

Web Object Attention Model For Both Image And Text Object

By

Prateek Rastogi

[10010146]

Under The Guidance Of
Dr. Samit Bhattacharya



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

ABSTRACT

Webpages have redefined the way in which information is published and consumed in today's world. With search engines becoming a predominant way to search information on web, any webpage should be able to convey its information to the random surfer within a short attention span. Hence, it would be beneficial if webpage authors can predict the sequence in which various information presented in a webpage is seen by the user.

Consequently, we divided the webpage contents into sub-units referred as object, which is basically lucid regions of webpages that convey same information. Based on the eye tracking data collected from 15 participants and 15 webpages, major factors affecting the user attention to a particular object are identified.

For the purpose of validation, we repeated the same procedure twice on two distinct sets of 15 subjects and 15 webpages.

Contents

1. Introduction	1
1.1 Motivation	2
1.2 Objective	2
1.3 Work Done	2
2. Previous Work	3
2.1 Guided Search Model	4
2.2 Saliency Model	4
2.3 Web Object Attention Model	5
2.4 Model To Predict Attention Sequence of Web Objects	5
2.4.1 Attention Factor	5
2.4.2 Drawbacks	5
2.5 Web Object Attention Model For Both Image and Text Objects	6
2.5.1 Types of Objects	6
2.5.2 Experiment Setup	10
2.5.3 Data Collected	10
3. Data Collection and Analysis	11
3.1 Data Collection	11
3.2 Data Analysis	12
4. Empirical Validation	14
5. Conclusion and Future Work	16
5.1 Conclusion	16
5.2 Future Work	16
Appendix A Webpages used in Empirical Data Collection	19
Appendix B Observed Ranks in Empirical Data	22
Appendix C Webpages used in Validation Data Set I	26
Appendix D Observed Ranks in Validation Data Set I	29
Appendix E Webpages used in Validation Data Set II	34
Appendix F Observed Ranks in Validation Data Set II	37

Chapter 1

Introduction

A webpage is a communication medium through which a user seeks information. A webpage can contain lot of information. A user may get overwhelmed by large amount of content and leave without acknowledging important content, or may waste ample amount of time searching for the relevant content.



Figure 1 A sample webpage with various entities labelled.

So, a web designer may want to present the information in a way that the most important information catches the user attention first. Fixation order of user's eye gaze can reveal sequence in which information presented in the webpage will attract user attention. Eye tracking equipment provides real time gaze plots of human eye, but they are expensive and the process of collecting data itself is time consuming task. An automatic design framework embedded in a web development environment might be the best way for a designer to determine view sequence of various entities accessible in a webpage.

1.1 Motivation

Guided Search Model [13] provides idea of human visual behaviour when user is searching for a particular item in a visual scene. However, a random surfer may not have any precise search task.

Visual Saliency Model [8] extends the idea presented in Guided Search Model. It explains the human attention behaviour in a more biological conceivable manner, but a major limitation of this model is its applicability on the images of only fixed resolutions.

As a result of those shortcomings, none of these models can match the accuracy of eye tracking system. Therefore, a web object attention model [11] has been proposed for predicting the order of eye fixation. This idea is further extended by [9]. Both these models consider only image as an object, but as web pages contain both image and textual objects, these models cannot predict the eye fixation order with adequate accuracy.

1.2 Objective

Our aim is to propose an attention factor model for webpages comprising both image and text for predicting the eye fixation order of random surfer in first few seconds.

1.3 Work Done

- To identify the major elements affecting a surfer's attention we have performed statistical analysis on the data collected from the experiment involving 15 participants and 15 web pages. From the analysis, we demarcated four major characteristics that affect user's attention.
- For validating our hypotheses we tested our theory on the data collected from two experiments, each involving 15 volunteers and 15 webpages.

Chapter 2

Previous Work

Human attention to a particular object is a complex process. As visual input is collected by eye, they play an important role in the process. It collects visual information and sent it to a part of brain known as visual cortex, which process it to direct attention.

Existing models use identifies different features to explain how visual attention is directed. All of them face challenges of merging features of different factors. In many cases, weights of different factors is learned and linear combination of the factors is performed.

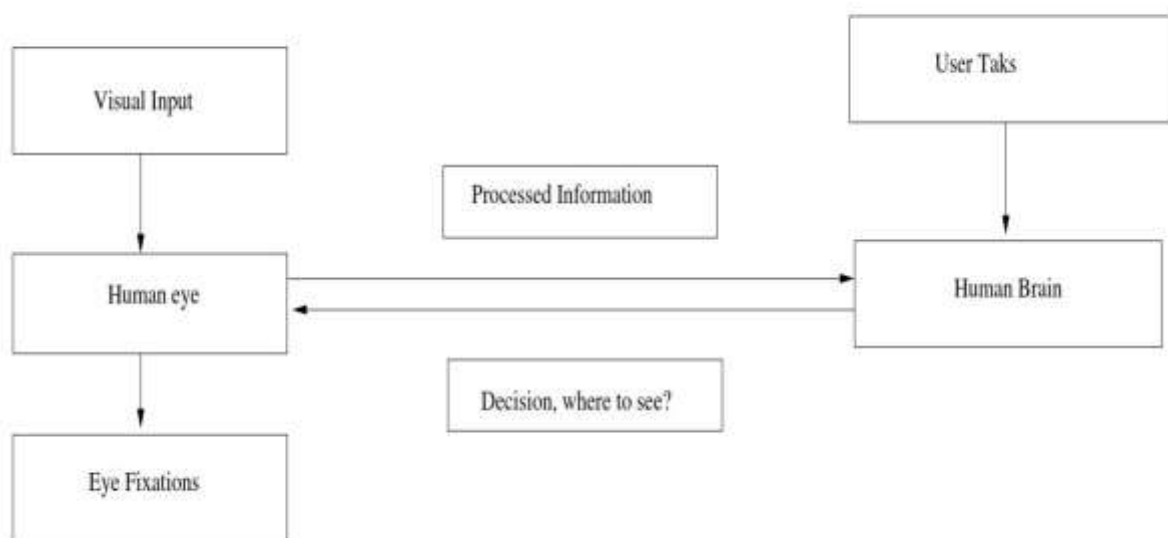


Figure 2.1 Human Visual Attention deployment

2.1 Guided Search Model

Guided Search Model [13] explains human visual search behaviour in two stages: parallel processing and limited capacity processing. In parallel processing whole visual scene is processed, but in limited capacity processing only a smaller area of the scene is handled. Initially, visual system processes entire visual scene in parallel. Parallel processing is done on the basis of simple basic properties of the scene only. The output of parallel processing is passed to limited capacity processors to narrow processing to a single item or a small portion in the scene e.g. Consider selecting a blue ball of brand X from the bag of balls.

Parallel processors create representations for basic visual features called feature maps. Limited capacity processors form activation map based upon user's task and the feature maps. There are two varieties of activation: bottom-up and top-down. Bottom-up activation is a measure of how different an item is from neighbouring items. Top-down activation is based on user task. Activation map of any particular object is the combination of both top-down and bottom-up activations. Peaks in the activation map show the order of items getting attended by the user. The item with maximum peak value is the most likely to be attended by the user.

2.2 Saliency Model

Visual Saliency is the distinctness of an object from its neighbouring objects [5]. As our visual system is incapable of processing all the features of all the locations in parallel, visually salient object will more easily grab attention compared to other non-salient objects.

In this model, feature maps, which give the comparative value of an item with neighbouring items in terms of colour, intensity, and orientation, are generated. Then those feature maps are normalised and merged to get a saliency map.

The models discussed above have limitations. The Guided Search Model [13] explains attention deployment only when user have particular search task in mind. On the other hand, Saliency Model [5] works only on images of 640 x 480 resolution. Therefore, the web object attention model was proposed.

2.3 Web Object Attention Model

It [11] tries to predict computationally the fixation order of the information present in the webpage. This model defines textual or pictorial information like a text box, an image, a table, or any other data as object. While calculating the attention factor, it takes intensity contrast, chromatic contrast, and size of an object into account.

2.4 Model To Predict Attention Sequence Of Web Page Objects

This model [9] extends the ideas presented in Web Object Attention Model [11]. It computes attention factor on the basis of chromatic contrast, size, x-position and y-position of an object. Higher value of attention factor represents higher attention drawing capability of an object.

2.4.1 Attention Factor

The proposed expression for attention factor according to the model is

$$AF_i = 0.16 * C_i + 0.40 * S_i + 0.43 * X_i + 0.1 * Y_i$$

Where C_i , S_i , X_i , and Y_i are normalized chromatic contrast, size, x-position factor and y-position factors respectively.

2.4.2 Drawbacks

The above model computes attention factor only for artificial webpages having six images with rectangular shape, but real world webpages contain not only images but also other objects, including textual objects, and interactive objects. Moreover, the number of objects in a real webpage may vary and is not fixed to six. Also, the presupposition that objects are only rectangular in shape is non-plausible in real world scenario.

2.5 Web Object Attention Model For Both Image And Text Objects

Web Object Attention Model for both Image and Text [17] extends the web attention model to cover both image and text objects. In doing so, it redefines object types and proposes two important factors, size and centre, that determines the rank of those objects. In our thesis, we will try to find other factors that play influential role in resolving the sequence in which web objects will be attended.

2.5.1 Types of Objects

The major object types identified in the above work [17] are:-

1. **Pure image:** A single image is an object. Example is show in figure 2.2.



Figure 2.2 Pure image Object

2. **Image with Caption:** It is defined as image having text nearer or just below, with a word limit of 15. In such cases, text accompanying image is usually an image description. Image along with text forms an object. Example is shown in figure 2.3.



Figure 2.3 Image with caption

3. **Image overlapped with Text:** An image with text embedded on it is identified as object. Example of this object is shown in figure 2.4.



Figure 2.4 Image overlapped with Text

4. **Paragraph:** Textual contents like paragraph, including its heading, forms a single object irrespective of variation in font style, for font family or font size inside text. Example of this object is shown in figure 2.5.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam cursus. Morbi ut mi. Nullam enim leo, egestas id, condimentum at, laoreet mattis, massa. Sed eleifend nonummy diam. Praesent mauris ante, elementum et, bibendum at, posuere sit amet, nibh. Duis tincidunt lectus quis dui viverra vestibulum. Suspendisse vulputate aliquam dui. Nulla elementum dui ut augue. Aliquam vehicula mi at mauris. Maecenas placerat, nisi at consequat rhoncus, sem nunc gravida justo, quis eleifend arcu velit quis lacus. Morbi magna magna, tincidunt a, mattis non, imperdiet vitae, tellus. Sed odio est, auctor ac, sollicitudin in, consequat vitae, orci. Fusce id felis. Vivamus sollicitudin metus eget eros.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. In posuere felis nec tortor. Pellentesque faucibus. Ut accumsan ultricies elit. Maecenas at justo id velit placerat molestie. Donec dictum lectus non odio. Cras ante vitae enim laculis aliquam. Mauris nunc quam, venenatis nec, euismod sit amet, egestas placerat, est. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Cras id elit. Integer quis urna. Ut ante enim, dapibus malesuada, fringilla eu, condimentum quis, tellus. Aenean porttitor eros vel dolor. Donec convallis pede venenatis nibh. Duis quam. Nam eget lacus. Aliquam erat volutpat. Quisque dignissim congue leo.

Figure 2.5 Paragraph Object

5. **List:** List along with its heading is considered an object. List can be ordered or unordered list. Example is shown in figure 2.6.

The ingredients:

- 100 g. flour
- 10 g. sugar
- 1 cup water
- 2 eggs
- salt, pepper

Figure 2.6 List

6. **Table:** Table along with its heading forms an object. Example such objects is shown in figure 2.7.

Company	Contact	Country
Alfreds Futterkiste	Maria Anders	Germany
Berglunds snabbköp	Christina Berglund	Sweden
Centro comercial Moctezuma	Francisco Chang	Mexico
Ernst Handel	Roland Mendel	Austria
Island Trading	Helen Bennett	UK
Königlich Essen	Philip Cramer	Germany
Laughing Bacchus Winecellars	Yoshi Tannamuri	Canada
Magazzini Alimentari Riuniti	Giovanni Rovelli	Italy
North/South	Simon Crowther	UK
Paris spécialités	Marie Bertrand	France
The Big Cheese	Liz Nixon	USA
Vaffeljernet	Palle Ibsen	Denmark

Figure 2.7 Table Object

7. **Heading:** We consider heading as an object. This object is different from heading of paragraph, table, and list. Example of is shown in figure 2.8.

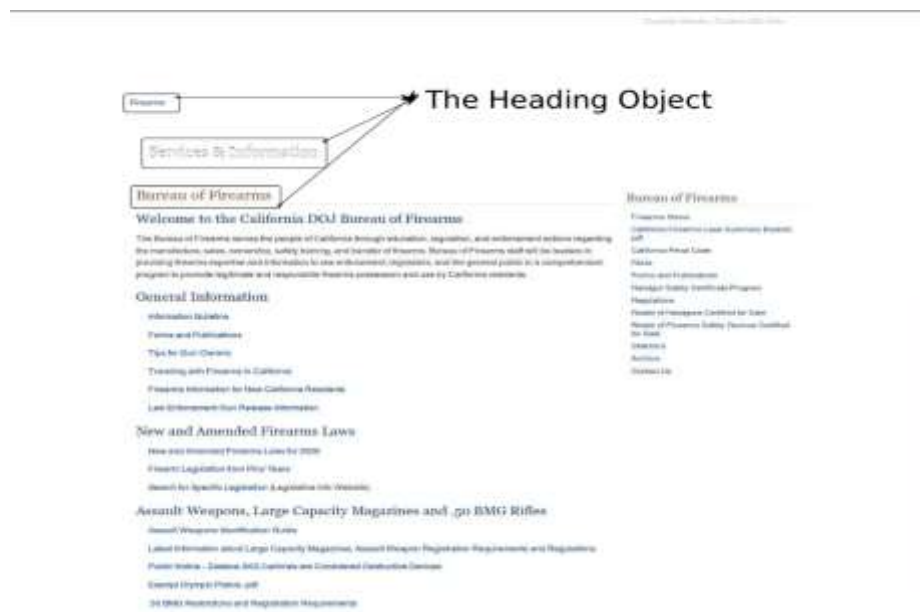


Figure 2.8 Heading Object

8. **Menu:** Menus are also identified as an object. They can be either at the top or at the sides of the page. Example of such object is shown in figure 2.9.



Figure 2.9 Menu

9. **Interactive Items:** These kinds of objects include items that require input by the user such as log-in field, search box, and radio button. Example is shown in figure 2.10.

Figure 2.10 Interactive Item

2.5.2 Experiment Setup

Eye Tracker: Tobii x50 Eye Tracker was used in experiment to collect the data. It provides real-time coordinates of the eye gaze. The eye tracker uses pupil centre reflection method to find the gaze directions. ClearView2.7 provided along with Tobii Eye Tracker helps in scrutinizing gaze data obtained from the eye tracker.



Figure 2.11 Tobii Eye Tracker

Monitor: 17inch LG Flatron having screen resolution of 1280 X 1024 was used.

Webpages and Volunteers: 15 Webpages were downloaded from the internet and 15 Volunteers participated in the experiment. With fixed time interval between successive webpages, volunteers were shown a slideshow of all 15 webpages. Blank slides were inserted between successive webpages to reset eye position of the volunteers. Also to avoid any learning bias, webpages were rearranged in the slideshow of each volunteer such that no two volunteer saw webpages in the same sequence.

2.5.3 Data Collected

For each webpage, objects were ranked for each user according to preferential voting scheme [2]. Then, observed ranks from different users were merged to obtain a combined index [2] of object ranks.

Chapter 3

Data Collection and Analysis

3.1 Data Collection

To identify the factors affecting the rank of web objects, we have collected data about following properties of web objects:

Area: Area of web object in square pixels. Maximum area of the webpage is 1310720 pixel² i.e. 1280*1024.

Mean: Average gray value of the object. This is the sum of the gray values of all the pixels in the object divided by the number of pixels.

StdDev: Standard deviation of the gray values used to generate the mean gray value i.e. Root Mean Square Contrast of the object.

Mode: Most frequently occurring grey value within the object.

Min: Minimum grey value within the object.

Max: Maximum grey value inside an object.

Centroid: The centre of the object. This is the average of the x and y coordinates of all of the pixels in the image or selection. Uses the X and Y Headings. Leftmost-Topmost point of the screen is taken as origin.

Centre of Mass: This is the brightness-weighted average of the x and y coordinates all pixels in the image or selection. Uses the XM and YM headings. These coordinates are the first order spatial moments.

Median: The median value of pixels in an object.

3.2 Data Analysis

Based on the combined index of object ranks, major factors responsible for the sequence in which objects were viewed in a webpage are identified. The key elements found in our analysis are discussed underneath.

1. **Y-axis:** Following Considerations were taken into account while analysing data to show the effect of Y-co-ordinate on the rank of web objects:-
 - Top-most point of the webpage is assumed as $Y=0$.
 - Positive Y-direction is from top to bottom.
 - To minimize the impact of size, only the pages that has 3-4 different object size dimensions are considered.
 - Effect of X-co-ordinate on the object rank is not considered.
 - From the ideal rank of object according to y-coordinate, an error margin of ± 2 is taken into account to reduce the irregularity in rank due to other factors.

In empirical data we have 15 webpages out of which 14 satisfied above constraints. In those 14 webpages, we have found **81** objects that are ranked correctly according to their Y distance from top i.e. rank of object increases from top to bottom, whereas **25** objects are not ranked appropriately.

2. **Position:** Entry position for users was center of screen 80% of the time. So, the objects positioned at or near the center of screen has more chances of receiving user attention early.
3. **X-axis:** Following assumptions were made while examining data for the consequence of X-Factor on the rank of web objects:-
 - Only those objects that are located at same y axis co-ordinate are considered.
 - For any particular Y-co-ordinate, centre of the screen is taken as $X=0$.
 - We are only considering the effect of x co-ordinate on the objects that are of nearly same size.
 - Till objects of same size and same y co-ordinate on any one side of the centre will not end we will make comparisons i.e. if there are two objects of same size and one is on the left of centre and another is on the right of centre, then

we will need only one comparison, but if there are four objects, then we may require two or three comparisons.

In empirical data we have 15 webpages out of which 13 satisfied above constraints. In those 13 webpages, **19** times a left side object is seen before a right side object and **13** times a right side object is seen before a left side object. Thus, we can conclude that X-co-ordinate position of an object has no effect on rank of objects.

4. **Size:** We got 15 objects that were ranked in their respective web pages. Out of those 15 objects, 9 were large images overlapped with text. These 9 image were attended 97 times out 135. Thus, we can determine that bigger size objects catch user's attention easily.
5. **Min and Max Grey Value:** Minimum grey scale value for most objects is 0 and maximum grey scale value for most objects is 255. Thus, min and max grey value of object has no considerable effect on object rank.
6. **Median:** Median grey scale value has little or no effect on the rank of object.
7. **Mode:** It is observed that mode grey value has no influence on the object rank.
8. **Contrast:** It is also observed that in 69.71% cases high contrast objects generally attracts user attention earlier than low contrast objects.
9. **Mean:** In 65.23% cases, we have seen that mean grey scale value increases for the object seen later, if all other factors are constant.

Chapter 4

Empirical Validation

For the validation of our hypothesis, we utilized the data gathered from repeating the same experiment twice on two different sets of 15 volunteers and 15 webpages. Validation results for our hypothesis are:-

- 1. Y-axis:** In validation data Set I we have 15 webpages out of which all webpages satisfied constraints mentioned in previous chapter. In those 15 webpages, we have found **94** objects that are ranked correctly according to their Y distance from top i.e. rank of object increases from top to bottom, whereas **25** objects are not ranked appropriately.
In validation data Set II we have 15 webpages out of which 8 webpages satisfied constraints mentioned in previous chapter. In those 8 webpages, we have found **54** objects that are ranked correctly according to their Y distance from top i.e. rank of object increases from top to bottom, whereas **19** objects are not ranked appropriately.
Thus, approximately **77%** of all the objects from validation data concurs with our hypothesis.
- 2. Position:** In **67.11 %** slides user's first fixation was at the center of the screen. As our success rate percentage is noteworthy, we can conclude that our hypothesis is valid.
- 3. X-axis:** In **validation data I** we have **12** webpages in which **13** times left side object is seen before a right side object and **14** times right side object is seen before a left side object.
In **validation data II** we have **9** webpages in which **9** times left side object is seen before a right side object and **5** times right side object is seen before a left side object.
Hence, on the whole in about **53%** of the cases left side object is seen earlier than right side object. Therefore, it validates our belief that x-position of an object has no effect on its rank.
- 4. Size:** In Validation data **set I** and **set II**, we had 24 webpages that had large images. Out of those 24 web pages, in 22 web pages large images were ranked first. This implies that in **91.67%** of cases we got precise match. So, our hypothesis appears true.

5. **Contrast:** In validation data **set I** and **set II**, we have found that in 72.31% cases high contrast objects are seen earlier than low contrast objects.
6. **Mean:** In 63.57% cases in validation data, we have seen that mean grey scale value is lesser for the object seen earlier.

Chapter 5

Conclusion and Future Work

6.1 Conclusion

In this thesis, we reviewed the previous work done to predict the sequence in which web objects will be attended by a user. We found some limitations in the earlier suggested models.

We performed statistical analysis on the empirical data collected from a previous work to propose our hypothesis about the factors affecting human attention. For validating our work, we applied our hypothesis on two different empirical data sets.

However, there are some limitations in our work such as the lack of automatic identification of objects. Also, Human attention depends upon two things -Visual features of object and User's task. User's task includes user's aim such as the information user for which the user is searching. We did not considered such task in our work.

6.2 Future Work

A mathematical equation can be formulated for ranking objects from the factors discussed above.

References

- [1] Ball L. J. and Poole A., Eye tracking in human computer interaction and usability research: Current status and future prospects, Encyclopedia of Human Computer Interaction. Pennsylvania, pp. 101-106, 2005.
- [2] Cook W. D. and Cress M.: B Data Envelopment Model for aggregating Preference Rankings. Management Science Vol 16, November 1990.
- [3] Gravetter F.J. and Wallnau L.J. :Statistics for the Behavioral Sciences (8th ed.), Belmont, CA: Wadsworth Cengage Learning, 2009
- [4] Hinkelmann K., Kempthorne O.: Design and Analysis of Experiments. I , II (Second ed.). Wiley. ISBN 978-0-470-38551-7. MR 2363107, 2008
- [5] Itti L.: Visual Saliency, Scholarpedia, vol. 2, no. 9, p. 3327, 2007
- [6] Itti L., Gold C. and Koch C.: Visual attention and target detection in cluttered natural scenes, Optical Engineering, vol. 40, no. 9, pp. 1784-1793, Sep. 2001.
- [7] Itti L. and Koch C.: Feature combination strategies for saliency-based visual attention systems, Journal of Electronic Imaging, vol. 10, no. 1, pp. 161-169, 2001
- [8] Itti L., Koch C. and Niebur E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254-1259, 1998.
- [9] Jana A.: Model To Predict The Attention Sequence Of Web Page Objects, MTP Thesis, Dept. of Computer Science & Engineering, Indian Institute of Technology Guwahati, 2011-12.
- [10] Lau H. Y. and Shi J.: Improved illumination invariance using a color edge representation based on double opponent neurons neural networks, IEEE International Joint Conference on Neural Networks (IJCNN), pp. 2735-2741, Jun.2008.

- [11] Priyansh V.: Web Object Attention Model, MTP Thesis, Dept. of Computer Science & Engineering, Indian Institute of Technology Guwahati, 2010-11.
- [12] Schnipke S. K. and Todd M. W., Trials and tribulations of using an eye-tracking system, in CHI 00 extended abstracts on Human factors in computing systems, ser. CHI EA 00. New York, NY, USA: ACM, pp. 273-274, 2000.
- [13] Wolfe AJ. M.: Guided search 2.0 a revised model of visual search. Psychonomic Bulletin & Review, vol. 1, no. 2, p. 202-238, 1994.
- [14] [online] <http://www.alexas.com>
- [15] [online] <http://www.tobii.com>
- [16] [online] <http://www.zweigmedia.com/RealWorld/simplex.html>
- [17] Khade G.: Web Object Attention Model for Both Image and Text Objects, MTP Thesis, Dept. of CSE, Indian Institute of Technology Guwahati, 2012-13.

Appendix A

Webpages used in Empirical Data



WebPage 1



WebPage 3



WebPage 2



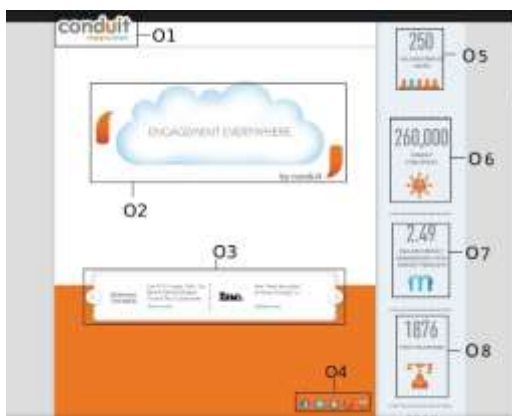
WebPage 6



WebPage 7



WebPage 5



WebPage 4



WebPage 12



WebPage 10



WebPage 11



WebPage 14



WebPage 15



WebPage 13



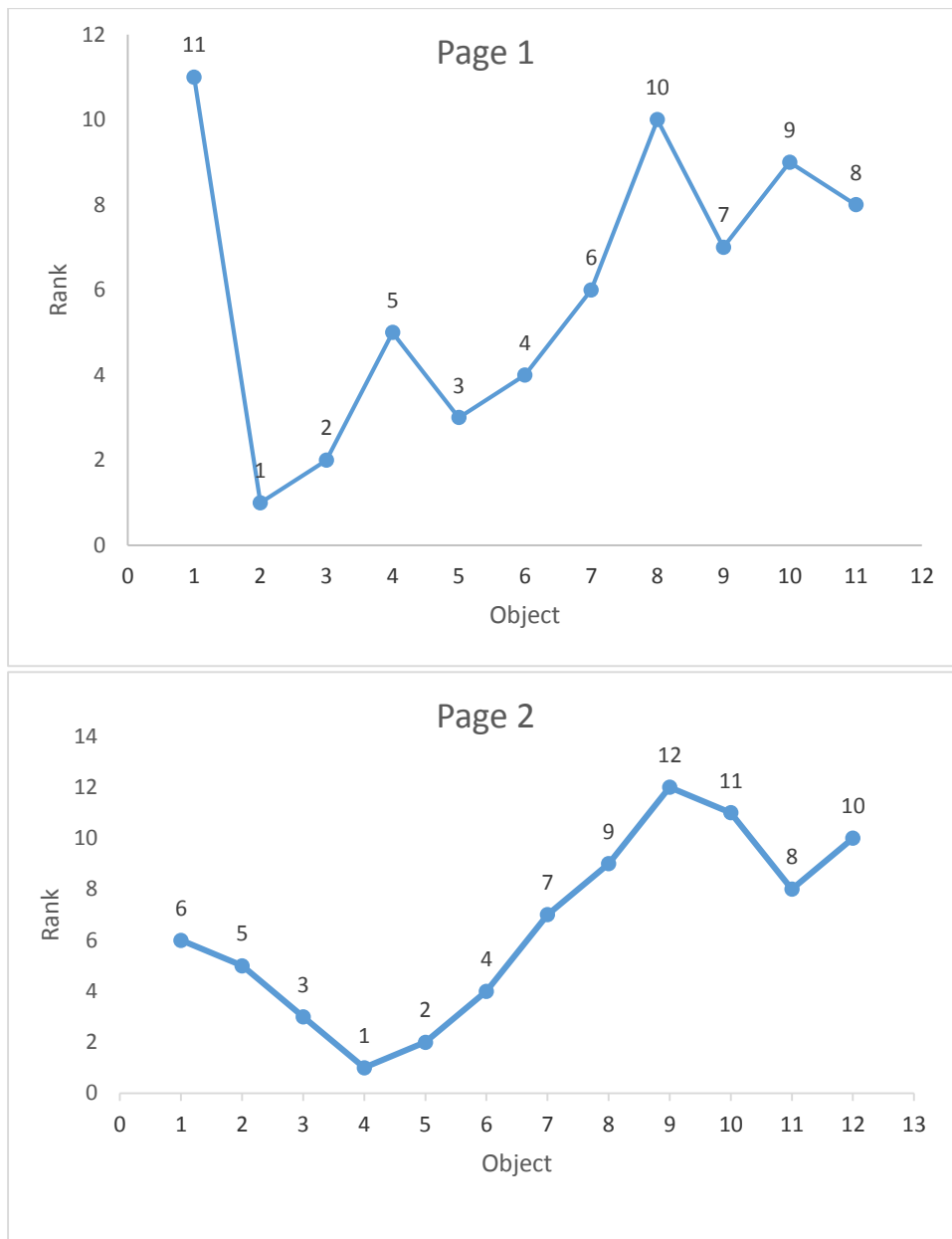
WebPage 8

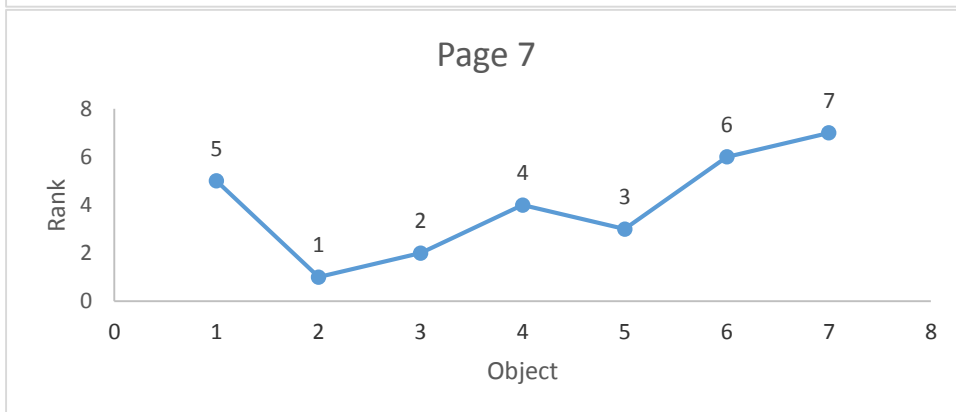
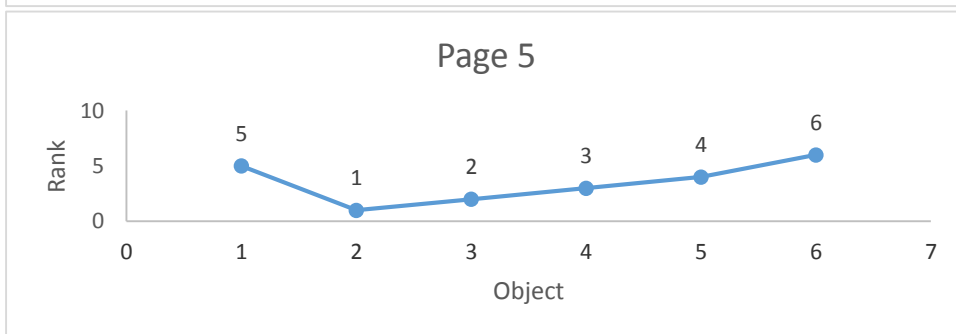
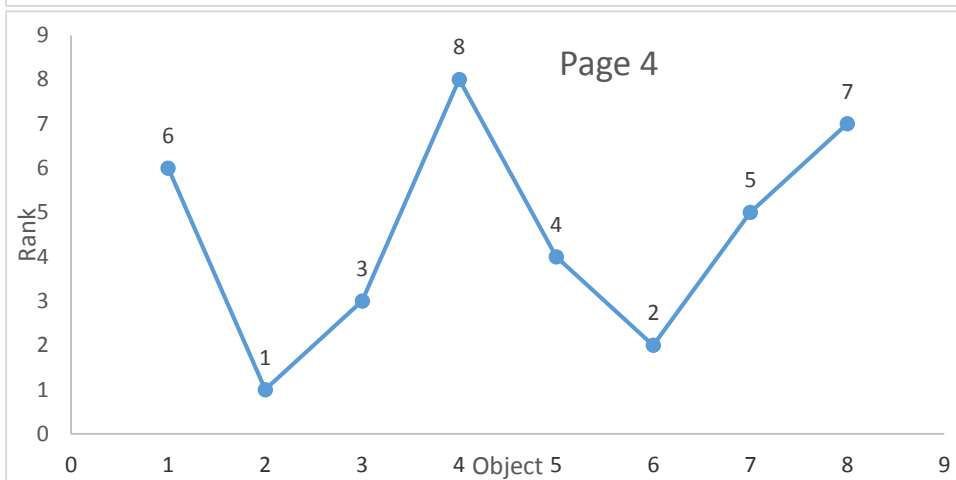
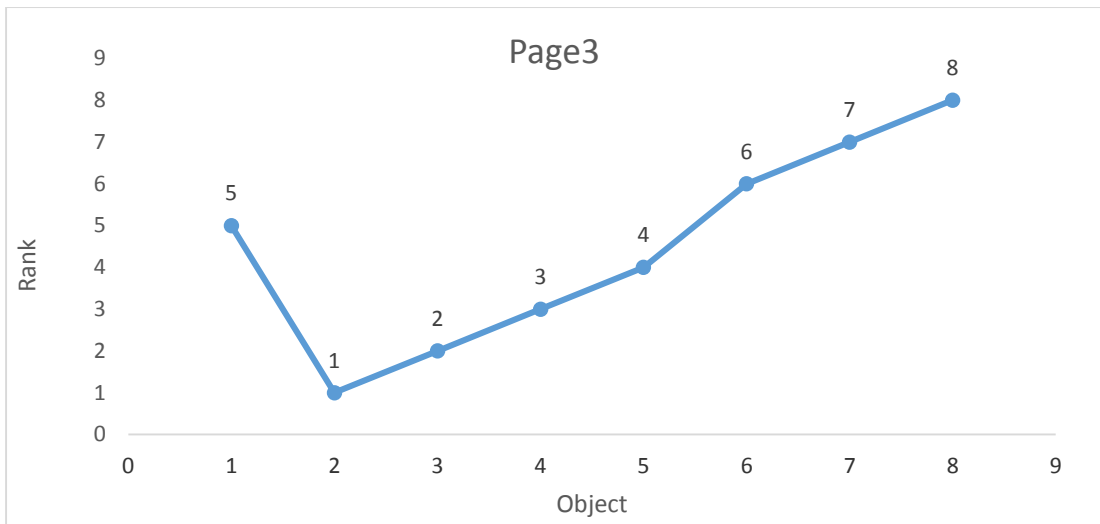


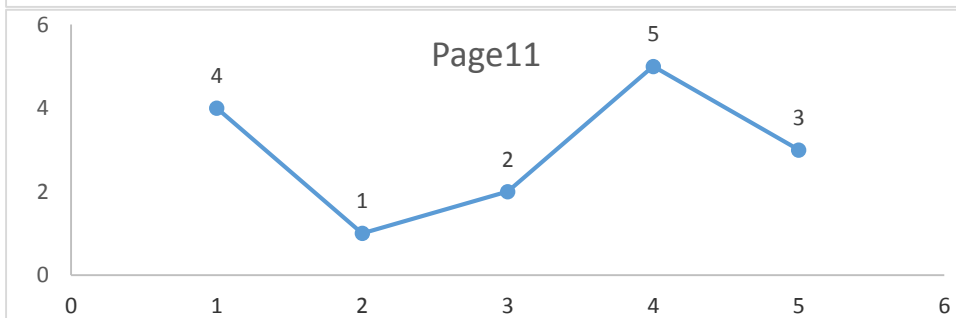
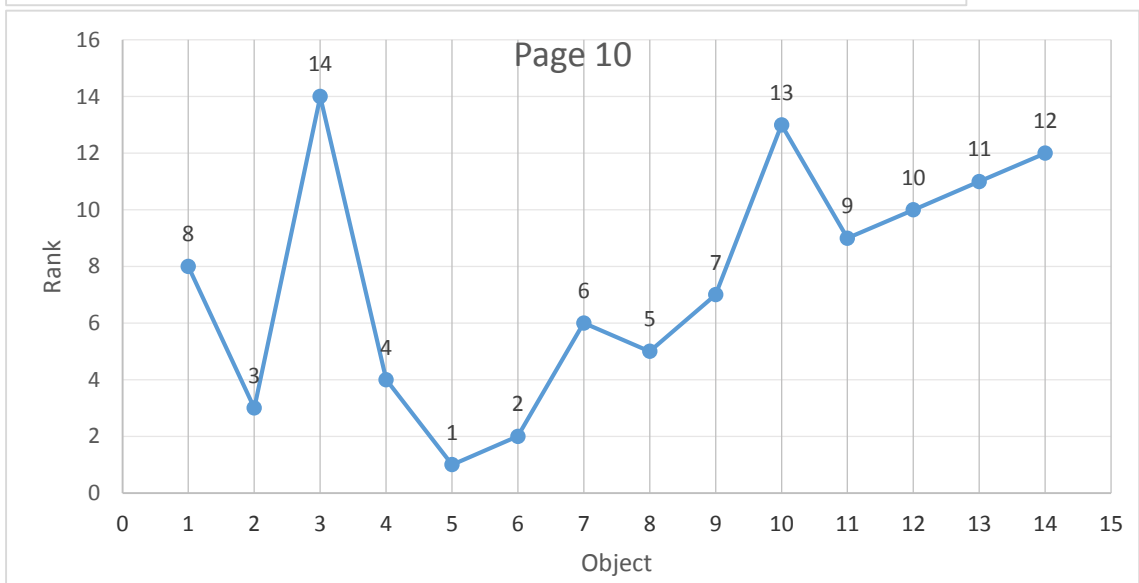
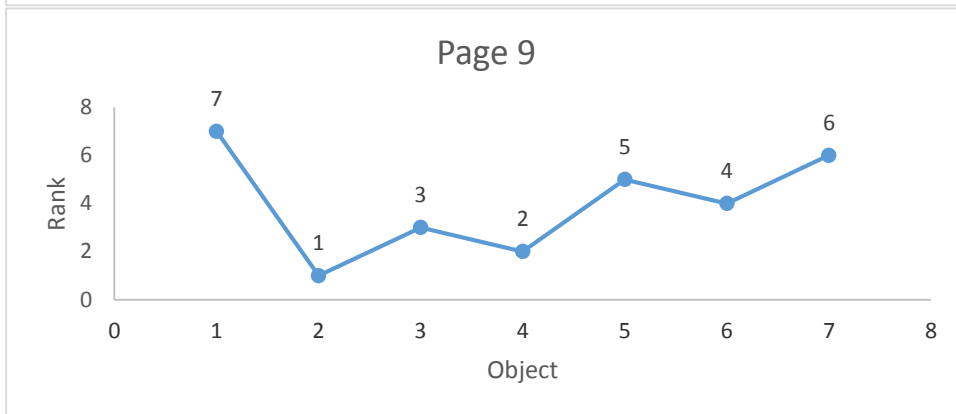
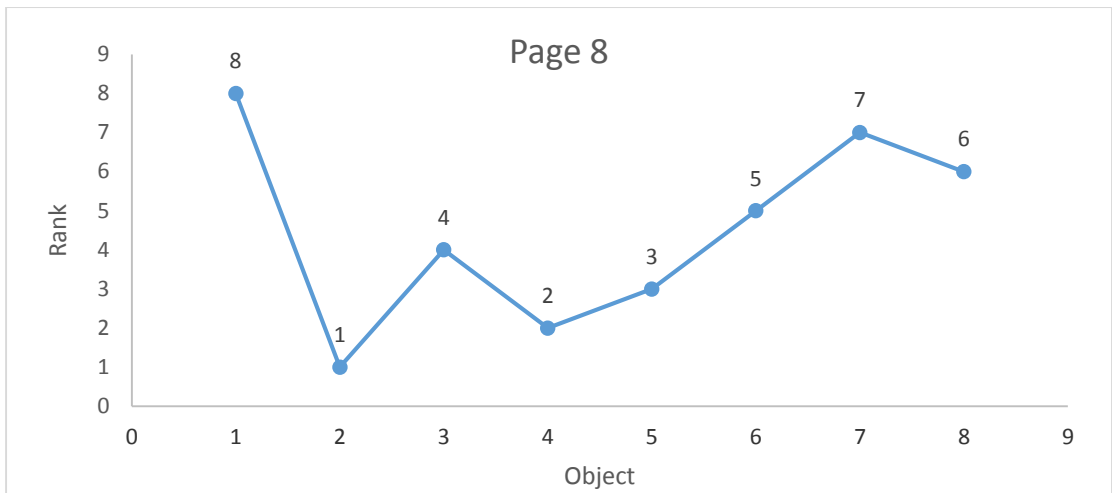
WebPage 9

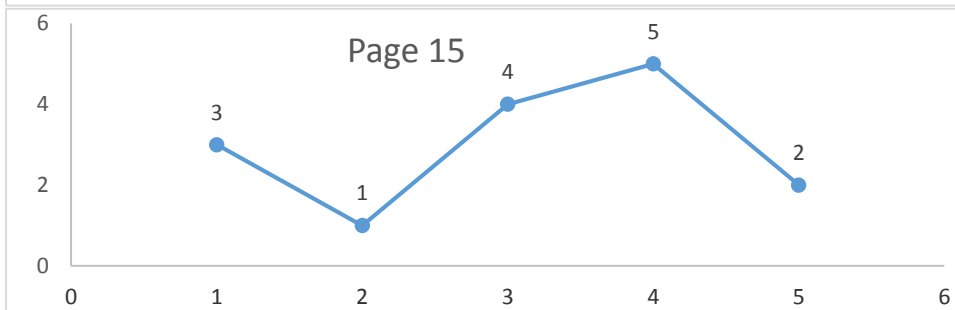
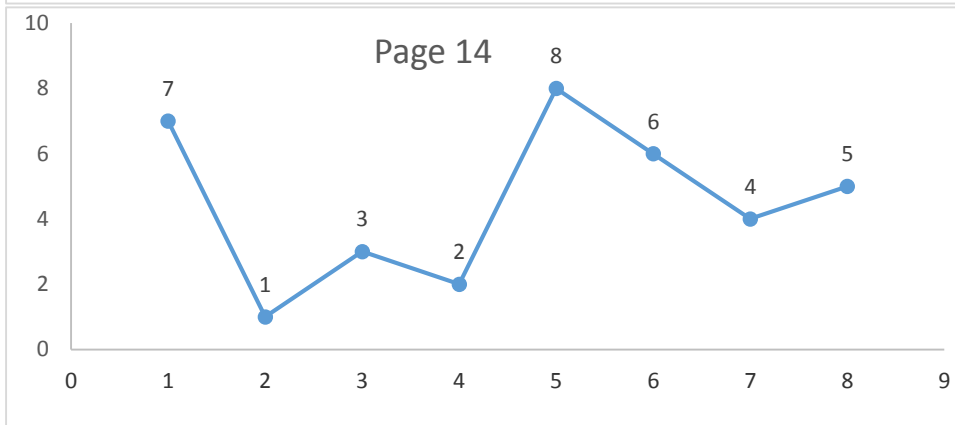
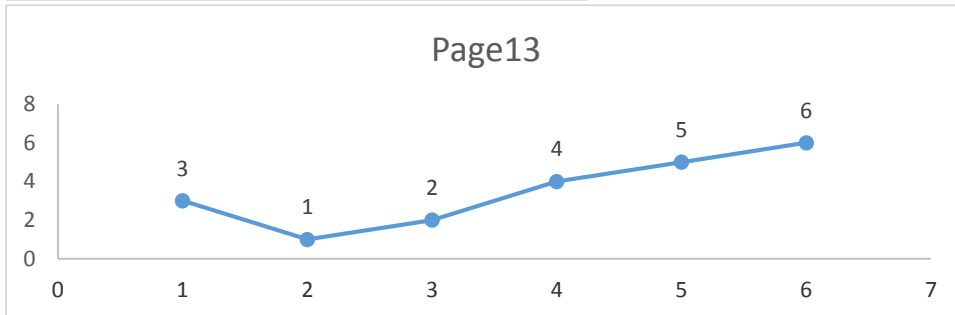
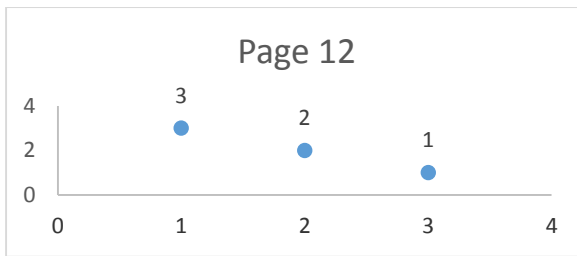
Appendix B

Observed Ranks of Objects in Empirical Data









Appendix C

Webpages used in Validation Data Set I



WebPage 1



WebPage 2



WebPage 3



WebPage 4



WebPage 5



WebPage 6



WebPage 7



WebPage 8



WebPage 9



WebPage 10



WebPage 11



WebPage 12



WebPage 13



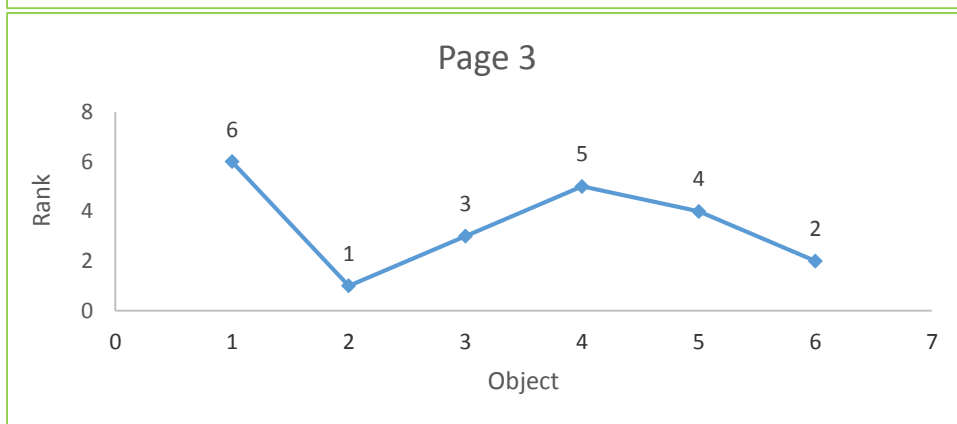
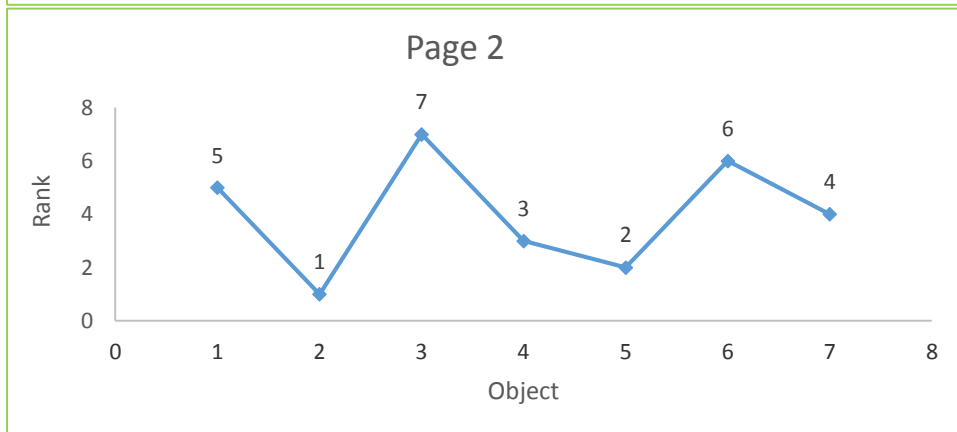
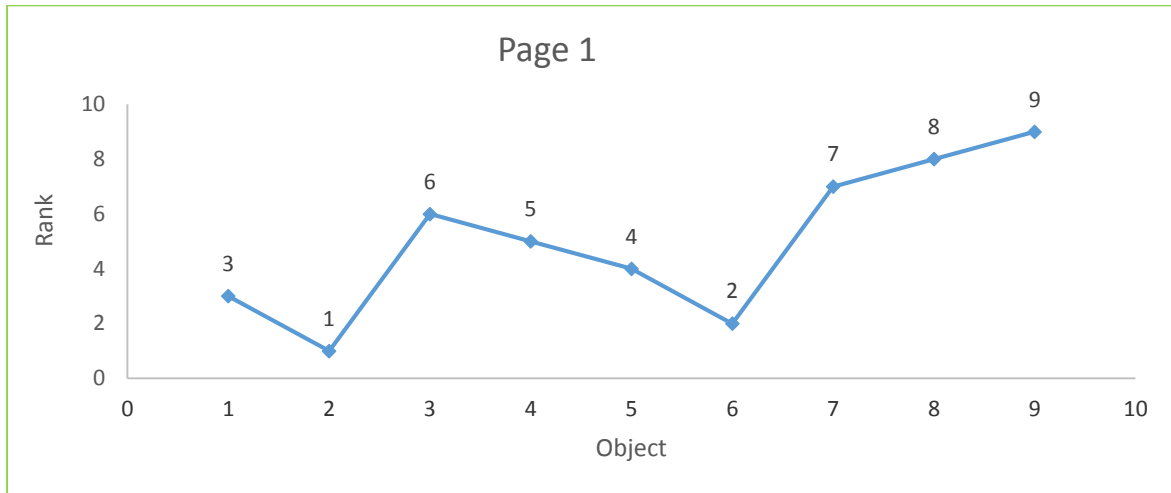
WebPage 14

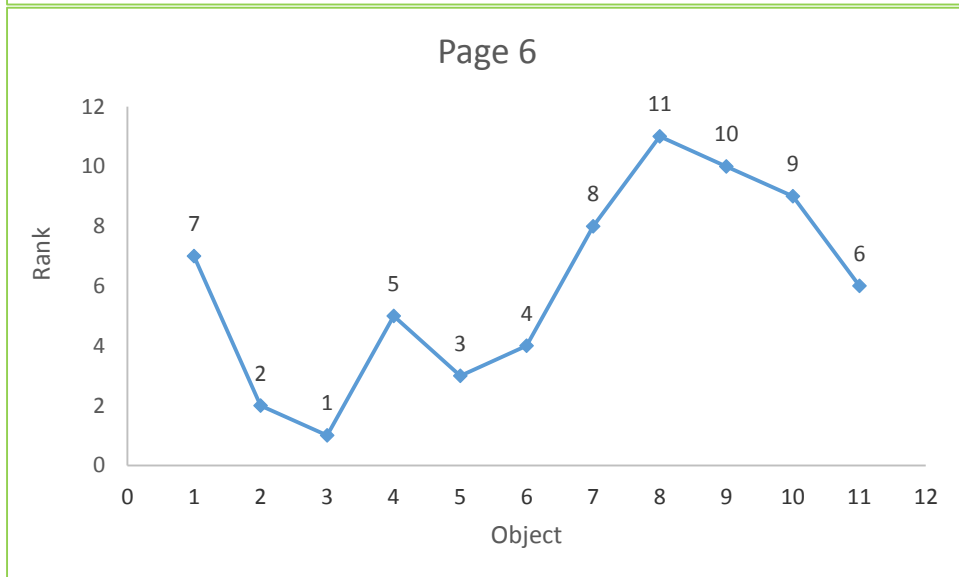
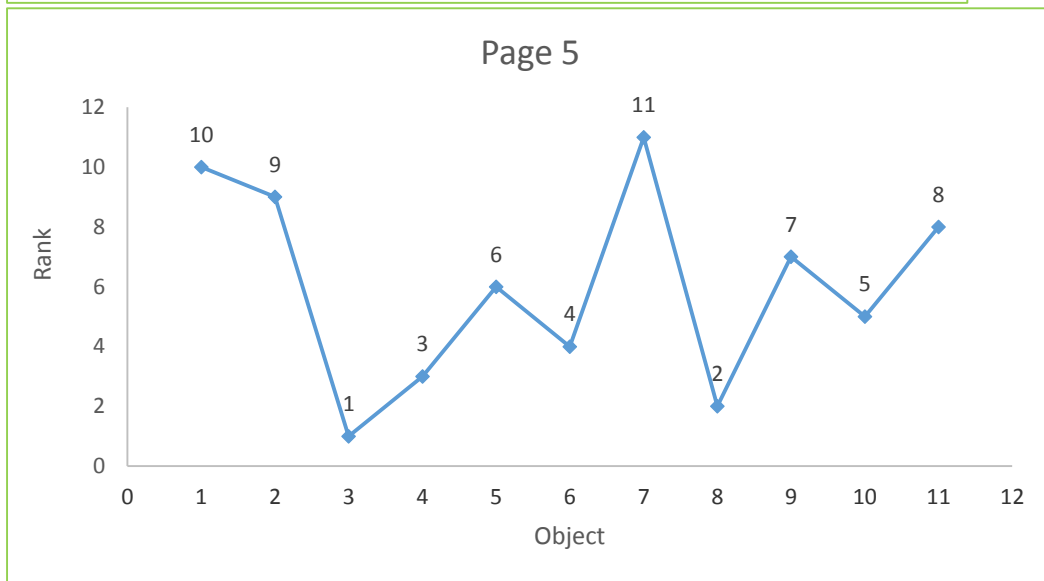
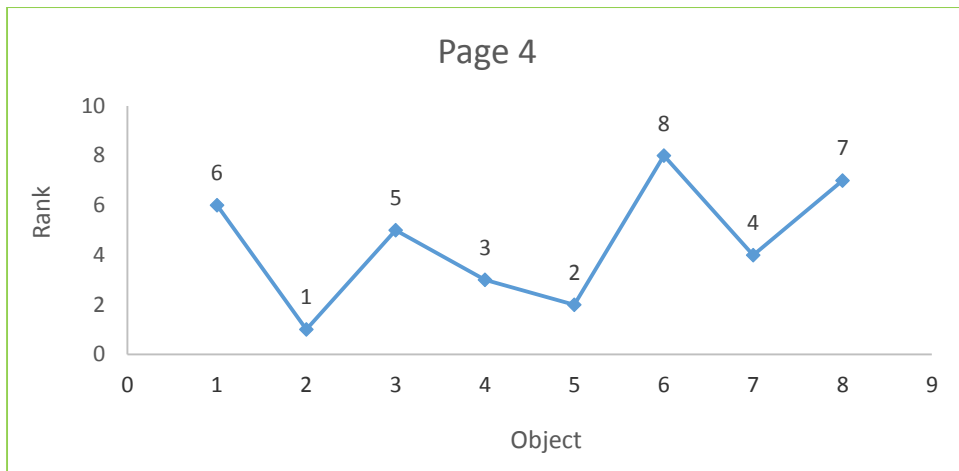


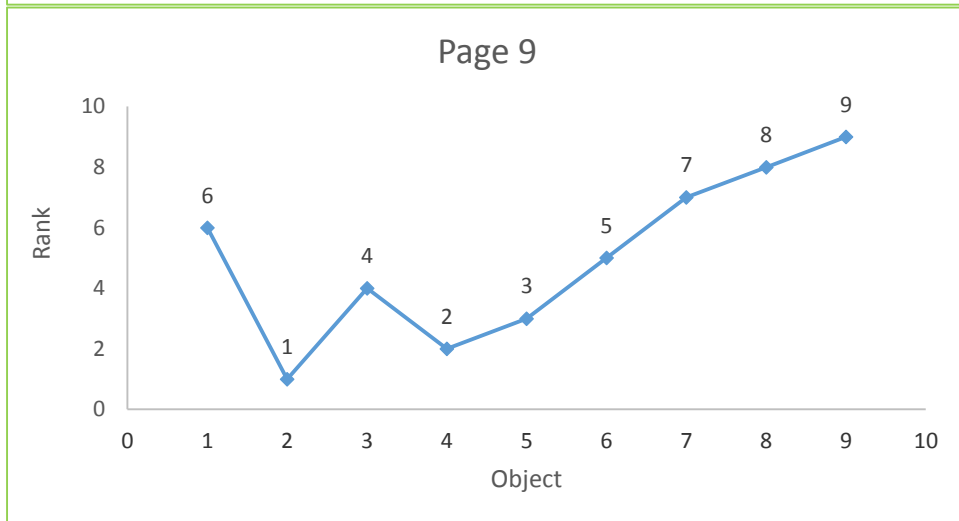
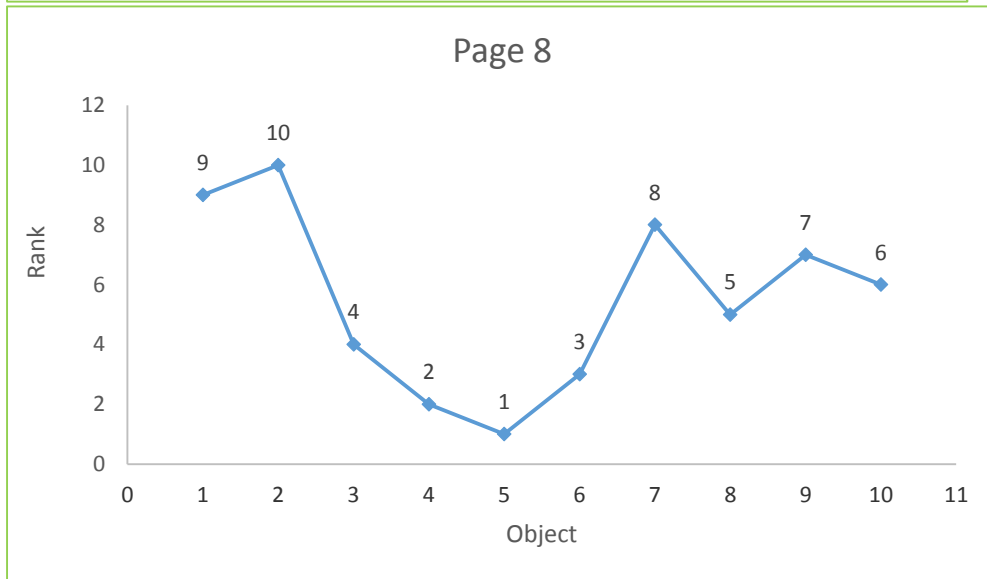
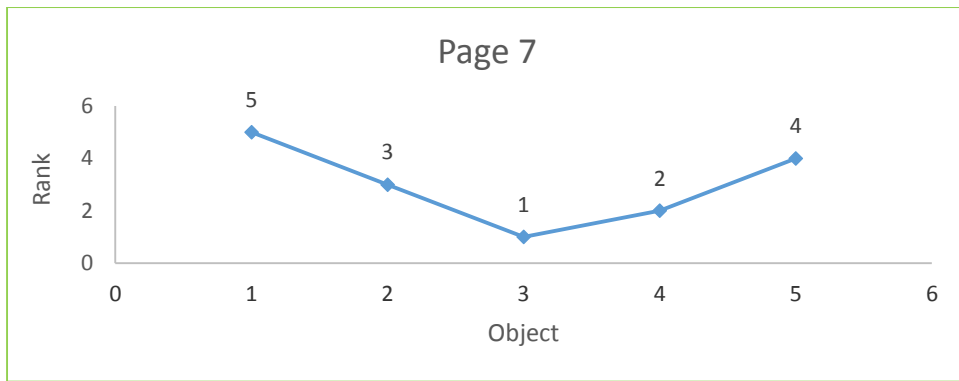
WebPage 15

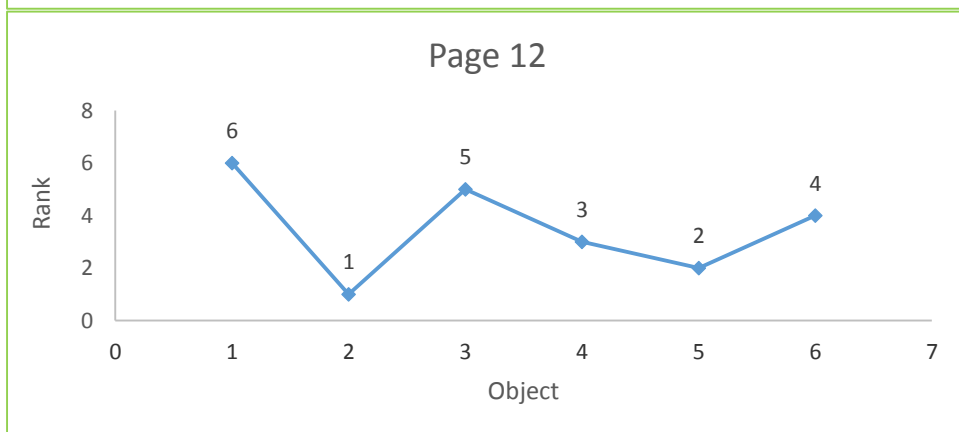
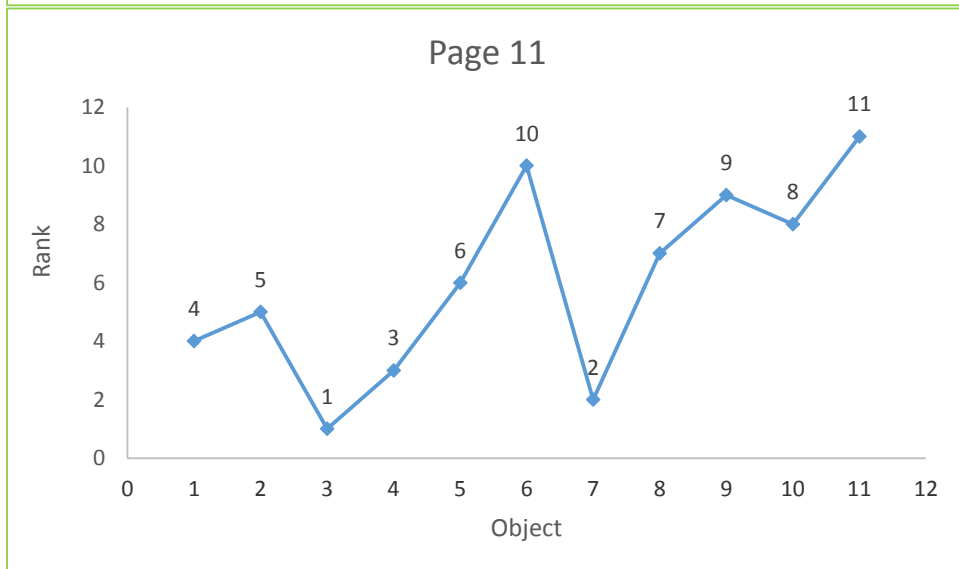
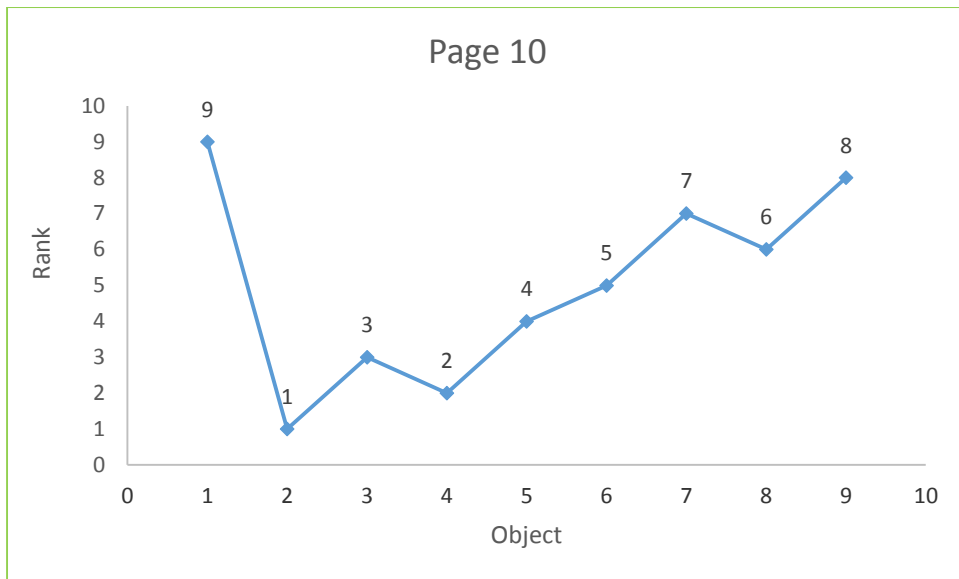
Appendix D

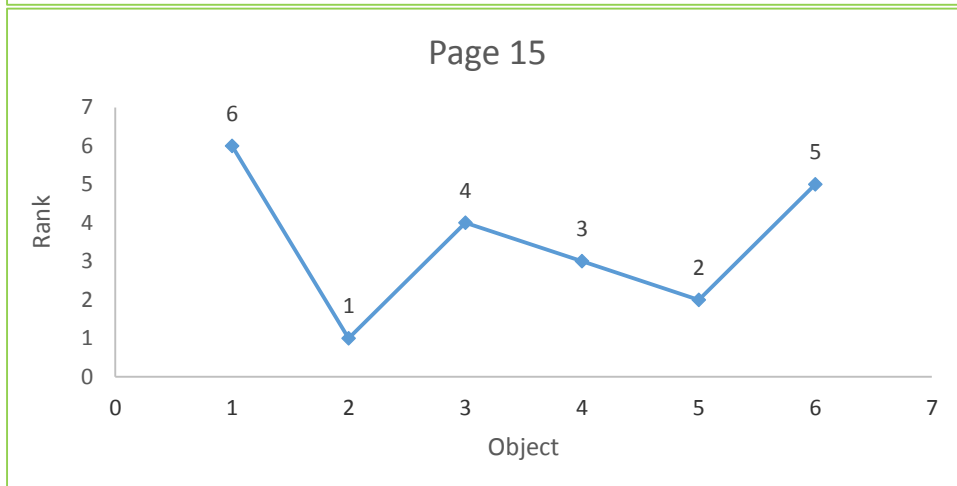
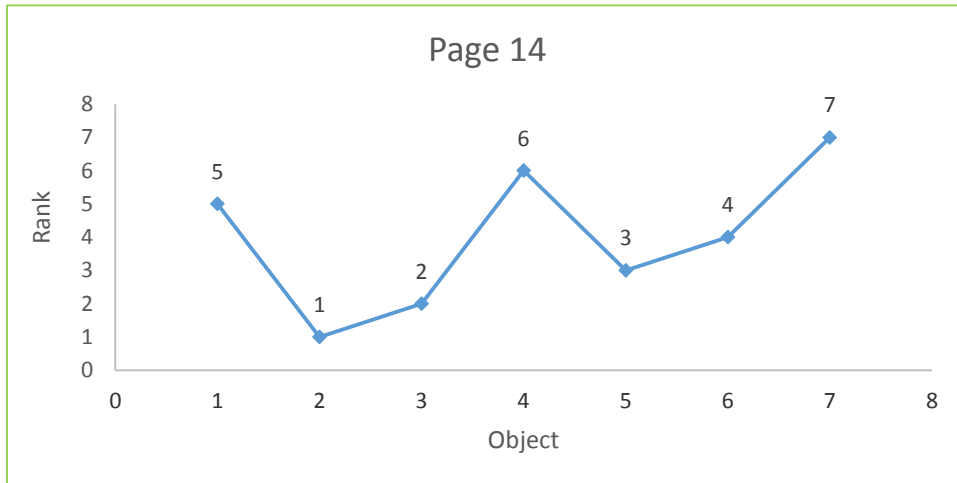
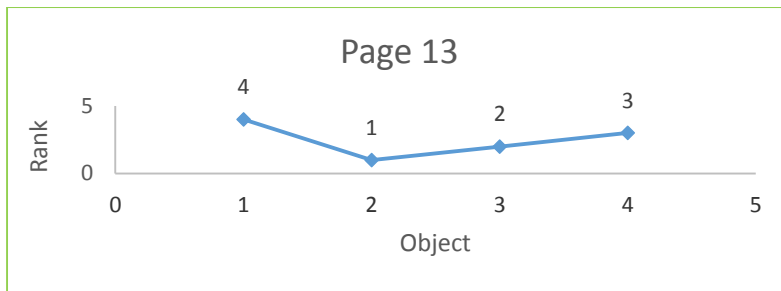
Observed Ranks of Objects in Validation Data Set I











Appendix E

Webpages used in Validation Data Set II



WebPage 1



WebPage 2



WebPage 3



WebPage 4



WebPage 5



WebPage 6



WebPage 7



WebPage 8



WebPage 9



WebPage 10



WebPage 11



WebPage 12



WebPage 13



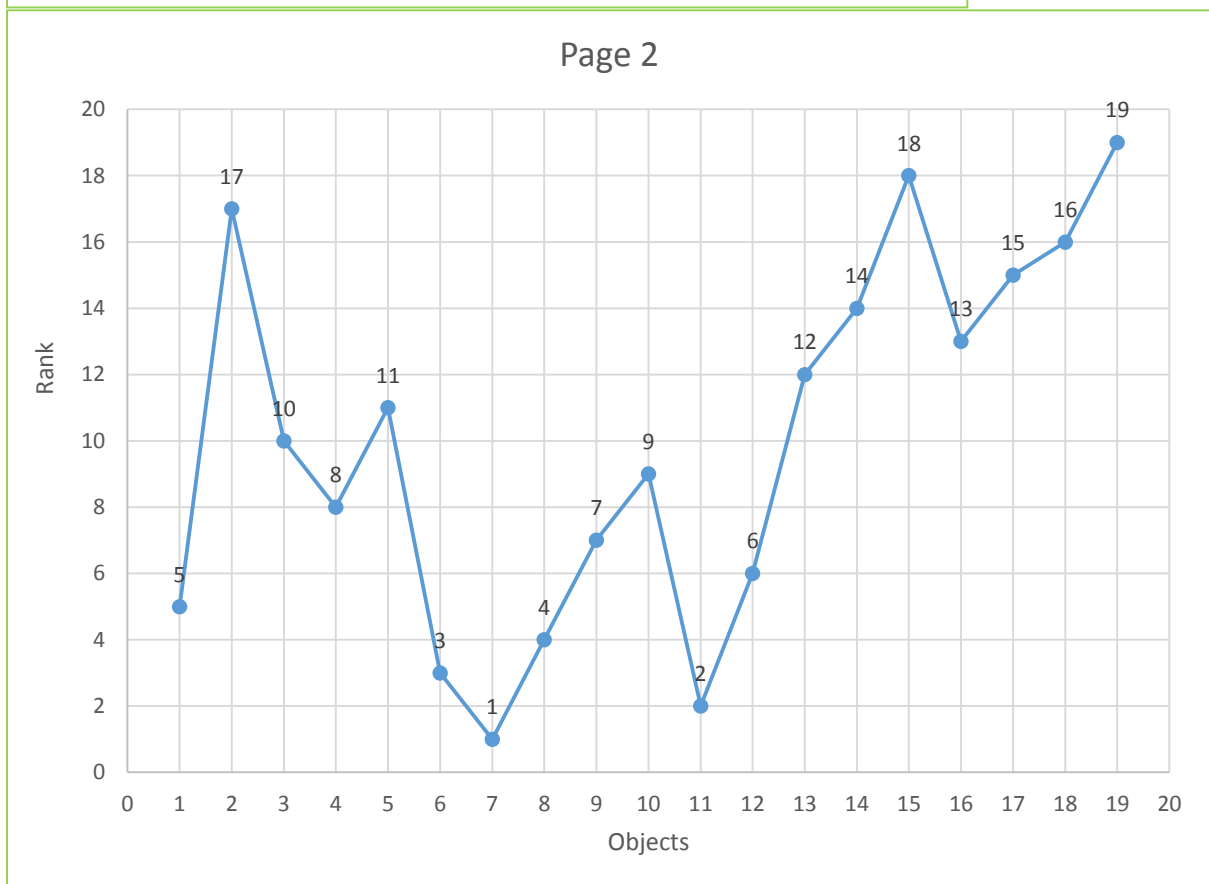
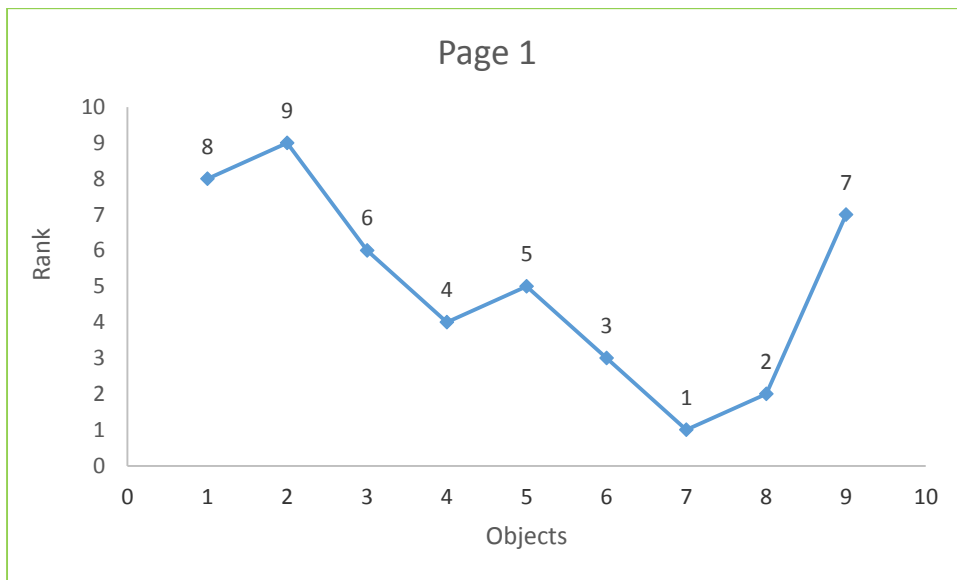
WebPage 14

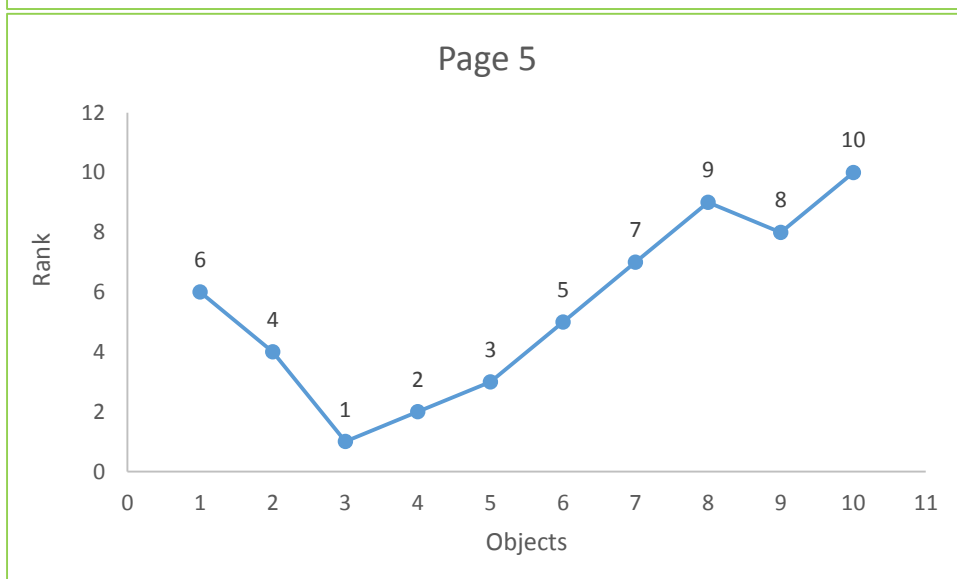
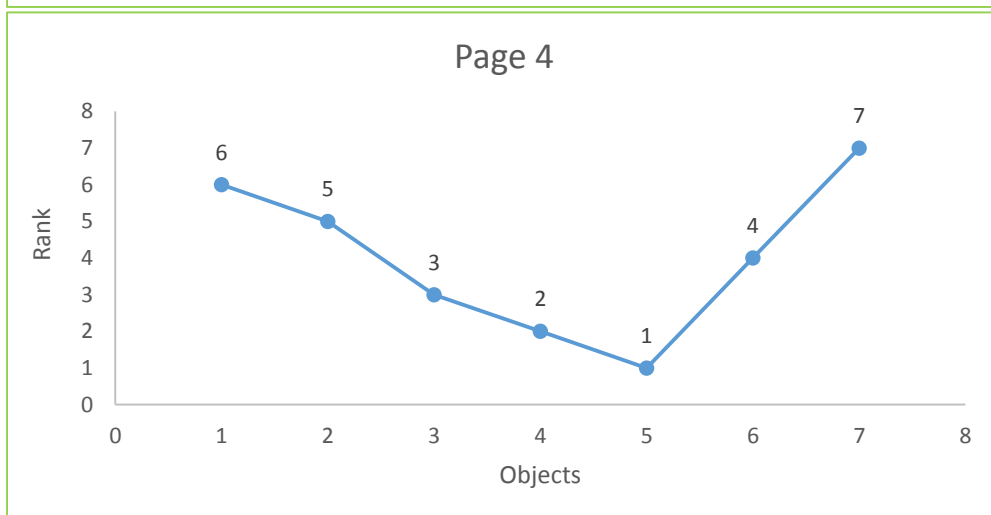
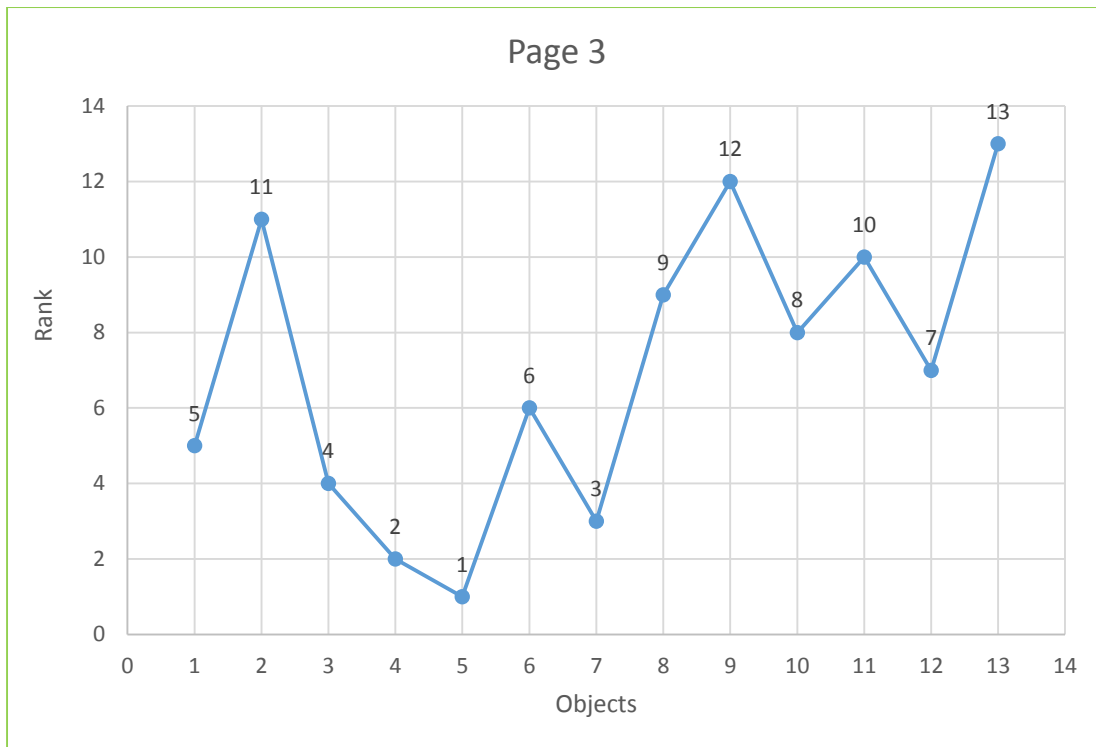


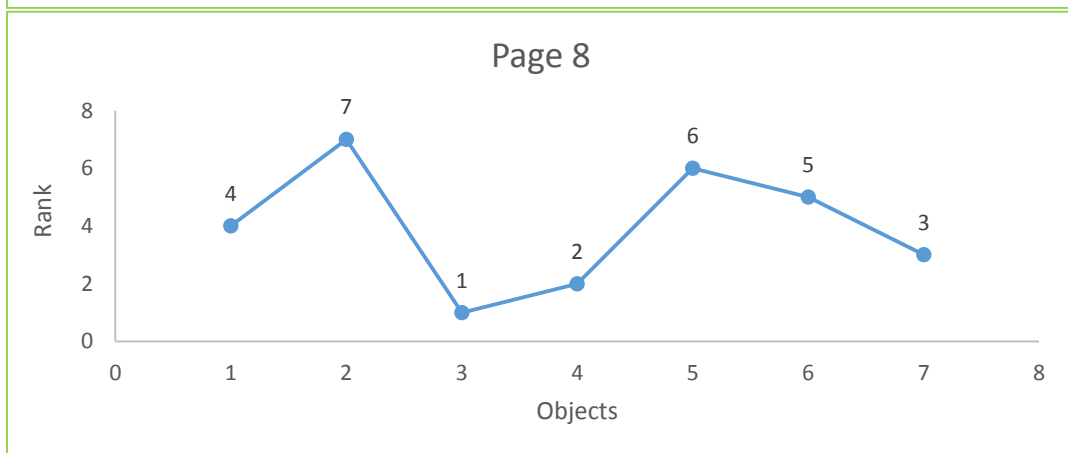
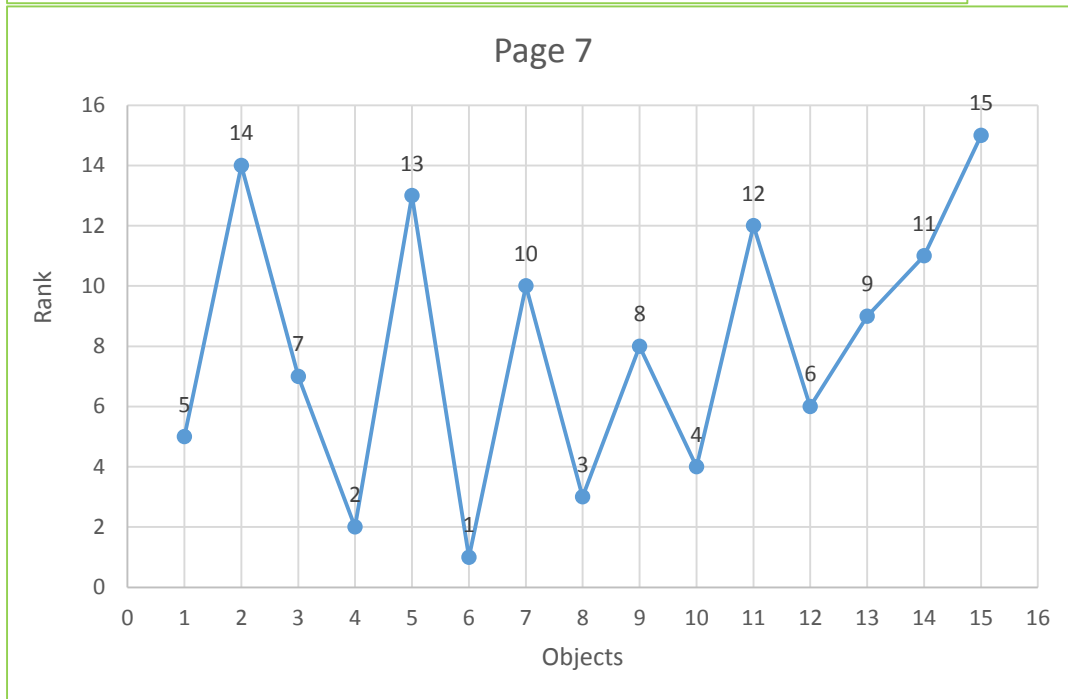
WebPage 15

Appendix F

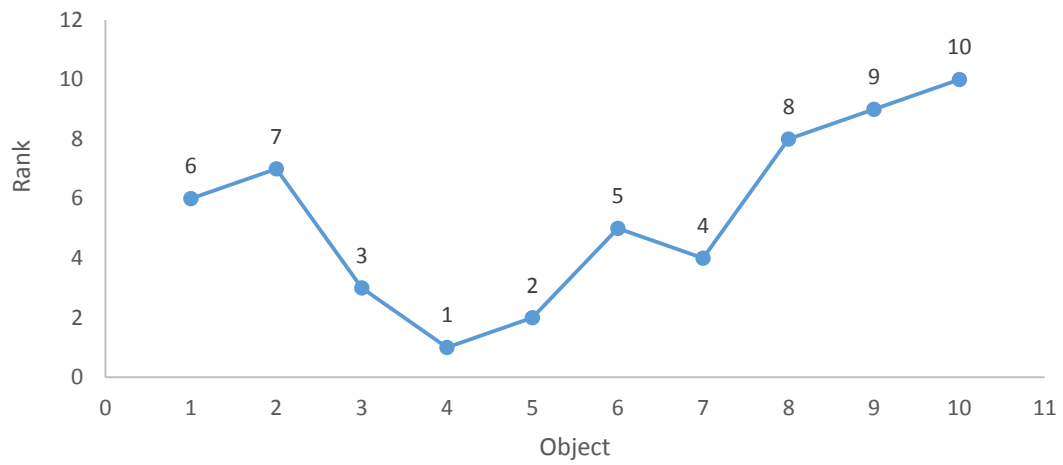
Observed Ranks of Objects in Validation Data Set II



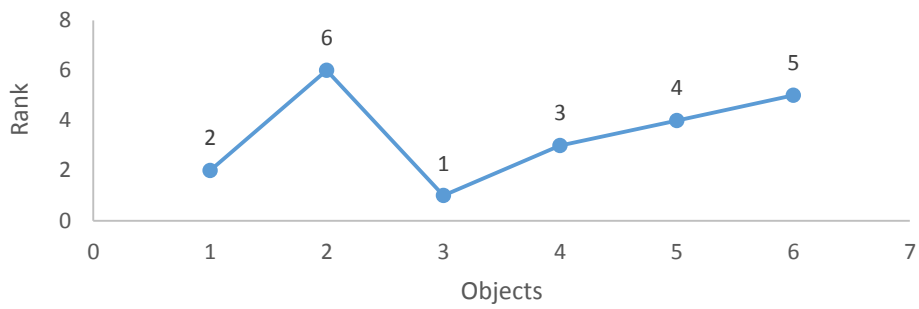




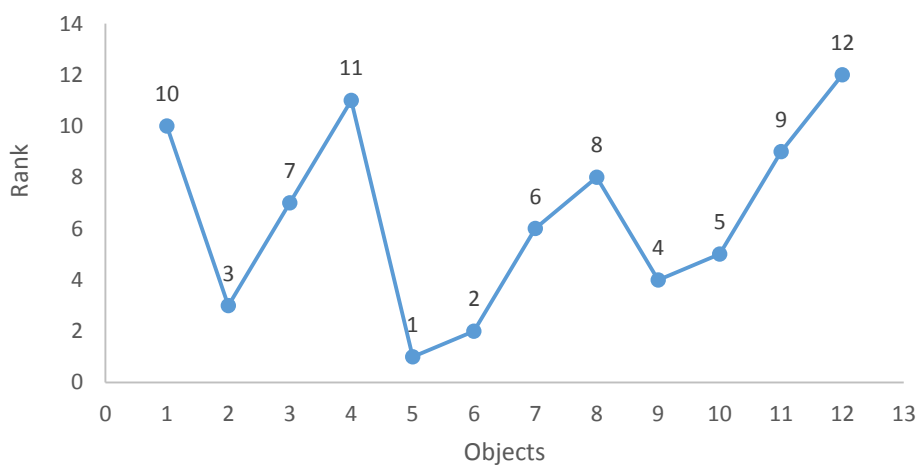
Page 9



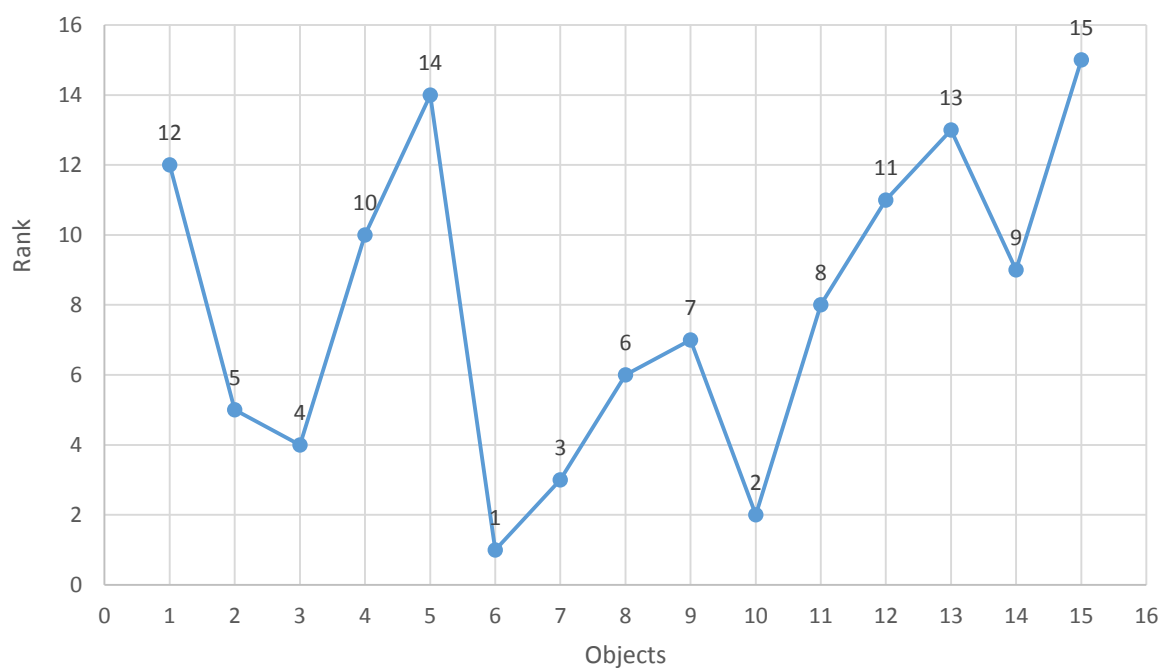
Page 10



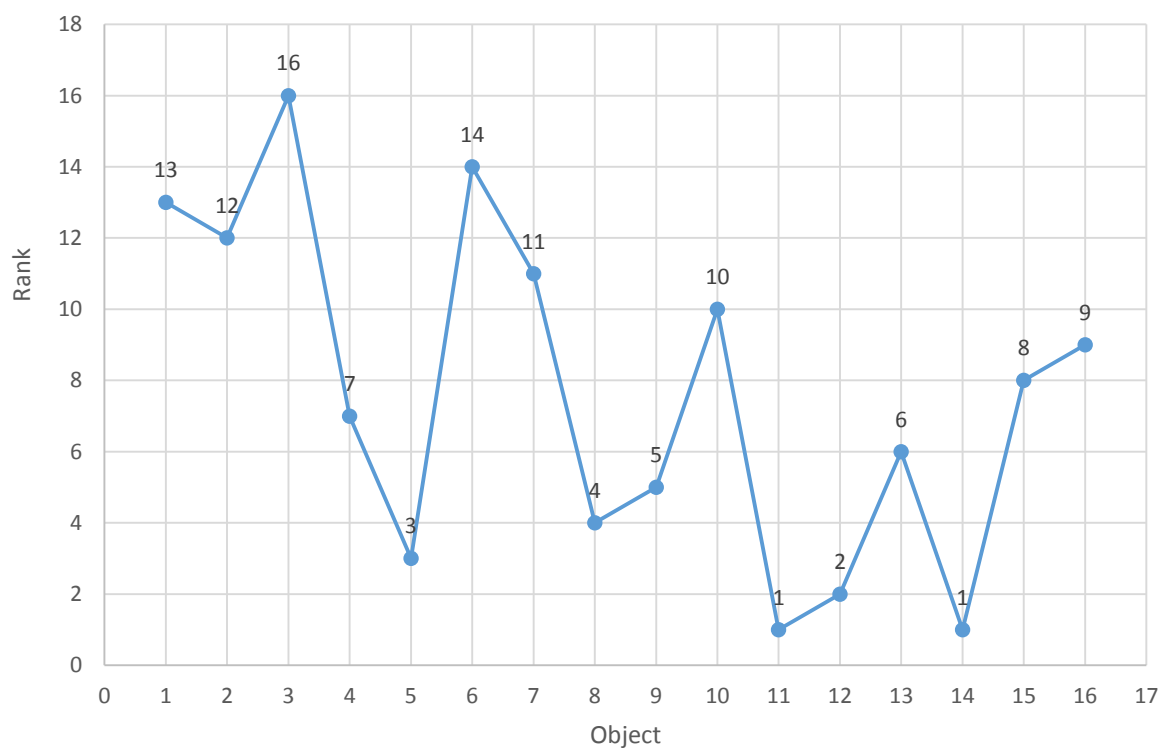
Page 11



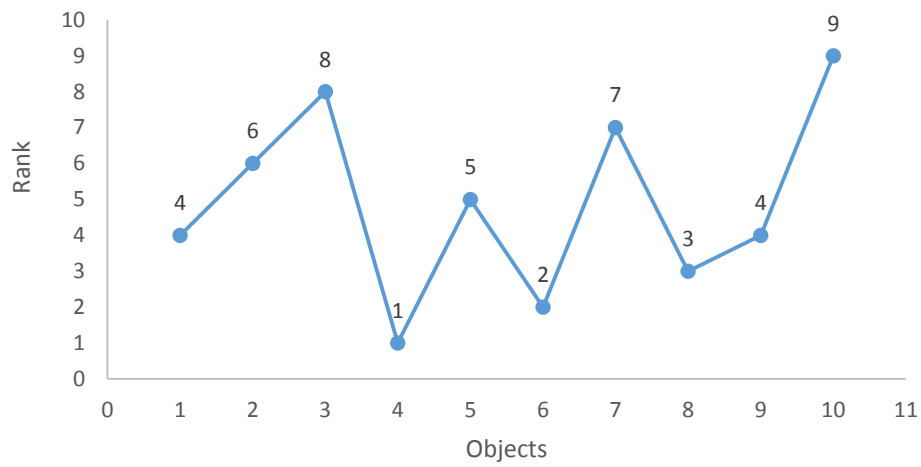
Page 12



Page 13



Page 14



Page 15

