FACET-TIME: USING INFORMATION SEEKING STRATEGIES TO SUPPORT EXPLORATION OF TIME SERIES DATA

A Thesis
Submitted to the Faculty of Graduate Studies and Research
In Partial Fulfillment of the Requirements
For the Degree of
Master of Science
in
Computer Science
University of Regina

By
Haider Ali Butt
Regina, Saskatchewan
December 2019

Copyright © 2019: Haider Ali Butt
Haider Ali Butt, candidate for the degree of Master of Science in Computer Science, has presented a thesis titled, *Facet-Time: Using Information Seeking Strategies To Support Exploration of Time Series Data*, in an oral examination held on November 26, 2019. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner: Dr. Christopher Oriet, Department of Psychology

Co-Supervisor: Dr. Orland Hoeber, Department of Computer Science

Co-Supervisor: Dr. Howard Hamilton, Department of Computer Science

Committee Member: Dr. R. Brien Maguire, Department of Computer Science

Chair of Defense: Dr. Timothy Maciag, Software Systems Engineering
Abstract

The notion of time has important value in people’s lives. By analyzing time-oriented data, we can understand underlying trends in the data. Graphically representing the data is useful for analysis because a single visualization can represent much information. Some tools providing visualizations for time-series data are well supported for lookup tasks; however, their support for exploratory search tasks are minimal. In exploratory search tasks, users may not have a specific search goal or they may have ill-defined information needs. They usually learn by iterating through and interpreting search results and perform investigation over them before integrating the obtained information into their knowledge.

In this thesis, an analytical tool (Facet-Time) is proposed that supports users in analyzing temporal data with a specific focus on supporting exploratory search. Facet-Time utilizes information visualization and interactive techniques to help users in finding information they are seeking and satisfying their search goals. It allows users with ambiguous search goals to interactively gain information about the search results. Facet-Time provides information scent support to assist users in choosing their path of exploration while they are defining queries. It also provides the visualization of time-series data where users can analyze the data and also make comparisons of subsets of data.

User evaluations were conducted to validate the design variations and features of Facet-Time. The outcome of the evaluations gave mostly mixed results, i.e., some experimental measures have significant value for some variations of Facet-Time and some measures do not have much significance. Improvements were observed in three subjective measures (usefulness, ease of use, and satisfaction) when using the sparklines feature versus when using no sparklines feature.
Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. Howard Hamilton and Dr. Orland Hoeber for their encouragement, guidance, and continued support throughout my Master’s study. Dr. Orland Hoeber consistently steered me in the right direction whenever he thought I needed it. I feel grateful for working under their supervision.

I gratefully acknowledge the financial support provided during my studies by the Faculty of Graduate Studies and Research, the Department of Computer Science, Celero, and NSERC (via Engage, Engage Plus, and Discovery grants awarded to my supervisors).

My thanks also go to my research group friends for all their support during the time we were working, and for all the fun we have had together. Finally, I must express my very profound gratitude to my parents for providing me with constant support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.
# Table of Contents

Abstract i

Acknowledgments ii

List of Figures v

List of Tables xi

1 Introduction 1
   1.1 Problem Statement .................................................. 1
   1.2 Approach Overview .................................................. 2
   1.3 Thesis Organization .................................................. 3

2 Background and Related Work 4
   2.1 Theoretical Frameworks ............................................. 4
       2.1.1 Information Visualization .................................. 4
       2.1.2 Information-Seeking ......................................... 11
       2.1.3 Information Scent ............................................. 12
   2.2 Exploratory Search ................................................ 13
       2.2.1 Faceted Navigation .......................................... 15
   2.3 Time Series Data Visualization ................................ 17
   2.4 Sparklines .......................................................... 20
   2.5 Exploratory Data Analysis Using Visualization ............... 21
   2.6 Discussion .......................................................... 22
3  Facet-Time

3.1 Introduction .................................................. 24
3.2 Software Design ............................................. 26
  3.2.1 Faceted Navigation ...................................... 26
  3.2.2 Timeline View ........................................... 28
  3.2.3 Sparklines ............................................... 34
3.3 Architecture .................................................. 38
  3.3.1 User Interface ........................................... 38
  3.3.2 Search Engine ........................................... 41
  3.3.3 Database ................................................ 42
3.4 Example ...................................................... 43
3.5 Discussion .................................................... 52

4  Evaluation ..................................................... 54
4.1 Methodology .................................................. 55
  4.1.1 Research Questions ..................................... 56
  4.1.2 Independent and Dependent Variables ................. 56
  4.1.3 Tasks ................................................... 58
  4.1.4 Data Collection Methods ................................ 60
4.2 Experimental Procedures ..................................... 61
4.3 Participant Recruitment ..................................... 62
4.4 Data Analysis ................................................. 62
  4.4.1 Analysis of Effectiveness ............................... 63
  4.4.2 Analysis of Efficiency .................................. 63
  4.4.3 Analysis of Perceived Usefulness and Perceived Ease of Use ........................................... 64
  4.4.4 Analysis of Perceived Satisfaction .................... 64
  4.4.5 Statistical Significance .................................. 64
4.5 Results ....................................................... 65
  4.5.1 Participants .............................................. 65
4.5.2 Comparison A: No Sparklines versus Sparklines (with Single Query) .................................................. 67
4.5.3 Comparison B: No Sparklines versus Sparklines (with Multiple Queries) ............................................ 71
4.5.4 Comparison C: Single Query versus Multiple Queries (with No Sparklines) ........................................ 75
4.5.5 Comparison D: Single Query versus Multiple Queries (with Sparklines) ............................................. 80
4.6 Discussion ........................................................................................................................................... 84

5 Conclusions and Future Research ........................................................................................................ 88
  5.1 Primary Contributions ..................................................................................................................... 88
  5.2 Limitations ...................................................................................................................................... 89
  5.3 Future Work .................................................................................................................................... 90

A Research Ethics Board Approvals ........................................................................................................ 103
B Consent Form ....................................................................................................................................... 105
C Pre-Study Questionnaire .................................................................................................................... 109
D Tasks and Post-task Questionnaire ....................................................................................................... 112
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>A three-stage model of human visual information processing [81]</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>An example of pre-attentively processed elements [39, 74]</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>Examples of the Gestalt Principles of similarity, proximity, connectedness, and enclosure</td>
<td>9</td>
</tr>
<tr>
<td>2.4</td>
<td>The three distinct channels of the Opponent Process Theory of Colour [81]</td>
<td>9</td>
</tr>
<tr>
<td>2.5</td>
<td>The two chromatic axes and the one achromatic axis and their roles in the HSB colour space [72]</td>
<td>10</td>
</tr>
<tr>
<td>2.6</td>
<td>The standard model of information seeking [22]</td>
<td>12</td>
</tr>
<tr>
<td>2.7</td>
<td>The dynamic model of information seeking [9]</td>
<td>12</td>
</tr>
<tr>
<td>2.8</td>
<td>A result page of the ACM digital library</td>
<td>13</td>
</tr>
<tr>
<td>2.9</td>
<td>Search activities [60]</td>
<td>15</td>
</tr>
<tr>
<td>2.10</td>
<td>Example showing the facets (on the left side) displayed in Amazon’s book shopping website</td>
<td>16</td>
</tr>
<tr>
<td>2.11</td>
<td>Instant based time representation</td>
<td>18</td>
</tr>
<tr>
<td>2.12</td>
<td>Interval-based time representation</td>
<td>18</td>
</tr>
<tr>
<td>2.13</td>
<td>Example of multiple granularities of time series data. Consider, the smallest possible unit is one day. In granularity 1, each granule represents a continuous set of seven days and here the smallest possible unit is one week. In granularity 2, each granule represents fortnights (a set of two consecutive weeks): Image source [4]</td>
<td>19</td>
</tr>
</tbody>
</table>
2.15 Sparkline showing the seasonal variation of common pediatric respiratory viruses prevalence [68] ................................. 20

3.1 The interface of the Facet-Time software ................................. 25
3.2 The faceted navigation panel is highlighted in the interface ........ 27
3.3 Selection of Payment and Cards Services from the Product facet . . 29
3.4 Selection of Champlin-Reichert from the Client facet and Payment
and Cards Services from the Product facet ................................. 30
3.5 Selection of Champlin-Reichert and Doyle-Zemlak from the Client
facet, Fulltime from the Resource Type facet, and Payment and
Cards Services from the Product facet ................................. 31
3.6 The timeline panel is highlighted in the interface ...................... 32
3.7 Colour map of 12 distinct colours along with their hexadecimal
values: Image Source [38] ............................................. 33
3.8 Selection of Kamron Little from the Human Resource facet and
Ward Ltd from the Client facet ........................................... 35
3.9 After a new query is created, Kamron Little is selected from the
Human Resource facet and Standard Work Activities is selected
from the Client facet .................................................. 36
3.10 The sparkline feature is highlighted in the interface ................. 37
3.11 Selection of Devin Edwards from the Human Resource facet ...... 39
3.12 The architecture of Facet-Time ........................................ 40
3.13 Selection of Contract from the Resource Type facet ................. 44
3.14 Blue timeline is for contract-based work on the Marketing Plans
product, and green timeline is for contract-based work on the Standard
Work Activities ....................................................... 45
3.15 Blue timeline is for contract-based work on the Marketing Plans
product, and green timeline is for contract-based work on the Cash
Management product ................................................... 45
3.16 Selection of the Contract from Resource Type facet and the Marketing Plans from Product facet ......................... 46
3.17 List of products that Porter Miles was working on ............... 48
3.18 Timelines for three queries are represented by light green, green, and orange colours ........................................... 49
3.19 After digging down into the timeline from March 11 to March 25 using the zoom-in control .............................. 50
3.20 Select March 22 on the timeline using the zoom-in control to get the number of hours spent on each product ............... 51

4.1 Four variations of Facet-Time ........................................ 57
4.2 Between-Subjects Design - Different participants are assigned to different conditions corresponding to a variable. Source [23] ........ 59
4.3 Within-Subjects Design - The same participant tests all conditions corresponding to a variable. Source [23] .................... 59
4.4 The flow of the user study to perform the tasks .................... 59
4.5 The average correctness values obtained of Tasks 1 and 2 for No Sparklines and Sparklines variations with Single Query ........ 67
4.6 The average time (in minutes) to complete Tasks 1 and 2 for No Sparklines and Sparklines variations with Single Query ........ 68
4.7 The frequency of responses received for the perceived usefulness of the No Sparklines and Sparklines variations for Task 1 with Single Query ....................................................... 69
4.8 The frequency of responses received for the perceived usefulness of the No Sparklines and Sparklines variations for Task 2 with Single Query ....................................................... 69
4.9 The frequency of responses received for the perceived ease of use of the No Sparklines and Sparklines variations for Task 1 with Single Query ....................................................... 69
4.10 The frequency of responses received for the perceived ease of use of
the No Sparklines and Sparklines variations for Task 2 with Single
Query ................................................................. 69

4.11 The frequency of the responses received for the perceived satisfac-
tion of the No Sparklines and Sparklines variations for Task 1 with
Single Query .......................................................... 70

4.12 The frequency of the responses received for the perceived satisfac-
tion of the No Sparklines and Sparklines variations for Task 2 with
Single Query .......................................................... 70

4.13 The average correctness values obtained for Tasks 1 and 2 for No
Sparklines and Sparklines variations with Multiple Queries ........ 71

4.14 The average time (in minutes) to complete Tasks 1 and 2 for No
Sparklines and Sparklines variations with Multiple Queries ........ 72

4.15 The frequency of responses received for the perceived usefulness of
the No Sparklines and Sparklines variations for Task 1 with Multiple
Queries ................................................................. 73

4.16 The frequency of responses received for the perceived usefulness of
No Sparklines and Sparklines variations for Task 2 with Multiple
Queries ................................................................. 73

4.17 The frequency of responses received for the perceived ease of use of
the No Sparklines and Sparklines variations for Task 1 with Multiple
Queries ................................................................. 74

4.18 The frequency of responses received for the perceived ease of use of
the No Sparklines and Sparklines variations for Task 2 with Multiple
Queries ................................................................. 74

4.19 The frequency of the responses received for the perceived satisfac-
tion of the No Sparklines and Sparklines variations for Task 1 with
Multiple Queries .......................................................... 75
4.20 The frequency of the responses received for the perceived satisfaction of the No Sparklines and Sparklines variations for Task 2 with Multiple Queries ............... 75
4.21 The average correctness values obtained for Tasks 1 and 2 for Single Query and Multiple Queries variations with No Sparklines ............ 76
4.22 The average time (in minutes) to complete Tasks 1 and 2 for Single Query and Multiple Queries variations with No Sparklines ............ 77
4.23 The frequency of responses received for the perceived usefulness of the Single Query and Multiple Queries variations for Task 1 with No Sparklines .................................................. 78
4.24 The frequency of responses received for the perceived usefulness of Single Query and Multiple Queries variations for Task 2 with No Sparklines .................................................. 78
4.25 The frequency of responses received for the perceived ease of use of the Single Query and Multiple Queries variations for Task 1 with No Sparklines .................................................. 79
4.26 The frequency of responses received for the perceived ease of use of the Single Query and Multiple Queries variations for Task 2 with No Sparklines .................................................. 79
4.27 The frequency of responses for the perceived satisfaction of the Single Query and Multiple Queries variations for Task 1 with No Sparklines .................................................. 79
4.28 The frequency of responses for the perceived satisfaction of the Single Query and Multiple Queries variations for Task 2 with No Sparklines .................................................. 79
4.29 The average correctness values obtained for Tasks 1 and 2 for Single Query and Multiple Queries variations with Sparklines ............ 81
4.30 The average time (in minutes) to complete Tasks 1 and 2 for Single Query and Multiple Queries variations with Sparklines ............ 81
4.31 The frequency of responses received for the perceived usefulness of the Single Query and Multiple Queries variations for Task 1 with Sparklines .......................................................... 82
4.32 The frequency of responses received for the perceived usefulness of Single Query and Multiple Queries variations for Task 2 with Sparklines .......................................................... 82
4.33 The frequency of responses received for the perceived ease of use of the Single Query and Multiple Queries variations for Task 1 with Sparklines .......................................................... 83
4.34 The frequency of responses received for the perceived ease of use of the Single Query and Multiple Queries variations for Task 2 with Sparklines .......................................................... 83
4.35 The frequency of the responses received for the perceived satisfaction of the Single Query and Multiple Queries variations for Task 1 with Sparklines .......................................................... 84
4.36 The frequency of the responses received for the perceived satisfaction of the Single Query and Multiple Queries variations for Task 2 with Sparklines .......................................................... 84
List of Tables

2.1 A list of visual features that can be pre-attentively processed [81] 5
2.2 User intents and related interaction controls [4] .................... 22

4.1 A list of comparisons of interfaces for data analysis ............... 63
4.2 Level of education of participants ................................. 66
4.3 Time-series data knowledge of participants ........................ 66
4.4 Exploratory data analysis experience of participants ............ 66
4.5 Participants’ experience with using interactive filters ............ 66
4.6 Participants’ experience with visual representation of time-series data 66
4.7 Statistical analysis (ANOVA) of the correctness of the answers to
the tasks for No Sparklines and Sparklines variations with Single
Query ................................................................. 67
4.8 Statistical analysis (ANOVA) of time(in minutes) to complete the
tasks for No Sparklines and Sparklines variations with Single Query 68
4.9 Statistical analysis (ANOVA) of the responses for the perceived
usefulness for No Sparklines and Sparklines variations with Single
Query ................................................................. 69
4.10 Statistical analysis (ANOVA) of the responses for the perceived ease
of use for No Sparklines and Sparklines variations with Single Query 70
4.11 Statistical analysis (Mann–Whitney U test) of the responses for the
perceived satisfaction for No Sparklines and Sparklines variations
with Single Query ..........................
4.12 Statistical analysis (ANOVA) of the correctness of the answers to
the tasks for No Sparklines and Sparklines variations with Multiple
Queries ........................................................... 72
4.13 Statistical analysis (ANOVA) of time (in minutes) to complete the
tasks for No Sparklines and Sparklines variations with Multiple
Queries ........................................................... 72
4.14 Statistical analysis (ANOVA) of the responses of the perceived use-
fulness for No Sparklines and Sparklines variations with Multiple
Queries ........................................................... 73
4.15 Statistical analysis (ANOVA) of the responses of the perceived ease
of use for No Sparklines and Sparklines variations .................. 74
4.16 Statistical analysis (Mann–Whitney U test) of the responses for the
perceived satisfaction for No Sparklines and Sparklines variations
with Multiple Queries ............................................ 75
4.17 Statistical analysis (ANOVA) of the correctness of the answers to
the tasks for Single Query and Multiple Queries variations with No
Sparklines ........................................................... 76
4.18 Statistical analysis (ANOVA) of time (in minutes) to complete the
tasks for Single Query and Multiple Queries variations with No
Sparklines ........................................................... 77
4.19 Statistical analysis (ANOVA) of the responses of the perceived use-
fulness for Single Query and Multiple Queries variations with No
Sparklines ........................................................... 78
4.20 Statistical analysis (ANOVA) of the responses of the perceived ease
of use for Single Query and Multiple Queries variations with No
Sparklines ........................................................... 79
4.21 Statistical analysis (Mann–Whitney U test) of the responses for
the perceived satisfaction for Single Query and Multiple Queries
variations with No Sparklines .................................... 80
4.22 Statistical analysis (ANOVA) of the correctness of the answers to the tasks for Single Query and Multiple Queries variations with Sparklines ........................................ 81

4.23 Statistical analysis (ANOVA) of time (in minutes) to complete the tasks for Single Query and Multiple Queries variations with Sparklines 82

4.24 Statistical analysis (ANOVA) of the responses of the perceived usefulness for Single Query and Multiple Queries variations with Sparklines ........................................ 82

4.25 Statistical analysis (ANOVA) of the responses of the perceived ease of use for Single Query and Multiple Queries variations with Sparklines ........................................ 83

4.26 Statistical analysis (Mann–Whitney U test) of the responses for the perceived satisfaction for Single Query and Multiple Queries variations with Sparklines ........................................ 84

4.27 Summary of the statistical analyses of Task 2 for all four comparisons. The tick-marks represent statistically significant differences. 85
Chapter 1

Introduction

The notion of time has essential value in many domains, including history, planning, project management, and science [4]. Analysis of time-oriented data helps us to learn from the past and enables us to plan for the future [4]. People have been doing time-series data analysis for centuries looking for patterns in the data. Existing computerized approaches work fine for focused and straightforward analysis of time-series data. However, if the analysis task itself is complex, then there is a need for analysis tools that support exploration.

Analysis of data appears to be similar to an information seeking process, i.e., users try to look for patterns in the data to understand what is in there. The focus of this research is to identify whether we can take methods and techniques out of information seeking and apply them to time-series data analysis to support exploration.

1.1 Problem Statement

Timelines are often used to help users understand the chronology of time and visualize spans of events. These representations can be used to analyze trends in past events and look for patterns in the data. Existing timeline visualization tools are well supported for simple and focused analysis tasks. These tools provide interaction techniques to support users in simple data analysis tasks.
When data analysis tasks are complex or users have ill-defined information needs, traditional timeline visualization tools are inadequate for data exploration. Such users require features beyond traditional timeline visualizations that can help them to explore time-series data. This research focuses on utilizing information-seeking strategies within time-series data to support exploration.

1.2 Approach Overview

Motivated by the problems stated above, this research aimed to improve users’ experiences when exploring the information within time-series data. In order to achieve the research goal, the following high-level research questions were formulated:

**RQ.a** How can we support exploratory search methods within time-series data analysis?

**RQ.b** What is the value of supporting exploratory search within time-series data analysis?

In order to answer RQ.a, a web-based analytical tool called Facet-Time was designed and developed. Facet-Time was developed by utilizing information-seeking methods along with time-series data visualization. It allows users to analyze temporal data with a specific focus on supporting exploratory search in the data.

The second research question (RQ.b) was answered through a user evaluation. The user evaluation was carried out to understand the value of the design decisions in the Facet-Time software. The user evaluation was conducted in a controlled laboratory environment. The methodology for the evaluation was designed around some specific research questions, which will be explained in Chapter 4.
1.3 Thesis Organization

The remainder of this thesis is organized as follows. In Chapter 2, the literature review of the existing studies related to this research is provided. In Chapter 3, the design and architecture details of the Facet-Time software are presented along with an example. The evaluation design, details of the controlled laboratory study, analysis techniques, and experimental results are explained in Chapter 4. In Chapter 5, the summary of the research contributions, limitations, and possible future work are discussed.
Chapter 2

Background and Related Work

This chapter situates the study described in this thesis in the context of previous research. First, a brief introduction to the theoretical frameworks that informed this research is presented. Discussions of exploratory search and faceted navigation are provided next. Then, existing studies about time series data visualization and data analysis are presented. Subsequently, a brief discussion on sparklines is given. Finally, this chapter concludes with a discussion about using information-seeking strategies to support the exploration of time series data.

2.1 Theoretical Frameworks

2.1.1 Information Visualization

People receive more information through vision than all of the other senses combined [81]. Information visualization is a discipline that deals with ways of visually representing abstract data to increase human cognition [81]. Information visualization presumes that visual representations take advantage of the link between human vision and the mind to allow users to see, analyze, and understand large amounts of information at once. It focuses on the creation of approaches for communicating information in intuitive ways.

Before designing visualizations for a system, it is essential to understand human
perception and the mechanism of visual information processing. *Perception* is the process used to identify and organize raw sensory information [13]. The brain then interprets that organized information using previous knowledge and understanding of the world to make sense out of it. In this whole process, users interpret the external world depending on internal factors, such as personality, motivation, and experience and external factors, such as size, contrast, motion, repetition, novelty, and familiarity of objects being perceived [81]. If the visual representation of an object does not align with users’ knowledge, then they may misinterpret the object [81].

*Visual information processing* is a series of steps that the human brain performs to interpret information from the outside world [62]. Visual information processing can be characterized as unfolding over three basic stages [81] (see Figure 2.1). In stage 1, the basic features of a visual object are processed (see Table 2.1). These features are also known as *pre-attentive features* since they can be detected rapidly by the human visual system without much mental effort [80]. In stage 2, the patterns present in the image are identified, and the visual scene is segmented into sections of different colour, texture, and motion patterns. Finally, in stage 3, a sequence of visual queries is answered through visual search strategies targeting goal-directed processing.

In the remainder of this section, the basic principles of information visualization are described: pre-attentive processing, Gestalt Principles, and the Opponent Process Theory of Colour.

<table>
<thead>
<tr>
<th>Visual Features</th>
<th>Added marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line orientation</td>
<td>Numerosity</td>
</tr>
<tr>
<td>Line length</td>
<td>2D position</td>
</tr>
<tr>
<td>Line width</td>
<td>Stereoscopic depth</td>
</tr>
<tr>
<td>Curvature</td>
<td>Convex</td>
</tr>
<tr>
<td>Hue</td>
<td></td>
</tr>
<tr>
<td>Intensity</td>
<td></td>
</tr>
<tr>
<td>Flicker</td>
<td></td>
</tr>
<tr>
<td>Direction of motion</td>
<td></td>
</tr>
<tr>
<td>Spatial grouping</td>
<td></td>
</tr>
<tr>
<td>Blur</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: A list of visual features that can be pre-attentively processed [81]
Figure 2.1: A three-stage model of human visual information processing [81]

Pre-attentive Processing

Pre-attentive processing is the first of the essential principles of information visualization. It refers to the subconscious accumulation of information by the human visual system from the environment [7]. Pre-attentive features can be identified with high accuracy in less than 250 ms [81].

Let us consider some examples of pre-attentive processing. In Figure 2.2a, there is a set of blue lines mixed with a single red line. Human vision can quickly identify the single red line from the other lines by looking at it. Here the visual property hue (colour) uniquely differentiates the red object from the other objects. Similarly, even though all objects have the same red colour in Figure 2.2b, the visual property curvature (shape) separates the circle from the squares. These examples show how objects that differ with respect to a single visual property may be noticed in a single glance if the difference is sufficiently large.

However, if multiple preattentive features are present in a single visualization and the differences among those features are too subtle, they cannot be preattentively processed, and the human brain will take time to identify the objects. So, careful attention should be paid to visual designs that support pre-attentive processing by avoiding such scenarios.
Figure 2.2: An example of pre-attentively processed elements [39, 74]

**Gestalt Principles**

The human brain tries to perceive objects as organized patterns because the mind has a natural disposition to perceive patterns in the stimulus based on specific rules. Psychologists have proposed theories to explain the various ways in which the human mind organizes visual elements into groups or associated entities. Based on these theories, a set of laws, named the Gestalt Laws, were developed [50]. The design principles that were created from the Gestalt Laws are called the *Gestalt Principles*. These principles are classified into the following categories: similarity, proximity, connectedness, and enclosure [10, 19, 26].

The principle of *similarity* states that objects that physically resemble each other are perceived as related. They are perceived as groups when they share the same attributes or same visual variables: shape, size, colour, texture, value, or orientation. For example, in Figure 2.3a, two different sets of circles can be found. One is a set of white circles and another a set of black circles. Here, the visual variable colour is used to identify the circles that are related (similar), which then leads the mind to perceive two groups of circles. While the principle of similarity helps to find objects that are related, it also helps to differentiate unrelated or different objects.

The principle of *proximity* states that elements placed close to one another are
perceived as being related. Even if their shapes, colour, and sizes are radically different, the objects will appear as a group if they are in close proximity. In Figure 2.3b, three groups of black circles can be seen because of the closeness to the circles in the same group and their distances from the circles in others groups.

The principle of connectedness states that elements joined to one another using lines are perceived as being related or in a group. In Figure 2.3c, the circles in the first, second, and third columns are connected using lines. These lines help the mind to perceive the first three columns as being related and interpreted as a group and to perceive the circles in the other columns as not being connected.

The principle of enclosure says that when elements are enclosed by lines or placed in a container, they are perceived as being related or in a group. In Figure 2.3d, the elements in the first two columns are enclosed in a bounding box and thus appear to be a group. Similarly, the elements in last two columns are enclosed by another bounding box and thus appear to be a separate group. Furthermore, a relationship can be perceived when objects are surrounded by an area shaded with colour.

**Opponent Process Theory of Colour**

Colours play a vital role in information visualization because they can be used to label, quantify, focus attention, and contribute to the interpretative sense of humans [73, 81]. The **Opponent Process Theory of Colour** states that three receptors with opposing actions control our ability to perceive colour [46]. These three receptors are translated into three channels or axes i.e., two chromatic axes, yellow-blue and red-green, and one achromatic black-white axis [86] (see Figure 2.4). The receptors allow the human mind to easily see the differences between the colours at the opposite ends; you do not see greenish-red because the opponent receptors can only detect one of these colours at a time.

These three distinct colour channels are used in the HSB (Hue-Saturation-Brightness) colour space (see Figure 2.5). In this colour space, hue is specified as
Figure 2.3: Examples of the Gestalt Principles of similarity, proximity, connectedness, and enclosure

Figure 2.4: The three distinct channels of the Opponent Process Theory of Colour [81]
an angular measure, starting at 0 degrees and ending at 360 degrees. It denotes the spectrum of colours using different angles, i.e. red starts at 0, yellow at 60, green at 120, and blue at 240 [37]. The radial distance from the intersection point of all three axes is used to measure saturation. It indicates the purity of colour, which ranges from 0% (no colour) to 100% (intense colour). Brightness is determined using the height of the achromatic axis, which can hold a value between 0% (black) to 100% (white).

Colour is utilized in the field of information visualization to distinguish one element from another [76]. In such cases, obtaining colours from the different channels, specified in the Opponent Process Theory of Colour, can help to produce a discrete set of colours. When dealing with categorical data, colours that are perceptually distinct from one another should be used in the design. For showing ordered data perceptually ordered colours can be used (e.g., increasingly light or dark colours). These type of principles can lead to better visual designs. Such designs can be compact, accurate, adequate for the purpose, easily understandable by users, and effective at communicating information. At the same time, care should be taken not to overload the interface with too many visual elements because they may lead to a cluttered design [29, 81].
2.1.2 Information-Seeking

The information-seeking process is all activities performed by humans and machines to find and obtain information [85]. It is important to understand the information-seeking process that people employ while exploring data, to provide support for good visual analysis and exploration. Numerous theoretical models have been proposed to explain the information seeking behaviour of humans. The most commonly discussed theoretical models are the standard model, the dynamic model, and sensemaking [40].

The standard model states that information seeking is a cyclic process consisting of identifying an information need, followed by the activities of query specification, examination of retrieved results, and if needed, repeating the cycle by reformulation of the query until a satisfactory result set is found [59]. One weakness of the standard model is that it assumes that users have a static information need and they try to refine a query until all and only those documents relevant to the original information need have been retrieved (Figure 2.6).

The dynamic model of information-seeking acknowledges that an information need may not be static. Instead, it can change as users interact with the search system and observe the results. Bates [9] proposed a dynamic model of information seeking named berry-picking. This model states that the searchers’ information needs, and consequently their queries, continually shift as they observe retrieved results (Figure 2.7). According to this model, information encountered at one point in a search may lead to further querying in a new, unanticipated direction.

Sensemaking is also an information seeking processes people go through to collect, organize and structure information to understand or solve a problem [69]. Sensemaking begins when the tasks to solve a problem are too big to handle within human cognitive limits, and so the process of collecting and organizing the information occurs. Sensemaking involves not only collection and organization of information but also requires learning about new domains, acquiring situational awareness of the problem, relating the situation with the previous knowledge about
it, and then trying to solve the problem [66]. According to Klein et al. sensemaking is a set of processes initiated when people realize that their current understanding of events is inadequate [49]. Pirolli and Russell have presented some examples of sensemaking activities in different domains [66]. For example, understanding features, costs, service plans, and trade-offs in consumer decision making (i.e., buying a computer); collecting, organizing, and comprehending information about a medical condition, and trade-offs to choose a treatment; or analyzing a subject’s domain to develop an efficient and effective training course.

### 2.1.3 Information Scent

In the early 1990s, information searching patterns and animal food foraging strategies were recognized as being similar, resulting in the emergence of information foraging [64]. This theory assumes that when searching for information, a searcher uses built-in foraging strategies, analogous to those that animals use to find food. The most crucial concept in information foraging theory is information scent. Information scent refers to the extent to which users can predict what they will find if they pursue a specific path in an environment [64]. As animals rely on scent to find their prey in the current space and utilize it to move towards other promising paths, so do humans rely on various hints in the environment to get desired information.
When a user is navigating within a visualization, to decide which information paths are promising to follow or which are better to leave, the visualization should present some clues about where to find useful information. Information scent provides users with cues and summarized information about content that is not immediately apparent [65]. Figure 2.8 shows a result page of the ACM Digital Library website, which provides an excellent example of information scent. It tells how many documents a user will encounter if he or she follows a particular path (i.e., if the user selects any of the given people).

2.2 Exploratory Search

Search is a fundamental activity in life or in using computers, and when we talk about tasks involved in the search, there are three kinds of searches; lookup, learn, and investigate [60]. Lookup search returns structured and discrete results such as numbers, names, short narratives, or specific files. Most people think of lookup searches as “fact retrieval” or “known item” searches. Lookup search and the standard model for information seeking appear to be similar ideas because in both
concepts the searcher tries to observe the search results by some known information. *Learning searching* involves multiple iterations and returns sets of objects that require cognitive processing and interpretation and often require the information seeker to spend time scanning/viewing, comparing, and making qualitative judgments [60]. As with learning search, *investigative search* involves multiple iterations but it takes place over possibly long periods of time and may return results that are critically assessed before being integrated into personal and professional knowledge bases [60].

Exploratory search is pertinent to learning and investigative search activities [60][63] (Figure 2.9). *Exploratory search* is an information seeking strategy [40] that represents the activities carried out by searchers who are unfamiliar with a domain, who may be unsure about the ways to achieve their goals, or who may have vague information needs [84]. It is a specialized form of information seeking, in terms of problem context or search strategies employed [27], in which people usually submit a tentative query to navigate nearer to relevant documents in the collection, then explore the environment to understand better how to exploit it, seeking cues about their next steps towards their goal [83]. The act of learning during the exploratory search process is characterized as knowledge gain, sense-making, or interpreting information, while the act of investigation lets users analyze, evaluate, or discover information [60].

Exploratory search requires several iterative steps to reach a complex search goal. In each step, users identify, analyze, and understand the relevant information they obtained from their current and previous steps to choose the next step along some path leading them to reach their search goals [9]. Users may sometimes deal with uncertainty with respect to their search goals because of their broad or ill-defined information needs [82]. The role of exploratory search systems is to help users with such goals to find relevant information that can satisfy their information needs [84].

In this research work, we use faceted navigation among other exploratory search
techniques, such as lookahead, relevance feedback, and interactive query refinement, to provide support for information seeking [54].

![Figure 2.9: Search activities [60]](image)

### 2.2.1 Faceted Navigation

Traditionally, the most popular strategy that users practice for data searching is described by standard model of information seeking [53]. It requires users to communicate with a machine, based on their preconceived information need. However, searching for the desired items can be time-consuming and exhausting with the standard model without any auxiliary facilities [87]. Problem structure, the complexity of the task, and prior knowledge have an interconnecting impact when searching [53]. Pertti said, “the more complex the task, the more ill-structured it is, and the less prior knowledge the actor has” [78]. Also, the standard model of information seeking does not explicitly support exploring connections that exist between data items [53].

*Faceted navigation* is a technique which allows users to narrow down search results by applying multiple filters based on various classifications of the data [77]. It is useful to explore data when there are a large number of objects that are well-structured along specific and meaningful categories [87]. We can think of facets as a set of categories, where each category corresponds to an individual dimension or property of the objects. Faceted navigation gives an abstract and overall view of the data. It addresses some weaknesses of the search models i.e., it allows exploration of associations between the data items. As users interact with facets
to specify queries, the data gets filtered out according to the selections that are made. Faceted navigation supports interactive information retrieval because it allows users to interactively define queries to search through a large collection of data [77]. It empowers users to elaborate their need progressively as they learn

Figure 2.10: Example showing the facets (on the left side) displayed in Amazon’s book shopping website
more about the topics [77]. It also allows users to have a feeling of control and understanding while exploring the data [40]. For example, the results of a query for purchasing a book online can be grouped under the labels (1) book category, (2) book format, (3) author, (4) release dates, (5) book series, and (6) customer reviews (see Figure 2.10), and all of these categories are facets.

Faceted navigation enables exploratory search [84]; specific facets may be evaluated and interactively selected to explore within the search results. It also allows users to view multiple perspectives of datasets and enables them to create queries interactively [89]. Although faceted navigation has clear benefits for end users, some challenges arise in the design of faceted navigation systems [77]. One of them is to determine where facets should be placed in a panel: to the left of the results, directly above the results, or below the results. Another challenge is to decide which strategy should be adopted for organizing facets, i.e., whether the order of presentation of facets should be kept static as the user navigates or the order should be based on ranking the facets according to their estimated utility. Selection of multiple values from a facet is also a thing to consider (i.e., should it be a disjunctive (OR) selection or a conjunctive (AND) selection?). Whichever approach we use, the interface should be self-consistent and adhere to familiar conventions. These are some key challenges to be faced while designing a faceted navigation system.

2.3 Time Series Data Visualization

The concept of time is of fundamental importance in many domains, including history, planning, project management, and science; and data that is inherently linked to time is called time-series data [4]. Time-series data represents a sequence of events, measuring the same thing over time e.g., historical events data, project planning data, or data gathered from a sensor over time [21]. A timeline is a visualization of a sequence of events or intervals in chronological order [4]. A
visualization is effective because it can present a wealth of information and facts, and can be processed by the human brain much more quickly than raw data. Timelines have been used for generations to communicate information about a series of events [35]. The goal of every timeline visualization is to effectively present information derived from the time series data and support users in data analysis, if required [21].

For any visualization, data is the main building block and types of data play a crucial role in designing the visualization. Time series data have two types: instant based time (also called point based time) and interval-based time [4]. Instant based time is based on a particular value of time (called a chronon) and can be represented as a single point, e.g., Aug 11, 1898 (Figure 2.11) represents a single day. In contrast, interval-based time is based on the notion of an interval, which can be specified by two instants that signify the beginning and end of the interval. Graphically, an interval can be represented by a line connecting both instants. For example Figure 2.12 represents an interval from September 13, 2004 to September 19, 2004. Another important thing in a timeline visualization is granularity. Granularity can be thought of as an abstraction of time (such as minutes, hours, days, weeks, months) used in order to make it easier to deal with time in every-day life [4]. Bettini et al. describe granularity as mappings from time values to larger or smaller conceptual units [15]. Low granularity results in an aggregated representation of the data, whereas high granularity leads to a detailed representation of the data (see Figure 2.13 for examples of time granularities).
Consider, the smallest possible unit is one day. In granularity 1, each granule represents a continuous set of seven days and here the smallest possible unit is one week. In granularity 2, each granule represents fortights (a set of two consecutive weeks): Image source [4]

Brehmer et al. described several representations of the timeline, including linear, radial, grid, spiral, and arbitrary-based representations [21]. The names of these representations refer to the overall form or shape of the path that time takes in a visualization (Figure 2.14). TimeLineCurator [33] presents a linear representation of a timeline that visualizes instant based data as well as interval-based data. Lensing Wikipedia [43] uses another representation of a linear timeline to show interval events and their overlaps over time. A radial representation is appropriate for presenting the periodic nature of time, such as cyclical weather information, daily routines of people, and yearly changes in ecosystems [21]. Grid-based representations are Month-Week-Day calendar based representations. This kind of representation is appropriate for presenting the cyclical (weekly) nature and high level of granularity of time; DateLens [11] is an example of grid-based representation. The spiral timeline is well suited to visualizing serial periodic time-oriented data, i.e., data that has a serial (continuous) dimension that exhibits periodicity. Diego et al. [3] used a spiral timeline in the design of an interactive visualization tool named Semantic Spiral Timelines (SST), which is intended to facilitate the exploration and analysis of academic information stored in e-learning platforms. Time Curves [8] uses an arbitrary timeline representation for visualizing diverse types of patterns in temporal data (e.g. large sudden changes or reversals of previous states).
2.4 Sparklines

A sparkline is a small line chart drawn without axes or coordinates [75]. It shows the general trend or the variation in the data. A sparkline can serve as a straightforward but useful visualization that simplifies cognitive tasks and highlights trends in the data [68]. Users can visually observe the overall trend with the help of this tiny visualization rather than manually exploring every aspect of the data. In the literature, the use of sparklines has been observed in many domains. They have been used in digital financial reports to assist novice investors [61]. Researchers have also used sparklines to represent medical data to reduce diagnostic errors [68] (Figure 2.15). A sparkline can also be embedded in text because of its small size and it can summarize the information from hundreds of data points in the space of a word or two [32].

Sparklines are ideal in situations where textual data representations obscure interpretation [68]. It has been observed that small visualizations with text help readers to memorize things during a brief lookup task [34]. Sparklines are potentially useful to highlight trends, simplify cognitive tasks, and provide the context of different aspects of the data [68].
2.5 Exploratory Data Analysis Using Visualization

Idreos et al. state, “Data exploration is about efficiently extracting knowledge from data even if we do not know exactly what we are looking for” [45]. Data exploration using visualization is useful when little is known about the data and the exploration goals are vague [47]. A visualization must support finding answers to the various questions that can likely arise while doing data exploration; Andrienko et al. referred to these questions as data exploration and analysis tasks [6]. A task consists of two parts, the target (i.e., a description of the unknown information that needs to be obtained) and the constraints (i.e., a specification of the known information). The latter are related to the target in a certain way and limit the set of items of information that are suitable as answers. For example, a task might require a direct comparison (i.e., determining the relations between different attributes in the data) or doing relation-seeking (i.e., determining whether a change in the value of one attribute causes a change in the value of another attribute). The main value of data exploration is that it allows users to satisfy their analytical needs and give direction to further reasoning [4].

A static visual representation alone is not enough to explore the data or to satisfy the analytical needs of users [80]. By providing users with an interaction mechanism as well as a good graphical representation, a visualization system can provide convenience in the data exploration. By supporting the interaction controls in a visualization, a system allows users to ask their questions explicitly (i.e., by specifying the target and the set of the constraints) and, in response, provides the required information. These controls can help users to reconfigure the visualization and explore the information according to their needs. Various interaction techniques can be embedded in visualizations to support several intents of users while they are exploring the data [4]. These intents describe what a visualization should support in terms of interaction in order to take full advantage of human perception and the machine’s capabilities. Visual exploration and analysis of time-oriented data require interaction methods that allow users to
<table>
<thead>
<tr>
<th>User Intent</th>
<th>Interaction Control</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark something as interesting</td>
<td>Select</td>
<td>TimeSearcher 3 [24] supports the selection of time intervals for similarity searches in the data.</td>
</tr>
<tr>
<td>Show something else</td>
<td>Explore</td>
<td>BinX [14] supports the exploration of different aggregations of a time-series. An aggregation is based on constructing bins where each bin holds a user-defined number of time points.</td>
</tr>
<tr>
<td>Show something conditionally</td>
<td>Filter</td>
<td>TimeFinder [42] provides a filter control timebox, rectangular regions that are placed and directly manipulated on a timeline to help users to specify a query visually and to filter the data.</td>
</tr>
<tr>
<td>Show related items</td>
<td>Connect</td>
<td>Continuum [5] is a timeline visualization tool suited to representing large amounts of hierarchically structured temporal data and their relationships. It provides connection lines for representing non-hierarchical relationships.</td>
</tr>
<tr>
<td>Show more or less detail</td>
<td>Abstract / Elaborate</td>
<td>FacetZoom [28] uses an interactive widget with a time axis to access data from different parts of the time domain at different levels of abstraction.</td>
</tr>
</tbody>
</table>

Table 2.2: User intents and related interaction controls [4]

manipulate the visual representation in different ways, including the selection of data of interest, filtering out irrelevant data, visualizing the connection between data elements, and many more [4]. Table 2.2 represents some intents that users could have while doing data analysis and the corresponding interaction controls that could be embedded in visualizations for the data exploration [88].

### 2.6 Discussion

All the existing tools for timeline visualization discussed in Section 2.3 provide interactive features for data exploration. Although these tools offer useful timeline
visualizations for lookup-based data search, they may not be suitable for satisfying the information needs of users whose search goals are ill-defined [52]. Such users wish to explore data by learning and investigating the results for which they perform multiple iterations of refining the query.

None of the tools described in the literature to date enable any data exploration support within timeline visualizations, which could help users explore and analyze time-series data. This research takes advantage of information-seeking strategies within time-series data exploration tasks. It adopts the benefits of the exploratory search model to allow users to explore time-series data interactively. Faceted navigation is a crucial enabler to support exploratory search [77]. Using faceted navigation, users learn more about the topic as they explore results, and formulate new questions as they learn from previously asked questions [40]. In Chapter 3, we describe a faceted navigation strategy well suited to timeline visualization.
Chapter 3

Facet-Time

In this chapter, the design and development of the Facet-Time software tool are presented. First, a brief introduction about the goals of Facet-Time is given. Next, the design of the Facet-Time interface along with its four variations, is presented. Subsequently, a description of its architecture is provided, followed by an explanation of the methods used for data access from MongoDB. Then, an example is presented to explain the various functionalities of the software. At the end of this chapter, the design decisions behind Facet-Time are discussed.

3.1 Introduction

Facet-Time is an analytical tool that supports users analyzing temporal data with a specific focus on supporting exploratory search in the data. It has been designed to support searchers’ knowledge discovery, temporal analysis, and comparison of subsets of the data. This tool encourages interactive information-seeking behaviour.

When users search for information or perform visual analysis, they utilize their general background knowledge to seek the information [69]. There is a chance that they might not be able to explore all aspects of data because of their ill-defined needs. This research focuses on helping users explore diverse aspects of time-series data utilizing information-seeking strategies. It uses an exploratory search
Figure 3.1: The interface of the Facet-Time software
Facet-Time utilizes faceted navigation adapted to a timeline (Figure 3.1). Faceted navigation enables users to explore multiple perspectives of datasets [77]. As users make selections in facets to define queries, the data get filtered according to those selections, allowing the users to create queries interactively for further analysis of data on the timeline.

For this research work, we are using a project management dataset of a company. Because of the privacy concerns, the data has been anonymized. The dataset has information about the products, human resources who work on those products, and the clients. It has a variety of types of data, i.e., categorical, numerical, and temporal. Based on different categories (i.e., products, human-resources, and clients), the categorical attributes have been utilized to produce faceted navigation. Temporal attributes have been used to draw the timeline.

### 3.2 Software Design

Facet-Time takes advantage of faceted navigation, timeline, and sparklines to provide support for visual exploration and analysis of time-series data. The details of these components are explained in the remainder of this section.

#### 3.2.1 Faceted Navigation

The core feature of Facet-Time is faceted navigation (Figure 3.2), which supports users while they specify their queries progressively. Whenever a user makes any selection in any facet, items which are not related to the current selection are removed, and only relevant items are presented. If the user makes selection across the facets an AND operation is performed on selected values. Whereas, if the selection is within a facet an OR operation is performed.

Suppose a user wants to see the human resources who have worked on the "Payment and Cards Services" product. By selecting that product from the Product
Figure 3.2: The faceted navigation panel is highlighted in the interface
facet, it will filter all the human resources in the Human Resources facet, and only those who have worked on this product will appear in the list (Figure 3.3). In the figure, one can also see the list of clients to whom this service has been provided. Now let’s say that out of those human resources, the user wants to see only those who have worked for a client named Champlin-Reichert. After the user selects Champlin-Reichert from the Client facet, all other facets will be refreshed. The user can now see the list of all human resources who have worked with Champlin-Reichert on Payment and Cards Services (Figure 3.4).

A user can also make multiple selections within a facet. Suppose, in the previous example, the user is interested in seeing those human resources who work full time and they have worked for Champlin-Reichert or another client named Doyle-Zemlak. As the user makes selections in the required facets, all other facets will be refreshed accordingly. Now the user can see the list of all human resources who have worked with either Champlin-Reichert or Doyle-Zemlak as full time resources on Payment and Cards Services (Figure 3.5).

3.2.2 Timeline View

The second main feature of Facet-Time tool is the timeline view to represent the value of an attribute of interest over time. In this case, the attribute of interest is the number of hours spent on products over time. Figure 3.6 shows the timeline view of the tool where the y-scale represents the number of hours worked and the x-scale represents the dates. The unit of time granularity of x-scale is a day, i.e., each granule represents a single day. The timeline also has a zoom-in/zoom-out control to obtain a detailed or an abstract view, respectively. Whenever users make selections in any of the given facets, the timeline is updated according to the selections made.

Figure 3.6 shows the timeline panel which represents a timeline for the overall number of hours spent by all human resources over time. Now let us say a user wants to analyze the timeline when “Kamron Little” was working with “Ward Ltd”. 
Figure 3.3: Selection of Payment and Cards Services from the Product facet
Figure 3.4: Selection of Champlin-Reichert from the Client facet and Payment and Cards Services from the Product facet
Figure 3.5: Selection of Champlin-Reichert and Doyle-Zemlak from the Client facet, Fulltime from the Resource Type facet, and Payment and Cards Services from the Product facet
Figure 3.6: The timeline panel is highlighted in the interface
As the user makes these selections in the Human Resources and Client facets, the timeline is updated accordingly (Figure 3.8). From the timeline, the user can observe that the only time that Kamron Little worked for Ward Ltd was in the month of March.

Another useful feature of the tool is that users can generate multiple timelines corresponding to different queries and then perform a visual comparison of these timelines with respect to the subsets of the data. An online tool ColorBrewer [38] has been used to choose distinct colours to draw multiple timelines. ColorBrewer provided 12 distinct colours, which have been used to draw multiple timelines (Figure 3.7).

Continuing from the example in Figure 3.8, suppose the user wants to compare the working hours spent by Kamron Little with Ward Ltd to those spent on Standard Work Activities. To do so, the user creates a new query by clicking the “New Query” button and then making the required selections in the facets to specify another timeline. Now on the timeline the user can visually compare the results of the two queries. The user can see that Kamron Little worked with Ward
Ltd as a client during March whereas he worked on Standard Work Activities mostly in May and to a lesser extent in April (Figure 3.9).

### 3.2.3 Sparklines

Another feature of the Facet-Time tool is the use of sparklines, one for each facet value (Figure 3.10). A sparkline represents the number of working hours for the associated facet value over the time of February 25, 2018 to May 31, 2018. The time granularity for sparklines is a day. The sparklines give a sense of the distribution of working hours for the facet values over time to provide information scent to users.

In every facet, the minimum value for the y-scale of sparklines is the same, i.e., zero. The maximum value for the y-scale of the sparklines in a facet can vary between facets. The highest value for any facet value determines the maximum value in a given facet. An attractive property of these sparklines is they are dynamic (i.e., whenever a user makes a selection in any of the facets, these sparklines change and present themselves according to the selected facet value(s)) and hence their maximum value also changes.

The colour of the sparklines in a facet is the same for all facet values but it is different from the colour of the sparklines in other facets. Different colours are used to indicate that the sparklines are conceptually different, i.e., the sparklines in every facet may have a different maximum value. Otherwise, users might perceive them as refering to the same thing if they had the same colours (Gestalt Principle of similarity). This use of colour reinforces the idea that users should only compare the sparklines in a single facet.

Let us say a user wants to analyze the distribution of hours of “Devin Edwards” (Human Resource) on different projects. The user selects Devin Edwards from the Human Resource facet and all other facet values change along with the sparklines (Figure 3.11). The user can perceive from the sparklines that Devin has some overlap of work on the Cash management project and the Payment and
Figure 3.8: Selection of Kamron Little from the Human Resource facet and Ward Ltd from the Client facet
Figure 3.9: After a new query is created, Kamron Little is selected from the Human Resource facet and Standard Work Activities is selected from the Client facet.
Figure 3.10: The sparkline feature is highlighted in the interface
Cards Services project during some initial time. Then during the middle of time, he stopped working on both projects and started working on the Standard Work Activities. And at the end, Devin stopped working on the Standard Work Activities and again started working on the Cash management project having a little overlap with the Standard Work Activities later.

### 3.3 Architecture

Facet-Time is web-based software that consists of loosely coupled components and follows a distributed client-server architecture. Its architecture is designed to allow users to have a smooth interaction with the software. Figure 3.12 depicts the architecture of Facet-Time, along with its process flow. The main components of this architecture are the user interface and the search engine. These components are explained in detail in this section.

#### 3.3.1 User Interface

The user interface of Facet-Time is implemented to work on the web. The user interface component was developed in HTML, CSS, JavaScript, and AngularJS. The D3 (Data-Driven Documents) library [20], which is implemented in JavaScript, was used to draw the timeline and sparklines.

**Component-Based Architecture**

The implementation of the user interface follows a component-based structure [41]. In AngularJS, the component-based architecture helps to design an application by integrating functional or logical components that represent well-defined communication interfaces containing methods, events, and properties. Every component follows the Model-View-Controller (MVC) architectural pattern [51]. This pattern allows components to be divided into three interconnected objects: model, view, and controller. Model objects are used for holding the application’s data and for
Figure 3.11: Selection of Devin Edwards from the Human Resource facet.
Figure 3.12: The architecture of Facet-Time

defining the logic that manipulates that data. View objects deal with ways of displaying data to users and handle user input and interaction with data. Controller objects are intermediary between view and model objects. A controller object updates a view object whenever data in the model object changes and updates a model object when there is a user-initiated change in the view object.

Interaction with Facets

The view object receives all interactions in the user interface. The first activity a user can perform with Facet-Time is to create a query. The facet visualization component is a view object which receives and responds to these interactions, and then sends a request to the controller. The service handler will then send a request to the search engine. After the controller receives a response from the search engine, it will send the response to the facets visualization and timeline visualization components to display particular facet values and the timeline.

Interaction with Timeline

Apart from the interaction with facets, there are other interactions that users can perform in the user interface. They can interact with a zoom-in/zoom-out control in the timeline to obtain a detailed or an abstract view of the timeline, respectively. The timeline visualization component handles any interaction in the
timeline. The service handler will then send a request to the search engine. After the controller receives a response from the search engine, it will send the response to the facet visualization component to update the sparklines and to the timeline visualization component to display the results of all queries defined by the user.

3.3.2 Search Engine

The data processing methods for Facet-Time are implemented on a search engine that runs on a server. Data processing is performed on the server to avoid overburdening the client-side and also to provide an abstraction of the underlying implementation and expose only objects or actions the developer of the user interface needs. The server is configured with an Intel Xeon 2.30 GHz processor, 8 GB memory, and runs on the Linux operating system version 7.2 (Nitrogen).

REST API Module

The controller in the user interface communicates with the search engine using Representational State Transfer (REST) services [31]. REST is an architectural style used in large-scale networked software that takes advantage of technologies and protocols of the web [31]. This style defines constraints that focus on the roles of components and their interaction with other components within the architecture.

The search engine uses POST methods of REST services for its information exchange because POST methods have no restrictions on data length and they do not expose query parameters in the URL. Here POST methods are used to retrieve the required response from the database and pass it to the controller in the user interface.

In Facet-Time, REST services were implemented using a framework called Spark [2]. Spark is a rapid development framework for Java that is built around Java 8 lambda expressions philosophy. A lambda expression is a function definition that is not bound to an identifier or name [30]. The use of lambda expressions
makes code less verbose than it would be if written in other Java frameworks. Facet-Time was developed using Spark, version 2.7.2.

Once the user interface is loaded or users submit a query in the interface, several REST requests are sent from the user interface to the search engine. Four REST requests are sent for fetching lists of facets (as we have four facets); their results are sent to the facet visualization component. Another request is sent by the user interface to the search engine to retrieve the temporal information, which is then sent to the timeline visualization component.

Data Access Module

The data access module is primarily responsible for communicating with the database to retrieve data for every REST API. All APIs interact with this module by passing the search query to get the matching results. The data access module initiates MongoClient and MongoDatabase objects (MongoDB classes for Java) which then communicate with the MongoDB database. Then, upon receiving the request, MongoDB searches against the query and returns the matching result to the data access module. After receiving replies for the query, the data access module sends the search results to the REST API module.

3.3.3 Database

Facet-Time uses a MongoDB version 4.0.0 database [1]. MongoDB is an Open-Source and cross-platform NoSQL database that provides high performance, high availability, and high scalability. MongoDB does not use tables like a traditional relational database; instead it stores data in the form of documents with dynamic schemas. Also, it avoids the use of joins which are frequently performed in a relational database, and provides indexing of data. These features facilitate the integration of new data and the retrieval of query results.
3.4 Example

In this section, the information-seeking behavior of a user using Facet-Time is described. It is further supported by screenshots of the software to correlate the users’ tasks with the software when performing the tasks.

Let us assume that a user, James, who is a product manager for a small software development company, has some concerns about the amount of contract work being done on his product (Marketing Plans) in comparison to other products in the company. He has laid out a general search strategy using Facet-Time to find the answer to this question, which consists of the following steps.

First, James wants to find products that have contract-based work done. He starts his search by selecting Contract from the Resource Type facet (see Figure 3.13). After he makes this selection, all other facets get updated, and he can check the Product facet to observe the products that have contract-based work done.

Next James plans to analyze whether the pattern of contract work of his Marketing Plans project is similar to the patterns of other contract-based projects, or different. First, he starts comparing the sparklines of all other products in Figure 3.13 with the sparkline of the Marketing Plans product. James observes that the values in the sparklines along y-axis are higher for the Payment and Cards Services product than for the Marketing Plans product, so their pattern of contract work is not similar. The sparkline for Standard Work Activities looks somewhat similar to the one for Marketing Plans product. For the Web Page Analysis product, James comprehends that the first part of the sparkline is a flat line, which means no work was done during that time. So from the sparklines, James can interpret that the pattern of work done on the Web Page Analysis product is not similar to that done for the Marketing Plans product. After comparing the sparklines of all products, James finds two products (Standard Work Activities and Cash Management), which have a sparkline that looks similar to the one for the Marketing Plans product in terms of the pattern of work.
Figure 3.13: Selection of Contract from the Resource Type facet
To further analyze and compare the patterns of these two products with the Marketing Plans product, James creates multiple queries to draw timelines for these products. Figure 3.14 shows the timeline for contract-based work on the Marketing Plans product in comparison to the Standard Work Activities, whereas Figure 3.15 shows the timelines for contract-based work on the Marketing Plans and the Cash Management products. James can observe that the pattern of work on Standard Work Activities is more similar to the pattern for the Marketing Plans product than for the Cash Management product.

To further explore, James is interested to know whether or not his contractors who were working on Marketing Plans product were also working on other prod-
Figure 3.16: Selection of the Contract from Resource Type facet and the Marketing Plans from Product facet
ucts of the company. First, James needs to know the identities of the contractors working on Marketing Plans product. James creates a new query, hides the previous two queries, and then selects Contract from Resource Type and Marketing Plans from products. After making those selections, James can observe from the Human Resources facet that there is only one contractor, Porter Miles, who was working on the Marketing Plans product (Figure 3.16).

Now to see whether Porter Miles was working on other products or not, James selects Porter Miles from the Human Resource facet and unselects the two previous selections (i.e., Contract from Resource Type and Marketing Plans from products). In the Product facet, James can see that there are two products other than Marketing Plans that Porter Miles was working on (Figure 3.17).

As James observed that Porter Miles was also working on two other projects, he is now interested to see whether there was any time when Porter Miles was working on all three products simultaneously. In Figure 3.17, James can observe from the sparklines for all three products that there is some overlap of work in the initial period. To further analyze the situation, James creates three different queries to compare the timelines for the three products (Figure 3.18). James observes that the timelines for all three products overlap somewhere from March 11 to March 25. Using the zoom-in control, James digs down in the timeline to focus March 11 to March 25, where he observes that on March 22, Porter Miles was working on all three products simultaneously (Figure 3.19).

Lastly, James is interested to know the number of hours Porter Miles spent on each product on March 22. For this purpose, as he zooms-in to March 22 on the timeline, he can observe the number beside each sparkline. In the Product facet, James can see that on March 22 Porter Miles spent 4.50 hours on Marketing Plans, 2.0 hours on Develop Advanced Analytics Strategies, and 1.0 hour on Bank Accounts Management (Figure 3.20).

In this example, the user (James) had underspecified information needs. To find the answer to his vague needs, he started with a general search strategy. Then
Figure 3.17: List of products that Porter Miles was working on
Figure 3.18: Timelines for three queries are represented by light green, green, and orange colours
Figure 3.19: After digging down into the timeline from March 11 to March 25 using the zoom-in control
Figure 3.20: Select March 22 on the timeline using the zoom-in control to get the number of hours spent on each product.
along the way of searching, James analyzed the results and learned from them. That helped guide his exploration activities.

3.5 Discussion

Time-series data is a common data type encountered in many domains. Analysis of time-series data helps us understand the underlying trends in the data points. Sometimes because of the ill-defined needs of users or the complexity of the search tasks, the data analysis become difficult. To address this problem, we need some exploratory search strategy. There are tools for timeline visualization, which are well supported for lookup tasks; their support for exploratory search tasks is minimal.

The research presents Facet-Time, web-based software that supports exploring and analyzing the time-series data. Some theoretical frameworks guided the design of Facet-Time, i.e., information visualization [81], pre-attentive processing [7], Gestalt Principles [50], colour theory [72], information seeking models [85], exploratory search [60], and information foraging theory [64].

Facet-Time utilizes faceted navigation [77], an exploratory search technique, that helps users to fulfill their information needs by progressively defining queries as they explore and learn more from search results. As users interact with facets to specify queries, the data gets filtered out according to the selections that are made and enables users to navigate information along multiple paths. Facet-Time also provides sparklines [75] along with faceted navigation. The sparklines give a sense of the distribution (information scent [64]) of facet values over time and assists users in choosing their path of exploration. To indicate that the sparklines are conceptually different, different colours have been used (Gestalt Principle of similarity).

Further, a timeline is provided to visualize the time-series data. A user can also draw multiple timelines corresponding to different queries and then perform
a visual comparison of these timelines. To distinguish each timeline from other
timelines a colour encoding scheme is used by following colour theory. This scheme,
based on the opponent process theory of colour [81], generates perceptually distinct
colours that are used to recognize each timeline uniquely. The Gestalt Principle of
similarity was also followed to show the correspondence between the timelines and
their respective legends. By visualizing various information, interactively creating
queries, and providing visual comparison of the subset of data users can explore
the dataset, understand different patterns in it, and find relevant information.

Overall, Facet-Time is designed to allow users to explore and analyze time-
series data using various automated methods, visual, and interactive techniques.
These can help users to achieve their goals (i.e., explore and learn about time-series
data) when they have under-specified needs.
Chapter 4

Evaluation

An evaluation is performed as an essential step to obtain feedback about the quality or value of a system and to highlight the changes required in the future design of that system [71]. Every system should undergo an evaluation process to validate its design choices.

Several evaluation methods exist to validate design decisions of a system [12] [17] [25]. The evaluation of interactive information retrieval (IIR) systems differs from the evaluation of traditional information retrieval (IR) systems [48]. The evaluation of a traditional IR system concentrates on the software without considering its users, whereas the evaluation of an IIR system considers both the software and its users with a particular focus on their behaviours, experience, and interactions with the system [48].

As discussed in Chapter 3, Facet-Time is an interactive data exploration tool for analyzing temporal data with a specific focus on supporting exploratory search within the data. Facet-Time utilizes sparklines along with faceted navigation to help users understand the distribution of each facet value over time. It also allows users to create queries and draw multiple timelines, corresponding to those queries, as a means of performing a visual comparison of subsets of the data. The purpose of this study is to assess variations of the proposed system against one another, i.e., compare the interface with sparklines to the interface with no
sparklines (keeping all other conditions the same for both interfaces), and compare the interface with multiple timelines to the interface with a single timeline (keeping all other conditions the same for both interfaces). This study will measure the effectiveness, efficiency, perceived usefulness, perceived ease of use, and perceived satisfaction of four possible variations of a design interface for exploration of time series data, which should permit the identification of the features of the interfaces that enable data analysts to discover the information they are seeking.

This chapter starts with a detailed description of the research methodology, which includes the specific research questions, the variables (independent and dependent), the tasks used in the experiment, and the data collection instruments. Then, the method used to assign tasks to participants is presented. Next, the experimental procedure and the participant recruitment process are described. Subsequently, the methods used for analyzing the collected data are explained. Finally, this chapter concludes with the results of the experiments and a summary of the key findings of the evaluation.

4.1 Methodology

One of the most commonly used methods of evaluating an IIR system is the controlled laboratory study [48]. A controlled laboratory study is a study conducted in an environment specifically designed for research that allows researchers to manipulate and observe the impact of various variables used in the experiment and to minimize the effects of other external factors [25].

In a controlled laboratory study, the researcher manipulates independent variables to observe changes (if any) in the dependent variables [44]. The values of independent variables reflect the experimental conditions that are controlled by the experimenters. Dependent variables are outcome variables that are measured during the experiment. In other words, the values of the dependent variables show the effects of the manipulation of the values of the independent variables.
Since the evaluation for this research was conducted as a controlled laboratory study the evaluation methodology involved developing specific research questions, identifying and controlling independent variables, measuring dependent variables, utilizing statistics methods, and answering the specific research questions. These steps are further explained in the remainder of this section.

4.1.1 Research Questions

Below are given the specific research questions that drove the evaluation. These questions compare the value of having a feature in the system as compared to not having it. For such questions, the value refers to effectiveness, efficiency, perceived usefulness, perceived ease of use, and perceived satisfaction.

RQ1 What is the value of adding sparklines into the facets as compared to not adding them?

RQ2 What is the value of providing multiple query visualizations of the data in contrast to providing only a single-query visualization?

4.1.2 Independent and Dependent Variables

The research questions in Section 4.1.1 imply that there are two independent variables, i.e., sparkline and query. For the sparkline variable, we have two possible values, i.e., "Sparklines" and "No Sparklines". Similarly, for the query variable, we have two possible values, i.e., "Single Query" and "Multiple Queries".

Because there are two independent variables and each variable has two possible values, there are four experimental conditions (or interfaces) to investigate (Figure 4.1):

1. No Sparklines + Single Query
2. No Sparklines + Multiple Queries
3. Sparklines + Single Query
4. Sparklines + Multiple Queries

From the specific research questions, the following are the dependent variables that were selected to measure the impact of manipulating the independent variables and to answering the specific research questions.

- Effectiveness
- Efficiency
- Perceived Usefulness
- Perceived Ease of Use
- Perceived Satisfaction

*Effectiveness* measures the ability of a system to produce the desired results. For this research, it was essential to assess the effectiveness of variations of the
proposed system against one another to observe the extent to which they help users to produce the desired results. Effectiveness is an objective variable. Objective data is observable and measurable data obtained through observation or physical examination.

For this study, *efficiency* refers to the speed with which a task can be completed [67]. A task is considered complete when a user has responded to all queries that were mentioned in the task. Efficiency is also an objective variable.

*Ease of use* refers to the degree to which users believe that using a particular system requires little effort to use, whereas *usefulness* refers to the degree to which users believe that a particular system enhances their job or task performance [79]. These two variables were measured to understand the participants’ opinions about the variations of Facet-Time. Ease of use and usefulness are both subjective variables that can be affected by other factors, including system effectiveness, user expectations and attitudes, system type, and task difficulty [36].

Another important measure used for evaluating the system was *satisfaction* [36, 48], which seeks to quantify the degree to which users believe an IIR system has satisfied all their information needs [58]. Satisfaction is a subjective variable.

### 4.1.3 Tasks

The task is a key instrument in the evaluation of IIR systems [16]. For this research, situated work tasks were designed to evaluate the system. A *situated work task* is a short cover story explaining a situation with realistic information [18]. Using such tasks increases experiment realism and encourages participants to search for information in the system [16]. They help participants to understand the context and motivation of a task by clearly explaining the environment of the situation and the problem to be solved.

As a way of exploring time-series data, each participant was asked to perform three situated work tasks, which included one training task and two primary tasks, using one of the four possible variations of a design interface (see Appendix D for
tasks). An interface was assigned to participants following a between-subjects design (Figure 4.2). The between-subjects approach is well suited to this study because the four interfaces work in a similar fashion. There was a reasonable chance that participants might have learned general information relevant to all interfaces if we had used an alternative approach (e.g., the within-subjects design (Figure 4.3)) to assign an interface. Figure 4.4 shows the flow of the tasks to be performed by participants. Qualtrics\(^1\), an online survey tool, was used to collect the data from participants in this study.

---

\(^1\)https://www.qualtrics.com/
4.1.4 Data Collection Methods

The effectiveness was measured using the precision or correctness of the answers to the given tasks. Every task includes sub-tasks, and for every sub-task, the user gets one point if the answer to the sub-task is correct and gets zero otherwise. For example, a task has five sub-tasks and the user gives the correct answers for only three of them, then the score for the effectiveness for that particular task will be $3/5 = 0.60$ (or 60%). To verify the precision of the answers to the given tasks, screen and video recording were used. All participant interactions with the interface were recorded using both screen recording software and a video recorder positioned over the participant’s shoulder and aimed at the keyboard, mouse, and screen. Since the researcher prepared the tasks, the answers to the sub-tasks were analyzed by him from the recorded videos to verify their relevance to the information-needs specified in the tasks. The precision value for each task for each participant was calculated separately since each task was different.

The efficiency was measured using the time (in minutes) that a user took to complete a task. The efficiency was measured separately for each task for each participant since each task was different.

Values were obtained for the perceived usefulness and perceived ease of use variables by administering questionnaires. A questionnaire is another standard method used for collecting data from participants during the evaluation of IIR systems [48]. It consists of a set of questions that can provide quantitative or qualitative data. To measure the perceived usefulness and the perceived ease of use, the Technology Acceptance Model 2 (TAM2) was adapted to create the questions that provide qualitative data [79]. TAM2 is widely used as an evaluation model to assess potential user acceptance of systems. Responses to questions in the questionnaires were gathered using the conventional five-point Likert scale, which has scale labels (1) strongly disagree, (2) disagree, (3) neutral, (4) agree, and (5) strongly agree [56]. To measure perceived satisfaction, TAM2 was used with one question regarding users’ satisfaction. The response to that question was
also gathered using the five-point Likert scale.

4.2 Experimental Procedures

Before running the user study, a pilot study was conducted with a few (three) participants in the planned environment to evaluate the full flow of the study with its instructions. The pilot study helped the investigator to verify that all procedures for the experiment were properly organized and worked smoothly. The verification included checking if the flow of questions in the questionnaire worked well, the software used for recording participants’ interactions worked correctly, and the hardware devices were functioning accurately. The hardware devices used in this experiment were a video camera and two laptops with built-in audio-recorders.

After the pilot study, the controlled laboratory experiment was administered as a single mode study, i.e., the participants completed the evaluation with only the researcher present. Before starting the study, the researcher had two copies of the consent form (see Appendix B for consent form). Participants reviewed the consent form and they were encouraged to clarify any doubts, if they had any. Then, the two copies of the consent form were signed by both the researcher and the participant. After the forms were signed, one copy was given to the participant and the other was retained by the researcher.

The experiment consisted of a sequence of steps. First, each participant was presented with a pre-study questionnaire (see Appendix C for the pre-study questionnaire). The focus of the pre-study questionnaire was to assess participants’ experience with time-series data, visual representations of time series data, and exploratory data analysis. Then, a video demonstration of the key features of the Facet-Time software along with a training task was presented. In order to become familiar with the operation of the software, each participant was asked to do the same training task using the Facet-Time software. Then, they were asked to perform two other tasks in order. After completing each task, a post-task
questionnaire was presented to the participants (see Appendix D for the post-task questionnaire). The entire experiment took between 60 to 90 minutes for each participant.

4.3 Participant Recruitment

Before starting the participants’ recruitment process, the required documents were prepared and submitted to the Research Ethics Board for approval (see Appendix A for Research Ethics Board approvals). After getting the approval, participants were recruited using posters, emails sent to the research student lists, and emails sent to advanced (3rd year, 4th year, and graduate) courses in Computer Science at the University of Regina during the Spring/Summer term of 2019. Interested participants contacted the investigator via email or text. An email was sent to them to coordinate information about the date, time, and place of the evaluation.

Although the data we used for this research was from a human resources domain, but we preferred to recruit computer science students as participants. We chose them for their expertise in complex data and analytical skills, but they do not have domain knowledge.

4.4 Data Analysis

In this section, the data analysis methods used in the evaluation are presented. Several sets of data were collected from the participants for measuring the values of the dependent variables.

As there are four variations of Facet-Time (Figure 4.1), the analysis was performed on the following interfaces (Table 4.1).

These comparisons were formed to help answer the specific research questions. The analysis of Comparison A and Comparison B helped assess the value of providing sparklines in the interface as compared to not providing them, while keeping all other conditions the same (RQ1). Similarly, the analysis of Comparison C and
<table>
<thead>
<tr>
<th>Name</th>
<th>Comparisons of Interfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison A</td>
<td>No Sparklines + Single Query vs. Sparklines + Single Query</td>
</tr>
<tr>
<td>Comparison B</td>
<td>No Sparklines + Multiple Queries vs. Sparklines + Multiple Queries</td>
</tr>
<tr>
<td>Comparison C</td>
<td>No Sparklines + Single Query vs. No Sparklines + Multiple Queries</td>
</tr>
<tr>
<td>Comparison D</td>
<td>Sparklines + Single Query vs. Sparklines + Multiple Queries</td>
</tr>
</tbody>
</table>

Table 4.1: A list of comparisons of interfaces for data analysis

Comparison D helped understand the value of providing multiple-queries in the interface as compared to providing only a single-query, while keeping all other conditions the same (RQ2).

The methods for analyzing the collected data are presented in this section.

### 4.4.1 Analysis of Effectiveness

In an evaluation, when one wants to compare differences between the means of different groups of data, the analysis of variance (ANOVA) method is used [48]. ANOVA is a statistical method to analyze differences among the means of groups of data.

Therefore, in this evaluation, to analyze the effectiveness (dependent variable) of interfaces in given comparisons as shown in (Table 4.1), ANOVA was used. Specifically, the one-way ANOVA between subjects method was used to find whether there were any statistically significant differences between the means of the possible values of the independent variable.

### 4.4.2 Analysis of Efficiency

In this evaluation, the efficiency of participants was determined by measuring the time (in minutes) to task completion for every evaluation task completed by a subject. The resulting numerical data was then used to perform ANOVA for given comparisons of interfaces (Table 4.1).
4.4.3 Analysis of Perceived Usefulness and Perceived Ease of Use

The responses for Likert questions related to perceived usefulness and perceived ease of use are ordinal values. Although they are reported using numbers, the distances between the possible responses are not equal. For example, the distance between neutral (3) and strongly agree (5) is not the same as the distance between disagree (2) and agree (4). Due to the non-equidistant factor, ANOVA analysis cannot be directly applied to the data. Instead, if there are several Likert questions, the numerical value of the responses can be summed and then be used in ANOVA analysis [55].

In this evaluation, there were four Likert questions related to perceived usefulness and four Likert questions related to the perceived ease of use. The participants’ responses to the two sets of four questions were summed to give two sets of four totals. Then ANOVA analysis was applied to each set for the given comparisons of interfaces (as shown in Table 4.1).

4.4.4 Analysis of Perceived Satisfaction

In order to measure perceived satisfaction, one question was presented to the participants. Since a single Likert item value cannot be used for ANOVA, a Mann–Whitney $U$ test [57] was used to analyze this measure. The Mann–Whitney $U$ test is a non-parametric test that can be used to compare differences between two independent groups when the dependent variable is either ordinal or continuous. In our case, this test is considered appropriate since the Likert scale returns ordinal data.

4.4.5 Statistical Significance

Statistical significance is the possibility that a relationship between two or more variables is caused by something other than chance. The analysis of five measure-
ments (effectiveness, efficiency, usefulness, ease of use, and satisfaction), for all four comparisons, is used to determine whether the results are statistically significant. The analysis methods used in this evaluation provide a \( p\)-value, representing the probability that random chance could explain the result. In general, a \( p\)-value of 5% or lower is considered to be statistically significant.

### 4.5 Results

This section reports the findings of the user study based on the analysis methods used. First, the participants’ responses to the pre-study questionnaire are summarized. Later the analysis results for the comparisons in Table 4.1 are provided.

#### 4.5.1 Participants

In total, 32 participants were recruited for the user evaluation. Among them, 78% were males and 22% were females. Their ages ranged between 20 and 32 years. In Tables 4.2 to 4.6, the responses of the participants to the pre-study questionnaire are summarized.

The majority of the participants were graduate students or undergraduate students in the third or fourth year of their degree (Table 4.2). Most of the participants agreed that they have familiarity with time-series data (Table 4.3). About 75% of participants reported that they had experience with exploratory data analysis while the remaining participants did not know of it (Table 4.4). More than 90% of participants had experience with using interactive filters (Table 4.5). Lastly, many participants had experience with the visual representation of time series data (for example viewing data on a timeline or in some other graphical format), whereas some participants were unsure or did not have any experience with timeline visualization (Table 4.6).
<table>
<thead>
<tr>
<th>Level of Education</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate Student (1st or 2nd year)</td>
<td>0%</td>
</tr>
<tr>
<td>Undergraduate Student (3rd or 4th year)</td>
<td>28%</td>
</tr>
<tr>
<td>Completed Undergraduate Degree</td>
<td>3%</td>
</tr>
<tr>
<td>Graduate Student (Master’s Degree)</td>
<td>56%</td>
</tr>
<tr>
<td>Graduate Student (Doctoral Degree)</td>
<td>3%</td>
</tr>
<tr>
<td>Completed Graduate Degree</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 4.2: Level of education of participants

<table>
<thead>
<tr>
<th>Familiarity with Time-Series Data</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>0%</td>
</tr>
<tr>
<td>Disagree</td>
<td>16%</td>
</tr>
<tr>
<td>Neutral</td>
<td>16%</td>
</tr>
<tr>
<td>Agree</td>
<td>59%</td>
</tr>
<tr>
<td>Strongly agree</td>
<td>9%</td>
</tr>
</tbody>
</table>

Table 4.3: Time-series data knowledge of participants

<table>
<thead>
<tr>
<th>Experience with Exploratory Data Analysis</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>0%</td>
</tr>
<tr>
<td>Disagree</td>
<td>7%</td>
</tr>
<tr>
<td>Neutral</td>
<td>25%</td>
</tr>
<tr>
<td>Agree</td>
<td>65%</td>
</tr>
<tr>
<td>Strongly agree</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 4.4: Exploratory data analysis experience of participants

<table>
<thead>
<tr>
<th>Experience with Using Interactive Filters</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>0%</td>
</tr>
<tr>
<td>Disagree</td>
<td>3%</td>
</tr>
<tr>
<td>Neutral</td>
<td>3%</td>
</tr>
<tr>
<td>Agree</td>
<td>63%</td>
</tr>
<tr>
<td>Strongly agree</td>
<td>31%</td>
</tr>
</tbody>
</table>

Table 4.5: Participants’ experience with using interactive filters

<table>
<thead>
<tr>
<th>Experience with the Visual Representation of Time-Series Data</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>0%</td>
</tr>
<tr>
<td>Disagree</td>
<td>6%</td>
</tr>
<tr>
<td>Neutral</td>
<td>31%</td>
</tr>
<tr>
<td>Agree</td>
<td>50%</td>
</tr>
<tr>
<td>Strongly agree</td>
<td>13%</td>
</tr>
</tbody>
</table>

Table 4.6: Participants’ experience with visual representation of time-series data
4.5.2 Comparison A: No Sparklines versus Sparklines (with Single Query)

This section provides the results of five measurements, namely effectiveness, efficiency, usefulness, ease of use, and satisfaction, to aid in assessing the value of providing sparklines in the interface as compared to not providing them, in the context of a single-query visualization.

Effectiveness: The effectiveness of the variations with No Sparklines and Sparklines are provided in Figure 4.5. From the figure, the participants appear to be slightly more accurate in completing the tasks using Sparklines than using No Sparklines. The ANOVA results obtained for the two tasks against the two variations is shown in Table 4.7. These results indicate that the apparent differences are not statistically significant for either task.

![Figure 4.5: The average correctness values obtained of Tasks 1 and 2 for No Sparklines and Sparklines variations with Single Query](image)

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 0.26, p = 0.62$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 0.17, p = 0.69$</td>
</tr>
</tbody>
</table>

Table 4.7: Statistical analysis (ANOVA) of the correctness of the answers to the tasks for No Sparklines and Sparklines variations with Single Query

Efficiency: The efficiency results for both variations are provided in Figure 4.6. It appears from the figure that the participants were slightly quicker in
completing the tasks using Sparklines than using No Sparklines. The ANOVA results obtained for the two tasks for the two variations are shown in Table 4.8.

These results indicate that the differences are not statistically significant for either task.

Figure 4.6: The average time (in minutes) to complete Tasks 1 and 2 for No Sparklines and Sparklines variations with Single Query

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>F(1,14) = 0.13, p = 0.72</td>
</tr>
<tr>
<td>Task 2</td>
<td>F(1,14) = 3.17, p = 0.09</td>
</tr>
</tbody>
</table>

Table 4.8: Statistical analysis (ANOVA) of time (in minutes) to complete the tasks for No Sparklines and Sparklines variations with Single Query

**Usefulness:** Figures 4.7 and 4.8 show the results of the perceived usefulness between the No Sparklines and Sparklines variations for Task 1 and Task 2, respectively. The responses of all participants for the usefulness of the Sparklines variation belonged to strongly agreeing and agreeing. In contrast for the No Sparklines variation, 80% of participants strongly agreed or agreed that it was useful for performing exploratory search tasks with a few neutral opinions. The ANOVA results for the perceived usefulness are reported in Table 4.9. These results indicate that the differences are not statistically significant for either task.
Figure 4.7: The frequency of responses received for the perceived usefulness of the No Sparklines and Sparklines variations for Task 1 with Single Query

Figure 4.8: The frequency of responses received for the perceived usefulness of the No Sparklines and Sparklines variations for Task 2 with Single Query

Table 4.9: Statistical analysis (ANOVA) of the responses for the perceived usefulness for No Sparklines and Sparklines variations with Single Query

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 2.25, p = 0.15$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 0.65, p = 0.43$</td>
</tr>
</tbody>
</table>

Ease of Use: The responses of the participants on the ease of use of No Sparklines and Sparklines variations for Task 1 and Task 2 are illustrated in Figures 4.9 and 4.10, respectively. All participants strongly agreed or agreed with the ease of use of the Sparklines variation. The No Sparklines variation received 80% strongly agreed and agreed responses and a few neutral responses. The ANOVA results in Table 4.10 indicate that the differences between the variations are not statistically significant for either task.

Figure 4.9: The frequency of responses received for the perceived ease of use of the No Sparklines and Sparklines variations for Task 1 with Single Query

Figure 4.10: The frequency of responses received for the perceived ease of use of the No Sparklines and Sparklines variations for Task 2 with Single Query

69
Table 4.10: Statistical analysis (ANOVA) of the responses for the perceived ease of use for No Sparklines and Sparklines variations with Single Query

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 2.35, p = 0.14$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 1.58, p = 0.23$</td>
</tr>
</tbody>
</table>

**Satisfaction:** The participants’ responses concerning their perceived satisfaction with the No Sparklines and Sparklines variations are presented in Figures 4.11 and 4.12, respectively. The figures show that the participants either strongly agreed or agreed that they were satisfied with both variations (No Sparklines and Sparklines). The statistical analysis for the perceived satisfaction measure is presented in Table 4.11. These results indicate that the differences between the variations are not statistically significant for either task.

Table 4.11: Statistical analysis (Mann–Whitney $U$ test) of the responses for the perceived satisfaction for No Sparklines and Sparklines variations with Single Query

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$U = 32, p = 0.96$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$U = 28, p = 0.71$</td>
</tr>
</tbody>
</table>

**Summary**

No statistically significant differences were observed between the No Sparklines and Sparklines variations in this comparison. That means adding the sparklines
to a single-query interface did not add value.

4.5.3 Comparison B: No Sparklines versus Sparklines (with Multiple Queries)

This section provides the results of five measurements, namely effectiveness, efficiency, usefulness, ease of use, and satisfaction, to aid in assessing the value of providing sparklines in the interface as compared to not providing them, in context of multiple-query visualization.

Effectiveness: The effectiveness of the variations of Facet-Time software with No Sparklines and Sparklines is provided in Figure 4.13. From the figure, the participants appear to be slightly more accurate in completing the tasks using Sparklines than using No Sparklines. Even for Task 2, the accuracy for Sparklines variation is 100%. The ANOVA results obtained for the two tasks against the two variations is shown in Table 4.12. These results indicate that the differences are not statistically significant for either task.

![Figure 4.13: The average correctness values obtained for Tasks 1 and 2 for No Sparklines and Sparklines variations with Multiple Queries](image)
Table 4.12: Statistical analysis (ANOVA) of the correctness of the answers to the tasks for No Sparklines and Sparklines variations with Multiple Queries

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>F(1,14) = 0.18, p = 0.67</td>
</tr>
<tr>
<td>Task 2</td>
<td>F(1,14) = 2.01, p = 0.17</td>
</tr>
</tbody>
</table>

Efficiency: The efficiency results for both variations are provided in Figure 4.14. It appears from the figure that the participants were slightly quicker in completing the tasks using Sparklines than using No Sparklines. The ANOVA results obtained for the two tasks for the two variations are shown in Table 4.13. These results indicate that the differences are not statistically significant for either task.

![Figure 4.14: The average time (in minutes) to complete Tasks 1 and 2 for No Sparklines and Sparklines variations with Multiple Queries](image)

Table 4.13: Statistical analysis (ANOVA) of time (in minutes) to complete the tasks for No Sparklines and Sparklines variations with Multiple Queries

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>F(1,14) = 0.47, p = 0.50</td>
</tr>
<tr>
<td>Task 2</td>
<td>F(1,14) = 1.91, p = 0.19</td>
</tr>
</tbody>
</table>

Usefulness: Figures 4.15 and 4.16 show the results of the perceived usefulness between the No Sparklines and Sparklines variations for Task 1 and Task 2,
respectively. The responses of all participants for the usefulness of the Sparklines variation belonged to strongly agreeing and agreeing. However, for No Sparklines variation, 80% participants strongly agreed or agreed that it was useful for performing exploratory search tasks with a few neutral, disagreed, and strongly disagreed responses too in both tasks. The ANOVA results for the perceived usefulness are reported in Table 4.14. From the results, it can be inferred that the differences were statistically significant for both tasks.

![Figure 4.15: The frequency of responses received for the perceived usefulness of the No Sparklines and Sparklines variations for Task 1 with Multiple Queries](image)

![Figure 4.16: The frequency of responses received for the perceived usefulness of No Sparklines and Sparklines variations for Task 2 with Multiple Queries](image)

Table 4.14: Statistical analysis (ANOVA) of the responses of the perceived usefulness for No Sparklines and Sparklines variations with Multiple Queries

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 8.79, p &lt; 0.05$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 6.19, p &lt; 0.05$</td>
</tr>
</tbody>
</table>

**Ease of Use:** The responses of the participants on the ease of use of No Sparklines and Sparklines variations for Task 1 and Task 2 are represented in Figure 4.17 and 4.18, respectively. All participants strongly agreed or agreed with the ease of use of the Sparklines variation. The No Sparklines variation received 80% strongly agreed and agreed responses with a few neutral, disagreed, and strongly disagreed responses. These results reveal that the participants perceived the Sparklines variation to be easier to use for performing the tasks than the No
Sparklines variation. Table 4.15 shows the ANOVA results for the perceived ease of use of both variations. From the results, it appears that the differences between the variations are statistically significant for both tasks.

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 2.35, p &lt; 0.05$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 1.58, p &lt; 0.05$</td>
</tr>
</tbody>
</table>

Table 4.15: Statistical analysis (ANOVA) of the responses of the perceived ease of use for No Sparklines and Sparklines variations

**Satisfaction:** The participants’ responses concerning their perceived satisfaction with the No Sparklines and Sparklines variations are presented in Figures 4.19 and 4.20, respectively. The figures show that all participants strongly agreed or agreed that they were satisfied with the Sparklines variation. In contrast, for the No Sparklines variation, 80% participants strongly agreed or agreed that they were satisfied with this variation with a few neutral and disagreed responses for both tasks. The statistical analysis for the perceived satisfaction measure is presented in Table 4.16. These results indicate that the differences between the variations are statistically significant for both tasks.
Summary

For the two objective measures, no statistically significant differences were observed between the No Sparklines and Sparklines variations in this comparison. However, for subjective measures (usefulness and ease of use, and satisfaction), participants’ experience is improved when sparklines are provided to a multiple-queries interface.

4.5.4 Comparison C: Single Query versus Multiple Queries (with No Sparklines)

This section provides the results of five measurements, namely effectiveness, efficiency, usefulness, ease of use, and satisfaction, to aid in assessing the value of providing Single Query as compared to Multiple Queries, in the context of having
no sparklines in the visualization.

**Effectiveness:** The effectiveness of the variations of Facet-Time software with the Single Query and Multiple Queries are provided in Figure 4.21. From the figure, it appears that the participants were able to complete both tasks more precisely using Multiple Queries than using Single Query. The ANOVA results obtained for the two tasks against the two variations are shown in Table 4.17. These results indicate that the differences between the variations are not statistically significant for either task.

![Figure 4.21: The average correctness values obtained for Tasks 1 and 2 for Single Query and Multiple Queries variations with No Sparklines](image)

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 0.19, p = 0.67$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 0.42, p = 0.52$</td>
</tr>
</tbody>
</table>

Table 4.17: Statistical analysis (ANOVA) of the correctness of the answers to the tasks for Single Query and Multiple Queries variations with No Sparklines

**Efficiency:** The efficiency results for both variations are provided in Figure 4.22. It appears from the figure that the participants were slightly quicker in completing Task 2 using Multiple Queries than using Single Query. However, for Task 1, the average time (in minutes) to complete the task using the Multiple Queries variation is more than the other variation. Apparently, the reason is that
a couple of the participants had difficulties in digging down in the timeline visualization to observe the overlap of different timelines, which led them to take more time to complete the task. These few instances affected the overall average time of participants to complete the task. The ANOVA results obtained for the two tasks for the two variations are shown in Table 4.18. These results indicate that the differences between the variations are not statistically significant for either task.

Figure 4.22: The average time (in minutes) to complete Tasks 1 and 2 for Single Query and Multiple Queries variations with No Sparklines

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 0.06, \ p = 0.81$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 1.42, \ p = 0.25$</td>
</tr>
</tbody>
</table>

Table 4.18: Statistical analysis (ANOVA) of time (in minutes) to complete the tasks for Single Query and Multiple Queries variations with No Sparklines

**Usefulness:** Figures 4.23 and 4.24 show the results of the perceived usefulness between the Single Query and Multiple Queries variations for Task 1 and Task 2, respectively. The responses for the usefulness for both variations are mostly strongly agreed or agreed. A few responses for the Multiple Queries variation were neutral, disagreed, or strongly disagreed. As stated earlier, the reason for such responses is that a couple of participants had difficulties in digging down
in the timeline visualization to observe the overlap of different timelines, which led them to think that the variation is not useful. The ANOVA results for the perceived usefulness are reported in Table 4.19. These results indicate that the differences between the variations are not statistically significant for either task.

Figure 4.23: The frequency of responses received for the perceived usefulness of the Single Query and Multiple Queries variations for Task 1 with No Sparklines

Figure 4.24: The frequency of responses received for the perceived usefulness of Single Query and Multiple Queries variations for Task 2 with No Sparklines

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>F(1,14) = 0.01, p = 0.92</td>
</tr>
<tr>
<td>Task 2</td>
<td>F(1,14) = 0.81, p = 0.38</td>
</tr>
</tbody>
</table>

Table 4.19: Statistical analysis (ANOVA) of the responses of the perceived usefulness for Single Query and Multiple Queries variations with No Sparklines

**Ease of Use:** The responses of the participants on the ease of use of the Single Query and Multiple Queries variations for Task 1 and Task 2 are presented in Figures 4.25 and 4.26, respectively. Almost all participants strongly agreed or agreed with the ease of use of both variations. A few participants disagreed with the ease of use of the Multiple Queries variation. Table 4.20 shows the ANOVA results for the perceived ease of use of both variations. These results indicate that the differences between the variations are not statistically significant for either task.
Figure 4.25: The frequency of responses received for the perceived ease of use of the Single Query and Multiple Queries variations for Task 1 with No Sparklines

Figure 4.26: The frequency of responses received for the perceived ease of use of the Single Query and Multiple Queries variations for Task 2 with No Sparklines

Table 4.20: Statistical analysis (ANOVA) of the responses of the perceived ease of use for Single Query and Multiple Queries variations with No Sparklines

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 0.44, p = 0.52$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 1.09, p = 0.31$</td>
</tr>
</tbody>
</table>

**Satisfaction:** The participants’ responses concerning their perceived satisfaction with the Single Query and Multiple Queries variations are presented in Figures 4.27 and 4.28, respectively. The figures show that many participants strongly agreed or agreed that they were satisfied with both variations. There are a couple of neutral or disagreed responses for the Multiple Queries variation. The statistical analysis results in Table 4.21 indicate that the differences between the variations are not statistically significant for either task.

Figure 4.27: The frequency of responses for the perceived satisfaction of the Single Query and Multiple Queries variations for Task 1 with No Sparklines

Figure 4.28: The frequency of responses for the perceived satisfaction of the Single Query and Multiple Queries variations for Task 2 with No Sparklines
<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>U = 24, p = 0.50</td>
</tr>
<tr>
<td>Task 2</td>
<td>U = 29, p = 0.75</td>
</tr>
</tbody>
</table>

Table 4.21: Statistical analysis (Mann–Whitney $U$ test) of the responses for the perceived satisfaction for Single Query and Multiple Queries variations with No Sparklines

Summary

No statistically significant differences were observed between Single Query and Multiple Queries variations in this comparison. This result suggests that providing multiple-queries support to an interface with no sparklines did not add value.

4.5.5 Comparison D: Single Query versus Multiple Queries (with Sparklines)

This section provides the results of five measurements, namely effectiveness, efficiency, usefulness, ease of use, and satisfaction, to aid in assessing the value of providing multiple-queries as compared to a single-query, in the context of having sparklines in the visualization.

**Effectiveness:** The effectiveness of the variations with Single Query and Multiple Queries are provided in Figure 4.29. From the figure, the participants appear to be slightly more accurate in completing the tasks using Multiple Queries than using Single Query. In fact for Task 2, the accuracy for the Multiple Queries variation is 100%. The ANOVA results obtained for the two tasks for the two variations is shown in Table 4.22. These results indicate that the differences are statistically significant for Task 2.
Figure 4.29: The average correctness values obtained for Tasks 1 and 2 for Single Query and Multiple Queries variations with Sparklines

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 0.37, p = 0.55$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 7.00, p &lt; 0.05$</td>
</tr>
</tbody>
</table>

Table 4.22: Statistical analysis (ANOVA) of the correctness of the answers to the tasks for Single Query and Multiple Queries variations with Sparklines

**Efficiency:** The efficiency results for both variations are provided in Figure 4.30. It appears that the participants were slightly quicker in completing the tasks using Multiple Queries than using Single Query. The ANOVA results obtained for the two tasks for the two variations are shown in Table 4.23. These results indicate that the differences are not statistically significant for either task.

Figure 4.30: The average time (in minutes) to complete Tasks 1 and 2 for Single Query and Multiple Queries variations with Sparklines
Table 4.23: Statistical analysis (ANOVA) of time (in minutes) to complete the tasks for Single Query and Multiple Queries variations with Sparklines

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 1.45, p = 0.24$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 3.93, p = 0.06$</td>
</tr>
</tbody>
</table>

**Usefulness:** Figures 4.31 and 4.32 show the results of the perceived usefulness between the Single Query and Multiple Queries variations for Task 1 and Task 2, respectively. All participants strongly agreed or agreed on the usefulness of both variations. The ANOVA results for the perceived usefulness are reported in Table 4.24. From the results, it can be inferred that the differences between the variations were statistically significant.

![Figure 4.31](image1)  
![Figure 4.32](image2)

Figure 4.31: The frequency of responses received for the perceived usefulness of the Single Query and Multiple Queries variations for Task 1 with Sparklines  
Figure 4.32: The frequency of responses received for the perceived usefulness of Single Query and Multiple Queries variations for Task 2 with Sparklines

Table 4.24: Statistical analysis (ANOVA) of the responses of the perceived usefulness for Single Query and Multiple Queries variations with Sparklines

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 10.60, p \leq 0.005$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 10.60, p \leq 0.005$</td>
</tr>
</tbody>
</table>

**Ease of Use:** The responses of the participants on the perceived ease of use of Single Query and Multiple Queries variations for Task 1 and Task 2 are represented in Figure 4.33 and 4.34, respectively. All participants strongly agreed or agreed
Figure 4.33: The frequency of responses received for the perceived ease of use of the Single Query and Multiple Queries variations for Task 1 with Sparklines.

Figure 4.34: The frequency of responses received for the perceived ease of use of the Single Query and Multiple Queries variations for Task 2 with Sparklines.

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$F(1,14) = 0.88$, $p = 0.36$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$F(1,14) = 1.10$, $p = 0.31$</td>
</tr>
</tbody>
</table>

Table 4.25: Statistical analysis (ANOVA) of the responses of the perceived ease of use for Single Query and Multiple Queries variations with Sparklines.

The results indicate that the differences between the variations are not statistically significant for either task.

Satisfaction: The participants’ responses concerning their perceived satisfaction with the Single Query and Multiple Queries variations are presented in Figures 4.35 and 4.36, respectively. The figures show that 90% of participants strongly agreed that they were satisfied with the Multiple Queries variation. In contrast, for the Single Query variation, most of the responses tend to agree for both tasks. The statistical analysis for the perceived satisfaction measure is presented in Table 4.26. These results indicate that the differences between the variations are statistically significant for both tasks.
Figure 4.35: The frequency of the responses received for the perceived satisfaction of the Single Query and Multiple Queries variations for Task 1 with Sparklines

Figure 4.36: The frequency of the responses received for the perceived satisfaction of the Single Query and Multiple Queries variations for Task 2 with Sparklines

<table>
<thead>
<tr>
<th>Task</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>$U = 12, \ p &lt; 0.05$</td>
</tr>
<tr>
<td>Task 2</td>
<td>$U = 12, \ p &lt; 0.05$</td>
</tr>
</tbody>
</table>

Table 4.26: Statistical analysis (Mann–Whitney $U$ test) of the responses for the perceived satisfaction for Single Query and Multiple Queries variations with Sparklines

Summary

Overall, the participants considered the Multiple Queries variation useful and satisfactory when provided with a sparkline interface. Also, they were more effective using the Multiple Queries variation.

4.6 Discussion

The research focuses on using information-seeking strategies to support the exploration of time-series data. The goal of this evaluation was to examine the value of some features of the variations of the Facet-Time interface. In order to evaluate those features, specific research questions focused on various measures were developed.

A controlled laboratory experiment was used in the evaluation to allow the researcher to control the characteristics of the evaluation. The situated work tasks used in the experiment helped the participants to get a broader view of
<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Comparison A</th>
<th>Comparison B</th>
<th>Comparison C</th>
<th>Comparison D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Usefulness</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Perceived Ease of Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Satisfaction</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 4.27: Summary of the statistical analyses of Task 2 for all four comparisons. The tick-marks represent statistically significant differences.

...the tasks along with the situation in which they were performed. To avoid bias towards a particular variation or task, the training task and the flow of tasks were the same for all the participants. During the experiment, the same laptop device was used by all participants. The same experimental procedure was followed for all participants.

Participants evaluating Facet-Time should have familiarity with time-series data and have some experience with exploratory data analysis. All participants in this evaluation met these basic requirements and were in different age groups and at different levels of study.

Several methods were involved in measuring and analyzing data accumulated during the evaluation. Statistical analyses of the results of these methods assessed the significance of apparent differences.

Since participants completed Task 1 and Task 2 consecutively, they were the more experienced using the interface while performing Task 2. As a result, their responses to this task may have been more valid. Table 4.27 presents a summary of the statistical analyses of Task 2 for all four comparisons.

The evaluation of Facet-Time for supporting the exploration of time-series data provided answers to the following research questions:

RQ1: What is the value of adding sparklines into the facets as compared to not adding them?
where \textit{value} refers to effectiveness, efficiency, perceived usefulness, perceived ease of use, and perceived satisfaction.

To answer this question, we made a comparison between dependent variables for the No Sparklines with Single Query variation and the Sparklines with Single Query variation (Comparison A). Then we performed a comparison between dependent variables for the No Sparklines with Multiple Queries and the Sparklines with Multiple Queries (Comparison B). The results show no statistically significant differences in the objective measures (i.e., accuracy and time) between variations of the interface with and without sparklines. However, improvements were observed in the subjective measures (perceived usefulness, perceived ease of use, perceived satisfaction) for interfaces with sparklines over those without sparklines, especially when they provided multiple-queries support.

**RQ2:** What is the \textit{value} of providing multiple query visualizations of the data in contrast to providing only single-query visualization?

where \textit{value} refers to effectiveness, efficiency, perceived usefulness, perceived ease of use, and perceived satisfaction.

To answer this question, we performed a comparison between dependent variables for the Single Query with No Sparklines variation and the Multiple Queries with No Sparklines variation (Comparison C). Then we compared dependent variables for the Single Query with Sparklines variation and the Multiple Queries with Sparklines variation (Comparison D). The results show no statistically significant differences in the values of the variations concerning the time to complete tasks. However, significant improvements were observed in accuracy and two subjective measures (perceived usefulness and perceived satisfaction) for the variations with multiple queries over those with a single query, especially when they provided sparklines support.

Although the outcomes of the evaluation were mixed, improvements were ob-
served in the subjective measures when sparklines and multiple-queries are used together. While it is essential to develop interfaces that allow people to be efficient and effective in their tasks, the subjective opinions of participants are also important. The diverse outcomes may also be a result of the differences between the individual participants.

The number of participants used to evaluate each interface was low (8) because the total number of participants was not large (32) and due to the between-subjects design, each participant was assigned to a single experimental condition. Repeating the user study with more participants would result in collecting more data, which could reveal more substantial differences between the results with the different experimental conditions.
Chapter 5

Conclusions and Future Research

In this chapter, the summary of the research contributions, limitations, and possible future work are discussed.

5.1 Primary Contributions

The primary contributions of this research are the design, development, and outcome of the study of software to support users analyzing temporal data with a specific focus on supporting exploratory search in the data. This research focused on addressing the challenges faced by users when performing complex or exploratory search tasks within time-series data. As part of this research, an analytical tool called Facet-Time was implemented to study the proposed approach for supporting exploratory search in time-series data.

The thesis covers and contributes to the area of information visualization, information seeking, exploratory search, and time-series data analysis.

One of the primary contributions is the design and implementation of the Facet-Time software. This tool was designed using information seeking techniques to support users’ exploratory search in time-series data when they have complex tasks to perform. A timeline is included in the software to allow users to observe the temporal distribution of the result set. The timeline also enables users to compare subsets of the data. Facet-Time provides sparklines to display pattern of working
hours for the facet values over time. While defining queries to explore the data, the sparklines may help users in deciding on their path of exploration.

Facet-Time was implemented as web-based software that is composed of loosely coupled components and follows a distributed client-server architecture. As these components are loosely coupled, they can be reused or replaced to incorporate modifications without affecting other components in the application.

The second contribution of this research is the study design and findings of the evaluation of Facet-Time. This software was evaluated by users in a controlled laboratory setting to study if it could support exploratory search in time-series data.

During the evaluation, design decisions related to Facet-Time were measured in terms of effectiveness, efficiency, perceived usefulness, perceived ease of use, and satisfaction. From the evaluation, various quantitative and qualitative data was collected and analytical methods were employed to assess the different aspects of the outcomes of using the software. The results provided interesting insights concerning Facet-Time.

Improvements have been observed in the effectiveness of using multiple queries when the interface is provided with sparklines. There are no significant differences in the interfaces in terms of efficiency across any of the comparisons. Having sparklines and multiple queries together in the interface was perceived as more useful and satisfactory than other variations of the Facet-Time. Lastly, sparklines were perceived as easy to use when the interface is provided with multiple queries.

5.2 Limitations

Even though this research proposes techniques that can help with exploring search within time-series data, it has some limitations. One of the limitations is that the time granularity is fixed for the timeline and sparklines. Currently, each granule on the x-scale of the timeline and sparklines represents a single day and it can
not be changed. If support were provided for dynamic granularity, users could adjust a visualization according to their needs to have an abstract or detailed representation of the data.

Another limitation is the comparison of sparklines across the facets. Where the purpose of sparklines is to give a sense of the distribution of data, users can also make comparisons between different sparklines. In Facet-Time, users can only compare sparklines within a specific facet. Because the sparklines are not aligned across the facets, it will be hard to compare them. Also, across several facets, the y-axes of the sparklines have different scales and it will be challenging to compare sparklines with different scales. Thus, improvements are needed to allow comparison of sparklines across facets.

The next limitation is related to the number of participants for the user evaluation, which limits the findings of the research. For the evaluation, the between-subjects study design was incorporated with 32 participants to avoid learning effects. There were four different variations of Facet-Time to evaluate; each variation was evaluated with only 8 participants. To obtain other significant findings from this user evaluation, the number of participants was low.

5.3 Future Work

This section describes possible future work to address the above limitations and to add new features to the Facet-Time software.

First, addressing the limitation of the software to support dynamic granularity would aid users with data analysis by adjusting the visualization to have an abstract or a detailed representation of the data. Users would be able to change the timeline view from days to weeks, weeks to months, and months to years. Dynamic granularity would help them to explore the underlying patterns which are only visible with abstract or detailed representations of the data.

A further improvement to Facet-Time would be to incorporate various selection
techniques in faceted browsing. Selection techniques in faceted browsing could help users perform their tasks more efficiently. One possible technique is to provide negative or inverted selection support. Currently, when users make selections in facets, the data gets filtered out according to the selections that are made. If negative selection was supported in faceted browsing, users would be able to eliminate the results relevant to items selected for exclusion.

Another possible improvement to faceted browsing would be to allow users to change between disjunctive (OR) selection and conjunctive (AND) selection within facets. For now, Facet-Time only supports disjunctive selection within facets. If support was provided for both disjunctive and conjunctive selection, users would be able to change between them according to their needs, and potentially be more efficient at performing their tasks.

Addressing some limitation of the evaluation may reveal further differences between the four variations of Facet-Time. Recruiting additional participants and collecting more data will make the results reliable. Further, we can perform other statistical techniques to analyze the data of user evaluation, i.e., factorial ANOVA. Factorial ANOVA can help to analyze the combined effect of more than one independent discrete variables on dependent variables. This way, we can observe whether or not sparklines and multiple query visualization together show significant improvements. Lastly, in further studies, more complex tasks can be designed to avoid the ceiling effect [70] and get stronger results. The results of effectiveness for all variations of Facet-Time indicate that tasks were not very complex to perform. More complex tasks may help to observe the differences between the four variations of Facet-Time.

In order to assess the generalizability of the approach, we can try different types of time-series data other than project management, i.e., historical data. Additional to the time factor, a historical dataset may contain other information, such as the location or description of a historical event. Such pieces of information can also be utilized in visualization along with a timeline and faceted navigation.
For example, we can render a map to show the location of events. The description of events can be presented as snippets along with the timeline. We can also use the description of events to create facets dynamically using Natural Language Processing (NLP) techniques. Later, user evaluations will help us to understand whether new features provide any value to users or not.
Bibliography


distortion: The gestalt illusions. *Journal of Experimental Psychology: Human

[27] V. Cox. Exploratory data analysis. In *Translating Statistics to Make Deci-

scale widget for navigating hierarchical metadata. In *Proceedings of the
SIGCHI Conference on Human Factors in Computing Systems*, pages 1353–

[29] M. J. Eppler and R. A. Burkhard. *Knowledge Visualization*. IGI Global,
Hershey, Pennsylvania, USA, 2008.


Architectures*. PhD Thesis, University of California, Irvine, California, USA,
2000.

information. In *Proceedings of the CHI Extended Abstracts on Human Factors

[33] J. Fulda, M. Brehmel, and T. Munzner. Timelinecurator: Interactive au-
thoring of visual timelines from unstructured text. *IEEE Transactions on

[34] P. Goffin, W. Willett, A. Bezerianos, and P. Isenberg. Exploring the effect
of word-scale visualizations on reading behavior. In *Proceedings of the CHI
Extended Abstracts on Human Factors in Computing Systems*, pages 1827–
1832, 2015.


Appendix A

Research Ethics Board Approvals
The University of Regina Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol, or related documents.

Any significant changes to your proposed method, procedures or related documents should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS
In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for the renewal and closure forms:
https://www.uregina.ca/research/for-faculty-staff/ethics-compliance/human/ethicsforms.html

Chris Street PhD
REB Chair
University of Regina
Appendix B

Consent Form
Project Title: A Study on the Exploration of Time Series Data

Researcher: Haider Ali Butt  
Department of Computer Science  
University of Regina  
butt200h@uregina.ca

Co-Supervisors:  
Dr. Orland Hoeber  
Department of Computer Science  
University of Regina  
orland.hoeber@uregina.ca  
http://www.cs.uregina.ca/~hoeber

Dr. Howard Hamilton  
Department of Computer Science  
University of Regina  
hamilton@cs.uregina.ca  
http://www2.cs.uregina.ca/~hamilton/

Purpose and Objectives of the Research:

- The purpose of this research is to study four design variations for an interface to support the exploration of time series data. Each participant in this study will be asked to use just one of these design variations.

Procedures:

- You will be provided with two printed copies of this consent form and will be asked to review the document in detail. You will be able to ask any questions you might have about the procedures and goals of the study, and your role within the evaluation.
- Once you are satisfied that you are informed about the study and wish to continue with your participation, you will be asked to sign both copies of the consent form. The researcher will also sign and date both copies. One will be kept with the researcher and one provided to you for your files.
- You will be asked to complete a pre-study questionnaire that measures your experience with time series data, data analysis, data exploration, visual representations of data, and searching for information.
- The researcher will then introduce you to a time series data exploration web application (interface), giving you a short video demonstration of the key features, then allowing you to operate/use the features yourself. You will be encouraged to ask questions about anything you don’t understand as you use the interface.
- You will then be provided with a training task, in which you can use the features of the application to explore time series data to solve a specific data analysis problem.
- You will then be asked to complete two tasks using the given interface.
- After each of these tasks, you will be asked to complete an online questionnaire, which will seek to measure your opinion on the satisfaction, usefulness, ease of use, and exploration support with the overall interface.
- Your interaction with the interface will be monitored using a video camera aimed over your shoulder at the keyboard, mouse, and display. We will also use screen monitoring software to record your session using the interface.
• The researcher will be available to help you to use the application should you run into any difficulties.
• You will have a 50% chance to win a $10 Tim Hortons gift card.
• The entire process should take approximately 60 minutes.

Potential Risks:
• Since this is a controlled laboratory evaluation that consist of using a computer system for exploration of the time series data, there are no anticipate risks associated with your participation in this study.

Potential Benefits:
• The benefit to the researcher is the information on how the time series data exploration tool is used to solve specific data analysis problems. This information will form the basis for the researcher’s thesis.

Anonymity & Confidentiality:
• Because you will undertake your tasks in the presence of the researcher, your participation will not be anonymous.
• All information you provide will be kept confidential and will in no way be linked to your identity.
• The data collected during this study will only be used by the researcher and his co-supervisors for the analysis of the interfaces in this specific project.

Storage of Data:
• The data for this research project will be stored on password-protected computers.
• When the data is shared between the researcher and his supervisor, it will be done in a secure and encrypted manner.
• All data from this project will be kept for a minimum of five years.
• All physical data will be shredded once it is no longer needed using the document disposal facilities provided by the University.
• At the conclusion of the project, all electronic files will be securely deleted.

Right to Withdraw:
• Your participation is voluntary, and you may choose not to answer any questions with which you are uncomfortable. You may withdraw from the research project for any reason without explanation or penalty of any sort.
• Your willingness to participate or not, and your specific answers to the questions will have no bearing on current or future interactions or dealings with the researcher and his co-supervisors.
• If you choose to withdraw, you may simply state that you no longer wish to participate. If you do so, we will delete all records of your participation in this study. However, we will keep you in the draw for one of the gift cards.
• Because no link between your identity and the data collected will be maintained, once you complete this session it will no longer be feasible to withdraw your participation and remove your data from the study.
Follow up:

- The results of this research will form the basis of the researcher’s M.Sc. Thesis. It may also be used in conference papers and academic journal articles.
- If you wish to obtain the results of this study, you can request them directly from the researcher (using the contact information at the top of this document).
- A link to the researcher’s thesis will be posted on the co-supervisors’ websites, where you may find further information about the study when it is complete.
- If you are selected in the draw for the gift cards, you will be contacted via the email or mobile phone number you used to coordinate your participation.

Questions or Concerns:

- You may contact the researchers using the information at the top of page 1 to answer any questions or address any concerns you might have.
- This project has been approved on ethical grounds by the University of Regina’s Research Ethics Board on (approval date). Any questions regarding your rights as a participant may be addressed to the committee at (585-4775 or research.ethics@uregina.ca). Out of town participants may call collect.

Your signature below indicates that you have read and understand the description provided; I have had an opportunity to ask questions and my/our questions have been answered. I consent to participate in the research project. A copy of this Consent Form has been given to me for my records.

<table>
<thead>
<tr>
<th>Name of Participant</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Researcher’s Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* A copy of this consent will be left with you, and a copy will be taken by the researcher.*
Appendix C

Pre-Study Questionnaire
The questions below collect your knowledge and interest about the given task.

1. How old are you? _______________

2. What is your gender?
   - Male
   - Female
   - Other

3. What is your current level of education?
   - Undergraduate Student (1st or 2nd year) _______________
   - Undergraduate Student (3rd or 4th year)
   - Completed Undergraduate Degree _______________
   - Graduate Student (Master’s Degree)
   - Graduate Student (Doctoral Degree)
   - Completed Graduate Degree

4. In which faculty or department are you pursuing your degree (or working)?
   _______________

5. I have some familiarity with data that has a component of time (for example, periodic measurements of temperature, changing stock market prices, or operating system event logs).
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

6. I have some experience with exploratory data analysis (for example, examining a data set to uncover previously unknown patterns and characteristics).
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree
7. I have some experience with using interactive filters when searching online (for example, size filters in online clothes shopping sites, brand or product category filters for online electronics shopping sites, etc.)

   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

8. I have some experience with the visual representation of time series data (for example viewing data on a timeline or in some other graphical format).

   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree
Appendix D

Tasks and Post-task Questionnaire
Training Task

Suppose you are a product manager for a small software development company. You have had some concerns about the amount of contract work being done on your project (Marketing Plans) in comparison to other projects in the company. Using the data analysis interface,

1. Find other products that have contract-based work done.
2. Assess whether their pattern of contract work is similar to your own, or different.
3. Identify if there is any overlap in who is doing the contract work for your project and any other projects for the company.
4. Is there any time at which your contractors have been working on multiple projects simultaneously?
5. If so, how many hours were worked on each project during that time?

These statements relate to the overall experience with the Facet-Time application for this task.

Perceived Satisfaction

1. I am satisfied with using the system for exploring time series data for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

Perceived Usefulness:

1. Using the system to explore time series data improved my performance for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

2. Using the system enhanced my effectiveness in exploring time series data for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

3. I found the system useful for exploring time series data for this task.
   - Strongly disagree
Disagree
Neutral
Agree
Strongly agree

4. Using the system increases my productivity doing this task.
   • Strongly disagree
   • Disagree
   • Neutral
   • Agree
   • Strongly agree

Perceived Ease of Use:

1. My interaction with the system was clear and understandable for this task.
   • Strongly disagree
   • Disagree
   • Neutral
   • Agree
   • Strongly agree

2. Interacting with the system did not require a lot of my mental effort for this task.
   • Strongly disagree
   • Disagree
   • Neutral
   • Agree
   • Strongly agree

3. I found the system to be easy to use for this task.
   • Strongly disagree
   • Disagree
   • Neutral
   • Agree
   • Strongly agree

4. I found it easy to get the system to do what I want it to do for this task.
   • Strongly disagree
   • Disagree
   • Neutral
   • Agree
   • Strongly agree
**Task 1:**

You are a General Manager of a product line (Web Page Analysis) that employs permanent full-time, permanent part-time, and contract human resources. You are planning to terminate the contract work on this product-line and re-assign this work to full-time human resources who already have work experience on this product-line with the same clients but are also working on other products. Your job here is to:

1. Find the clients who have funded the Web Page Analysis product on contracts.
2. Find all full-time human resources who have worked on the Web Page Analysis product for those clients.
3. Identify those human resources who have spent 80 to 95 hours in total in May on this product for those clients.
4. For those resources:
   a. Assess whether their pattern of work on the Web Page Analysis product for those clients matches one another (they are working a similar number of hours over similar periods of time)
   b. Are there any other common products they were working on during May for these same clients?
   c. If so, how many hours did they spend on those products?

These statements relate to the overall experience with the Facet-Time application for this task.

**Perceived Satisfaction**

1. I am satisfied with using the system for exploring time series data for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

**Perceived Usefulness:**

1. Using the system to explore time series data improved my performance for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree
2. Using the system enhanced my effectiveness in exploring time series data for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

3. I found the system useful for exploring time series data for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

4. Using the system increases my productivity doing this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

Perceived Ease of Use:

1. My interaction with the system was clear and understandable for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

2. Interacting with the system did not require a lot of my mental effort for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

3. I found the system to be easy to use for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

4. I found it easy to get the system to do what I want it to do for this task.
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree
Task 2:

The Funk Group (client) has some concerns about the work done by our company on the Cash Management product, and they are unhappy with the progress made during a couple of months (April and May 2018) compared to the preceding month (March 2018). Yet they have been paying overtime for some human resources during those months.

1. Find human resources who were working on the Cash Management product for the Funk Group.
2. Of those human resources, identify the ones who spent more than 8 hours a day on the Cash Management product for the Funk Group during any day in April or May 2018.
3. Find the total number of hours spent working on the Cash Management product for the Funk Group in April and May together.
4. Is there any human resource who did work with the Funk Group on the Cash Management product during March but who did not work on it in April?
5. If so, for each such human resource:
   a. Assess whether the human resource was working on any other products during April.
   b. If so, how many hours did the human resource spend on all other products during April?
   c. Identify any period in the period from March to May when the human resource was working on both the Cash Management product and any other products of our company.

These statements relate to the overall experience with the Facet-Time application for this task.

Perceived Satisfaction

1. I am satisfied with using the system for exploring time series data for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

Perceived Usefulness:

1. Using the system to explore time series data improved my performance for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
Strongly agree

2. Using the system enhanced my effectiveness in exploring time series data for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

3. I found the system useful for exploring time series data for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

4. Using the system increases my productivity doing this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

Perceived Ease of Use:

1. My interaction with the system was clear and understandable for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

2. Interacting with the system did not require a lot of my mental effort for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree

3. I found the system to be easy to use for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
4. I found it easy to get the system to do what I want it to do for this task.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly agree