

USING GOOGLE ANALYTICS AND THINK-ALOUD STUDY FOR IMPROVING  
THE INFORMATION ARCHITECTURE OF METU INFORMATICS INSTITUTE  
WEBSITE: A CASE STUDY

A THESIS SUBMITTED TO  
THE GRADUATE SCHOOL OF INFORMATICS  
OF  
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

SEHER DEMİREL KÜTÜKÇÜ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE  
IN  
THE DEPARTMENT OF INFORMATION SYSTEMS

SEPTEMBER 2010

Approval of the Graduate School of Informatics

\_\_\_\_\_  
Prof. Dr. Nazife Baykal  
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

\_\_\_\_\_  
Assist. Prof. Dr. Tuğba Taşkaya Temizel  
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

\_\_\_\_\_  
Assist. Prof. Dr. Tuğba Taşkaya Temizel  
Supervisor

Examining Committee Members

Assist. Prof. Dr. Erhan Eren (METU, IS) \_\_\_\_\_

Assist. Prof. Dr. Tuğba Taşkaya Temizel (METU, IS) \_\_\_\_\_

Dr. Ali Arifoğlu (METU, IS) \_\_\_\_\_

Assist. Prof. Dr. Sevgi Özkan (METU, IS) \_\_\_\_\_

Assist. Prof. Dr. Pınar Şenkul (METU, CENG) \_\_\_\_\_

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

**Name, Last name : Seher Demirel Kütükçü**

**Signature : \_\_\_\_\_**

## **ABSTRACT**

### **USING GOOGLE ANALYTICS AND THINK-ALOUD STUDY FOR IMPROVING THE INFORMATION ARCHITECTURE OF METU INFORMATICS INSTITUTE WEBSITE: A CASE STUDY**

Demirel Kütükçü, Seher

M.S., Department of Information Systems

Supervisor: Assist.Prof. Dr. Tuğba Taşkaya Temizel

September 2010, 176 pages

Today, web sites are important communication channels that reach a wide group of people. Measuring the effectiveness of these web-sites has become a key issue for researchers as well as practitioners. However, there is no consensus on how to define web site effectiveness and which dimensions need to be used for the evaluation of these web sites. This problem is more noteworthy for information driven web sites like academic web sites. There is limited academic literature in this predominant application area. The existing studies measured the effectiveness of these academic web sites by taking into account their information architecture mostly using think-aloud methodology. However, there is limited study on web analytics tools which are capable of providing valuable information regarding the web site users such as their navigation behaviours and browser details. Although web analytics tools provide detailed and valuable information, the existing studies have utilized their very basic features.

In this thesis, we have explored web analytic tools and think-aloud study method to improve information architecture of web sites. Taking METU Informatics Institute web site as a case study, we have used the reports of Google Analytics which is a commercial web analytics tool owned by Google and think-aloud study results to improve the information architecture of our case study web-site.

Keywords: Information Architecture, Web Analytics, Google Analytics, Landing Page Optimization, Think-Aloud

## ÖZ

### ODTÜ ENFORMATİK ENSTİTÜSÜ WEB SAYFASININ BİLGİ MİMARİSİNİ İYİLEŞTİRMEK İÇİN GOOGLE ANALİTİK VE SESLİ DÜŞÜN YÖNTEMİNİN KULLANILMASI: ÖRNEK OLAY ÇALIŞMASI

Demirel Kütükçü, Seher

Yüksek Lisans, Bilişim Sistemleri Bölümü

Tez Yöneticisi: Yrd. Doç. Dr. Tuğba Taşkaya Temizel

Eylül 2010, 176 sayfa

Günümüzde web sayfaları birçok kişiye ulaşabilmek için önemli bir iletişim kanalıdır. Bu web sayfalarının etkinliğini ölçmek araştırmacılar ve uygulayıcılar için kilit bir konu haline gelmiştir. Fakat, web sayfası etkinliğinin nasıl tanımlanması gerektiği ve web sayfalarını değerlendirirken hangi boyutların (dimension) kullanılması gerektiği konusunda fikir birliği yoktur. Bu problem akademik web sayfaları gibi bilgi ağırlıklı web sayfalarında daha dikkate değerdir. Uygulama ağırlıklı olan bu alanda, kısıtlı akademik çalışma yer almaktadır. Mevcut çalışmalarda bu akademik web sitelerinin başarısı bu sayfaların bilgi mimarileri dikkate alınarak çoğunlukla sesli-düşün yöntemi kullanılarak ölçülmektedir. Fakat, gezinme davranışları ve kullanılan tarayıcı programı detayları gibi web sayfası kullanıcılarına ait değerli bilgiler sunma kapasitesine sahip web analitik araçları ile ilgili kısıtlı çalışma bulunmaktadır. Web analitik araçları detaylı ve değerli bilgi sunmalarına karşın, mevcut çalışmalar bu araçların çok temel özelliklerinden faydalanmaktadır.

Bu tez çalışmasında, web sayfalarının bilgi mimarisini iyileştirmek için web analitik araçlarını ve sesli-düşün yöntemini araştırdık. ODTÜ Enformatik Enstitüsü web sayfasını örnek olay olarak ele alarak bu sayfanın bilgi mimarisini iyileştirmek amacıyla Google firmasına ait ticari bir web analitik aracı olan Google analitik raporlarını ve sesli-düşün çalışması sonuçlarını kullandık.

Anahtar Kelimeler: Bilgi Mimarisi, Web Analitik, Google Analitik, Açılış Sayfası Optimizasyonu, Sesli Düşün

To the memory of my grandfather Ahmet Demirel,

## **ACKNOWLEDGMENTS**

This was a long study for me during which I got financial, technical and motivational support from many people around me. I want to express sincere appreciation to Assist. Prof. Dr. Tuğba Taşkaya Temizel for her guidance, insight and patience throughout the research as my advisor. My financial support was from The Scientific and Technological Research Council of Turkey (TUBİTAK) without whom this study can not be finalized, thanks for its support. I cannot forget the technical support of the Research Assistants Süleyman Özarlan, Fatih Ömrüzun and İbrahim Arpacı, thanks for their cooperation. My dear friends Birgül Çakır and Güliz Karaaslan worked like my secretariat for finding participants to the think-aloud study, many thanks goes to them from me. Last thank goes to my husband who really worked hard to test my motivation by asking me to go out at every occasion.

# TABLE OF CONTENTS

|  |      |
|--|------|
| ABSTRACT .....   | iv   |
| ÖZ .....   | vi   |
| DEDICATION.....  | viii |
| ACKNOWLEDGMENTS.....   | ix   |
| TABLE OF CONTENTS .....  | x    |
| LIST OF TABLES.....  | xiv  |
| LIST OF FIGURES .....  | xvii |
| CHAPTER  |      |
| 1. INTRODUCTION .....  | 1    |
| 2. LITERATURE REVIEW .....   | 4    |
| 2.1 Website Evaluation .....   | 4    |
| 2.2 Information Architecture .....   | 12   |
| 2.3 Landing Page Optimization .....  | 16   |
| 2.4 Web Analytic Tools and Google Analytics .....  | 17   |
| 2.4.1 Web Analytic Tools.....  | 18   |
| 2.4.2 What is Google Analytics? .....  | 20   |
| 2.4.3 How does Google Analytics function?.....   | 21   |
| 2.4.4 What Google Analytics can and cannot do?.....                                      | 22   |
| 2.4.5 What are dimension and metric and how to get reports in Google<br>Analytics? ..... | 26   |
| 2.4.6 How Google Analytics is used and what results are obtained? .....                  | 29   |

|       |  |    |
|-------|--|----|
| 2.4.7 | Used Dimensions and Metrics in the Literature for Information<br>Driven Web Sites.....   | 32 |
| 3.    | METHODOLOGY: GOOGLE ANALYTICS .....  | 41 |
| 3.1   | Properties of the Institute Web-Site Chosen As Case Study .....  | 42 |
| 3.2   | How did we prepare Google Analytics reports?.....  | 50 |
| 3.3   | Time Period Selection .....  | 50 |
| 3.4   | Important Findings about Google Analytic Reports .....   | 50 |
| 3.4.1 | Misleading Keyword and LandingPagePath Pairs .....   | 52 |
| 3.4.2 | Double Pageview Reporting Problem and Its Reason.....  | 53 |
| 3.5   | Detailed Explanations of Google Analytics Dimensions and Metrics .....   | 55 |
| 3.5.1 | LandingPagePath.....   | 55 |
| 3.5.2 | NewVisits and VisitorType.....   | 58 |
| 3.5.3 | TimeonSite, TimeOnPage, Visits and Entrances .....   | 59 |
| 3.5.4 | PageDepth Dimension vs PageViews Metric .....  | 61 |
| 3.5.5 | ReferralPath or Source?.....   | 63 |
| 3.5.6 | Visitor Path Identification using LandingPagePath, Keyword,<br>PagePath, PreviousPagePath, NextPagePath Dimensions and<br>Entrances and Exits Metrics..... | 64 |
| 3.5.7 | Summary of What We Learnt on Google Analytics Dimensions and<br>Metrics .....  | 67 |
| 3.6   | Matching Of What Visitors Possess And What The Web-Site Requires .....   | 69 |
| 3.7   | Landing Page Optimization .....  | 70 |
| 3.8   | Problematic Keyword Clusters Which Landed on Informatics Online<br>Program Web pages.....  | 74 |
| 3.8.1 | Problematic Keyword Cluster 1 .....  | 74 |
| 3.8.2 | Problematic Keyword Cluster 2 .....  | 79 |
| 3.8.3 | Problematic Keyword Cluster 3 .....  | 80 |
| 3.8.4 | Problematic Keyword Cluster 4 .....  | 81 |
| 3.8.5 | Problematic Keyword Cluster 5 .....  | 82 |

|        |  |     |
|--------|--|-----|
| 3.8.6  | Problematic Keyword Cluster 6 .....  | 83  |
| 3.8.7  | Problematic Keyword Cluster 7 .....  | 84  |
| 3.8.8  | Problematic Keyword Cluster 8 .....  | 85  |
| 3.9    | Problematic Keyword Clusters Which Landed on Information Systems<br>Program Web pages.....   | 86  |
| 3.9.1  | Problematic Keyword Cluster 1 .....  | 87  |
| 3.9.2  | Problematic Keyword Cluster 2 .....  | 88  |
| 3.10   | Problematic Keyword Clusters Which Landed on Software Management<br>Program Web-pages .....  | 89  |
| 3.11   | Problematic Keyword Clusters Which Landed on Work Based Learning<br>Program Web pages.....   | 91  |
| 3.11.1 | Problematic Keyword Cluster 1 .....  | 91  |
| 3.11.2 | Problematic Keyword Cluster 2 .....  | 93  |
| 3.11.3 | Problematic Keyword Cluster 3 .....  | 94  |
| 3.12   | Problematic Keyword Clusters Which Landed on the Web pages related to<br>Informatics Online, Information Systems and Software Management<br>Programs ..... | 95  |
| 3.13   | Problematic Keyword Clusters Which Landed on Web pages of both<br>Informatics Online and Software Management Programs.....                                 | 96  |
| 3.14   | Summary Of The Results.....  | 97  |
| 4.     | METHODOLOGY: THINK-ALLOUD STUDY .....  | 100 |
| 4.1    | How We Prepared the Questionnaire? .....   | 100 |
| 4.1.1  | Promotion activities of the on-line Survey: .....  | 105 |
| 4.2    | Participants .....   | 106 |
| 4.3    | Materials .....  | 107 |
| 4.4    | Procedures.....  | 108 |
| 4.5    | Sample Size .....  | 109 |
| 4.6    | Testing Validity of the Results .....  | 111 |
| 4.7    | Summary Of The Results .....   | 113 |

5. CONCLUSION..... 128

REFERENCES ..... 134

APPENDICES ..... 144

    A. METU Human Subjects Ethics Committee (HSEC) Approval ..... 145

    B. Voluntary Participation Information ..... 146

    C. 2009 Traffic From Search Engine Report..... 147

    D. Web-Site Addresses Of The Chosen Programs ..... 148

    E. Code Used To Get The Report..... 151

    F. Think-Aloud Study Questions..... 156

## LIST OF TABLES

|  |    |
|--|----|
| Table 1: Used Dimensions and Metrics in the Literature for Content Sites .....                 | 34 |
| Table 2: Google Analytics Dimension and Metric Abbreviation List.....                          | 42 |
| Table 3: 2009 Period Top 30 Landing Pages; Their Entrances, Bounces and Bounce Rate .....      | 46 |
| Table 4 : Selected Period Top 30 Landing Pages; Their Entrances, Bounces and Bounce Rate ..... | 49 |
| Table 5: Misleading Keyword-LandingPagePath Pair Report .....                                  | 53 |
| Table 6: Double PageView Reporting .....   | 53 |
| Table 7: Reason Behind Double PageView Reporting.....  | 54 |
| Table 8: Keyword-LandingPagePath Pairs Test 1.....   | 56 |
| Table 9 : Keyword-LandingPagePath Pairs Test 2.....  | 57 |
| Table 10: Detecting the Entrance Keyword .....   | 57 |
| Table 11: New Visits and VisitorType Example.....  | 59 |
| Table 12: TimeonSite, TimeonPage, Entrances Cumulative Report.....                             | 60 |
| Table 13: Distinction between TimeonSite and TimeonPage Detailed Report.....                   | 60 |
| Table 14: PageDepth vs PageViews Cumulative Report.....  | 62 |
| Table 15: Distinction between PageDepth and PageViews Detailed Report.....                     | 62 |
| Table 16: ReferralPath Example Report .....  | 64 |

|   |    |
|---|----|
| Table 17: Source Example Report .....   | 64 |
| Table 18: Visitor Session Information, Source: google, Keyword: aspects of control and<br>complementation in Turkish, NetworkLocation: comcast cable communications holdings<br>inc, LandingPagePath: /thesis/2009/aspects-control-and-complementation-turkish and<br>PageDepth: 14 ..... | 66 |
| Table 19: Single Session Information .....  | 67 |
| Table 20: Example Complex Report .....  | 68 |
| Table 21: Example from Custom Report for Landing Page Optimization .....  | 71 |
| Table 22: Selected keywords to be analyzed.....   | 73 |
| Table 23: “uzaktan master” Problematic Keyword Cluster Report.....  | 75 |
| Table 24: “master degree” Problematic Keyword Cluster Report .....  | 79 |
| Table 25: “odtü tezsiz yüksek lisans” Problematic Keyword Cluster Report.....   | 80 |
| Table 26: “informatics online course” Problematic Keyword Cluster Report.....   | 81 |
| Table 27: URL of the Informatics Online program Problematic Keyword Cluster Report<br>.....   | 82 |
| Table 28: “bilişim yüksek lisans ankara” Problematic Keyword Cluster.....   | 83 |
| Table 29: “master Ankara” Problematic Keyword Cluster.....  | 84 |
| Table 30: “uzaktan eğitim veren üniversiteler” Problematic Keyword Cluster.....   | 85 |
| Table 31: “sağlık bilişim sistemleri” Problematic Keyword Cluster.....  | 87 |
| Table 32: “health information systems phd programs” Problematic Keyword Cluster..   | 88 |
| Table 33: “software management odtu” Problematic Keyword Cluster.....   | 89 |
| Table 34: “yaşam temelli öğrenme” Problematic Keyword Cluster .....   | 91 |
| Table 35: “middlesex üniversitesi” Problematic Keyword Cluster Report .....   | 93 |
| Table 36: “institute for work based learning” Problematic Keyword Cluster .....   | 94 |

Table 37: “information systems and programming metu” Problematic Keyword Cluster ..... 95

Table 38: “online class metu” Problematic Keyword Cluster..... 96

Table 39: Think-Aloud Study Questions Category and Source ..... 101

Table 40: Branding Category Questions Results..... 117

Table 41: Functionality Category Questions Results ..... 118

Table 42: Content & Ease of Use Category Questions Results..... 118

Table 43: Challenge & Skill Category Questions Results ..... 119

Table 44: Self-Efficacy Category Questions “Yes” or “No” Reply Results..... 120

Table 45: Self-Efficacy Category Confidence Level Reply Results ..... 121

Table 46: Overall Evaluation Category Reply Results ..... 122

Table 47: Qualitative Evaluation of Institute Web-Site..... 122

## LIST OF FIGURES

|  |     |
|--|-----|
| Figure 1: Google Analytics Custom Reporting Screen View .....        | 27  |
| Figure 2: Data Feed Query Explorer Screen View .....                 | 28  |
| Figure 3: 2009 Period Keywords Histogram.....                        | 45  |
| Figure 4: Selected Period Keywords Histogram .....                   | 48  |
| Figure 5: Performance Results of Participants for Each Task.....     | 116 |
| Figure 6: Institute Web-Site View during the Think-Aloud Study ..... | 124 |
| Figure 7: t-test Output of Hypothesis 1 .....                        | 125 |
| Figure 8: t-test Output of Hypothesis 2.....                         | 126 |
| Figure 9: t-test Output of Hypothesis 3.....                         | 127 |

# CHAPTER 1

## INTRODUCTION

Today, web sites are important communication channels that reach a massive audience. Measuring the effectiveness of these web-sites has become a key issue for researchers as well as practitioners. However, there is no consensus on how to define web site effectiveness and which dimensions need to be used for the evaluation of these web sites. This problem is more noteworthy for information driven web sites like academic web sites. Effectiveness of information driven web sites are defined by the success of their information architecture in the literature. And, success of information driven web sites are measured by classical methods like questionnaires and think-aloud study. Web analytic tools are newly emerging tools to evaluate web sites compared to *observations, think-aloud, questionnaires, eyetracking methods*. These classical methods are time consuming and costly compared to newly emerging web analytic tools. The preliminary aim of this study was testing whether newly emerging web analytic tools which save time and which are cost effective compared to classical methods can be used instead of or to support classical methods to improve the information architecture of information driven web sites We took METU Informatics Institute web site as a case study. To reach our aim, we benefit from the results of a newly emerging tool and also a classical tool. Among web analytic tools, we choose to use Google Analytics which is a commercial web analytics tool owned by Google in our study. Web analytics is a predominant application area and so there is limited academic literature in this sector.

The existing studies measured the effectiveness of these academic web sites by taking into account their information architecture mostly using think-aloud methodology. Bearing this in mind, we choose to use think-aloud among the classical methods. However, in the middle of our study we realize that we are interpreting the reports of Google Analytics inaccurately. When we search for literature to interpret the Google Analytics reports, we see a gap in this area. There is limited study to interpret Google analytics reports and so very few properties of this web analytic tool are being used. Since time allocated to this study is limited, we spent most of our time to interpret the reports of Google Analytics. We collected the results of Think-Aloud study but could not compare with the results of Google Analytics reports statistically. Testing whether web analytic tools can be used instead of or to support classical methods to improve the information architecture of information driven web sites which is the next step of this study can be a topic of further studies in this area. This study differs from other studies in the sense that it uses both Google Analytics which is a commercial web analytics tool owned by Google and think-aloud study results to improve the information architecture of a web site. In this study, we give special importance to form an understanding of our case study web site in the minds of the readers. To reach this, in the beginning of our study, we introduce our case study web site using the language of Google Analytics. We give statistics about our case study web site taken from Google Analytics. If we consider web analytic tools which are known predominantly in application area, this study gathers together the limited academic literature in this area with the expertise of people in this application area by collecting expert opinions published in blogs and whitepapers. We discover that valuable features of web analytic tools are not used because they are complex and confusing. To overcome this problem, we explored Google Analytics metrics and dimensions in detail which is another contribution of this study. We believe this will increase the usage of valuable features served by web analytic tools. Landing page optimization technique via Google Analytics as a tool for improving the information architecture of web sites is recommended in many books but we have not encountered any created methodology on this issue. In this study, we form a

methodology to detect the most problematic keywords which land on our pre-specified web pages and analyze these keywords one by one. We create a new questionnaire for our think-aloud study by tailoring the questionnaires in the literature to our needs. This questionnaire can be used in future studies related to the analysis of content based web sites. Think-aloud studies in the literature were carried out with a maximum 24 participants and the statistical reason behind choosing this sample size was not explained. In our study, we use 32 participants to reach statistically significant results and explain why we choose this sample size in Section 4.5.

The rest of the thesis includes four chapters. In the second chapter, we explain the literature on website evaluation, information architecture, web analytic tools and specifically Google Analytics. In the third chapter, we explore the methodology we applied while using Google Analytics. The fourth chapter explains the methodology we used during think-aloud study. In the last chapter, we present the conclusion.

## **CHAPTER 2**

### **LITERATURE REVIEW**

Our aim in this thesis study is to improve the success of our case study web site so in this chapter we first search the literature on how to evaluate success of web sites. We explored the literature on information architecture. Lastly, we explore the qualitative and quantitative tools used in the literature like Web Analytic Tools, particularly Google Analytics and think-aloud study to improve the information architecture of web sites.

#### ***2.1 Website Evaluation***

Pressure on companies to document their website's value has been increasing at present (Patton, 2002). Therefore measuring the effectiveness of a web-site has become a key issue for researchers as well as practitioners (Hong, 2007). A web site value can be measured based on usability and other key design criteria such as navigation, response time, credibility and content according to Nielsen's study (as cited in Atkinson, 2007) and Hong (2007).

Usability evaluation methods can be broadly categorized into qualitative methods which include *observations, think-aloud, questionnaires and eyetracking* and quantitative

methods which include *questionnaires, web log data and web analytics tools* (Atkinson, 2007).

Eye tracking is used to determine what percentage of participants in a usability test fixated on a specific element or region of interest (Tullis & Albert, 2008). A combination of an infrared video camera and infrared light sources to track where the participant is looking is required. To apply eye tracking method, “areas of interest”, “look-zones” and a “minimum total fixation time” need to be defined. For the same element, the effectiveness of different locations on a web page can be compared by the use of eye tracking.

Think aloud usability testing with real participants is one of the *most fundamental evaluation methods* according to Nielsen’s study (as cited in Atkinson, 2007). While participants are using a system or prototype to complete a predetermined set of tasks, they are being observed and they are kindly requested to verbalise their thoughts and comments during performing the tasks. The performances of the participants are also recorded by a video to analyse later the comments and observations of the participants. These observations and comments help to define problems in the system which may negatively affect the user experience according to Addwise’s study (as cited in Atkinson, 2007).

To monitor and improve the quality of web-sites, heuristic expert evaluation and think-aloud usability testing are the most current laboratory approaches (Elling, Lentz & de Jong, 2007). The results of these approaches may be used to revise a website or certain web pages. However, to focus more predominantly on the overall quality of websites, online questionnaires are better tools. Although online questionnaires are suggested, there is no agreement about the question what web-site quality exactly is and which dimensions or items a questionnaire should contain (Elling, Lentz & de Jong, 2007). In the case of information driven web sites, it seems logical to connect web site quality to usability. In 2000, Webby Awards which evaluated web sites under 27 category and

their judges rated web sites based on six criteria: *content, structure & navigation, visual design, functionality, interactivity, and overall experience* (Ivory, Sinha & Hearst, 2001). Factors of web site quality are *content, interactivity and navigation* according to an another study (Bauer & Scharl, 2000). Site evaluation metrics are broadly divided into three categories which are *functional/navigational issues, content and style* and lastly *contact information* for another study (Misic & Johnson, 1999). Whereas, Schubert & Dettling (2001) make use of *usefulness, ease of use, trust factors* for evaluating web site quality. To evaluate success of e-commerce sites, Olson & Boyer (2003) use *attitude, perceived usefulness, perceived ease of use* and lastly *comfort level* factors. Mateos, Mera, Gonzales, & Lopez (2001) developed Web Assessment Index (WAI) which has the following dimensions: *site content, speed, accessibility and navigability*. Waite and Harrison's study (as cited in Hong, 2007) add technical aspects as factors such as efficiency, speed, and reliability which are collectively termed performance. Negash, Ryan & Igbaria, (2003) define web site quality in terms of user satisfaction and they use 3 distinct categories to evaluate web site quality which were *information quality, system quality and service quality*. Liu & Arnett (2000) conclude in their study that web site success in the context of E-commerce sites is related to four major factors which are *quality of information and service, system use, playfulness, and system design quality*. Signore (2005) aims at defining a quality model for web sites. In this quality model, he uses *correctness, presentation, content, navigation and interaction* as dimensions of his model. Zviran, Glezer & Avni (2006) define web site quality as *user satisfaction*. They empirically search the effect of user-based design and Web site usability on user satisfaction. They investigate the relationship among four dimensions on commercial web sites: *user-satisfaction, usability, user-based design, and Web site type*. As Zviran et al. (2006) give importance to *Web site type* during the evaluation of Web site quality, Stolz, Viermetz, Skubacz & Neuneier (2005) distinguish performance indicators of information driven web sites and transaction based e-commerce sites. Stolz et al. (2005) argues that results of the metrics could be used for only statistical analysis if the users intention was not considered.

“The main difference between information driven web sites and transaction based e-commerce sites are given as the knowledge of users intention. Although intention of transaction based e-commerce sites’ users could be known by the user feedback like a purchase, it is not possible for information driven web sites. Users of transaction based e-commerce sites do not stay anonymous since they provide personal billing data, making the user identifiable when returning to the site whereas a user on an information driven web site stays anonymous and it remains uncertain whether the user was interested in the content s/he visited (Stolz et al., 2005)”.

This situation requires differentiated indicators for these two types of web pages to measure the success.

In the academic literature we have found 5 different questionnaire composed of different dimensions which aim at evaluating the quality of web-sites. We have also found a commercial study in this area from a well-known expert working in the market (Rubinoff, 2004).

Kirakowski, Claridge & Whitehand (1998) has formed Website Analysis Measurement Inventory (WAMMI) which is a questionnaire consisting of 60 Likert scale questions. WAMMI is later on reduced to 20 questions for simplicity (Elling, Lentz & de Jong, 2007). These 20 questions are listed under 5 dimensions which were the result of an analysis of the feedback that was produced by a large group of website designers and users. Website quality is defined as *website usability* and the five dimensions are explained as the degree to which users:

- feel efficient
- like the system
- find the system helpful
- feel in control of the interactions
- can learn to use the system

The second questionnaire is prepared by Skadberg & Kimmel (2004). In this study, Website quality is defined as *optimal experience*. The determinants of the optimal experience in the hypermedia Web environment for a web-site are defined as:

- contents
- design
- performance
- visitor's individual differences

The third evolution questionnaire is formed by Muylle, Moenaert & Despontin (2004). In this study, Website quality is defined as *user satisfaction*. The name of this questionnaire is Website User Satisfaction (WUS) questionnaire. This questionnaire is formed of 60 questions which are categorized under 4 dimensions and 11 sub dimensions. These dimensions and subdimensions are:

- connection
  - ease of use
  - entry guidance
  - structure
  - hyperlink connotation
  - speed
- quality of information
  - relevance
  - accuracy
  - comprehensibility
  - comprehensiveness
- layout
- language

The fourth evaluation questionnaire we have found in this area is prepared by van Schaik & Ling (2005). It is composed of 30 questions which is filled after the respondents visited a university web-site and performed three information retrieval tasks. The dimensions used in this evaluation questionnaire are:

- Perceived ease of use
- Disorientation
- Flow
  - Involvement
  - Control
- Perceived Usefulness
- Aesthetic Quality

The fifth evaluation questionnaire is formed by Elling et al. (2007). Their aim is developing and validating a generic Website Evaluation Questionnaire (WEQ) which may be used to evaluate municipal and other governmental websites. To decide on the dimensions of website quality for the WEQ, they benefit from the studies of Kirakowski et al. (1998), van Schaik & Ling (2005) and Muylle et al. (2004). WEQ uses 3 dimensions and 9 sub dimensions which are:

- Content
  - Relevance
  - Comprehensibility
  - Comprehensiveness
- Navigation
  - Ease of use
  - Structure
  - Hyperlinks

- Speed
- Search Engine
- Layout

In the application area, “many looked user experience as an overall indicator of web site success” (Rubinoff, 2004). However, since how effectively a Website provides net positive experience often turns into a subjective affair, Rubinoff (2004) has formed a “quick and dirty methodology” in his words for quantifying the user experience. This methodology helps him in providing his clients with a quick, objective, visual representation of where their site stands compared to their competitors and the site’s past performance if an improvement has been realized. Different from the questionnaires provided by academic literature, Rubinoff (2004) has formed his methodology for the use of designers, developers, clients to share a common understanding of the site in question. For him, the user experience is made up of four independent dimensions and he bases his methodology on the following dimensions:

- branding
- usability
- functionality
- content

Under each dimension, he creates a series of statements or parameters against which the Website in question will be measured. To give an example, under “branding” dimension he uses “The site provides visitors with an engaging and memorable experience” statement which will be answered by the user of his methodology. A scale of 1 to X is created for each of the statements. X is changing according to the number of statements under each measure. By this, he means that for example under “branding” dimension, one can create many statements but total of scales for these statements need to be equal to 100. This means that if one creates 5 statements under a dimension, the scale is 1 to

20 and if 10 statements are created the scale will be 1 to 10. In the two conditions, total scores will be equal to 100. After assessing the site according to each statement, he gives each statement a score within the specified scale. As a second task, he calculates total scores for each dimension. Rubinoff (2004) prefers spider chart to display the results. In the spider chart, each axes are represented by a dimension and overall scores for each dimension are pointed and then plotted. By following the same steps for the competitors and the previous version of the same websites, the spider chart shows a clear picture of the concerning website in terms of each measure and compared to its competitors or its previous version. This methodology created by Rubinoff (2004) can be extended as to include new dimensions with each having a different weight if preferred.

Semi Automated Web Site Evaluation Tool (SAWSET) is developed by Genç (2006) to help save time and money in design. The approach used to develop this tool consists of two main topics which are Structural Evaluation and Content Evaluation. Structural evaluation covers similar dimensions used in the literature which we explained in the above paragraphs. Dimensions used in Structural Evaluation part of this tool are identity, loading and viewing, navigation, interactivity, comprehensibility, personalization and content, information quality and, up to datedness and security. Totally 98 questions are used in Structural Evaluation and each question and dimension are given a weight as Rubinoff (2004) applied. 21 out of 98 questions are automatically replied by the system designed which uses spiders to crawl the web site and seeks answer for selected questions depending on predefined rules. Content evaluation is carried out to compare the content of the web sites in the same scope of business using a simple same content clustering method.

## 2.2 *Information Architecture*

In the emergence of this discipline, there are two milestones which are publication of a book by Rosenfeld and Morville in 1998 and the organization of a preliminary summit by the American Society for Information Science and Technology (ASIS&T) in May 2000 (Dillon & Turnbull, 2000). Theme of this submit is defining *information architecture*. Rosenfeld & Morville (1998) have four different definitions of information architecture. These definitions are:

- “1. The combination of organization, labeling, and navigation schemes within information system.
2. The structural design of an information space to facilitate task completion and intuitive access to content.
3. The art and science of structuring and classifying web sites and intranets to help people find and manage information.
4. An emerging discipline and community of practice focusing on bringing principles of design and architecture to the digital landscape.”

The reasoning they provided for preparing four different definions is the relationship between words and meaning’s being tricky. They believe that no document fully and accurately represents what their author intended to mean and so it is so hard to design good web-sites.

Information architecture is “simply a set of aids that match user needs with information resources” according to Davenport’ study (cited in Gullikson et al., 1999). Wurman (1996) cited again in Gullikson et al. (1999) defines as “structure or map of information which allows others to find their personal paths to knowledge”. Whereas Gullikson et al. (1999) explains information architecture as “How information is categorised, labelled and presented and how navigation and access are facilitated”. Gullikson et al. (1999)

also claims that information architecture determines not only whether users will and can find what they need, but also affects user satisfaction and influences return visits. He bases his claim on the studies of Nielsen (1999) and Koman (1998). Nielsen (1999) cited at Gullikson et al. (1999) concludes in his study after many usability tests that people do not come to the web for an “experience”- they come for information. Similarly, Koman (1998) cited at Gullikson et al. (1999) has reported that “roughly two thirds of users are looking for specific information”. Information architecture is the term used to describe the process of designing, implementing and evaluating information spaces that are humanly and socially acceptable to their intended stakeholders (Dillon, 2002).

Lash (2002) defines information architecture as covering *usability* but more than that. He does not agree with defining information architecture as equal to User-Centered Design. He agrees with the definition of Lou Rosenfeld and Peter Morville in the second edition of their book, information architecture for the World Wide Web. They described information architecture as composed of three circles which are *Users, Context and Content*. They mean that information architecture needs to take into account the information itself (content), the people using the information (users), and the business issues (context) in which the information is being presented (Lash, 2002). To take into account users, content and context; Information Architects need to have knowledge about these three. “Architects want to know the context for making design decisions, validation of heuristic assumptions and understand why visitors come to the site and what they seek according to interviewees (Wiggins, 2007)”.

Web Analytics data help Information Architects via answering their questions stated above. Web Analytics can be used to improve information architecture (Wiggins, 2007). Hallie Wilfert is a senior Information Architect at SRA International, a technology and strategic consulting firm in Arlington, VA. She works mainly with government clients and she claims that an information architect should care about web analytics because it allows to broaden the scope of the architect’s research (Wilfert, 2008). She discusses

that to improve a particular section of site, the architect will ideally need some interviews with the target audience, build a prototype and test it with users which will cost a tremendous research budget. “Whereas, web analytics research is on the architect’s entire site and with the entire population and the cost is negligible compared to classical research budgets. With web analytics, architects can identify problems or issues that they might have not known or not cared about otherwise (Wilfert, 2008).”

Another way of improving information architecture is using Questionnaires and *think-aloud* studies. Common methodology is giving tasks to the participants and also using perception tests. Gullikson et al. (1999) apply a different version of think-aloud to assess the impact of information architecture on academic web site usability taking Dalhousie University’ web site ([www.dal.ca](http://www.dal.ca)) as a case study. He gives a task which is composed of finding the answers of 6 questions and then he applies a perceptions test containing four Likert-scaled and three open ended questions related with the participants’ ability to use the web-site. As a final task, participants verbally explain the approaches they took in responding the questions during watching their videos which were recorded when they were doing the task. The only difference from think-aloud study is the final task, in think-aloud study participants think loudly during finding the replies of the questions. At the end of their study, they find out that potential user groups and needs of them are so diverse. Consequently, they recommend the site designers to provide the visitors with multiple pathways and multiple ways of accessing those pathways. The designers can add search engine, sitemap, alphabetical index, multiple categorical menu structures, FAQ (Frequently Asked Questions) part and navigational aids to the web site. Navigational aids can be in the form of a standart menu on each page which will provide access to the top-level menu from anywhere on the site, to use any of the Access tools from anywhere on the site and to determine easily one’s location within the hierarchy. About information design, categorical menu structures which are user pathways need to contain the following characteristics: scheme, categories, labelling and presentation. Hybrid approach for organizing the categories for this web site needs to be used.

Categories defined within the scheme need to be distinct and mutually exclusive (Gullikson et al. , 1999). Labels need to be written in the language of the user population and they need to be unambiguous and directly related to the concept. Menu structure need to relieve the visitors. They need to be broad and multiple menus can be accommodated simlutenously.

Maloney &Bracke (2004) propose a framework that combines elements of information architecture with approaches to incremental system design and implementation considering also the needs of libraries which have legacy systems that were not designed for the Web environment. Information architecture definition is extended and additional constraints that characterize library web sites are incorporated to the definition and it is named as Extended Information Architecture (EIA) (Maloney & Bracke, 2004). In their framework, components of information architecture and system architecture are determined and their relations are described. EIA which describes the user needs is composed of coordinating structure and service elements. Coordinating structure is composed of organization, navigational structure and labelling whereas service elements are composed of functional specifications and content specifications. Coordinating structure informs service elements but service elements constrains coordinating structure. Similarly, system architecture constrains information architecture but information architecture informs system architecture.

There are initiatives in this area. Information Architecture (IA) Institute is established in 2005. Formerly, it was Asimolar Institute for information architecture (AIfIA) which was formed in 2003 by a self-identified, dedicated IA professionals to advance and promote the field (Dillon & Turnbull, 2000). Another initiative is American Society for Information Science and Technology (ASIS&T) which was established in 1937 as the American Documentation Institute (ADI). In 2000, its title change has been realized from ADI to ASIS&T to reflect the range of its members. It has members in over 50 countries worldwide and there are 20 Special Interest Groups in a variety of fields. There exists an information architecture *Special Interest Group* under ASIS&T. It provides “a

forum for practitioners, researchers, and educators working in the multidisciplinary areas of information architecture where they can continue the conversation since March 2000 (ASIS&T, 2010)”. From May 2000 to today, it hosts information architecture summits annually.

### ***2.3 Landing Page Optimization***

Ash (2008) alleges that well-optimized landing pages can change the economics of your business overnight and turbocharge your online marketing programs. Researchers in Canada conduct three studies to ascertain how quickly people form an opinion about web page visual appeal and they find out that visual appeal can be assessed within 50 milliseconds (Lindgaard, Fernandes, Dudek, & Brown, 2006). Gofman (2007) concludes from the results of this recent study that what makes a difference is mostly the main features and the general appearance of the landing page, not necessarily the actual content.

In his book Ash (2008), to detect and uncover the problems about landing pages, proposes the following methods which are audience role modelling, web analytics, onsite search, usability testing, usability reviews, focus groups, eye-tracking studies, customer service representatives, surveys, forums and blogs. Among the web analytic features, the following ones can be used to discover common problems about landing pages: visitors, map, languages, technical capabilities, visible browser window, new vs. returning visitors, depth of interaction, traffic sources, and content.

Landing page optimization is a subset of Search Engine Optimization (non-paid search, also known as organic search) (Clifton, 2008a). Tim Ash, CEO of [SiteTuners.com](http://SiteTuners.com), a landing page optimization firm that offers conversion consulting, full-service guaranteed-improvement tests, and software tools to improve conversion rates, states

that by using landing page optimization and testing one can often produce double-digit conversion rate improvements and change the economic position of an on-line business dramatically (Ash, 2008). If you are targeting relevant core keywords, *Bounce Rates* for Search Engine Optimization (SEO) ideally will not be more than 20% to 25%, more than 35% *Bounce Rate* for SEO is an alarming signal to examine usability factors on the site (Kumar, 2009b).

## **2.4 *Web Analytic Tools and Google Analytics***

In this part of our study, we describe web analytic tools, their visitor collection methods and specifically explain Google Analytics. How Google Analytics functions and its limitations are also explained in this chapter. As Wiggins (2007) concluded in her presentation in information architecture Summit 2007, there are “*not much out there in the academic literature on using web analytics (hopefully to change!)*”. We also faced with this problem during literature search so we mostly benefited from expert opinions, blogs, whitepapers on this topic. While benefiting from blogs, we do not forget that blogs are not as reliable information sources as books and journal articles and they are valid only when they were written due to fast changing feature of this web analytics sector. To increase the reliability of information from blogs, we give particular attention to the blogs of known experts in this area. While citing information from these blogs, we also give a little bit more information about the writer of these blogs. To overcome the validity issue of the disseminated information from blogs, we test them using the current version of Google Analytics.

### 2.4.1 Web Analytic Tools

There are five main methods which can be used to collect visitor data: Page Tags (client-side data collection), Logfiles (server-side data collection), Hybrid solutions, Network Data Collection devices or “packet sniffers”, using a Web-server API Loadable Module-programs (Clifton, 2008b; Clifton, 2010). In recent years, the popularity of Page Tags has increased due to its allowing the analysis to be outsourced, “Hosted” solution.

**Page tagging** method uses data collected by user’s browser. The technique is known as client-side data collection. Information is usually captured by Javascript code placed on each page of the web site. This javascript code is known as beacons or tags. This technique is used by outsourced, hosted vendor solutions (Clifton, 2008b; Clifton, 2010). Free Page Tag Vendors: (because hard disk space and bandwidth are so cheap) are Google Analytics, Microsoft Adcenter Analytics and Yahoo Index Tools. Advantages of this method are explained by Clifton (2008b) as: provides more accurate session tracking (caching and proxies are not problem), collects and processes visitor data nearly in real time, data storage and archiving can be outsourced, program updates can be done by the vendor. Disadvantages of this method are explained by Clifton (2008b) as: tracking code set up errors lead to data loss and so going back and re-analyzing are impossible, firewalls can restrict activation of tags, completed downloads and bandwidth cannot be tracked because only request of page or file can be tracked, robots-spiders cannot be tracked, and lastly the terminal of the visitor needs to be capable of understanding Javascript and/or cookies (mobile users cannot be tracked because most of them are not capable of understanding Javascript and/or cookies). StatCounter (StatCounter, 2010), Sitemeter (Site Meter, 2009), Nedstat and Omniture (Clifton, 2008c; Clifton, 2010) utilize “JavaScript Page Tagging” technology.

**Logfile method** uses data collected by your web-server. It is known as server-side collection. This technique captures all requests made to your web-server including pages, images and PDF's (Clifton, 2008b; Clifton, 2010). It is mostly used by “stand-alone” software vendors. Advantages of this method are explained by Clifton (2008b) as: historical data can be reprocessed easily, firewall is not a problem for tracking, completed downloads and bandwidth can be tracked and also completed or partial downloads can be reported, search engine spiders and robots can be tracked by default, mobile visitors can be tracked by default. Disadvantages of this method are explained by Clifton (2008b) as: served pages from proxies and/or caches cannot be logged and so reported, event tracking is not available, program updates need to be done by your own team (no outsourcing), robots multiply visits, data storage and archiving cannot be outsourced. Deep Metrics -LiveStats (George, 2007) , Webalizer (Barrett, 2009), W3Perl (W3Perl, 2010), AWStats utilize “Log File” technology.

**Hybrid method** combines Page Tagging and Logfile methods. As it can be understood from the above explanations, a disadvantage of page-tagging method is an advantage of logfile method. Combining the two methods result in a more reliable visitor tracking. Clickstream, Coremetrics, Webtrends, NetTracker, Visual Sciences –HBX Analytics (George & Heimann, 2007) uses Hybrid method. Google Analytics uses page tagging visitor data collection method but it can be configured as a hybrid data collector (Clifton, 2008c).

**Network data collection devices (packet sniffers)** gather web traffic information from routers to “black box”appliances. The disadvantage of this method is that it is expensive and complicated. Additionally, it is not offered by many vendors (Clifton, 2008b; Clifton, 2010).

The last technique we will explain here is using **Web Server Application Programming Interface (API)** or **Loadable Module**. These programs are used for extending the capabilities of web servers. The logged fields are enhanced and extended.

The captured data are streamed to a reporting server in real time (Clifton, 2008b; ; Clifton, 2010). Woopra (iFusion Labs LLC, 2009), W3Counter (W3Perl, 2010) utilize API technology.

Fang (2007) claimed that “web analytics offer objective and multi-faceted statistical data in a visual way for webmasters to better understand the interaction between their visitors and their websites”. Additionally, Arendt & Wagner (2010) explained that compared to other site evaluation tools, usage statistics reported by web analytic tools have some advantages because they report not what the users tell what they would do instead they monitor how users actually work with a site. Analyzing usage statistics provided by web analytic tools did not resolve all controversies about web site redesign but it supported certain design decisions more than other methods according to Arendt & Wagner (2010).

Atkinson (2007) uses both web analytics and think-aloud evaluation methods to examine the quality of an e-commerce site. She defines web site quality as *users' experience*. She uses William Hill website, an online sports betting web site, as a case study. She makes use of both live web analytics data collection method and controlled data collection by web analytics and finally she runs and analyzes think aloud evaluation sessions. She uses a customized web analytic software by RedEye International. Web analytics data are collected by RedEye tagging technology. At the end of her study, she recommends on how to improve the web-site and how each method can be used to overcome the limitations of each other.

#### **2.4.2 What is Google Analytics?**

It is one of the analytics software packages available in the market like WebTrends, Omniture, HBX/WebSideStory, CoreMetrics, ClickTracks, DeepMetrix and Google

Analytics is based on Urchin v6 (ActualMetrics.com, 2009). It is using Page tagging visitor data collection method among the methods we explained above. It is a hosted solution. All the data will be kept and reported from the Google servers. It is free of charge up to 5 million pageviews per month per account according to Google Analytics Terms of Service Article 2 (Google, 2010e). Although Jasra (2006) had heard that Google's numbers (page views/visitors) were slightly lower compared to other analytics vendors and he noticed this almost consistently when he compared his Google numbers to that of his StatCounter's, he also commented that he personally had not noticed a big discrepancy in his site's data. Dyrli (2006) cited at Fang (2007) claimed that by far the most sophisticated web analytics tool was Google Analytics.

### **2.4.3 How does Google Analytics function?**

As we mentioned above, Google Analytics uses Page tagging visitor data collection method. There is a Google Analytics tracking code which needs to be inserted to the pages to be tracked. The page is served and the Google Analytics Tracking Code JavaScript is executed. The Google Analytics Tracking Code calls the `trackPageView()` method. At this point, the Google Analytics first-party cookies are read and/or written. The webpage then sends an invisible gif request containing all the data to the Google Analytics reporting server, where the data is captured and processed. Data is processed regularly throughout the day and you can see the results in your reports. Google collects and reprocesses all data for a 24-hour period at the day's end because there can be interruptions during logfile transfers which will result in partial logfile processing to create reports (Clifton, 2010).

Google Analytics is very easy to use; you need to first create an account and agree the terms of conditions. By agreeing on the Terms of Conditions, you agree on sharing your

web-sites reports with Google. One thing you need to consider which is not written anywhere but learned by application is Google uses *Bounce Rate* values of your each page to improve its search engine, if your page has high *Bounce Rate* value, your ranking position will be dropped after you enable Google Analytics (Kumar, 2009a). After agreeing on the Terms of Conditions, you need to insert the Google Analytics tracking code to each and every page of your web-site. This step is important because for web-sites including many pages, there can be forgotten web-pages which won't be tracked and during update of the webpages tracking code can be erased unintentionally. Second issue is where to insert the tracking code. Google recommends to insert it at the end of the page because as we explained above Google Analytics Tracking Code is a JavaScript and usually JavaScript errors occur when an element of a web page's script contains an error or fails to execute correctly, by putting the code at the end, Google Analytics will not be tracking and so reporting about an error page will give more reliable results.

#### **2.4.4 What Google Analytics can and cannot do?**

Who cannot be tracked by Google Analytics?

- Google Analytics uses first-party cookies and so someone who blocks all cookies cannot be tracked by Google Analytics because all the data is passed to the Google Analytics servers via the first-party cookies.
- Since Google Analytics Tracking Code is a Java Script code and needs Javascript in the visitor's PC to be activated, a visitor who disables JavaScript cannot be tracked.

- Since cached pages are saved on a visitor's local machine, Google Analytics will not track visits to the cached pages if the visitor is not connected to the internet.
- JavaScript errors occur when an element of a web page's script contains an error or fails to execute correctly. If an error occurs before the Google Analytics Tracking Code is executed, the visit to the page won't be tracked.

The users who cannot be tracked because of the above listed reasons result in the lower visitor number compared to Google Analytics reports. This situation is also supported by what Manos Jasra, Web Analytics Analyst / Software Developer, stated: "It is impossible to get 100% data integrity with any solution and I have heard complaints that Google's traffic data is at times much lower than other vendors but I personally haven't noticed a big discrepancy in my site's data (Jasra, 2006)".

Who and what will be reported incorrectly by Google Analytics?

- Someone who deletes their cookies will still be tracked, but they'll be identified as a new visitor to the site which will result in higher new visitor number and lower returning visitor number.
- Daily and hourly reports of Google Analytics have higher absolute unique visitors than visits which is also a bug discussed in Google Analytics support forums (Kumar, 2009c). It is recommended by the Google Analytics Online Marketing Manager & Web Analyst Lalit Umar that to be on the safer side 2 days time lag difference would give proper results and added that Google Analytics team is still working on this bug.
- JavaScript errors occur when an element of a web page's script throws an error or fails to execute correctly. If an error occurs before the Google Analytics Tracking Code is executed, the visit to that page won't be tracked and since

timeOnPage is calculated by subtracting the time of the new accessed page from the previously accessed page, this metric will also be calculated incorrectly.

- A page refresh is counted as a new page by Google Analytics which results in inflating the pageviews (NMmarketing, 2009). Consequently, NextPagePath is reported by Google Analytics as the same page with visited PagePath and PreviousPagePath which will be misleading information. If the refreshed page is the landing page and also the exit page, this means that the *Bounce Rates* will be higher than the reality.
- Owning or sharing multiple computers will lead inaccurate reports in Google Analytics. If users share the same computer under the same user account, Google Analytics will report this as 1 unique visit. If someone accesses to your site from different computers for example from home, from office, from internet cafe et al., although this is only one unique visitor, it will be reported by Google Analytics as more than 1 unique visitor which is not reflecting the reality. Since Google Analytics reports are based on anonymous users, this problem cannot be solved.
- There is also an issue on visit count which needs to be discussed. According to (Dainow, 2008). Google Analytics is calculating visits in an inaccurate way. He, an independent web analytics and marketing consultant working in the U.K. and Ireland, does not agree with the calculation of Google Analytics' bounces and visit metrics as each bounce is also counted as a visit. He explains that as bounces are like people looking at the window of your shop; you cannot count them as visitors until they enter the shop. He blames Google Analytics increasing the visit number deliberately which results in erroneously reported average duration of a visit. He backs his belief by explaining that in July 2007, Google changed the calculation of average duration so that it did not include bounces anymore as he suggested but a month later Google put it back to the old way of

calculation. The reason behind this putting back was in the words of Brett Crosby, senior manager at Google Analytics who explained in Google blog that it was because people complained the change meant the new (accurate) numbers were out of line with the old (inaccurate) ones. (That blog has since been removed by Google, but you can find it copied on many sites.) (Dainow, 2008).

To overcome the accuracy problem of Google Analytics reports, we made use of YMMV web stats. YMMV webstats is a tool to be used in conjunction with Google Analytics (Google, 2010). It creates a method to determine the accuracy of the data being retrieved from Google Analytics (Google, 2010). YMMV webstats is focused on tracking three areas: Users who block ads; users who block Google Analytics; users who use no script. YMMV also offers a control log. As it can be easily understood, YMMV uses logfile method for visitor data collection.

We set up the program on our web-site and as of 25.06.2010 the accuracy rate of Google Analytics reports for our web-site is reported as 98.25% by YMMV.

As can be seen above, Google Analytics results are not 100% accurate but as Clifton (2010) indicated focusing on measuring trends instead of precise numbers will help you answer many questions to improve your site. Novo (2006) also agrees with Clifton (2010), he uses this statement “even if the data is not 100% accurate, in some way, as long as you continue to use data collected in the same way each time, you can still build trend charts.” These ideas also support Google in deciding the way opposite of Dainow (2008) for whom, it is putting consistency more important than accuracy.

## 2.4.5 What are dimension and metric and how to get reports in Google Analytics?

Once the data collected by Google Analytics is processed for the reports, it appears in two primary formats: *metrics* and *dimensions*. **A metric is a numeric summary** of user behavior to your website (Google, 2010b). It is calculated in two basic ways: as overview totals and in association with one or more dimensions. For example, *pageviews* is a metric that summarizes the total pageviews for a particular page. *Bounce Rate* summarizes the percentage of single-page visits to your site. *Visits* summarize the number of sessions on your site. When metrics are viewed without a dimension, they provide site-wide or *aggregate* values (Google, 2010b).

**A dimension is a data key or field** typically in the form of a string. Dimensions by themselves are not generally meaningful, but when paired with metrics, they can divide or *segment* the metric from the perspective of that dimension (Google, 2010b). Chosen dimensions will determine the granularity of the report.

All the dimensions and metrics cannot be used in the same report because they won't be valid combinations. If you ask for dimensions or metrics from two different groups, your options are limited to select from the same metrics and dimensions in the intersection of the two groups. For example, if you select a visit dimension and a visit metric, then your options to build your custom report are now reduced to visit metrics and dimensions, internal search dimensions and campaign dimensions (Google, 2010c).

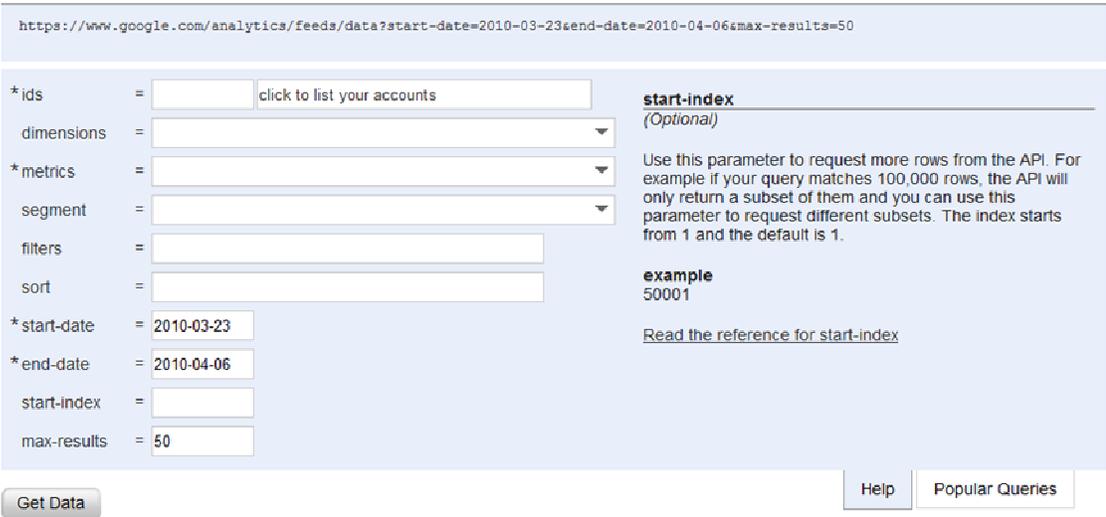
Google Analytics has standard format reports and it also allows users *custom reporting*. The screen view of this custom reporting is given in Figure 1. A user can choose his/her own dimensions and metrics from the menu at the left side of the window. Users are limited to select a maximum of five dimensions - one for top-level table segment, and up to four dimensions to drill down to. The dimensions chosen will stay the same for all the metrics chosen which are located in different tabs. The screen is capable of including a

maximum of 10 metrics and to overcome this limitation, tabs are used. Each tab has 10 metrics capacity. Users can add new tabs by clicking on “Add Tab” link which is placed in the middle top of the screen. This function helps adding more metrics to the custom created report. Maximum of 5 tabs are allowed which means that a custom report having up to 50 metrics (5x10) can be created. Drag and drop can be used to choose dimensions and metrics in this screen. Double clicking on each metric and dimension categories results in opening the metrics and dimensions available under each category. Valid combinations of dimensions and metrics can be seen automatically by the user, the interface is designed accordingly, and invalid combinations cannot be selected. These reports have also filtering and reporting period choice options after the custom report is prepared. Although custom reporting provides a convenient way of showing results, it has certain limitations. For example, a maximum of 500 rows and 2 dimensions can be shown on each report and data export is a real restriction for the users.



Figure 1: Google Analytics Custom Reporting Screen View

However, another service by Google Analytics, *Data Feed Query Explorer* which can be accessed at <http://code.google.com/apis/analytics/docs/gdata/gdataExplorer.html>, allows users more flexibility and it is better for users who need to export the Google Analytics data and analyze it in other tools. This is a user friendly designed page and it has a good interface as can be seen in Figure 2. Users are allowed to choose a maximum of 7 dimensions and 10 metrics. Up to 10,000 rows can be reported per each query. Filtering, segmentation, sorting, starting index assigning, reporting period selection and maximum result row number assigning options are available. One deficiency of this reporting mechanism is that valid combinations of dimensions and metrics need to be known in advance by the user. The reporting mechanism only gives error when the combination is not valid and it does not inform the user about the problematic metric or dimension.



**Figure 2: Data Feed Query Explorer Screen View**

Users can benefit from the valid dimension-metric combination list available at this link: <http://code.google.com/apis/analytics/docs/gdata/gdataReferenceValidCombos.html>.

Another deficiency of this service is the maximum allowable filter expression; the limit is 128 characters (Google, 2010a).

There exists a more flexible way of taking Google Analytics reports other than Data Feed Query Explorer. This method is creating applications in Java by downloading the source code and examples created by the Google Analytics API team (Google, 2010g). The Analytics Data Export API returns a maximum of 10,000 entries per request like Data Feed Explorer. If your returned report includes 10001 rows, this means that the full report can include more rows, you need to change the `query.setStartIndex ( )` value to the multiples of 10000 plus 1, to retrieve the left rows of your report.

#### **2.4.6 How Google Analytics is used and what results are obtained?**

In the academic literature, we have found two studies which have used Google Analytics to improve information driven web sites taken as case studies.

Fang (2007) used Google Analytics to improve the content and design of the Rutgers-Newark Law Library (RNLL) main website. Rutgers-Newark Law Library is part of Rutgers School of Law-Newark. They tracked the web site for nearly one and a half months. They have mainly monitored Site Overlay, Content by Titles, Funnel Navigation, Visitor Segmentation, Visualized Summaries, information on visitors' connection speed and computer configuration. Digital Services Librarian received and interpreted the Google Analytics reports and together with librarians and administrators, they decided on the actions for improvement. According to the Google Analytics reports, they firstly synchronized what visitors possess as connection speed and browser capabilities with what the web-site requires. To give an example, they decided to keep their 800x600 webpage template based on the screen resolution information of their

visitors. Related with the design, from the Google Analytics reports, they discovered that most popular items were reached by visitors using search engine because some of the popular content was deeply buried. They also discovered that the right hand menu on their main website which took up about 20% of the web page layout and provides clickable news headlines from JURIST: “web-based legal news and real-time legal research service” hosted by the University of Pittsburgh School of Law generated very few clicks, that is underused. To retain first-time visitors, make popular content easily accessible and generate more clicks on the right hand menu, they added a *Most Viewed Items* section on the right-hand menu using *Content by Titles* report of Google Analytics. They secondly discovered while using *Site Overlay* feature that *Quick links* section on the left hand menu was hard to differentiate, all links were underlined and there was no change when mouse was over. They additionally discovered that the site where they want the visitors to refer was not used by the visitors according to *Google Analytics Funnel Visualization* reports. They decided to add an *Other Links of Interest* section on the right-hand menu bar which was including firstly their major content, the Internet Law Guide where they want the visitors to refer and two popular external links suggested by reference librarians based on the information collected from Google Analytics. Secondly, to overcome the non differentiating problem, for all the items on the right-and left-hand menu bars, they added a mouse- over effect, increased the font size and bulleted the items. They reduced the spaces of less clicked on the main site links if they were necessary to stay on the main web-site. They also used Google Analytics to test whether or not the redesign worked. They compared pre-modification 22 days reports with post-modification 22 days reports. They used *new visitors, returning visitors, return visits, number of pageviews, page depth, and number of people who viewed more than three pages metrics* to measure the improvement.

Second study we have found is also using a library web site as case study. Arendt & Wagner (2010) explained how Google Analytics was implemented by Morris Library at Southern Illinois University Carbondale on its web site and how they used the reports to

redesign their web site. They used a one year period Google Analytics reports in their case study. As Fang (2007) did, by using basic reports Arendt (2010) also synchronized the *screen resolutions* of visitors and the web site. They found the reason of high *Bounce Rate* by looking at Google Analytics reports that excluded visits where the user immediately bounced from the homepage. Since in that report *the number of visits* dropped drastically, *the percentage of visitors with on-campus IP addresses* dropped in a considerable value; they discovered that many of the bounces may have come from in-library users because on the library computers, library' home page was opened by default and patrons were allowed to use these computers other than browsing library resources. Using *Content by Title* report which eliminates double reporting of a web page if it has more than one URL, they rearranged the links on the main page of the web site. They analyzed *Top Landing Pages*, *Top Exit Pages* report which did not yield any ideas for site improvement in their case. They tried to analyze *Navigation Summary* report but since they did not have the resource to examine *Navigation Summary for each page*, the library staff detected just a few parts of the web site that they thought confusing and deserved analysis. Navigation summary showed that the library staff were right, the detected sites were confusing for the visitors because they were navigating away before reaching the target of each page. They decided to add better cues on these pages in order to guide the visitors. According to the *Keywords* report, they ensured that these keywords were incorporated into the site to optimize search results. They checked for the terminology the visitors used and the web site houses for links.

Hasan, Morris & Proberts (2009) investigated whether Google Analytics metrics could be used to evaluate the overall usability of e-commerce sites and also to identify potential usability problem areas. Their research involved three e-commerce case studies. They compared heuristic evaluation of experts for each page with usability findings indicated by Google Analytics software. They used three months period Google Analytic reports. A total of 5 experts evaluated the sites using the heuristic guidelines according to six major categories: *navigation, internal search, architecture, content and design*. They

identified thirteen key web metrics that could either individually or in combination provide alternative to heuristic evaluation in determining usability issues. Heuristic evaluation results mostly overlapped with web metrics results. They concluded in their research that specific web metrics used by them in this research can provide quick, easy and cheap indications of general potential usability problem areas on e-commerce web-sites.

Since Google Analytics is a high level web-metric tool, it is limitedly used. In this thesis study, we tried to overcome these limits and make use of most of the metrics reported by Google Analytics. All the studies we could find that has been done using Google Analytics are summarized above. Our thesis study adds keyword analysis and more detailed, tested metric, dimension explanations which we believe will be a step to overcome limited use of Google Analytics reports.

#### **2.4.7 Used Dimensions and Metrics in the Literature for Information Driven Web Sites**

Patton (2002) explained that today's Chief Information Officers as a whole believe that "web metrics are no longer one-size-fits-all." In today's world, web metrics you used needs to match your web site's business and audience. According to Hong (2007), research on web site metrics show that metrics vary with web-site categories. For example, Patton (2002) has suggested metrics for three types of web sites which were: Business to consumer (B2C)/ retail sites, content sites and Business to Business (B2B) sites. For B2C sites, she suggested using metrics which will help track the customer's likelihood to buy products like conversion rate and average order value in order to measure success levels. For free content sites, she advised using metrics that show a customer's commitment to a web-site like like page views and unique visitors (those

who visit a site more than once during a specific period of time) in order to satisfy advertisers. For B2B sites, she found site performance, user efficiency and average time spent on system metrics more beneficial. Additionally, she publicized that the most valuable metrics will depend on what you are trying to do with your site. The global trend is firms' tying web site metrics to business objectives (Hong, 2007). By this way, metrics will enable firms to measure the extent to which they approach their predetermined objectives. As Hong (2007) explained, it is logical that metrics can be linked to web site categories because web site of each category shares common business objectives and so their success can be measured by a common set of metrics.

Peterson (2004) from his own experience categorized online business models under 4 groups: e-commerce, advertising, lead generation, customer support. As Stolz et al.(2005) indicated, information driven web sites necessitate use of different metrics than transaction based e-commerce web-sites.

Because of the reasons above, since our case study is on an information driven web site, we analyzed the literature to find the suggested metrics for content sites. In Table 1, you can find the expectation and related suggested metric together with its critique and resource. Suggested metrics which are present in Google Analytics are written in italic format. Percentage Engagement is defined by Clifton (2008a) as visitors who contact you or leave a comment on your website over total visits. Percentage Brand Engagement is the ratio of visitors who know your brand before visiting your site over total visits. It is calculated by adding the visits which uses search terms containing your brand names with number of direct access visits and then dividing this to the sum of search engine directed visits and direct visits.

**Table 1: Used Dimensions and Metrics in the Literature for Content Sites**

| <b>Expectation</b>  | <b>Suggested Metrics</b>   | <b>Critique</b>  | <b>Resource</b>                      |
|---|--|--|--------------------------------------|
| See more visitors access your site from search engines with respect to time | ▪ <i>Percentage of visits from search engines</i>  | If any of these metrics are high, it can mean great search engine marketing strategy or other marketing channels are not working   | (Clifton, 2008a) pg: 219, Table 10.1 |
|   | ▪ <i>Percentage of conversions from search engine visitors</i>                           |  |                                      |
| See visitors engaging with your website more                                | ▪ <i>Percentage of visits that leave a blog comment or download a document</i>           | High values in these metrics indicate visitors more engaging in our website in other words, increase in these values indicate that success of our web page is improved.  | (Clifton, 2008a) pg: 219, Table 10.1 |
|   | ▪ <i>Percentage of visits that complete a Contact Us form or click on a mailto: link</i> |  |                                      |
|   | ▪ <i>Average time on site per visit</i>  |  |                                      |
|   | ▪ <i>Average page depth per visit</i>  |  |                                      |
| Improve the customer experience   | ▪ <i>Percentage of visits who Bounces (single-page visits)</i>                           | Higher values in these metrics are the signals of problems in customer experience. These metric values need to be set at low values to improve the customer experience.  | (Clifton, 2008a) pg: 219, Table 10.1 |
|   | ▪ <i>Percentage of internal site searches that produce zero results</i>                  |  |                                      |
|   | ▪ <i>Percentage of visits that result in a help/support page visit</i>                   |  |                                      |
| Measure the success of content sites  | ▪ <i>Visit volume with respect to time</i>   | Increase in these metrics except the last two ones: percentage new versus returning visitors and percentage bounce rate means an increase in the success of the web sites. Increase in percentage new versus returning visitors is favourable if increasing the visitor number is more important than increasing the customer loyalty and also high values in this metric are favourable if the web site is newly established and it needs to increase the visitor number. However, to have a successful web site, percentage bounce rate needs to be low. Increase in this metric is a negative signal. | (Clifton, 2008a) pg: 246             |
|   | ▪ <i>Average time on site</i>  |  |                                      |
|   | ▪ <i>Average pageviews per visit</i>   |  |                                      |
|   | ▪ <i>Average number of advertisements clicked</i>  |  |                                      |
|   | ▪ <i>Percentage engagement</i>   |  |                                      |
|   | ▪ <i>Percentage brand engagement</i>   |  |                                      |
|   | ▪ <i>Percentage new versus returning visitors</i>  |  |                                      |
| ▪ <i>Bounce Rate</i>  |  |  |                                      |

Table 1 (cont.)

| Expectation  | Suggested Metrics   | Critique  | Resource                         |
|--|---|---|----------------------------------|
| Learn your visitor's expectations before arriving on your website. | <ul style="list-style-type: none"> <li>▪ <i>Bounce Rate</i></li> </ul>        | High <i>Bounce Rates</i> here (greater than 50 percent) is an indicator that something may be missing with your web site. You need to look from where the visitors are coming: search engines, referrals or direct. Check the bounce rate for each entrance sources, detect the most problematic ones and analyze the problem. Analyze the keywords used as entrance keywords from the source: search engines. Try to understand what the visitors are expecting from your web site with the used keywords. | (Clifton, 2008a) pg: 280-282     |
|  | <ul style="list-style-type: none"> <li>▪ <i>Entrances keywords</i></li> </ul> |   |                                  |
|  | <ul style="list-style-type: none"> <li>▪ <i>Entrances sources</i></li> </ul>  |   |                                  |
| Use widely used metrics by public institution web-sites            | <ul style="list-style-type: none"> <li>▪ <i>Visits</i></li> </ul>             | Visits metric, most commonly measured as a quick indicator of web site success, is used for gathering simple usage statistics. Increase in this metric is a positive signal for a web page success.   | (Hong, 2007) pg: 283-284 Table 4 |
|  | <ul style="list-style-type: none"> <li>▪ <i>Pageviews</i></li> </ul>          | The "pageviews" metric is particularly useful as a tool to learn about overall web site usage patterns. It allows one to estimate the level of visitors' interest in the web site, and determine whether to launch an advertisement or publicity campaigns on that site. Increase in this metric is a positive signal for a web page success.   |                                  |
|  | <ul style="list-style-type: none"> <li>▪ <i>Best page</i></li> </ul>          | This metric gives "hot content" pages. It shows what the users are mostly interested in so this metric is helpful during new construction or remodelling of a web site.   |                                  |
|  | <ul style="list-style-type: none"> <li>▪ <i>Navigation paths</i></li> </ul>   | This metric shows what the users are looking for and how they reach their goal. If popular web pages which can be accessed only by more than 2 steps are identified, improvements in the information architecture can be done to have these web pages easily accessible by the visitors.  |                                  |

Table 1 (cont.)

| Expectation   | Suggested Metrics  | Critique  | Resource                                       |  |
|---|--|---|--|--|
| <p>Aim your site to be highly rated by Webby awards judges who are internet professionals</p> | <ul style="list-style-type: none"> <li>▪ Word Count</li> </ul>                                   | <p>Decide which of these three groups of web page sizes your web page belongs: <b>low</b> (avg:66.8 wordcount), <b>medium</b> (avg:229.87 wordcount or <b>high</b> (827.15 wordcount).</p>  | <p>(Ivory, Sinha &amp; Hearst, 2001) pg: 4</p> |  |
|   | <ul style="list-style-type: none"> <li>▪ Body Text %</li> </ul>                                  |   |  |  |
|   | <ul style="list-style-type: none"> <li>▪ Emphasized Body Text %</li> </ul>                       |   |  |  |
|   | <ul style="list-style-type: none"> <li>▪ Text Positioning Count</li> </ul>                       | <p>If your web page is a low size one; high word count means more content, smaller page sizes, lower graphics %, graphics count and higher font counts are positive signals.</p>  |  |  |
|   | <ul style="list-style-type: none"> <li>▪ Text Cluster Count</li> </ul>                           | <p>These features result in low download time and high font</p>   |  |  |
|   | <ul style="list-style-type: none"> <li>▪ Link Count</li> </ul>                                   | <p>variations used between header and body text.</p>  |  |  |
|   | <ul style="list-style-type: none"> <li>▪ Page Size</li> </ul>                                    |   |  |  |
|   | <ul style="list-style-type: none"> <li>▪ Graphic %</li> </ul>                                    | <p>If your web page is a medium size one; lower emphasized body text %, higher text clustering count, lower graphic %, higher colour count are positive signals.</p>  |  |  |
|   | <ul style="list-style-type: none"> <li>▪ Graphics Count</li> </ul>                               |   |  |  |
|   | <ul style="list-style-type: none"> <li>▪ Color Count</li> </ul>                                  | <p>If your web page is a high size one; lower body text %, lower emphasized body text %, higher link count, higher page size, lower graphic % and higher colour count are positive signals.</p>   |  |  |
| <p>Predict user characteristics</p>   | <ul style="list-style-type: none"> <li>▪ <i>Revisits of pages in the same session</i></li> </ul> | <p>Spatial ability is constructing and using a visual mental model of the space. 20 % of the variance in spatial ability of the visitors and 15% of the variance in entertainment purpose internet user visitors can be predicted by revisits. Higher revisit is a signal of low spatial ability and low interest in internet for entertainment purposes.</p> | <p>(Juvina &amp; Oostendorp, 2006) pg:267</p>  |  |

Table 1 (cont.)

| Expectation                          | Suggested Metrics   | Critique   | Resource                 |
|--------------------------------------|---|--|--------------------------|
| Improve the information architecture | <ul style="list-style-type: none"> <li>▪ <i>Overall traffic volumes</i></li> </ul>            | If high, it is a positive signal.  | (Peterson, 2004), pg:233 |
|                                      | <ul style="list-style-type: none"> <li>▪ <i>Percent new visitors</i></li> </ul>               | New visitors are those who may struggle more to find information as they learn your information architecture. The greater the volume of new visitors, the more likely your help/support page will be visited. It is a good signal to have high percent of new visitors.  |                          |
|                                      | <ul style="list-style-type: none"> <li>▪ Page “stickiness”</li> </ul>                         | It is equal to 1-Bounce Rate. Increase of this metric is a positive signal; this means that visitors are not leaving your site from the entrance page.   |                          |
|                                      | <ul style="list-style-type: none"> <li>▪ <i>Search keywords and phrases</i></li> </ul>        | Keeping track of the top search engine words and phrases that are driving traffic to you will provide insight into what visitors are expecting to find in the website.   |                          |
|                                      | <ul style="list-style-type: none"> <li>▪ <i>Percent of visits under 90 seconds</i></li> </ul> | It is unlikely that a visitor can do much on a content based site in less than 90 seconds other than find contact information. Increase in this metric is a negative signal because it shows that visitors are not engaged in the website.   |                          |
|                                      | <ul style="list-style-type: none"> <li>▪ <i>Entry pages and content</i></li> </ul>            | Keeping track of the top 10 entry pages and content that visitors are most likely to begin their visit will help you identify the type of information visitors are most interested in.   |                          |
|                                      | <ul style="list-style-type: none"> <li>▪ <i>“Information Find” Conversion Rate</i></li> </ul> | If you track the process of visitors moving from your home page to specific information you can generate an "information find" conversion rate. If you define the path in the right way, increasing conversion rate is a positive signal because this means that visitors understand the information architecture and find the data they search for. |                          |

Table 1 (cont.)

| Expectation                                  | Suggested Metrics  | Critique   | Resource                                 |
|--|--|--|--|
| Improve the information architecture         | <ul style="list-style-type: none"> <li>▪ <i>Top Pages and Content Requested by new visitors</i></li> </ul> | Knowing what first-time ever visitors are looking for on your web-site can help you tailor content and drive visitors towards the most popular web pages.  |  |
| Investigate the general usability of a site  | <ul style="list-style-type: none"> <li>▪ <i>Average page views per visit</i></li> </ul>                    | If low, there is navigational problem  | (Hasan, Morris & Proberts, 2009) pg.701  |
|  | <ul style="list-style-type: none"> <li>▪ <i>Bounce Rate</i></li> </ul>                                     | If high, navigational problem  |  |
|  | <ul style="list-style-type: none"> <li>▪ <i>Average searches per visit</i></li> </ul>                      | If low, it means the site has good navigation so that a search facility is not needed or alternatively that there are problems with the search facilities. If high, information architecture of the web site is too complex that visitors cannot find what they look for without using insite search service.      |  |
|  | <ul style="list-style-type: none"> <li>▪ <i>Percent of visits using search</i></li> </ul>                  |  |  |
|  | <ul style="list-style-type: none"> <li>▪ <i>Percentage of click depth visits</i></li> </ul>                | Decide which of these three categories your web page belongs: low (<2 pages), medium (3 to average page views), high (>more than average page views). If your web page is in low page depth category, there is navigational problem. Higher is better.   |  |
| Examine the usability of the internal search | <ul style="list-style-type: none"> <li>▪ <i>Average searches per visit</i></li> </ul>                      | If these two metric and average page views per visit are low, there is a problem. If these two metrics are low and average page views per visit is high, this can mean that visitors rely on navigation rather than the internal search of the site to find what they needed which means that there is no problem. | (Hasan, Morris, & Proberts, 2009) pg.701 |
|  | <ul style="list-style-type: none"> <li>▪ <i>Percent of visits using search</i></li> </ul>                  |  |  |
|  | <ul style="list-style-type: none"> <li>▪ <i>Search results to site exits ratio</i></li> </ul>              | Needs to be low, If high, this indicates that users are leaving the sites immediately after conducting a search and that the site probably does have usability problems related to the inaccuracy of the search results  |  |

Table 1 (cont.)

| Expectation   | Suggested Metrics   | Critique   | Resource                                |
|---|---|--|---|
| Test usability problems with the information architecture | <ul style="list-style-type: none"> <li>▪ <i>Percentage of time spent visits</i></li> </ul>  | Decide which of these three categories your web page belongs: (Little (0 seconds-3 minutes), Medium (3 minutes-10 minutes), High (>10 minutes) If your web page belongs to high category, it is a positive signal. Large number of visitors who spent little time on the sites indicates potential usability problems with the information architecture. | (Hasan, Morris & Proberts, 2009) pg.702 |
|   | <ul style="list-style-type: none"> <li>▪ <i>Average searches per visit</i></li> </ul>       | The low value of these metrics together with the high percentages of visits with medium click depth provides a potential indication that the architecture of site has fewer problems as visitors were able to navigate through the site, implying that search facilities may not be needed.  |   |
|   | <ul style="list-style-type: none"> <li>▪ <i>Percent of visits using search</i></li> </ul>   |  |   |
|   | <ul style="list-style-type: none"> <li>▪ <i>Percentage of click depth visits</i></li> </ul> | Low value of the average page views per visits metric together with the high percentage of visits with low click depth provides a potential indication that site has a complex architecture and that users could not navigate within it.   |   |

Table 1 (cont.)

| Expectation                               | Suggested Metrics   | Critique   | Resource                                    |
|---|---|--|---|
| Test interest in the content of the sites | <ul style="list-style-type: none"> <li>▪ <i>Percentage of click depth visits</i></li> </ul>                                 | Needs to be high, Low value of the percentages of visits in terms of the number of pages viewed indicates that visitors are not engaged in your web site.  | (Hasan, Morris & Proberts, 2009) pg.702-703 |
|   | <ul style="list-style-type: none"> <li>▪ <i>Bounce Rate</i></li> </ul>  | High <i>Bounce Rate</i> implies that either users are uninterested in the sites' content or that the design is unsuitable for the users. From the metrics it is difficult to determine if a high Bounce rate is due to content or design problems. |   |
|   | <ul style="list-style-type: none"> <li>▪ <i>Percentage of time spent visits</i></li> </ul>                                  | Check with percentage of click depth visits, if more than 3 pages are visited only upto 3 minutes are spent during visits, this means there are content problems. It is better to have high values in this metric.                                 |   |
| Test help/support service                 | <ul style="list-style-type: none"> <li>▪ <i>Information find conversion (ranges for the selected pages) rate</i></li> </ul> | The low information find conversion rate) is a signal of access to help/support pages. It needs to be increased to improve the customer service.   | (Hasan, Morris & Proberts, 2009) pg.703     |

## CHAPTER 3

### METHODOLOGY: GOOGLE ANALYTICS

In this thesis study, we use both a Web Analytic tool and *think-aloud* study to improve the information architecture of our case study web-site <http://www.ii.metu.edu.tr> as suggested by the literature in Section 2.4. Among the web analytic tools, we decide using Google Analytics which we analyzed in detail in Section 2.4.2. In this chapter, we will describe the methodology we applied for using Google Analytics reports. For Google Analytics metrics and dimensions, abbreviations are used in the example tables throughout this chapter. The abbreviations are given in Table 2: Google Analytics Dimension and Metric Abbreviation List.

We start this chapter by introducing our case study web site using Google Analytics reports. In the second step, we will explain how we prepare Google Analytics reports and how we choose the time period of these reports. We will end up this chapter by explaining in detail the metrics and dimensions which have confusing explanations from our point of view, apply synchronization of what visitors possess and what our web site requires and lastly apply landing page optimization technique.

**Table 2: Google Analytics Dimension and Metric Abbreviation List**

| Dimension                      | Abbreviation |
|--------------------------------|--------------|
| City                           | C            |
| Date                           | Dt           |
| Day                            | D            |
| ExitPagePath                   | EPP          |
| Hour                           | H            |
| Keyword                        | K            |
| LandingPagePath                | LPP          |
| NetworkLocation                | NL           |
| NextPagePath                   | NPP          |
| PageDepth                      | PD           |
| PagePath                       | PP           |
| PreviousPagePath               | PPP          |
| ReferralPath                   | RP           |
| Source                         | S            |
| TargetLandingPage <sup>1</sup> | TLP          |
| VisitorType                    | VT           |

| Metric          | Abbreviation |
|-----------------|--------------|
| Bounces         | B            |
| Entrances       | E            |
| ExitPagePath    | EPP          |
| Exits           | EX           |
| NewVisits       | NV           |
| PageViews       | PV           |
| TimeonPage      | ToP          |
| TimeonSite      | ToS          |
| UniquePageViews | UPV          |
| Visits          | V            |

### ***3.1 Properties of the Institute Web-Site Chosen As Case Study***

To improve the institute web site, Google Analytics is the first tool we have used to reach our aim in this study. Above, we touched upon the explanation of dimensions and

---

<sup>1</sup> *TargetLandingPage* is not a dimension defined by Google Analytics. We created and used this dimension and its abbreviation in our analysis from Section 3.8 to 3.13.

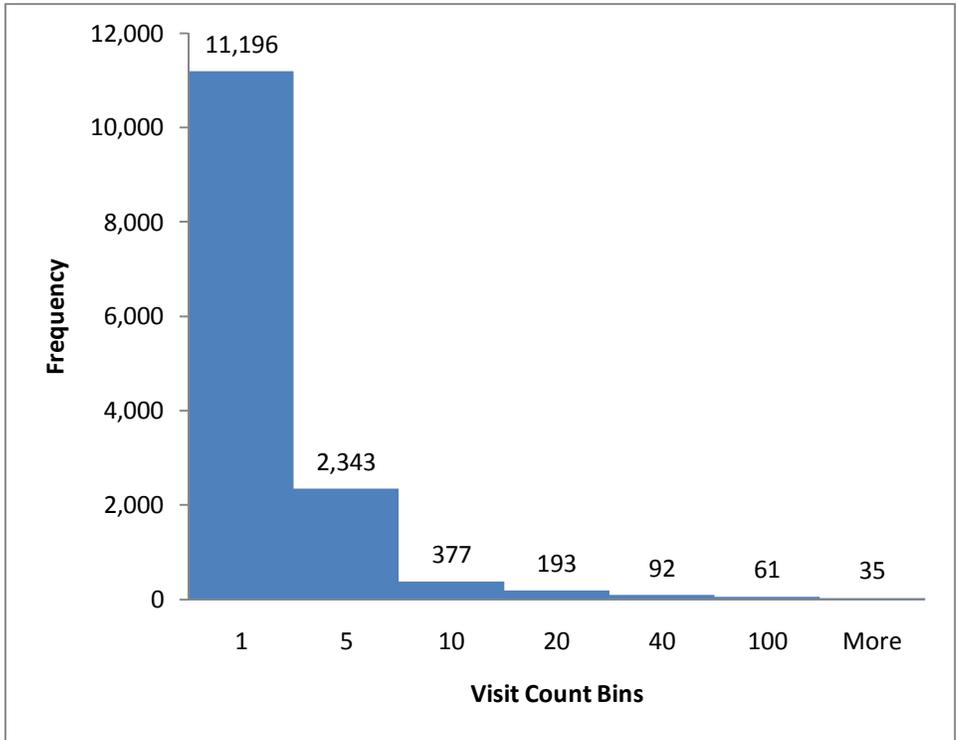
metrics. Like a tailor, we need to know our customer very well to solve its problems. We cannot use the same methodology for different web-sites which have different aims and traffic conditions.

Our case study is based on our institute web site. It is a content based web-site. It is established to communicate with the potential, current and graduate students, potential and current faculty. It provides information about the programs, courses, application procedure, and faculty. Google Analytics tracking code has been added to our web-site on 06.09.2007. A new web-site has been put into service on May 2009. To form the new web site, web sites of other famous universities and university web sites that took awards are analyzed. Content management systems are examined considering white-papers, forums and expert comments. As a result, Drupal which is an open source Content Management System is decided. Metatags Nodewords module of Drupal is also installed to enable adding Metatags functionality. Designers of the web site added Metatags to each program web page. Metatags are decided as keywords of each web page. To the visitors who use one of these Metatags as keywords to land on the Informatics Institute web site, a custom prepared web page by Drupal is served. Drupal forms this custom web page by appending content of each web page including the searched meta tag one under the other and putting the Metatag as the title of the custom page. Metatag functionality is used to increase the number of visitors and serve customized web pages to each visitor.

Although we aim to add requirements of all the user groups: faculty members, students, potential students, research assistants; requirement analysis of the new web-site are restricted with the experiences of faculty members and project assistants. After putting into service the new-website, a *think-aloud* study within the framework of this thesis study is realized by taking into account the requirements of current and potential students. Since putting into service the new web-site, it is constantly being improved. Our case study web site is an information driven web site, it is composed of about 6,733 web pages and our concern in this thesis study is to improve its *information architecture*.

During the year 2009; 133,120 visits, 387,797 pageviews, 2.90 pages/visit, 55.89 % Bounce Rate, 2 minutes 54 seconds Average Time on Site, 48.12% New Visits have been realized. Top ten web-pages during 2009 are the main page, information on Application page, web page of Information Systems program, information on full-time faculty page, web page of Cognitive Science program, web page of Modeling and Simulation program and contact page. Nearly 40% of visits are direct, about 28% is from referring sites and 32% of visits is from search engines. Our web site receives visits from search engines with 14,331 different keywords. 1,599 of these keywords have zero visits which means after the visit was realized to our web-site these keywords are used for subsequent searches and so we cannot see the usage frequency of these keywords (Anil, 2010). The most popular keyword is “ii.metu.edu.tr”. However, as can be seen from

Figure 3, the remaining 12,732 keywords have a lower tailed shape frequency graph. The horizontal axis shows the visit count bins and vertical axis shows the frequency of these visit counts. 11,196 keywords are not used as the entrance keywords (meaning of zero visit) or used only in 1 visit whereas 35 of the keywords among 12,732 are used more than 100 times.



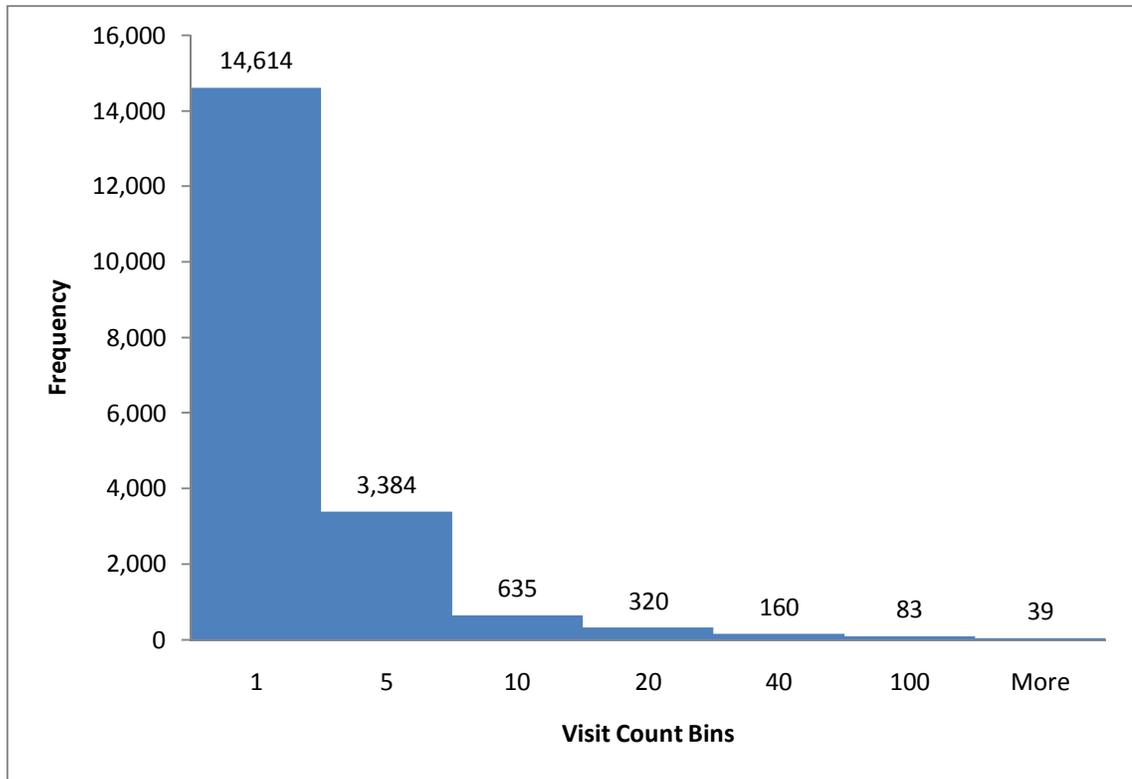
**Figure 3: 2009 Period Keywords Histogram**

The most popular web-page is the home page: “<http://www.ii.metu.edu.tr/>”. Below table gives the top 30 landing pages and their Entrances, Bounces and Bounce Rate (BR) values for 2009 period. Landing pages which are studied in depth in this study are given in bold and italic style.

**Table 3: 2009 Period Top 30 Landing Pages; Their Entrances, Bounces and Bounce Rate**

|    | <b>Landing Page</b>  | <b>E</b> | <b>B</b> | <b>BR</b> |
|----|--|----------|----------|-----------|
| 1  | /  | 74,380   | 42,578   | 57%       |
| 2  | /index.php   | 13,938   | 5,791    | 42%       |
| 3  | <i>/academic_program/informatics-online</i>  | 1,827    | 724      | 40%       |
| 4  | /Graduate_Programs/Modeling_and_Simulation/GameTechnologies.php  | 1,760    | 1,215    | 69%       |
| 5  | /content/darwin-day  | 1,549    | 1,229    | 79%       |
| 6  | /Application/Application_Home_TR.php   | 1,418    | 918      | 65%       |
| 7  | /Application/Application_Home.php  | 1,263    | 442      | 35%       |
| 8  | /Graduate_Programs/Cognitive_Science/CognitiveScience.php  | 1,209    | 628      | 52%       |
| 9  | /is100   | 1,042    | 141      | 14%       |
| 10 | /is100/instructors   | 925      | 184      | 20%       |
| 11 | /News/donemizni.php  | 754      | 640      | 85%       |
| 12 | <i>/Graduate_Programs/Information_Systems/Information_Systems/InformationSystems.php</i>   | 735      | 395      | 54%       |
| 13 | /News/Academic_P.php   | 612      | 460      | 75%       |
| 14 | /Graduate_Programs/Health_Informatics/HealthInformatics.php  | 559      | 355      | 64%       |
| 15 | /announcements/2009/12/2009-2010-akademik-y%C4%B1%C4%B1-bahar-d%C3%B6nemi-lisans%C3%BCst%C3%BC-programlara-ba%C5%9Fvuru-tarihleri    | 482      | 363      | 75%       |
| 16 | <i>/Graduate_Programs/Information_Systems/Software_Management/SoftwareManagement.php</i>   | 454      | 206      | 45%       |
| 17 | /Faculty/Research_Assistants.php   | 451      | 314      | 70%       |
| 18 | /Faculty/Full_Time.php   | 425      | 254      | 60%       |
| 19 | /CMS/drupal-6.10/  | 417      | 73       | 18%       |
| 20 | /News/thesis_submission.php  | 403      | 320      | 79%       |
| 21 | /Statistics/Publications.php   | 373      | 270      | 72%       |
| 22 | /Alumni/Alumni.php   | 347      | 183      | 53%       |
| 23 | /Application/OYP.php   | 337      | 251      | 74%       |
| 24 | /Faculty/Former_Assistants.php   | 307      | 229      | 75%       |
| 25 | /404.html?page=/is100/course_calendar.htm&from=http://kepler.ii.metu.edu.tr/~is100/announcements.htm                                 | 305      | 233      | 76%       |
| 26 | /tr/announcements/2009/12/2009-2010-akademik-y%C4%B1%C4%B1-bahar-d%C3%B6nemi-lisans%C3%BCst%C3%BC-programlara-ba%C5%9Fvuru-tarihleri | 287      | 239      | 83%       |
| 27 | /tr/content/100-introduction-information-technology-and-applications   | 287      | 152      | 53%       |
| 28 | /content/applying-informatics-institute  | 268      | 130      | 49%       |
| 29 | <i>/departments/department-work-based-learning-studies</i>   | 255      | 88       | 35%       |
| 30 | /Contact/Contact.php   | 240      | 156      | 65%       |

During the selected period *01.09.09 to 21.03.10*; 132,409 visits, 441,412 page views, 3.3 pages/visit, 47.31% *Bounce Rate*, 3 minutes 12 seconds Average Time on Site, 55.95% New Visits have been realized. Almost 32% of visits was direct, about 24% was from referring sites and 44% of visits was from search engines. 20,166 keywords have been used in visits from search engines. 909 of these keywords has zero visits which means after the visit was realized to our web-site these keywords were used for subsequent searches and so we cannot observe the usage frequency of these keywords (Anil, 2010). However, as can be seen from the Figure 4, the remaining 19,257 keywords have a lower tailed shape frequency graph. The most popular keyword is “is100”. The horizontal axis shows the visit count bins and the vertical axis shows the frequency of these visit counts. 14,614 keywords are not used as the entrance keyword (meaning of zero visit) or used only in 1 visit whereas only **39** of the keywords among 19,257 are used more than 100 times.



**Figure 4: Selected Period Keywords Histogram**

The most popular web-page is the main home page: "<http://www.ii.metu.edu.tr/index.php>." Below table gives the top 30 landing pages and their Entrances, Bounces and Bounce Rate (BR) values for the selected period. Landing pages which are the main page of analyzed programs in this study are given in bold and italic style. This information will be beneficial to have an understanding of what kind of a web-page we have studied on.

**Table 4 : Selected Period Top 30 Landing Pages; Their Entrances, Bounces and Bounce Rate**

|    | <b>Landing Page</b>  | <b>E</b> | <b>B</b> | <b>BR</b> |
|----|--|----------|----------|-----------|
| 1  | /index.php   | 32,375   | 12,107   | 37%       |
| 2  | /  | 17,247   | 9,107    | 53%       |
| 3  | /is100   | 15,691   | 2,170    | 14%       |
| 4  | <i>/academic_program/informatics-online</i>  | 5,641    | 2,396    | 42%       |
| 5  | /content/online-master   | 2,565    | 1,624    | 63%       |
| 6  | /content/darwin-day  | 1,557    | 1,235    | 79%       |
| 7  | /is100/presentations_and_tutorials   | 1,414    | 601      | 43%       |
| 8  | /is100/instructors   | 1,234    | 296      | 24%       |
| 9  | /tr/content/100-introduction-information-technology-and-applications   | 732      | 209      | 29%       |
| 10 | /course/current_term_course_list   | 721      | 248      | 34%       |
| 11 | /announcements/2009/12/2009-2010-akademik-y%C4%B1%C4%B1-bahar-d%C3%B6nemi-lisans%C3%BCst%C3%BC-programlara-ba%C5%9Fvuru-tarihleri    | 680      | 513      | 75%       |
| 12 | /content/applying-informatics-institute  | 471      | 241      | 51%       |
| 13 | <i>/academic_program/work-based-learning</i>   | 454      | 216      | 48%       |
| 14 | /content/information-newcomers   | 365      | 226      | 62%       |
| 15 | /404.html?page=/is100/course_calendar.htm&from=http://kepler.ii.metu.edu.tr/~is100/announcements.htm                                 | 345      | 255      | 74%       |
| 16 | /content/questionnaire-testing-informatics-institute-web-site  | 321      | 145      | 45%       |
| 17 | /tr  | 307      | 94       | 31%       |
| 18 | /News/thesis_submission.php  | 296      | 222      | 75%       |
| 19 | /content/who-may-apply   | 295      | 182      | 62%       |
| 20 | /tr/announcements/2009/12/2009-2010-akademik-y%C4%B1%C4%B1-bahar-d%C3%B6nemi-lisans%C3%BCst%C3%BC-programlara-ba%C5%9Fvuru-tarihleri | 292      | 244      | 84%       |
| 21 | /Graduate_Programs/Cognitive_Science/CognitiveScience.php  | 290      | 151      | 52%       |
| 22 | /tr/category/tags/uzaktan-e%C4%9Fitim-veren-%C3%BCniversiteler   | 290      | 200      | 69%       |
| 23 | /Graduate_Programs/Modeling_and_Simulation/GameTechnologies.php  | 283      | 176      | 62%       |
| 24 | <i>/departments/department-work-based-learning-studies</i>   | 283      | 107      | 38%       |
| 25 | <i>/academic_program/software-management</i>   | 278      | 79       | 28%       |
| 26 | /News/donemizni.php  | 254      | 209      | 82%       |
| 27 | /academic_program/cognitive-science  | 253      | 97       | 38%       |
| 28 | <i>/academic_program/information-systems</i>   | 251      | 138      | 55%       |
| 29 | /content/100-introduction-information-technology-and-applications  | 238      | 114      | 48%       |
| 30 | /content/metu-turkish-corpus   | 221      | 93       | 42%       |

### ***3.2 How did we prepare Google Analytics reports?***

We chose Java language and used NetBeans IDE 6.8 and exported reports in Microsoft Windows 2007 Excel to analyze. The code generates reports in text files. Later, we import them into Excel by selecting the “File Origin” as “1254 Turkish (Windows)”.

### ***3.3 Time Period Selection***

Think-aloud study is carried out for the period 25.02.2010-21.03.2010. To have the Google Analytics reports of the realized study, we need to include this period as reporting period.

Our motivation in this study is to assess whether potential and current students are able to find the relevant information about the programs on our web site. Therefore, questions in the questionnaire are prepared as to meet this requirement. The time period selected for the study includes pre-application and application period for our programs according to the academic calendar. Application to the graduate programs for the spring semester according to Academic Program was between 01.12.2009 and 28.12.2009. We start pre-application period from September’09 to include more visits and we conclude to take reports of the period between **01.09.09 and 21.03.10**.

### ***3.4 Important Findings about Google Analytic Reports***

While reading the Google Analytics reports, there are two issues which need to be considered to understand the reports accurately. Google Analytics reports metrics zero

instead of not applicable if that metric is not calculated for the selected dimension. For example, if you take a report which includes *PagePath* dimension, *TimeonSite* and *Entrances* metrics, *TimeonSite* value will be reported more than zero only in the tuples which have *Entrances* values more than zero. Because *TimeonSite* is a session based calculated metric and it is reported only at the *Entrances* page of each session, this metric is not applicable for every *PagePath*. Instead of not applicable, Google Analytics reports *TimeonSite* zero for the other *PagePaths* which are not entrance pages. A second issue is all dimension-metric combination tuples of the requested custom report is not reported by Google Analytics. To be reported, value of one of the metrics needs to be more than zero. For example; if you select *PagePath* as dimension and select *Bounces*, *Entrances* and *Exits* as metrics, *PagePath*'s for which *Bounces*, *Entrances* and *Exits* values are all equal to zero will not be reported. However, this does not necessarily mean that the indicated page has not been visited at all. It only means that no one either entered or exited from it. To get the report of all the visited pages, *PageViews* metric need to be added to the custom report.

Google Analytics reports are session based. "A visitor session ends after 30 minutes of inactivity on your website, or when the browser exits. Google Analytics is able to determine the start of a new session by the absence of "-utmb" or "-utmz" session cookies (Google, 2010h)." This means that if cookies are not deleted, browser is not closed and a visit to the same page before 30 minutes of expiration time is realized, this new visit will be reported by Google Analytics as part of previously opened session.

A user can visit a web page in three different ways:

1. Access the website directly writing the URL or domain name at the browser's address bar which will be reported by Google Analytics as direct traffic.
2. Access the website from a referral link which will be reported by Google Analytics as referring site traffic.

3. Access the website from a search engine by using a keyword which will be reported by Google Analytics as a traffic from search engine.

A user can navigate within a web site again in three different ways:

1. By using the links on web pages
2. By doing multiple searches with different keywords to find specific information on the web pages
3. By using internal search service if available on the website.

### **3.4.1 Misleading Keyword and LandingPagePath Pairs**

Google Analytics reports misleading *Keyword and LandingPagePath* pairs as it is shown in Table 5 if *the landing page does not include Google Analytics tracking code or if the page is tracked by Google Analytics with a different account*. We applied a test in which we searched “tugba temizel deneme1” keyword in Google and clicked on the search result “/~tugbatt/index.html” web page which resulted in landing to that page and this page is not tracked by the same Google Analytics account. Next we visited two more non-tracked web pages and then visited “/academic\_program/informatics-online” as the first web page being tracked by Google Analytics. As it can be seen in Table 5, Google Analytics reports this session as a visit which used “tugba temizel deneme1” *keyword* and landed to “/academic\_program/informatics-online” which is inaccurate. Google Analytics reported “/academic\_program/informatics-online” web page inaccurately as *LandingPagePath* because our Google Analytics account starts tracking the visitor session at that page. The real *LandingPagePath* “/~tugbatt/index.html” and the following web pages we have visited before the reported *LandingPagePath* are tracked by another Google Analytics account. Users of Google Analytics encounter reported impossible *Keyword-LandingPagePath* pairs although it is impossible to land

on the page reported with the keyword reported, the example we give in Table 5 also explains this situation. If the number of non-tracked web pages are high in a web site or if they are tracked by other Google Analytics accounts, possibility of encountering similar inaccurate *Keyword-LandingPagePath* pair reports by Google Analytics increases.

**Table 5: Misleading Keyword-LandingPagePath Pair Report**

| NL                               | H  | K                           | LPP                                  | PP                                   | PPP         | B | E | EX | NV | PV |
|----------------------------------|----|-----------------------------|--------------------------------------|--------------------------------------|-------------|---|---|----|----|----|
| ulastirma<br>bakanligi<br>ankara | 16 | tugba<br>temizel<br>deneme1 | /academic_program/informatics-online | /academic_program/informatics-online | (Entrances) | 0 | 1 | 0  | 0  | 1  |

### 3.4.2 Double Pageview Reporting Problem and Its Reason

Table 6 shows the session information of a visit done by us to test whether double *PageView* reporting by Google Analytics exist or not and then analyze the reasons behind. We searched “ii.metu.edu.tr newwindow” *Keyword* in Google and among the search results clicked on the “/index.php” to land on the institute web site, visited 5 different pages only once and exited from the “/research-groups” web page. However, as you can see from Table 6, *PageViews* for each visited page are reported as two although we visit each page only once. This report is taken after 4 days from the date of the test.

**Table 6: Double PageView Reporting**

| K                           | LPP        | PP                                   | EPP              | E | PV | EX |
|-----------------------------|------------|--------------------------------------|------------------|---|----|----|
| ii.metu.edu.tr<br>newwindow | /index.php | /academic_program/informatics-online | /research-groups | 0 | 2  | 0  |
| ii.metu.edu.tr<br>newwindow | /index.php | /content/contact                     | /research-groups | 0 | 2  | 0  |
| ii.metu.edu.tr<br>newwindow | /index.php | /content/who-may-apply               | /research-groups | 0 | 2  | 0  |
| ii.metu.edu.tr<br>newwindow | /index.php | /index.php                           | /research-groups | 1 | 2  | 0  |
| ii.metu.edu.tr<br>newwindow | /index.php | /research-groups                     | /research-groups | 0 | 2  | 1  |

To find the reason behind double pageview reporting of Google Analytics, we added the dimensions *PreviousPagePath* and *NextPagePath* to identify the path of the visit in our custom report. As you can see in Table 7, double pageview reporting is done by Google Analytics because all the pages are reported as each page is **refreshed** which is not the real case. We can conclude “page refresh” when the *PreviousPagePath* and *NextPagePath* are same for the same tuple.

**Table 7: Reason Behind Double PageView Reporting**

| PP  | NPP   | PPP   | E        | EX       | PV       | Comment        |
|---|---|---|----------|----------|----------|----------------|
| /index.php                                  | /index.php                                  | (Entrances)                                 | 1        | 1        | 1        |                |
| /index.php                                  | /index.php                                  | (Entrances)                                 | 1        | 0        | 1        |                |
| <b>/index.php</b>                           | <b>/index.php</b>                           | <b>/index.php</b>                           | <b>0</b> | <b>0</b> | <b>1</b> | <b>REFRESH</b> |
| /academic_program/informatics-online        | /academic_program/informatics-online        | /index.php                                  | 0        | 0        | 1        |                |
| <b>/academic_program/informatics-online</b> | <b>/academic_program/informatics-online</b> | <b>/academic_program/informatics-online</b> | <b>0</b> | <b>0</b> | <b>1</b> | <b>REFRESH</b> |
| /content/contact                            | /content/contact                            | /academic_program/informatics-online        | 0        | 0        | 1        |                |
| <b>/content/contact</b>                     | <b>/content/contact</b>                     | <b>/content/contact</b>                     | <b>0</b> | <b>0</b> | <b>1</b> | <b>REFRESH</b> |
| /content/who-may-apply                      | /content/who-may-apply                      | /content/contact                            | 0        | 0        | 1        |                |
| <b>/content/who-may-apply</b>               | <b>/content/who-may-apply</b>               | <b>/content/who-may-apply</b>               | <b>0</b> | <b>0</b> | <b>1</b> | <b>REFRESH</b> |
| /research-groups                            | /research-groups                            | /content/who-may-apply                      | 0        | 0        | 1        |                |
| <b>/research-groups</b>                     | <b>/research-groups</b>                     | <b>/research-groups</b>                     | <b>0</b> | <b>1</b> | <b>1</b> | <b>REFRESH</b> |

### **3.5 Detailed Explanations of Google Analytics Dimensions and Metrics**

Google Analytics dimensions and metrics are not explained in detail in anywhere which may lead to misunderstanding as we have already experienced at the beginning of this study. Also, the logic behind valid dimension-metric combinations is not provided which results in time consuming trial and errors for custom reporting. This also leads to lower use rate of these services. Here, we will explain with examples these metrics and dimensions in detail which have the most confusing definitions. Examples are given in tables and abbreviations are used for these metrics and dimensions. The abbreviations used throughout this chapter are given in Table 2. We will also explain why there exists valid dimension-metric combination terminology and how to decide whether a dimension and metric is a valid combination or not. We believe that this study will fill a big gap in this area.

#### **3.5.1 LandingPagePath**

It is defined as “The path component of the first page in a user's session, or "landing" page (Google, 2010f)” but from this explanation we cannot conclude easily if in a session more than one search is done, which of the visited pages will be labeled as *LandingPagePath*. We test this question and other possible scenarios in this study, however, Google Analytics reports of these tests are not consistent. If someone searches more than one keyword and lands different pages in the same session, we can see them in the analytics reports as can be seen in Table 8. It is important to note that **in one session** *Entrances* metric is only considered for the first landing page. If the searches are not done in the same session, each keyword will be reported as entrance of each session. On the other hand, Exits *metric* is independent of the session information and based on page information. This metric gives how many times that page is used as an exit page.

In a test, we use four different keywords which land to four different web pages. This situation is reported by Google Analytics as 3 different keywords are used to land on four different web pages as can be seen in Table 8. In real case, “ion metu land3” *Keyword* is used to land on “/academic\_program/informatics-online” web page. All the visits are accepted as a new session since each has *Entrances* value as 1.

**Table 8: Keyword-LandingPagePath Pairs Test 1**

| <b>K</b>                       | <b>LPP</b>                            | <b>PD</b> | <b>E</b> | <b>EX</b> |
|--------------------------------|---------------------------------------|-----------|----------|-----------|
| information systems metu land1 | /index.php                            | 8         | 1        | 1         |
| work based learning metu land2 | /academic_program/work-based-learning | 10        | 1        | 1         |
| work based learning metu land2 | /academic_program/informatics-online  | 6         | 1        | 1         |
| software management metu land4 | /tr/category/tags/software-management | 6         | 1        | 1         |

In another test in which we use two keywords “ion metu keyword1” and “software management metu keyword2” consecutively, only the second keyword is reported by Google Analytics. The web pages visited after the first search are reported as they are direct traffic. However, *LandingPagePath* of the session is accurately reported as the landing page of the first search keyword which is “academic\_program/informatics-online” although reported *Keyword-LandingPagePath* pair is not accurate. We actually land on “/tr/category/tags/software-management” web page after we searched “software management metu keyword2”.

**Table 9 : Keyword-LandingPagePath Pairs Test 2**

| <b>K</b>                          | <b>LPP</b>                           | <b>PP</b>   | <b>NPP</b>  | <b>E</b> | <b>EX</b> |
|-----------------------------------|--------------------------------------|---|---|----------|-----------|
| (not set)                         | /academic_program/informatics-online | /academic_program/informatics-online                        | /academic_program/informatics-online                        | 1        | 0         |
| (not set)                         | /academic_program/informatics-online | /content/who-may-apply                                      | /content/who-may-apply                                      | 0        | 0         |
| (not set)                         | /academic_program/informatics-online | /full_time_faculty  | /full_time_faculty  | 0        | 0         |
| software management metu keyword2 | /academic_program/informatics-online | /tr/category/tags/software-management                       | /tr/category/tags/software-management                       | 0        | 0         |
| software management metu keyword2 | /academic_program/informatics-online | /tr/node/51   | /tr/node/51   | 0        | 0         |
| software management metu keyword2 | /academic_program/informatics-online | /course_tr?field_academic_program_nid=51&field_type_nid=All | /course_tr?field_academic_program_nid=51&field_type_nid=All | 0        | 1         |

So if one is interested in the first entry pages of search engine traffic sessions on the website, the *LandingPagePath* dimension will be insufficient. The best way to use is add *PreviousPagePath* and *PagePath* dimensions to your custom report and filter the report for *PreviousPagePath* which takes the value “(entrance)” and Keyword which does not take the value “(not set)”. The reported keywords will be the entrance keywords of the sessions. If we apply this technique to Table 9, we will encounter that the entrance is done by a direct visit, there is no keyword as reported in Table 10.

**Table 10: Detecting the Entrance Keyword**

| <b>K</b>  | <b>LPP</b>                           | <b>PPP</b> | <b>E</b> | <b>EX</b> |
|-----------|--------------------------------------|------------|----------|-----------|
| (not set) | /academic_program/informatics-online | (entrance) | 1        | 0         |

The problem in reporting can be because of opening the results in a new tab, in a new window or the time between two searches are so little to be identified by Google Analytics. This can be research objective of further studies.

### 3.5.2 NewVisits and VisitorType

Another confusing dimension and metric are that of *NewVisits* and *VisitorType*. According to Google, *NewVisits* metric is defined as the number of visitors whose visit to your website was marked as a first-time visit. *VisitorType* dimension is a boolean indicating if visitors are new or returning. Possible values for *VisitorType* are: New Visitor, Returning Visitor.

We have made the following Google search on the website: we searched “information systems metu nvt ” on the website on 23.06.2010 at hour 21:53 . Before doing that, we cleaned the cache and cookies in order to appear as “New Visitor” on the site. As there was no one who did the same search (keyword is unexpected one), we could easily pick the details of our session from analytics as can be seen below.

We landed to the main page in the same window. Then visited 4 different pages. We have visited one of the pages twice and one of the pages we have visited is not tracked by Google Analytics.

The report of the Google Analytics is given in Table 11. We have omitted *PageViews* metric as it is 1 for all the rows in the table to pay attention how *NewVisits* and *VisitorType* metrics differ.

**Table 11: New Visits and VisitorType Example**

| <b>K</b>                            | <b>VT</b>          | <b>LPP</b>        | <b>PP</b>                               | <b>NV</b> | <b>B</b> | <b>E</b> | <b>EX</b> |
|-------------------------------------|--------------------|-------------------|---|-----------|----------|----------|-----------|
| information systems metu nvt        | New Visitor        | /index.php        | /academic_program/information-systems   | 0         | 0        | 0        | 0         |
| information systems metu nvt        | New Visitor        | /index.php        | /content/applying-informatics-institute | 0         | 0        | 0        | 0         |
| information systems metu nvt        | New Visitor        | /index.php        | /content/general-information            | 0         | 0        | 0        | 1         |
| <b>information systems metu nvt</b> | <b>New Visitor</b> | <b>/index.php</b> | <b>/index.php</b>                       | <b>1</b>  | <b>0</b> | <b>1</b> | <b>0</b>  |

It is very clear that the last tuple is the entrance point to the site (by observing the *Entrances* metric). Pay attention to the *NewVisits*. It appears as “1” in the last tuple. But in the other tuples it is 0. But the visitor is marked as “New Visitor” in the *VisitorType* dimension. This is due to the fact that *NewVisits* metric appears only for the entrances pages. So if one is interested in the movements/paths of the user (user segmentation) in terms of new and returning user, it is sufficient to use *VisitorType*. But if you are interested in reasoning the problems of the first Entrances pages to be associated with the new visits, then *NewVisits* can be used.

### 3.5.3 TimeonSite, TimeOnPage, Visits and Entrances

Google defines *TimeonPage* as “How long a visitor spent on a particular page or set of pages. It is calculated by subtracting the initial view time for a particular page from the initial view time for a subsequent page. Thus, this metric does not apply to exit pages for your website (Google, 2010f).” On the other hand, Google defines *TimeonSite* as “The total duration of visitor sessions over the selected dimension.”

**Table 12: TimeonSite, TimeonPage, Entrances Cumulative Report**

| K         | S      | PP                                   | VT        | B | E  | ToS  | ToP  | PV | NV | EX |
|-----------|--------|--------------------------------------|-----------|---|----|------|------|----|----|----|
| part-time | google | /academic_program/information-system | Returning | 1 | 12 | 1419 | 2437 | 28 | 0  | 25 |

In Table 12, *TimeonPage* appears to be greater than *TimeonSite*. *TimeonSite* metric works on dimensions and for only the entrances pages with the selected dimension. It means that visitors that satisfy the four criteria in terms of *Keyword*, *Source*, *PagePath* and *VisitorType* (these are the dimensions we selected for our initial query) as shown. By satisfying the four criteria, we mean the *Returning Visitors* who came from Google search engine using the keyword “part-time” and visited “/academic\_program/information-system” page as the *Entrances* page have spent 1419 seconds in total on the website. On the other hand, 2437 seconds as *TimeonPage* is equal to the total time spent on the “/academic\_program/information-system” page by *Returning Visitors* who came from Google search engine using the keyword “part-time” and visited “/academic\_program/information-system” page and it might not necessarily be the *Entrances* page. For example, a dummy query we performed below shows the distinction:

**Table 13: Distinction between TimeonSite and TimeonPage Detailed Report**

| K                   | PP                                   | E | EX | PV | ToP | ToS | V |
|---------------------|--------------------------------------|---|----|----|-----|-----|---|
| ii.metu.edu.tr tab1 | /academic_program/informatics-online | 0 | 0  | 2  | 24  | 0   | 0 |
| ii.metu.edu.tr tab1 | /content/contact                     | 0 | 0  | 2  | 23  | 0   | 0 |
| ii.metu.edu.tr tab1 | /content/who-may-apply               | 0 | 0  | 2  | 30  | 0   | 0 |
| ii.metu.edu.tr tab1 | /index.php                           | 1 | 0  | 2  | 54  | 132 | 1 |
| ii.metu.edu.tr tab1 | /research-groups                     | 0 | 1  | 2  | 0   | 0   | 0 |

As can be seen, *TimeonSite* for the first entry page is the total of *TimeonPage* of all the pages visited during that session ( $24+23+30+54=132$  is *TimeonSite* for the session which is visited with “ii.metu.edu.tr tabl” query).

What about a tuple with zero *TimeonSite* value? Should we discard it? Is it an error? Consider the above example. As can be seen, *TimeonSite* is zero for the pages in which the *Visits* metric has zero value. *TimeonPage* metric works on the selected dimension whereas *TimeOnSite* takes value only if *Visits* metric takes value for that tuple. This means that *TimeOnSite* is a visit based calculated metric, not *page* based. Additionally, *Visits* metric takes value only if *Entrances* metric takes value. Also bear in mind that *Visits* is also an indicative metric for the entrance pages for the selected dimension. *Visits* metric is defined as “The total number of visits over the selected dimension. A visit consists of a single-user session (Google, 2010f).”

#### 3.5.4 PageDepth Dimension vs PageViews Metric

*PageDepth* is described by Google as “The number of pages visited by visitors during a session (visit). The value is a histogram that counts pageviews across a range of possible values. In this calculation, all visits will have at least one pageview, and some percentage of visits will have more whereas *PageViews* is explained as “The total number of pageviews for your website when aggregated over the selected dimension (Google, 2010f). For example, if you select this metric together with *PagePath*, it returns the number of page views for each page.” *PageDepth* is a dimension and *PageViews* is a metric.

Table 14 shows a visit from the same *NetworkLocation*, on the same *Day*, same *Hour*. Since only the second tuple includes *Entrances* value, we can certainly say that these two visits are part of the same session. Here, *PageDepth* is equal to 17, which is the total of *PageViews* in each visit.

**Table 14: PageDepth vs PageViews Cumulative Report**

| NL                         | H  | K                      | D          | PD | PV | E | V |
|----------------------------|----|------------------------|------------|----|----|---|---|
| eser telekomunikasyon a.s. | 14 | hasan ogul             | 06.05.2010 | 17 | 5  | 0 | 0 |
| eser telekomunikasyon a.s. | 14 | http://ii.metu.edu.tr/ | 06.05.2010 | 17 | 12 | 1 | 1 |

Table 14 gives the details of the visits stated above in the two tuples in Table 14 Here, the total number of *PageViews* is equal to 17, which is the value for *PageDepth*. We can here conclude that value of *PageDepth* dimension is calculated similar to *TimeonSite* which is the sum of *TimeonPage* visited in the same session. The difference is *TimeonSite* takes value only for the entrance pages but *PageDepth* takes value for each selected dimension.

**Table 15: Distinction between PageDepth and PageViews Detailed Report**

| NL                         | H  | K                      | Dt         | PD | LPP        | PP                                      | PV | E |
|----------------------------|----|------------------------|------------|----|------------|---|----|---|
| eser telekomunikasyon a.s. | 14 | hasan ogul             | 06.05.2010 | 17 | /index.php | /alumni                                 | 1  | 0 |
| eser telekomunikasyon a.s. | 14 | hasan ogul             | 06.05.2010 | 17 | /index.php | /biblio/author/352                      | 1  | 0 |
| eser telekomunikasyon a.s. | 14 | hasan ogul             | 06.05.2010 | 17 | /index.php | /people/algan-uskarci                   | 1  | 0 |
| eser telekomunikasyon a.s. | 14 | hasan ogul             | 06.05.2010 | 17 | /index.php | /people/hasan-ogul                      | 2  | 0 |
| eser telekomunikasyon a.s. | 14 | http://ii.metu.edu.tr/ | 06.05.2010 | 17 | /index.php | /academic_program/information-systems   | 1  | 0 |
| eser telekomunikasyon a.s. | 14 | http://ii.metu.edu.tr/ | 06.05.2010 | 17 | /index.php | /content/applying-informatics-institute | 1  | 0 |

**Table 15 (cont.)**

| NL                               | H  | K                      | Dt         | PD | LPP        | PP  | PV | E |
|----------------------------------|----|------------------------|------------|----|------------|---|----|---|
| eser<br>telekomunikasyon<br>a.s. | 14 | http://ii.metu.edu.tr/ | 06.05.2010 | 17 | /index.php | /content/doğum-günü-kutlaması   | 1  | 0 |
| eser<br>telekomunikasyon<br>a.s. | 14 | http://ii.metu.edu.tr/ | 06.05.2010 | 17 | /index.php | /content/informatics-institute-<br>website-users-guide                      | 1  | 0 |
| eser<br>telekomunikasyon<br>a.s. | 14 | http://ii.metu.edu.tr/ | 06.05.2010 | 17 | /index.php | /events/2010/05/doktora-phd-<br>yeterlilik-sözlü-jürileri-25-<br>mayıs-2010 | 1  | 0 |
| eser<br>telekomunikasyon<br>a.s. | 14 | http://ii.metu.edu.tr/ | 06.05.2010 | 17 | /index.php | /events/2010/05/doktora-<br>yeterlilik-sınavı-17-18-mayıs-<br>2010          | 2  | 0 |
| eser<br>telekomunikasyon<br>a.s. | 14 | http://ii.metu.edu.tr/ | 06.05.2010 | 17 | /index.php | /index.php  | 2  | 1 |
| eser<br>telekomunikasyon<br>a.s. | 14 | http://ii.metu.edu.tr/ | 06.05.2010 | 17 | /index.php | /portal   | 3  | 0 |

### 3.5.5 ReferralPath or Source?

*Referral Path* dimension is explained as “The path of the referring URL. If someone places a link to your website on their website, this element contains the path of the page that contains the referring link (Google, 2010f).” *Source* dimension is explained as “The domain (e.g. google.com) of the source referring the visitor to your website. The value for this dimension sometimes contains a port address as well. (Google, 2010f).” If you are interested in which search engines your visitors are coming from and if you want to get a report showing both the keywords used to land on your specified pages and the used search engines, you need to choose *Keyword*, *LandingPagePath* and *Source* as dimensions of your report. If you choose *ReferralPath* dimension instead of *Source* dimension, *ReferralPath* will give the URL of the webpage which the visitor clicked the link on and so come from. In your report, since the visitor does not use any keyword, you will get “not set” as value of *ReferralPath* which can be seen in Table 16. If you

choose *Source* as in the example report Table 17 , you will get the Search engine which the visitor is coming from.

**Table 16: ReferralPath Example Report**

| K              | LPP                                  | RP        | B | EX |
|----------------|--------------------------------------|-----------|---|----|
| ali kantar     | /academic_program/informatics-online | (not set) | 0 | 1  |
| ayfer kütük    | /academic_program/informatics-online | (not set) | 1 | 1  |
| ceyhan temürcü | /academic_program/informatics-online | (not set) | 0 | 1  |

**Table 17: Source Example Report**

| K                     | LPP                                  | S      | B | EX |
|-----------------------|--------------------------------------|--------|---|----|
| "ali kantar"          | /academic_program/informatics-online | google | 0 | 1  |
| "ayfer kütük"         | /academic_program/informatics-online | google | 1 | 1  |
| "ceyhan temürcü"      | /academic_program/informatics-online | google | 0 | 0  |
| "course section" metu | /academic_program/informatics-online | google | 0 | 0  |

### 3.5.6 Visitor Path Identification using LandingPagePath, Keyword, PagePath, PreviousPagePath, NextPagePath Dimensions and Entrances and Exits Metrics

Obtaining the complete path information of a user is not possible with standart Google Analytics reports. But it may be possible with Data Feed Query.

To demonstrate, we have prepared a custom report using Data Feed Query Explorer to retrieve the complete path information of the users. We chose *NetworkLocation*, *Keyword*, *PageDepth*, *LandingPagePath*, *PagePath*, *PreviousPagePath* and *NextPagePath* as dimensions and *Entances*, *Exits* and *PageViews* as metrics, start-date and end-date as 11.05.2009. Table 18 shows a visitor's session information from this custom report. To form this table, first of all, we decide which of the tuples in our custom report belongs to same session or not. If two tuples belong to same session,

*NetworkLocation* and *PageDepth* are expected to be the same. In addition, if *Keyword* and *LandingPagePath* are also same, we will be more certain to decide that they belong to same session.

If one is interested in the browsed pages in relation with *Keywords*, one can use *PagePath* dimension and *PageViews* metric in the queries. By this way, all the web pages visited in the same session which used “aspects of control and complementation in Turkish” as keyword to land on “thesis/2009/aspects-control-and-complementation-turkish” page in our case in Table 18 will be reported.

After selecting the tuples which belong to the same visitor session, by looking at *PreviousPagePath*, *PagePath* and *NextPagePath* dimensions and *Entrances* and *Exits* metrics, one can identify path information of a user session. The *PagePath* which takes *Entrances* value 1 and has *PreviousPagePath* value as “(entrance)”, that page is the first page of the visit path. On the other hand, *PagePath* which takes *Exit* value 1 or the value of *ExitPagePath* which we do not report here is the last visited page of this session. The other web pages visited during this session are detected by looking at *PreviousPagePath* and *NextPagePath*. Special attention needs to be given to web pages which are viewed more than once in the same session. In Table 18, we arrange the tuples according to the visit sequence of the web pages in this session. A total of 13 web pages are visited, one page which is “/tr/node/2333” is visited twice. It is visited as the second page and then fourth page. As you can see from Table 18, *PagePath* column shows the visited web pages. The first tuple has *PreviousPagePath* value as “(entrance)” and *Entrances* value as 1. These two indicate that *PagePath* in this tuple which is “/thesis/2009/aspects-control-and-complementation-turkish” is the first visited page of this session. The last page visited in this session is:

“thesis/resources/II\_thesis\_preparation\_guideline\_June2006.pdf?.....”. We understand this from its having *Exits* value as 1. “/tr/node/2343” is the third visited page in this session because it is the web page reported as the *NextPagePath* of second visited

*PagePath* and *PreviousPagePath* of it is the web page reported as second visited *PagePath*.

**Table 18: Visitor Session Information, Source: google, Keyword: aspects of control and complementation in Turkish, NetworkLocation: comcast cable communications holdings inc, LandingPagePath: /thesis/2009/aspects-control-and-complementation-turkish and PageDepth: 14**

|    | PPP  | PP  | NPP   | E | EX | PV |
|----|--|---|---|---|----|----|
| 1  | (entrance)   | /thesis/2009/aspects-control-and-complementation-turkish  | /thesis/2009/aspects-control-and-complementation-turkish  | 1 | 0  | 1  |
| 2  | /thesis/2009/aspects-control-and-complementation-turkish       | /tr/node/2333   | /tr/node/2333   | 0 | 0  | 1  |
| 3  | /tr/node/2333  | /tr/node/2343   | /tr/node/2343   | 0 | 0  | 1  |
| 4  | /tr/node/2343  | /tr/node/2333   | /tr/node/2333   | 0 | 0  | 1  |
| 5  | /tr/node/2333  | /tr/node/57   | /tr/node/57   | 0 | 0  | 1  |
| 6  | /tr/node/57  | /tr   | /tr   | 0 | 0  | 1  |
| 7  | /tr  | /index.php  | /index.php  | 0 | 0  | 1  |
| 8  | /index.php   | /content/job-postdoc  | /content/job-postdoc  | 0 | 0  | 1  |
| 9  | /content/job-postdoc   | /content/information-newcomers  | /content/information-newcomers  | 0 | 0  | 1  |
| 10 | /content/information-newcomers                                 | /academic_program/cognitive-science   | /academic_program/cognitive-science   | 0 | 0  | 1  |
| 11 | /academic_program/cognitive-science                            | /Thesis/resources/II_thesis_preparation_guideline_June2006.pdf  | /Thesis/resources/II_thesis_preparation_guideline_June2006.pdf  | 0 | 0  | 1  |
| 12 | /Thesis/resources/II_thesis_preparation_guideline_June2006.pdf | /Thesis/resources/II_thesis_preparation_guideline_June2006.pdf  | /Thesis/resources/II_thesis_preparation_guideline_June2006.pdf  | 0 | 0  | 2  |
| 13 | /Thesis/resources/II_thesis_preparation_guideline_June2006.pdf | /thesis/resources/II_thesis_preparation_guideline_June2006.pdf?field_department_name_nid=1499&field_msc_phd_value_many_to_one=All&field_year_value_many_to_one=All&title=&field_student_alumni_nid=All&field_keywords_value=&uid= | /thesis/resources/II_thesis_preparation_guideline_June2006.pdf?field_department_name_nid=1499&field_msc_phd_value_many_to_one=All&field_year_value_many_to_one=All&title=&field_student_alumni_nid=All&field_keywords_value=&uid= | 0 | 1  | 1  |

For example, Table 19 shows a single session data recorded on *Day*: 13, *Hour*: 13, with *Keyword*: “bilişim sistemleri doktora” from *NetworkLocation*: Middle East Technical University:

**Table 19: Single Session Information**

| PP   | PPP  | B | E | EX | NV | PV |
|--|--|---|---|----|----|----|
| /academic_program/information-system         | index.php                                    | 0 | 0 | 0  | 0  | 1  |
| /academic_program/information-system         | /tr/academic_program/bili%C5%9Fim-sistemleri | 0 | 0 | 1  | 0  | 3  |
| index.php                                    | /portal                                      | 0 | 0 | 0  | 0  | 1  |
| /portal                                      | (Entrances)                                  | 0 | 1 | 0  | 0  | 1  |
| /tr/academic_program/bili%C5%9Fim-sistemleri | /academic_program/information-system         | 0 | 0 | 0  | 0  | 3  |

The tuple with “PPP= (Entrances)” is the first entry the user has visited. Then the user visited “/portal” followed by the main home page, English version of “Information Systems” department web page, the Turkish version of the same page and revisited the English version again. *PageViews* are often 1 but here we have seen the tuple having 3 *PageViews* both in the second and fifth tuple. It may indicate that the page can be refreshed for 3 times or the same sequence of *PageViews* has been realized three times which does not need to be successively as in the case of page refreshment. To detect this, *NextPagePath* dimension needs to be added. By this way, for every visited page, the page visited before that page and after that page is reported.

### 3.5.7 Summary of What We Learnt on Google Analytics Dimensions and Metrics

We can analyze Table 20 which includes originally the dimensions: *Keyword*, *LandingPagePath*, *City*, *VisitorType*, *PageDepth*, *Source*, *NextPagePath*. To illustrate our case, we took a portion of this report as can be seen below. Table 20 is the report of a visit from Pittsburgh which searched “a tool to support personal software process”

keyword in Google search engine and landed to “/thesis/2001/improving-individual-software-engineering-skills-tool-support-personal-software-process” web page of our institute.

**Table 20: Example Complex Report**

| VT          | PD | NPP  | B | E | EX | PV | ToP | UPV | NV |
|-------------|----|--|---|---|----|----|-----|-----|----|
| New Visitor | 4  | /people/alpay-karagöz  | 0 | 0 | 1  | 2  | 0   | 1   | 0  |
| New Visitor | 4  | /thesis/2001/improving-individual-software-engineering-skills-tool-support-personal-software-process | 0 | 1 | 0  | 2  | 22  | 1   | 1  |

- 1) *NewVisits* is equal to zero although *VisitorType* is a new visitor because *NewVisits* takes value only for the pages having *Entrances* value more than zero.
- 2) All the values of the dimensions except *NextPagePath* are same for the two tuples which means that in this session, there were two different web pages labeled as *NextPagePath* by Google Analytics. Although there is a visit reported in the first tuple by a New Visitor with “a tool to support personal software process” keyword landed to the “/thesis/2001/improving-individual-software-engineering-skills-tool-support-personal-software-process” web page from Pittsburgh, it is reported that there is **no *Entrances* (*E value equal to zero*)** to this *LandingPagePath* because all landing pages do not need to be also entrance pages and all visits do not need to be entrance visits of a session. A session can be composed of more than one visits and a session can start with a search keyword, a direct visit or a referral visit. There can be more than one keyword search during a session. In this example, session starts with a search keyword because second tuple which is a visit with a keyword reports the entrance visit of the session since *Entrances* value is 1.

- 3) As you can see *PageDepth* dimension is reported as 4 for the two tuples which is equal to the sum of *PageViews* reported in the two tuples. Equal *PageDepth value* for the two tuples is a sign of their being part of the same session.

### **3.6 Matching Of What Visitors Possess And What The Web-Site Requires**

The Google Analytics reports help to identify the visitors' hardware, network properties and software preferences so that it is possible to couple them with the requirements of the web-site. We considered the following statistical information about the visitors from 01.01.2009 to 31.12.2009:

**Browser:** about 54% Internet Explorer, 36% Firefox, 7% Chrome

**Operating System:** 94% Windows, 4% Linux, 2% Macintosh

**Screen Resolution:** 29% 1280x800 , 23% 1280x1024 , 21% 1024x768 and 7% 1440x900 and left 20% are all above 1024x768, viewing content of our website now necessitates a minimum resolution of 1024x768 pixels which is proper for most of the visitors.

**Connection Speed:** About 95% of visitors used high-speed internet connections, such as cable, DSL or corporate networks, 5 % of visitors still used dial-up or other low speed connections.

### **3.7 Landing Page Optimization**

In this study, we applied landing page optimization using Google Analytics reports on METU Informatics Institute. For the site we considered [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr), *Bounce rate* for Search Engine Optimization (SEO) is in 2009 59.85%, the details of which can be seen in APPENDIX C. Nearly 60% *Bounce Rate* for SEO means that we need to examine our site's usability as Kumar (2009c) suggests. To be more focused, among all the programs under Informatics Institute, the four programs which are *Information Systems, Software Management, Informatics Online and Work Based Learning* websites were chosen to be improved using landing page optimization method. All TR and English web-site addresses of the selected programs together with the tagged addresses were firstly determined, the list of which you can reach in the APPENDIX D. The list is composed of 88 web-page addresses in total. The web-page of the remaining programs and departments can be the focus of further studies.

Following, a custom report was taken from the Google Analytics using NetBeans IDE 6.8 for which the dimensions were specified as *Keyword, LandingPagePath* and metrics were specified as *Bounces, Entrances, PageViews, UniquePageviews*. Filtering property was also used to form this custom report. The report was filtered in order to take only visits using a keyword that is traffic from Search engines and only the visits which were landed to the specified page addresses in APPENDIX D. The code used to get the report is given in APPENDIX E. Three rows from the report is given as an example in Table 21. Abbreviations listed in the beginning of this chapter in Table 2 is used in the coming tables.

**Table 21: Example from Custom Report for Landing Page Optimization**

| <b>K</b>             | <b>LPP</b>                           | <b>B</b> | <b>E</b> | <b>PV</b> | <b>UPV</b> |
|----------------------|--------------------------------------|----------|----------|-----------|------------|
| ion metu             | /academic_program/informatics-online | 24       | 37       | 67        | 51         |
| özden özcan top      | /academic_program/informatics-online | 23       | 45       | 69        | 62         |
| middlesex university | /category/tags/middlesex-university  | 21       | 34       | 59        | 46         |

In the previous studies such as in Clifton (2008a), to learn your visitor's expectations before arriving on your website, keywords are used. High *Bounce Rates* here (greater than 50 percent) are used as an indicator to determine the problematic keywords. However, in this study, we do not want to consider the keywords that are rarely used but always bounced; we want to find out the keywords that are frequently used and have high *Bounce Rate* values. Bounces give the number where that page is at the same time landing page and exit page and no other pages are visited. *Bounce Rate* is the ratio of *Bounces* over *Entrances*. Although it seems that *Bounce Rate* takes into account the usage frequency of the keywords, the reality is if a keyword is used in only one search throughout the selected time period and that visit bounced, the *Bounce Rate* for that keyword will be 100%. Since information driven web sites will have plenty of keywords used only once, there will be many keywords detected as problematic if only bounce rate is used. As we explained in 3.1, Metu Informatics Institute website has 14,614 keywords which are used in zero or 1 visit during the selected period. If we only consider the *Bounce Rate* to detect problematic keywords, we will then come up with infrequent keywords. So, high *Bounce Rate* does not always help to identify the problematic keywords. The frequency of the keyword is also need to be considered. To overcome this problem, a two step filtering method is used in this study.

In the first step, we sort the report according to the number of *Bounces* from the largest to the smallest. The *Bounces* higher than 1 are selected. Choosing the *Bounces* greater than 1 is totally related to the properties of our site and the time allocated to the study. As you can remember from Section 3.1, for the selected period only 39 of the keywords has frequency over 100. More than half of the keywords are used only once.

In the second step, we calculate the *Bounce Rate* for the selected tuples which is equal to the number of *Bounces* over the *Entrances* and we sort the tuples from the largest *Bounce Rate* to the lowest *Bounce Rate*. It appears that only 57 keywords have *Bounces* more than 1. Among them, 22 keywords have 100 % *Bounce Rate*. Among them to analyze, we select the first 20 of these 100% *Bounce Rate* keywords. These keywords according to their landing page group are given in Table 22.

Choosing *Bounce Rates* over 1 and then selecting the first 20 top Bounce Rate keywords among them is heuristically decided in this two step problematic keyword filtering method. To include more cases and so overcome the restrictions of this heuristically decided two step problematic keyword filtering method, we use a clustering method. We form Keyword clusters based on the top 20 problematic selected keywords. These keyword clusters need to be prepared manually by experts because the intended meanings by these keywords have to be properly understood before forming the clusters and string matching tools lack this property. We manually identify the keywords in our custom report which refer to the same meaning in English and Turkish with the selected top 20 problematic keywords given in Table 22. We make use of the following *heuristic* procedure: a keyword in our custom report is added to the cluster of each above listed problematic keyword cluster if:

- It is the translation of that keyword in English or Turkish.
- It is exactly the same word but including spelling errors.
- If we believe that the intended meanings of the keywords are same. By this, we mean not putting “master programs” and “master programs Ankara” in the same cluster because with “master programs Ankara”, the visitor searches “master programs” only in “Ankara”, it is a more restricted search however “master program” and “ms degree” can be put in the same cluster because their intended

meanings are the same, both of the visitors search for the programs having master degree.

**Table 22: Selected keywords to be analyzed**

|    | <b>Keywords (tags)</b>  | <b>Landing Web-Page Category<br/>Comprising The Keyword or Tag</b>     |
|----|---|--|
| 1  | uzaktan master  | <i>Informatics Online</i>  |
| 2  | master degree   |  |
| 3  | master program  |  |
| 4  | eğitim yüksek lisans  |  |
| 5  | odtü tezsiz yüksek lisans   |  |
| 6  | informatics online course   |  |
| 7  | <a href="http://www.ii.metu.edu.tr/academic_program/informatics-online">http://www.ii.metu.edu.tr/academic_program/informatics-online</a> |  |
| 8  | bilişim yüksek lisans ankara  |  |
| 9  | master ankara   |  |
| 10 | master programları ankara   |  |
| 11 | yuksek lisans ankara  |  |
| 12 | uzaktan eğitim veren üniversiteler  |  |
| 13 | sağlık bilişim sistemleri   | <i>Information Systems</i>   |
| 14 | health information systems phd program  |  |
| 15 | software management odtu  | <i>Software Management Program</i>                                     |
| 16 | yaşam temelli öğrenme   | <i>Work Based Learning</i>   |
| 17 | middlesex üniversitesi  |  |
| 18 | institute for work based learning   |  |
| 19 | information systems and programming metu  | <i>Informatics Online, Information Systems and Software Management</i> |
| 20 | online class metu   | <i>Informatics Online and Software Management</i>                      |

After forming the problematic keyword clusters, we go over exploring whether the clusters also have high *Bounce Rates* and *Bounces* in order to search for the reasons behind high *Bounce Rates* and *Bounces* for these clusters. We compare the Targeted Landing Pages (TLP) for these keyword clusters with the realized ones also considering whether the keyword is used in English or Turkish. We search the reasons of bounces by analyzing whether the landing pages include these keywords in the same intended meaning or not. Additionally, we apply the recommendation of Clifton (2008a). We check whether keywords are placed within the first 200 human-readable words (that is, not HTML code) or not. We will analyze the problematic keyword clusters under each landing web page category comprising the keyword or tag following the order given in Table 22. While analyzing the keywords in the following sections, we will use the abbreviations listed in the beginning of this chapter in Table 2.

### ***3.8 Problematic Keyword Clusters Which Landed on Informatics Online Program Web pages***

There are 12 problematic keywords which land on Informatics Online program web pages as shown in Table 22. In this part, we will analyze the clusters we formed based on these keywords. We will use the abbreviations listed in the beginning of this chapter in Table 2.

#### **3.8.1 Problematic Keyword Cluster 1**

The language of the keywords is all decided as “Turkish”.

**Table 23: “uzaktan master” Problematic Keyword Cluster Report**

| K                              | B | E | LPP  | TLP   | PV | UPV |
|--------------------------------|---|---|--|---|----|-----|
| uzaktan master                 | 5 | 5 | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans    | “tr/node/38”,<br>“academic_program/informatics-online”, “tr/academic_program/internet-%C3%BCzerinden-bili%C5%9Fim”and “node/38” | 5  | 5   |
| uzaktan master                 | 8 | 9 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans |   | 13 | 11  |
| master uzaktan                 | 0 | 1 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans |   | 2  | 2   |
| master uzaktan eğitim          | 1 | 1 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans |   | 1  | 1   |
| uzaktan eğitimle master        | 1 | 1 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans |   | 1  | 1   |
| uzaktan eğitim master          | 1 | 2 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans |   | 4  | 4   |
| uzaktan eğitim master          | 1 | 1 | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans    |   | 1  | 1   |
| uzaktan yüksek lisans          | 4 | 5 | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans    |   | 7  | 7   |
| uzaktan yüksek lisans          | 1 | 3 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans |   | 6  | 5   |
| uzaktan yüksek lisans          | 1 | 1 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans |   | 1  | 1   |
| uzaktan yüksek lisans          | 2 | 3 | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans    |   | 6  | 5   |
| uzaktan yüksek lisans          | 0 | 1 | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans    |   | 1  | 1   |
| uzak eğitim yüksek lisans      | 1 | 1 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans |   | 1  | 1   |
| uzaktan eğitim + yüksek lisans | 1 | 1 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans |   | 1  | 1   |
| uzaktan eğitim yüksek lisans   | 0 | 1 | /academic_program/informatics-online                     |   | 3  | 3   |

Table 23 (cont.)

| K                             | B  | E  | LPP   | TLP   | PV  | UPV |
|-------------------------------|----|----|---|---|-----|-----|
| uzaktan eğitim yüksek lisans  | 9  | 12 | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans     | "tr/node/38",<br>"academic_program/informatics-online",<br>"tr/academic_program/internet-%C3%BCzerinden-bili%C5%9Fim" and "node/38" | 19  | 17  |
| uzaktan eğitim yüksek lisans  | 0  | 1  | /tr/academic_program/internet-%C3%BCzerinden-bili%C5%9Fim |   | 6   | 4   |
| uzaktan eğitim yüksek lisans  | 14 | 27 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans  |   | 113 | 85  |
| uzaktan öğrenim yüksek lisans | 1  | 1  | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans     |   | 1   | 1   |
| uzaktan öğretim yüksek lisans | 1  | 1  | /academic_program/informatics-online                      |   | 1   | 1   |
| uzaktan öğretim yüksek lisans | 1  | 1  | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans  |   | 1   | 1   |
| yüksek lisans uzaktan eğitim  | 1  | 1  | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans     |   | 1   | 1   |
| yüksek lisans uzaktan eğitim  | 0  | 1  | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans     |   | 4   | 2   |
| yüksek lisans uzaktan eğitim  | 0  | 1  | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans  |   | 12  | 11  |
| yükseklisans uzaktan eğitim   | 1  | 1  | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans     |   | 1   | 1   |
| onilene master                | 1  | 1  | /tr/category/tags/online-master-program                   |   | 1   | 1   |
| online master                 | 6  | 13 | /tr/category/tags/online-master-program                   |   | 41  | 32  |
| online master                 | 1  | 1  | /tr/category/tags/online-master-program                   |   | 1   | 1   |
| onilene master                | 1  | 1  | /tr/category/tags/online-master-program                   |   | 1   | 1   |
| online master bilgi           | 0  | 1  | /tr/category/tags/online-masters-informatics              |   | 2   | 2   |
| internet master               | 1  | 1  | /category/tags/internet-%C3%BCzerinden-bili%C5%9Fim       | 1   | 1   |     |
| internet üzerinden master     | 2  | 3  | /category/tags/internet-%C3%BCzerinden-bili%C5%9Fim       | 5   | 4   |     |

Table 23 (cont.)

| K   | B  | E  | LPP   | TLP   | PV | UPV |
|---|----|----|---|---|----|-----|
| internet<br>üzerinden<br>master             | 1  | 1  | /tr/academic_program/internet-<br>%C3%BCzerinden-<br>bili%C5%9Fim | "tr/node/38",<br>"academic_program/informatics-<br>online", "<br>tr/academic_program/internet-<br>%C3%BCzerinden-<br>bili%C5%9Fim"and "node/38" | 1  | 1   |
| internet<br>üzerinden<br>yükseklisans       | 1  | 1  | /tr/node/38   |   | 1  | 1   |
| online<br>master<br>program                 | 0  | 1  | /tr/category/tags/online-master-<br>program                       |   | 6  | 6   |
| online<br>master<br>programları             | 10 | 13 | /tr/category/tags/online-master-<br>program                       |   | 30 | 22  |
| online<br>yüksek<br>lisans                  | 1  | 1  | /category/tags/tezsiz-<br>y%C3%BCksek-lisans-ankara               |   | 1  | 1   |
| online<br>yüksek<br>lisans                  | 1  | 1  | /category/tags/uzaktan-<br>e%C4%9Fitim-<br>y%C3%BCksek-lisans     |   | 1  | 1   |
| online<br>yüksek<br>lisans                  | 0  | 1  | /tr/category/tags/uzaktan-<br>e%C4%9Fitim-<br>y%C3%BCksek-lisans  |   | 2  | 1   |
| online<br>yüksek<br>lisans                  | 1  | 1  | /category/tags/uzaktan-<br>e%C4%9Fitim-<br>y%C3%BCksek-lisans     |   | 1  | 1   |
| online<br>yüksek<br>lisans                  | 1  | 1  | /category/tags/uzaktan-<br>e%C4%9Fitim-<br>y%C3%BCksek-lisans     |   | 1  | 1   |
| master<br>program on<br>internet            | 1  | 1  | /tr/category/tags/online-master-<br>program                       |   | 1  | 1   |
| masters<br>degree<br>online in<br>internet  | 1  | 1  | /category/tags/masters-degree-<br>online                          |   | 1  | 1   |
| online<br>master<br>programs                | 0  | 4  | <b>/tr/category/tags/online-<br/>master-program</b>               |   | 33 | 24  |
| distance<br>education<br>master<br>programı | 1  | 1  | /tr/category/tags/online-master-<br>program                       |   | 1  | 1   |

**Discussion:** Our first keyword to be analyzed is “uzaktan master”, which is used in 14 visits during the selected period. 132 Entrances in total have been realized by the keywords in this cluster. The total Bounce Rate is 65%. Bounce Rate for each landing page is given below:

- /academic\_program/informatics-online; 50%
- /tr/academic\_program/internet-%C3%BCzerinden-bili%C5%9Fim; 50.00%
- /category/tags/internet-%C3%BCzerinden-bili%C5%9Fim; 75%
- /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans; 79.41%
- /category/tags/masters-degree-online; 100%
- /category/tags/tezsiz-y%C3%BCksek-lisans-ankara; 100%
- /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans; 60%
- /tr/category/tags/online-master-program; 58.33%
- /tr/category/tags/online-masters-informatics; 0%
- /tr/node/38; 100%

All the landing pages in Turkish content includes “uzaktan” and “internet üzerinden” but *does not include “online” “and “distance” among the first 200 words.* The opposite situation stands for the landing pages served in English.

For this keyword cluster, although the more related content is served as landing pages in all of the cases, *Bounce Rate* is high; *this can be because of the visitor’s low interest for informatics online master program. Users may be looking for an online master program on a different topic. We could not identify any problem with the content of the web page.*

### 3.8.2 Problematic Keyword Cluster 2

Table 24: “master degree” Problematic Keyword Cluster Report

| L   | K                                | B | E | LPP  | TLP | PV | UPV |
|-----|----------------------------------|---|---|--|-----|----|-----|
| ENG | master degree                    | 3 | 3 | /category/tags/masters-degree-online                     | NA  | 3  | 3   |
| ENG | masters degree                   | 1 | 1 | /tr/category/tags/masters-degree-online                  |     | 1  | 1   |
| ENG | master's degree                  | 0 | 1 | /tr/category/tags/masters-degree-online                  |     | 6  | 3   |
| ENG | master digry                     | 1 | 1 | /tr/category/tags/masters-degree-online                  |     | 1  | 1   |
| ENG | ms degree                        | 1 | 1 | /tr/category/tags/masters-degree-online                  |     | 1  | 1   |
| TR  | Master                           | 0 | 1 | /tr/category/tags/masters-degree-online                  |     | 12 | 7   |
| TR  | yüksek lisans                    | 0 | 1 | /category/tags/tezsiz-y%C3%BCKsek-lisans-ankara          |     | 5  | 5   |
| TR  | yüksek lisans                    | 3 | 4 | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCKsek-lisans    |     | 6  | 5   |
| TR  | yukse lisans                     | 1 | 1 | /category/tags/tezsiz-y%C3%BCKsek-lisans-ankara          |     | 1  | 1   |
| ENG | master program                   | 2 | 2 | /tr/category/tags/online-master-program                  |     | 2  | 2   |
| TR  | yüksek lisans master programları | 1 | 1 | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCKsek-lisans    |     | 1  | 1   |
| TR  | eğitim yüksek lisans             | 0 | 1 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCKsek-lisans |     | 5  | 3   |
| TR  | eğitim yüksek lisans             | 2 | 2 | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCKsek-lisans |     | 2  | 2   |

**Discussion:** “master degree” is our second problematic keyword. While forming the keyword cluster, 3<sup>th</sup> and 4<sup>th</sup> problematic keywords which are respectively “master program” and “eğitim yüksek lisans” were added to “master degree” keyword cluster. These three keywords were used respectively in 3, 2 and 1 visits during the selected period. Other keywords added to this cluster from the custom report are “ms degree”, “master” and “yüksek lisans” as can be seen from Table 24. As the keywords are very generic, there is not a specific target web page which we want the users to land when they visit the web site with any of these keywords. The *Bounce Rate* for this cluster is

75%. There is no language problem for the keywords in this cluster, that is if the keyword searched is in English, served page is also in English, they are in harmony. *The problem stems from the content of the landing pages.* Visitors landing with the keywords in this cluster are looking for master programs available under Informatics Institute however *the provided content is information about Informatics Online program.* A new page needs to be designed to capture visitors using the keywords in this cluster. This new page needs to give information on all the available master programs under Informatics Institute. The main home page may also include these keywords.

### 3.8.3 Problematic Keyword Cluster 3

The language of the keywords is in “Turkish”.

**Table 25: “odtü tezsiz yüksek lisans” Problematic Keyword Cluster Report**

| K                         | B | E | LPP  | TLP | PV | UPV |
|---------------------------|---|---|--|-----|----|-----|
| odtü tezsiz yüksek lisans | 3 | 3 | /tr/category/tags/tezsiz-y%C3%BCksek-lisans-ankara | NA  | 3  | 3   |
| odtu tezsiz yüksek lisans | 1 | 1 | /tr/category/tags/tezsiz-y%C3%BCksek-lisans-ankara |     | 1  | 1   |
| tessiz master odtu        | 0 | 1 | /tr/category/tags/tezsiz-y%C3%BCksek-lisans-ankara |     | 3  | 2   |
| metu tezsiz yüksek lisans | 0 | 1 | /tr/category/tags/tezsiz-y%C3%BCksek-lisans-ankara |     | 11 | 11  |

**Discussion:** “odtü tezsiz yüksek lisans” is our fifth problematic keyword. This keyword was used in 4 visits during the selected period. “tessiz master odtu” and “metu tezsiz yüksek lisans” keywords are added to this cluster. There is no existing target web-page which includes this set of keywords. The total *Bounce Rate* for this cluster is 66.67%. The language of served web-page as landing page and the language of used keywords in this cluster are the same that is they are all in Turkish. Served landing page is a tag page as we described tag pages in Drupal in Section 3.1. In this tag page, although “yüksek

lisans” keyword is used among the first 200 words, “tezsiz” keyword is not used among the first 200 words and “master” keyword is not used in anywhere of the page. This can be the first reason of high *Bounce Rate* but the most important one is that the content of the served web page is not appropriate for these keywords. The content of the landing pages is information about Informatics Online program although there are other programs which have non-thesis master options. *A new page needs to be designed to serve for these visitors’ needs.*

### 3.8.4 Problematic Keyword Cluster 4

**Table 26: “informatics online course” Problematic Keyword Cluster Report**

| K                            | B | E | LPP                                      | TLP   | PV | UPV |
|------------------------------|---|---|--|---|----|-----|
| informatics online course    | 2 | 2 | /academic_program/informatics-online     | /academic_program/informatics-online-course | 2  | 2   |
| informatics online course    | 1 | 1 | /category/tags/informatics-online-course |   | 1  | 1   |
| online course in informatics | 1 | 1 | /category/tags/informatics-online-course |   | 1  | 1   |

**Discussion:** “informatics online course” is our sixth problematic keyword. This keyword is used in 3 visits during the selected period. From the custom report “online course in informatics” keyword is added to this cluster as can be seen in Table 26. The *Bounce Rate* for this cluster is 100% although the served landing page for this keyword is the same with the targeted pages. When we searched in Google, these keywords do not land on the reported pages here. The reason behind this high *Bounce Rate* can be the misleading report taken by Google Analytics as explained in Section 3.4.1. *This keyword needs to be deleted from the selected problematic keywords list.*

### 3.8.5 Problematic Keyword Cluster 5

The language of the keywords is all selected as “English”.

**Table 27: URL of the Informatics Online program Problematic Keyword Cluster Report**

| K  | B | E  | LPP                                  | TLP                                  | PV | UPV |
|--|---|----|--------------------------------------|--------------------------------------|----|-----|
| http://www.ii.metu.edu.tr/academic_program/informatics-online      | 8 | 22 | /academic_program/informatics-online | /academic_program/informatics-online | 87 | 58  |
| :<br>http://www.ii.metu.edu.tr/academic_program/informatics-online | 0 | 2  | /academic_program/informatics-online |                                      | 19 | 12  |

**Discussion:** The URL of the Informatics Online program is the problematic keyword of this cluster. This keyword is used in 26 visits during the selected period. The *Bounce Rate* for this cluster is 34% although the landing page is the targeted landing page. However, this *Bounce Rate* is not as high as the compared problematic keywords. Link of the informatics-online program for promotion purposes is put on the walls of many Facebook groups and the link does not work so people need to copy the link address and search for it.

### 3.8.6 Problematic Keyword Cluster 6

The language of the keywords is all selected as “Turkish”.

**Table 28: “bilgişim yüksek lisans ankara” Problematic Keyword Cluster**

| K  | B | E | LPP  | TLP   | PV | UPV |
|--|---|---|--|---|----|-----|
| bilgişim yüksek lisans ankara            | 0 | 1 | /academic_program/informatics-online               | "academic_program/information-systems" and<br>"/academic_program/informatics-online " | 5  | 5   |
| bilgişim yüksek lisans ankara            | 2 | 2 | /category/tags/tezsiz-y%C3%BCKsek-lisans-ankara    |   | 2  | 2   |
| bilgişim yüksek lisans ankara            | 0 | 3 | /tr/category/tags/tezsiz-y%C3%BCKsek-lisans-ankara |   | 19 | 16  |
| bilgişim yüksek lisans anakar            | 0 | 1 | /tr/category/tags/tezsiz-y%C3%BCKsek-lisans-ankara |   | 5  | 3   |
| ankara+bilgişim+yüksek lisans            | 0 | 1 | /tr/category/tags/tezsiz-y%C3%BCKsek-lisans-ankara |   | 8  | 7   |
| bilgişim sistemleri yüksek lisans ankara | 1 | 1 | /tr/category/tags/tezsiz-y%C3%BCKsek-lisans-ankara |   | 1  | 1   |

**Discussion:** “bilgişim yüksek lisans Ankara” is the problematic keyword of this cluster. It is used in 20 visits during the selected period. The *Bounce Rate* for this cluster is 33%.The problem stems from the content of the tag page and the name of the tag as “tezsiz yüksek lisans Ankara (non thesis master program Ankara)”. If the visitors are searching for a master programme with a thesis option, the header of the web page which is the tag “tezsiz yüksek lisans Ankara (non thesis master program Ankara)” will make them feel as they are in the wrong web-page. The content of the tag page is structured in three parts: first part is the header which is the name of the tag, second part gives information about Informatics Online program in Turkish and third part is the translation of second part in English. User needs to pull down the cursor to view the entire page. *The heading of the tag page needs to be revised or a new tag page needs to be created which will also include information about “Information Systems Program”.*

### 3.8.7 Problematic Keyword Cluster 7

The language of the keywords is all selected as “Turkish”.

**Table 29: “master Ankara” Problematic Keyword Cluster**

| K                         | B | E  | LPP  | TLP | PV | UPV |
|---------------------------|---|----|--|-----|----|-----|
| master ankara             | 2 | 2  | /category/tags/tezsiz-y%C3%BCKsek-lisans-ankara    | NA  | 2  | 2   |
| master ankara             | 1 | 2  | /tr/category/tags/tezsiz-y%C3%BCKsek-lisans-ankara |     | 4  | 3   |
| ankara master             | 1 | 1  | /category/tags/tezsiz-y%C3%BCKsek-lisans-ankara    |     | 1  | 1   |
| ankara master             | 0 | 1  | /tr/category/tags/tezsiz-y%C3%BCKsek-lisans-ankara |     | 2  | 2   |
| yüksek lisans ankara      | 3 | 5  | /category/tags/tezsiz-y%C3%BCKsek-lisans-ankara    |     | 13 | 13  |
| yüksek lisans ankara      | 9 | 11 | /tr/category/tags/tezsiz-y%C3%BCKsek-lisans-ankara |     | 47 | 33  |
| master programları ankara | 2 | 2  | /category/tags/tezsiz-y%C3%BCKsek-lisans-ankara    |     | 2  | 2   |
| yuksek lisans ankara      | 2 | 2  | /category/tags/tezsiz-y%C3%BCKsek-lisans-ankara    |     | 2  | 2   |
| yuksek lisasn ankara      | 1 | 1  | /category/tags/tezsiz-y%C3%BCKsek-lisans-ankara    |     | 1  | 1   |

**Discussion:** 10<sup>th</sup> problematic keyword “master programları ankara” and 11<sup>th</sup> problematic keyword “yüksek lisans ankara” are added to form the “master ankara” keyword cluster. “master programları ankara” keyword is used *twice*, “yüksek lisans ankara” is used *twice* and “master ankara” is used four times in all the visits during the selected period. The *Bounce Rate* for this cluster is 78% .The problem stems from these keywords’ being so generic and the content of the tag page. As can be seen in Table 29, visitors using this keyword cluster land to “tezsiz yüksek lisans Ankara (non thesis master program Ankara)” tag page. Although visitors intend to find information about a master program in Ankara, the served tag page includes aggregated information about Informatics

Online program in Turkish and its English version. ION is not the only master program under Informatics Institute so as for the 7<sup>th</sup> problematic keyword cluster; *a new page needs to be designed to capture visitors' interest using the keywords in this cluster. This new page needs to give information on all the available master programs under Informatics Institute and this page needs to include a link to the web page of METU where other master programs in METU are presented.*

### 3.8.8 Problematic Keyword Cluster 8

The language of the keywords is all selected as “Turkish”.

**Table 30: “uzaktan eğitim veren üniversiteler” Problematic Keyword Cluster**

| K                                      | B | E | LPP   | TLP                                  | PV | UPV |
|--|---|---|---|--------------------------------------|----|-----|
| uzaktan eğitim veren üniversiteler     | 1 | 3 | /academic_program/informatics-online                      | /academic_program/informatics-online | 6  | 4   |
| uzaktan eğitim veren üniversiteler     | 2 | 2 | /category/tags/uzaktan-eğitim-veren-üniversiteler         |                                      | 2  | 2   |
| uzaktan eğitim veren üniversiteler     | 0 | 1 | /tr/academic_program/internet-%C3%BCzerinden-bili%C5%9Fim |                                      | 5  | 5   |
| uzaktan eğitim veren üniversiteler     | 2 | 3 | /tr/category/tags/uzaktan-eğitim-veren-üniversiteler      |                                      | 6  | 6   |
| uzaktan eğitim veren üniversiteler     | 0 | 1 | /academic_program/informatics-online                      |                                      | 2  | 2   |
| internetten eğitim veren üniversiteler | 1 | 1 | /category/tags/uzaktan-eğitim-veren-üniversiteler         |                                      | 1  | 1   |

**Discussion:** “uzaktan eğitim veren üniversiteler” is the base keyword in this cluster. It is used in 281 visits during the selected period. The *Bounce Rate* for this cluster is 70 %. For the similar keywords listed in Table 30, three different web pages are served as landing pages. Although the keywords used are in Turkish, all the served landing pages except the third tuple are in English. The tag pages served are empty pages which includes only “There are currently no posts in this category.” text message as content. The served landing pages in the first, third and fifth tuple are the most proper web pages that can be served for this keyword cluster. However, since these web pages are designed to introduce Informatics Online program, they do not include the keyword “uzaktan eğitim veren üniversiteler”. *A new page needs to be designed to capture visitors’ interest using the keywords in this cluster. This new page needs to give information about METU, Informatics Institute and all online programs, or links to the web-pages of these programs.*

### ***3.9 Problematic Keyword Clusters Which Landed on Information Systems Program Web pages***

There are two problematic keyword clusters which landed on Information Systems program web page. These keyword clusters are based on respectively “sağlık bilişim sistemleri” and “health information systems phd program” as problematic keywords. We will use the abbreviations listed in the beginning of this chapter in Table 2 while analyzing these keyword clusters.

### 3.9.1 Problematic Keyword Cluster 1

The language of the keywords is all selected as “Turkish”.

**Table 31: “sağlık bilişim sistemleri” Problematic Keyword Cluster**

| K                           | B  | E  | LPP  | TLP  | PV | UPV |
|-----------------------------|----|----|--|--|----|-----|
| sağlık bilişim sistemleri   | 10 | 10 | /tr/academic_program/bili%C5%9Fim-sistemleri | tr/node/45;<br>departments/departme<br>nt-health-informatics | 10 | 10  |
| sağlıkta bilişim sistemleri | 1  | 1  | /tr/academic_program/bili%C5%9Fim-sistemleri |  | 1  | 1   |
| tıbbi sistemler ve bilişim  | 0  | 1  | /tr/category/tags/y%C3%BCksek-lisans         |  | 26 | 14  |
| tıbbi sistemler ve bilişim  | 0  | 1  | /tr/category/tags/y%C3%BCksek-lisans         |  | 3  | 3   |

**Discussion:** Problematic keyword to be analyzed is “sağlık bilişim sistemleri”, which was used in 29 visits during the selected period. From the custom report we find out that there is also the keyword “tıbbi sistemler ve bilişim” which is also used to search the same information. We added this keyword also to the this keyword cluster as there were no visitors using English translation of this keyword. By “sağlık bilişim sistemleri“ and different spelled versions that can be seen in the above table, a total of 11 visitors landed to the “/tr/academic\_program/bili%C5%9Fim-sistemleri” page which includes information about the Information Systems program and all of them bounced. When we analyze the landing page, among 200 human-readable words “sağlık bilişim sistemleri” is present. However, these words are used to explain the areas where information systems are used; *it is not giving the response what the visitors using these keywords are searching for*. Visitors who used “tıbbi sistemler ve bilişim” and different spelled versions of this keyword, landed to the “/tr/category/tags/y%C3%BCksek-lisans” page. From these visitors, two of them go on surfing although the first 200 hundred words in this page does not include these keywords. However if the visitor pulls the cursor

downwards, he/she can find information on what she/he searched. Another issue that needs to be touched upon is *although keywords used are Turkish, the served landing pages content are in English*. This can be another reason of the bounces. The *Bounce Rate* of this cluster is 85%. *The tag page needs to be revised to include search keywords in the first 200 words and the keywords in Information Systems program web page needs to be deleted or a link to the “tr/node/45” and “departments/department-health-informatics” web-pages need to be given.*

### 3.9.2 Problematic Keyword Cluster 2

The language of the keyword is selected as “Turkish”.

**Table 32: “health information systems phd programs” Problematic Keyword Cluster**

| <b>K</b>                               | <b>B</b> | <b>E</b> | <b>LPP</b>                            | <b>TLP</b>  | <b>PV</b> | <b>UPV</b> |
|--|----------|----------|---------------------------------------|---|-----------|------------|
| health information systems phd program | 2        | 2        | /academic_program/information-systems | “tr/node/45”, “departments/department-health-informatics” | 2         | 2          |

**Discussion:** This keyword was used *two times* in visits during the selected period. The *Bounce Rate* for this cluster is 100%. The served landing page for this keyword is “/academic\_program/information-systems”. The discrepancy between the content of the served and the targeted landing page is the reason behind this high *Bounce Rate*. The served landing page is the web-page of “information systems program” because it includes “healthcare information system” word among first 200 words. The targeted landing page does not include “health information systems phd program” among its content. In the target landing page, the wording is used as “health informatics” so *this*

keyword needs to be added and repeatedly used in the content of this targeted page and a link to this target web page needs to be put in Information Systems program web page.

### 3.10 Problematic Keyword Clusters Which Landed on Software Management Program Web-pages

In this part, there is only one problematic keyword which landed on Software Management Program web pages. We will use the abbreviations listed in the beginning of this chapter in Table 2 while analyzing this keyword cluster.

**Table 33: “software management odtu” Problematic Keyword Cluster**

| L   | K                          | B | E | LPP  | TLP   | PV  | UPV |
|-----|----------------------------|---|---|--|---|---|-----|
| TR  | software management odtu   | 0 | 1 | /tr/node/39  | "/tr/node/39",<br>"/tr/category/tags/software-management",<br>"/tr/academic_program/yaz%C4%B11%C4%B1m-y%C3%B6netimi", "/tr/category/tags/software-management" | 9   | 6   |
| TR  | software management odtü   | 2 | 2 | /tr/category/tags/software-management                |   | 2   | 2   |
| TR  | odtu software management   | 0 | 3 | /tr/category/tags/software-management                |   | 26  | 20  |
| TR  | odtü "software management" | 0 | 1 | /tr/node/39  |   | 3   | 3   |
| TR  | odtü software management   | 0 | 6 | /tr/category/tags/software-management                |   | 48  | 40  |
| TR  | odtü software management   | 0 | 1 | /tr/node/39  |   | 4   | 3   |
| TR  | yazılım yönetimi metu      | 0 | 0 | /tr/academic_program/yaz%C4%B11%C4%B1m-y%C3%B6netimi |   | 6   | 5   |
| TR  | yazılım yönetimi odtü      | 0 | 1 | /tr/academic_program/yaz%C4%B11%C4%B1m-y%C3%B6netimi |   | 1   | 1   |
| ENG | metu software management   | 1 | 8 | /tr/category/tags/software-management                |   | "/academic_program/software-management", "/node/39" | 62  |

**Table 33 (cont.)**

| L   | K  | B | E  | LPP                                   | TLP   | PV | UPV |
|-----|--|---|----|---------------------------------------|---|----|-----|
| ENG | middle east technical university "software management" | 0 | 1  | /academic_program/software-management | "/academic_program/software-management", "/node/39" | 2  | 2   |
| ENG | middle east technical university software management   | 0 | 1  | /academic_program/software-management |   | 7  | 5   |
| ENG | middle east technical university software management   | 1 | 1  | /tr/category/tags/software-management |   | 1  | 1   |
| ENG | software management metu                               | 0 | 4  | /academic_program/software-management |   | 57 | 35  |
| ENG | software management metu                               | 2 | 14 | /tr/category/tags/software-management |   | 61 | 50  |
| ENG | software management metu program                       | 0 | 1  | /academic_program/software-management |   | 3  | 2   |

**Discussion:** This problematic keyword cluster is based on “software management odtu” keyword. This keyword was used 6 times in visits during the selected period. Although *Bounce Rate* for the base keyword is 100%, the *Bounce Rate* for the cluster is 13%. This shows that clustering the keywords is necessary to be able to focus on the real problematic keywords. This keyword cluster does not necessitate special attention since 13% *Bounce Rate* is in the acceptable range according to Kumar (2009b).

### 3.11 Problematic Keyword Clusters Which Landed on Work Based Learning Program Web pages

There are 3 problematic keywords which landed on Work Based Learning program web page that we will analyze under this topic. These keywords are “yaşam temelli öğrenme”, “middlesex üniversitesi” and “institute for work based learning”. We will use the abbreviations listed in the beginning of this chapter in Table 2 while analyzing these keyword clusters.

#### 3.11.1 Problematic Keyword Cluster 1

The language of the keywords is all selected as “Turkish”.

**Table 34: “yaşam temelli öğrenme” Problematic Keyword Cluster**

| K                     | B | E | LPP   | TLP  | PV | UPV |
|-----------------------|---|---|---|--|----|-----|
| yaşam temelli öğrenme | 3 | 3 | /category/tags/i%C5%9F-ya%C5%9Fam%C4%B1-temelli-%C3%B6%C4%9Frenme | “tr/node/46” and “/tr/academic_program/work-based-learning/i%C5%9F-ya%C5%9Fam%C4%B1-temelli-%C3%B6%C4%9Frenme” | 3  | 3   |
| yaşam temelli öğrenme | 1 | 1 | /tr/node/46   |  | 1  | 1   |

**Discussion:** “yaşam temelli öğrenme” is the base problematic keyword of this cluster. This keyword was used 6 times in visits during the selected period. This cluster consists of only this keyword as you can see in Table 34. The *Bounce Rate* for this cluster is 100%. There is no problem with the language of served page as the keyword is in Turkish and the served pages content is also in Turkish. The problem is with the tag

page served. The content of this page starts with an announcement about another department. The first 200 words do not include the keyword; visitor needs to pull the cursor down to see the related content with his/her search. The second landing page “/tr/node/46” includes the keyword in the first sentence and the content is directly related with the keyword so the bounces may be because the visitor obtained all the information s/he asked for on the page .

### 3.11.2 Problematic Keyword Cluster 2

Table 35: “middlesex üniversitesi” Problematic Keyword Cluster Report

| L   | K                      | B  | E  | LPP                                    | TLP   | PV | UPV |
|-----|------------------------|----|----|--|---|----|-----|
| TR  | middlesex üniversitesi | 3  | 3  | /category/tags/middlesex-university    | “tr/node/46” and<br>“/tr/academic_program/work-based-learning/i%C5%9F-ya%C5%9Fam%C4%B1-temelli-%C3%B6%C4%9Frenme” | 3  | 3   |
| TR  | middlesex universitesi | 1  | 1  | /category/tags/middlesex-university    |   | 1  | 1   |
| TR  | middle sex universtesi | 1  | 1  | /category/tags/middlesex-university    |   | 1  | 1   |
| TR  | middlesex              | 4  | 6  | /category/tags/middlesex-university    |   | 10 | 7   |
| TR  | middlesex,             | 1  | 1  | /category/tags/middlesex-university    |   | 1  | 1   |
| TR  | middle sex             | 1  | 1  | /category/tags/middlesex-university    |   | 1  | 1   |
| ENG | middlesex university   | 1  | 1  | /category/tags/middlesex-university    |   | 1  | 1   |
| ENG | middlesex.university   | 1  | 2  | /tr/category/tags/middlesex-university |   | 2  | 2   |
| ENG | middle sex university  | 1  | 1  | /category/tags/middlesex-university    |   | 1  | 1   |
| ENG | middlesex university   | 21 | 34 | /category/tags/middlesex-university    |   | 59 | 46  |
| ENG | middlesex university   | 3  | 4  | /tr/category/tags/middlesex-university |   | 8  | 6   |
| ENG | middlesex universty    | 1  | 1  | /category/tags/middlesex-university    |   | 1  | 1   |
| ENG | middlesex.university   | 0  | 2  | /tr/category/tags/middlesex-university |   | 5  | 2   |
| ENG | mdlesex university     | 1  | 1  | /category/tags/middlesex-university    |   | 1  | 1   |

**Discussion:** “middlesex üniversitesi” is the base problematic keyword of this cluster. This keyword was used 3 times in visits during the selected period. This cluster consists of only this keyword and its translation into English as can be seen in Table 35. The *Bounce Rate* for this cluster is 67.80%. *No matter the language of the keyword, the served landing pages content is in Turkish.* This can be one reason of the high *Bounce Rate*. Served landing pages are tag pages which have the same content. In these tag pages, “middlesex” keyword is used among the first 200 words but the most important point which is “*Work Based Learning Programme’s being a joint programme between METU and Middlesex University leading to a dual diploma given by these universities*” is not stated among the first 200 keywords. This can be the second reason of high *Bounce Rate*. Third reason can be the keywords’ being so generic.

### 3.11.3 Problematic Keyword Cluster 3

The language of the keywords is all selected as “English”.

**Table 36: “institute for work based learning” Problematic Keyword Cluster**

| K  | B | E | LPP   | TLP   | PV | UPV |
|--|---|---|---|---|----|-----|
| institute for work based learning        | 1 | 1 | /departments/department-work-based-learning-studies | "/departments/department-work-based-learning-studies";"/tr/node/46" | 1  | 1   |
| institute of workbased learning          | 1 | 1 | /departments/department-work-based-learning-studies |   | 1  | 1   |
| institute work based learning            | 2 | 2 | /departments/department-work-based-learning-studies |   | 2  | 2   |
| intitute of work bsased learning         | 0 | 1 | /departments/department-work-based-learning-studies |   | 3  | 2   |
| iş yaşamı temelli öğrenme ana bilim dalı | 1 | 1 | /tr/node/46   |   | 1  | 1   |

**Discussion:** This problematic keyword cluster is based on “institute for work based learning” keyword. This keyword was used 3 times in visits during the selected period. The *Bounce Rate* for this cluster is 83%. *Bounce Rate* is high although the landing pages are proper for the keywords. The reason can be the visitors finding the information they are looking for on the landing page. Low number of page views supports our theory. *There is no problem with the keyword.* To increase the visitor satisfaction, a link to the web page of Work Based Learning Institute at Middlesex University can be added to the content.

**3.12 Problematic Keyword Clusters Which Landed on the Web pages related to Informatics Online, Information Systems and Software Management Programs**

There is only one problematic keyword which landed on the web pages related to Informatics Online, Information Systems and Software Mangement programs. The language of the keyword is selected as “English”. We will use the abbreviations listed in the beginning of this chapter in Table 2 while analyzing this keyword cluster.

**Table 37: “information systems and programming metu” Problematic Keyword Cluster**

| K  | B | E | LPP                                | TLP  | PV | UPV |
|--|---|---|------------------------------------|--|----|-----|
| information systems and programming metu | 2 | 2 | /category/tags/information-systems | /category/tags/information-systems and /academic_program/information-systems | 2  | 2   |

**Discussion:** This keyword was used *4 times* in visits during the selected period. The *Bounce Rate* for this cluster is 100 % although the landing page is the Targeted Landing Page. The problem stems from the content of the tag page. It starts with information about “MIN 528 Fundamentals Mathematics for Information Systems “course which is given by Medical Informatics program and also which is totally irrelevant with the keyword. The tag page includes information about Information Systems program but the user needs to pull down the cursor. The tag page needs to be revised to lower the *Bounce Rate* for this keyword.

### ***3.13 Problematic Keyword Clusters Which Landed on Web pages of both Informatics Online and Software Management Programs***

There is one problematic keyword which landed on the web pages of both Informatics Online and Software Management programs. In this part, we will analyze this keyword cluster which is “online class metu”. The language of the keywords is selected as “English”. We will use the abbreviations listed in the beginning of this chapter in Table 2 while analyzing this keyword cluster.

**Table 38: “online class metu” Problematic Keyword Cluster**

| <b>K</b>            | <b>B</b> | <b>E</b> | <b>LPP</b>                                  | <b>TLP</b>                  | <b>PV</b> | <b>UPV</b> |
|---------------------|----------|----------|---|-----------------------------|-----------|------------|
| online class metu   | 2        | 2        | /tr/category/tags/informatics-online-course | https://online.metu.edu.tr/ | 2         | 2          |
| metu online classes | 1        | 1        | /tr/category/tags/informatics-online-course |                             | 1         | 1          |
| netclass odtu       | 1        | 1        | /academic_program/software-management       |                             | 1         | 1          |

*Discussion:* This problematic keyword cluster is based on “online class metu” keyword. This keyword was used *3 times* in visits during the selected period. *Bounce Rate* for this keyword cluster is 100%. The two different served landing pages which can be seen in Table 38 are not related with the keyword cluster. The landing page in the first two tuples does not have content, it warns the visitor by “There are currently no posts in this category.” message. The landing page in the third tuple informs the visitors about the Software Management program, it includes only “online” keyword but not in the first 200 human readable words. When we made a new Google search with this keyword, none of the landing pages listed exists in the results. From this, we conclude that keyword-landing page match was reported wrongly by Google Analytics for this keyword cluster. Since the first result of Google Analytics is the Targeted Landing Page, *there is no problem with this keyword cluster.*

### **3.14 Summary Of The Results**

Landing page optimization method we applied from Section 3.8 to 3.13 helped us provide recommendations about the information architecture of our case study web site: metu informatics institute.

During our analysis above, we discover that the success of landing pages differ according to their type whether they are tagged pages or normal pages. To test this, we distinguish tagged landing pages from the normal ones. We have found that 103 of the total 137 landing pages are tagged pages, the left 34 landing pages are normal pages. *Bounce Rate* for the tagged pages is 63% whereas *Bounce Rate* for the normal pages is 50% which shows that tag pages are lowering the success of our case study web site. There needs to be a rearrangement for these tagged pages as we recommended above for each keyword cluster.

All the web pages about informatics online program needs to include “uzaktan”, “internet üzerinden”, “online” and “distance” keywords among the first 200 words of their content to lower the bounce rate of these keywords.

To solve the problems identified in Section 3.8.2, 3.8.3, 3.8.7 and 3.8.8, a new page which gives short description of all available master programs under Informatics Institute and includes links to each program web page needs to be designed both in English and Turkish. The master programs in this page need to be categorized as thesis and nonthesis. A second category needs to be online programs, normal programs and evening education programs. The new designed page needs to include “yüksek lisans”, “ms degree”, “ master degree”, “yüksek lisans master programları”, “eğitim yüksek lisans”, “odtu tezsiz yüksek lisans”, “tessiz master odu”, “metu tezsiz yüksek lisans”, “master ankara”, “yükseklisans ankara”, “master programları ankara”, “uzaktan eğitim veren üniversiteler” and “internetten eğitim veren üniversiteler” keywords.

“bilşim yüksek lisans Ankara” keyword needs to be added by the web site designers as a metatag to each related program web page of informatics institute to solve the problem in Section 3.8.6.

To solve the problem discussed in Section 3.9.1 and 3.9.2, the keywords “sağlık bilşim sistemleri”, “tıbbi sistemler ve bilşim”, “health information systems phd program” present in information systems program web page needs to have a link to the health informatics program and the web page of health informatics program also needs to include these keywords.

The problems identified under Section 3.11.1 and 3.11.2 can be solved by deleting metatags “iş yaşamı temelli öğrenme” and “Middlesex university” because the tagged page content is not proper for these keywords. If these metatags are deleted by the website designers from the institute web site, Work based learning program web page will be the only served landing page to the visitors coming with these keywords. This

web page includes both of these keywords and its content is proper to the needs of these visitors. Work based learning program web page includes information about its being a joint programme between METU and Middlesex University leading to a dual diploma given by these universities which will grab the attention of visitors. Since the program web page is prepared both in English and Turkish, the language of the served page problem will be also solved.

To solve the problem identified in Section 3.12, metatag “information systems” needs to be deleted by the web site designers because the tagged web page’s content is not proper. By this way, information systems program web page will be the only page served to the visitors using “information systems and programming metu” keyword. Information systems program web page is adequate enough to fill the information needs of these visitors.

## **CHAPTER 4**

### **METHODOLOGY: THINK-ALOUD STUDY**

We applied *think-aloud* study which was suggested by the literature explained in Section 2.1. In this chapter, we will describe the methodology we applied during think-aloud study. We will explain how we prepared the Questionnaire, chose the participants, the materials we used, procedures we applied. We will finalize this chapter after explaining the reason behind our sample size, validity test results and results of the think-aloud study.

#### ***4.1 How We Prepared the Questionnaire?***

As Atkinson (2007), Avoris et al. (2003), Gwizdka & Spence (2007), Brinck et al. (2003) used in their studies, we make use of tasks to evaluate our web page. And we benefit from their questions to prepare our ones. Tasks used for testing need to be selected that are critical to users and according to the goals of the web site in question (Brinck et al., 2003). The information sought by these tasks are intended to be representative of the set of common possible information questions that potential and current students of METU Informatics Institute might ordinarily ask when visiting this site as Gwizdka & Spence (2007) did in their studies. A total of nine tasks are formed

considering the task categorization done by Kellar, Watters, & Shepherd (2007). The first three tasks are aiming “information gathering” and the rest are aiming “fact finding”. To ensure comparable data, these tasks are created for the same departmental web pages we analyzed with Google Analytics. After each task, the perceived hardness level of that task is evaluated by the participant according to 5 point Likert scale and lastly each task is finalized with an open ended question “If it was hard to find, why? Where did you think the information should be?” to collect the reasoning and problem solving methodology of each participant. Other than tasks, our questionnaire includes demographic questions, short assessment questions and open ended questions. We prepare short assessment questions benefiting from the Rubinoff (2004), Skadberg & Kimmel (2004) and Compeau & Higgins (1995) and Gullikson et al. (1999) studies. Short assessment questions are categorized as Branding, Functionality, Contents and Ease of Use, Challenge and Skill, Self-Efficacy and overall evaluation. These categories are a combination of the ones specified in the referenced studies. In our short assessment questions, we use a 5 point Likert Scale except three of overall evaluation questions. You can see in Table 39 the type of each question, category, rank in the webform and source if exists; in the questionnaire we prepared for think-aloud study.

**Table 39: Think-Aloud Study Questions Category and Source**

| Type       | Category              | Rank | Question or Task                                       | Source                   |
|------------|-----------------------|------|--|--------------------------|
| Demografik | Manually collected    | 0    | Sex  | (Gwizdka & Spence, 2007) |
|            |                       | 0    | Age  | (Gwizdka & Spence, 2007) |
|            |                       | 0    | Average Internet Use in Years                          | (Gwizdka & Spence, 2007) |
|            |                       | 0    | Average daily Internet Use in hrs                      | (Gwizdka & Spence, 2007) |
|            | Collected via webform | 1    | What is your purpose for visiting this page?           | -                        |
|            |                       | 2    | What is your Last degree (obtained or in preparation)? | -                        |

Table 39 (cont.)

| Type           | Category                 | Rank | Question or Task   | Source |
|----------------|--------------------------|------|--|--------|
| Demogra<br>fik | Collected via<br>webform | 3    | What is your Faculty and Department that you study at? Please choose from the list.  | -      |
|                |                          | 4    | Please select the option/options below that best suits for you. Language of education:   | -      |
| Tasks          | Information Gathering    | 1    | Imagine that you decide to do a doctorate study. Among the programs <i>Information Systems IS</i> , <i>Software Management SM</i> and <i>Informatics-OnLine ION</i> , which one/ones offer/s doctorate study?  |        |
|                |                          | 2    | Imagine that you look for a NON-thesis master programme. Assume that you live and work in Muğla. You were graduated from economics department. You DO NOT have any experience in software industry. Which of the programs among Information Systems IS, Software Mangement SM and Informatics-OnLine ION are you eligible for application?         | -      |
|                |                          | 3    | Which of the programs among Information Systems IS, Software Mangement SM and Informatics-OnLine ION are evening education programs?   | -      |
|                | Fact Finding             | 4    | Imagine that you were graduated from Electrical and Electronics Engineering Department but you are working as a software programmer in a company for two years. You are interested in Software Management (SM) programme. Now, please attempt to find whether you are eligible for this program or not by looking at our application requirements. | -      |
|                |                          | 5    | Imagine that you are accepted for a program at Informatics Institute and you are now at the course registration period. You need to choose the courses that you will take this semester.Task 5: Now please attempt to find the elective courses of Work Based Learning WBL Programme. Choose the courses from the list below                       | -      |
|                |                          | 6    | Now please attempt to find the courses that were opened in 2008-2009 Fall Semester   | -      |
|                |                          | 7    | Imagine you are accepted for a program which includes a Thesis Study at Informatics Institute. You decided on your thesis subject.You are supposed to have an advisor. Please attempt to find the faculty member who participates into two research groups about Software Technologies and Data Mining.  | -      |
|                |                          | 8    | Imagine you are about to finish your master studies this semester at Informatics Institute. You need information on the procedure of Thesis Study.Please attempt to find information on what you need to do before your thesis defense.  | -      |
|                |                          | 9    | Please attempt to find information on the maximum time limit you can use to do the minor corrections on your thesis study pointed out by your Jury and Informatics Institute.  | -      |

**Table 39 (cont.)**

| Type   | Category              | Rank                               | Question or Task  | Source   |
|--|-----------------------|------------------------------------|---|--|
| Short Assessment Questions- 5 Point Likert Style | Branding              | 1                                  | The visual impact of the site is consistent with the brand identity       | (Rubinoff, 2004)   |
|  |                       | 2                                  | Graphics, Collaterals and Multimedia add value to the experience.         | (Rubinoff, 2004)   |
|  |                       | 3                                  | The website does NOT deliver on the perceived promise of the brand-METU   | (Rubinoff, 2004)   |
|  | Functionality         | 4                                  | The Web site speed is fast.   | (Skadberg & Kimmel, 2004) Appendix A. Online survey instrument Que2  |
|  |                       | 5                                  | There is little waiting time for the Web pages to load.                   | (Skadberg & Kimmel, 2004) Appendix A. Online survey instrument Que3  |
|  |                       | 6                                  | Users receive timely responses to their queries / submissions.            | (Rubinoff, 2004)   |
|  | Content & Ease of Use | 8                                  | Navigation of the Web site was simple and easy                            | (Skadberg & Kimmel, 2004) Appendix A. Online survey instrument Que7  |
|  |                       | 9                                  | Link density provides clarity and easy navigation                         | (Rubinoff, 2004)   |
|  |                       | 10                                 | Interacting with the Web site was NOT easy.                               | (Skadberg & Kimmel, 2004) Appendix A. Online survey instrument Que12 |
|  |                       | 11                                 | The website provides visitors with an engaging and memorable experience   | (Rubinoff, 2004)   |
|  |                       | 12                                 | I did NOT feel comfortable using this web site for these tasks            | (Gullikson et al., 1999) Appendix II Perceptions Test Que1           |
|  |                       | 13                                 | The website helps its visitors accomplish common goals and tasks          | (Rubinoff, 2004)   |
|  |                       | 14                                 | Content structured in a way that facilitates the attainment of user goals | (Rubinoff, 2004)   |
|  |                       | 15                                 | I felt that I had the freedom to go anywhere in the Web site              | (Skadberg & Kimmel, 2004) Appendix A. Online survey instrument Que11 |
|  |                       | 16                                 | It was NOT easy to find the required information on the web site.         | (Gullikson et al., 1999) Appendix II Perceptions Test Que2           |
|  |                       | 17                                 | The organisation of information on the web site was clear to me.          | (Gullikson et al., 1999) Appendix II Perceptions Test Que3           |
|  | 18                    | Content is up-to-date and accurate | (Rubinoff, 2004)  |  |
|  | Challenge and Skill   | 7                                  | How often have you visited Informatics Institute Website?                 | (Skadberg & Kimmel, 2004) Appendix A. Online survey instrument Que22 |
|  |                       | 19                                 | Content is appropriate to customer needs and business goals               | (Rubinoff, 2004)   |
|  |                       | 20                                 | Content across multiple languages is NOT comprehensive                    | (Rubinoff, 2004)   |
|  |                       | 21                                 | The website prevents errors and helps the user recover from them          | (Rubinoff, 2004)   |
|  |                       | 22                                 | How often did you use "Search" service available on the web-site?         | -  |
|  |                       | 23                                 | How useful was the "Search" service available on the web-site?            | -  |

**Table 39 (cont.)**

| Type   | Category           | Rank | Question or Task  | Source  |
|--|--------------------|------|---|---|
| Short Assessment Questions- 5 Point Likert Style | Self-Efficacy      | 24   | I could perform better using this website, If I had navigated similar web pages before this page. | (Compeau & Higgins, 1995) Appendix Computer Self-Efficacy Measure Que10 |
|  |                    | 25   | I could perform better using this website, If someone else had helped me get started              | (Compeau & Higgins, 1995) Appendix Computer Self-Efficacy Measure Que6  |
|  |                    | 26   | I could perform better using this website, If I had a lot of time to accomplish the tasks.        | (Compeau & Higgins, 1995) Appendix Computer Self-Efficacy Measure Que7  |
|  |                    | 27   | I could perform better using this website, If my English was better                               | -   |
|  |                    | 28   | Please rate your computer literacy level  | -   |
| Open-Ended Questions                             | Overall Evaluation | 29   | Overall I was satisfied with this web site.   | (Gullikson et al., 1999) Appendix II Perceptions Test Que4              |
|  |                    | 30   | What did you like most about our Institute web site?  | (Gullikson et al., 1999) Appendix II Perceptions Test Que5              |
|  |                    | 31   | What did you like least about our Institute web site ?  | (Gullikson et al., 1999) Appendix II Perceptions Test Que6              |
|  |                    | 32   | Other Comments  | (Gullikson et al., 1999) Appendix II Perceptions Test Que7              |

Firstly, we plan to use this questionnaire to make an on-line survey which we will disseminate from the institute web site and as a second step apply the same questionnaire using think-aloud technique.

Since studies conducted in Middle East Technical University (METU) and/or studies conducted by METU personnel/students, which involve collecting data from participants, are subject to review by the METU Human Subjects Ethics Committee (HSEC), for both type of the studies: on-line survey and think-aloud study, we applied for METU Human Subjects Ethics Committee and first took their approval. This approval can be seen in APPENDIX A.

We prepare the on-line survey, the content of which is the Questionnaire listed above except the manually collected category demographic questions, using the webform

option of the content management system Drupal. Each task is put in a separate web-page to be able to track and measure the time taken for each task. The on-line applied version of the survey is attached in the APPENDIX F. Our applied think-aloud study is composed of 8 demographic questions, 9 tasks, 29 short assessment questions and 3 open ended questions.

While preparing the on-line version, only the demographic questions are set as mandatory, the other questions are set as optional. We set the other questions as optional to prevent so many noise in the results and so a long preprocessing step while analysing these results.

#### **4.1.1 Promotion activities of the on-line Survey:**

Promotion activities are based on two channels: e-mail and social networking website facebook. The link of the survey is sent via e-mail to:

- all of the friends who know English.
- all the students of the Informatics Institute.

On the facebook side, on the wall of the groups, the member numbers of which are given in the parenthesis at the time of promotion which are: Metu (7375), Odtü (6159 ), IAL (1896), BİLİŞİM GÜVENLİĞİ [İZMİR] ( 1294), Türkiye Bilgisayar Mühendisleri ve Programlamacılar Derneği (399), BiLiŞiM TEKNOLOJİLERİ(590), T.A.B.A. (Turkish American Business Academy)-www.tabaturkey.com(367), Bilişim Destek(551), Bilişim Dergi (270), Bilişim Platformu(313), CHIP, Bilgisayar, Genç Liderler Akademisi, Istanbul Proje Yönetim Derneği ( IPYD )(263), TBD ( Türkiye Bilişim Derneği )(666), Bilişim İK(198), KalDer(369), Web Analytics(687), the following message is posted.

“I am studying on my M.S. thesis study. My hypothesis is:

Web analytic tools can be used instead of/ with classical tools like surveys and think-aloud to test the success of information architecture of web sites which are prepared for information disposal.

I prepared an on-line survey for this purpose to collect data. It is composed of 9 tasks and 32 short assessment questions. It has a different style than normal surveys. You need to find the answer of the questions within the web site of METU Informatics Institute web-site. Can you please help my academic study by participating in this survey? This survey is ethically approved by my university.

<http://www.ii.metu.edu.tr/content/questionnaire-testing-informatics-institute-web-site>”

Although by the help of facebook groups the post above reached in total thousands of people, the results are not satisfying, we cannot reach even 50 people in 2 months and the submitted surveys are mostly empty and wrongly understood. From the comments of my friends who participated in this on-line survey, we realize that tasks in this survey are misunderstood as users often skipped reading the explanation part of the survey on the first page. Instead of looking for the answers inside the institute web-site, people tried to answer these questions from their own knowledge. We realize that people are not used to apply tasks in this type of on-line surveys. Because of these reasons, we decide to use only the results of think-aloud study.

## ***4.2 Participants***

32 adults (18 females and 14 males) have participated in this study. Since the programs offered under Informatics Institute are interdisciplinary, having participants from different undergraduate faculty backgrounds are an important concern for us. We manage this by collecting participants from 8 different faculties where they are studying or have studied during undergraduate study. Average age of this participant group is about 26. For 12 of the participants, “undergraduate” is the last degree obtained or in

preparation, for 10 of them “Master of Science” and for 10 of them “doctorate”. Since a plenty of current students at Informatics Institute are both working and also studying at the Institute, to have a more reliable sample we include 7 participants among the working people. Average internet use experience of this participant group is 8.9 years. Average daily internet use for this group is 4.7 hours. The website was not familiar for the participants because the content of the web-site and the visual design was recently changed at the time of think-aloud study which helps us getting rid of familiarity issue of usability studies. 23 of the participants have never visited the Institute web-site in the past, 7 of them have visited once in every several months and 2 of them have visited more than once a week. Participation was voluntarily and no incentives were given for participation.

### **4.3 *Materials***

Participant sessions were captured using SONY HDV Handycam Digital HD Video Camera Recorder HDR-FX1E model camera which enabled us collecting more information by asking more questions about why the participant believed the task was hard or easy, how the problem can be solved from her/his point of view or why the information cannot be found instead of concentrating on taking notes of what the participants says. Using camera recording also helped recovering unexpected electricity cuts as we experienced with one of the participants. A second before submitting the webform, electricity has been cut and we then fill the webform according to camera recording and submit on behalf of the participant.

#### **4.4 Procedures**

The study was conducted from 25.02.2010 to 21.03.2010 in the Vision Lab of the Informatics Institute. Participants accessed the Questionnaire prepared for the think-aloud study using Internet Explorer loaded on a Pentium 4 PC with a 17 inch monitor using Windows XP with good connection to the institute web-site through the high band university network. The PC and the room used during the study have never been changed to remove the effects of environment. Think-aloud study was realized with each participant one by one and with the same moderator.

Institute web-site which includes a link to the Questionnaire was opened by us and then clicked to open the Questionnaire. Questionnaire was opened with a tracking number which can be seen in the right part of the same line with “Yes”, “No” buttons. As can be seen in the APPENDIX B, the questionnaire starts with the Voluntarily Participation Information which includes all the information required by METU Human Subjects Ethics Committee and the warnings of us which we experienced in the on-line survey as common misunderstandings. The most striking parts of this page was written in red font and also the Turkish translation of these red sentences were given at the end of each sentence. This “Voluntarily Participation Information” was read to the participant. Different from the studies realized by Gwizdka (2007) and Gullikson (1999), participants were free to use “sitemap” , “in-site search” service and “google” search engine because the users of the web-site may use all of these services while searching for an information in real life.

Additionally, think-aloud method and the expectations from the participant was explained as:

“Please use only this single browser window for navigating. We want you to feel like you are at home searching for something on the web. Feel like you are alone in this room and loudly tell what you are thinking during searching for the information the task requires, we want also from you to tell us where did you look at the information, where do you think it should be incorporated. Please, tell your ad hoc

comments loudly. For the tasks, you have 3.20 minutes time restriction and feel relaxed with the time restriction because researches showed that 3 minutes is sufficient to find an information from an information driven web site. People got bored before 3 minutes if they cannot find and we are giving you 20 seconds more.”

Gullikson (1999) proved when the answer was humanly “findable”, it was locatable in much less time than 3 minutes and we were giving in this study 20 seconds grace as Gwizdka (2007) used in time-limit web-navigation study.

Lastly, consent of the participant is asked. If the participant agrees on the participation, he/she clicks on the “Yes” button to go further to start the study, the tracking code was noted by us. Before answering the demographic questions included in the Questionnaire, we asked the demographic questions of which answers were manually collected. These were: age, sex, average internet use in years, average daily internet use. Except the cases participant needs explanations, demographic questions and short-assessment questions were read by the participant and replied on his/her own. If participant requests to translate the question to Turkish, that kind of request is fulfilled. Especially, for the participants who are doing their undergraduate studies, the meaning of Master of Science and doctorate studies were explained. In the task part, we read and explained the task whilst the participant could read from the webform and we checked for the time limit for each task. For the tasks part, participants could see only one task at the same time. If the participant could not find the answer within the time limit, we went on the next step which is participant answering the hardness level and the comments on not being able to find the answer. We proceeded with the next task.

#### ***4.5 Sample Size***

Think-aloud studies are more costly than surveys because the moderator applies the study one by one, at the same location and each study is much more time consuming

than the surveys. Generally these studies are done with less than 20 participants. Atkinson (2007) used 8 participants, Brinck et al. (2003) used 24 participants, Gullikson et al. (1999) similarly used 24 participants but his study was a bit different than the standard think-aloud methodology, after participants finished tasks and the perceptions test, participants of the study verbally explained the approaches taken in responding to the questions while the previous session is replayed from the Microsoft Camcorder.

Statistical power analysis exploits the relationships among the four variables involved in statistical inference: sample size ( $N$ ), significance criterion ( $\alpha$ ), population effect size (ES) and statistical power (Cohen, 1992a; Cohen, 1992b). According to him, research methodologists agree about the desirability of power analysis in research planning and assessment. However, progress in application of this method over the last quarter century has been slow. Sedlmeier and Gigerenzer's study cited at (Cohen, 1992a) is based on the review of 54 articles in terms of power analysis. In 7 of the articles they found out that the authors of these articles mistakenly concluded accepting the null hypothesis because power analysis was not done in these articles. If these authors had done power analysis, they would have seen that their chance of rejecting their null hypothesis in the presence of substantial population effects was only 25%. Not to be among these mistakenly concluded hypotheses, during the decision of sample size, we made use of statistical power analysis and specifically the table prepared by Cohen (1992a) and named as "N for Small, Medium, and Large ES at Power=0.80 for  $\alpha=0.01, 0.05$  and  $0.10$ " which is a short rule-of-thumb treatment. This table states that if multiple regression/correlation analysis will be done and the significance tests will be applied at  $\alpha=0.05$  level and the researcher expects a large EF, candidate independent variables are expected to be 2, the researcher needs a minimum of 30 participants to reach 80% statistical power.

We decided to use more than 30 participants and could reach 32 participants in our study which is the highest number compared to the participant number in think-aloud studies we took as reference.

#### **4.6 Testing Validity of the Results**

The answer of the participants may change according to how the question is asked which makes the answers unreliable. It would be nice to know that the questionnaire you are using will always result in consistent and reliable responses even if the questions were replaced with similar ones (Santos, 1999). If you can generate a variable from the set of similar questions that returns a stable response, it can be concluded that your variable is reliable. "Cronbach's alpha is an index of reliability associated with the variation accounted for by the true score of the hypothetical variable we generated. As a result, Cronbach's alpha determines the internal consistency or average correlation of items in a survey instrument to test its reliability (Santos,1999)." Alpha coefficient ranges from 0 to 1. Reliability increases as the coefficient approaches to 1. Nunnaly's study (as cited in Santos (1999)) has indicated 0.7 is an acceptable reliability coefficient but lower thresholds are sometimes used in the literature. Gliem & Gliem (2003) concluded it is indispensable to calculate and report Cronbach's alpha coefficient for internal consistency reliability for any scales or subscales one may be using when Likert-type scales are used.

To test the validity of the results, for the 5 point Likert Scale short assesment questions Cronbach alpha test is applied considering the conclusions of Gliem & Gliem (2003) and questions that need to be excluded are detected to reach 0.7 reliability coefficient which is stated as acceptable by Nunnaly's study (as cited in Santos (1999)) and George and Mallery's study (as cited in Gliem and Gliem ( 2003)).

A total of the replies of 32 participants to the 37 5-point Likert scale questions are the base of our reliability test. However, since Cronbach alpha does not accept missing replies, a data preprocessing step is needed to prepare the matrix that will be used for the test. Because of missing values, we have to exclude all the replies of one participant and 11 questions. These excluded questions are:

1. Perceived Hardness Level of Task 8
2. Perceived Hardness Level of Task 9
3. Content is up-to-date and accurate QUE:18
4. Content across multiple languages is NOT comprehensive QUE:20
5. The website prevents errors and helps the user recover from them QUE:21
6. How often did you use “Search” service available on the web-site? QUE:22
7. How useful was the “Search” service available on the web-site? QUE:23
8. Level of Confidence for QUE:24
9. Level of Confidence for QUE:25
10. Level of Confidence for QUE:27
11. Please rate your computer literacy level QUE:28

This reliability test is applied using “R” free statistical package which provides a language and environment for statistical computing and graphics. We used “reliability” function of “CTT” package in “R”. This function performs reliability analyses, providing coefficient alpha and item statistics. The most important feature of this function is that

it provides “alpha.if.deleted” statistics. This statistics give Cronbach’s alpha value if the corresponding item was deleted. Using this function, we increased the reliability of our study from 0,3551 to 0,7670 which is an acceptable reliability level according to George and Mallery’s study (as cited in Gliem and Gliem ( 2003). To reach this reliability level, we applied a total of 6 step reliability calculation. At each step, we deleted the item that most increases the reliability level. The deleted items are listed below giving also at which step they were deleted:

1. Perceived Hardness Level of Task 3
2. Perceived Hardness Level of Task 6
3. Perceived Hardness Level of Task 2
4. The website does NOT deliver on the perceived promise of the brand-METU  
QUE:3
5. Interacting with the Web site was NOT easy QUE:10
6. I did NOT feel comfortable using this web site for these tasks QUE:12
7. It was NOT easy to find the required information on the web site QUE:16

#### ***4.7 Summary Of The Results***

Different from Gullikson et al. (1999)’s study, we evaluate answers not only as ‘found’ or ‘not found’ but also as ‘false’ answers to find out the reasons behind the problems of the web site. Figure 5 shows the number of right, false, empty replies and also average perceived hardness level of each task.

Perceived hardness level of Task 2 and Task 3 are deleted from the list respectively at the third step and the first step for increasing the reliability level to acceptable levels of Cronbach alpha value calculations in part 4.6.

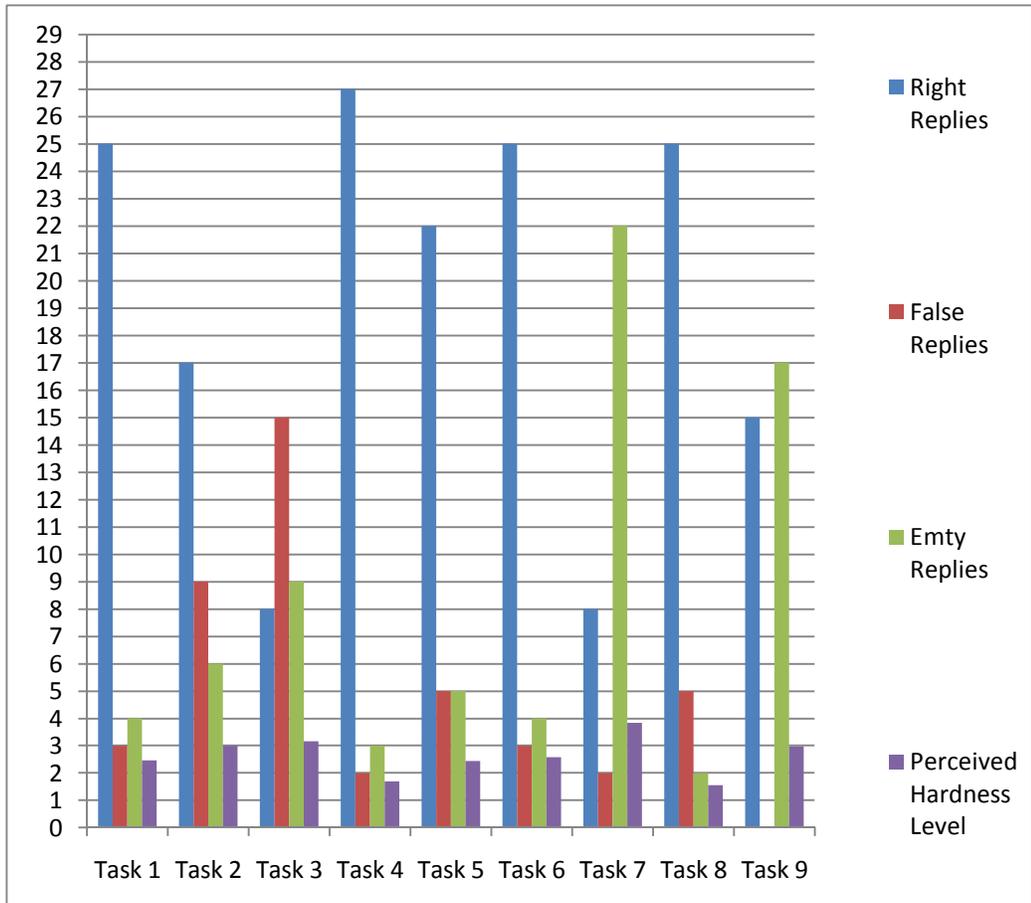
Task 3 has the highest false replies; this task asks from the participants to find which of the programs among Information Systems (IS), Software Management (SM) and Informatics-Online (ION) are evening education programs. This was an information gathering task. The highest number of wrong replies can stem from this task' being wrongly understood by the participants or since it is an information gathering task, one of the program web-page may not be including this information and so partial replies may be collected by the participants. To test which of the reason is more accurate we analyse the wrong replies. Although the right reply was "SM and ION", 9 of the participants choose "SM" and 5 of the participants choose "ION", this means that since 14 out of 15 false replies include part of the right reply, we can conclude that the task is well understood by the participants. If the question is well understood, the only left reason behind partial replies is the program web pages' not including the reply. When we checked whether "evening education" keyword or a similar keyword is included in the first 200 human-readable words of the program web-pages, we found out that "evening education" appears at the bottom of the Software Management program web-page. However, participant needs to pull down the cursor to view it. Informatics Online program web-page includes "who need continuing education at anytime and anywhere without the need to come to the METU campus for lectures." phrase in the first paragraph but all the participants may not understand this phrase as "evening education". Looking whether the keyword is used in the content of program web pages shows that participants need to have given "ION" reply more than "SM" although the opposite took place. We need to test also the results of in-site and Google searches. In-site search gave "Software Management" program and Google gave first ION and secondly SM web-pages but you could not find any "evening education" keyword in the opened "ION" tagged web-page. From these facts, our observations during the think-

aloud study and camera recordings, we can conclude that participants could not decide whether “no need to come to the Metu campus” also means “evening education” or not. They mostly used insite search service to find the answer of this task instead of pulling down the cursor and Google searches.

As can be seen in Figure 5, Task 4 has the highest right replies which means that it was comparatively easy for users to reach application requirements of “Software Management” program. Perceived hardness level of Task 6 is deleted from the list at the second step of increasing the reliability level to acceptable levels of Cronbach alpha value calculations in part 4.6.

Task 7 has the highest empty replies. Not surprisingly, the average perceived hardness level for Task 7 is the highest. This is a fact finding task which asks the participants to find the faculty member who participates into two research groups about Software Technologies and Data Mining. The highest number of empty replies can be the result of tasks’ not being understood or its necessitating so many clicks to reach.

Since Cronbach alpha reliability test does not accept missing values and perceived hardness level of Task 8 and Task 9 include missing values for some of the participants, they are not included in the Cronbach alpha reliability test in part 4.6.



**Figure 5: Performance Results of Participants for Each Task**

After performing the tasks, the participants are asked 5 point Likert scale short assessment questions. These questions assess the experience of the participants about the informatics institute web site. As described in Section 4.1, short assessment questions included in the think-aloud study belongs to 6 different web site evaluation categories which are respectively: branding, functionality, content & ease of use, challenge & skill, self-efficacy and lastly overall evaluation. The results of each question will be given as tables of each category. These tables include category, category average, rank of the question in the webform, the question itself, average replies to each question, median

and mode of replies. In the tables, questions which are not included in the reliability calculations because of missing values are written in *italic* and the ones which are deleted during reliability calculations to reach acceptable Cronbach alpha reliability levels are written in **bold**. Both types of these questions are not included in the calculation of category average values below.

The first three questions are used to evaluate branding property of the web site. Questions and statistics about them are given in Table 40. 1 stands for strong disagreement and 5 stands for strong agreement with the statement. Branding category reliable questions average is 3.69 which means that participants agree that informatics institute web site is better than average in terms of branding dimension of web site success.

**Table 40: Branding Category Questions Results**

| Category | Category Average | Rank | Question   | Average | Median | Mode |
|----------|------------------|------|--|---------|--------|------|
| Branding | 3.69             | 1    | The visual impact of the site is consistent with the brand identity            | 3.53    | 4      | 4    |
|          |                  | 2    | Graphics, Collaterals and Multimedia add value to the experience.              | 3.84    | 4      | 5    |
|          |                  | 3    | <b>The website does NOT deliver on the perceived promise of the brand-METU</b> | 2.06    | 2      | 2    |

Functionality category questions and their statistics can be seen in Table 41. 1 and 5 stands for the same meaning as it is stated for branding category. Functionality of the web site is assessed by the participants as better than average by 3.91 category average. The web site speed is thought to be better than the other functionality category statements.

**Table 41: Functionality Category Questions Results**

| Category      | Category Average | Rank | Question   | Average | Median | Mode |
|---------------|------------------|------|--|---------|--------|------|
| Functionality | 3.91             | 4    | The Web site speed is fast.                                    | 4.16    | 4      | 5    |
|               |                  | 5    | There is little waiting time for the Web pages to load.        | 3.69    | 4      | 4    |
|               |                  | 6    | Users receive timely responses to their queries / submissions. | 3.88    | 4      | 5    |

1 and 5 stand for the same meaning as it is stated for branding category in Table 42. Content & Ease of use category reliable questions have 3.61 average means that participants evaluate our case study web page as better than average in terms of content & ease of use dimension of web site success.

**Table 42: Content & Ease of Use Category Questions Results**

| Category              | Category Average                          | Rank | Question  | Average | Median | Mode |
|-----------------------|---|------|---|---------|--------|------|
| Content & Ease of Use | 3.61                                      | 8    | Navigation of the Web site was simple and easy                            | 3.34    | 3      | 4    |
|                       |   | 9    | Link density provides clarity and easy navigation                         | 3.56    | 4      | 4    |
|                       |   | 10   | <b>Interacting with the Web site was NOT easy.</b>                        | 2.38    | 2      | 2    |
|                       |   | 11   | The website provides visitors with an engaging and memorable experience   | 3.31    | 3      | 3    |
|                       |   | 12   | <b>I did NOT feel comfortable using this web site for these tasks</b>     | 2.47    | 2      | 2    |
|                       |   | 13   | The website helps its visitors accomplish common goals and tasks          | 3.91    | 4      | 4    |
|                       |   | 14   | Content structured in a way that facilitates the attainment of user goals | 3.66    | 4      | 4    |
|                       |   | 15   | I felt that I had the freedom to go anywhere in the Web site              | 3.94    | 4      | 5    |
|                       |   | 16   | <b>It was NOT easy to find the required information on the web site.</b>  | 2.81    | 3      | 3    |
|                       |   | 17   | The organisation of information on the web site was clear to me.          | 3.56    | 4      | 4    |
| 18                    | <i>Content is up-to-date and accurate</i> | 4.37 | 5   | 5       |        |      |

Challenge & skill category questions results are given in Table 43. Category average of reliable questions is 3.97. This is evaluated as same as the above categories, the participants are more close to strong agreement with the statement under this category. 7<sup>th</sup> rank question in the webform asks for frequency and so it does not have average, median, mode values. 23 of the participants have never before visited our case study web site. 4 of the participants have visited the informatics institute web site once in every several months and 3 of the participants have visited once in every several weeks. Only 2 of the participants have visited the informatics institute web site more than once a week. Since 23 of 32 participants have never visited the case study web page before, we could evaluate the information architecture of our web site for the first time visitors. In 22<sup>th</sup> and 23<sup>rd</sup> ranked questions, 1 stands for never and 5 stands for always. 2.77 average value of 22<sup>nd</sup> ranked question means that search service available on the web site is used less than average and 3.50 average value of 23<sup>th</sup> question stands for better than average useful search service.

**Table 43: Challenge & Skill Category Questions Results<sup>2</sup>**

| Category          | Category Average | Rank | Question   | Average | Median | Mode |
|-------------------|------------------|------|--|---------|--------|------|
| Challenge & Skill | 3.97             | 7    | How often have you visited Informatics Institute Website?                | NA      | NA     | NA   |
|                   |                  | 19   | Content is appropriate to customer needs and business goals              | 3.97    | 4      | 4    |
|                   |                  | 20   | <i>Content across multiple languages is NOT comprehensive</i>            | 2.38    | 2      | 2    |
|                   |                  | 21   | <i>The website prevents errors and helps the user recover from them</i>  | 3.56    | 4      | 4    |
|                   |                  | 22   | <i>How often did you use "Search" service available on the web-site?</i> | 2.77    | 3      | 2    |
|                   |                  | 23   | <i>How useful was the "Search" service available on the web-site?</i>    | 3.50    | 4      | 4    |

---

<sup>2</sup> NA refers to Not Applicable

Self-efficacy category questions are two step questions. Firstly, the existence of any effect of the statement during applying the tasks is asked and if the participant believes that there is an effect, confidence level of the participant is asked. The number of the participants who accept the effect of each statement under self-efficacy category are given in Table 44. All of the participants are more than moderately confident that they could perform better using the web site if they had a lot of time to accomplish the tasks. This means that 3 minutes 20 seconds time restriction is less than needed for the tasks included in this think-aloud study. Table 45 gives the confidence level of the participants and the statistical values for self-efficacy category. 1 stands for not all confident, 3 stands for moderately confident and 5 stands for totally confident for the 24<sup>th</sup> to 27<sup>th</sup> question. For 28<sup>th</sup> question, 1 means basic and 5 means excellent. Most of the participants accept their computer literacy level as 4 and average of the replies is 3.53 which means that participants believe that their computer literacy level is better than average.

**Table 44: Self-Efficacy Category Questions “Yes” or “No” Reply Results**

| Category      | Rank | Question  | Number of "Yes" | Number of "No" |
|---------------|------|---|-----------------|----------------|
| Self-Efficacy | 24   | I could perform better using this website, If I had navigated similar web pages before this page. | 24              | 8              |
|               | 25   | I could perform better using this website, If someone else had helped me get started              | 27              | 5              |
|               | 26   | I could perform better using this website, If I had a lot of time to accomplish the tasks.        | 32              | 0              |
|               | 27   | I could perform better using this website, If my English was better                               | 22              | 10             |

**Table 45: Self-Efficacy Category Confidence Level Reply Results**

| Category      | Category Average | Rank | Question   | Average | Median | Mode |
|---------------|------------------|------|--|---------|--------|------|
| Self-Efficacy | 4.38             | 24   | <i>I could perform better using this website, If I had navigated similar web pages before this page.</i> | 3.79    | 4      | 3    |
|               |                  | 25   | <i>I could perform better using this website, If someone else had helped me get started</i>              | 3.93    | 4      | 5    |
|               |                  | 26   | I could perform better using this website, If I had a lot of time to accomplish the tasks.               | 4.38    | 5      | 5    |
|               |                  | 27   | <i>I could perform better using this website, If my English was better</i>                               | 4.09    | 4      | 5    |
|               |                  | 28   | <i>Please rate your computer literacy level</i>  | 3.53    | 4      | 4    |

Overall evaluation category questions and their statistical values are given in Table 46. There is only one short assessment question in this category; the other three are open ended questions so they do not have statistical values. 1 stands for strongly disagree and 5 stands for strongly agree. Overall satisfaction from our institute web site is on average 3.78 which means that participants in the think-aloud study assess our web site better than average. The results of the open ended questions under this category are given in Table 47. It presents the qualitative evaluation of the institute web site. While analyzing the qualitative evaluation of the institute by the participants of think-aloud study, Figure 6 which has the view of the institute web page during think-aloud study will be beneficial.

**Table 46: Overall Evaluation Category Reply Results<sup>3</sup>**

| Category           | Category Average | Rank | Question   | Average | Median | Mode |
|--------------------|------------------|------|--|---------|--------|------|
| Overall Evaluation | 3.78             | 29   | Overall I was satisfied with this web site.            | 3.78    | 4      | 4    |
|                    |                  | 30   | What did you like most about our Institute web site?   | NA      | NA     | NA   |
|                    |                  | 31   | What did you like least about our Institute web site ? | NA      | NA     | NA   |
|                    |                  | 32   | Other Comments   | NA      | NA     | NA   |

**Table 47: Qualitative Evaluation of Institute Web-Site**

| Most Liked Attributes   | Least Liked Attributes  |
|---|---|
| ✓ Color choice  | ✓ The color of the links made them unnoticeable.  |
| ✓ Categorization of the content   | ✓ In the leftside menu, links are being opened over white background and so their sub-headings are not noticeable and subheadings are like part of the content of the viewed page. Subheadings view time is also very less to be noticed. |
| ✓ The menu on the left side   | ✓ An additional different menu at the top is not noticeable.  |
| ✓ Sub-headings are very clear to understand what they include and target can be reached easily by the help of them.   | ✓ Links under the banner are useless.   |
| ✓ "Current Student' menu .  | ✓ Some important web-pages' links are out of vision and so hard to find .   |
| ✓ The links in "Information for" part are organized very well and it helps to find what you look for.   | ✓ In the home page, distribution of the frames at the bottom can be better. The calender on the right part and part about symposium could be arranged under the middle part.  |
| ✓ The flash in the centre of the main page is successfull in terms of attention grabbing by the photos, appereance of the main page and general information provided. | ✓ A 2 frame design structure will be better than the current 3 frame structure. The left and main frame are adequate since the right frame is not used too much.  |

<sup>3</sup> NA refers to Not Applicable

**Table 47 (cont.)**

| Most Liked Attributes   | Least Liked Attributes  |
|---|---|
| ✓ Program introduction videos by instructors which exist on program web pages.  | ✓ Too much content and too small font   |
| ✓ It is a simple web-page. Everything is clear to understand and easy to find.  | ✓ There are no photos about the department that will grab attention in the logo part like the ones used on METU main page.  |
| ✓ Content is sufficient.  | ✓ Lots of content necessitates more time while looking for an information.  |
| ✓ Faculty member information can be easily found  | ✓ Lots of information, hard to find what you look for .   |
| ✓ All the academic programs and M.S. and P.h.d programs available under each program has been explained in detail and it is easy to reach them. | ✓ A lot of information is located under deep links  |
| ✓ “Who may apply” page is designed very well, content is structured in a way that you can find what you look for easily.                        | ✓ Accessing the core information about the department was harder than accessing the details.  |
| ✓ Announcements part in the right-hand side provides fast information gathering without navigating in the site.                                 | ✓ The headings used for the left side menu does not define what is included under that topic.   |
| ✓ Site’s including no advertisement is good.  | ✓ Searching for information in this site necessitates familiarity with the topics.  |
| ✓ In-site search returns successful results   | ✓ Faculty staff information needs to be put under Academics category. In “Application to Graduate Programs ”web-page, content needs to be added about which undergraduate programs are accepted. Under each department, we need to see as a separate heading thesis and nonthesis programs offered. |
|   | ✓ Too complex to navigate, need to click multiple links for reaching information, it was full of links which the visitor needs to read to detect the useful ones.   |
|   | ✓ For the same level data, we need to turn back to the previous page, there are links everywhere but there is no link to access Information systems department page from the programs under Information Systems.  |
|   | ✓ The links about the programs should be given inside the webpage of that program. Not as a new link under another category! For example, the curriculum list of a master program could have been given as a link under the program's webpage.  |

Table 47 (cont.)

| Most Liked Attributes | Least Liked Attributes   |
|-----------------------|--|
|                       | <ul style="list-style-type: none"> <li>✓ Some web-pages which are written in Turkish language.</li> </ul>  |
|                       | <ul style="list-style-type: none"> <li>✓ For some pages, when the language is changed to Turkish, the content is again served in English.</li> </ul>   |
|                       | <ul style="list-style-type: none"> <li>✓ Language preference of the visior is not saved and the language of the served page automatically turns to English which slow downs the navigation.</li> </ul> |

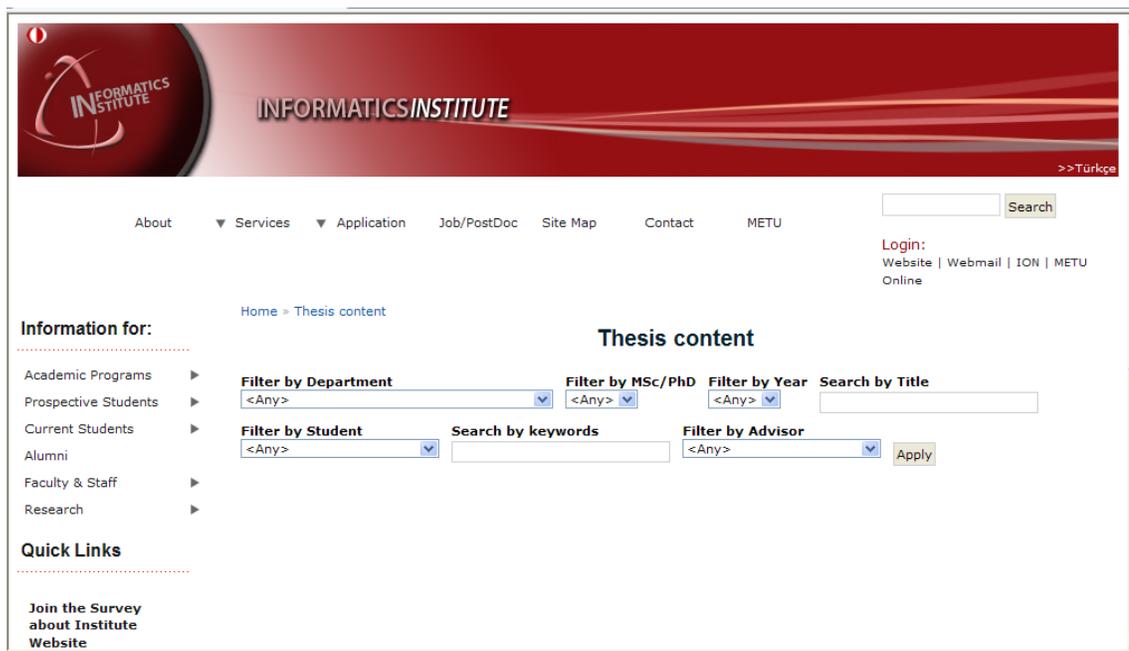


Figure 6: Institute Web-Site View during the Think-Aloud Study

Additional to the studies above, we tested the following three null hypothesis using Microsoft Office 2007 Excell Data Analysis Toolpark. We conducted an independent samples t test to determine if there is a significant difference between the sum of tasks accomplished by undergraduate degree participants and graduate degree participants which is composed of master of science and doctorate degree. The resulting *t* test output is given in Figure 7.

| <u>t-Test: Two-Sample Assuming Equal Variances</u> |                           |   |
|--|---------------------------|---|
|  | <i>Undgraduate Degree</i> | <i>Master of Science &amp; Doctorate Degree</i> |
| Mean   | 4.7500                    | 5.7895  |
| Variance   | 2.7500                    | 2.7310  |
| Observations                                       | 12                        | 19  |
| Pooled Variance                                    | 2.7382                    |   |
| Hypothesized Mean Difference                       | 0                         |   |
| df   | 29                        |   |
| t Stat   | -1.7036                   |   |
| P(T<=t) one-tail                                   | 0.0496                    |   |
| t Critical one-tail                                | 1.6991                    |   |
| P(T<=t) two-tail                                   | 0.0992                    |   |
| t Critical two-tail                                | 2.0452                    |   |

**Figure 7: t-test Output of Hypothesis 1**

The significant *t* test indicates at alpha=0.05 level that we can accept the null hypothesis of graduate degree's having no effect on the sum of accomplished tasks. Because our hypothesis was directional, a one-tailed test is appropriate. Graduate degree was not a significant demographic feature of the participants who accomplished more tasks,  $t(29) = -1.7036, p < .0496$ , one tailed.

Secondly, we conducted the same test to determine if there is a significant difference between the sum of tasks accomplished by participants whose daily average internet use in hours is more than 2 hours and equal to and lower than 2 hours. We select 2 because it is the mode of this sample. The resulting  $t$  test output is given in Figure 8.

| <u>t-Test: Two-Sample Assuming Equal Variances</u> |   |   |
|--|---|---|
|  | <i>Daily Average Internet Use in hrs &gt; 2 hrs</i> | <i>Daily Average Internet Use in hrs &lt;=2 hrs</i> |
| Mean   | 5.8947  | 4.5833  |
| Variance   | 1.3216  | 4.6288  |
| Observations                                       | 19  | 12  |
| Pooled Variance                                    | 2.5761  |   |
| Hypothesized Mean Difference                       | 0   |   |
| df   | 29  |   |
| t Stat   | 2.2159  |   |
| P(T<=t) one-tail                                   | 0.0174  |   |
| t Critical one-tail                                | 1.6991  |   |
| P(T<=t) two-tail                                   | 0.0347  |   |
| t Critical two-tail                                | 2.0452  |   |

**Figure 8: t-test Output of Hypothesis 2**

The significant  $t$  test indicates at  $\alpha=0.05$  level that we can reject the null hypothesis of daily average internet use more than 2 hours having no effect on the sum of accomplished tasks. Because our hypothesis was directional, a one-tailed test is appropriate. Daily average internet use more than 2 hours was a significant demographic feature of the participants who accomplished more tasks,  $t(29) = 2.2159$ ,  $p < 0.0174$ , one tailed.

Latly, we conducted the same test to determine if there is a significant difference between the sum of tasks accomplished by participants whose average internet use in

years is more than 10 years and equal to and lower than 10 years. We select 10 because it is the median and mode of this sample. The resulting  $t$  test output is given in Figure 9.

| <u>t-Test: Two-Sample Assuming Equal Variances</u> |   |   |
|--|---|---|
|  | <i>Average internet use<br/>in years &gt;10 yrs</i> | <i>Average internet use<br/>in years &lt;=10yrs</i> |
| Mean   | 5.7500  | 5.2609  |
| Variance   | 4.2143  | 2.5652  |
| Observations                                       | 8   | 23  |
| Pooled Variance                                    | 2.9633  |   |
| Hypothesized Mean Difference                       | 0   |   |
| df   | 29  |   |
| t Stat   | 0.6923  |   |
| P(T<=t) one-tail                                   | 0.2471  |   |
| t Critical one-tail                                | 1.6991  |   |
| P(T<=t) two-tail                                   | 0.4943  |   |
| t Critical two-tail                                | 2.0452  |   |

**Figure 9: t-test Output of Hypothesis 3**

The significant  $t$  test indicates at  $\alpha=0.05$  level that we can accept the null hypothesis of average internet use more than 10 years having no effect on the sum of accomplished tasks. Because our hypothesis was directional, a one-tailed test is appropriate. Average internet use more than 10 years was not a significant demographic feature of the participants who accomplished more tasks,  $t(29) = 0.6923$ ,  $p < 0.2471$ , one tailed.

## **CHAPTER 5**

### **CONCLUSION**

In this thesis study, taking METU Informatics Institute web site as a case study, we have focused on forming recommendations to restructure and classify information driven web sites to help visitors easily find and manage information. To form recommendations, we make use of a newly emerging tool for evaluation of web sites which is web analytic tool and a classical tool which is think-aloud study.

Among the web analytic tools, we prefer to use Google Analytics which is free of charge up to 5 million pageviews per month per account according to Google Analytics Terms of Service Article 2 (Google, 2010e) and it is recommended by experts in this area. Google Analytics has valuable reports which are not used because of their complexity. In this thesis study, we explain our findings on the issues that need to be considered while reading Google Analytics reports. Additionally, we explain the dimensions and metrics which have the most confusing definitions using Google Analytics report examples. To improve the information architecture of information driven web sites, we use landing page optimization method which is available in Google Analytics. Since we have limited time for this study, we specify the most important program web pages in the institute web site and apply landing page optimization method to these web pages. We form a methodology to detect the most problematic keywords which land these pre-specified web pages and analyze these keywords one by one.

To apply think-aloud study, we create a new questionnaire by tailoring the questionnaires in the literature to our needs. This questionnaire can be used in future studies which can be applied to content sites. It is composed of 8 demographic questions, 9 tasks, 29 short assesment questions and 3 open ended questions. Since studies conducted in Middle East Technical University (METU) and/or studies conducted by METU personnel/students, which involve collecting data from human participants, are subject to review by the METU Human Subjects Ethics Committee (HSEC), we take their approval before application. We apply think-aloud study to 32 participants from 25.02.2010 to 21.03.2010.

We conclude in this study by using Google Analytics results that tag pages are lowering the success of our case study web site and so a rearrangement for these tagged pages needs to be done. All the web pages about informatics online program need to include “uzaktan”, “internet üzerinden”, “online” and “distance” keywords among the first 200 words of their content to lower the bounce rate of these keywords. A new page which gives short description of all available master programs under Informatics Institute and includes links to each program web page needs to be designed both in English and Turkish. The master programs in this page need to be categorized as thesis and nonthesis. A second category needs to be defined for online programs, normal programs and evening education programs. The new designed page needs to include “yüksek lisans”, “ms degree”, “ master degree”, “yüksek lisans master programları”, “eğitim yüksek lisans”, “odtü tezsiz yüksek lisans”, “tessiz master odtü”, “metu tezsiz yüksek lisans”, “master ankara”, “yükseklisans ankara”, “master programları ankara”, “uzaktan eğitim veren üniversiteler” and “internette eğitim veren üniversiteler” keywords. “bilgi yüksek lisans Ankara” keyword needs to be added by the web site designers as a metatag to each related program web page of informatics institute. The links to the health informatics program need to have the following keywords: “sağlık bilgi sistemleri”, “tıbbi sistemler ve bilgi”, “health information systems phd program” which are already present in information systems program web page. Since the content

of the tagged pages are not proper and the work based learning program web page is sufficient, metatags “iş yaşamı temelli öğrenme” and “Middlesex university” need to be deleted. Similarly, since information systems program web page is adequate, metatag “information systems” needs to be deleted by the web site designers.

From the think-aloud study results, we conclude that non thesis master programs, online programs, evening education programs are not clearly explained in our case study web site which supports our recommendation of creating a new page which gives short description of all available master programs under Informatics Institute according to thesis-nonthesis and online-evening-normal education categories by using Google Analytics results. Second issue necessitates a revision on faculty member web sites. The research groups that the faculty member exists in are not given within the first 200 human readable words and so it is not read by the visitors. This information should be given at the top of the page. Thesis preparation guideline which can be accessed from the institute web site is too complex for the visitors to detect important points in a limited time such as 3 minutes 20 seconds as we applied in the think-aloud study. This guideline needs to be simplified. 6 different web site evaluation categories exist in the think-aloud study which are respectively: branding, functionality, content & ease of use, challenge & skill, self-efficacy and lastly overall evaluation. According to think-aloud study results, our case study web site is better than average in all of these categories. Self-efficacy category average is the highest one which shows that participants are confident that they have the capability to perform the tasks given to them. According to the independent samples t-tests we conducted per each demographic feature we choose, we concluded that graduate degree and average internet use more than 10 years were not significant demographic features of the participants who accomplished more tasks. Only daily average internet use more than 2 hours appeared to be a significant demographic feature of the participants who accomplished more tasks.

According to the answers given by participants to open ended questions and my notes on their thoughts about the web site during the study, we group the recommendations according to content and design. Content related recommendations are:

- A lot of information is located under deep links, there is too much content and used font is too small; this problem needs to be solved.
- Labels used for the left side menu are written in academic terminology, they are not understood by people having only undergraduate degree and so these labels are ambiguous for these visitors. These labels need to be revised. For example, some of the visitors are used to access web pages of the academicians under Academics category and Program sub-category in which the academician delivers lectures. They first look for these labels, faculty and staff labels on our institute web site misguide these visitors.
- Some web-pages do not have their Turkish translations, these pages need to be created.
- In “Application to Graduate Programs ” web-page, content needs to include the required student background information such as the graduated program for application to our programs.
- Photos about the department that will grab attention in the logo part like the ones used on METU main page need to be added.

Design related recommendations are:

- In the leftside menu, links are opened over white background and so their sub-headings are not noticeable and subheadings are like part of the content of the viewed page and also subheadings view time is also very little to be noticed, this design to be changed with a more usable one.

- In the home page, distribution of the frames at the bottom can be better. The calendar on the right part and part about symposium could be arranged under the middle part. A 2 frame design structure will be better than the current 3 frame structure. The left and main frame are adequate since the right frame is not used too much.
- Too complex to navigate, need to click multiple links for reaching information, full of links which the visitor needs to read to detect the useful ones. Navigation needs to be simplified, navigational aids need to be provided.
- Language preference need to be saved.
- An additional different menu at the top is not noticeable and links under the banner are useless so they need to be deleted.

A/B test which is a direct comparison between two pieces of equipment, or two design alternatives can be applied by the Informatics Institute web site administrators to measure the success of recommendations provided by this study. Since our study relies on Google Analytics reports for the period **01.09.09 to 21.03.10**, to make the comparison it will be better to rely on the same period next year reports not to be affected by the seasonality. But as a minimum standard, a 6 month data collection period needs to be allocated after realizing the recommendations to comply with our data collection period in this study. The steps which need to be realized by the Informatics Institute web site administrators after applying the recommendations of this study and accumulating 6 months data are:

- 1 Using the report in APPENDIX D as the baseline, same report for the again 6 months period needs to be taken. If there is decrease in the bounce rates of the selected web sites, it means that our recommendations which are the result of this study are effective in improving the information architecture of the institute web site.

- 2 Using the code in APPENDIX E after updating the “query.setStartDate” and “query.setEndDate”, the updated version of the Custom Report we used in section 3.7 to 3.14 will be obtained. The Problematic Keyword statistics of the new period will be accumulated by applying the methodology we explained in section 3.7 which are successively,
  - Sort the Custom Report according to bounces and choose the keywords which have bounces over 1
  - Calculate the bounce rates for the selected keywords and choose keywords which have 100% bounce rates.

In our study we have counted only 57 keywords which have *Bounces* over 1 and among them, 22 keywords have 100 % *Bounce Rate*. If the number of keywords having *Bounces* over 1 and the count of 100 % *Bounce Rate* keywords decreases, this will show the success of our recommendations for the improvement of the information architecture of Informatics Institute web site.

Further studies can be done both in Google Analytics and think-aloud study part. One venue for future research can be finding the ways on fixing the inaccurate *Keyword-LandingPagePath* pair report of Google Analytics. Another possible route can be benefiting from other valuable reports of Google Analytics such as goal conversion, visitor trending, map overlay, visitor loyalty for improving the information architecture of information driven web sites. In the think-aloud study direction, further studies can be analyzing the results using partial least squares modeling to detect the correlations between properties of users, their assesment of the website and their success in performing the tasks.

## REFERENCES

(2010). Retrieved July 6, 2010, from W3Perl: <http://www.w3perl.com/>

ActualMetrics.com. (2009). *Compare Google Analytics to Urchin Software*. Retrieved March 26, 2009, from actualmetrics.com:  
<http://www.actualmetrics.com/compare/google-analytics/urchin-5-software/>

Agichtein, E., Brill, E., & Dumains, S. (2006). Improving web search ranking by incorporating user behavior information. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. Seattle: ACM.

Anil, W. A. (2010). *Why does Google analytics shows zero visits for keywords ? What does it mean ?* Retrieved July 7, 2010, from Google Analytics Help Forum:  
<http://www.google.com/support/forum/p/Google+Analytics/thread?tid=0d04370f123bc16c&hl=en>

Arendt, J., & Wagner, C. (2010). Beyond Description: Converting Web Site Usage Statistics into Concrete Site Improvement Ideas. *Journal of Web Librarianship* , 4 (1), 37--54.

Ash, T. (2008). *Landing page optimization*. Indianapolis, Indiana: Wiley Publishing, Inc.

Ash, T. (2008, June 11). *Uncovering Site Problems for Landing Page Optimization, Part I*. Retrieved April 11, 2010, from Search Engine Watch:  
<http://searchenginewatch.com/3629878>

ASIS&T. (2010). *About ASIS&T*. Retrieved July 26, 2010, from asis & t The Information Society for the Information Age: <http://www.asis.org/index.html>

Atkinson, E. (2007). *Web analytics and think aloud studies in web evaluation: understanding user experience*. University College London, Faculty of Life Sciences. London: UCL.

Avoris, N., Tselios, N., Fidas, C., & Papachristos, E. (2003). Website evaluation: A usability-based perspective. *Advances in Informatics* , 217-231.

Awio Web Services LLC. . (n.d.). Retrieved July 6, 2010, from W3Counter: <http://www.w3counter.com>

Barrett, B. (2009, November 11). *The Webalizer*. Retrieved July 6, 2010, from <http://www.webalizer.org/>

Bauer, C., & Scharl, A. (2000). Quantitive evaluation of Web site content and structure. *Internet Research: Electronic Networking Applications and Policy* , 10 (1), 31--43.

Brinck, T., Ha, S. S., Pritula, N., Lock, K., Sperdelozzi, A., & Monan, M. (2003). Making an iMPact: redesigning a business school Web site around performance metrics. *Proceedings of the 2003 conference on Designing for user experiences* (pp. 1--15). ACM.

Burby, J., Brown, A., & Committee, W. S. (2007, August 16). *Web Analytics Definitions*. Retrieved April 4, 2009, from Web Analytics Association: [http://www.webanalyticsassociation.org/resource/resmgr/PDF\\_standards/WebAnalyticsDefinitionsVoll.pdf](http://www.webanalyticsassociation.org/resource/resmgr/PDF_standards/WebAnalyticsDefinitionsVoll.pdf)

Cao, M., Zhang, Q., & Seydel, J. (2005). B2C e-commerce web site quality: an empirical examination. *Industrial Management and Data Systems* , 105 (5), 645--661.

Chi, E., Rosien, A., & Heer, J. (n.d.). LumberJack: Intelligent discovery and analysis of web user traffic composition. *WEBKDD 2002-MiningWeb Data for Discovering Usage Patterns and Profiles* , 1--16.

Clifton, B. (2008a). *Advanced web metrics with Google Analytics*. Indianapolis, Indiana: Wiley Publishing, Inc.

Clifton, B. (2010, April). *E-Metrics Whitepaper: Understanding Web-Analytics Accuracy*. Retrieved June 4, 2010, from Advanced-Web-Metrics.com: <http://www.advanced-web-metrics.com/docs/accuracy-whitepaper.pdf>

Clifton, B. (2008b, February). *Web Analytics Whitepaper: Increasing Accuracy for Online Business Growth*. Retrieved May 4, 2009, from Measuring Success: Official site for the book *Advanced Web Metrics with Google Analytics* by Brian Clifton: [www.estudaringlesnaaustralia.com/descarga.php?id=7](http://www.estudaringlesnaaustralia.com/descarga.php?id=7)

Clifton, B. (2008c, May). *Web Analytics: Web traffic data sources & vendor comparison*. Retrieved March 4, 2009, from Measuring Success: Official site for the book *Advanced Web Metrics with Google Analytics* by Brian Clifton: <http://www.advanced-web-metrics.com/docs/web-data-sources.pdf>

Cohen, J. (1992a). A power primer. *Psychological bulletin* , 112 (1), 155--159.

Cohen, J. (1992b). Statistical power analysis. *Current Directions in Psychological Science* , 1 (3), 98--101.

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *Mis Quarterly* , 19 (2), 189--211.

Dainow, B. (2008, November 18). *The disturbing inaccuracy behind Google Analytics*. Retrieved April 14, 2010, from IMEDIA CONNECTION Connecting the Marketing Community: <http://www.imediaconnection.com/content/21145.asp>

DeSoto, A., Clifton, B., Novo, J., Mason, N., Singh, A., Crosby, B., et al. (2006). *Google Analytics – Conversion University – Learn. Explore. Profit*. Retrieved from [http://www.google.com/analytics/conversionuniversity\\_dt.html](http://www.google.com/analytics/conversionuniversity_dt.html): <http://www.cumbrowski.com/CarstenC/pdf/Google-Analytics-Conversion-University.pdf>

Dillon, A. (2002). Information architecture in JASIST: Just where did we come from? *Journal of the American Society for Information Science and Technology* , 53 (10), 821--823.

Dillon, A., & Turnbull, D. (2000). Information architecture: why, what & when. *Presentation at ASIS Summit* , 1--9.

Dong, J., Martin, S., & Waldo, P. (2001). A user input and analysis tool for information architecture. *CHI'01 extended abstracts on Human factors in computing systems* (pp. 23--24). ACM.

Elling, S., Lentz, L., & de Jong, M. (2007). Website evaluation questionnaire: development of a research-based tool for evaluating informational. *Electronic Government* , 293--304.

Facca, F. M., & Lanzi, P. L. (2005). Mining interesting knowledge from weblogs: a survey. *Data & Knowledge Engineering* , 53 (3), 225--241.

Fang, W. (2007). Using Google Analytics for Improving Library Website Content and Design: A Case Study. *Library Philosophy and Practice* , 9 (3), 1--17.

Genç, A. Ş. (2006). *Web site evaluation*. The Middle East Technical University, Graduate School of Informatics. Ankara: METU.

George, R., & Heimann, A. (2007, October 26). *simpleview Whitepaper: Log Based vs. JavaScript Page Tagging Web Stats – Explained*. Retrieved July 06, 2010, from Simpleview:  
[http://www.simpleviewinc.com/docs/simpleview\\_Whitepaper\\_Log\\_Files\\_vs\\_JavaScript\\_Tagging\\_Web\\_Stats3.pdf](http://www.simpleviewinc.com/docs/simpleview_Whitepaper_Log_Files_vs_JavaScript_Tagging_Web_Stats3.pdf)

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. (pp. 82-88). Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

Gofman, A. (2007). Consumer Driven Multivariate Landing Page Optimization: Overview, Issues, and Outlook. *The IPSI BgD Transactions on Internet Research* , 3 (2), 7--9.

Goodhue, D., Lewis, W., & Thompson, R. (2006). PLS, small sample size, and statistical power in MIS research. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences - 2006 HICSS'06*, 8.

Google. (2010g). *Client Libraries & Sample Code*. Retrieved May 2, 2010, from Google code lab: <http://code.google.com/apis/analytics/docs/gdata/gdataLibraries.html>

Google. (2010h). *Cookies & Google Analytics*. Retrieved August 7, 2010, from Google code: <http://code.google.com/apis/analytics/docs/concepts/gaConceptsCookies.html>

Google. (2010a). *Data API - Data Feed*. Retrieved April 6, 2010, from Google code labs: <http://code.google.com/apis/analytics/docs/gdata/gdataReferenceDataFeed.html>

Google. (2010f). *Dimensions & Metrics Reference*. Retrieved April 6, 2010, from Google code labs:  
<http://code.google.com/apis/analytics/docs/gdata/gdataReferenceDimensionsMetrics.html>  
1

Google. (2010b). *Functional overview*. Retrieved 6 23, 2010, from Google code: <http://code.google.com/apis/analytics/docs/concepts/gaConceptsOverview.html>

Google. (2010e). *Google Analytics Terms of Service*. Retrieved June 5, 2010, from Google Analytics Official Website: <http://www.google.com/intl/en/analytics/tos.html>

Google. (2010c). *What are valid dimension-metric combinations in a custom report?* Retrieved April 6, 2010, from Google Analytics Help:  
<http://www.google.com/support/analytics/bin/answer.py?answer=99174>

Google. (2010d). *Why YMMV web stats?* Retrieved June 23, 2010, from YMM-  
WEBSTATS: <http://code.google.com/p/ymmv-webstats/>

Gullikson, S., Blades, R., Bragdon, M., McKibbin, S., Sparling, M., & Toms, E. G. (1999). The impact of information architecture on academic web site usability. *The Electronic Library* , 17 (5), 293--304.

Gwizdka, J., & Spence, I. (2007). Implicit measures of lostness and success in web navigation. *Interacting with Computers* , 19 (3), 357--369.

Hasan, L., Morris, A., & Proberts, S. (2009). Using Google Analytics to Evaluate the Usability of E-Commerce Sites. *Human Centered Design* , 697--706.

Hollink, V., van Someren, M., & Wielinga, B. J. (2007). Navigation behavior models for link structure optimization. *User Modeling and User-Adapted Interaction* , 17 (4), 339--377.

Hong, I. (2007). A survey of web site success metrics used by Internet-dependent organizations in Korea. *Internet research* , 17 (3), 272--290.

iFusion Labs LLC. (2009). *Woopra*. Retrieved July 6, 2010, from Woopra:  
<http://www.woopra.com/>

Ivory, M. Y., Sinha, R. R., & Hearst, M. A. (2001). Empirically validated web page design metrics. *Proceedings of the SIGCHI conference on Human factors in computing systems March 31-April 4*. Seattle: ACM.

Jasra, M. (2006, August 15). *Web Analytics Comparison – Google vs. VisiStat*. Retrieved March 26, 2009, from Enquiro Search Solutions:  
<http://www.enquiro.com/marketing-monitor/Web-Analytics-Comparison-Google-VisiStat.asp>

Juvina, I., & Oostendorp, H. (2006). Individual differences and behavioral metrics involved in modeling web navigation. *Universal Access in the Information Society* , 4 (3), 258–269.

Kaushik, A. (2007). *Web analytics: An hour a day*. Indianapolis, Indiana: Wiley Publishing, Inc.

Kellar, M., Watters, C., & Shepherd, M. (2007). A goal-based classification of Web information tasks. *Proceedings 69th Annual Meeting of the American Society for Information Science and Technology*, 43. Austin, TX,: Citeseer.

Kirakowski, J., Claridge, N., & Whitehand, R. (1998). Human centered measures of success in web site design. *Proceedings of the Fourth Conference on Human Factors & the Web*. NJ, USA: AT&T Labs.

Kumar, L. (2009a, November 21). *Does Google Analytics affect SEO*. Retrieved April 2, 2010, from Google Analytics Help Forum:  
<http://www.google.com/support/forum/p/Google+Analytics/thread?tid=2b810d1cf1680e1c&hl=en>

Kumar, L. (2009b, May 25). *What's considered a good or bad bounce rate?* Retrieved April 2, 2010, from Google Analytics Help Forum:  
<http://www.google.com/support/forum/p/Google+Analytics/thread?tid=3301d79d3932b447&hl=en>

Kumar, L. (2009c, May 14). *Why is my daily absolute unique visitors more than my visits?* Retrieved March 23, 2010, from Google Analytics Help Forum: 5.  
<http://www.google.com/support/forum/p/Google+Analytics/thread?tid=15ef18a2781b3d88&hl=en>

Lash, J. (2002, November 5). *Information Architecture is not Usability*. Retrieved July 27, 2010, from Digital Web Magazine: [http://www.digital-web.com/articles/information\\_architecture\\_is\\_not\\_usability/](http://www.digital-web.com/articles/information_architecture_is_not_usability/)

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & information technology*, 25 (2), 115--126.

Liu, C., & Arnett, K. P. (2000). Exploring the factors associated with Web site success in the context of electronic commerce. *Information & management*, 38 (1), 23--33.

Marcoulides, G. A., Chin, W., & Saunders, C. (2009). A Critical Look at Partial Least Squares Modeling. *Management Information Systems Quarterly*, 33 (1), 171-175.

Mateos, M. B., Mera, A. C., Gonzales, F. J., & Lopez, O. R. (2001). A new Web assessment index: Spanish universities analysis. *Internet Research: Electronic Networking Applications and Policy* , 11 (3), 226--234.

Misic, M. M., & Johnson, K. L. (1999). Benchmarking: a tool for web site evaluation and improvement. *Internet Research: Electronic Networking Applications and Policy* , 9 (5), 383--392.

Muyllé, S., Moenaert, R., & Despontin, M. (2004). The conceptualization and empirical validation of web site user satisfaction. *Information & Management* , 41 (5), 543--560.

Negash, S., Ryan, T., & Igbaria, M. (2003). Quality and effectiveness in web-based customer support systems. *Information & Management* , 40 (8), 757--768.

NMmarketing. (2009, January 1). *how analytics works? will it inflate*. Retrieved April 2, 2010, from Google Analytics Help Forum:  
<http://www.google.com/support/forum/p/Google+Analytics/thread?tid=5f6d803aa06c2bd2&hl=en>

Olson, J. R., & Boyer, K. K. (2003). Factors influencing the utilization of Internet purchasing in small organizations. *Journal of Operations Management* , 21 (2), 225--245.

Patton, S. (2002, November 15). *E-Commerce Tools: Web Metrics That Matter*. Retrieved July 19, 2010, from CIO:  
[http://www.cio.com/article/31511/E\\_Commerce\\_Tools\\_Web\\_Metrics\\_That\\_Matter?page=2&taxonomyId=3181](http://www.cio.com/article/31511/E_Commerce_Tools_Web_Metrics_That_Matter?page=2&taxonomyId=3181)

Peterson, E. T. (2006). *The Big Book of Key Performance Indicators* (First Edition ed., Vol. Book Two in the Web Analytics Demystified Series).  
<http://www.webanalyticsdemystified.com>.

Peterson, E. (2004). *Web analytics demystified: a marketer's guide to understanding how your web site affects your business*. Celilo Group Media and CafePress.

Prom, C. J. (2007). Understanding On-line Archival Use through Web Analytics. *ICA-SUV Seminar, August 14, 2007*. Dundee, Scotland.

Rosenfeld, L., & Morville, P. (1998). *Information architecture for the world wide web*. Sebastopol, CA, USA: O'Reilly & Associates, Inc. .

Rubinoff, R. (2004, April 21). *How To Quantify The User Experience*. Retrieved May 3, 2009, from sitepoint: <http://www.sitepoint.com/print/quantify-user-experience/>

Santos, J. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of Extension* , 37 (2), 1--5.

Schubert, P., & Dettling, W. (2001). Web site evaluation: Do Web applications meet user expectations? Music, consumer goods and e-banking on the test bed. *Proceedings of the 14th International Bled Electronic Commerce Conference, Bled, Slovenia, June* (pp. 383-403). Bled, Slovenia: IAB – Institute for Business Economics, FHBB – University of Applied Sciences, Basel.

Seva, R., Wu, J., & Yi, X. L. (2006). Evaluation of Cinema Website. *IEEE International Conference on Systems, Man and Cybernetics, 2006. SMC'06. 1*. Taipei, Taiwan: IEEE.

Signore, O. (2005). A comprehensive model for web sites quality. *Seventh IEEE International Symposium on Web Site Evolution, 2005.(WSE 2005)* (pp. 30--36). IEEE Computer Society.

Site Meter. (2009). *How does it work?* Retrieved July 6, 2010, from sitemeter: <http://www.sitemeter.com/>

Skadberg, Y. X., & Kimmel, J. R. (2004). Visitors' Flow experience while browsing a web site: its measurement, contributing factors and consequences. *Computers in Human Behavior* , 20 (3), 403--422.

StatCounter. (2010). *How does it work?* Retrieved July 6, 2010, from StatCounter.com: [http://www.statcounter.com/how\\_it\\_works.html](http://www.statcounter.com/how_it_works.html)

Stolz, C., Viermetz, M., Skubacz, M., & Neuneier, R. (2005). Guidance performance indicator-Web metrics for information driven Web sites. *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on* (pp. 186--192). IEEE.

Thelwall, M., Christel, M. G., White, R. W., Roth, R. A., & Lankes, R. D. (2009). *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*. Morgan & Claypool.

Tullis, T., & Albert, W. (2008). *Measuring The User Experience: collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann Pub.

U.S. Department of Health and Human Services (HHS) and the U.S. General Services Administration (GSA). *Research-Based Web Design & Usability Guidelines*. Washington: U.S. Government Printing Office.

van Schaik, P., & Ling, J. (2005). Five psychometric scales for online measurement of the quality of human-computer interaction in web sites. *International Journal of Human-Computer Interaction* , 18 (3), 309--322.

Weischedel, B., & Huizingh, E. R. (2006). Website optimization with web metrics: a case study. *Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet* (pp. 463--470). Fredericton, Canada: ACM.

Wiggins, A. (2007). Data Driven Design: Using Web Analytics to Improve Information Architectures. *Paper presented in the research track at ASIST Information Architecture Summit 22-26 March, 2007*. Las Vegas: NV.

Wilfert, H. (2008, December). *Web Analytics and Information Architecture* . Retrieved July 29, 2009, from fumsi: <http://web.fumsi.com/go/article/manage/3460>

Zviran, M., Glezer, C., & Avni, I. (2006). User satisfaction from commercial web sites: the effect of design and use. *Information & Management* , 43 (2), 157--178.

## **APPENDICES**

## Approval

  
1956

Orta Doğu Teknik Üniversitesi  
Middle East Technical University

Öğrenci İşleri Dairesi  
Başkanlığı  
Registrar's Office

06531 Ankara, Türkiye  
Phone: +90 (312) 2103417  
Fax: +90 (312) 2107960  
www.oidb.metu.edu.tr

B.30.2.ODT.72.00.00 9410 - 579

20/11/2009

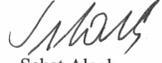
ENFORMATİK ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi: İlgi: 06/11/2009 tarih ve 674- 13802 sayılı yazınız.

Üniversitemiz Bilişim Sistemleri Anabilim Dalı Yüksek Lisans öğrencisi Seher Demirel Kütükçü'nün, 2009- 2010 eğitim-öğretim yılı 1. ve 2. dönemlerinde tezi ile ilgili "Bilgilendirme Amaçlı Hazırlanmış Web Sayfalarının Bilgi Mimarisinin (Information Architecture) Başarısının Ölçümünde Sesli Düşün (Think-Aloud) ve Anket Yönetiminin Yanında/Yerinde Web Analitik Araçlarının Nasıl Kullanılabileceği Konusunda Enstitü Web Sayfasını Örnek Olay Olarak Ele Alan Bir Araştırma" başlıklı araştırma çalışmasına ilişkin olarak Orta Doğu Teknik Üniversitesinde uygulama yapmak için, öğrencinin isteği doğrultusunda görevlendirilmesi Etik Komite onayı ile uygun görülmüştür.

Gereğini bilgilerinize arz ederim

Saygılarımla.

  
Sebat Alanlı  
Öğrenci İşleri Dairesi Başkanı V.

## **APPENDIX B**      **Voluntary Participation Information**

### **Questionnaire for Testing the Informatics Institute Web-site**

In this survey, you will be given tasks which are aimed at finding information from the web-site of METU Informatics Institute./ Bu çalışmada, sizlere ODTÜ Enformatik Enstitüsü web sayfasından bir takım bilgileri bulmanızı amaçlayan görevler verilecek.

Please note that this survey does not test your computer literacy or your knowledge about the Informatics Institute. All the information you will look for exists in the Institute web-site. / Bu çalışma kesinlikle sizlerin bilgisayar becerilerini ya da Enformatik Enstitüsü hakkındaki bilgi birikimini test etmeye yönelik değildir. Bulmanız istenen bilgilerin tamamı Enstitü web sayfasında mevcuttur.

Our aim is to service the right information to the right people within minimum time from our Institute web-site. You can support us by reserving approximately your 45 minutes to participate in our questionnaire which is composed of 9 tasks and 32 short assessment questions. We will appreciate you participation.

Now, please attempt to do the following tasks. While carrying out the tasks, you can browse our web pages to find the requested information, use the “Search” option or “Sitemap” in our web pages. Please DO NOT use your existing knowledge while answering the questions. **You MUST try to find the answer on our web pages at all times.** After you finished each task, we need your feedbacks regarding the answer, the level of its hardness you perceive and lastly your reasoning and suggestion.

The scale below each task shows the hardness you perceive. ( 1:very easy & 5: very hard ) Please mark the hardness level that best suits for you.

Your actions on the web will be recorded in our log files only to be used for academic purposes.

If you **agree** on these conditions, please check the “**Yes**” button below to continue with the survey .

If you **disagree** on these conditions, please check the “**No**” button below to redirect to frontpage.

YES NO      Survey ID: 774875

## APPENDIX C 2009 Traffic From Search Engine Report

http://www.ii.metu.edu.tr  
**Search Engines**

Jan 1, 2009 - Dec 31, 2009  
 Comparing to: Site



Search sent 43,144 total visits via 15 sources

| Site Usage                 |                            |                                 |                             |                             |             |
|----------------------------|----------------------------|---------------------------------|-----------------------------|-----------------------------|-------------|
| Visits                     | Pages/Visit                | Avg. Time on Site               | % New Visits                | Bounce Rate                 |             |
| 43,144                     | 2.87                       | 00:02:22                        | 49.74%                      | 59.85%                      |             |
| % of Site Total:<br>32.24% | Site Avg:<br>2.90 (-7.78%) | Site Avg:<br>00:02:54 (-18.04%) | Site Avg:<br>48.12% (3.37%) | Site Avg:<br>55.88% (7.12%) |             |
| Source                     | Visits                     | Pages/Visit                     | Avg. Time on Site           | % New Visits                | Bounce Rate |
| google                     | 38,213                     | 2.81                            | 00:02:33                    | 53.47%                      | 57.28%      |
| live                       | 2,260                      | 1.45                            | 00:00:48                    | 7.51%                       | 82.62%      |
| bing                       | 1,892                      | 1.58                            | 00:00:52                    | 12.26%                      | 80.87%      |
| yahoo                      | 416                        | 2.27                            | 00:01:36                    | 87.74%                      | 69.47%      |
| search                     | 257                        | 1.90                            | 00:01:30                    | 69.26%                      | 68.87%      |
| msn                        | 63                         | 1.70                            | 00:00:56                    | 63.49%                      | 69.84%      |
| ask                        | 24                         | 2.38                            | 00:03:29                    | 75.00%                      | 66.67%      |
| aol                        | 11                         | 1.64                            | 00:00:15                    | 100.00%                     | 81.82%      |
| altavista                  | 7                          | 2.57                            | 00:03:26                    | 100.00%                     | 57.14%      |
| mynet                      | 6                          | 1.33                            | 00:00:07                    | 83.33%                      | 66.67%      |
| alltheweb                  | 1                          | 1.00                            | 00:00:00                    | 100.00%                     | 100.00%     |
| kvaslr                     | 1                          | 1.00                            | 00:00:00                    | 100.00%                     | 100.00%     |
| lycos                      | 1                          | 1.00                            | 00:00:00                    | 100.00%                     | 100.00%     |
| netscape                   | 1                          | 4.00                            | 00:01:11                    | 0.00%                       | 0.00%       |
| yandex                     | 1                          | 3.00                            | 00:01:55                    | 100.00%                     | 0.00%       |

1 - 15 of 15

**APPENDIX D**

**Web-Site Addresses Of The Chosen Programs**

| Analyzed Program (Landing Page Category) | Web-site Addresses of the Chosen Programs                   |    | E     | B     | BR   | PV         |
|--|---|----|-------|-------|------|------------|
| <b>Informatics Online Program</b>        | /academic_program/informatics-online                        | 1  | 5,641 | 2,396 | 42%  | 91,675,643 |
|  | /tr/academic_program/internet-%C3%BCzerinden-bili%C5%9Fim   | 2  | 65    | 24    | 37%  | 184,565    |
|  | /tr/node/38   | 3  | 22    | 8     | 36%  | 56,222     |
|  | /node/38  | 4  | 0     | 0     | NA   | 0          |
|  | /category/social-tags/magazines                             | 5  | 0     | 0     | NA   | 0          |
|  | /category/social-tags/poetry                                | 6  | 0     | 0     | NA   | 0          |
|  | /category/social-tags/technology                            | 7  | 2     | 2     | 100% | 152        |
|  | /category/tags/informatics-online                           | 8  | 21    | 11    | 52%  | 2,821      |
|  | /category/tags/informatics-online-course                    | 9  | 7     | 5     | 71%  | 117        |
|  | /category/tags/internet-%C3%BCzerinden-bili%C5%9Fim         | 10 | 44    | 22    | 50%  | 8,444      |
|  | /category/tags/ion  | 11 | 0     | 0     | NA   | 10         |
|  | /category/tags/masters-degree-online                        | 12 | 19    | 13    | 68%  | 3,019      |
|  | /category/tags/online-course                                | 13 | 0     | 0     | NA   | 0          |
|  | /category/tags/online-informatics-degree                    | 14 | 0     | 0     | NA   | 0          |
|  | /category/tags/online-master-program                        | 15 | 0     | 0     | NA   | 0          |
|  | /category/tags/online-masters-informatics                   | 16 | 0     | 0     | NA   | 0          |
|  | /category/tags/tezsiz-y%C3%BCksek-lisans-ankara             | 17 | 40    | 31    | 78%  | 4,940      |
|  | /category/tags/uzaktan-e%C4%9Fitim-veren-%C3%Bcniversiteler | 18 | 0     | 0     | NA   | 0          |
|  | /category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans       | 19 | 55    | 39    | 71%  | 7,855      |
|  | /tr/category/social-tags/magazines                          | 20 | 0     | 0     | NA   | 0          |
|  | /tr/category/social-tags/poetry                             | 21 | 0     | 0     | NA   | 0          |
|  | /tr/category/social-tags/technology                         | 22 | 2     | 2     | 100% | 32         |
|  | /tr/category/tags/informatics-online                        | 23 | 1     | 1     | 100% | 71         |
|  | /tr/category/tags/informatics-online-course                 | 24 | 11    | 8     | 73%  | 1,611      |
|  | /tr/category/tags/internet-%C3%BCzerinden-bili%C5%9Fim      | 25 | 2     | 1     | 50%  | 102        |
|  | /tr/category/tags/ion                                       | 26 | 12    | 4     | 33%  | 1,612      |
|  | /tr/category/tags/masters-degree-online                     | 27 | 32    | 17    | 53%  | 4,532      |
|  | /tr/category/tags/online-course                             | 28 | 2     | 1     | 50%  | 22         |
|  | /tr/category/tags/online-informatics-degree                 | 29 | 4     | 0     | 0%   | 54         |
|  | /tr/category/tags/online-master-program                     | 30 | 66    | 38    | 58%  | 8,366      |
|  | /tr/category/tags/online-masters-informatics                | 31 | 4     | 2     | 50%  | 74         |

| Analyzed Program (Landing Page Category) | Web-site Addresses of the Chosen Programs                      |    | E            | B            | BR         | PV                |
|--|--|----|--------------|--------------|------------|-------------------|
|  | /tr/category/tags/tezsiz-y%C3%BCksek-lisans-ankara             | 32 | 48           | 29           | 60%        | 6,248             |
|  | /tr/category/tags/uzaktan-e%C4%9Fitim-veren-%C3%BCniversiteler | 33 | 0            | 0            | NA         | 0                 |
|  | /tr/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-lisans       | 34 | 85           | 51           | 60%        | 12,385            |
| <b>Total</b>                             | <b>34 web-pages</b>  |    | <b>6,185</b> | <b>2,705</b> | <b>44%</b> | <b>91,978,897</b> |
| <b>Information Systems Program</b>       | /academic_program/information-systems                          | 1  | 251          | 138          | 55%        | 4,434,249         |
|  | /tr/academic_program/bili%C5%9Fim-sistemleri                   | 2  | 67           | 37           | 55%        | 48,767            |
|  | /node/21   | 3  | 0            | 0            | NA         | 0                 |
|  | /tr/node/21  | 4  | 49           | 19           | 39%        | 100,049           |
|  | /category/tags/alanda  | 5  | 0            | 0            | NA         | 0                 |
|  | /category/tags/bili%C5%9Fim-sistemleri                         | 6  | 5            | 2            | 40%        | 195               |
|  | /category/tags/derin   | 7  | 0            | 0            | NA         | 0                 |
|  | /category/tags/dersler   | 8  | 2            | 1            | 50%        | 22                |
|  | /category/tags/doktora   | 9  | 5            | 2            | 40%        | 85                |
|  | /category/tags/master  | 10 | 4            | 1            | 25%        | 74                |
|  | /category/tags/phd   | 11 | 0            | 0            | NA         | 0                 |
|  | /category/tags/sistemleri                                      | 12 | 0            | 0            | NA         | 0                 |
|  | /category/tags/temel   | 13 | 0            | 0            | NA         | 0                 |
|  | /category/tags/veri  | 14 | 0            | 0            | NA         | 0                 |
|  | /category/tags/y%C3%BCksek-lisans                              | 15 | 7            | 3            | 43%        | 107               |
|  | /category/tags/yeni  | 16 | 0            | 0            | NA         | 0                 |
|  | /tr/category/tags/alanda                                       | 17 | 0            | 0            | NA         | 10                |
|  | /tr/category/tags/bili%C5%9Fim-sistemleri                      | 18 | 2            | 1            | 50%        | 52                |
|  | /tr/category/tags/derin  | 19 | 0            | 0            | NA         | 0                 |
|  | /tr/category/tags/dersler                                      | 20 | 0            | 0            | NA         | 20                |
|  | /tr/category/tags/doktora                                      | 21 | 0            | 0            | NA         | 0                 |
|  | /tr/category/tags/master                                       | 22 | 4            | 4            | 100%       | 74                |
|  | /tr/category/tags/phd  | 23 | 9            | 6            | 67%        | 209               |
|  | /tr/category/tags/sistemleri                                   | 24 | 0            | 0            | NA         | 0                 |
|  | /tr/category/tags/temel  | 25 | 0            | 0            | NA         | 0                 |
|  | /tr/category/tags/veri   | 26 | 0            | 0            | NA         | 0                 |
|  | /tr/category/tags/y%C3%BCksek-lisans                           | 27 | 10           | 4            | 40%        | 1,410             |
|  | /tr/category/tags/yeni   | 28 | 0            | 0            | NA         | 0                 |
| <b>Total</b>                             | <b>28 web-pages</b>  |    | <b>415</b>   | <b>218</b>   | <b>53%</b> | <b>4,585,323</b>  |
| <b>Software</b>                          | /academic_program/software-management                          | 1  | 278          | 79           | 28%        | 2,104,278         |

| Analyzed Program (Landing Page Category)                                       | Web-site Addresses of the Chosen Programs   |    | E          | B          | BR         | PV               |
|--|---|----|------------|------------|------------|------------------|
| <b>Management Program</b>  | /tr/academic_program/yaz%C4%B1%C4%B1m-y%C3%B6netimi   | 2  | 53         | 17         | 32%        | 30,552           |
|  | /node/39  | 3  | 0          | 0          | NA         | 0                |
|  | /tr/node/39   | 4  | 13         | 3          | 23%        | 44,513           |
|  | /category/tags/sm   | 5  | 0          | 0          | NA         | 0                |
|  | /category/tags/software-management  | 6  | 3          | 3          | 100%       | 43               |
|  | /category/tags/yaz%C4%B1%C4%B1m-y%C3%B6netimi   | 7  | 0          | 0          | NA         | 0                |
|  | /tr/category/tags/sm  | 8  | 2          | 0          | 0%         | 52               |
|  | /tr/category/tags/software-management   | 9  | 56         | 15         | 27%        | 6,656            |
|  | /tr/category/tags/yaz%C4%B1%C4%B1m-y%C3%B6netimi  | 10 | 1          | 0          | 0%         | 31               |
| <b>Total</b>   | <b>10 web-pages</b>   |    | <b>406</b> | <b>117</b> | <b>29%</b> | <b>2,186,125</b> |
| <b>Work Based Learning Program</b>   | /departments/department-work-based-learning-studies   | 1  | 283        | 107        | 38%        | 789,283          |
|  | /tr/academic_program/work-based-learning/i%C5%9F-ya%C5%9Fam%C4%B1-temelli-%C3%B6%C4%9Frenme | 2  | 7          | 3          | 43%        | 1,397            |
|  | /node/46  | 3  | 0          | 0          | NA         | 0                |
|  | /tr/node/46   | 4  | 32         | 16         | 50%        | 15,132           |
|  | /category/tags/i%C5%9F-ya%C5%9Fam%C4%B1-temelli-%C3%B6%C4%9Frenme                           | 5  | 23         | 19         | 83%        | 6,523            |
|  | /category/tags/middlesex-university   | 6  | 84         | 56         | 67%        | 11,784           |
|  | /category/tags/wbl  | 7  | 6          | 4          | 67%        | 86               |
|  | /category/tags/work-based-learning-studies  | 8  | 29         | 17         | 59%        | 6,729            |
|  | /tr/category/tags/i%C5%9F-ya%C5%9Fam%C4%B1-temelli-%C3%B6%C4%9Frenme                        | 9  | 3          | 3          | 100%       | 273              |
|  | /tr/category/tags/middlesex-university  | 10 | 16         | 10         | 63%        | 2,916            |
|  | /tr/category/tags/wbl   | 11 | 4          | 0          | 0%         | 54               |
|  | /tr/category/tags/work-based-learning-studies   | 12 | 3          | 1          | 33%        | 63               |
| <b>Total</b>   | <b>12 web-pages</b>   |    | <b>490</b> | <b>236</b> | <b>48%</b> | <b>834,240</b>   |
| <b>Informatics Online, Information Systems and Software Management Program</b> | /category/tags/   | 1  | 0          | 0          | NA         | 0                |
|  | /category/tags/information-systems  | 2  | 32         | 16         | 50%        | 5,132            |
|  | /tr/category/tags   | 3  | 3          | 2          | 67%        | 53               |
|  | /tr/category/tags/information-systems   | 4  | 12         | 9          | 75%        | 2,312            |
| <b>Total</b>   | <b>4 web-pages</b>  |    | <b>47</b>  | <b>27</b>  | <b>57%</b> | <b>7,497</b>     |

## **APPENDIX E      Code Used To Get The Report**

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package googleanalyticsproject;
import com.google.gdata.client.analytics.AnalyticsService;
import com.google.gdata.client.analytics.DataQuery;
import com.google.gdata.data.analytics.AccountFeed;
import com.google.gdata.data.analytics.DataEntry;
import com.google.gdata.data.analytics.DataFeed;
import java.io.BufferedOutputStream;
import java.io.FileOutputStream;
import java.io.OutputStream;
import java.io.OutputStreamWriter;
import java.io.Writer;
import java.net.URL;
public class Main {
    /**
     * @param args the command line arguments
     */
    // write your google account e-mail address
    private static final String CLIENT_USERNAME = "*****";
// write your google account password
    private static final String CLIENT_PASS = "*****";
    private static final String TABLE_ID = "ga:4784817";
    private static String dataFile = "myresults.txt";
    public static void main(String[] args) {
        // TODO code application logic here
        try{
            AnalyticsService analyticsService = new AnalyticsService("gaExportAPI_acctSample_v1.0");
            analyticsService.setUserCredentials(CLIENT_USERNAME, CLIENT_PASS);
            // Construct query from a string.
            URL queryUrl = new URL(
                "https://www.google.com/analytics/feeds/accounts/default?max-results=50");
            // Make request to the API, using AccountFeed class as the second parameter.
            AccountFeed accountFeed = analyticsService.getFeed(queryUrl, AccountFeed.class);
            // Create a query using the DataQuery Object.
            DataQuery query = new DataQuery(new URL(
                "https://www.google.com/analytics/feeds/data"));
            // Output data to the screen.
            FileOutputStream fos=new FileOutputStream(dataFile);
            OutputStream bout = new BufferedOutputStream(fos);
            Writer ort = new OutputStreamWriter(bout,"Cp1254");
            ////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
            //While importing data to Excel 1254 Turkish (Windows) needs to be selected
            //Additionally, during data import text qualifier " needs to be set as disabled and the first column
//needs to be selected as "text".

```

```

////////////////////////////////////

dataFile = "myresults1.txt";
query = new DataQuery(new URL(
    "https://www.google.com/analytics/feeds/data"));
query.setStartDate("2009-09-01");
query.setEndDate("2010-03-21");
query.setDimensions("ga:keyword,ga:landingPagePath ");
query.setMetrics("ga:bounces,ga:entrances, ga:pageviews, ga:uniquePageviews");
query.setSort("ga:keyword");
query.setFilters("ga:keyword!=(not set);ga:landingPagePath==/academic_program/informatics-
online,ga:landingPagePath==/academic_program/information-
systems,ga:landingPagePath==/academic_program/software-
management,ga:landingPagePath==/category/social-
tags/magazines,ga:landingPagePath==/category/social-tags/poetry,ga:landingPagePath==/category/social-
tags/technology,ga:landingPagePath==/category/tags,ga:landingPagePath==/category/tags/alanda,ga:landi
ngPagePath==/category/tags/bili%C5%9Fim-sistemleri,ga:landingPagePath==/category/tags/bili%C5%9Fim-
sistemleri,ga:landingPagePath==/category/tags/derin,ga:landingPagePath==/category/tags/dersler,ga:landi
ngPagePath==/category/tags/doktora,ga:landingPagePath==/category/tags/i%C5%9F-
ya%C5%9Fam%C4%B1-temelli-%C3%B6%C4%9Frenme,ga:landingPagePath==/category/tags/i%C5%9F-
ya%C5%9Fam%C4%B1-temelli-%C3%B6%C4%9Frenme,ga:landingPagePath==/category/tags/informatics-online");
    query.setMaxResults(40000);
    query.setIds(TABLE_ID);
    DataFeed dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    // Make a request to the API, using DataFeed class as the second parameter.
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    // Output data to the screen.
    fos=new FileOutputStream(dataFile);
    bout = new BufferedOutputStream(fos);
    ort = new OutputStreamWriter(bout,"Cp1254");
    ort.write("ga:keyword \t ga:landingPagePath \t ga:bounces \t ga:entrances \t ga:pageviews \t
ga:uniquePageviews \n");
    System.out.println("----- Data Feed Results -----");
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa
lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t"+
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+" \n");
        ort.write(retrievedLine);
    }
    query.setFilters("ga:keyword!=(not set);ga:landingPagePath==/category/tags/informatics-online-
course,ga:landingPagePath==/category/tags/information-
systems,ga:landingPagePath==/category/tags/internet-%C3%BCzerinden-
bili%C5%9Fim,ga:landingPagePath==/category/tags/internet-%C3%BCzerinden-
bili%C5%9Fim,ga:landingPagePath==/category/tags/ion,ga:landingPagePath==/category/tags/master,ga:landi
ngPagePath==/category/tags/masters-degree-online,ga:landingPagePath==/category/tags/middlesex-
university");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa

```

```

lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t"+
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+"\n");
    ort.write(retrievedLine);
}
    query.setFilters("ga:keyword!=(not set);ga:landingPagePath==/category/tags/software-
management,ga:landingPagePath==/category/tags/temel,ga:landingPagePath==/category/tags/tezsiz-
y%C3%BCksek-lisans-
ankara,ga:landingPagePath==/category/tags/temel,ga:landingPagePath==/category/tags/tezsiz-yüksek-
lisans-ankara,ga:landingPagePath==/category/tags/uzaktan-e%C4%9Fitim-veren-
%C3%Bcni-versiteler,ga:landingPagePath==/category/tags/uzaktan-eğitim-veren-
üniversiteler,ga:landingPagePath==/category/tags/uzaktan-e%C4%9Fitim-y%C3%BCksek-
lisans,ga:landingPagePath==/category/tags/uzaktan-eğitim-yüksek-
lisans,ga:landingPagePath==/category/tags/veri,ga:landingPagePath==/category/tags/wbl");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa
lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t"+
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+"\n");
        ort.write(retrievedLine);
    }
    query.setFilters("ga:keyword!=(not set);ga:landingPagePath==/category/tags/work-based-learning-
studies,ga:landingPagePath==/category/tags/y%C3%BCksek-
lisans,ga:landingPagePath==/category/tags/yüksek-
lisans,ga:landingPagePath==/category/tags/yaz%C4%B1%C4%B1m-
y%C3%B6netimi,ga:landingPagePath==/category/tags/yazılım-
yönetimi,ga:landingPagePath==/category/tags/yeni,ga:landingPagePath==/departments/department-work-
based-learning-studies,ga:landingPagePath==/node/21,ga:landingPagePath==/node/38");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa
lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t"+
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+"\n");
        ort.write(retrievedLine);
    }
    query.setFilters("ga:keyword!=(not
set);ga:landingPagePath==/node/39,ga:landingPagePath==/node/46,ga:landingPagePath==/tr/academic_pr
ogram/bilişim-sistemleri,ga:landingPagePath==/tr/academic_program/bili%C5%9Fim-
sistemleri,ga:landingPagePath==/tr/academic_program/internet-%C3%BCzerinden-
bili%C5%9Fim,ga:landingPagePath==/tr/academic_program/internet-üzerinden-
bilgişim,ga:landingPagePath==/tr/academic_program/work-based-learning/i%C5%9F-
ya%C5%9Fam%C4%B1-temelli-
%C3%B6%C4%9Frenme,ga:landingPagePath==/tr/academic_program/work-based-learning/iş-yaşamı-
temelli-öğrenme,ga:landingPagePath==/tr/academic_program/yaz%C4%B1%C4%B1m-
y%C3%B6netimi,ga:landingPagePath==/tr/academic_program/yazılım-
yönetimi,ga:landingPagePath==/tr/category/social-tags/magazines");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa

```

```

lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t"+
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+"\n");
    ort.write(retrievedLine);
}
    query.setFilters("ga:keyword!=(not set);ga:landingPagePath==/tr/category/social-
tags/poetry,ga:landingPagePath==/tr/category/social-
tags/technology,ga:landingPagePath==/tr/category/tags,ga:landingPagePath==/tr/category/tags/alanda,ga:l
andingPagePath==/tr/category/tags/bili%C5%9Fim-
sistemleri,ga:landingPagePath==/tr/category/tags/biliřim-
sistemleri,ga:landingPagePath==/tr/category/tags/derin,ga:landingPagePath==/tr/category/tags/dersler");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa
lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t"+
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+"\n");
        ort.write(retrievedLine);
    }
    query.setFilters("ga:keyword!=(not
set);ga:landingPagePath==/tr/category/tags/doktora,ga:landingPagePath==/tr/category/tags/i%C5%9F-
ya%C5%9Fam%C4%B1-temelli-%C3%B6%C4%9Frenme,ga:landingPagePath==/tr/category/tags/iř-
yařamı-temelli-öğrenme,ga:landingPagePath==/tr/category/tags/informatics-
online,ga:landingPagePath==/tr/category/tags/informatics-online-
course,ga:landingPagePath==/tr/category/tags/information-
systems,ga:landingPagePath==/tr/category/tags/internet-%C3%BCzerinden-
bili%C5%9Fim,ga:landingPagePath==/tr/category/tags/ion");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa
lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t"+
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+"\n");
        ort.write(retrievedLine);
    }
    query.setFilters("ga:keyword!=(not
set);ga:landingPagePath==/tr/category/tags/master,ga:landingPagePath==/tr/category/tags/masters-degree-
online,ga:landingPagePath==/tr/category/tags/middlesex-
university,ga:landingPagePath==/tr/category/tags/online-
course,ga:landingPagePath==/tr/category/tags/online-informatics-
degree,ga:landingPagePath==/tr/category/tags/online-master-
program,ga:landingPagePath==/tr/category/tags/online-masters-informatics");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa
lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t"+
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+"\n");
        ort.write(retrievedLine);
    }
    query.setFilters("ga:keyword!=(not
set);ga:landingPagePath==/tr/category/tags/phd,ga:landingPagePath==/tr/category/tags/sistemleri,ga:landi

```

```

ngPagePath==/tr/category/tags/sm,ga:landingPagePath==/tr/category/tags/software-
management,ga:landingPagePath==/tr/category/tags/temel,ga:landingPagePath==/tr/category/tags/tezsiz-
y%C3%BCksek-lisans-ankara,ga:landingPagePath==/tr/category/tags/tezsiz-yüksek-lisans-
ankara,ga:landingPagePath==/tr/category/tags/uzaktan-e%C4%9Fitim-veren-
%C3%Bcniversiteler,ga:landingPagePath==/tr/category/tags/uzaktan-eđitim-veren-üniversiteler");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa
lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t" +
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+" \n");
        ort.write(retrievedLine);
    }
    query.setFilters("ga:keyword!=(not set);ga:landingPagePath==/tr/category/tags/uzaktan-
e%C4%9Fitim-y%C3%BCksek-lisans,ga:landingPagePath==/tr/category/tags/uzaktan-eđitim-yüksek-
lisans,ga:landingPagePath==/tr/category/tags/veri,ga:landingPagePath==/tr/category/tags/wbl,ga:landingP
agePath==/tr/category/tags/work-based-learning-
studies,ga:landingPagePath==/tr/category/tags/y%C3%BCksek-
lisans,ga:landingPagePath==/tr/category/tags/yüksek-
lisans,ga:landingPagePath==/tr/category/tags/yaz%C4%B1l%C4%B1m-
y%C3%B6netimi,ga:landingPagePath==/tr/category/tags/yazılım-
yönetimi,ga:landingPagePath==/tr/category/tags/yeni");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa
lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t" +
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+" \n");
        ort.write(retrievedLine);
    }
    query.setFilters("ga:keyword!=(not
set);ga:landingPagePath==/tr/node/21,ga:landingPagePath==/tr/node/38,ga:landingPagePath==/tr/node/39
,ga:landingPagePath==/tr/node/46");
    dataFeed = analyticsService.getFeed(query.getUrl(), DataFeed.class);
    for (DataEntry entry : dataFeed.getEntries()) {
        String retrievedLine = new String
(entry.stringValueOf("ga:keyword")+ "\t"+entry.stringValueOf("ga:landingPagePath")+ "\t"+entry.stringVa
lueOf("ga:bounces")+ "\t"+entry.stringValueOf("ga:entrances")+ "\t" +
entry.stringValueOf("ga:pageviews")+ entry.stringValueOf("ga:uniquePageviews ")+" \n");
        ort.write(retrievedLine);
    }
    ort.close();
    fos.close();
}
catch (Exception ex) {
    ex.printStackTrace();
}
}
}
}

```

## **APPENDIX F**      **Think-Aloud Study Questions**

### DEMOGRAPHIC QUESTIONS

**QUE 1.** What is your purpose for visiting this page? \*

- To get information about the programmes opened by the Institute
- Merely surfing, browsing the web in the leisure time
- Directed by the survey
- Searching for a specified information
- For work purposes
- Other

**QUE 2.** What is your Last degree (obtained or in preparation)?: \*

- Undergraduate
- Master of Science
- Doctorate

**QUE 3.** What is your Faculty that you study at? Please choose from the list. \*

**QUE 4.** Please select the option/options below that best suits for you. Language of education:: \*

In my master and doctorate studies were English

In my university was English.

In my high school was English.

Other

\*Compulsory questions

**MAIN SURVEY PART 1:**

Firstly please, open the Institute web-site at a new window by clicking on this link [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr).

**TASK 1**

( You are supposed to find the answer from our institute web-site [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr) )

Imagine that you decide to do a doctorate study. Among the programs Information Systems IS, Software Management SM and Informatics-OnLine ION, which one/ones offer/s doctorate study? Now, please attempt to find the answer. :

- IS
- SM and ION
- All
- IS and SM

Please mark the hardness level that best suits for you:

1      2      3      4      5

Perceived hardness level

1:very easy & 5: very hard

If it was hard to find, why? Where did you think the information should be?:

## TASK 2

( You are supposed to find the answer from our institute web-site [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr) )

Imagine that you look for a NON-thesis master programme. Assume that you live and work in Muğla. You were graduated from economics department. You DO NOT have any experience in software industry

Which of the programs among Information Systems IS, Software Mangement SM and Informatics-OnLine ION are you eligible for application? Now, please attempt to find the answer.:

- IS
- SM
- ION
- SM and ION

Please mark the hardness level that best suits for you:

1 2 3 4 5

|                          |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Perceived hardness level | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|

1:very easy & 5: very hard

If it was hard to find, why? Where did you think the information should be?:

### TASK 3

( You are supposed to find the answer from our institute web-site [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr) )

Which of the programs among Information Systems IS, Software Mangement SM and Informatics-OnLine ION are evening education programs? Now, please attempt to find the answer of this question.:

IS

SM

ION

SM and ION

Please mark the hardness level that best suits for you:

1      2      3      4      5

Perceived hardness level

1:very easy & 5: very hard

If it was hard to find, why? Where did you think the information should be?:

#### TASK 4

( You are supposed to find the answer from our institute web-site [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr) )

Imagine that you were graduated from Electrical and Electronics Engineering Department but you are working as a software programmer in a company for two years.

You are interested in Software Management (SM) programme. Now, please attempt to find whether you are eligible for this program or not by looking at our application requirements.:

Yes

No

Please mark the hardness level that best suits for you:

1 2 3 4 5

Perceived hardness level

1:very easy & 5: very hard

If it was hard to find, why? Where did you think the information should be?:

**MAIN SURVEY PART 2: TASK 5**

(You are supposed to find the answer from our institute web-site [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr) )

Imagine that you are accepted for a program at Informatics Institute and you are now at the course registration period. You need to choose the courses that you will take this semester.

Now please attempt to find the elective courses of Work Based Learning WBL Programme. Choose the courses from the list:

Please mark the hardness level that best suits for you:

**1      2      3      4      5**

Perceived hardness level

1:very easy & 5: very hard

If it was hard to find, why? Where did you think the information should be?:

**TASK 6**

( You are supposed to find the answer from our institute web-site [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr) )

Now please attempt to find the courses that were opened in 2008-2009 Fall Semester. Provide the web

page link that contains the answers below.:

Please mark the hardness level that best suits for you:

**1      2      3      4      5**

Perceived hardness level

1:very easy & 5: very hard

If it was hard to find, why? Where did you think the information should be?:

## TASK 7

( You are supposed to find the answer from our institute web-site [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr) )

Imagine you are accepted for a program which includes a Thesis Study at Informatics Institute. You decided on your thesis subject.

You are supposed to have an advisor. Please attempt to find the faculty member who participates into two research groups about Software Technologies and Data Mining.:

Please mark the hardness level that best suits for you:

1 2 3 4 5

|                          |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Perceived hardness level | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|

1:very easy & 5: very hard

If it was hard to find, why? Where did you think the information should be?:

## TASK 8

( You are supposed to find the answer from our institute web-site [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr) )

Imagine you are about to finish your master studies this semester at Informatics Institute. You need information on the procedure of Thesis Study.

Please attempt to find information on what you need to do before your thesis defense. Copy the web page link (or the document link containing the information) where you have found the information below.:

Please mark the hardness level that best suits for you:

1 2 3 4 5

|                          |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Perceived hardness level | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|

1:very easy & 5: very hard

If it was hard to find, why? Where did you think the information should be?:

**TASK 9**

( You are supposed to find the answer from our institute web-site [www.ii.metu.edu.tr](http://www.ii.metu.edu.tr) )

Please attempt to find information on the maximum time limit you can use to do the minor corrections on your thesis study pointed out by your Jury and Informatics Institute.:

1 week

3 weeks

1 month

6 weeks

Please mark the hardness level that best suits for you:

1      2      3      4      5

Perceived hardness level

1:very easy & 5: very hard

If it was hard to find, why? Where did you think the information should be?:



5: 1 2 3 4 5

There is little waiting time for the Web pages to load.

1: Strongly Disagree  
5: Strongly Agree

6: 1 2 3 4 5

Users receive timely responses to their queries / submissions.

1: Strongly Disagree  
5: Strongly Agree

7: Never Before Once Every Several Months Once Every Several Weeks Once a Week More Than Once a Week

How often have you visited Informatics Institute Website?

8: 1 2 3 4 5

Navigation of the Web site was simple and easy.

1: Strongly Disagree  
5: Strongly Agree

9: 1 2 3 4 5

Link density provides clarity and easy navigation

1: Strongly Disagree

5: Strongly Agree

10: 1 2 3 4 5

Interacting with the Web site was NOT easy.

1: Strongly Disagree

5: Strongly Agree

11: 1 2 3 4 5

The website provides visitors with an engaging and memorable experience

1: Strongly Disagree

5: Strongly Agree

12: 1 2 3 4 5

I did NOT feel comfortable using this web site for these tasks.

1: Strongly Disagree

5: Strongly Agree

**13:** 1 2 3 4 5

The website helps its visitors accomplish common goals and tasks

1: Strongly Disagree  
5: Strongly Agree

**14:** 1 2 3 4 5

Content structured in a way that facilitates the attainment of user goals

1: Strongly Disagree  
5: Strongly Agree

**15:** 1 2 3 4 5

I felt that I had the freedom to go anywhere in the Web site

1: Strongly Disagree  
5: Strongly Agree

**16:** 1 2 3 4 5

It was NOT easy to find the required information on the web site.

1: Strongly Disagree  
5: Strongly Agree

17: 1 2 3 4 5

The organisation of information on the web site was clear to me.

1: Strongly Disagree

5: Strongly Agree

18: 1 2 3 4 5

Content is up-to-date and accurate

1: Strongly Disagree

5: Strongly Agree

19: 1 2 3 4 5

Content is appropriate to customer needs and business goals

1: Strongly Disagree

5: Strongly Agree

20: 1 2 3 4 5

Content across multiple languages is NOT comprehensive

1: Strongly Disagree

5: Strongly Agree

21:

1 2 3 4 5

The website prevents errors and helps the user  
recover from them

1: Strongly Disagree

5: Strongly Agree

22:

1 2 3 4 5

How often did you use “Search” service available  
on the web-site?

1: Never

5: Always

23:

1 2 3 4 5

How useful was the “Search” service available on  
the web-site?

1: Never

5: Always

24: I could perform better using this website, If I had navigated similar web pages before this page:

Yes

No

If Yes,: **1 2 3 4 5**

Please mark your level of confidence

- 1 Not at all confident
- 3 Moderately confident
- 5 Totally confident

25:

I could perform better using this website, If someone else had helped me get started.:

Yes

No

If Yes, : **1 2 3 4 5**

Please mark your level of confidence

- 1 Not at all confident
- 3 Moderately confident
- 5 Totally confident

26:

I could perform better using this website, If I had a lot of time to accomplish the tasks.:

Yes

No

If Yes, :

**1      2      3      4      5**

Please mark your level of confidence

1 Not at all confident

3 Moderately confident

5 Totally confident

27: I could perform better using this website, If my English was better.:

Yes

No

If Yes, :

**1      2      3      4      5**

Please mark your level of confidence

1 Not at all confident

3 Moderately confident

5 Totally confident

28:

1 2 3 4 5

Please rate your computer literacy level

1: Basic

5: Excellent

29:

1 2 3 4 5

Overall I was satisfied with this web site.

1: Strongly Disagree

5: Strongly Agree

30: What did you like most about our Institute web site ?:

31: What did you like least about our Institute web site ?:

**32: Other Comments:**

