. The shared data between pairs of sectors are depicted as lines. The thickness of a line is based on the number of shared dots and on the associations these dots represent as shown in Fig. 2.15 [12].



	movie \ occupation	K-12 student	self-employed	scientist	executive	writer	homemaker	academic/educator	programmer	unemployed	artist	tradesman	customer service	retired	doctor/health care
1	1193	25	79	47	196	96	23	191	107	20	86	22	33	42	72
2	661	29	19	10	42	39	10	35	29	7	35	4	10	7	23
3	914	19	30	15	62	43	17	81	32	8	38	1	8	15	26
4	3408	41	57	26	165	59	18	126	75	13	49	9	18	39	50
5	2355	72	73	49	159	80	32	103	117	25	78	18	33	19	65
6	1197	59	86	75	227	118	28	202	173	30	113	23	40	20	76
7	1287	19	40	26	94	34	12	63	41	7	32	6	12	21	32
8	2804	24	54	30	119	73	20	107	76	21	67	16	33	18	49
9	594	30	30	21	59	40	15	65	38	12	41	5	15	11	27
10	919	50	71	46	178	101	23	168	99	21	98	19	34	27	52
11	595	51	37	21	97	65	21	93	58	15	49	8	22	10	40
12	938	6	9	3	19	11	6	27	4	5	12	0	1	8	9
13	2398	6	21	9	39	25	4	39	24	6	21	2	7	8	11
14	2918	38	63	33	158	80	21	114	84	25	72	16	33	15	43
15	1035	26	32	20	106	56	16	94	46	10	43	5	20	20	36
16	2791	32	87	39	203	86	19	143	113	20	75	18	41	38	57
17	2687	30	16	10	38	27	6	31	26	3	12	3	10	1	18
18	2018	23	32	15	43	31	9	45	21	6	28	7	13	7	22
3700	1843	0	0	0	0	0	0	0	0	1	0	0	0	0	0
3701	286	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3702	3530	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3703	2198	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3704	2703	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3705	2845	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3706	3607	0	0	0	0	1	0	0	0	0	0	0	0	0	0

Figure 2.15: (a) A large contingency table, (b) the corresponding dot-based Contingency Wheel. (adapted from Kriglstein et al. [12])

Evaluation

A qualitative evaluation to test the prototype has been conducted. The goal was to assess the clarity of the conceptual design, to find out the advantages and drawbacks of the representation, and how users interact with the new method. For the evaluation semi-structured interviews has been conducted [58, 59].

Users: Ten participants who studied computer science have been recruited. They were familiar with statistical methods. Five participants were visualization experts. Each session lasted 90 minutes.

Dataset: The dataset used for the evaluation was the answers of a standardized psychological test. From this dataset a 94 x 9 contingency table has been extracted. The rows represent the questions and the columns represent the 3 x 3 possible combinations of two answers on a question.

Tasks: The evaluation study was divided in four parts. It starts with an introduction about the dataset followed by a tutorial on the technique. The main study included the tasks to be solved by the users. The main focus of the tasks was the usability issues and to test if the visualization idea was clear. For example, to merge all sectors representing particular answer-combinations or to observe how the lines changed when the slider moves. After finishing the tasks, users have been asked for their impressions about the visualization [12].

The interviews have been recorded and the results have been analyzed. The methods of Bortz and Döring [60] have been used to compare the users' answers. Based on the results of the evaluation, a redesign of the dot-based Contingency Wheel has been introduced, called Contingency Wheel++ [61]. It simplifies the visualization by replacing the dots with histograms along the radial dimension [12].

2.3 Method

This thesis aims to provide an empirical evidence of the effectiveness of Radial Sets in performing the pattern-finding tasks mentioned in chapter3. Therefore, this work is based on the paper; Radial Sets: Interactive Visual Analysis of Large Overlapping Sets [6].

Some research resources for this chapter have mainly been explored through Radial Sets paper. Additional resources were found using Google scholar search engine, IEEE Digital Library, ACM library, and Vienna University of technology library. The following keywords have been used: Information Visualization, Overlapping sets, intersection between sets, and Information Visualization evaluation.

The resulted papers were divided into two parts according to the topic: Evaluation of InfoVis and visualizing overlapping sets. Some papers that were introducing new techniques but use different visual representation and support different tasks than Radial Sets were excluded. The focus was on the techniques that can be used to visualize overlapping sets. The techniques were divided into two groups according to the evaluated method: Techniques with user study and techniques without user study.

2.4 Discussion

This chapter provided an overview of visualization techniques that are related to Radial Sets. The techniques have been categorized into two groups according to the evaluation method.

- The first part presented a group of visualization techniques that have a user study. A description of each technique and the evaluation method were introduced. The evaluation was discussed according to the tasks it supports, the users, and the data set. The limitations of the experiment and the results have been presented.

- The second part presented the visualization techniques that don't have a user study. The techniques have either been evaluated by means of a case study or no evaluation method was introduced.
- In the third part an evaluation of dot-based contingency wheel [57] were presented. The evaluation method was introduced because this technique has the same visual metaphor as Radial Sets.

In the next section a summary of the presented techniques along with the scalability of each technique in terms of the number of sets and the elements they can handle will be presented.

Several techniques have been presented for visualizing different kinds of overlapping sets. Alsallakh et al. [3] provided an overview of such techniques. The techniques have been classified into 7 categories based on the visual representations they use and the tasks they support. The categories have been compared to provide guidance for choosing an appropriate technique for a given problem.

Finally, visualizing overlaps between sets is a challenging problem because of the exponential growth of possible overlaps between them. The presented methods use different visual representations for visualizing overlaps between sets. Some techniques are severely limited in the number of sets they can handle. Other techniques bypass this problem by limiting the number of sets and overlaps that can be visualized at once or avoid visualizing the overlaps explicitly [6].

2.5 Conclusion and Results

The evaluation of Information Visualization is very important to examine the effectiveness and the usability of a new visualization tool. Choosing the appropriate method for the evaluation depends on the purpose. Defining appropriate evaluation questions and methodology pose a challenge for evaluators to fulfill the objective of the evaluation. Moreover, selecting an appropriate data set to test or users and the right tasks is a nontrivial procedure [8].

A summary of some selected techniques for visualizing overlapping sets are shown in Table. 2.13. The techniques are categorized according to the evaluation method. The scalability of each technique is presented in terms of the number of sets and elements they can depict [3].

Finally, The aim of the study of the previous techniques was not to compare the performance of these techniques with Radial Sets. The goal was to survey a common evaluation standards that we need to address in our evaluation

Technique	With User study	Without User study	Scalability	
			Sets	Elements
ComED [32]	✓	-	10 to 20	Hundreds
DupED [32]	✓	-	About 10	Tens
Bubble Sets [34]	-	✓	About 10	Tens
LineSets [35]	✓	-	10 to 100	Hundreds
Kelp diagrams [36]	-	✓	About 10	Tens
ConSet [46]	✓	-	About 100	About 100
PixelLayer [62]	-	✓	Tens	Hundreds
Frequency grids [63]	✓	-	3 to 5	Hundreds
KMVQL [64]	✓	-	4 to 6	Not applicable
Mosaic displays [65]	✓	-	Up to 4 sets	Large (agg.)
Double-Decker [66]	-	✓	4 to 6	Large (agg)
Anchored maps [55]	-	✓	20 to 50	Hundreds
PivotPaths [67]	✓	-	50 to 100	Hundreds
Sets'o'grams [56]	-	✓	50 to 100	Large (agg.)
Radial Sets [6]	✓	-	20 to 30	Large (agg.)

Table 2.13: A summary of techniques for visualizing overlapping sets

CHAPTER 3

Radial Sets

In this chapter the Radial Sets [6] technique for visualizing overlapping sets is presented. This chapter is based on the article" Radial Sets: Interactive Visual Analysis of Large Overlapping Sets" [6, p. 2496–2505]. The visual metaphor, the used visual representation, and the interactive exploration environment will be introduced. The data, users, and the list of analysis tasks that Radial Sets supports will be discussed. Finally, the functionalities and features involved in the visualization technique will be described.

3.1 Introduction

Radial Sets (Fig. 3.3) is a new InfoVis technique for analyzing set memberships of large number of elements. It employs frequency-based representations that aggregate the elements in the sets and in the set's overlaps. The frequency-based representation is used to depict how the elements belong to the sets and how the sets overlap. This provides an easy and quick way to find and to analyze different kinds of overlaps between the sets. Furthermore Radial Sets supports relating the overlaps to the attributes of the elements, which results in enabling a scalable visualization of large and complex overlapping sets.

In addition, Radial Sets supports various interactions for selecting elements of interest. This facilitate finding out if the selected elements are over-represented in specific sets or overlaps, and detecting if the selected elements exhibit a different distribution for a specific attribute compared to the rest of the elements. Such interactions provide a useful method to formulate highly-expressive visual queries on the elements based on their set memberships and attribute values. For example, it is possible to query exclusive markers that belong to a specific gene.

3.2 Data, Users and Tasks

Radial Sets has been designed for the following data, users, and tasks:

- **Data:** Large overlapping sets (represented as memberships of n elements in m sets).
- **Users:** Data analyst domain experience.
- **Tasks:** Several pattern finding tasks in overlapping sets:
  - **T1:** Analyze the distribution of elements in each set according to their degrees (the number of sets they belong to).
  - **T2:** Find elements in a specific set that are exclusive to this set, or that belong at least, at most, or exactly to (k) other sets.

  These two tasks (T1, T2) are concerned with the element memberships in the sets. For example, for a product (as a set) it is possible to find the features that come exclusively with it or the features that are shared between multiple products.

  - **T3:** Analyze overlaps (intersections) between groups of k sets.
  - **T4:** Analyze overlaps between pairs of sets: find which pairs of sets exhibit higher overlap than other pairs (related to the previous task).
  - **T5:** Find elements that belong to a specific overlap.

  These tasks are concerned with the overlaps between the sets. One example is, finding out which marker combinations are shared between the genes.

  - **T6:** Analyze how an attribute of the elements correlates with their memberships to the sets and the overlaps.
  - **T7:** Analyze how set memberships and attribute values for a selected subset of elements differ from the rest of the elements.

  These tasks are concerned with attributes analysis. One example is to determine if the product's price depends on some features and if some features combinations increase or decrease it.

These seven pattern finding tasks are supported by Radial Sets. They are a selected subset of a more comprehensive list supported by different techniques [3]. Such tasks often arise when dealing with large overlapping sets. They have been addressed and proposed by many state-of-art methods [6, 32, 56, 68].

3.3 The Visual Metaphor

Radial Sets supports analyzing and discovering overlap patterns between large intersecting sets. To avoid the topological constraints of Euler diagrams, Radial Sets uses separate visual items for the sets and for the overlaps. As shown in Fig. 3.1 three kinds of visual elements are used [6]:

1. **Regions** to represent the sets.
2. **Histograms** inside the regions to represent the elements in each set.

3. **Links** between the regions to represent overlaps between the sets.

The sets are depicted as non overlapping regions with radial arrangement as shown in Fig. 3.1a. The overlaps are depicted as links between these regions (Fig. 3.1c). A thick link indicates a large overlap between the respective sets. Overlaps between three or more sets can also be depicted as links of higher order, which show the number of shared elements between the respective sets [6].

Radial Sets encode the overlaps using frequency-based representations of proportional size. These representations are used to depict the absolute or the normalized sizes of the overlaps. The set elements and the overlaps are visualized using area-based representations. This allows using colors to represent information about the elements. This is useful to support attribute-analysis tasks [6].

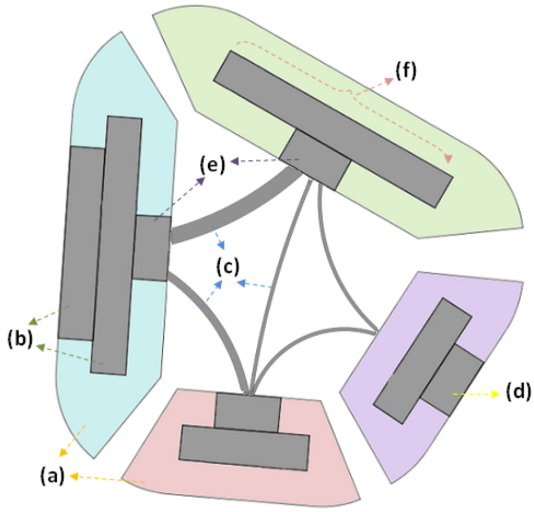


Figure 3.1: The visual items used in Radial Sets: (a) The regions, (b) The histograms, (c) The links, (d) The outermost histogram bar, (e) The innermost histogram bar, (f) The size of the group (adapted and simplified from Alsallakh et al. [6]).

The elements are depicted as histogram bars in the respective set regions they belong to as shown in Fig. 3.1b. The radial histograms encode the elements' degrees. The degree of an element represents the number of sets it belongs to. The elements are arranged in each set based on their degrees. The outermost histogram bar (Fig. 3.1d) contains elements that belong exclusively to the respective sets . The second histogram bar in each set contains the elements that belong to this set and one other set too. The innermost (Fig. 3.1e) histogram bar contains elements that are shared between as many sets as possible [6].

The size of a histogram bar (Fig. 3.1f) is proportional to the number of elements in it. Therefore, even if the elements are not depicted individually the distribution of the elements in each set by degree remains visible. This reveals which sets tend to have more exclusive elements and how many elements are unique in each set. This also exposes which sets tend to share elements with one or more other sets and how many elements are shared with one, two of more other sets [6].

The histogram bars can be colored according to an attribute of the elements they represent. Likewise the links can also be colored by an attribute of the overlaps they represent as shown in Fig. 3.9. This reveals how this attributes correlates with the sets membership. For example, we can easily find out which overlaps represent a higher disproportionality [6].

The elements aggregated in the bars or in the links can be explored in details using interaction. For example, by clicking on a link between two sets the elements contained in the respective overlap are listed for details on demand as shown later in the next section.

Finally, The frequency-based representations can depict either the absolute sizes or the relative sizes of the sets, elements and the overlaps. The absolute sizes represent the real sizes. For example, the absolute size of an overlap between two sets is 100 means that there are 100 elements shared between these two sets regardless the size of other overlaps. The relative sizes make it easier to compare the overlaps between sets that have different sizes. This can be done by representing the portions of the respective sets the overlaps represent as shown in Fig. 3.2.

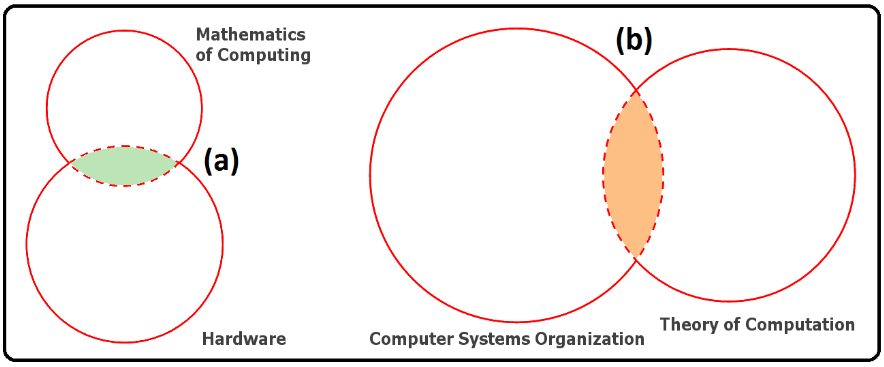


Figure 3.2: Two overlaps of 2nd-degree, having different absolute sizes, but nearly equal relative sizes (adapted from Alsallakh et al. [6])

3.4 The Interactive Exploration Environment

The interactive exploration environment allows analyzing and revealing information at different levels of detail. The user interface consists of coordinated and multiple views that enable users to formulate highly-expressive and visually-guided queries on the sets, overlaps, and elements. The query results can be analyzed in details through these views [6].

The user interface is composed of the following views, as shown in Fig. 3.3:

- **The Radial Sets view**

  This view (Fig. 3.3c) is the central part of the interface. The other views show more summarized or more detailed information about the sets, the elements, and the overlaps. The Radial Sets view provides an overview of the sets, the distribution of the elements in the sets, and how the sets overlap [6].

  Users can extract more details about the elements and the overlaps on demand by using the detail views. The tooltips also can be used by moving the mouse pointer over a visual item to obtain more information about it. The visual item can be a set, a histogram bar in a set (subset) or an overlap between two or more sets.

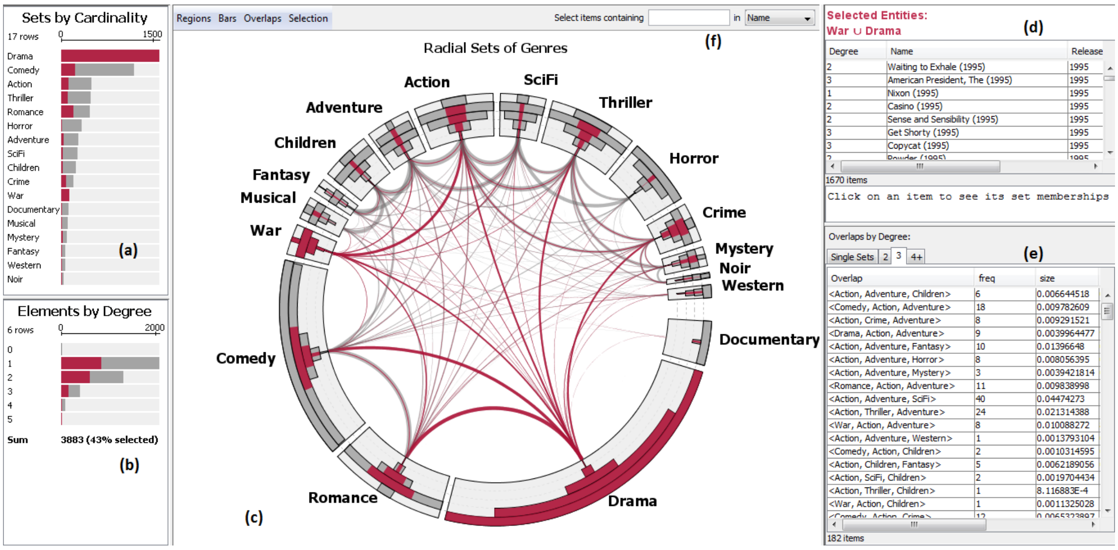


Figure 3.3: Radial Sets: (a) The sets bar chart, (b) The degree histogram, (c) The Radial Sets view (d) The selection view (e) The overlap analysis view, (f) A search box to select elements containing a specific text.

  The tooltips contain information such as:

  - A description of the set, subset as shown in Fig. 3.4 or of the overlap as shown in Fig. 3.5.
  - Information about the elements in the respective set, subset or overlap. Information such as, whether the elements are exclusive to the set or shared with one or more sets.
  - The absolute and the relative sizes of the set, overlap or of a selected portion in a set.
  - The dis-proportionality of the elements in the set or in the overlap.

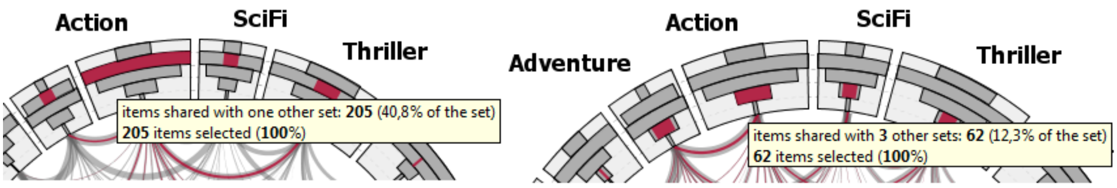


Figure 3.4: Tooltips showing various information about the sets or subsets represented by the regions and the bars.

In addition, the Radial Sets view offers several functionalities to manipulate the sets, for example, to merge the sets or to change the order of the sets by using drag and drop, or to merge two sets and replace them by their union. The menu bar in the top of the view is used to modify and to configure the order of the sets. The commands in this bar are used to colour the bars and links, and to specify the histogram scaling, and the size of the overlaps (absolute / relative). The selection commands are used to manipulate a subset of selected elements in the sets [6].

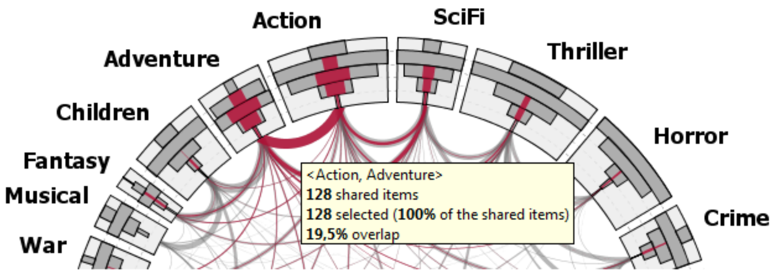


Figure 3.5: Tooltips showing various information about an overlap between two sets represented by the links.

- **The Summary views**

  The summary views (Fig. 3.3a, b) show summary information about:

  - The sets. For example, the number of elements in each set
  - The elements. For example, the degree of the elements contained in the sets.

  Two views are used to show the summary information:

  - **The sets bar chart:** The sets are ordered by their cardinalities. This view depicts the set sizes in descending order, along with the selected portions of these sets as shown in Fig. 3.3a.
  - **The degree histogram:** The elements are grouped according to their degrees as shown in Fig. 3.3b.

In addition to provide summary information, the summary views are used to define which sets are depicted in the Radial-Sets view. This can be performed by using the (show/ hide) functionality. By right clicking on a set in the sets bar chart, a pop-up menu will appear. This menu includes the (show/ hide) functionality (Fig. 3.6a). Based on the selected function, a set will be included/ excluded from the Radial Sets view respectively [6].

Furthermore, the summary views are used to define which elements to incorporate in the computations. This can be performed by using the include/ exclude functionality. By right clicking on a degree-group in the degree histogram, a pop-up menu containing four options will appear as shown in Fig. 3.6b. The options are: exclude, exclude all selected, include only this, and include all but this. Based on the selected options the receptive group of elements will be processed [6].

Finally, both views can be used to gain an overview on the elements under selection, and to define or to refine a specific selection.

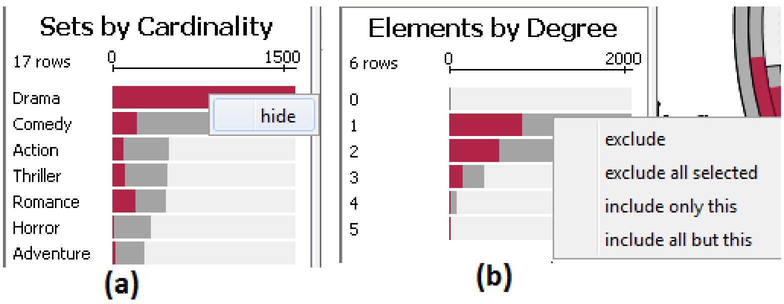


Figure 3.6: (a) The show/hide menu to define the depicted sets, (b) The include/exclude menu to define the involved elements in the computations.

- **The Selection view**

This view provides detailed information about selected elements (Fig. 3.3d). An expression that externalizes how the selection was defined is shown at the top on this view. This expression uses the common set-theory notation. Additional extensions are used to express the conditions related to the degrees of the elements and the values of the attributes. This view uses a tabular list of the elements contained in the selection. The values of the attributes are listed along with the respective elements in this tabular list. The list can be sorted according to an element's attribute [6].

An additional view can be used to analyze and explore the attributes in more details as shown in Fig. 3.7. The set memberships for a specific element can be examined by clicking on the element in the tabular list. The selected element will be highlighted.

An element's set memberships can be indicated in two ways:

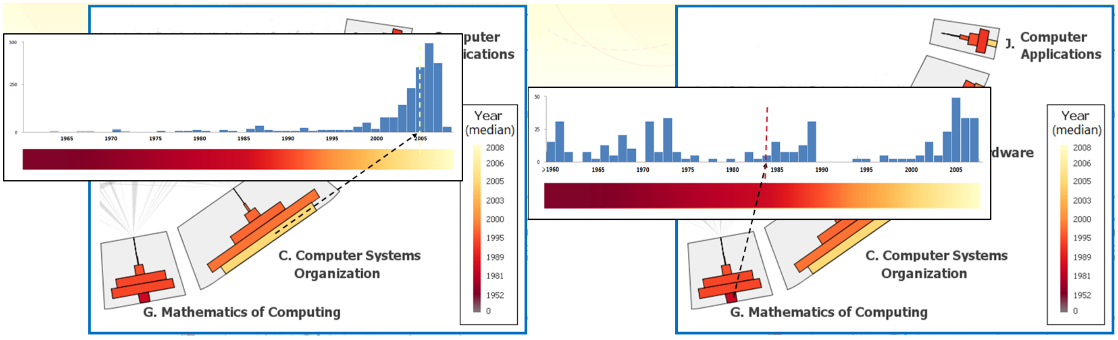


Figure 3.7: A linked view showing more details according to the median published date of the ACM papers.

- **Graphically:** As a star graph in the Radial-Sets view. The graph exhibits to which sets and to which bars in these sets the highlighted element belongs as shown in Fig. 3.8a.
- **In text:** As a comma-separated list of the element's set memberships, e.g., highlighted item belongs to k sets: Set(1), Set(2),..., Set(k). The text is shown at the bottom of the selection view as shown in Fig. 3.8b.

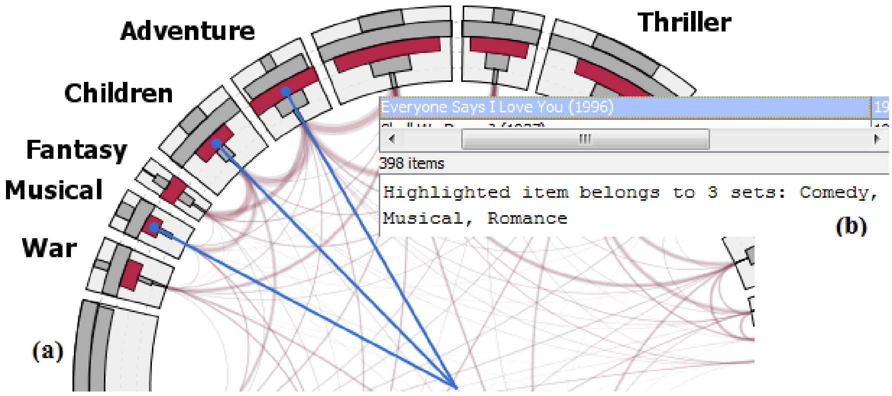


Figure 3.8: Indicating element's set memberships. (a) Graphically, (b) In text

Finally, the selection view provides detailed information about specific elements. Additionally the interactive selection allows users to filter and manipulate the data. It offers several functionalities to hide or to exclude specific elements form the analysis. These functionalities are performed based on the elements' attributes, set memberships, and degrees. For example, some data sets contain skewed distributions of set sizes (e.g., few sets contain the majority of the elements) or of element degrees (e.g., most elements are exclusive in their sets). By filtering out such portions users can discover more information about the rest of the data [6].

- **The overlap analysis view**

  This view (Fig. 3.3e) is used to analyze and to compare the overlaps between the sets in more details. The overlaps are shown in tabular lists. Detailed information about the 2-sets, 3-set, and k-sets overlaps are listed along with the respective overlapping sets. This includes, for example, the size of the overlapping sets [6].

  The Radial-Sets view is updated when an overlap in the lists is chosen. A new visual item is presented to define the involved sets and the size of the overlap as shown in Fig. 3.9. Also, the overlap analysis view is updated when the selection in the Radial-Sets view changes.

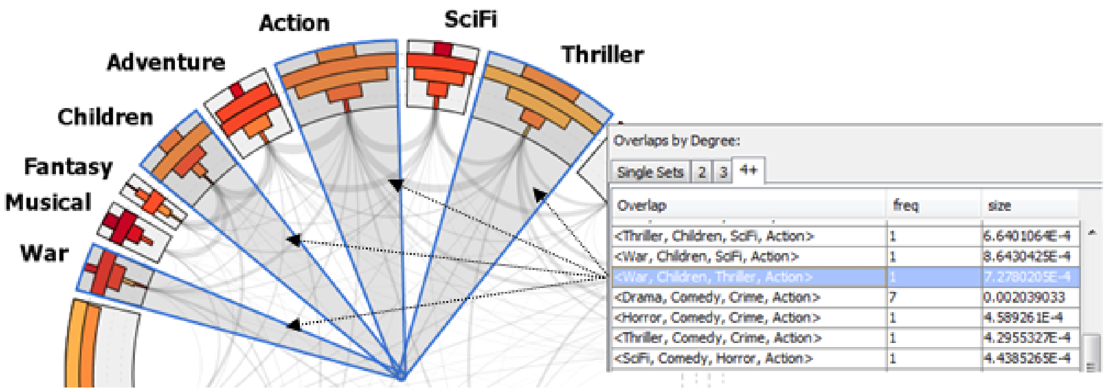


Figure 3.9: The Radial-Sets view is updated according to the selected overlap from the tabular lists.

In the Radial-Sets view the arcs and the bubbles depict the overlaps. The overlaps between pairs of sets (e.g., overlaps of degree 2) are depicted as arcs between the respective regions. The thickness of an arc encodes the absolute/normalized size of the overlap. The overlaps between more than two sets (e.g., overlaps of degree more than 2) are depicted as bubbles. A bubble is created in the inner area of the Radial Sets. The size of the bubble is proportional to the size of the overlap. The bubble is connected with the respective sets via arrow heads. The bubble along with the arrow heads form a hyperedge that denotes to the overlapping sets [6].

A ”bubble chart” of the overlaps can be presented by showing only the bubbles of the hyperedges. By clicking on a bubble the links to the sets involved in the corresponding overlap are revealed. The bubbles can be scaled either by using the same scaling factor as the histograms which presents an overlap in proportion of the involved sets, or to fit in the inner area which supports the interaction with the bubbles and to compare the bubbles’ sizes.

The arcs and bubbles offer an overview of the existing overlaps and the sets involved in them. They facilitate selecting a specific overlap, and are useful to analyze overlap patterns [6].

3.5 Functionalities and Features

The Radial Sets view along with the summary views allow defining several subsets of elements in the sets. To define a selection over the elements, it is possible to brush these subsets. This selection can be specified using set operations such as union and intersection as shown in Fig. 3.10. A variety of combinations of subsets can be created by means of set operations.

The selection possibilities enable selecting the elements by their set memberships and degrees. The selection is specified iteratively, which results in updating the selected items presented in the Radial Sets view and in the summary views during the selection. This provides an immediate feedback and a guidance on how to refine the selection [6].

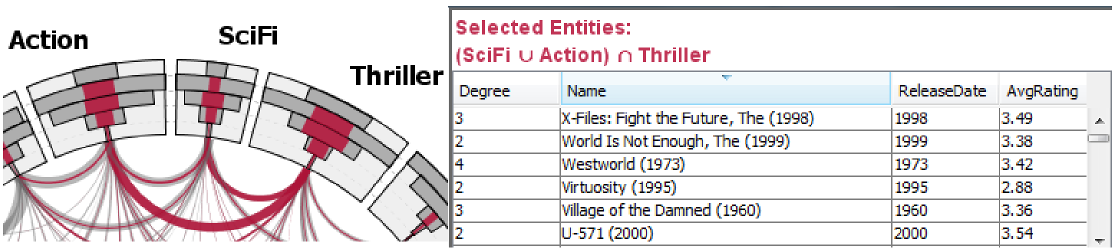


Figure 3.10: Using the set operations to define a subset of elements in Radial Sets.

The selection can also be defined based on the element's attribute values. This can be performed via textual search in the attribute values (Fig. 3.3f), or via coordinated views that allow brushing the elements that have certain attribute values.

Brushing the elements in Radial Sets can be performed in two ways:

1. Clicking on the individual bars in the set region

2. Dragging the mouse to define a range over the bars as shown in Fig. 3.11. The same interactions can be performed with the bars in the summary views. The selection is set to the brushed elements if no set operation is activated during brushing. The set operation can be activated via keyboard modifiers [6].

The keyboard modifiers are used to perform the following set operations:

- **Set union:** To add the brushed elements to the existing selection. This operation can be performed using the SHIFT button.

- **Set intersection:** To specify if the brushed elements should be intersected with the existing selection. This can be done clicking the CTRL button between the selections of two or more sets.

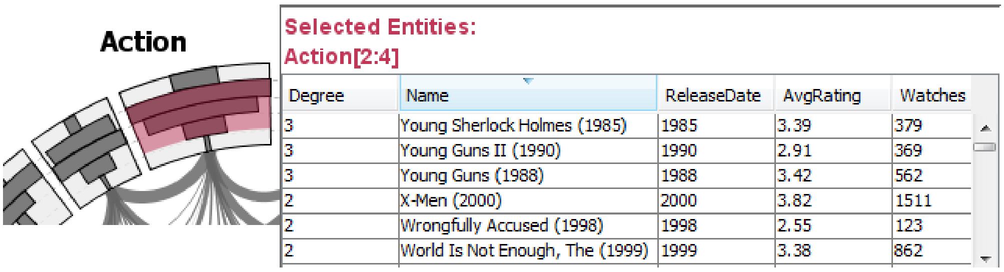


Figure 3.11: Brushing the elements in Radial Sets by dragging the mouse over the bars.

- **Set difference:** To subtract the brushed elements to the existing selection. By clicking the ALT button, the second set will be subtracted from the first set. It is possible to subtract multiple sets from one set. The union of the subsequent sets will be subtracted from the former set. The expression presented in the summary view will be like, $(Set1/Set2 \cup Set3)$ as shown in Fig. 3.12.

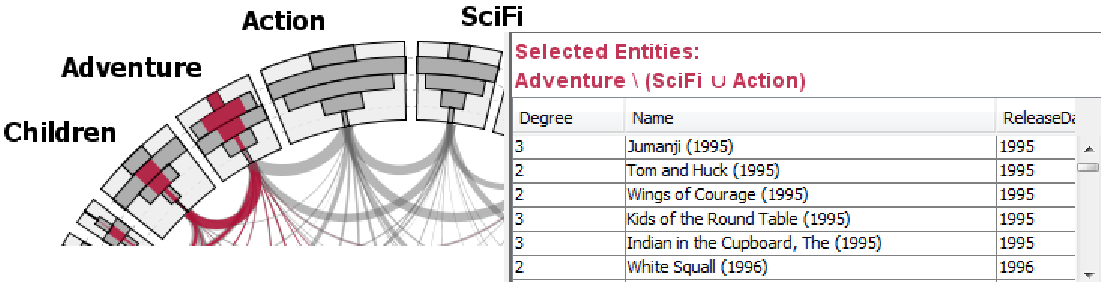


Figure 3.12: Set difference, subtraction in Radial Sets.

The overlaps and the histogram bars depict a subset of elements. The size of this subset is encoded by thickness or area. Detailed information about the elements in a certain subset can be revealed by coloring the bars or the areas.

Performing a selection operation over a subset of elements highlights the selected portions. Users can also define which information to present via colors. This can be performed by selecting an attribute of the elements as source of the coloring. For example, in Fig. 3.13 color represents the median publication date of ACM papers.

An overview of the distribution of the attribute's values in the subsets is presented. This attribute along with the elements' membership support differentiating the sets and the overlaps [6]. For example, in Fig. 3.13 it is easy to detect that papers which are exclusively mathematics of computing have relatively old publishing date on average. On the other hand, papers which are mathematics of computing and computer system are recent.

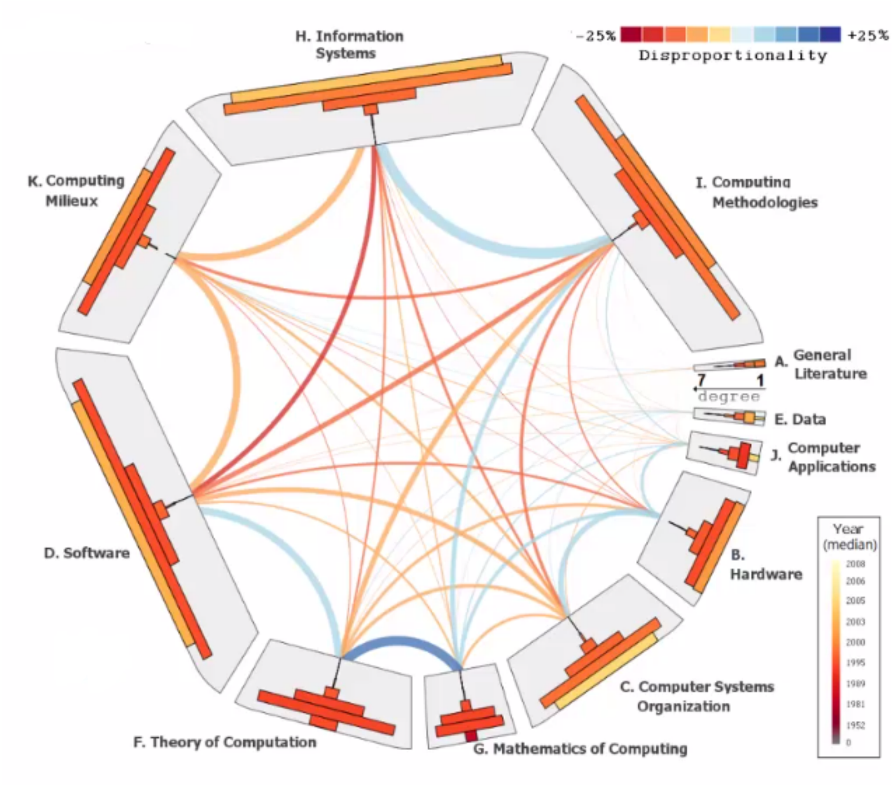


Figure 3.13: Radial Sets depicting ACM papers according to their genres, the bars and the arcs are colored according to the median published date of the papers [6].

To analyze the exclusiveness of the overlaps their visual items can be coloured. This items are colored by the average degree of the overlaps' elements. The overlaps that are more exclusive are colored corresponding to the values which are closer to their degrees. The exclusiveness can also be analyzed via interaction. For example, analyzing the exclusiveness of 4-degree overlaps will be performed by selecting the 4-degree elements (e.g., elements that belong to 4 sets) [6].

The bubbles and the analysis of overlap exclusiveness were not included in the evaluation of Radial Sets. This is because the evaluation is focused on the most important features and functions of the new technique, and also because we wanted to limit the sessions time.

CHAPTER 4

Evaluation

This chapter describes the design of the evaluation in details, which is the main part of this thesis. First, the hypotheses will be introduced. Then, the evaluation method will be described. An overview on the tasks, users, and the used data set will also be presented. In the last section of this chapter the tool used to collect tasks completion time and error data; EvalBench [11] will be introduced.

4.1 Introduction

For the evaluation of Radial Sets a quantitative study, known also as laboratory experiment has been conducted. The plan was to perform the following evaluations:

- **Internal evaluation:** As an empirical evidence for the effectiveness of the new visual technique in performing the pattern-finding tasks mentioned in chapter. 3.
- **External evaluation:** As a comparative user study to compare the Radial Sets technique against Set'o'grams [56]. Sets'o'grams has been selected since it is one of the few state-of-art methods that can handle large number of elements. And because it supports similar tasks to the one mentioned in chapter. 3.

The goal of the evaluation was to detect the pros and the cons of the selected visualization techniques in solving the pattern-finding tasks (see Chapter. 3). But due to some technical problems with Sets'o'grams it was not possible to conduct the comparison. Therefore, the first part was the main part for the experiment, which evaluate Radial Sets performance for each task.

The internal evaluation has been conducted focusing on how well Radial Sets perform the tasks. This includes measuring the time and error made by the users when they are solving the tasks. In addition, the new technique has been qualitatively evaluated in order to assess the clarity of the conceptual design and to elicit usability issues.

4.2 Hypotheses

My assumption is that Radial Sets is effective for dealing with large overlapping sets and in solving the pattern finding tasks mentioned in chapter. 3. The tasks that involve analyzing and comparing the sets, elements or overlaps can be solved fast with a high correctness. Based on this assumption I created three hypotheses.

The first hypothesis is focused on the performance of the visualization technique when investigating the element memberships in the sets, for example, the number of sets an element belongs to. The second hypothesis refers to the capability of the visualization to identify and explore the overlaps between the sets, for example, the number of elements in a specific overlap. The third hypothesis is concerned with the attributes analyzing, for example, to define the sets that contain a group of elements that tend to have the same attribute values.

The evaluation was based on the recorded times and the correctness of the tasks. These data have been recorded using EvalBench [11] as explained in details later in this chapter.

- **Hypothesis 1:** Radial Sets enable to quickly analyze the distribution of elements in the sets, exploring the elements in each set according to their degrees, and determining the exclusive or the shared elements in the sets. This hypothesis is related to the first and second tasks mentioned in chapter. 3.

- **Hypothesis 2:** Radial Sets support revealing and facilitate analyzing the overlaps between large sets. This include determining the sets that tend to have high or low overlaps and exposing the elements in the overlaps. This hypothesis is based on the third, fourth, and fifth tasks mentioned in chapter. 3.

- **Hypothesis 3:** Radial Sets enable analyzing the elements' attributes and revealing how they correlates with the elements' set memberships or overlaps. This hypothesis is related to the sixth and the seventh tasks mentioned in the chapter. 3.

4.3 Method

At the beginning, the participants received a short introduction about set-typed data and how such type of data can be represented. Then an introduction to the visualization technique and how set-typed data can be depicted using it (the visual metaphor) was presented.

The introduction was followed by a tutorial about Radial Sets to give the users the chance to get acquainted with it. The tutorial provided users with a guidance showing how the main functions work, how to deal with the tool, and how to obtain information using it.

Additionally, the participants got an overview on the evaluation process. The types and the answer modality of the tasks that will be solved by the users were described (e.g., multiple-choice, determine or define, Likert scale). Instructions on how to use EvalBench and how to proceed form a task to the next one were also presented.

Finally, the participants answer a questionnaire about their experiences with Information Visualization. The questionnaires also included questions about participants' demographic data. The study design and the process will be described in details in the next section.

4.4 Design

The evaluation was set up based on the following aspects (see Table. 4.1):

- **The visualization technique:** Radial Sets.
- **The task type:** This defines the kind of the task (analyze, compare, determine, find).
- **The answer modality:** This characterizes the type of the task's answer (single value, multiple values, text field, behavior).
- **Time:** The completion time needed to solve each task in milliseconds.
- **Error:** The correctness of each task. A correct answer is rated by 1 while a wrong or a missing answer is rated by 0.

Data	**Description**
Visualization technique	Radial Sets
Task type	Analyze, Compare, Determine, Find
Task category	Single value, Multiple values, Text field, Behavior
Task time	In Milliseconds
Task error	1 for correct, 0 for wrong

Table 4.1: A summary of the data used in the experiment.

Users

All recruited participants for this user study had reasonable computer experience. This was because the tasks require interaction with the visualization tool (search, to apply the set operations, drag and drop). 13 users had some basic knowledge about Information Visualization techniques, the others had no experience with it.

The knowledge about sets and the basic operations on sets was necessary. To match this requirements university students with different backgrounds were recruited. The students' backgrounds ranged between computer science, economics, law, and engineering (see Table. 4.3).

A pilot evaluation with 8 users has been conducted before the experiment as shown in Table. 4.2 and Table. 4.3. The results of this study was excluded from the evaluation. The aim of the pilot study was to test the evaluation design and to control the sessions time. Users' preference and comments from the pilot study have been used to improve the evaluation design and the procedure. For example, in the pilot study the users complained about the number of tasks (73) and the session time (90m), which have been reduced in the actual experiment.

	Users	**Male**	**Female**	**Color-blind**	**Age range**
Pilot study	8	6 (75%)	2 (25%)	0	21- 45
Experiment	32	21 (65.6%)	11 (34.4%)	0	20- 32

Table 4.2: Summary of the participants' characteristics by age and gender.

32 students participated in the experiment, 21 male and 11 female persons. The age of these participants ranged between 20 and 32, the average age was (24,78). All users had normal or corrected-to-normal vision, none was color-blind. The users have been classified according to age and gender as shown in Table. 4.2. Table. 4.3 classifies them according to their background.

	Computer science		**Engineering**		**Economics**		**Law**		2* **Total**
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	
Pilot study	4	1	2	1	-	-	-	-	8
Experiment	11	6	4	3	4	2	2	-	32

Table 4.3: Summary of the participants' characteristics by backgrounds.

In summary, the preconditions for recruiting the participants were

- Reasonable computer experience.
- Knowledge about sets and the basic operations on sets.
- Normal or corrected-to-normal vision, no color-blindness.

Apparatus

A laptop has been used for the evaluation. It was Lenovo Ideapad with Windows7 (32-bit) as operating system, Intel centrino2 (1.3GHz) as processor and 2GB RAM. An external LG Monitor LCD (19 inch/ 48,3 cm) has been connected with the laptop and used for better resolution.

The users used both a mouse and the keyboard to solve the tasks. The mouse was a standard HP optical mouse. All participants performed the experiment using the same laptop, monitor, and mouse.

Content

The materials used for the experiment comprise of a questionnaire, a video, an introductory presentation, the training tasks, and the evaluation tasks. All materials, except the video, are listed in appendix. A and appendix. C.

The aim of the questionnaire was to collect demographic data about the users. It included questions about age, gender, occupation, and sight disorder. In addition, It involved a part asking the participants if they have experience with Information Visualization. In case a user has an experience, he/she was asked to estimate his/her knowledge with InfoVis. The estimation is classified into three levels: beginner, intermediate, and advanced.

The video was used to introduce set-typed data and how they can be represented. The users first got an illustration how the sets, elements and overlaps can be depicted using Euler diagrams. Thereafter an explanation about Radial Set's metaphor and which visual items it uses to represent the sets, elements, and overlaps was presented.

The introductory presentation introduced the main functions and features of the tool to the users. Coloring the bars or the arcs according to a selected attribute and how to employ such function to extract information was explained. The set operations and how to apply them by interacting with the tool and by using the keyboard modifiers were presented.

The video used in the experiment can be found on internet[1]. The training tasks and the evaluation tasks will be discussed later in details in this chapter.

Dataset

The tasks were defined over movies data. The data come from the MovieLens database. The data set used for the experiment comprises 3883 movies that are produced between 1919 and 2000. It contains various of information about movies (see Table. 4.4).

- **The movies genres:** The style of the movie or subject matter. 17 movie genres have been defined: Adventure, Action, Children, Comedy, Crime, Documentary, Drama, Fantasy, Horror, Musical, Mystery, Noir, Romance, Sci-Fi, Thriller, War, and Western. A movie can have multiple genres.
- **The release dates:** Refers to the date on which a movie was made available to watch for public. The release dates of the movies used for the evaluation ranged between 1919 and 2000.

[1] http://www.radialsets.org/. last accessed on the 12th of September 2014

- **Average rating:** Represents the rating of the movies by audience. A movie rate ranges on a scale from 1 to 5, with 1 as the lowest rate and 5 as the highest rate.

- **Number of watches:** The number of audience who watched the movie and rated it.

	Movies genres	**Movies**	**Release date**	**Average rating**	**Number of watches**
Value range	17	3883	1919- 2000	1- 5	1-3428

Table 4.4: The main attributes of the MovieLens data set used in the experiment.

Each genre defines a set over the movies and is represented by a set region. Movies, the elements, which belong to a genre are represented by histograms in the corresponding region. Genres can overlap since one movie can belong to more than one genre. No constraints have been applied on the number of sets, elements, and overlaps in the experiment. Table. 4.5 lists the values that have been used for the evaluation.

	Sets	**Elements**	**2-set intersections**	**3-set intersections**	**4-set intersections**	**5-set intersections**
Experiment	17	3883	104	182	88	7

Table 4.5: The number of the set, elements, and set intersections in the experiment.

Such real world data were selected because users can easily become familiar with it. Also the background of the data should be easy to understand for non experts in the movies industry.

Tasks

The tasks used in the experiment were classified into two groups, training and testing. Each group contained a variety of questions covering a certain topic. The main goal of the experiment was to evaluate Radial Sets performance for each of the seven pattern finding tasks (chapter. 3). Therefore, the questions were derived as instances of them.

The total number of the tasks were 60 tasks. Table. 4.6 shows the number of tasks in each group. A detailed description of the tasks are listed in the appendix. A.

	# Training questions	**# Evaluation questions**	**Total number of tasks**
Experiment	7	53	60

Table 4.6: The number of questions used in each group of tasks for the evaluation.

A task type and a task category have been assigned to each evaluation question. The task type defines the kind of the task and the task category characterizes the type of the task's answer.

Task type

The task type has been defined according to the tasks supported by Radial Sets. This was because of the lack of a task taxonomy related to overlapping sets, a recent survey [3] is doing the first steps toward such one. Moreover, the analytic task taxonomy introduced by Amar et al. [69] and Andrienko et al. [70] have also been used. The task type used in the experiment has been defined (see Table. 4.7) as following:

- **Analyze:** The analyze tasks were used to expose the sets, subset of elements or overlaps. For example, when a user has to expose an overlap between two movie genres to infer if it contains old or recent movies.
- **Compare:** The compare tasks were used in order to compare multiple sets or overlaps. For example, in some tasks users have to compare the overlaps to detect which two sets have the highest overlap.
- **Determine:** The determine tasks aimed to specify the size of the sets and overlaps. For example, to specify how many movies an overlap contains.
- **Find:** The find tasks were used to search for sets or elements that are contained in a set or in an overlap. For example, to find the genres a movie belongs to.

Answer modality

This has been defined according to the interface the users used to answer the questions. Four answer modalities have been defined (see Table. 4.7) as following:

- **Single value:** This mode is used for tasks that have only one value as an answer, for example, Asking about how many movies does a certain genre contain.
- **Multiple values:** This mode is used for tasks that have two or more answers, for example, to name the genres to which a movie belongs to.
- **Behavior:** This mode is used for tasks that tend to have a scale of values as an answer, for example, to identify if a subset of movies has a high, medium or low average rating.
- **Text field:** This mode is used when users have to use a text box to solve the task, for example, to name a movie that belongs to an overlap.

	Task type				Task category			
	Analyze	**Compare**	**Determine**	**Find**	**Single value**	**Multiple values**	**Behavior**	**Text field**
Evaluation questions	8	11	19	15	27	7	7	12
Sum of tasks	53				53			

Table 4.7: The task types and answer modality for the evaluation questions.

The tasks have been classified into two groups as following:

- **Training questions:** Tasks for the participants to explore and get acquainted with the tool.
- **Evaluation questions:** To evaluate the effectiveness of Radial Sets in performing the pattern-finding tasks (see chapter. 3).

The training questions

The main goal of these questions was that users can get started exploring the tool and get familiar with the visualization. The questions were also derived, same as the evaluation questions, as instances of the seven pattern-finding tasks (see chapter. 3). The main difference between this questions and the evaluation questions was that the users got instructions and hints to accomplish the tasks. Furthermore, they were allowed to ask questions about performing a certain task or using a function when they were solving this group of tasks. The collected data regarding the training tasks was excluded from the evaluation results. Table. 4.8 shows a description of this group of questions.

Nr.	Task Description
1	Click on the “Horror” region.
2	Click on the top bar in the “Horror” region.
3	Click on the arc that connects the “Comedy” region and the “Romance”.
4	Click on the thickest arc.
5	Move the mouse pointer to the top bar in the” Drama” region. - Notice the number of the items in the most top bar.
6	Click on the “Drama” region, notice: - The number of the exclusive movies - The number of the shared movies with one another region
7	With how many genres does the “Documentary” genre overlap? - Name them?

Table 4.8: The Training questions.

The evaluation questions

This group of questions focused on dealing with large overlapping sets. They covered investigating the element memberships in the sets, exploring the overlaps between the sets, and analyzing the attributes. The evaluation questions have been defined as instances of the seven pattern-finding tasks (see chapter. 3) to provide an evidence of the defined hypotheses.

Each task addresses one of the three hypotheses (see section. 4.2). These tasks have been further divided into three groups according to the level of difficulty: easy, intermediate or hard questions.

The criteria for determining a level of difficulty for each question were defined in a similar manner as in the task taxonomy of Brehmer et al. [71] as follows:

- A question is defined as an easy level of difficulty question if it requires a query on the data base that involves one or no set operation and at most one step of de-aggregation.
- A question is of intermediate level of difficulty if it requires a query on the data base that involves two or three set operations on the data.
- A question is defined as a hard level of difficulty question if it requires a query on the data base that can be performed with four or more set operations on the data.

Table. 4.9 shows the distribution of the evaluation questions to the levels of difficulty and hypotheses along with the respective pattern finding tasks.

Hypotheses	**Related pattern-finding tasks**	**Level of difficulty**			**Sum tasks**
		Easy	Intermediate	Hard	
H1	T1- T2	7	6	5	18
H2	T3- T4- T5	6	10	6	22
H3	T6- T7	3	6	4	13
Sum tasks		16	22	15	53

Table 4.9: The distribution of the evaluation questions to levels of difficulty and hypotheses.

Hypothesis 1 deals with elements-set memberships. Therefore, the first group of the evaluation questions is concerned with analyzing the distribution of elements in the sets and specifying the exclusive and the shared elements in each set. Table. 4.10 shows the first group of the evaluation questions along with the respective pattern-finding tasks, level of difficulty, task type, and task category.

Hypothesis 2 focuses on the overlaps between the sets. Therefore, the second group of the evaluation questions is concerned with exposing the overlaps and defining the elements that belong to them as shown in Table. 4.11.

Hypothesis 3 deals with the elements attributes. Consequently, the third group of the evaluation questions covers analyzing how these attributes correlate with the elements-set memberships. Table. 4.12 presents the evaluation tasks related to the third hypothesis.

Some questions are divided into two tasks (e.g., questions 1 and 2). For one task a user has to define the number of the elements in the sets and for the second task the user has to specify one element. This aims to ensure that the users understand the task and are able to interact easily with the tool in order to extract more detailed information on demand.

Hypothesis 1					
Nr.	**Type**	**Category**	**Task**	**Difficulty**	**Description**
1	D	S	T2	Easy	How many movies does the genre Action contain?
2	F	T	T2	Easy	Name one of them (Action's movies)?
3	D	S	T2	Easy	How many movies come exclusively with the Action genre?
4	F	T	T2	Easy	Name one of them (exclusive Action's movies)?
11	D	S	T1	Easy	How many genres does the movie Bad Boys belong to?
12	F	M	T1	Easy	Name the genre/s (the movie Bad Boys belongs to)?
13	D	S	T1	Intermediate	What is the highest degree of the elements in the Action genre?
14	D	S	T1	Intermediate	The highest degree of the elements in the Documentary genre is 2.
15	D	S	T1	Intermediate	What is the degree of the movie Casino?
16	D	S	T1	Intermediate	The degree of the movie Casino is 2, which means it belongs to 3 genres?
17	D	S	T1	Intermediate	The degree of the movie Twister is 4, which means it belongs to 4 genres?
18	F	M	T1	Intermediate	Name the genres, to which the movie Twister belongs?
26	C	S	T2	Hard	Which one of the following genres, whose movies are mostly exclusive to it?
27	C	S	T2	Hard	Which genre of the following has the largest number of degree 2?
28	C	S	T2	Hard	Which one of the following genres, whose movies are mostly shared with other genres?
29	F	T	T2	Easy	Name a movie that belongs ONLY to Thriller genre?
34	F	T	T2	Hard	Name a movie belongs to Action and to at most 2 other genres.
35	F	T	T2	Hard	Name a movie belongs to Drama and to at least 2 other genres.

Table 4.10: The evaluation questions related to the first hypothesis. (D: Determine, F: Find, C: Compare, S: Single value, M: Multiple value, T: Text field).

Hypothesis 2					
Nr.	**Type**	**Category**	**Task**	**Difficulty**	**Description**
5	D	S	T5	Easy	How many movies belong to Musical and Children at the same time?
6	F	T	T5	Easy	Name one of them (Children and Musical movies)?
7	F	T	T5	Easy	Name a Movie that belongs to Romance and Drama at the same time?
8	D	S	T5	Easy	How many movies are from Romance or Comedy (or both)?
9	F	T	T5	Easy	Name a movie from Romance or Comedy (or both)?
10	D	S	T3	Intermediate	How many movies are Comedy but not Drama?
19	F	M	T3	Easy	Name two genres that overlap?
20	A	B	T4	Intermediate	How is the overlap between Drama and Documentary?
21	A	B	T4	Intermediate	How is the overlap between Action and Adventure?
22	C	M	T4	Hard	Which two genres have the highest overlap?
23	C	M	T4	Hard	Which two genres have a low overlap?
24	C	S	T4	Intermediate	With which genre do Comedy movies have the highest overlap?
25	C	S	T4	Intermediate	Which genre has the least overlaps with all other genres?
30	F	T	T5	Intermediate	Name a movie that belongs to exactly two genres?
31	F	T	T5	Intermediate	Name a movie that belongs to AT MOST two genres?
32	F	T	T5	Hard	Name a movie that belongs to AT LEAST three genres?
33	F	T	T5	Hard	Name a movie that belongs to exactly 4 genres?
36	D	S	T5	Intermediate	How many movies are from the Romance, Comedy and Drama?
37	D	S	T5	Intermediate	How many movies are from Romance, Horror or Comedy?
38	D	S	T5	Intermediate	How many movies are from Romance and Comedy but not Drama?
39	D	S	T5	Hard	How many movies are either Romance or Drama but not both?
40	D	S	T5	Hard	How many movies are both Action and Drama but nothing else?

Table 4.11: The evaluation questions related to the second hypothesis.

Hypothesis 3					
Nr.	**Type**	**Category**	**Task**	**Difficulty**	**Description**
41	D	S	T6	Easy	What is the Median Release Date of the movies in Action?
42	D	S	T6	Easy	What is the Median Release Date of the exclusive movies in Action?
43	A	S	T7	Hard	In the Sci-Fi genre, which movies are more recent the shared or the exclusive movies?
44	C	M	T7	Hard	Which genre has the oldest exclusively movies?
45	C	M	T7	Hard	Which genre/s has the most recent exclusive movies?
46	C	S	T7	Intermediate	Name a genre that tends to have a high average rating.
47	C	S	T7	Intermediate	Name a genre that tends to have a low average rating.
48	A	B	T7	Hard	Does the number of the watches increase when a movie has more genres?
49	D	S	T6	Easy	What is the Median Release Date of the shared movies between Romance and Drama?
50	A	B	T7	Intermediate	The movies from Children and Musical tend to be?
51	A	B	T7	Intermediate	The movies from Action and Thriller tend to be?
52	A	B	T7	Intermediate	Movies form Horror and Sci-Fi have average rating that is:
53	A	B	T7	Intermediate	Movies form Drama and Romance have average rating that is:

Table 4.12: The evaluation questions related to the third hypothesis. (D: Determine, A: Analyze, C: Compare, S: Single value, M: Multiple value, B: Behavior.)

The qualitative feedback

These questions were designed to qualitatively evaluate Radial Sets. The goal of these questions is to elicit usability and understandability feedback based on users opinion. Six qualitative questions are presented to the users after solving the evaluation questions.

The questions were formulated as follows: The first question inquired about the users' opinion on the usability of the tool. The second, third, and fourth questions focused on the clarity of sets, elements, and overlaps representation respectively. The fifth question covered the interaction with the tool (e.g., search or brushing operations). The final question concentrated on applying the operations on sets using the tool. Table. 4.13 shows the description of each question used in the experiment.

The qualitative feedback	
Nr.	**Task description**
1	How did you find the tool?
2	How intuitive was the representation of the sets as regions?
3	How intuitive was the representation of the elements as bars in the sets?
4	How intuitive was the representation of the overlaps as arcs between the regions?
5	How intuitive was the interaction with the tool (search, click)?
6	How intuitive was applying the operations on sets using the tool (union, intersection)?

Table 4.13: The question used to qualitatively evaluate Radial Sets.

The users answered these questions using a Likert scale. Each question was graded on a scale ranging from 1 to 5. Each value of the scale represents a description based on the respective question. Table. 4.14 shows the values of the scale with the corresponding description.

Values	1	2	3	4	5	**Question**
Description 1	Very easy	Easy	Neutral	Hard	Very hard	54, 58, 59
Description 2	Very clear	Clear	Neutral	Not clear	Not clear at all	55, 56, 57

Table 4.14: The Likert scale values and the corresponding descriptions.

Procedure

The experiment started by asking the users to fill in a questionnaire. This questionnaire contained questions about personal data (e.g., age, gender, occupation, and sight disorder) and self-assessment of visualization experience and knowledge about sets.

Then every user got a 20-minute introduction covering the following topics:

- Set-typed data and how to depict such data using Euler diagrams
- Radial Sets and its visual metaphor, and
- How the evaluation will be carried out, briefly explaining EvalBench [11]).

The introduction included presenting the main functions and features of the visualization technique. The evaluation process and how to use EvalBench were covered in the introduction. The main interfaces of EvalBench the users used to perform the tasks and how to proceed from a task to another were described.

After the introduction a five-minute demonstration on how to interact with the tool was presented. It included solving some example tasks using the keyboard modifiers and other functionality.

Before starting the tasks, users were offered to take a 5-minute break. This aimed to avoid any confusion with the tutorial part, for mind refreshing and to stay alert then continue with the training questions.

The evaluation consisted of 60 questions and comprised a training and an evaluation session. Before every evaluation session a training session includes seven tasks was performed. This aimed to give the users a chance to get acquainted with Radial Sets. Users got instructions with the tasks during the session and feedback if their answers were correct or not after it. The users were informed to solve the questions correctly and as fast as possible and had the possibility to ask question and get clarification. At the end users' answers were reviewed and the wrong answered questions were discussed and corrected.

To start the evaluation session, Radial Sets were presented along with EvalBench. The presentation was in full screen mode to provide enough space for the visualization, the task description and the answers. A detailed description on how EvalBench works will be presented in the last section of this chapter. The user then start solving the evaluation questions by interacting with the tool and submitting the answers. There were various options to submit an answer, for example, either by selecting one or more answers from a list or by entering the answer in a text box.

After completing the evaluation questions the user started with the qualitative feedback. For answering these questions no interactions with the tool were required. Users were asked about their opinion on some usability issues, the interaction with the tool, and the visualization.

Finally, the evaluation ended by asking the user to provide her/his feedback about the visualization technique. What advantages and disadvantages she/he found while interacting with the tool. What recommendations on how to improve the representation they can suggest. User comments have been recorded, reviewed, and listed in the next chapter.

The training questions, evaluation questions, and qualitative feedback were presented for all users in the same order. The time and error of the evaluation questions and qualitative feedback have been recorded using EvalBench [11]. The data related to the training question were not included in the result. Furthermore, the collected data regarding the evaluation and feedback questions were analyzed and presented in chapter. 5.

The pilot study

Before the evaluation has been conducted, a pilot study has been carried out. The goal of the study was to test if the questions were understandable, to find flaws in the design, and to assess the time to accomplish the tasks. The study has been performed with 8 users (see Table. 4.2 and Table. 4.3).

The users of the pilot study went through the same procedure as the users of the experiment, i.e., an introduction has been presented (20 minutes), tutorial on how to interact with Radial Sets (five minutes), a break (five minutes), and then they started solving the questions. The difference between the pilot study and the experiment was that in the pilot study the introduction and the tutorial were presented for all 8 users at the same time, while in the experiment they were presented for each user individually.

The task completion time and the task correctness were recorded using EvalBench individually, as in the experiment. The collected data from this study was excluded from the evaluation results. User comments and notes have also been recorded and reviewed.

The results of the pilot study were used to improve the evaluation design. Based on these results the session time and the number of questions have been reduced from 90 minutes to 60 minutes and from 73 questions to 59 respectively. Moreover, the wording of some questions has been simplified to make it easier to understand. Some new question were introduced to ensure that users understood the tasks. For example, two new questions were added to the tasks which aimed to ensure that users understood the concept of the degree of an element (tasks 16 and 17).

4.5 EvalBench

EvalBench [11] is a software library for visualization evaluation. The library was developed using the Java programming language. It can be integrated with visualization prototypes that need to be evaluated via loose coupling. It supports both quantitative and qualitative evaluation methods such as controlled experiments and laboratory questionnaires.

Evalbench has been used in this work to record the tasks time and error the users made when they were solving the questions. The software has been integrated with Radial Sets as shown in Fig. 4.1.

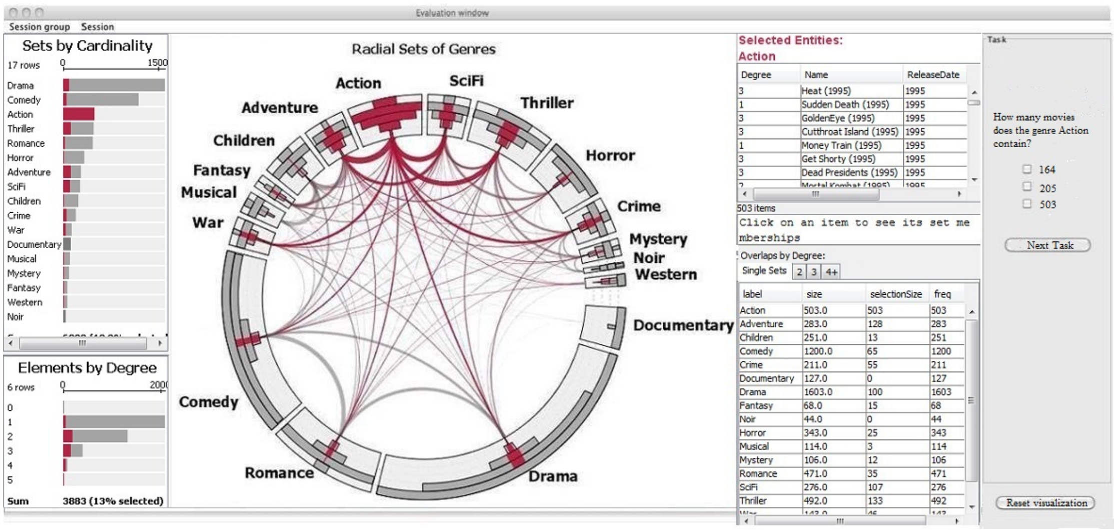


Figure 4.1: Screenshot of EvalBench along with Radial Sets.

In order to load the questions into EvalBench, an XML file comprising the questions list has been created. This file is listed in details in appendix. B. The file contains information describing each question, the type of the question, and the type of answers (e.g., text, multiple choice).

The following attributes have been defined for the evaluation of Radial Sets.

- Question ID: An identifier of a task which is unique.
- Question category: Describes the type of the question (e.g., Analyze).
- Question description: A textual description of the action that has to be performed by the user (e.g., Name a movie that belongs to Romance and Drama at the same time).
- Question configurations: These configurations define the data set and the visualization mode.

- Correct answers: To define the correct answer for the question. It might be a numerical value or multiple values (e.g., Action, as answer to the question, what is the genres of the movie Bad Boys).

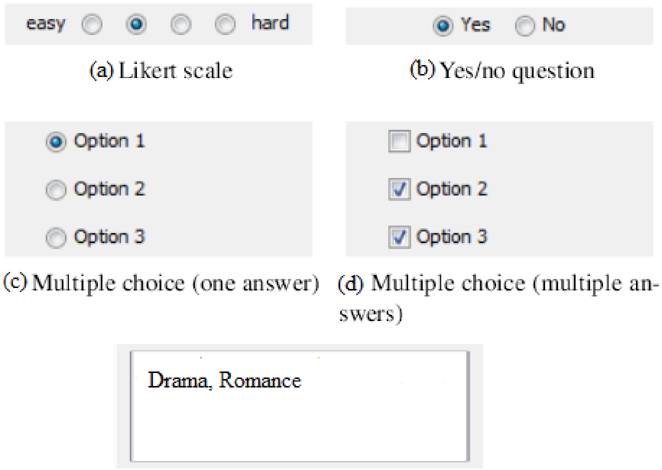
While a user is performing the experiment, EvalBench stores answer information in a new file when a the user solves the question. For each question the file contains the following information:

- Start date: Represents the date when a user started to solve the question (measured in milliseconds).
- End date: Represents the date when the question was finished (measured in milliseconds).
- Given response: The user's answer for a task.
- Task correctness: To compare the defined correct answer with the given response by a user.

When the evaluation session starts, a pop-up message is shown with the task description. After reading the description and pressing the 'OK' button, the visualization was presented and a timer for the task was started. The visualization was presented on the left side of the screen along with task description and the possible answers on the right side of the screen. The user had to solve the task by interacting with the tool and submitting the answer.

EvalBench provided various options to answer the questions and a related user interface (see Fig. 4.2). For the evaluation of Radial Sets the following options have been offered to the users to submit the answers during the experiment:

- Check boxes: The user can submit one or more right answer from a list of answers (Fig. 4.2d), for example, "Which two genres have the highest overlap, Which two genres have the highest overlap?".
- Radio Buttons: The user had to submit only one answer (Fig. 4.2c), for example, ||"Which genre has the least overlaps with all other genres?".
- Text box: The user can use it to enter a text string (Fig. 4.2e). For example, Name a Movie that belongs to Romance and Drama at the same time.
- Likert scale: This input form has been used to answer the qualitative feedback questions (Fig. 4.2a), for example, "representing the elements as bars in the sets was (very easy, easy, medium, hard or very hard)".
- Yes/no question: This option was used to answer true/false questions (Fig. 4.2b), for example, "The degree of the movie Casino is 2, which means that it belongs to 3 genres".



(a) Likert scale (b) Yes/no question (c) Multiple choice (one answer) (d) Multiple choice (multiple answers) (e) Text answer

Figure 4.2: Questions answering options offered by EvalBench and used in the experiment (adopted and simplified from Aigner et al. [11]).

A task is completed after the user enters the answer and presses the 'Next task' button. This turns the timer to stop off and brings up a new pop-up message with the description of the next question. The same procedure was repeated for all questions. The state of the visualization has been changed for the questions concerned with attribute analyzing (e.g., coloring the bars or overlaps).

For every evaluation session, EvalBench created a comma-separated file (CSV). This file has been imported into a statistics package (e.g., R) to analyze the evaluation results as shown in the next chapter. The file contains the run time attributes of each question. The question completion time and the question correctness along with some of the design time attributes were also recorded as shown in Fig. 4.3.

	B	D	E	F	J	K	L	M
1	task	taskType	startTime	duration_ms	questionText	correctness	givenAnswer	correctAnswer
2	1	7	01.09.2013 19:21	23052	How many movies does the genre Action contain?	1.0	option3	option3
3	2	9	01.09.2013 19:22	23089	Name one of them (Action's movies)?	1.0	heat	heat
4	3	7	01.09.2013 19:22	25135	How many movies come exclusively(ONLY) with the Action genre?	1.0	option1	option1
5	4	9	01.09.2013 19:23	26115	Name one of them (exclusive Action's movies)?	1.0	badboys	badboys
6	5	9	01.09.2013 19:23	22054	How many movies belong to Musical and Children in the same time?	1.0	option2	option2
7	6	2	01.09.2013 19:23	23708	Name one of them (Children and Musical movies)?	1.0	sleepingbeauty	sleepingbeauty
8	7	5	01.09.2013 19:24	32155	Name a Movie that belongs to Romance and Drama in the same time?	1.0	algiers	algiers
9	8	7	01.09.2013 19:24	24779	How many movies are from Romance or Comedy (or both)?	1.0	option3	option3
10	9	8	01.09.2013 19:25	12618	Name a movie from Romance or Comedy(or both)?	1.0	aboutadam	aboutadam
11	10	8	01.09.2013 19:25	21196	How many movies are Comedy but not Drama?	1.0	option1	option1

Figure 4.3: Screenshot of the recorded attributes opened in Microsoft Excel.

CHAPTER 5

Results

This chapter presents the results of an empirical evaluation of Radial Sets. The evaluation results contained collected time and error data of 32 users. This data was collected during the experiment when the participants were performing the tasks. The data was prepared and processed for analysis. The R software package was used as a tool to analyze the collected data. The results of the experiment were separately grouped and analyzed according to the hypothesis they are related to. Moreover, the results were divided for each group of tasks based on the level of difficulty assigned to the respective questions.

The results were collected by recording time and error made by users during the experiment. For each question the average time and the percentage of correct answers are reported. Additionally, the respective confidence interval (CI) for the average time of each question is reported. The confidence interval has been calculated using the formula [72]:

$$\bar{X} \pm 1.96 \times \frac{\alpha}{\sqrt{n}}$$

Where:

- ($\bar{X}$) is the sample mean
- (α) is the significance level and is equal to (0.05)
- (n) is the sample size and equal to (32)

5.1 Hypothesis H1

H1: Radial Sets enable to quickly analyze the distribution of elements in the sets, exploring the elements in each set according to their degrees, and determining the exclusive or the shared elements in the sets.

To test the first hypothesis, the experiment includes 18 questions with 3 different levels of difficulty (easy, intermediate, and hard). In chapter. 3 I mentioned seven tasks supported by Radial Sets. These questions have been formulated as instances of the first (T1) and the second (T2) pattern finding tasks (see chapter. 3). The questions focused on elements memberships in the sets and covered both analyzing the distribution of the elements in each set according to their degrees as well as finding elements in a specific set that are exclusive to it or shared with one or more sets.

Easy-Difficulty Tasks

This group of tasks consists of seven questions (Table. 5.1). The users' completion times for solving these questions have been imported into R. A box-plot for each question has been generated as shown in Fig. 5.1.

In chapter. 3 the criteria for assigning a level of difficulty to each question has been presented. A question is defined as an easy question if it requires a query on the data base that can be performed with one or no set operation on the data and at most one level of de-aggregation.

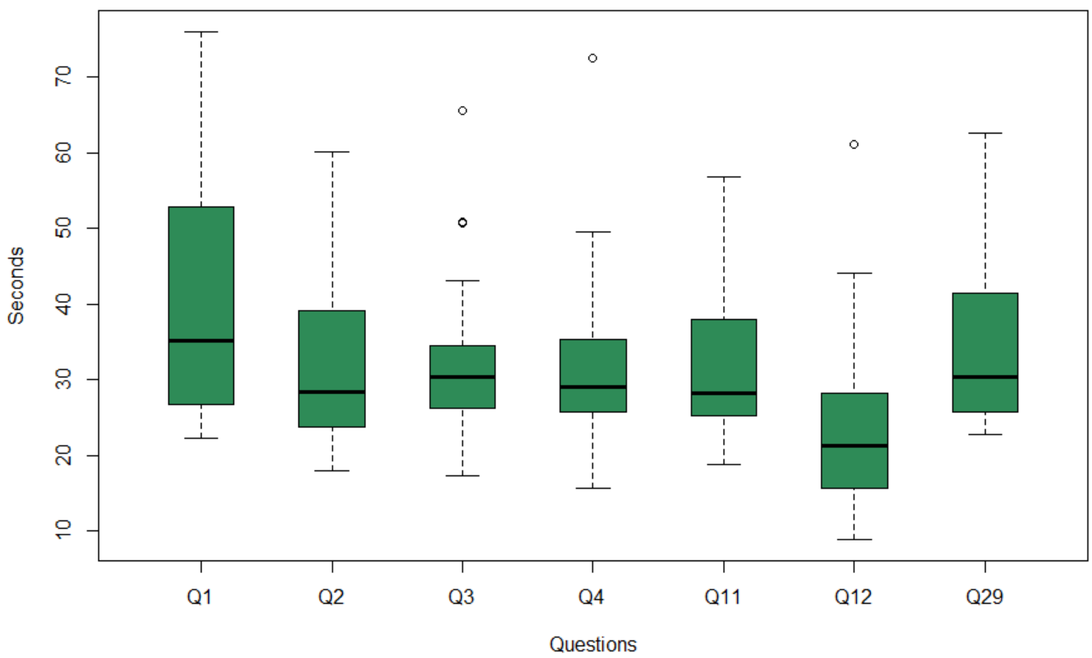


Figure 5.1: Box plots of the completion times for the easy difficulty questions of H1.

The bar chart in Fig. 5.2 shows the percentage of users who answered the questions correctly (out of 32 users in total). The correctness is equal to the percentage of users who answered the questions correctly. The questions are of easy difficulty and are related to H1. All users answered this group of questions correctly.

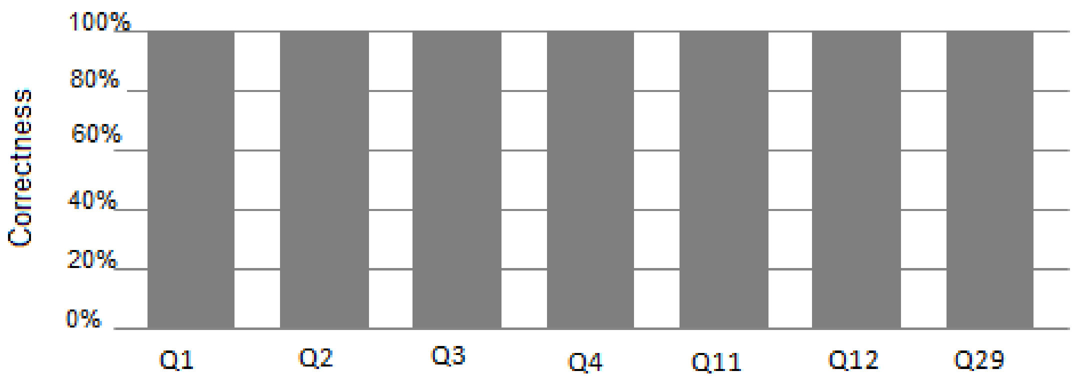


Figure 5.2: The percentage of users who answered the easy difficulty questions of H1 correctly.

Table. 5.1 summarizes the average completion time along with the respective confidence interval and the correctness rate for each easy difficulty question of the first hypothesis.

Question Nr.	**Time ± CI (sec.)**	**Correctness (pct.)**
1	**39.748** ± 5.18	100%
2	32.551 ± 4.24	100%
3	32.168 ± 3.36	100%
4	31.478 ± 3.59	100%
11	32.655 ± 3.62	100%
12	**23.376** ± 3.70	100%
29	33.687 ± 3.52	100%
Overall	32.238	100%

Table 5.1: Summary of the results of the easy difficulty questions of H1.

Table. 5.1 shows that all users answered this group of questions correctly. The highest average time was 39.748 seconds for solving Q1, whereas the lowest average time was 23.376 seconds for solving Q12. The overall average time of the easy questions is 32.238 seconds, and the overall average correctness is 100%.

Intermediate-Difficulty Tasks

This group of tasks encompasses six questions. Fig. 5.3 shows the generated box plots of the completion times for each question. The correctness is equal to the percentage of users who answered the questions correctly.

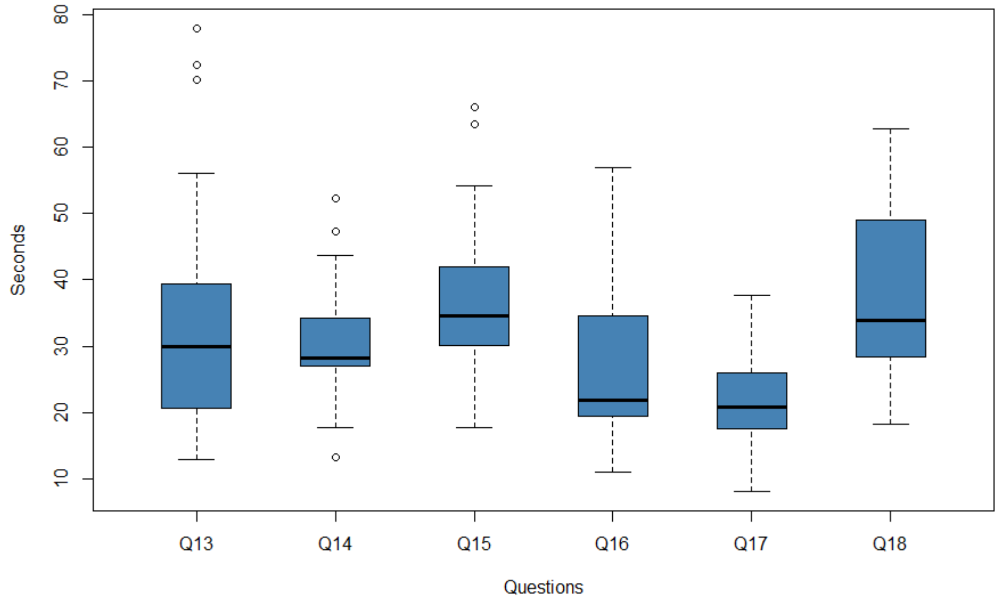


Figure 5.3: Box plots of the completion times for the intermediate difficulty questions of H1.

The bar chart in Fig. 5.4 shows the percentage of users who answered the questions correctly (out of 32 users in total). The questions are of intermediate difficulty and are related to H1.

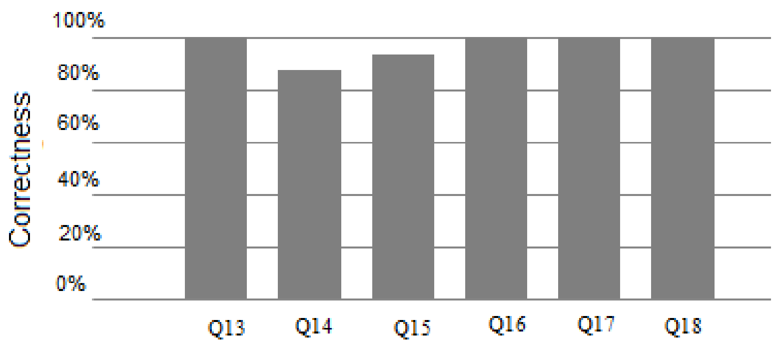


Figure 5.4: The percentage of users who answered the intermediate difficulty questions of H1 correctly.

All users answered Q13, Q16, Q17, and Q18 correctly, whereas 87.500% (28 out of 32) and 93.750% (30 out of 32) of the users answered Q14 and Q15 correctly respectively.

A question is assigned as an intermediate level of difficulty if it requires a query on the data base that can be performed with two or three set operations on the data (see chapter. 3). Table.

5.2 summarizes the average completion time and correctness rate for each of these intermediate question related to the first hypothesis.

Question Nr.	**Time ± CI (sec.)**	**Correctness (pct.)**
13	33.299 ± 5.82	100%
14	30.683 ± 2.80	**87.500%**
15	36.163 ± 4.03	93.750%
16	26.941 ± 3.96	100%
17	**21.703** ± 2.24	100%
18	**37.845** ± 4.61	100%
Overall	31.106	96,875%

Table 5.2: Summary of the results of the intermediate difficulty question related to of H1.

The highest and the lowest average time the users needed to solve this kind of questions were 37.845 seconds for Q18 and 21.703 seconds for Q17 respectively. Moreover, the users solve this type of tasks with a high success rates. The overall average time of the intermediate questions is 31.106 seconds, and the overall average correctness is 96,875%.

From Table. 5.2 we notice that Q14 has an average completion time but a relatively low correctness. Comparing Q14 with Q13 which has 100% correctness we notice the following:

- Both questions have the same type.
- They focus on the elements' degree.
- Q13, Q14 are instances of the same task, T1.
- Q13 needed more completion time (33.299 sec) than Q14 (30.683 sec) with no errors.

The reason for such result might be that the users wanted to finish the question as fast as possible, assuming both questions have the same answer. The data items included in these might also have an impact of the correctness.

Hard-Difficulty Tasks

This group of tasks composed of five questions. The generated box-plots of the completion times for each question are presented in Fig. 5.5.

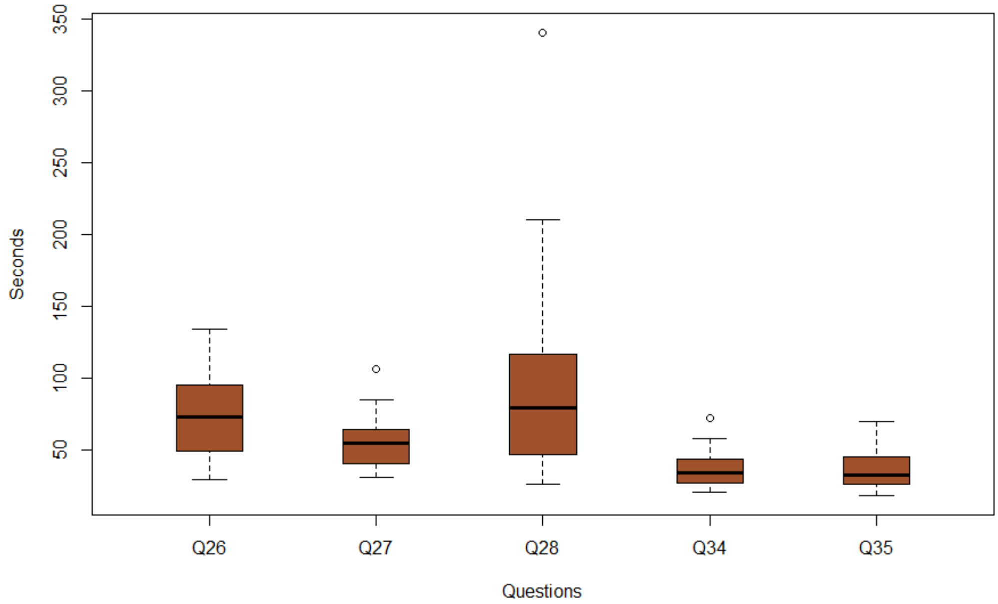


Figure 5.5: Box plots of the completion times for the hard difficulty questions of H1.

Fig. 5.6 shows the percentage of users who answered the hard difficulty questions related to H1 correctly.

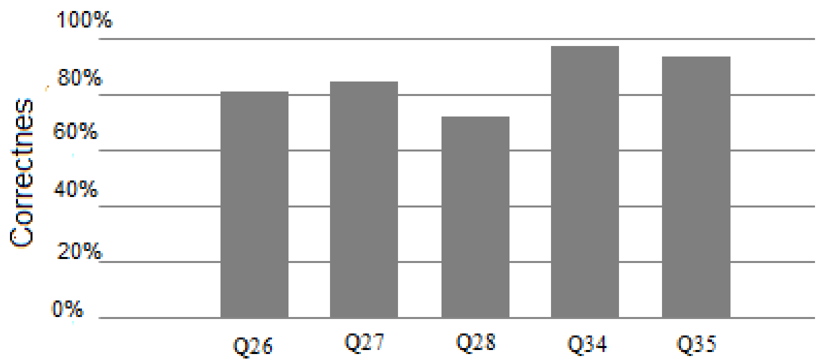


Figure 5.6: The percentage of users who answered the hard difficulty questions of H1 correctly.

A question is defined as a hard level of difficulty question if it requires a query on the data base that can be performed with four or more set operations on the data (see chapter. 3). Table. 5.3 summarizes the average of the completion time and the correctness rate for the hard difficulty question.

Question Nr.	Time ± CI (sec.)	Correctness
26	74.214 ± 9.79	81.250%
27	54.869 ± 5.98	84.375%
28	**91.379** ± 21.71	**71.875%**
34	37.676 ± 4.22	**96.875%**
35	**36.283** ± 4.42	93.750%
Overall	58.884	85,625%

Table 5.3: Summary of the results of the hard difficulty question related to of H1.

The highest and the lowest average time the users needed to solve this kind of question were 91.379 seconds for Q28 and 36.283 seconds for Q35 respectively. Moreover, the users solve this type of questions with relatively good success rates. The overall average time of the hard difficulty questions is 58.884 seconds and the overall average correctness is 85,625%.

From Table. 5.3 we noticed that Q26, Q27, and Q28 needed more time to be accomplished than the other questions with a relatively low correctness. These questions were concerned with finding and comparing elements in the sets.

The reason for such result might be that Radial Sets can either depict the absolute sizes of the overlaps and the elements in a set, or their normalized sizes. Moreover, both sizes were presented simultaneously via a tooltip. This presentation might have caused confusion to the users while answering the questions. Some users asked which presented size should be considered although the difference between both sizes and the purpose of each one was explained in the introduction.

Discussion of hypothesis H1

The first hypothesis is concerned with elements-set memberships. The evaluation questions focus on analyzing the distribution of elements in the sets and specifying the exclusive and the shared elements in each set. 18 questions with 3 different levels of difficulty (easy, intermediate, and hard) have been performed. The questions have been formulated as instances of the first (T1) and the second (T2) tasks (see chapter. 3).

The results in Table. 5.1, Table. 5.2, and Table. 5.3 summarize the average completion time and correctness for the questions according to their levels of difficulty. The results show that Radial Sets is effective for solving the easy and the intermediate questions, even for users who have no experience in visualization. The technique might also be considered for solving the hard level of difficulty questions, taking into account the overall correctness rate of 85,625% for these questions. The results of the questions that require a comparison between the elements of multiple sets can considerably be improved by modifying the visual presentation of the depicted absolute and normalized sizes of the overlaps and the elements in a set. In summary the results of the first hypothesis provide evidence that Radial Sets can quickly enable:

- Analyzing the distribution of elements in the sets,
- Exploring the elements in each set according to their degrees and
- Determining the exclusive or the shared elements in each set with relatively high success rates (98.125%).

5.2 Hypothesis H2

H2: Radial Sets support revealing and facilitate analyzing the overlaps between large sets. This include determining the sets that tend to have high or low overlaps and exposing the elements in the overlaps.

This hypothesis was tested by means of 22 questions categorized into three groups based on their level of difficulty. The questions have been defined as instances of the third task (T3), the fourth task (T4), and the fifth task (T5) (see chapter. 3). The questions focused on the overlaps between the sets. They covered analyzing the overlaps, exposing elements that belong to a specific overlap, and finding which pairs of sets have higher overlap than other pairs.

Easy-Difficulty Tasks

This group comprised six questions. The respective box plots for users' completion times have been generated using the R software package as shown in Fig. 5.7.

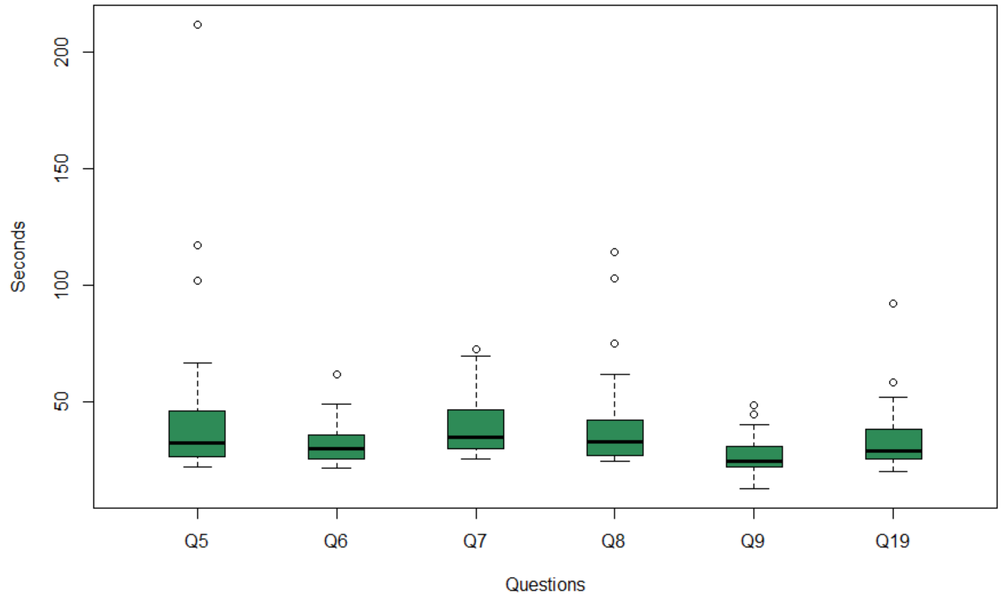


Figure 5.7: Box plots of the completion times for the easy difficulty questions of H2.

The bar chart in Fig. 5.8 shows the percentage of users who answered the questions correctly (out of 32 users in total). The questions are of easy difficulty and are related to H2.

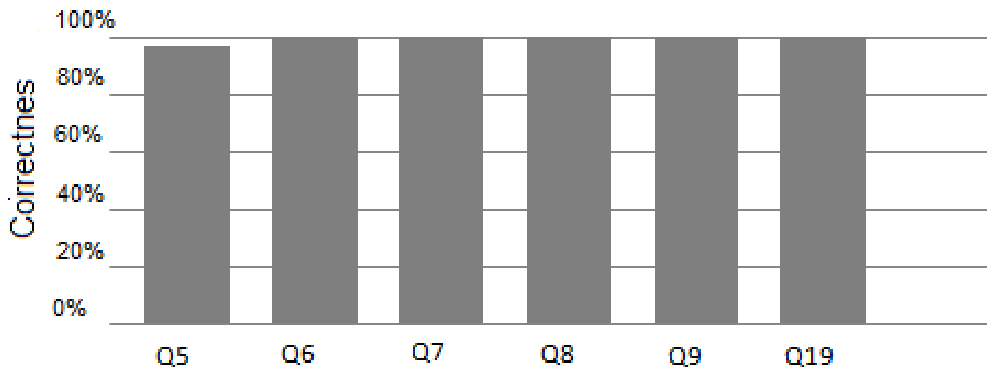


Figure 5.8: The percentage of users who answered the easy difficulty questions of H2 correctly.

Table. 5.4 summarizes the average completion time and the correctness rate for each easy difficulty question of the second hypothesis H2. All users answered this group of questions correctly except Q5, 31 users out of 32.

Question Nr.	Time ± CI (sec.)	Correctness
5	**46.028** ± 12.95	**96.875%**
6	32.082 ± 3.06	100%
7	39.974 ± 4.80	100%
8	39.897 ± 7.42	100%
9	**26.563** ± 2.59	100%
19	34.500 ± 5.18	100%
Overall	36.507	99,480%

Table 5.4: Summary of the results of the easy difficulty questions of H2.

Table. 5.4 shows that the users answered this group of questions correctly. The highest average time needed to solve this kind of questions was 46.028 seconds for solving Q5, whereas the lowest average time the users needed to solve them was 26.563 seconds for solving Q9. The total average time of the easy questions is 36.507 seconds, and the total average correctness is 99,48%.

Intermediate-Difficulty Tasks

This group of tasks encompasses ten questions. Fig. 5.9 shows the generated box plots of the completion times for each question.

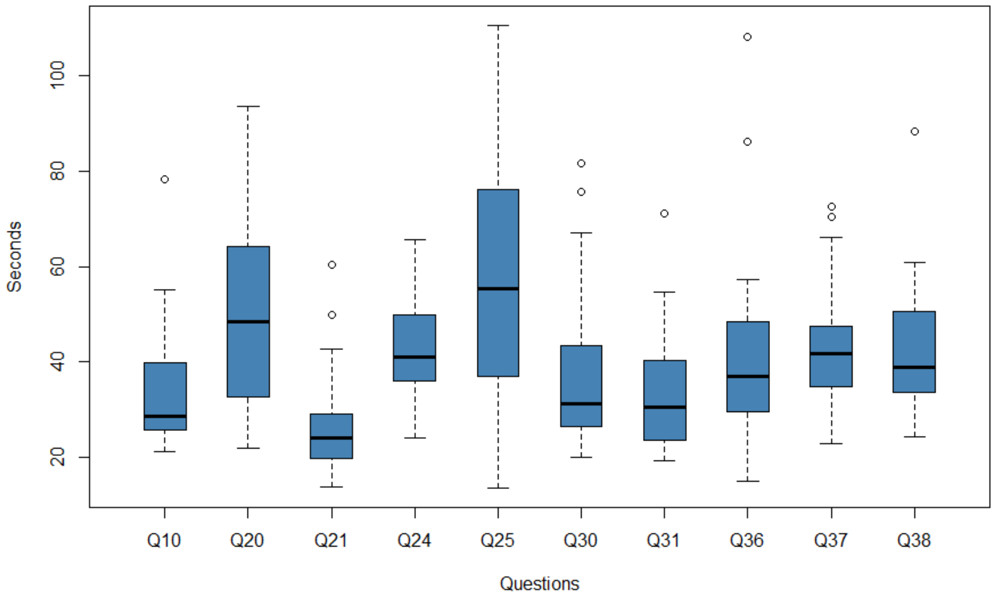


Figure 5.9: Box plots of the completion times for the intermediate difficulty questions of H2.

The bar chart in Fig. 5.10 shows the percentage of users who answered the questions correctly. The questions are of intermediate difficulty and are related to H2.

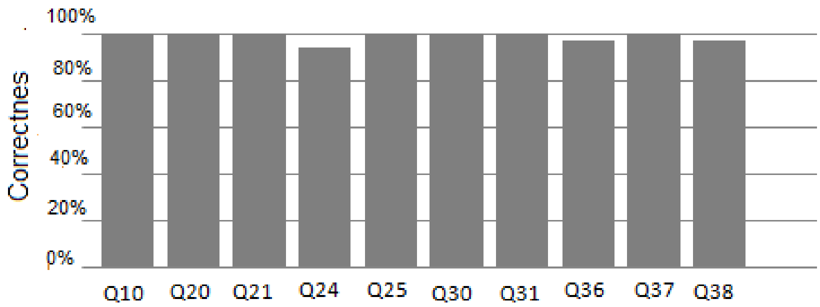


Figure 5.10: The percentage of users who answered the intermediate difficulty questions of H2 correctly.

Table. 5.5 summarizes the average completion time and the correctness rate for each intermediate question related to the second hypothesis H2. The highest and the lowest average time the users needed to solve this kind of questions were 56.309 seconds for Q25 and 26.638 seconds for Q21 respectively. Moreover, the users solve this type of tasks with a very high success rates. The overall average time of the intermediate questions is 40.778 seconds, and the overall average correctness is 98,751%.

Question Nr.	**Time ± CI (sec.)**	**Correctness**
10	34.041 ± 4.30	100%
20	50.144 ± 6.53	100%
21	**26.638** ± 3.56	100%
24	43.343 ± 3.87	**93,750%**
25	**56.309** ± 8.44	100%
30	37.451 ± 5.62	100%
31	33.541 ± 4.27	100%
36	41.064 ± 6.38	96,875%
37	42.916 ± 4.39	100%
38	42.336 ± 4.54	96,875%
Overall	40.778	98,751%

Table 5.5: Summary of the results of the intermediate difficulty questions related to H2.

Table. 5.5 shows that the performance of Radial Sets is consistent across different questions in this category with relatively high correctness rate. The tool is very suitable for the intermediate difficulty questions.

Hard-Difficulty Tasks

This group of tasks encompass six questions. The generated box plots of the completion times for each question are presented in Fig. 5.11.

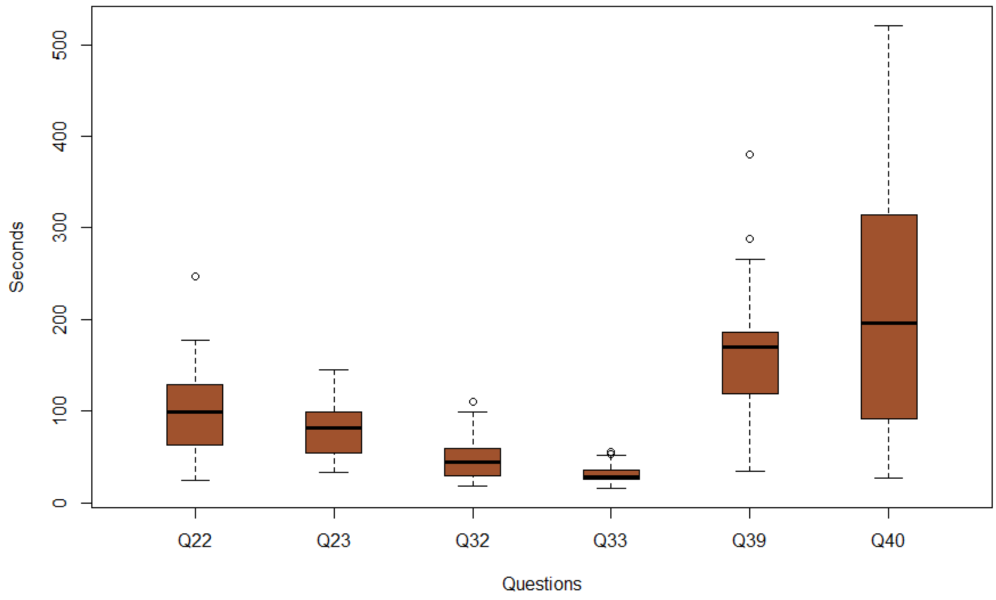


Figure 5.11: Box plots of the completion times for the hard difficulty questions of H2.

Fig. 5.12 shows the percentage of users who answered the hard difficulty questions related to H2 correctly.

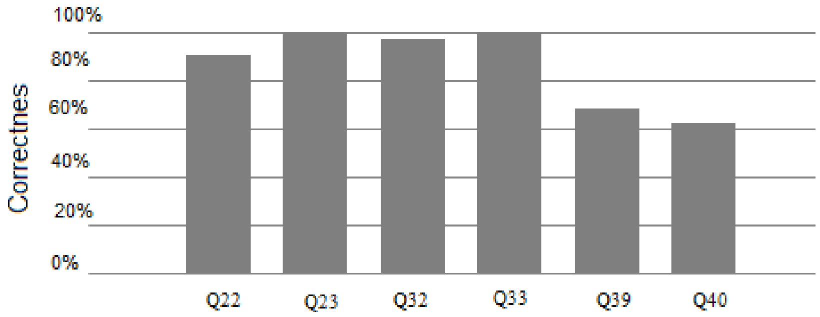


Figure 5.12: The percentage of users who answered the hard difficulty questions of H2 correctly.

Table. 5.6 summarizes the average completion time and correctness rate for the hard difficulty questions.

Question Nr.	Time ± CI (sec.)	Correctness
22	101.245 ± 16.36	90.625%
23	78.823 ± 9.82	100%
32	47.971 ± 7.58	96.875%
33	**31.804** ± 3.54	100%
39	158.290 ± 26.61	68.750%
40	**255.594** ± 50.37	**62.500%**
Overall	112.288	86,460%

Table 5.6: Summary of the results of the hard difficulty tasks related to H2.

The highest and the lowest average time the users needed to solve this kind of question were 255.594 seconds for Q40 and 31.804 seconds for Q33 respectively. The total average time of the hard questions is 112.288 seconds and the total average correctness is 86,46%

From Table. 5.6 we noticed that users solved Q39 and Q40 in a long time with a low correctness rate compared to the other questions. These two questions were concerned with finding elements that belong to a specific overlap using keyboard modifiers to perform set operations. Although solving these questions required higher knowledge about both operations on sets and visualizations, some users with less knowledge and experience solved them correctly.
Moreover, Q22 and Q23 have been solved in a relatively long time but with high success rates. These two questions were concerned with finding which pairs of sets have higher or lower overlap than other pairs. To solve these questions users had to go through the depicted overlaps and to compare them.

Discussion of hypothesis H2

The second hypothesis covers the overlaps between the sets. The evaluation questions focus on analyzing the overlaps between two sets or between groups of sets, finding elements that belong to a specific overlap, and finding which pairs of sets have higher or lower overlap than other pairs. 22 questions with 3 different levels of difficulty (easy, intermediate, and hard) have been performed. The questions have been formulated as instances of T3, T4, and T5 pattern finding tasks (see chapter. 3).

The results in Table. 5.4, Table. 5.5, and Table. 5.6 summarize the average completion time and correctness for the questions according to their levels of difficulty. The results show that Radial Sets can effectively be used to solve both the easy and the intermediate questions. In order to solve the hard questions effectively, the domain expert should have experience in visualization. The results of the second hypothesis provide evidence that Radial Sets supports:

- Revealing the overlaps between large sets,
- determining the sets that tend to have high or low overlaps and
- exposing the elements in the overlaps.

5.3 Hypothesis H3

H3: Radial Sets enable analyzing the elements' attributes and revealing how they correlates with the elements' set memberships or overlaps.

This hypothesis includes 13 questions categorized into three groups based on their level of difficulty. The questions have been defined as instances of task T6 and the task T7 from the tasks (see chapter. 3). The questions focus on elements attributes. They covered analyzing how an attribute of the elements correlates with their memberships and with the overlaps. Additionally, they cover analyzing how these correlations for a subset of elements differ from the rest of the elements.

Easy-Difficulty Tasks

This group encompasses three questions. Fig. 5.13 shows the generated box-plots of the users' completion times for each question.

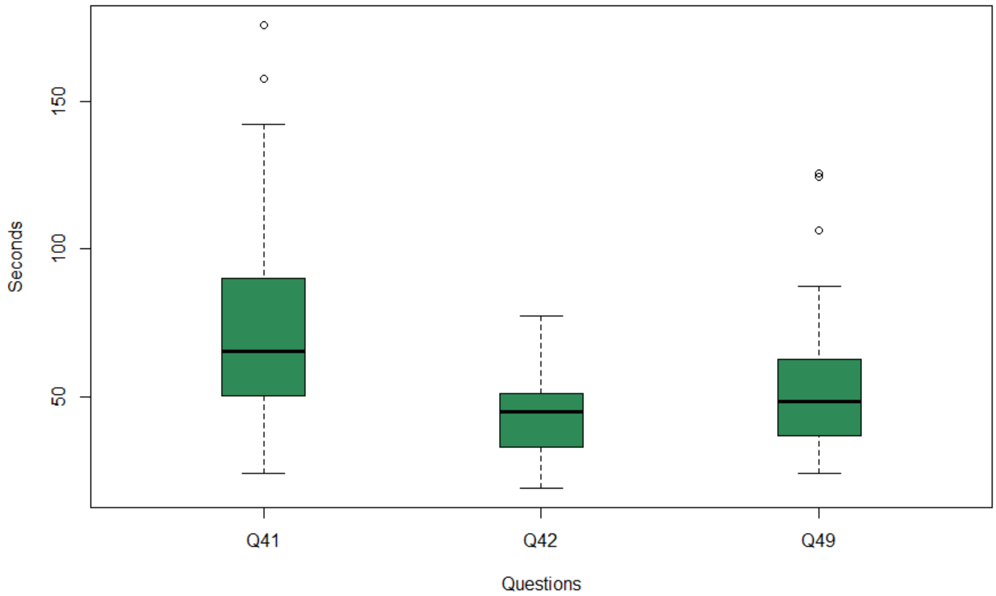


Figure 5.13: Box plots of the completion times for the easy difficulty questions of H3.

The bar chart in Fig. 5.14 shows the percentage of users who answered the questions correctly. The questions are of easy difficulty and are related to H3.

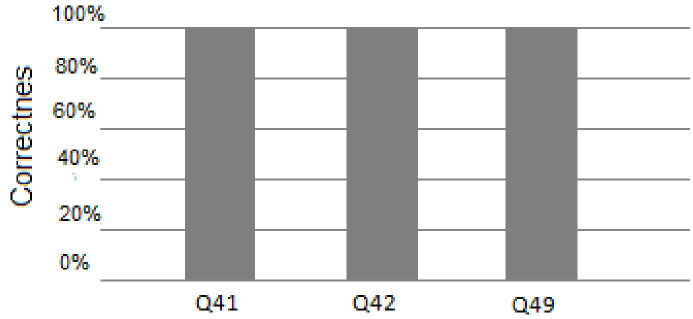


Figure 5.14: The percentage of users who answered the easy difficulty questions of H3 correctly.

Table. 5.7 summarizes the average completion time and correctness for each easy question of hypothesis H3.

Question Nr.	**Time $\pm$ CI (sec.)**	**Correctness**
41	**74.501** $\pm$ 13.20	100%
42	**42.953** $\pm$ 4.49	100%
49	54.606 $\pm$ 8.90	100%
Overall	36.507	100%

Table 5.7: Summary of the results of the easy difficulty questions related to H3.

Table. 5.7 shows that all 32 users answered this group of questions correctly. The highest average time needed to solve this kind of questions was 74.501 seconds for solving Q41, whereas the lowest average time was 42.953 seconds for solving Q42. The overall average time of the easy difficulty questions is 57.353 seconds, and the overall average correctness is 100%.

Intermediate-Difficulty Tasks

This group of tasks encompasses six questions. Fig. 5.15 shows box plots of the completion times for each question.

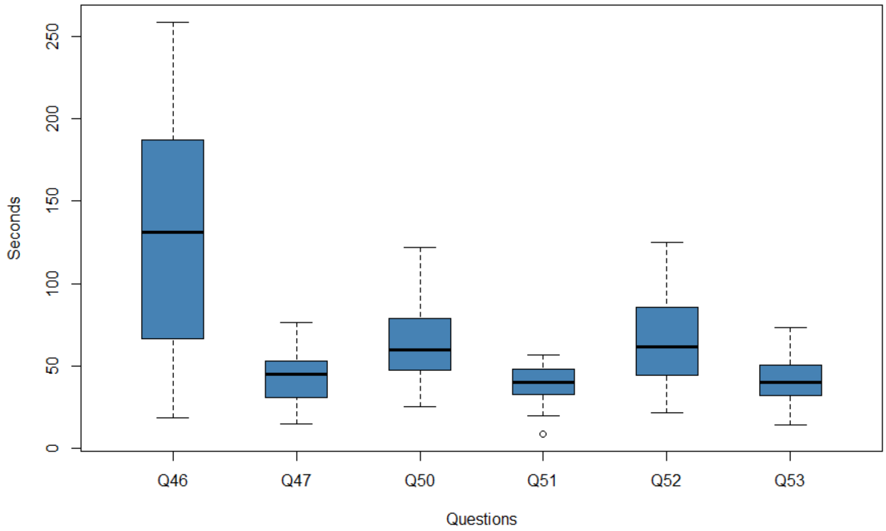


Figure 5.15: Box plots of the completion times for the intermediate difficulty questions of H3.

The bar chart in Fig. 5.16 shows the percentage of users who answered the questions correctly. The questions are of intermediate difficulty and are related to H3.

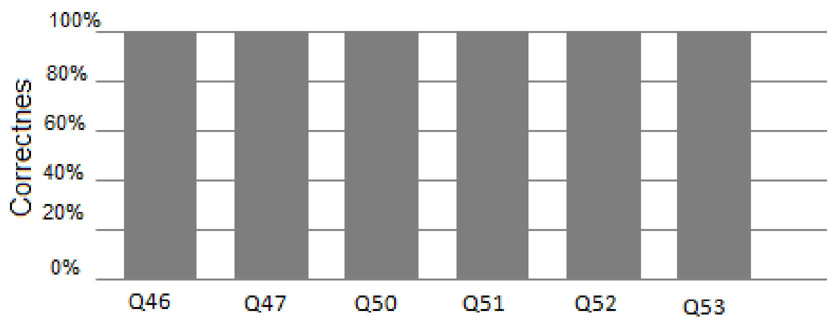


Figure 5.16: The percentage of users who answered the intermediate difficulty questions of H3 correctly.

Table. 5.8 summarizes the average completion time and correctness for each intermediate-difficulty question of the hypothesis H3. The highest and lowest average time the users needed to solve this kind of questions were 133.59 seconds for Q46 and 38.31 seconds for Q51 respectively. Moreover, all users solved this type of questions with no errors. The overall average time of the questions is 64.39 seconds, and the overall average correctness is 100%.

Question Nr.	Time ± CI (sec.)	Correctness
46	**133.59** ± 24.53	100%
47	44.661 ± 5.49	100%
50	63.701 ± 8.43	100%
51	**38.310** ± 3.91	100%
52	65.132 ± 9.52	100%
53	40.946 ± 4.95	100%
Overall	64.390	100%

Table 5.8: Summary of the results of the intermediate difficulty questions related to H3.

Table. 5.8 shows that Radial Sets is very effective for solving this kind of task. Users required longer time to solve Q46 than other questions. This question focused on finding the genre that tends to have a high average rating. Users had to check all genres and compare the color of the respective bars. Some users needed more time because there was more than one genre that has a high average rating and they wanted to find the genre with the highest average rating. Other users wanted to be sure of the answer, therefore, they used both the color representation and the value of the average rating shown by means of tooltips.

Hard-Difficulty Tasks

This group of tasks composed of four questions. The generated box plots of the completion times for each question are presented in Fig. 5.17.

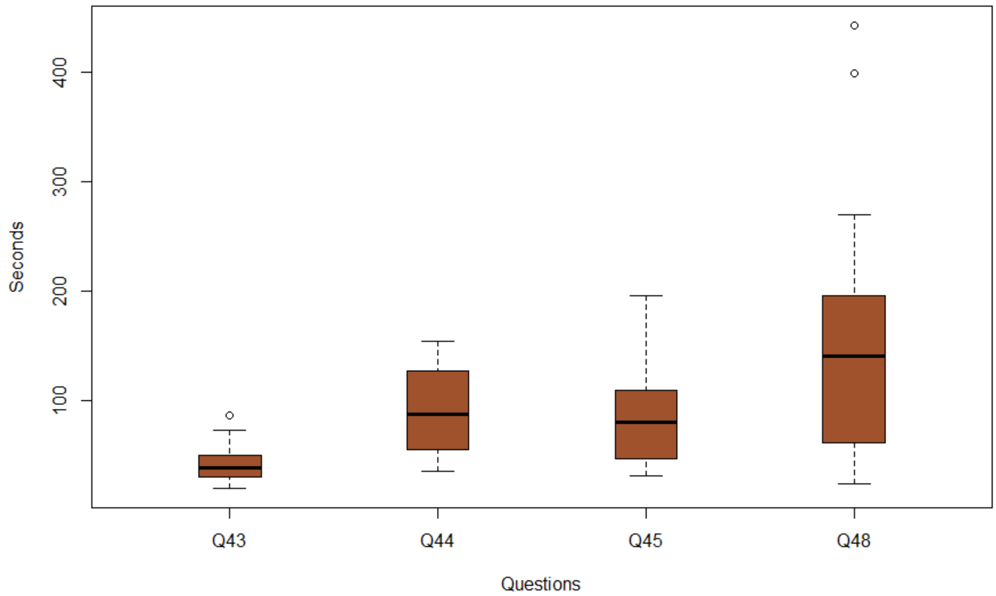


Figure 5.17: Box plots of the completion times for the hard difficulty questions of H3.

Fig. 5.18 shows the percentage of users who answered the hard difficulty questions related to H3 correctly.

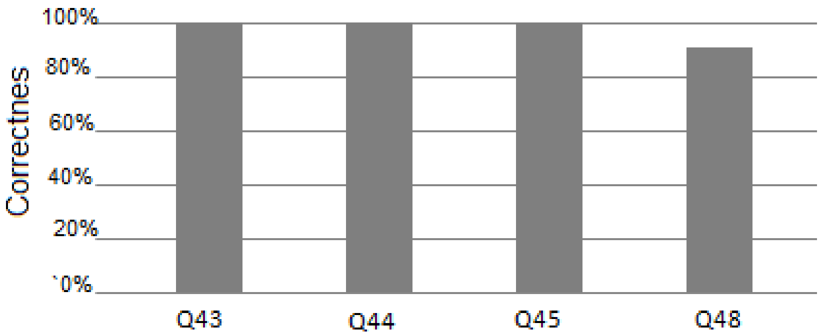


Figure 5.18: The percentage of users who answered the hard difficulty questions of H3 correctly.

Table. 5.9 summarizes the average of the completion time and the correctness rate for the hard difficulty question.

The highest and the lowest average time the users needed to solve this kind of question were 145.243 seconds for Q48 and 42.091 seconds for Q43 respectively. The overall average time of the hard questions is 90,646 seconds and the overall average correctness is 97,658%. Moreover, the users solve this type of tasks with high correctness rates.

Question Nr.	Time ± CI (sec.)	Correctness
43	**42.091** ± 5.35	100%
44	90.061 ± 12.92	100%
45	85.188 ± 14.22	100%
48	**145.243** ± 35.38	**90.625%**
Overall	90,646	97,658 %

Table 5.9: Summary of the results of the hard difficulty questions related to H3.

From Table. 5.9 we noticed that users solved Q48 in a longer time with a lower correctness compared to the other questions. This question was concerned with analyzing how the attribute ”number of watches” correlates with the set memberships and overlaps to detect if the number of watches increases when a movie has more genres. Users had to compare the color of the exclusive and the shared elements in each genre. Then based on the results of all comparisons, a user can find the answer of the question. Such procedure is not trivial and may require a long time especially for users with less experience in visualizations.

On the other hand, Q43 has been solved in a short time with no error. This question covered analyzing how the set memberships and the attribute (release date) values can be used to differentiate the exclusive from the shared movies in a specific genre (Sci-Fi). This question can easily be solved by comparing the color, which represent the attribute values of the exclusive and the shared movies.

The difference between both questions is, Q48 requires multiple comparisons at the level of all sets, while Q43 requires a single comparison at the level of a single sets. This explains the reason why users needed longer time for solving Q48 and shorter time for Q43.

Discussion of hypothesis H3

The third hypothesis is concerned with attributes analysis. The evaluation questions focus on: Analyzing how the attributes correlate with the elements-set memberships or with the overlaps, and how they can be used to distinguish a subset of elements from the rest of them. Users performed 13 questions with 3 different levels of difficulty (easy, intermediate, and hard). The questions have been formulated as instances of tasks T6 and T7 (see chapter. 3).

The results in Table. 5.7, Table. 5.8, and Table. 5.9 summarize the average completion time and correctness for the questions according to their levels of difficulty. The discussions show that even users with less knowledge or experience in visualization can solve the questions with high correctness rates. The results are consistent across different questions providing an evidence for the effectiveness of Radial Sets in solving this kind of tasks. Consequently, the results of third hypothesis suggest that Radial Sets:

- Enable analyzing the elements' attributes.
- allow analyzing how an attribute of the elements correlates with their memberships to the sets and the overlaps.

5.4 Qualitative feedback results

In this section the results of the qualitative feedback described in chapter. 4 are presented. After solving the evaluation questions, users started to answer six qualitative questions categorized into two groups. The first group covers usability issues and the second group covers the clarity of the visual representation. Each group included 3 questions to be answered using a Likert scale ranging from 1 to 5. The values of the scale with the corresponding description are listed in chapter. 4.

Usability

To evaluate the usability of Radial Sets, three questions have been formulated. These questions address the ease of use of the tool (Q54), the interaction with it (Q58), and the application of set operations (Q59). The results of the questions have been analyzed. Fig. 5.19 shows the bar charts of user answers of the first group of questions.

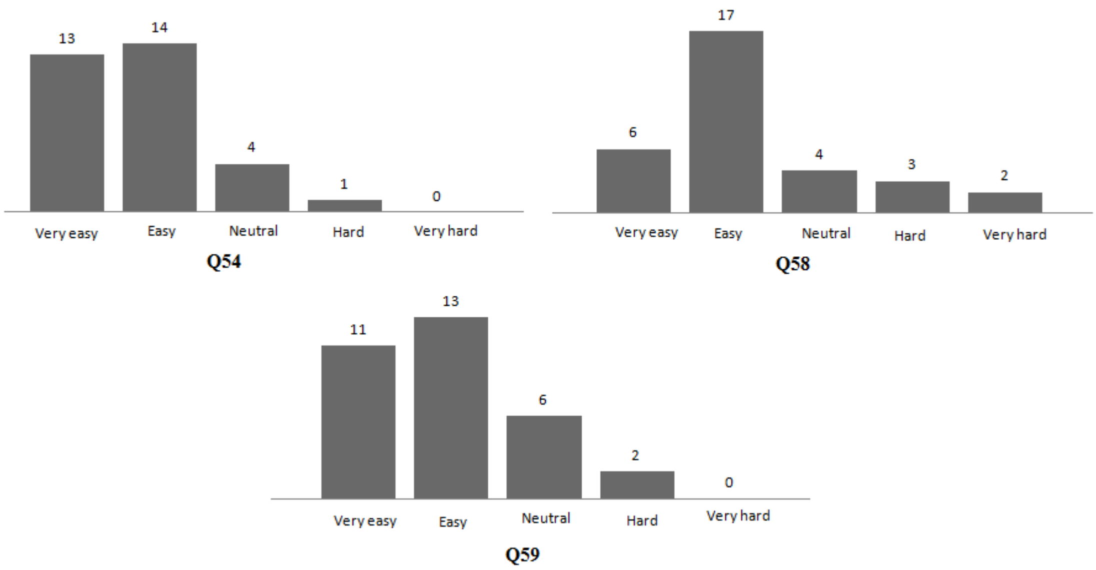


Figure 5.19: Users' opinion on the usability of Radial Sets. The ease of use (Q54), the interaction with the tool (Q58), and applying the operations on sets (Q59).

Table. 5.10 shows the results of users' answers on the three questions addressing

Description	Value	#of Users	Percentage
		Q54	
Very easy	1	13	40.63%
Easy	2	14	43.75%
Neutral	3	4	12.50%
Hard	4	1	3.13%
Very hard	5	0	0%
		Q58	
Very easy	1	6	18.75%
Easy	2	17	53.13%
Neutral	3	4	12.50%
Hard	4	3	9.38%
Very hard	5	2	6.25%
		Q59	
Very easy	1	11	34.38%
Easy	2	13	40.63%
Neutral	3	6	18.75%
Hard	4	2	6.25%
Very hard	5	0	0%

Table 5.10: Summary of the users' opinion on the usability of Radial Sets.

Table. 5.10 shows that, more than (84%) of the users found Radial Sets easy to use. Moreover, more than (71%) rated the interaction with it as easy. Finally, more than (75%) of them found applying set operations using, easy.

On the other hand, only (3.13%), (15.63%), and (6.25%) found the tool hard to use, to interact with, and to apply set operations using respectively. The most frequently reported difficulty was concerned with interaction (Q58). Some Users mentioned claimed that the reason for such rating was, because of the presence of too many arcs between the sets. This makes it sometimes hard to find and to click on the right arc from the first time, although the same users found the interaction with the bars or regions easy.

Based on these results, Radial Sets can be considered to be a relatively easy to use tool with the main functions and features being easy to learn. However, there is a need to improve interaction with the arcs.

Clarity of the visual representation

To evaluate the clarity of the design of Radial Sets, three questions have been formulated. the users have been asked to assess the intuitiveness of visual representation, in particular, representing the sets as regions (Q55), elements as bars in the regions (Q56), and overlaps as arcs between the regions (Q57). The results of the question have been analyzed. Fig. 5.20 shows bar charts of user answers on the second group of questions.

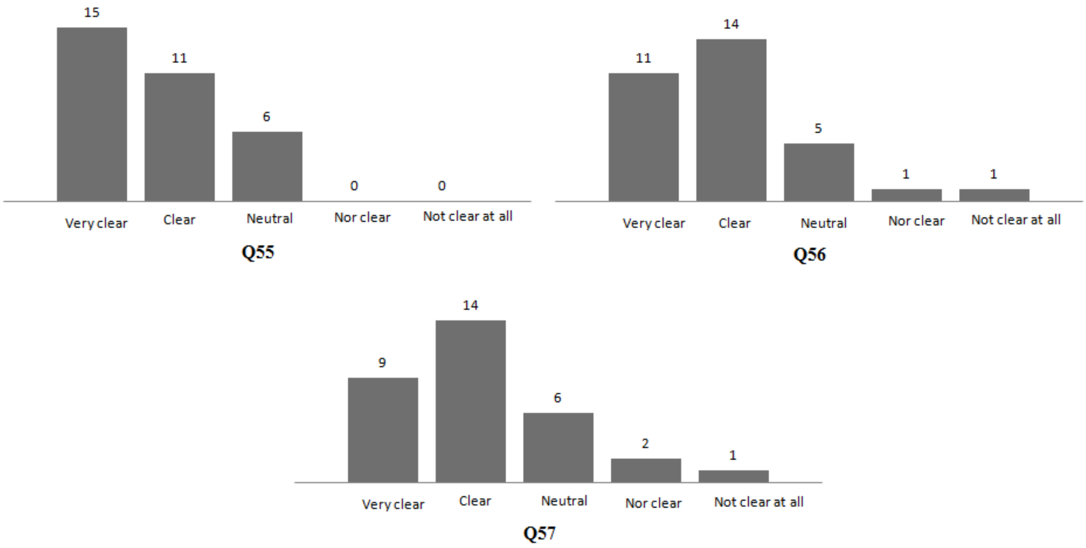


Figure 5.20: Users' opinion on the clarity of the design of Radial Sets. The visual representation of the sets (Q55), the elements (Q56), and the overlaps (Q57).

Table. 5.11 shows the users' answers on the three questions regarding the clarity of the visual representation of the sets, elements, and overlaps. From Table. 5.11 in total (81.25%), (78.13%), and (71.83%) of the users assessed the visual representation of the sets as regions, depicting elements as bars in the regions, and the overlaps as arcs between the regions as clear respectively.

Description	Value	#of Users	Percentage
		Q55	
Very clear	1	15	46.88%
Clear	2	11	34.38%
Neutral	3	6	18.75%
Not clear	4	0	0%
Not clear at all	5	0	0%
		Q56	
Very clear	1	11	34.38%
Clear	2	14	43.75%
Neutral	3	5	15.63%
Not clear	4	1	3.13%
Not clear at all	5	1	3.13%
		Q57	
Very clear	1	9	28.13%
Clear	2	14	43.75%
Neutral	3	6	18.75%
Not clear	4	2	6.25%
Not clear at all	5	1	3.13%

Table 5.11: Summary of the users' opinion on the clarity of the visual design of Radial Sets.

On the other hand 6.25% of the users found the visual representation of the elements as bars in the sets hard. 9.38% found the visual representation of the overlaps as arcs between the regions hard. Moreover, none of the users rated the visual representation of the sets as regions as hard.

These results suggest that the sets and elements in Radial Sets are relatively easy to perceive as such, while arcs can be harder to perceive as overlaps. Users' comments and feedback about Radial Sets will be listed in the next section.

Summary- Qualitative feedback questions

These questions aimed to qualitatively evaluate Radial Sets by assessing usability issues and eliciting design feedback. They included six questions to be answered after performing the experiment. They focused on users' opinion on the ease of use of Radial Sets, the visual representation of the sets, elements and overlaps, the interaction with it and finally, its functions and the features.

Concerning the usability of Radial Sets, Table. 5.10 shows that most of the users found using it relatively easy. Even users with less experience in visualization needed short time to learn its functions and how to interact with it. This provide an evidence of the usability of Radial Sets.

For the clarity of the design, Table. 5.11 shows that the majority of users were able to understand the representations of the sets as regions, the elements as bars in these regions, and the overlaps as arcs between the regions. Some users provided comments and suggestions to improve the visual representations, as listed in the next section.

Further feedback and user comments

The evaluation ended by asking the user to provide her/his feedback and recommendations about Radial Sets and what advantages and limitations have been found when they were interacting with the tool. In the following, users' comments and suggestions are listed:

- A user was expecting by clicking on an arc between two sets to see only the shared elements between these sets (shared elements with the degree 2), without including the elements that belongs to other sets too.
- A uses recommended presenting an additional column in the overlap analysis view that provides more information about the overlaps as well as the possibility to sort the overlaps ascendency or descendancy by this information.
- A user claimed that the arcs were confusing. He recommended hiding the arcs in the beginning, and by clicking on a region, only the related arcs will be shown. It might be easier to show only the visual elements or the desired functions of the tool according to the tasks at hand. For example, for analyzing how an attribute value correlates with the set memberships there is no need to present the arcs.
- Some questions requested showing the absolute sizes for tasks on the level of a single set. The normalized sizes are then used when only to compare between the elements of two or more sets (tasks on the level of multiple sets). Some users were confused because of presenting both absolute sizes and normalized sizes of the overlaps and the elements in a set simultaneously.
- Most users found the keyboard modifiers easy to use. Only seven users asked for an annotation that contains the keyboard modifiers and the respective set operation.
- Some users recommended that it will be better if they have the option to choose other colors scales than red-yellow for the bars and overlaps.

- Three users found the overlap analysis view complicated, especially the size, selection size, and the frequency columns. They asked about reason for presenting them and what they could be used for.
- A user was so impressed with the tool. He asked about the chance of how to promote the tool and if it is possible for him to do that.

5.5 Conclusion

The evaluation carried out during this thesis was aimed to quantitatively assess the effectiveness of Radial Sets in performing the pattern finding tasks mentioned in chapter. 3. There were 32 participants who solved 60 questions each. The correctness and completion times were recorded using EvalBench [11]. Moreover, additional questions aimed to qualitatively assess usability issues and the visual design based on users' feedback.

To answer the research questions:

- **Q1:** *Is the Radial Sets technique effective in performing tasks it is designed to support?* Three hypotheses based on the pattern finding tasks covered the elements-set membership, the overlaps, and attribute analysis were defined. The effectiveness of Radial Sets in solving each task is reported along with discussion of each hypothesis.
- **Q2:** *How can the Radial Sets technique be improved to satisfy the objectives of the design?* The qualitative feedback questions were defined. The usability and the clarity of the visual representations of its elements were observed in the discussion of these question.

It has been found that Radial Sets is suited for dealing with large overlapping sets and solving the tasks. Furthermore, It is relatively easy to learn tool and interact with it for all users with technical background even if they had no experience with visualization.

Hypothesis H1 stated that Radial Sets is effective for analyzing the distribution of elements in the sets according to their degrees and determining the exclusive or the shared elements in the sets. Hypothesis H2 addressed that Radial Sets support analyzing the overlaps between large sets and determining the sets that tend to have high or low overlaps. Hypothesis H3 stated that Radial Sets enable analyzing the elements' attributes and revealing how they correlates with the elements' set memberships or overlaps. The qualitative feedback questions showed that Radial Sets is relatively easy to use tool with the main functions and the representations of the visual elements are clear.

It has been noticed that some improvements can significantly increase the effectiveness of the tool in solving the tasks, in particular, simplifying the representation of the overlaps as arcs. Presenting certain arcs based on the targeted task or for specific sets only instead of presenting the whole arcs between the sets makes the set overlaps clearer and easier to comprehend and interact.

Finally, for the future work, a comparative evaluation with another tool that supports the same tasks might be conducted. The evaluation could be performed by mean of quantitative and qualitative methods with same type of questions. The results might be analyzed by mean of significance analysis. Other data sets could be used, new tasks, and larger number of users or tasks. Future work on the presentation of some visual elements, such as the tooltips and on some features, such as colors can be done for the sake of improving the tool.

Bibliography

[1] S. K. Card, J. D. Mackinlay, and B. Schneiderman, *Readings in Information Visualization: using vision to think*. Morgan Kaufmann, 1999.

[2] L. Euler, "Lettres a une princesse d'allemagne. sur divers sujets de physique et de philosophie, 1761. letters no. 102-108 vol 2 basel."

[3] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers, "Visualizing sets and set-typed data: State-of-the-art and future challenges," in *Eurographics conference on Visualization (EuroVis) State of The Art Reports*, pp. 1–21, Eurographics, 2014.

[4] K. Wittenburg, A. Malizia, L. Lupo, and G. Pekhteryev, "Visualizing set-valued attributes in parallel with equal-height histograms," in *Proceedings of the International Working Conference on Advanced Visual Interfaces AVI*, pp. 632–635, ACM, 2012.

[5] C. M. Freitas, P. R. Luzzardi, R. A. Cava, M. Winckler, M. S. Pimenta, and L. P. Nedel, "On evaluating information visualization techniques," in *Proceedings of the working conference on Advanced Visual Interfaces (AVI)*, pp. 373–374, ACM, 2002.

[6] B. Alsallakh, W. Aigner, S. Miksch, and H. Hauser, "Radial Sets: Interactive visual analysis of large overlapping sets," *IEEE Transactions on Visualization and Computer Graphics (Proceedings Information Visualization 2013)*, vol. 19, no. 12, pp. 2496–2505, 2013.

[7] C. D. Wickens and C. M. Carswell, "The proximity compatibility principle: Its psychological foundation and relevance to display design," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 3, pp. 473–494, 1995.

[8] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 9, pp. 1520–1536, 2012.

[9] C. Plaisant, "The challenge of information visualization evaluation," in *Proceedings of the working conference on Advanced visual interfaces (AVI)*, pp. 109–116, ACM, 2004.

[10] A. Kerren, J. Stasko, and J.-D. Fekete, *Information Visualization: human-centered issues and perspectives LNCS*, vol. 4950. Springer, 2008.

[11] W. Aigner, S. Hoffmann, and A. Rind, "EvalBench: A software library for visualization evaluation," in *Computer Graphics Forum*, vol. 32, pp. 41–50, Wiley Online Library, 2013.

[12] M. Pohl, F. Scholz, S. Kriglstein, B. Alsallakh, and S. Miksch, "Evaluating the dot-based contingency wheel: Results from a usability and utility study," in *Human Interface and the Management of Information. Information and Knowledge Design and Evaluation*, pp. 76–86, Springer, 2014.

[13] K. Andrews, "Evaluation comes in many guises," in *Workshop on Beyond time and errors: novel evaLuation methods for Information Visualization (BELIV)*, 2008.

[14] G. Ellis and A. Dix, "An explorative analysis of user evaluation studies in information visualisation," in *Proceedings of the AVI workshop on BEyond time and errors: novel evaluation methods for Information Visualization BELIV*, pp. 1–7, ACM, 2006.

[15] S. Carpendale, "Evaluating information visualizations," in *Information Visualization*, pp. 19–45, Springer, 2008.

[16] J. E. McGrath, "Human-computer interaction," ch. Methodology Matters: Doing Research in the Behavioral and Social Sciences, pp. 152–169, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995.

[17] D. M. Hilbert and D. F. Redmiles, "Extracting usability information from user interface events," *ACM Computing Surveys (CSUR)*, vol. 32, no. 4, pp. 384–421, 2000.

[18] M. Y. Ivory and M. A. Hearst, "The state of the art in automating usability evaluation of user interfaces," *ACM Computing Surveys (CSUR)*, vol. 33, no. 4, pp. 470–516, 2001.

[19] T. Munzner, "A nested model for visualization design and validation," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 921–928, 2009.

[20] U. Net, "Methods table," 2006.

[21] O. Kulyk, R. Kosara, J. Urquiza, and I. Wassink, "Human-centered aspects," in *Human-Centered Visualization Environments*, pp. 13–75, Springer, 2007.

[22] L. Barkhuus and J. Rode, "From mice to men–24 years of evaluation in CHI," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - ALT CHI*, Citeseer, 2007.

[23] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage, 2013.

[24] J. J. Thomas and K. A. Cook, *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press, 2005.

[25] J. Venn, "On the diagrammatic and mechanical representation of propositions and reasonings," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 10, no. 59, pp. 1–18, 1880.

[26] J. Flower, A. Fish, and J. Howse, “Euler diagram generation,” *Journal of Visual Languages & Computing*, vol. 19, no. 6, pp. 675–694, 2008.

[27] P. Simonetto, *Diagrammes d’Euler pour la visualisation de communautés et d’ensembles chevauchants*. PhD thesis, Université Sciences et Technologies-Bordeaux I, 2011.

[28] L. Micallef and P. Rodgers, “eulerape: Drawing area-proportional 3-venn diagrams using ellipses,” *PLoS ONE (to appear)*, 2014.

[29] M. Alqadah, G. Stapleton, J. Howse, and P. Chapman, “Evaluating the impact of clutter in euler diagrams,” pp. 108–122, 2014.

[30] P. Chapman, G. Stapleton, P. Rodgers, L. Micallef, and A. Blake, “Visualizing sets: An empirical comparison of diagram types,” in *Diagrammatic Representation and Inference* (T. Dwyer, H. Purchase, and A. Delaney, eds.), vol. 8578 of *Lecture Notes in Computer Science*, pp. 146–160, Springer Berlin Heidelberg, 2014.

[31] A. Verroust and M.-L. Viaud, “Ensuring the drawability of extended [e]uler diagrams for up to 8 sets,” in *Diagrammatic Representation and Inference*, pp. 128–141, Springer, 2004.

[32] N. H. Riche and T. Dwyer, “Untangling Euler diagrams,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 16, no. 6, pp. 1090–1099, 2010.

[33] P. Rodgers, L. Zhang, and A. Fish, “General Euler diagram generation,” in *Diagrammatic Representation and Inference*, pp. 13–27, Springer, 2008.

[34] C. Collins, G. Penn, and S. Carpendale, “Bubble sets: Revealing set relations with isocontours over existing visualizations,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 1009–1016, 2009.

[35] B. Alper, N. H. Riche, G. Ramos, and M. Czerwinski, “Design study of LineSets, a novel set visualization technique,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2259–2267, 2011.

[36] K. Dinkla, M. J. van Kreveld, B. Speckmann, and M. A. Westenberg, “Kelp diagrams: Point set membership visualization,” in *Computer Graphics Forum*, vol. 31, pp. 875–884, Wiley Online Library, 2012.

[37] W. Meulemans, N. H. Riche, B. Speckmann, B. Alper, and T. Dwyer, “Kelpfusion: a hybrid set visualization technique,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 11, pp. 1846–1858, 2013.

[38] M. Harrower and C. A. Brewer, “Colorbrewer.org: an online tool for selecting colour schemes for maps,” *The Cartographic Journal*, vol. 40, no. 1, pp. 27–37, 2003.

[39] J. Bertin, *Graphics and graphic information processing*. Walter de Gruyter, 1981.

[40] J. Bertin and M. DARU, “Matrix theory of graphics: Jacques Bertin’s theories,” *Information Design Journal*, vol. 10, no. 1, pp. 5–19, 2000.

[41] L. Wilkinson and M. Friendly, “The history of the cluster heat map,” *The American Statistician*, vol. 63, no. 2, 2009.

[42] I. Liiv, “Seriation and matrix reordering methods: A historical overview,” *Statistical analysis and data mining*, vol. 3, no. 2, pp. 70–91, 2010.

[43] H. Siirtola and E. Mäkinen, “Constructing and reconstructing the reorderable matrix,” *Information Visualization*, vol. 4, no. 1, pp. 32–48, 2005.

[44] J. Talbot, B. Lee, A. Kapoor, and D. S. Tan, “EnsembleMatrix: interactive visualization to support machine learning with multiple classifiers,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 1283–1292, ACM, 2009.

[45] N. Henry and J.-D. Fekete, “MatrixExplorer: a dual-representation system to explore social networks,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 12, no. 5, pp. 677–684, 2006.

[46] B. Kim, B. Lee, and J. Seo, “Visualizing set concordance with permutation matrices and fan diagrams,” *Interacting with computers*, vol. 19, no. 5-6, pp. 630–643, 2007.

[47] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, “Cluster analysis and display of genome-wide expression patterns,” *Proceedings of the National Academy of Sciences*, vol. 95, no. 25, pp. 14863–14868, 1998.

[48] H. A. Kestler, A. Müller, T. M. Gress, and M. Buchholz, “Generalized venn diagrams: a new method of visualizing complex genetic set relations,” *Bioinformatics*, vol. 21, no. 8, pp. 1592–1595, 2005.

[49] M. Ghoniem, J.-D. Fekete, and P. Castagliola, “A comparison of the readability of graphs using node-link and matrix-based representations,” in *IEEE Symposium on Information Visualization (INFOVIS).*, pp. 17–24, IEEE, 2004.

[50] K. Wittenburg, T. Lanning, M. Heinrichs, and M. Stanton, “Parallel bargrams for consumer-based information exploration and choice,” in *Proceedings of the 14th annual ACM symposium on User interface software and technology UIST*, pp. 51–60, ACM, 2001.

[51] J. Heer, S. K. Card, and J. A. Landay, “Prefuse: a toolkit for interactive information visualization,” in *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI)*, pp. 421–430, ACM, 2005.

[52] H. Rosling, “Gapminder,” *GapMinder Foundation http://www. gapminder. org*, 2009.

[53] P. Simonetto, D. Auber, and D. Archambault, “Fully automatic visualisation of overlapping sets,” in *Computer Graphics Forum*, vol. 28, pp. 967–974, Wiley Online Library, 2009.

[54] J. Flower and J. Howse, “Generating Euler diagrams,” in *Diagrammatic Representation and Inference*, pp. 61–75, Springer, 2002.

[55] K. Misue, "Drawing bipartite graphs as anchored maps," in *Proceedings of the Asia-Pacific Symposium on Information Visualisation-Volume 60*, pp. 169–177, Australian Computer Society, Inc., 2006.

[56] W. Freiler, K. Matkovic, and H. Hauser, "Interactive visual analysis of set-typed data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 6, pp. 1340–1347, 2008.

[57] B. Alsallakh, E. Gröller, S. Miksch, and M. Suntinger, "Contingency wheel: Visual analysis of large contingency tables," *Proc. EuroVA*, pp. 53–56, 2011.

[58] C. Courage and K. Baxter, *Understanding your users: A practical guide to user requirements methods, tools, and techniques*. Gulf Professional Publishing, 2005.

[59] D. Stone, C. Jarrett, M. Woodroffe, and S. Minocha, *User interface design and evaluation*. Morgan Kaufmann, 2005.

[60] J. Bortz and N. Döring, *Forschungsmethoden und Evaluation: für Human-und Sozialwissenschaftler*. Springer Deutschland, 2006.

[61] B. Alsallakh, W. Aigner, S. Miksch, and M. E. Gröller, "Reinventing the contingency wheel: scalable visual analytics of large categorical data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2849–2858, 2012.

[62] R. Sadana, T. Major, A. Dove, and J. Stasko, "Onset: A visualization technique for large-scale binary set data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 20, no. 12, p. to appear, 2014.

[63] L. Micallef, P. Dragicevic, and J. Fekete, "Assessing the effect of visualizations on bayesian reasoning through crowdsourcing," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2536–2545, 2012.

[64] J. Huo, "KMVQL: a visual query interface based on karnaugh map," in *Proceedings of the working conference on Advanced visual interfaces*, pp. 243–250, ACM, 2008.

[65] H. Hofmann, "Exploring categorical data: interactive mosaic plots," *Metrika*, vol. 51, no. 1, pp. 11–26, 2000.

[66] H. Hofmann, A. P. Siebes, and A. F. Wilhelm, "Visualizing association rules with interactive mosaic plots," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, pp. 227–235, ACM, 2000.

[67] M. Dörk, N. H. Riche, G. Ramos, and S. Dumais, "Pivotpaths: Strolling through faceted information spaces," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2709–2718, 2012.

[68] P. Simonetto and D. Auber, "Visualise undrawable euler diagrams," in *Information Visualisation (IV). 12th International Conference*, pp. 594–599, IEEE, 2008.

[69] R. Amar, J. Eagan, and J. Stasko, “Low-level components of analytic activity in information visualization,” in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pp. 111–117, IEEE, 2005.

[70] N. Andrienko and G. Andrienko, *Exploratory analysis of spatial and temporal data*. Springer, 2006.

[71] M. Brehmer and T. Munzner, “A multi-level typology of abstract visualization tasks,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 12, pp. 2376–2385, 2013.

[72] D. R. Cox and D. V. Hinkley, *Theoretical statistics*. CRC Press, 1979.

List of Figures

List of Tables

APPENDIX A

User Tasks

The description of users questions along with the possible answers is listed here. The right answer/answers is/are marked in bold. Some questions required more than one choice to be ticked in order to submit the right answer. An answer is considered as a right one, only if the user submitted all required choices. On the other side if a question has more than one right answer, submitting only one of them is enough to considered it as a right one. Others questions required that the user submits the right answer as a text.

A.1 Evaluation questions

Q.Nr	Questions description	Possible answers
1	How many movies does the genre Action contain?	[164, 205, **503**]
2	Name one of them (Action's movies)?	Text answer
3	How many movies come exclusively with the Action genre?	[**65**, 71, 76]
4	Name one of them (exclusive Action's movies)?	Text answer
5	How many movies belong to Musical and Children at the same time?	[**21**, 37, 81]
6	Name one of them (Children and Musical movies)?	Text answer
7	Name a Movie that belongs to Romance and Drama at the same time?	Text answer
8	How many movies are from Romance or Comedy (or both)?	[**598**, 809, 1467]
9	Name a movie from Romance or Comedy (or both)?	Text answer

Table A.1: The evaluation questions.

Q.Nr	Questions description	Possible answers
10	How many movies are Comedy but not Drama?	[**974**, 1401, 1613]
11	How many genres does the movie Bad Boys belong to?	[**1**, 3, 4]
12	Name the genre/s (the movie Bad Boys belongs to)?	[**Action**, Adventure, Children, Comedy, Musical, Romance, Drama, Documentary, Horror, Western, Thriller, Crime]
13	What is the highest degree of the elements in the Action genre?	[1, **5**, 6]
14	The highest degree of the elements in the Documentary genre is 2.	[**True**, False]
15	What is the degree of the movie Casino?	[0, **2**, 5]
16	The degree of the movie Casino is 2, which means it belongs to 3 genres?	[True, **False**]
17	The degree of the movie Twister is 4, which means it belongs to 4 genres?	[**True**, False]
18	Name the genres, to which the movie Twister belongs?	[**Action**, **Adventure**, Children, Comedy, Crime, Documentary, Drama, Fantasy, Horror, Mystery, **Romance**, **Thriller**]
19	Name two genres that overlap?	Text answer
20	How is the overlap between Drama and Documentary?	[**Low**, Medium, High]
21	How is the overlap between Action and Adventure?	[Low, Medium, **High**]
22	Which two genres have the highest overlap?	[**Action**, **Adventure**, Children, Comedy, Crime, Documentary, Drama, Fantasy, Horror, Musical, Romance, SciFi]
23	Which two genres have a low overlap?	[Action, Adventure, Children, Comedy, Crime, **Documentary**, Drama, Fantasy, Horror, Musical, Romance, **Western**]

Table A.2: The evaluation questions.

Q.Nr	Questions description	Possible answers
24	With which genre do Comedy movies have the highest overlap?	[Action, Adventure, Children, Fantasy, Musical, War, **Romance**, Drama, Crime, Horror, Thriller, SciFi]
25	Which genre has the least overlaps with all other genres?	[Action, Adventure, Children, Mystery, Crime, **Documentary**, Drama, Fantasy, Horror, Musical, Nior, Western]
26	Which one of the following genres, whose movies are mostly exclusive to it?	[Action, Adventure, Fantasy, Drama, **Documentary**]
27	Which genre of the following has the largest number of degree 2?	[Drama, Action, **Romance**]
28	Which one of the following genres, whose movies are mostly shared with other genres?	[Thriller, **Action**, Drama]
29	Name a movie that belongs ONLY to Thriller genre.	Text answer
30	Name a movie that belongs to exactly two genres.	Text answer
31	Name a movie that belongs to AT MOST two genres.	Text answer
32	Name a movie that belongs to AT LEAST three genres.	Text answer
33	Name a movie that belongs to exactly 4 genres.	Text answer
34	Name a movie belongs to Action and to at most 2 other genres.	Text answer
35	Name a movie belongs to Drama and to at least 2 other genres.	Text answer
36	How many movies are from the Romance, Comedy and Drama?	[14, **34**, 51]
37	How many movies are from Romance, Horror or Comedy?	[**1766**, 1801, 1867]
38	How many movies are from Romance and Comedy but not Drama?	[93, **170**, 181]
39	How many movies are either Romance or Drama but not both?	[1194, 1372, **1666**]
40	How many movies are both Action and Drama but nothing else?	[22, **39**, 46]

Table A.3: The evaluation questions.

Q.Nr	Questions description	Possible answers
41	What is the Median Release Date of the movies in Action?	[**1993**, 1997, 1998]
42	What is the Median Release Date of the exclusive movies in Action?	[1993, **1994**, 1998]
43	In the Sci-Fi genre, which movies are more recent the shared or the exclusive movies?	[Exclusive, **Shared**]
44	Which genre has the oldest exclusively movies?	[Adventure, Musical, War, Western, **Noir**, Horror, SciFi, Action, Romance, Drama, Thriller, Documentary]
45	Which genre/s has the most recent exclusive movies?	[Action, Adventure, Children, Comedy, **Romance**, Drama, **Documentary**, **Crime**, Thriller, SciFi, Fantasy, Horror]
46	Name a genre that tends to have a high average rating.	[Action, Adventure, Children, Fantasy, Musical, War, Drama, Documentary, Horror, Western, **Noir**, Crime]
47	Name a genre that tends to have a low average rating.	[Action, Adventure, Children, Comedy, Musical, Romance, Drama, Documentary, **Horror**, Western, Thriller, Crime]
48	Does the number of the watches increase when a movie has more genres?	[**Yes**, No]
49	What is the Median Release Date of the shared movies between Romance and Drama?	[**1995**, 1997, 1999]
50	The movies from Children and Musical tend to be?	[**Old**, Recent]
51	The movies from Action and Thriller tend to be?	[Old, **Recent**]
52	Movies form Horror and Sci-Fi have average rating that is:	[**Low**, High]
53	Movies form Drama and Romance have average rating that is:	[Low, **High**]

Table A.4: The evaluation questions.

A.2 Qualitative feedback questions

Q.Nr	Questions description	Possible answers
54	How did you find the tool?	[Very easy, Easy, Neutral, Hard, Very hard]
55	How intuitive was the representation of the sets as regions?	[Very clear, Clear, Neutral, Not clear, Not clear at all]
56	How intuitive was the representation of the elements as bars in the sets?	[Very clear, Clear, Neutral, Not clear, Not clear at all]
57	How intuitive was the representation of the overlaps as arcs between the regions?	[Very clear, Clear, Neutral, Not clear, Not clear at all]
58	How intuitive was the interaction with the tool (search, click)?	[Very easy, Easy, Neutral, Hard, Very hard]
59	How intuitive was applying the operations on sets using the tool (union, intersection)?	[Very easy, Easy, Neutral, Hard, Very hard]

Table A.5: The qualitative feedback questions.

A.3 Training questions

Q.Nr	Questions description
1	Click on the "Horror" region.
2	Click on the top bar in the "Horror" region.
3	Click on the arc that connects the "Comedy" region and the "Romance".
4	Click on the thickest arc.
5	Move the mouse pointer to the top bar in the" Drama" region. - Notice the number of the items in the most top bar.
6	Click on the "Drama" region, notice: - The number of the exclusive movies - The number of the shared movies with one another region
7	With how many genres does the "Documentary" genre overlap? - Name them?

Table A.6: The training questions.

APPENDIX B

XML file for EvalBench

In this section the XML file that has been created in order to load the tasks to Evalbench is listed. An example of each task type will be presented.

```
<?xml version="1.0" encoding="UTF-8"?>
<taskList>
   -<task id="TaskId1">
      <configuration></configuration>
      <task-description></task-description>
      <task-instruction>
        Select one of the following options.
      </task-instruction>
      <task-type>TaskTypeChoice</task-type>
      <questions>
        <choice-selection id="q1">
        <question-text>
          How many movies does the genre “Action” contain?
        </question-text>
          <correctAnswers>
            <correctAnswer>option3</correctAnswer>
          </correctAnswers>
          <maxChoices>1</maxChoices>
          <possibleAnswers>
            <possibleAnswer label="option1">
              <displayLabel>164</displayLabel>
            </possibleAnswer>
            <possibleAnswer label="option2">
              <displayLabel>205</displayLabel>
```

```
          </possibleAnswer>
          <possibleAnswer label="option3">
            <displayLabel>503</displayLabel>
          </possibleAnswer>
        </possibleAnswers>
      </choice-selection>
    </questions>
  </task>

 -<task id="TaskId2">
    <task-description></task-description>
    <task-instruction></task-instruction>
    <task-type>TaskTypeText</task-type>
    <configuration></configuration>
    <questions>
      <text-input id="q1">
        <question-text>
          Name one of them (Action’s movies)?
        </question-text>
        <correctAnswer>heat</correctAnswer>
        <regEx>[a-z]*</regEx>
        <singleLine>false</singleLine>
      </text-input>
    </questions>
  </task>

-<task id="TaskId18">
    <configuration></configuration>
    <task-description></task-description>
    <task-instruction>
       Select one of the following options.
    </task-instruction>
    <task-type>TaskTypeChoice</task-type>
    <questions>
      <choice-selection id="q1">
        <question-text>
          Name the genres, to which the movie “Twister” belongs?
        </question-text>
        <correctAnswers>
          <correctAnswer>option1</correctAnswer>
          <correctAnswer>option2</correctAnswer>
          <correctAnswer>option11</correctAnswer>
          <correctAnswer>option12</correctAnswer>
```

```
      </correctAnswers>
      <maxChoices>4</maxChoices>
      <possibleAnswers>
        <possibleAnswer label="option1">
          <displayLabel>Action</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option2">
          <displayLabel>Adventure</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option3">
          <displayLabel>Children</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option4">
          <displayLabel>Comedy</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option5">
          <displayLabel>Crime</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option6">
          <displayLabel>Documentary</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option7">
          <displayLabel>Drama</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option8">
          <displayLabel>Fantasy</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option9">
          <displayLabel>Horror</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option10">
          <displayLabel>Mystery</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option11">
          <displayLabel>Romance</displayLabel>
        </possibleAnswer>
        <possibleAnswer label="option12">
          <displayLabel>Thriller</displayLabel>
        </possibleAnswer>
        </possibleAnswers>
      </choice-selection>
    </questions>
</task>
```

```
-&lt;task id="TaskId54"&gt;
       &lt;task-description&gt;
         Task containing a Likert question.
       &lt;/task-description&gt;
       &lt;task-instruction&gt;
         Select any item between easy and hard.
         (Qualitative question: no correctness needed)
       &lt;/task-instruction&gt;
       &lt;task-type&gt;TaskTypeLikert&lt;/task-type&gt;
       &lt;configuration&gt;&lt;/configuration&gt;
       &lt;questions&gt;
         &lt;likert id="q1"&gt;
           &lt;question-text&gt;
             How did you find the tool?
           &lt;/question-text&gt;
           &lt;count-options&gt;5&lt;/count-options&gt;
           &lt;left-label&gt;Very easy&lt;/left-label&gt;
           &lt;right-label&gt;Very hard&lt;/right-label&gt;
         &lt;/likert&gt;
       &lt;/questions&gt;
     &lt;/task&gt;

&lt;/taskList&gt;
```

APPENDIX C

Questionnaire

The questionnaire was used to collect data about the users. User were asked to estimate their experience with visualization.

Participant Name (optional):	
Age:	
Gender :	[ ] Male, [ ] Female
Occupation:	
Color blindness?	[ ] Yes, [ ] No
Do you have experience with Information Visualization?	[ ] Yes, [ ] No
If yes, estimate your experience:	[ ] Beginner, [ ] Intermediate, [ ] Advanced

Table C.1: The questionnaire used in the experiment.