Developing a Framework for Web Analytics

A thesis submitted for the

Master of Science in Information Management

(University of Koblenz-Landau)

By

Verena Hausmann

Student ID: 206210057

vhausmann@uni-koblenz.de

Faculty 4: Computer Science

Institute for IS Research

University of Koblenz-Landau, Germany

Supervisors:

Prof. Dr. Petra Schubert,

Prof. Dr. Susan P. Williams

March 2012

Declaration/Erklärung

I declare that:

This thesis presents work carried out by myself and does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university. To the best of my knowledge, it does not constitute any previous work published or written by another person except where due reference is made in the text.

Ich versichere,

dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden. Der Veröffentlichung dieser Arbeit im Internet stimme ich zu.

Verena Hausmann

March 2012

Abstract

The World Wide Web (WWW) has become a very important communication channel. Its usage has steadily grown within the past. Interest by website owners in identifying user behaviour has been around since Tim Berners-Lee developed the first web browser in 1990. But as the influence of the online channel today eclipses all other media the interest in monitoring website usage and user activities has intensified as well. Gathering and analysing data about the usage of websites can help to understand customer behaviour, improve services and potentially increase profit. It is further essential for ensuring effective website design and management, efficient mass customization and effective marketing. Web Analytics (WA) is the area addressing these considerations.

However, changing technologies and evolving Web Analytic methods and processes present a challenge to organisations starting with Web Analytic programmes. Because of lacking resources in different areas and other types of websites especially small and medium-sized enterprises (SME) as well as non-profit organisations struggle to operate WA in an effective manner.

This research project aims to identify the existing gap between theory, tool possibilities and business needs for undertaking Web Analytic programmes. Therefore the topic was looked at from three different ways: the academic literature, Web Analytic tools and an interpretative case study. The researcher utilized an action research approach to investigate Web Analytics presenting an holistic overview and to identify the gaps that exists. The outcome of this research project is an overall framework, which provides guidance for SMEs who operate information websites on how to proceed in a Web Analytic programme.

Keywords:
Web Mining, Web Analytics, Web Analytics Framework, Tool Evaluation, Criteria Matrix

Das World Wide Web (WWW) ist heute zu einem sehr wichtigen Kommunikationskanal geworden, dessen Nutzung in der Vergangenheit stetig gestiegen ist. Websitebesitzer haben schon seit der Entwicklung des ersten Web Browser von Tim Berners-Lee im Jahre 1990 Interesse daran, das Verhalten von Benutzern zu erkennen und zu verstehen. Durch den Einfluss, den der Onlinekanal heute erzielt und welcher alle anderen Medien übersteigt, ist auch das Interesse im Beobachten von Website-Benutzungen und -Benutzeraktivitäten noch weiter gestiegen. Das Sammeln und Analysieren von Daten über die Benutzung von Websites kann helfen, Benutzerverhalten zu verstehen, Services zu verbessern und Gewinn zu steigern. Darüber hinaus ist es Voraussetzung für effektives Website Design und Management, effektive Mass Customization und effektives Marketing. Das Themengebiet, welches diese Aspekte betrachtet, heißt Web Analytics (WA).

Allerdings führen sich ändernde Technologien und sich entwickelnde Web Analytics Methoden und Prozesse zu großen Herausforderungen für Unternehmen, die Web Analytic Programme durchführen. Aufgrund fehlender Ressourcen in den verschiedensten Bereichen, sowie einer hier oft aufzufindenden anderen Art von Websites, ist es vor allem für Klein- und Mittelständige Unternehmen (KMU) sowie Non-Profit Organisationen schwer, Web Analytics in einer effektiven Weise zu betreiben.

Dieses Forschungsvorhaben zielt daher darauf ab, die vorhandene Lücke zwischen der Theorie, den Möglichkeiten die Tools bieten und den betrieblichen Anforderungen zu identifizieren. Hierfür wird das Thema von drei verschiedenen, jedoch aufeinander aufbauenden Richtungen betrachtet: der akademischen Literatur, Web Analytic Programmen und einer Fallstudie. Mit Hilfe eines Action Research Ansatzes war es möglich, einen ganzheitlichen Überblick des Themengebiets Web Analytics zu erhalten und bestehende Lücken aufzudecken. Das Ergebnis dieser wissenschaftlichen Arbeit ist ein Framework, welches KMUs die Informations-Websites betreiben dabei hilft, Web Analytic Initiativen durch zu führen.

Keywörter:
Web Mining, Web Analytics, Web Analytics Framework, Tool Evaluation, Criteria Matrix

Table of Contents

List of Figures

List of Tables

Glossary

ASP	Application Service Provide
EMBER	Effective Marketing for Business in European Regions
IS	Information Systems
KPI	Key Performance Indicator
P2T	Passport to Trade
RQ	Research Question
SaaS	Software-as-a-Service
SME	Small and Medium-sized Enterprises
WA	Web Analytics
WWW	World Wide Web

1 Introduction

Web Mining or Web Analytics can be described as the discovery and analysis of useful information from the World Wide Web (Cooley et al. 1997). This first chapter provides a brief introduction to the reasons for Web Mining Usage and this research investigation. It thereby starts with an outline of the problem statement which explains the motivation that led to this research project in Section 1.1. It further describes the research aim and objective in Section 1.2 before presenting the research questions that guide the whole work (Section 1.3). At the end of this chapter in Section 1.4 a complete overview of the thesis is presented.

1.1 Problem Statement

Within recent years the use of the Internet by companies as well as private households has increased dramatically. In the past, advertisements on TV were very important parts of marketing communication and have been a key marketing channel. The TV as a marketing channel is still important, but the time people spend on the Internet has already overtaken the amount of time people spend watching TV. As the influence of the online channel exceeds all other media (Lovett et al. 2009) the Internet and its websites are becoming more and more important. For personal use websites serve for example as an information source, a general store and also as a place to stay connected with friends. Companies and organisations also use websites to inform people about their company, products and services, selling goods and marketing. They use the Internet as a strategic channel to their customers (Gassman 2009a, p.2). With the wide spread adaption of the Internet and its increasing use, the Internet today has become a very important communication channel.

In order to use the web more effectively organisations are increasingly gathering data about the usage of their websites to understand customer behaviour (Gassman 2009a, p.2). The interest in using usage data generated from the web and transferring it into Business Intelligence[1] (BI) can help them to understand their customers, improve their

[1] The Term Business Intelligence is very broad. It represents the data and knowledge gathering, storage and management in order to present complex internal and competitive information to planners and decision makers (Negash 2004).

services and thereby increase profit. Even though businesses now have great interest in Web Mining, the theory supporting Web Mining is quite fragmented and not user focused. The first papers around that topic were published in the mid 1990s and by 2001 the number of papers published already exceeded 150. This shows the great interest in Web Mining from the beginning (Facca & Lanzi 2005, p.226). The whole topic around the Internet and Web Mining in particular represents a complex and changing landscape. More and more new and different technologies are available that offer opportunities for even more applications and the number and diversity of analytic tools continues to grow (Herschel 2010, p.4). One problem is that the differences between all the WA tools often are not clear. And also the literature in this area is quite broad and complex. Papers mainly describe advantages and limitations of specific algorithms for analysis and the development of improved methods for Web Mining. Many different methods of how to conduct Web Mining are discussed, but no integrated view is given. Little attention is directed towards the user and organisational aspects of Web Analytics. But they are faced with significant challenges to identify which tool to use, understanding the differences between tools and metrics, and what must be considered when undertaking Web Analytic initiatives. Even though access to tools is available, "*unfortunately, many organisations are not getting the most value out of their initiatives*" (Gassman 2009). Especially small and medium-sized enterprises (SMEs) as well as non-profit organizations which operate small websites can easily be overwhelmed. They often do not have the resources concerning money and staff to operate Web Analytics in a large scale manner. It seems that it is even harder for them to describe and measure whether their site is successful or not compared to a large e-commerce site.

In order to help these businesses to make sense of the wider fields of Web Mining this thesis seeks to bridge the rift that exists in these two data sources: the needs of practice or business and the research literature. The aim and the objective that motivate this project are further outlined in the next section.

1.2 Scope of this Research Project

This thesis aims to broaden the focus of academic research by addressing the wider organisational and user issues and providing greater guidance towards selecting appropriate tools and methods for Web Analytics. It further seeks to assist organisations to implement an effective process for Web Mining. Special attention is thereby given to small businesses that operate information websites. The scope is not to give detailed

guidelines for the Web Usage Mining, but to provide an overview of current practices. On this basis a framework for Web Usage Mining is developed that guides the development and maintaining of a reliable, usable and sustainable Web Analytic programme. It should assist organisations to gain a clear overview of what business relevant aspects need to be addressed. Therefore three different research objectives can be distinguished.

- First, to understand the current theory and to identify key trends in Web Analytic methods, metrics and techniques that exist, a literature review, analysis and synthesis will be conducted.
- In a second step analytic tools will be investigated and evaluated in order to get an overview of the existing functionalities and scopes such tools have and the type of methods they support.
- The third objective brings together the knowledge from the previous described steps. It investigates the issues, challenges and stages in Web Analytic implementations. From this, a framework for the practice that will assist organisations to implement an effective process for Web Mining will be developed.

A case website will be used in order to exemplify the issues and guide the recommendations for the future application. Different types of websites have different goals. Therefore it will be analysed to see if there is a need to differentiate the applicability of methods and metrics according to a site. An appropriate way for visualising the process e.g. a tree structure or a diagram will be designed to give an even more understandable way to familiarise organisational users with the topic.

The investigation will focus on the usage data which can be gathered from the web and its analysis for better website usability in terms of reducing click depth and finding information more easily. Figure 1 shows the outcomes which this thesis will deliver.

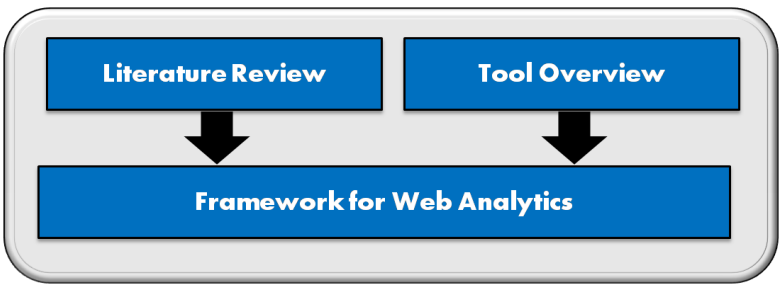


Figure 1: Outcomes of this Research Project

1.3 Research Questions

To reach the described objectives the following research question was designed:

RQ: What is required in a framework to support small and medium-sized enterprises (SME) to provide an easy entry point for undertaking Web Analytic programmes?

Because of its complexity the over-arching question cannot be answered as it is. Therefore, the following subsidiary research questions were developed which guided the whole project.

RQ1a: How can Web Mining be classified?
RQ1b: What different methods for conducting Web Mining are described in the academic literature?
RQ1c: Which data capturing method is the best to use today?

RQ2a: What does the Web Analytic market look like?
RQ2b: What are the best Web Analytic tools to be used?
RQ2c: Which criteria can be used to describe and differentiate these tools?

RQ3a: Is there a need for distinguishing the kind of website for what Web Analytics is used in order get clear results?
RQ3b: Are there differences between Web Mining for small or big businesses that need to be taken into account?

The questions are grouped into three categories dealing with (1) the literature, (2) the tools and (3) the framework development. Each category represents one stage in this thesis which was designed to answer the overall question.

1.4 Outline of this Thesis

This thesis consists of 9 chapters that describe the background and the research methodology, dive in deeper into the concepts of Web Analytics and provide a conclusion to the above outlined research questions. This section briefly describes the content of those chapters. An overview can be seen in Figure 2.

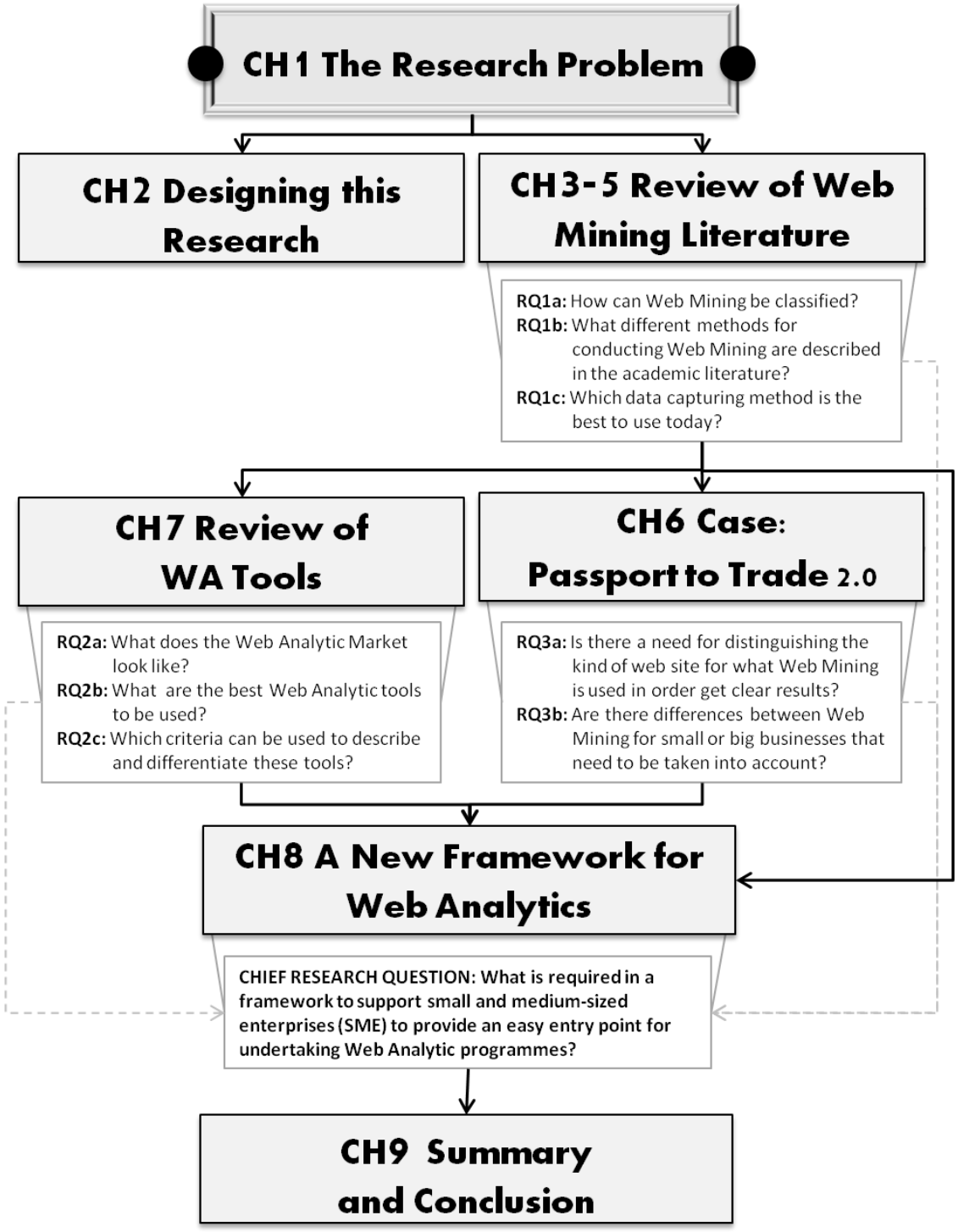


Figure 2: Overview of Research Project

Chapter 2 gives an overview of the research methodologies used in Information Systems and classifies this thesis. It further presents the research design and discusses its limitations.

Chapter 3-5 review existing literature in the areas of websites, Web Mining in general and Web Analytics in particular. They represent the theoretical body of this thesis and provide the basis for the framework development in Chapter 8.

Chapter 6 gives an overview of the Web Analytic market and existing tools. Here a criteria matrix is being developed that should assist in the process of choosing a tool.

Chapter 7 exemplarily shows the conduction of Web Analytics on a real world project.

Chapter 8 finally brings together the previously obtained findings. On their basis a process framework for Web Analytic investigations is developed and described.

Chapter 9 is the final chapter and summarises the findings of this research study and gives recommendations for possible future work.

2 Designing this Research

Within the research literature the terminology for different terms such as methodologies and methods is quite confusing and not consistent. Often the same terms are used in different ways and it cannot be said that there is one single correct way (Crotty 1998, p.1). For example, a case study can be a methodology or a method, depending on the view of the individual researcher. Most commonly a distinction is made between the ontology, epistemology, theoretical perspective, methodology and methods used. Figure 3 shows how the different building blocks of research are connected with each other.

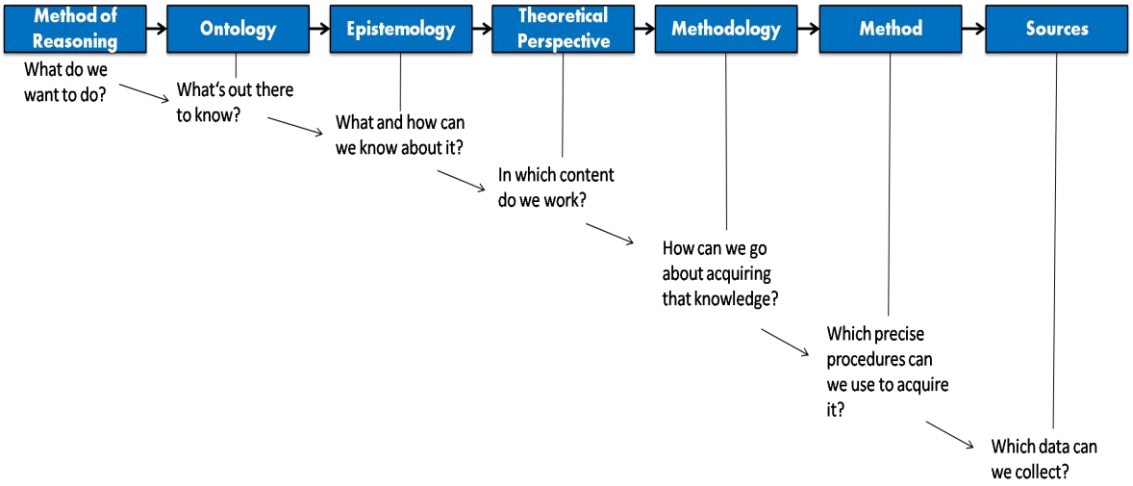


Figure 3: The Building Blocks of Research (adapted from Hay 2002, p.64)

After defining the method of reasoning which asks about what someone wants to do in general the ontology needs to be defined. "*Ontology is the study of being*" (Crotty 1998, p.10) and is concerned with the reality and the question of 'what is'. The knowledge of that reality, thus what it means to know is called epistemology (Crabtree & Miller 1999; Crotty 1998). Together the ontology and epistemology guide the theoretical perspective which build the philosophical stance and provides a context for the process (Crotty 1998). Going further the ways of getting to know about that reality is the methodology (Crabtree & Miller 1999) and the approaches used to get the data are the methods (Crotty 1998). Within social science research there is no one individual and correct way of how to do research (Neuman 2003, p.90), but it is important to be clear about what to use in one's own research project and to be able to justify the choice. Thereby the justification should result from the developed research questions (Crotty 1998, p.2).

This chapter shortly discusses the strategies used in Information System (IS) research in general and therefore outlines the methods of reasoning (Section 2.1), the possible theoretical perspectives (Section 2.3) as well as the methodologies (Section 2.3) and methods (Section 2.4) which can be used. Section 2.5 presents the research design of this thesis and a discussion about the limitations of this research project in Section 2.6 concludes the chapter.

2.1 Methods of Reasoning

At the very beginning of each research project it can be asked which method of reasoning was followed. Generally two broad ones are separated: deductive and inductive (Trochim 2006; Neuman 2006; Blaikie 2000). Deductive reasoning, also called the 'top-down' approach begins with an abstract theory. It works from the more general to the more specific. It is used for theory testing, as at the beginning a theory is developed, which leads to a more specific hypothesis which then is confirmed (or not) through data collection (Trochim 2006; Lee 1999; Neuman 2006).

In contrast inductive reasoning can be seen as the 'bottom-up' approach (Trochim 2006). It starts the other way around. Here observations of the reality are conducted before abstract ideas are generated and general conclusions or theories are developed at the end (Trochim 2006; Neuman 2006). Within the inductive approach knowledge is built in order to analyse and describe 'what is', understand 'how' and 'why' and design an action: 'how to do' something. Therefore it is used for theory building (Gregor 2002).

Although the distinction of the two described approaches are quite clear and it can clearly be argued which one someone follows, it can be said, that most research projects involve both methods of reasoning at some time in the project (Trochim 2006). Beside this two traditional methods a third method, abduction can be found within the literature (Burks 1946). Within an abductive approach to reasoning an explanatory hypothesis is developed. Conclusions are used in order to support requirements which could explain the conclusion. The difference to deductive and inductive thus is the extension of knowledge.

2.2 Ontology, Epistemology and Theoretical Perspective

The **ontology** represents the picture of social reality on which a theory is based. It asks questions like what assumptions are made, what exists and how do units interact with

each other in the nature of social reality. It is concerned with the beliefs of the social reality (Blaikie 2000, p.9). It can for example be distinguished between a realist and idealist view (Blaikie 2000, p.95). Within the idealistic approach it is assumed, that all knowledge is constructed. In contrast realism sees the world as it is (Hannes & Lockwood 2011, p.6).

Epistemology is the underlying philosophy and the underlying assumption that for example can be objectivism, constructionism or subjectivism (Crotty 1998). Objectivism believes that knowledge exists whether we are aware of it or not. When conducting objectivist research attempts are made to find effects and explanations. Events are predicted and theories and hypotheses are tested. Everything revolves around the explanation of things. The opposite of objectivism represents subjectivism. Within subjectivity it is believed that everyone has a different understanding of what we know. Taking into account the culture as one influencing factor of people's interaction with the reality it is argued within a constructionist approach. Here "all meaningful reality is socially constructed" (Crotty 1998, p.55) thus knowledge emerges through interaction in social contexts.

Ontology and epistemology always come together. The ontology can be seen as being alongside epistemology, informing the **theoretical perspective**. Together they intertwine as the theoretical perspective that shapes the way of how things can be understood (Crotty 1998). Therefore the theoretical perspective often is referred to as the research paradigm. Walsham (1993, p.21) for example merges the epistemology with the theoretical perspective, however most commonly it is distinguished between the three approaches positivist, critical and interpretive (Neuman 2003; Myers 2011). The differentiation between these theoretical perspectives in practice is not always clear cut but they can be described as following.

Positivist

When conducting positivist research it is assumed that the reality is objectively given. The world is not capable of being influenced by the researcher or his instruments (Lincoln & Guba 1985). Conclusions are directly drawn from observations (Myers 2011) and everyone experiences the world in the same way (Neuman 2006). The "*reality is real; it exists "out there" and is waiting to be discovered*" (Neuman 2006, p.83). When analysing the same phenomena twice, the result should be the same or similar (Myers 2011).

In the past the research in IS was mostly dominated by positivist research (Orlikowski & Baroundi 1991), but a shift towards more empirical studies and qualitative research can be observed (W. Chen & Hirschheim 2004).

Critical theory

Within critical research the social reality can be changed through interaction of people but is limited by historical, political, social, cultural and economic circumstances (Orlikowski & Baroundi 1991; Myers 2011). Critical theory is often used in IS research in order to develop system specification, design and implementation (Orlikowski & Baroundi 1991).

Interpretive

Same as critical research, interpretive research also supports the idea that people have influence on the social and economic world and can change it. People live in a world where they construct their own reality as they interact with the world around them (Novak & Hoffman 1997; Orlikowski & Baroundi 1991). The reality is getting accessible through the different social constructions such as language, consciousness and shared meanings (Myers 2011). Data which is collected within the interpretive approach is the presentation of other people's constructions about what they are up to (Geertz 1973, p.9).

Quantitative vs. Qualitative approaches

Even though the specific research paradigms are often aligned with one specific approach e.g. positivist with quantitative research and interpretivist with qualitative (Remenyi 2002) it is not ruled out that each assumptions in general can either been seen from a qualitative or quantitative perspective.

Quantitative research has evolved from natural science. "*Natural Science is concerned with explaining how and why things are*' (March & Smith 1995, p.253). The basis for this kind of research generally relies on formal instruments and numerical measures (Myers 1997) which are mainly analysed using mathematically based methods (in particular statistics) (Aliaga & Gunderson 2002). But this does not mean that the analysed phenomenon has to be in a quantitative form. Information about attitudes or beliefs for example can be gathered using questionnaires and then be translated using statistics (Muijs 2004). However, the researcher's goal mainly relies in getting numerical output, constructing a meaning from it (Straub et al. 2005) and being able to replicate the same findings again in the future. Therefore the process is mostly predefined and well-documented (Neuman 2006).

In contrast to the quantitative approach, the qualitative research approach emerged from context of social and cultural science and is used to explain social phenomena in natural settings (Pope & Mays 1995; Myers 2011). The researcher tries to understand particular issues within the context and therefore explores the perspective and behaviour of acting people within the situation (Kaplan & Maxwell 2005). Here the concepts are being established during and through the research itself and not before the actual research (Neuman 2006). Major data collection methods within qualitative research are among other observations, document analysis, the researcher's impression or interviews (Kaplan & Maxwell 2005; Myers 2011). Therefore, data in the form of words rather than numbers are important.

In the past most IS research studies were based on quantitative outcomes (Kaplan & Duchon 1988), but a clear distinction between quantitative and qualitative research does not exist (de Vaus 2002). Furthermore, qualitative research in Information System is often assigned to interpretivism or positivism (W. Chen & Hirschheim 2004).

As already discussed earlier the distinction between the different phrases are not clear cut and used differently within the literature. Figure 4 gives a short overview of how the researcher sees the connections between the method of reasoning, the ontology, the epistemology, the theoretical perspective, quantitative and qualitative research, the methodology and the research methods.

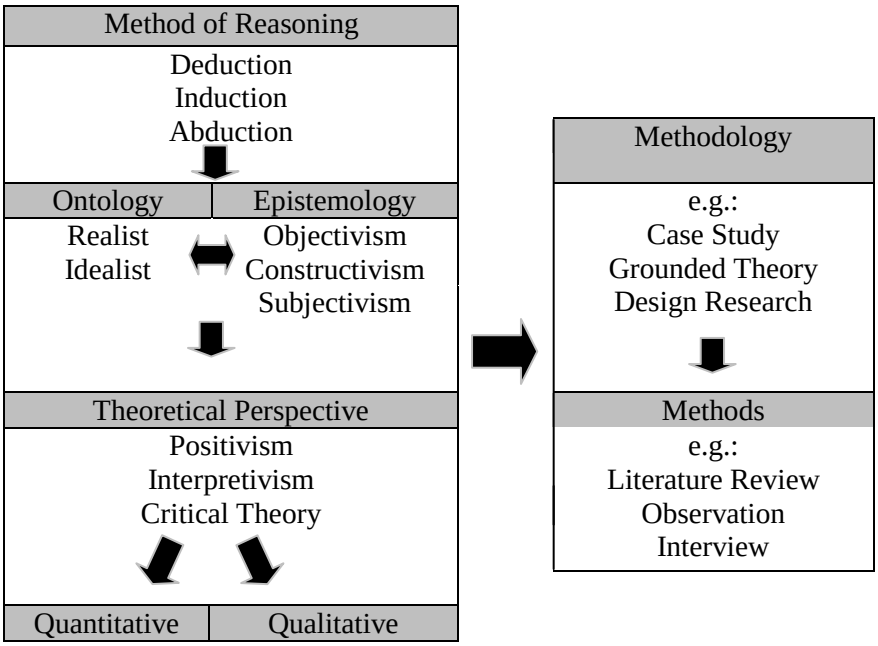


Figure 4: Methodological Connections

The overall method of reasoning determines what is done/which way is taken within the whole research project. The ontology and the epistemology are interlinked and

interdependent. Together they build the basis for the theoretical perspective used in a research approach. Thus, even though both are possible the theoretical perspective mainly is dominated by qualitative or quantitative research aspects. All this classification then lead to the methodology which is used and this in turn determines the data collection methods.

2.3 Methodology

Following Crotty's (1998, p.3) definition a methodology is "*the strategy, plan of action, process or design lying behind the choice and use of particular methods and linking the choice and use if methods to the desired outcome*". Thus the methodology frames the research and determines the research methods which can be used. Within the research literature many different methodologies like ethnography, action research, heuristic inquiry, experimental research, etc. can be found. It is not the aim of this thesis to outline all of them. Following four examples of possible research methodologies are shortly discussed in order to get an idea of what it means to use a specific methodology.

2.3.1 Case Study

Within IS research the case study methodology is used within many situations as it can be applied to research towards description, discovery and testing of theories and can involve qualitative data as well as quantitative data (Cavaye 1996). It is strongly associated with and the most frequently used qualitative research methodology as a researcher explores a process or a phenomenon within its real-life setting using many different data collection techniques (Creswell 2003, p.15; Orlikowski & Baroundi 1991). There is no standard definition of case study (Benbasat et al. 1987), however the overall aim of the case study methodology is to understand a phenomenon within its original, natural context (Whitman & Woszezynski 2004, p.312). Benbasat et al. (1987, p.371) pointed out the following 11 key characteristics of case studies:

1. *"Phenomenon is examined in a natural setting.*
2. *Data are collected by multiple means.*
3. *One or few entities (person, group, or organization) are examined.*
4. *The complexity of the unit is studied intensively.*
5. *Case studies are more suitable for the exploration, classification and hypothesis development stages of the knowledge building process; the investigator should have a receptive attitude towards exploration.*
6. *No experimental controls or manipulation are involved.*

7. *The investigator may not specify the set of independent and dependent variables in advance.*
8. *The results derived depend heavily on the integrative powers of the investigator.*
9. *Changes in site selection and data collection methods could take place as the investigator develops new hypotheses.*
10. *Case research is useful in the study of "why" and "how" questions because these deal with operational links to be traced over time rather than with frequency or incidence.*
11. *The focus is on contemporary events."*

As the analysed subject often is only looked at from one perspective e.g. the view of one organisation and the gathered data typically is subjective, findings from case study research are not always generalisable. Data collection methods used within case studies are for example document analysis, interviews, direct observation, participant observation or archival records (Yin 2009). An in depth description of case study can be found in Yin (2009) and Benbasat et al (1987).

2.3.2 Grounded Theory

The beginnings of grounded theory can be traced back to Glaser & Strauss (1967) and originates from the discipline of Sociology. It is a form of ethnographic investigation. Through the systematic collection and analysis of data a theory for the purpose of prediction and explanation can be developed (Crotty 1998). Important to note is that the goal is not to establish 'the' theory, but 'a' theory that can help to understand a specific actions (Heath & Cowley 2004). The primary purpose of the theory generation can be defined as following:

- To enable prediction and explanation of behaviour
- To provide theoretical process in sociology
- To be usable in practical applications, in prediction and explanation for the practitioner
- To provide a perspective on behaviour
- To guide the researcher's style for particular areas of behaviour (Glaser & Strauss 1967, p.3)

One key process is induction which means that generated ideas must be verified by the data (Heath & Cowley 2004). This fact is, were the name comes from. Everything is "grounded" in the data from which the theory was build (Power & Moynihan 2003). It

is a highly structured and systematic approach which uses data collection methods like interviews, literature review, or participant observation. Neumann (2003, p.52) stated that grounded theory often is positivistic as it searches for a theory that is precise, provable, replicable and generalisable. It is very useful when developing context-based and process-oriented descriptions and explanations of phenomena (Myers 2011).

2.3.3 Design Research

To understand design research as a methodology it helps to start with a definition of the term design. Within the Merriam-Webster Online Dictionary design as a verb is defined as "*to create, fashion, execute, or construct [something] according to plan*". As a noun it is defined as "*a plan or protocol for carrying out or accomplishing something*" and "*the process of preparing this*" (Merriam-Webster 2001). Looking at these and most other definitions of design it is noticeable that most definitions share three characteristics: (1) design refers to a process, (2) the process is goal-oriented and (3) the goal aims to solve a problem, meet needs, improve situations or create something new or useful (Friedman 2003, p.508). Within IS research, design science basically is about problem solving which addresses business needs (Hevner et al. 2004). It is particular useful in new areas and in systems where deficiencies could be identified (Rossi & Sein 2003).

It can be classified as (applied) research as it uses knowledge to solve practical problems (Nunamaker et al. 1990). Fallman (2003) distinguishes two different kinds of design research. On the one hand, design-oriented research which means to produce new knowledge by using design activities within the research process. On the other hand, research-oriented design where the primary objective is the creation of artefacts which can answer real world problems (Fallman 2003).

In IS design research is used in order to address so called wicked problems (Hevner et al. 2004). Wicked problems were introduced by Rittel and Webber (1973). They contrasted wicked problems which occur in the social world and are related to planning and policy vs. tame problems which occur in the natural science world. Wicked problems are being described as: ill-defined, ambiguous and associated with strong moral, political and professional issues (Ritchey 2011, p.2).

The process of design research can be described in an iterative cycle including the steps: awareness of problem, suggestion, development, evaluation and conclusion (Takeda et al. 1990; Vaishnavi & Kuechler 2001). Rossi and Stein (2003) use different terms (identify a need, build, evaluate, learn and theorise) but mean the same when describing

the design science cycle. Within the step development/build artefacts should be developed. Such artefacts can include algorithms, human/computer interfaces, system design methodologies or language (Vaishnavi & Kuechler 2001). Within IT they can be characterised as following. They are:

- not natural, neutral, universal or given but designed, constructed and used by people and shaped through interest, values and assumptions.
- always attached to some time, place, discourse and community.
- always builded through multiplicity of often fragile and fragmentary components
- neither fixed nor independent but emerge from ongoing social and economic practices.
- not static, but dynamic as existing artefacts are exchanged by new ones (Orlikowski & Iacono 2001, p.131).

Besides the above outlined characteristics for artefacts Table 1 outlines the design science research guidelines which can be found within the literature. However it is important to notice, that these guidelines are not fixed or mandatory (Hevner et al. 2004, p.83).

Guideline	Description
Design as an Artefact	Design-science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation.
Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Design Evaluation	The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods.
Research Contribution	Effective design-science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies.
Research Rigor	Design-science research relies upon the application of

	rigorous methods in both the construction and evaluation of the design artefact.
Design as a Search Process	The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Table 1: Design-Science Research Guidelines (Hevner et al. 2004, p.83)

Rossi and Sein (2003) suggest that the ontological standpoint of design research should be realist because a real world exists and we can intervene to improve it. The research mostly is qualitative and according to Hevner et al. (2004) design research fundamentally is an interpretative approach.

2.3.4 Action Research

Kurt Lewin first introduced the term 'action research' in 1946 explaining a new approach in social research were the generation of theory within a social system was developed through the actions of the researcher in the system (Susman & Evered 1978, p.586). He further characterized it as "*a comprehensive research on the conditions and effects of various forms of social action and research leading to social action*" (Lewin 1946). Within today's literature the most used definition of action research can be found by Rapoport (Susman & Evered 1978). He defines action research as aiming "*to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science by joint collaboration within a mutually acceptable ethical framework*" (Raporort 1970, p.499). Thereby, one of the key characteristics as described by Checkland & Holwell (1998, p.11) is the deep involvement of the researcher. In order to find a solution to a specific problem the researcher completely immerses himself in the situation regardless of the path it takes. The researcher himself becomes part and agent of the organisational change thus today it is a feasible methodology in organisational development (Chishom & Elden 1993).

A cyclical process for action research developed by Susman and Evered (1978) give an overview of how action research is been conducted by dividing the process of development into 5 steps namely: diagnosing, action planning, action tacking, evaluating and specific learning (see Figure 5).

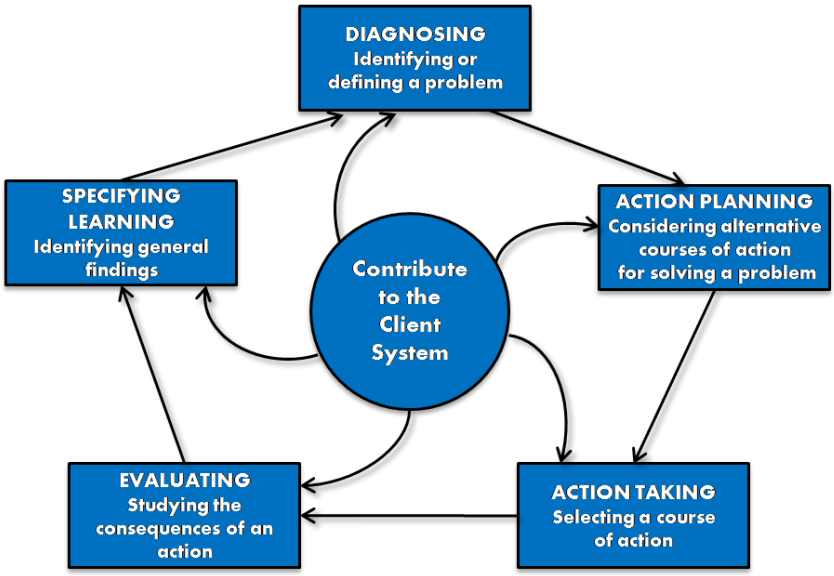


Figure 5: The Action Research Cyclical Process (adapted from Susman & Evered 1978, p.588)

All steps thereby contribute to the development within the client system which is the social system where the researcher faces the problem. As the process shows, action research is applied research which merges research with social actions (Neuman 2006). It tries to solve practical problems through the actions of the researcher (Davidson et al. 2004) who at the same time is gathering data to add to the body of knowledge (Myers 2011).

Within their paper "The field of action research" Chein, Cook and Harding (1948) distinguish four different kinds of action research. First, 'diagnostic action research' which occurs when the researcher only collects data for diagnosing and brings it back to the system. Second, 'empirical action research' where the researcher evaluates all actions of the system and feeds it back to it. Third, 'participant action research' when both diagnosing and action planning are carried out in collaboration between the researcher and the system. And last, 'experimental action research' when the researcher and the client collaborate in nearly all or all phases to develop an experiment to evaluate its consequences (Susman & Evered 1978, p.588).

2.4 Research Methods

Having chosen a methodology for the research project, the next step is to think about possible research methods. Research methods are "*the techniques or procedures used to gather and analyse data related to some research question or hypothesis*" (Crotty 1998, p.3). Quite often research methods are also referred to as data collection techniques. In order to get different views and to increase the robustness of results it is appropriate to

use multiple data collection techniques in one project. Marshall and Rossman for example (2011) distinguish between participant and direct observation, in-depth interview, document analysis, survey, etc. It is not the aim of this thesis to name and discuss all methods. Therefore below 3 methods are outlined exemplarily.

Review of Literature

The literature review is one of the earliest steps in any research process and has the goal to integrate and summarise everything that is known in a specific area (Neuman 2006). It is "*a systematic, explicit, and reproducible method for identifying, evaluation and synthesizing the existing body of completed and recorded work produced by researchers, scholars, and practitioners*" (Fink 2009, p.3). It provides the background as well as the justification for the research questions (Bruce 1994).

Document Analysis

When conducting document analysis all kind of documents including, minutes, reports, files etc. can be used. The documents are either produced during the everyday action or specifically for the researcher (Marshall & Rossman 2011). Thus it needs to keep in mind that the documents may not be accurate but can help to understand an organisational and/or personnel background.

Content Analysis

Locking at Holsti's Book about content analysis shows that many different definitions of content analysis can be found within the literature (Holsti 1969, pp.2–3). These definitions include aspects like 'statistical semantics', 'classification', 'description' and 'coding'. Content thereby quite often is referred to as any kind of communication. Neuendorf's definition of content analysis brings these aspects together by describing it as the "*systematic, objective, quantitative analysis of message characteristics*" (2002, p.1).

2.5 Research Design

Thus far, the described possible research strategies were outlined in a general manner. Following it is argued which strategies and methods were uses within this research project. However, the aim of the researcher is not to categorise this thesis into the whole structure outlined above. This is based on the fact that it is not always possible to clearly place someone's work into each category.

As already mentioned everything is dependent on the outcome which should be achieved and the related research question. After the research problem was identified and described in Section 1.1 the research objective and related questions have been developed. These drive the research project and limit the possibility for used strategies.

Looking at the chief research question it is obvious that it is not a phrased hypothesis which should be confirmed or refuted. Thus the deductive approach was not suitable for this research project. Also the inductive approach does not fit completely. It is not tried to generalise the findings, but rather to help understanding and informing the development of the final framework. The method of reasoning thus is an abductive approach.

Classifying this research into an epistemology it very quick becomes clear that for example the findings of the tool evaluation and the analysis of the exemplary case cannot be treated as facts which are objective. Therefore objectivism was not suitable for this research. Also subjectivism does not completely fit to this research. Within subjectivism it is assumed that knowledge does not develop from an interplay between subject and object, but is projected by the subject onto the object (Crotty 1998, p.9). Rather than projecting a meaning to something, it is assumed that knowledge is contingent upon human and their interaction which leads to a constructionism approach. Looking for example to research question 3b: "Are there differences between Web Mining for small or big businesses that need to be taken into account?" this question explicitly asks for differences that can arise through human interaction. Resulting from the method of reasoning and the epistemology the theoretical perspective is clearly set, as they go hand in hand. The interpretive approach takes into account the possible influences humans can have on the social and economic world and therefore is the right one for this thesis.

The research question also defines the needed research methodology, data sources and thus also the data collection methods. Looking at the different supporting research questions it can be said, that they are quite diverse. But it is possible to answer them by using the action research approach. Thereby different parts as e.g. the analysis of Web Analytic tools can be seen as independent action research cycles which nevertheless contribute to the whole. All phases from diagnosing over action taking to evaluating were carried out during the whole process of developing the final process framework which in turn will contribute to the understanding of Web Analytics.

Within the different steps different research methods were used. The literature review provided the basis for understanding the topic. Furthermore, a document analysis was

conducted while analysing the exemplary cased. Files like project reports and descriptions on websites as well as the content on the analysed website itself helped to understand the goal of the case website. Furthermore, a content analysis was used in order to evaluate the WA tools by identifying and testing the different functions. In addition different statistics and metrics were used in order to describe the user behaviour discovered through the log files.

In order to give an overview of the whole research design for this thesis Table 2Figure 5 summarises everything by mapping the research objectives, questions and strategies.

<table>
<tr><td colspan="2">Problem
Increase Internet usage and interest in Web Analytics, but fragmented and not user focused literature; Enormous availability and diversity of WA tools; Missing integrated view of Web Analytics for organisational usage</td></tr>
<tr><td>Research Objectives

The overall objective of this research project is to develop a framework for Web Usage Mining which guides organisations in the development and maintaining of a reliable, usable and sustainable Web Analytic programmes

▪ Identify key trends in Web Analytic methods, metrics and techniques
➔ Literature Review

▪ To investigate into an better understanding of Web Analytic tool functionality and differences
➔ Tool Overview

▪ To contribute to the development of a framework for effective Web Analytic usage
➔ Framework for Web Analytics</td><td>Research Questions
Chief Research Question
What is required in a framework to support small- and medium- sized enterprises (SME) to provide an easy entry point for undertaking Web Analytic projects?

Supporting Research Questions
RQ1a: How can Web Mining be classified?
RQ1b: What different methods for conducting Web Mining are described in the Web Mining literature?
RQ1c: Which data capturing method is the best to use today?

RQ2a: What does the Web Analytic market look like?
RQ2b: What tools for Web Mining are available on the market?
RQ2c: Which criteria can be used to describe and differentiate these tools?

RQ3a: Is there a need for distinguishing the kind of website for what Web Mining is used in order get clear results?
RQ3b: Are there differences between Web Mining for small or big businesses that need to be taken into account?</td></tr>
<tr><td colspan="2">Scope and Context
Assist SMEs in the process of Web Analytics with the help of current literature, tool evaluations and thought the use of one example case with an information website.</td></tr>
<tr><td colspan="2">Method of Reasoning
Abductive
Epistemology
Constructionism

Theoretical Perspective
Interpretive
Methodology
Action Research</td></tr>
</table>

Method: Literature Review
Theoretical Data: Web Analytics: context and process description (systematic analysis)
Method: Content Analysis
Subjective Data: Functionality matrix and tool comparison (Tool analysis)
Method: Document, Content and Statistical Analysis
Exemplary Data: Website analysis and analytic statistics (Case evaluation)
Developing Key Outcome and Conclusion
Respond to the finding of the gap analysis and findings from the case evaluation. Interpretation of the findings from the supporting research questions. Addressing chief research question.

Table 2: Research Design

2.6 Limitations

This research project has a number of limitations. Even if the topic of Web Analytics is quite young, it is broad and diverse. Many academic papers have already been published and hundreds of different tools are available. This in connection with time limitations causes the fact that it was not possible for the researcher to exhaustively review the topic. For example, the processes of how the web data is analysed, e.g. which algorithms can be used was not addressed deeply. Furthermore, a selection of tools to be analysed had to be made and the number was set to 15 for practical reasons. In addition the author was dependent on access and availability of the tools. However, this does not mean that excluded tools were not equally appropriate to study.

Another limitation had to be made in the usage of the interpretative case study. Different to planned beforehand, the data for analysis which was available for the researcher was not available in the amount and quality as expected. Therefore the insights which were expected from the case needed to be revised. Finally, the developed framework has not yet been tested in a real-world scenario. It was developed using theoretical and practical input but is not yet verified in practice, this constitutes a further cycle of future research.

3 Websites: An Introduction

"Suppose all the information stored on computers everywhere were linked, I thought. Suppose I could program my computer to create a space in which anything could be linked to anything. All the bits of information in every computer at CERN, and on the planet, would be available to me and to anyone else. There would be a single, global information space."(Berners-Lee & Fischetti 2000, p.4f)

Within his book "Weaving The Web: The Past, Present and Future of the World Wide Web" Tim Berners-Lee the inventor of the World Wide Web (WWW) describes how he came to the idea of developing the beginnings of the WWW and how it emerged further.

This first theoretical chapter will give a short overview of how websites in the Internet evolved and how they can be classified today. Following this the usage/success of websites is discussed and key challenges are outlined. The aim of this chapter is not to give a detailed analysis of websites. It rather serves as a starting point for the subsequent empirical research and discussion.

3.1 Evolution

It can be said, that the major step in the evolution of the Internet started with Tim Berners-Lee. Even though he wasn't the first who had the idea of connecting two or more computers his idea of having links between nodes was the one which started the bigger development, the WWW. Within a project of US research laboratories and US universities the first communication between two computers took place via the ARPANET (Advanced Research Projects Agency Network) in October 1969. In 1971/1972 the first e-Mails were sent. Eight years later, when Tim Berners-Lee was consulting for CERN in 1980 he wrote a program called "Enquire-Within-Upon-Everything" which could handle bidirectional links between arbitrary nodes. He developed a program called Enquire because he was searching for something that could help him within his work. However, when he left CERN, he didn't take the code of this program with him. It got lost. Another nine years later, in 1989 when Berners-Lee was again working for CERN he wrote the first paper about the idea of the WWW "Information Management: A Proposal". Unfortunately the paper didn't get attention at CERN and nothing happened. In 1990 he then again tried to publish the paper, but still

no one really listened, except Mike Sendall. From him he got the acceptance to program his idea. The development for the first web browser began and the term World Wide Web was born (Berners-Lee 1989; Berners-Lee & Fischetti 2000; Connolly 2000). In March 1991 the program was given to the first people within CERN. Only five months later when Berners-Lee started to keep logs of how many times the first web server info.cern.ch was accessed he counted 10 to 100 hits a day. From that point the web grow rapidly and CERN is seen as the place "*Where the web was born!*" (CERN 2008).

It was not only used by universities, very soon public and private sector organizations became interested in using the technology. It was the beginning of the hype which is still continuing today. The Internet is becoming more and more important and the time people spend on the internet in the meantime became larger than the time people watch TV. It is believed that this trend will grow further (Hassler 2010, p.15).

3.2 Classification

When surfing through the web many different kinds of websites can be found. As the different sites have different purposes it is helpful to classify websites into a specific category. Five basic types are often distinguished: Commercial, Personal, Informational, Organisational and Political (P. Crowder & D. A. Crowder 2008, p.15f).

- **Personal websites**

  Even though the goal of a personal website is largely to introduce an individual, his interest or opinion, the implementation can be very different. Personal websites range from simple text sites over to picture sharing sites to blogs or forums. One person usually administrates them.

- **Commercial websites**

  The term itself already explains which types of site fall into this category. All websites which are built in order to generate profit are called commercial websites. Company websites, which do not generate profit but for example just represent a business often are not included in the category commercial website, but in the category organisational website.

- **Organisational websites**

  Organisational websites present information about a particular organisation. They are quite similar to informational websites but are administrated by an organisational body.

- **Informational websites**

  The purpose of informational websites is to provide information on a particular topic (free or paid). Many of these sites are created by public-minded organisations who want to make people aware of a special topic.

- **Political websites**

  Websites on which political issues are shown or discussed are called political websites. This includes sites dealing with political parties, candidates, legislations, etc.

As can be seen from the example of a company website the distinction between the categories is not always clear cut. Often it depends on the perspective whether a website is categorized in one or another type. However, beside the above outlined classification it is also common to classify websites in terms of the content and features they offer. Examples for these classifications are search engines, blogs, wikis, directories, forums, online shop, social networking sites, file sharing sites, etc. It is not the objective of this thesis to explain all these types.

But no matter which kind of website someone looks at, they all have something in common. They consist of three basic elements: texts, pictures and links (Wu et al. 2009, p.168). The home page thereby often is the first contact a user has with a specific site. It is important, that this page picks up the intention of the user to further use the website (Wu et al. 2009, p.167). The structure then is created through hypertext links between different page views (Cooley 2003, p.98). In particular, these links can lead to a complicated network of pages. Users should be assisted so that they can find searched information without further help (Spiliopoulou et al. 2000, p.142).

3.3 Usage/Success

Today, the World Wide Web is still growing rapidly. In 2009 about 1 million pages were released each day. As more and more information is available online the usage is also growing (Kumar Jain et al. 2009, p.1). The size of the web increases with its users. The number of websites and visitors to websites has increased exponentially (Pani et al. 2011, p.15). By the end of March 2011 2,095,006,005 people had already used the internet. This accounts for 30.2% of the world population (Miniwatts Marketing Group 2011).

An early study of e-commerce websites showed that only 56% of users successfully performed the task they intended (Nielsen & Norman 2000). Furthermore more than half (65%) of all online shopping efforts failed (Souza et al. 2000). The reasons for this are diverse. One example is that people cannot find what they are looking for because it is not obviously placed within the site. Looking from a user perspective, Spiliopoulou (2000, p.128) outlines three factors that can be identified that influence the way people perceive and evaluate a site. The first is the content, which is represented by the products or services that are offered. The second and third factor, web page design and overall site design, refer to the accessibility of the content. This confirms Sullivan (1997). He recommends measuring the success of a website with the help of the navigability and accessibility of a page. Chi (2002, p.64) stated "*usability can make or break a Web site*". Usability is an important and essential character for a site to survive (Nielsen & Norman 2000). Therefore it is becoming essential for providers of websites to support better service and increase the quality of websites (Pani et al. 2011, p.15). Berthon et al. (1996) state that the success of a website can be measured by the 'contact efficiency', thus through the percentage of users that engage in exploring the site and by the 'conversion efficiency', the percentage of the visitors that finally become customers.

Today websites are a key communication channel for organisations (Norguet et al. 2006, p.430). One big advantage of electronic publishing in contrast to newspapers or television broadcasts which are hardly measurable directly, is that websites can be measured directly (Ogle 2010, p.2604). Through the possibility of logging all user events on a site and analysing this information, user behaviours can become visible and used for improvements in websites, marketing, etc. (Burby & Atchison 2007). Therefore it is not surprising that the interest in monitoring user activities on websites is said to be as old as the web itself (Spiliopoulou & Pohle 2001, p.86).

However, it is important that a website has a clearly defined outcome. What is the intent of the website? Only by answering this question can analysts begin to define whether a website is successful and the outcome is achieved or not. They do this with the help of metrics which can be used to measure a website and which are explained in details in Section 5.2. Besides these metrics qualitative behaviour analysis is also needed to really understand visitors. Questions like "*Why do our visitors do the things they do, and what is their motivation for engaging with our website?*" (Kaushik 2007, p.10) need to be asked. There is an ongoing trend to try to understand users and to use this information in the best way (Pani et al. 2011, p.22). In order to get answers to all these diverse questions and attempts a new research area has emerged: Web Mining.

3.4 Challenges

Over 10 years ago Kosala and Blockeel (2000, p.1) identified that the web is huge, diverse and dynamic. Since then the web has grown even bigger. The problems in using and administering it are significant. Within their paper Kosala and Blockeel outline the following problems, of which aspects can still be applied today.

- **Finding relevant information**

  Even though since the year 2000 the search engines within the web have become much better it sometimes is quite hard to find the information someone is searching for. If we for example have two sites about the same topic and the one is search engine optimized and the other not, we might not find the second site.

- **Developing new knowledge with the information available**

  Finding information is one side of the coin, using it in an appropriate manner is the other. Especially today, where everyone is able to share information in the web the question of reliability needs to be asked.

- **Information personalization**

  The possibilities to personalize websites are huge. But who wants to see what, and who is interested in what? People need to be analysed in order to serve them with the right information.

- **Learning about users**

  As already described in the point above, understanding the users is key for, for example, efficient mass customization, effective Website design and management, effective marketing, etc.

One attempt to solve these problems by getting insight in the content, structure and usage of websites is Web Mining. Web Mining is further outlined in the next Chapter, Chapter 4.

4 Web Mining

As websites are a key communication channel not only for companies, but also for private individuals trying to find diverse information, it is important to find ways to make the web more usable. In order to, for example, understand user behaviour or results of search engines it is necessary to analyse the information available in/from the web. The field that describes these tasks is called Web Mining. In the past the innovation within Web Mining was often derived from possibility-driven innovation and not from customer-driven innovation. This means that it was looked at the data available and the question of what else could be possibly to get to know with this data was asked (Kaushik 2007, p.7). After a short introduction of how Web Mining evolved and a more detailed definition of Web Mining this chapter outlines which questions were asked and why. Following this, the process which can be used for Web Mining is explained. At the end a classification of Web Mining is presented and the different types are described.

4.1 Definition of Web Mining

The process of getting knowledge through analysing data is not new. Already before computers were invented people tried to analyse huge amounts of data (Berry & Linoff 2011, p.2). The methodology called Data Mining (often although referred to as Knowledge Discovery or Business Intelligence) has been used for many years. Data Mining can be defined as "*the exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules*" (Berry & Linoff 2011, p.5). As these analyses can help to find possibilities for improvement in different business fields their adoption within the business area is growing. Therefore, in a more recent definition of Data Mining it is even defined as a business process that needs to be undertaken in order to be able to react to the results and generate a long-term benefit (Berry & Linoff 2011, p.2).

Within Data Mining three different types can be distinguished: Data Mining, Text Mining and Web Mining. Data Mining is concerned with data kept in databases in a structured and organised way. In contrast Text Mining mainly deals with unstructured data/text. Between these two mentioned types of Data Mining we can find Web Mining (also referred to as Web Data Mining or Web Data Extraction).

"Web mining is the use of data mining techniques to automatically discover and extract information from World Wide Web documents and services"
(Etzioni 1996, p.65).

Cooley et al. (1997) describe Web Mining as the application of Data Mining techniques to discover and analyse useful information from the World Wide Web. Simplified Web Mining refers to the usage of Data Mining to analyse web data (Pierrakos et al. 2003, p.319). Within Web Mining research different areas such as database, Information Retrieval (IR) and Artificial Intelligence (AI) incorporate together (Kosala & Blockeel 2000, p.1). This shows how diverse and overarching the field is. It handles semi-structured and/or unstructured data (Zhang & Segall 2008, p.683f). Because of the huge amount of heterogeneous data available it is one of the most challenging areas of Data Mining. Sceptics even say that the web is too unstructured to accomplish efficient Web Mining. They argue that the web was designed for human usage and therefore uses natural language text which is hard to analyse (Etzioni 1996, p.1). This may be one reason why Web Mining can be seen as a minefield. But Greening (2002) also outlines that it simultaneously can be seen as a goldmine. It is possible to obtain a huge amount of data and the sooner this data is analysed and actions are undertaken it is possible to be ahead of the competitors.

4.2 Reasons for Using Web Mining

As mentioned before Web Mining mostly is a possibility-driven innovation (Kaushik 2007, p.7). When developing Web Mining it only considered the data that is available and from this different questions could be answered. Following an overview of the types of questions that could be answered is given. This in turn also gives an overview of why Web Mining is used today.

When looking through the Web Mining literature (e.g. Spiliopoulou 2000; Cooley 2003; Kaushik 2007) it quickly becomes obvious, that the overall goal is to increase the success of a website. Success thereby means different things. For an e-commerce site this often means a higher sales volume. But it can also mean more clicks on ads or more visitors who extract information. When planning for a project today Return on Investment (ROI) studies become more and more important. Setting up a website or redesigning a site now also often need to be based on a ROI study in order to get the necessary financial investments (Hassler 2010, p.28). In order to get to know how a website is used and to improve a site knowing the user's behaviour is a key indicator. When using Web Mining techniques questions like the following often lead the course of action:

- How did the users browse through the website?
- Where did the users come from to our website?
- How long did users spend on a specific page?
- Which page was the entry/exit page?
- Did errors occur during the use of the website?
- Do we have pages that users never come to?
- Why did the user behave like this?

By answering these questions through Web Analytics (techniques and metrics therefore are explained in Chapter 5) it is possible to get an idea of what the user is doing. But the important question of why the user is behaving like this cannot be answered only through existing data captured on a server or somewhere else. To answer the last question qualitative analysis like surveys or usability testing are needed.

As websites are one of the key communication channels for organisations today (Norguet et al. 2006, p.430) the main goal of Web Mining is to get to know a way of how to use a website in the most efficient way. By understanding the user's needs it can be possible to redesign the site, personalize services and optimize campaigns in order to get pleasant users which may become customers.

When analysing all these questions one very important aspect which should always be kept in mind is the need to ask the question if the website is doing well from the perspective of the visitor (Kaushik 2007, p.77). People who are disappointed might talk badly about the site. Each good sales person asks the customer about his wishes and is not just telling him what he thinks might fit. The internet is a quite quick medium and it is not enough only to look at the customer every two or three years when redesigning the page. It is important to not fall behind the trend. Websites are constantly changing. It's always better to recognize by yourself that something is broken instead of getting to know it from the customer. Web Mining can help to identify issues which can be addressed at an appropriate page like an FAQ or support area (Hassler 2010, p.36f). On the other hand it is also important to keep in mind the business objectives of the website owner (Spiliopoulou & Pohle 2001, p.89). Only when combining both views it is possible to ensure that the website is successful for both, customers and owner.

4.3 Web Mining as a Process

When conducting Web Mining, the overall method used is quite similar. The process can be divided into the following subtasks (Kosala & Blockeel 2000; Zhang & Segall 2008; Etzioni 1996; Spiliopoulou 2000):

- **Resource finding**

  Within this first step data which we do not know about yet should be located.

- **Information extraction (selection and pre-processing)**

  Information from discovered resources should be extracted automatically (e.g. with automatically generated log files). Furthermore, this information needs to be pre-processed in order to remove irrelevant parts and prepare the data for further analysis.

- **Generalization/pattern analysis and recognition**

  This step uncovers general patterns of a site and can be used to compare them with patterns from other sites.

- **Analysis/validation and interpretation**

  The patterns discovered need to be analysed and validated in order to derive feasible statements and make legitimate conclusions.

- **Visualization**

  Visualization of the results can help to generate an overview of the discovered data.

Apart from this short overview more details about the process and techniques for Web Analytics is presented in Section 5.3.

4.4 Classification of Web Mining

In early attempts to categorize Web Mining two different ways were identified: Web Content Mining and Web Usage Mining. Web Content Mining uses all information available in the World Wide Web. Web Usage Mining deals with the discovery of user browsing and access patterns (Cooley et al. 1997). Later, a third category was distinguished: Web Structure Mining (Kosala & Blockeel 2000; Madria et al. 1999). It is concerned with the underlying link structure of the web itself. This categorization aligns with the three different types of data that can be identified in Web Mining. First

there is the data on the World Wide Web. This includes all texts, images, movies, etc. Second, web log data which track user activities and third, web structure data, thus linking structures (Madria et al. 1999). All three types of Web Mining have in common that they process information from the web. Content Mining therefore uses information from texts, images etc. Structure Mining uses information from linkages of web pages and Usage Mining information from users who actually used a particular site (Pani et al. 2011, p.15). Figure 6 shows a schematic diagram which classifies Web Mining into the three subgroups.

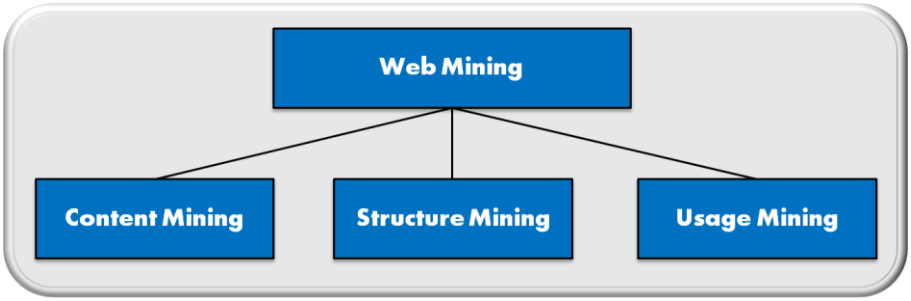


Figure 6: Web Mining Classification
(adapted from Cooley et al. 1997; Madria et al. 1999;
Kosala & Blockeel 2000; Pani et al. 2011)

The diagram above shows the classification of Web Mining but should not be seen as clear cut. All areas can have interrelations with each other and overlap (Kosala & Blockeel 2000, p.4; Masand & Spiliopoulou 2000).

It is not the objective of this thesis to give a complete overview of the whole area of Web Mining. However, a short overview of each of the three different areas of Web Mining is provided below.

4.4.1 Web Content Mining

The content available on the web is very diverse. Data such as text, images, audio, video, metadata and hyperlinks basically make up the web content. These data are mostly in an unstructured or semi-structured form. When thinking about the web content it is not only web pages that should be taken into account. Today many companies use e.g. digital libraries or web applications which also allow users to access company data from the web. Within Web Content Mining (also referred to as text mining, text data mining or knowledge discovery in texts) the attempt is to discover useful information from these data and to improve information finding and filtering so that it can be found more easily (Kosala & Blockeel 2000, p.3; Pierrakos et al. 2003). It includes information extraction, identification, match and integration of semantically

similar data (Zhang & Segall 2008, p.685). One problem of Web Content Mining is the semantic integration across web documents. For example the price of a product on one page is shown in numbers, on another page this same price is included in a picture (Madria et al. 1999, p.310). Each Web Mining attempt should at least include a small Web Content Mining part in order to get to know the website and to build together with Web Structure Mining an information architecture. Further information on Web Content Mining can be found in Kosala and Blockeel (2000), Zhang and Segall (2008) or Cooley (2003).

4.4.2 Web Structure Mining

This type of Web Mining deals with the structure, the web graph of a site and applies Data Mining techniques to them (Pierrakos et al. 2003). The information source thus is the hyperlink structure of a site (Zhang & Segall 2008, p.691). Web Structure Mining often is used for the categorization of web pages in order to find similarities and relationships between them (Kosala & Blockeel 2000). But it is also used to describe the organisation of the content of a website and the intrapage structure (Cooley 2003). Identifying a map or a graph of a web page can be quite simple, but this is dependent on the technology used for the site. If the site is generated dynamically and/or has personalized links (and content) included it get more complicated.

Web Structure Mining and Web Content Mining go hand in hand. They are interrelated tasks which are needed in order to process reasonable Web Usage Mining. For example, if a user cannot be uniquely identified through IP-Addresses or cookies the path a user went through the information structure may help still identify him. In his article Cooley (2003, p.93) goes even further. He claims that Web Structure and Web Content Mining not only greatly enhance Web Usage Mining, but that "*it cannot be completed without it*". The general understanding of the site is necessary in order to understand and interpret the metrics.

4.4.3 Web Usage Mining

The goal of Web Usage Mining is to discover and understand users' behaviour, compare the actual use with the expected use and get ideas for redesigning the site in order to meet the interest of users (Norguet et al. 2006, p.430; Jaideep Srivastava et al. 2000). The data which are looked at within this approach is data which is generated while the user interacts with the web (Kosala & Blockeel 2000; Etminani et al. 2009, p.396). Thus the data is not generally available in the web, but must be actively

collected. Secondary data which is collected through e.g. log files or tagging often is referred to as clickstream data.

Beside Web Usage Mining the term Web Analytics can be found in the literature. However, despite extensive search the author couldn't find a clear distinction between these two terms. It seems that the term Web Analytics is used more, when describing the whole process from data gathering to recommendations for website redesign. As this whole process will be applied within this research effort the author further uses the term Web Analytics which is described in more details within the next chapter.

5 Web Analytics

Already at the beginning of the Internet era people asked the question 'Is someone really looking at the site?' And if yes the next question followed soon 'How many are looking at it?' (Hassler 2010). Within the evolution of Web Analytics a lot of innovation was driven through asking the question of what is possible to do with the captured data (Kaushik 2007, p.7). These questions formed the first ideas of Web Analytics. But even though that was the beginnings of Web Analytics today it is much more. From the beginnings in the 1990s, it took more than 10 years before a standard definition of Web Analytics was given by the Web Analytics Association in 2006 which shows the broad spectrum of Web Analytics today (Kaushik 2007, p.2). They defined Web Analytics as

> *"the objective tracking, collection, measurement, reporting, and analysis of quantitative Internet data to optimize websites and web marketing initiatives" (http://www.webanalyticsassociation.org/ 2006).*

Looking at this definition it becomes clear, that Web Analytics can help to understand where problems exists, but it does not explain how to solve these problems or optimize the website (Wu et al. 2009, p.164). This further part needs to be done by an analyst who understands the metrics produced through WA and can translate them into actions.

Historically Web Analytics was part of the work of the IT department as it worked with the web servers log files and databases. But today, with the help of the analytic tools, this has changed. Often it is still located within IT, but the way people look at it has become more of a business function (Hassler 2010; Kaushik 2007). The data is used for website improvements on the one hand, but as the definition above shows it is also used for marketing, sales and support on the other hand (Kaushik 2007, p.93). This means, the strategic goals of Web Analytics are not only the prediction of user behaviour, but also the redesign of the website according to user interests (Norguet et al. 2006, p.430), the improvement of the quality of a website (Weischedel & Huizingh 2006, p.463) and getting insights for better disposal. In order to achieve all this and to improve the usefulness and the usability of a website it is key to assess the user's reaction and interaction (Sullivan 1997). And "*The only information left behind by many users visiting a Web site is the trace through the pages they have accessed*" (Spiliopoulou 2000, p.128).

By 2002 Web Analytics was already a billion dollar a year industry (Chi 2002, p.64). Companies spend millions of dollars for Web Analytics to optimize websites and campaigns which in turn should bring them billions of dollars in online revenue (Kaushik 2007, p.xxviii). Web Analytics is a booming business (Weischedel & Huizingh 2006, p.463). But as Albert Einstein argued: "*Not everything that counts can be counted, and not everything that can be counted counts*". Therefore it is important to really understand the data, metrics which are drawn from this data and how to interpret them in a meaningful way.

This chapter will first give an overview of the different data collection techniques available for Web Analytics today. Following this, different metrics are outlined before the process of Web Analytics is described in more detail. A short section then discusses the legal aspects of Web Analytics. At the end the challenges are summarised and a summary is given.

5.1 Data Collection

The way data is collected for Web Analytics can be distinguished into quantitative and qualitative data collection methods (see Table 3). Within the quantitative methods it is further distinguished between server-side data collection, client-side data collection and alternative methods which use a third level. When collecting server-side data this means that the data captured are from the view of the server and what the server works on. Log files are the data captured here. The client-side collection methods page tagging and web beacons show what happened on the website from a user perspective. Methods which use an alternative area to capture data are for example packet sniffer or reverse proxies. Here, the data collection is introduced between the server and the user (Hassler 2010; Kaushik 2007). Within the qualitative data collection methods investigations like lab usability testing or surveys are common methods.

Quantitative	Server-side Log files
	Client-side Page tagging Web beacon
	Alternative methods Packet sniffer Reverse proxies
Qualitative	Lab usability testing
	Heuristic evaluation
	Site visits
	Survey

Table 3: Data Collection Methods

Beside these methods Kaushik (2007, pp.24&39) outlines a different grouping of data. He distinguishes between 4 groups: clickstream (web logs, web beacons, JavaScript tags, packet sniffing), outcomes (visitors, page view, time, referrers),

research/qualitative (surveys, heuristic evaluation, usability testing, site visits) and competitive data.

Within this research project the aim is to get an overview of the quantitative methods (clickstream) and the metrics (outcomes) as these are the data used by Web Analytic tools. Therefore only the quantitative methods are further outlined here.

The first method used for Web Analytics data collection within history was to capture log files generated on the web server (Kaushik 2007; Hasan et al. 2009). However, as other methods have been developed which try to overcome the issues of log files, all methods are outlined below.

5.1.1 Log Files

Originally log files were developed for debugging purposes (Suneetha & Krishnamoorthi 2009, p.328). The web log is a file that is being created by the web server and each time a particular resource is requested it writes information to the log (Pani et al. 2011, p.19). When a user requests data from a server via an URL the server accepts the request and creates an entry (log) for this particular request before sending the web page to the visitor (Kaushik 2007). Thus it records the information about the activities on the server. But the first logs, server error logs, were developed to discover errors during the transmission of files through the web (Kaushik 2007) and not to provide information for Web Analytics (Kohavi et al. 2004, p.86). However, they were the original source of data collection of the web and their potential was recognised quite early. In 1995 Dr. Stephen Turner invented the first log analyser tool called Analog (Wu et al. 2009, p.163). As the number of accesses to web pages increases rapidly also the number of log entries increased (Kumar Jain et al. 2009). Peacock (2002, p.6) stated that "*No other survey technique generates as much data for so little effort. The trick is to turn it into useful information and practical applications*" (Peacock 2002, p.6). When managed properly log files can be a good data source for extracting interesting patterns about user behaviour (Pani et al. 2011, p.15) and for getting information to improve decision making (Kumar Jain et al. 2009). Used systematically log files can be a very effective tool for gathering passive feedback from users (Sullivan 1997).

Web server logs are plain texts (ASCII) and independent from the server platform. Different web servers may save different information within logs. Table 4 gives a short overview of this basic information.

Data	Example
IP address of the user who made the request	192.168.2.104
Time Stamp in the format [day/month/year: hour: minute: second zone]	[09/11/2011:09:15:38 - 0700]
Request of the client: Method information requested used protocol	GET/xxx.gif HTTP/01
Status code	e.g. status codes of the Hypertext Transfer Protocol: 200 – ok; 300 – multiple choices; 400 – bad request; 404 – not found; 500 – server error
Size of transferred object (in bytes)	2568
Referrer: URL of the site that the user have been referred from (preceding page)	http://www.xxx.de/index.thml
User Agent	Mozilla/7.0.1 [en] (Windows NT 6.0; I ;Nac)
User Name (for pages with registration)	Hans

Table 4: Information Captured within Log Files
(Suneetha & Krishnamoorthi 2009, p.329; Grace et al. 2011, p.102; Pani et al. 2011, p.20; Ogle 2010, p.2605; Sen et al. 2006, p.86)

In addition to the examples outlined above demographic information such as country information can also be captured. It is important to notice that it is not the personal information of the user that is captured at this time, but the technological details, hence the location of the server used by the user (Weischedel & Huizingh 2006, p.465).

The format this data is captured and displayed in thereby differs from the log file format used on the web server. Beside custom log formats which can be individually configured common log formats which also are a standard format for web server log files are the W3C Common Log Format (CLF) and the W3C Extended Log File Format (ECLF) (Pani et al. 2011). The ECLF thereby is the most common format used (Sen et al. 2006, p.85). The following gives an example of what an ECLF log file can look like:

> www.lyot.obspm.fr - - [01/Jan/97:23:12:24 +0000] "GET /index.html HTTP/1.0" 200 1220 "http://www.w3perl.com/softs/" "Mozilla/4.01 (X11; I; SunOS 5.3 sun4m)" (W3perl 2011)

Notwithstanding that different web server may save different types of information the basic information maintained are often similar. Because of this standardised way web server logs are saved it is also possible to analyse data from the past (Hassler 2010, p.58).

Besides the advantages of log files being able to collect huge amounts of data with little effort a number of challenges and limitations exist. As server logs capture each activity

of the server when requesting a site, one site request of a user can lead to several logs. If for example a site is requested which includes a picture, loading the picture will be a separate log. The same applies to CSS data or flash videos. Log image entries and similar ones need to be identified correctly and be removed for the purpose of usage mining (Pani et al. 2011, p.20; Hassler 2010, p.46). In addition logs that have been created because of web robots, spiders and crawlers (bots) need to be removed (Pani et al. 2011, p.21). Bots are automated programs which visit a website (Heaton 2002). They are for example used by web search engines or site monitoring software in order to see what is available at a site. As the number of bots today is enormous they can dramatically distort the results (Kohavi et al. 2004, p.95). Furthermore, if a page was cached and a user gets the cached page no server log will be written as the server does not get the request. Thus the measures would be underestimated. The same applies if users use the back button of the browser to switch between pages (Pani et al. 2011, p.21; Cooley et al. 1997).

Another limitation of server logs is the fact that the time spent on the last page cannot be measured. The request of the following website will generate a server log on the specific web server, but no hint is sent to the first server. Thus the time spent on the exit page cannot be recognised (Hofacker 2005, p.233). The identification of individual users is a different problem that exists. If users are not logged in to a site, the IP-address often is the information which is used to identify a user. But in a big company network for example multiple users share one IP-address and it cannot be separated between individual users (Weischedel & Huizingh 2006, p.465). Furthermore, with dynamic IP-addresses it happens that a user visiting a site for the second time will be seen as a new visitor only because his/her IP-address has changed since the last visit (Hassler 2010, p.47). Safe distinction of users can only be guaranteed by user authentification or when using cookies (Spiliopoulou 2000, p.129). A cookie is as "*a message given to a web browser by a web server. The browser stores the message in a text file. The message is then sent back to the server each time the browser requests a page from the server*" (Web Analytics Association). The problem here is that more and more users do not allow the storing of cookies or delete them regularly. Hassler (2010) refers to a study of an American Internet analysis company comScore (Abraham et al. 2007) which draws the conclusion that about one third of all Internet users regularly delete cookies and around 12 percent do not allow saving cookies at all. These users will distort/underestimate the measurements.

A few years ago the majority of web metrics were generated with the use of server log files. They are able to create an overview of the users' behaviour, existing problems

within the website and the used technology (Weischedel & Huizingh 2006, p.464). But within recent years other data collection techniques for Web Analytics have been developed (e.g. page tagging) which try to overcome the challenges and limitations of web server logs and even try to analyse more/different data. Log files only capture the server-side data within the moment of a request. But they do not have any information about what the user is doing in between clicking a new side and they also do not know which setting are used (Kaushik 2007, p.54). Because of little development in web logs and other positive innovations such as JavaScript tags, Kaushik (2007, p.27) recommends to now only use web logs to analyse search engine robot behaviour to measure success in search engine optimization. In all other cases other data capturing methods should be used.

5.1.2 JavaScript Tags

When using JavaScript tagging each website needs to have a short JavaScript code included. When a visitor requests an URL from a web server, the web server sends back the page, including the JavaScript code. This code is executed while the page loads. It captures different data such as page views and cookies and sent them to a data collection server. The variety of data which can be captured is huge. It ranges from clicks and position of the curser to mouse moves and keystrokes to the window size of the browsers and installed plug-ins. Besides, any information that can be captured by log files can also be captured with tagging (Kaushik 2007, p.54). Following is an example of the JavaScript code which needs to be included in each page when using the tool Piwik:

```
<!-- Piwik -->

<script type="text/javascript">

var pkBaseURL = (("https:" == document.location.protocol) ?
"https://{$PIWIK_URL}/" : "http://{$PIWIK_URL}/");

document.write(unescape("%3Cscript src='" + pkBaseURL + "piwik.js'
type='text/javascript'%3E%3C/script%3E"));

</script><script type="text/javascript">

try {

var piwikTracker = Piwik.getTracker(pkBaseURL + "piwik.php", {$IDSITE});

piwikTracker.trackPageView();

piwikTracker.enableLinkTracking();
```

```
} catch( err ) {}

</script>

<!-- End Piwik Code -->
```
(Source: Piwik 2011)

When using this data collection method it is possible to capture more data more accurately than with web beacons or log files. JavaScript tagging was developed at the end of the 1990's and at the moment it seems to be the most used data gathering technique (Kaushik 2007, p.30ff; Hassler 2010, p.29).

One clear benefit of JavaScript tagging is the possibility to also get data from cached pages. As the code is executed each time a websites loads it is also executed if the site is loaded from a cached space. In contrast, if users have switched off JavaScript, which 2 to 6 percent of users have, the data of these users will not be captured at all. The implementation effort of JavaScript tagging normally is easy. Indeed the Java code needs to be included in every single page, but it's only a few lines and therefore it is possible to control what data is being collected. But it should always be implemented within the footer of a web page because it should not happen that a page is not loading just because the tagging causes problems. Furthermore, it is possible that an ASP vendor will collect the data. Here the question of who owns the data needs to be clarified. But beside all this benefits JavaScript also has limitations. What was described about the cookies for log files is the same for tagging. If users switch them off or delete them, this information is lost. Furthermore, even if it is possible, it is much harder to capture data from downloads with tagging than with log files. PDF files for example do not include executable Java code. If they are requested through an URL of the web page, the request can be captured. But if they are directly opened from a search engine then the request will not be recognised. In addition, if a website already uses a lot of JavaScript on its site the tagging can cause conflicts and sometimes is not even possible (Kaushik 2007, p.32f; Hassler 2010, p.60f).

5.1.3 Web Beacons

Web beacons are 1 x 1 pixel transparent images which are sent from the web server along with the website and are then executed to send data to a data collection server. They were developed and are mostly used to measure data about banners, ads and e-mails and to track user across multiple websites. Web beacons are easy implemented in web pages around an <img src> thus an image HTML tag. They can capture data such as the page viewed, time, cookie values or referrers. As robots do not execute

image requests they won't be visible in web beacon data collection. But if users turn off image request in their e-mail programs and web browsers they also won't be measured. Furthermore, as web beacons are often coming from third-party servers they raise privacy issues and for example antispyware programs will delete them.

Web beacons are not as powerful as JavaScript tags. Thus they shouldn't be used as the main data collection methods. But they stand out when trying to get data across multiple websites (Kaushik 2007, p.28f).

5.1.4 Packet Sniffing

Packet sniffers are implemented between the user and the website server. A user requests a page. Before this request is processed by the website server it runs through software- or hardware-based packet sniffer which collects data. It is then passed through the website server. On the way back to the user the packet sniffer again is between them to collect data. Using this method of data collection will bring the most comprehensive amount of data as every communication is watched. Technical as well as business related data can be captured. Only cached pages will not be measured unless they have additional JavaScript tags. One clear benefit of packet sniffing in contrast to JavaScript tagging is that there is no need to touch the website, nothing on the actual website needs to be changed. But on the other hand additional hard- and software is needed which need to be installed, controlled and maintained. Furthermore, packet sniffing raises privacy issues. Using this method, raw packets of the Internet web server traffic is collected. This means that information such as passwords and addresses, etc. will be saved. It needs carful investigations to handle this data correctly (Kaushik 2007, p.35f).

Only a few Web Analytic tools support packet sniffing today and because of additional hard- and software it can become quite expensive. Therefore Kaushik (2007) recommends to only use packet sniffing if JavaScript tagging has significant shortfalls.

5.1.5 Data Collection Methods: Summary

The methods discussed above for data collection have limitations. And there are a few factors which all of them need to keep in mind. Probably the most important point is to make sure that the website is working for the user. The customer is first, not the analysis. If the analysis is not working that might be bad, but if the site is not working because of problems with the data collection this is even worse. Furthermore, the user privacy needs to be maintained. Privacy policies need to be in place and should be observed. Using web logs on the own server directly leads to the ownership of the log files. But with all other collection methods attention should be paid to the data

ownership if the analysis is given to a third party. Furthermore, if the website is hosted on several servers the data needs to be aggregated in order to get the whole overview. The issues with cookies and the measurement of the time spent on the last page as already outlined for log files are issues true for each collection method (Kaushik 2007, p.36f).

Regardless of the method that is used for data collection no way is 100% accurate. There is no one way to go (Kaushik 2007, p.37). As already outlined within the explanation of the different methods above each method has its own benefits and challenges. Web logs and JavaScript tags are the most used data collection techniques at the moment (Kaushik 2007, p.100) but there are big discussions whether to use the one or the other technique. Hints about when to use what method were given above, but a question to ask is "what do we want to get out of the data?". No technique will collect 'all' data and it is also hard to make solid statements about the quality of clickstream data. The web is a nonstandard environment. Websites and their technologies are constantly changing. Visitors use different mediums to access pages and not everything is captured in clickstream data due to different reasons (Kaushik 2007, p.108). Kaushik very clearly expressed the data quality in general by saying: "*Data quality on the Internet absolutely sucks*" (2007, p.109). But besides the importance of data quality it is more important how confident someone is with the data. If someone knows how the data was generated and what drawbacks the method has it is possible to still use the data in a good way. The important point is that the data leads to a decision so that it is possible to move forward (Kaushik 2007, p.110).

The data itself is not very useful until it is analysed. But even when coming up with different metrics the interpretation can be problematic. For example, does a user who looked at a lot of pages represent a satisfied user because he found a lot of information or is it a lost user who clicks around because he couldn't find what he was searching for (Sullivan 1997)? Or what does a long viewing time of a page say? The visitor may actually be reading the text, but it is also possible that he is making a phone call or is getting a coffee and the page is just open in the meantime (Weischedel & Huizingh 2006, p.465).

Even though it is not further outlined here, qualitative data collection techniques can further help to understand users. With the help of metrics the quantitative data can be analysed but they only show what happened at a specific site/a given timeframe. But they all miss to tell why something happened. To really understand user behaviour the quantitative analysis answering the 'what' (clicks, visitor count, etc.) and the qualitative

analysis answering the 'why' (intent, motivation) need to be combined (Kaushik 2007, p.13f). Besides, having competitive data and to know how competitors are doing within the business help to estimate whether the own business is doing well or not (Kaushik 2007, p.44).

5.2 Metrics

As outlined in the above section there are different ways to collect data for Web Usage Mining. The effort to get these data is not too high, but the analysis and interpretation are very time-consuming (Hasan et al. 2009, p.697). The first and most asked questions when talking about Web Analytics often are: "*How many visitors came to our site? How long did they stay? How many pages did they visit?*" (Kaushik 2007, p.132). In order to get objective answers to this questions defined measurable attributes can be used. These quantifiable attributes are called web metrics (Hassler 2010, p.90) or key performance indicators (KPI) (Burby & Atchison 2007). They help to understand how visitors are using a specific site (Weischedel & Huizingh 2006, p.463) and to answer these questions.

Defining business metrics is one of the first steps in the analytical process (Burby & Atchison 2007, p.49). Within their book Burby & Atchison (2007) write about business metrics as measurements of performance, specific for one organisation, which are based on the most important business and web goals and which lead to long-term considerations. They argue that most KPIs are financial in nature and give examples, such as the top-line revenue or higher average revenue per purchase. They do so because they look at Web Analytics only from a commercial business perspective having an e-commerce website.

However the types of metrics can be distinguished by the quantity, the ratio which looks at metrics and their relation and the value which e.g. represents an URL or search referrer (Hassler 2010, p.105). Most commonly the following web metrics are used:

Page View: The page view represents the number of times a single page was viewed (Web Analytics Association 2008). In the past **hits** which represented requests to the server were measured from log files. They are probably the oldest metric in Web Analytics (Hassler 2010, p.90). But since web pages nowadays include more than just text, one request can easy lead to 5 or more hits as every picture for example is counted separately. Thus, the metric hit is not that important any more. Instead the page as a whole is looked at. It is a good metric to estimate the general demand of a website. The question to ask is if it is good or bad to have a high number of page views. If the

structure of a website is not leading to answers the number of page views might be quite high. But people might not be happy with the site. Only from the number of page views it cannot be stated if a site is successful or not (Kaushik 2007, pp.9, 140). Kaushik points this out by saying *"... a page view is a page view; it is just a viewed page. Nothing more, nothing less*" (2007, p.77). One aspect which may become more and more important in the future is the information value of page views when using Rich Internet Applications like AJAX. They can lead to an underestimate of page views as it can happen, that a user only request one page and parts of the site are only reloaded while surfing (Hassler 2010, p.93).

Most viewed Page: Also called most popular pages or most requested URL the most viewed page simply counts which page had the greatest number of visits (Kaushik 2007, p.150).

Visitors and Visits/Sessions: These metrics are some of the basic ones in Web Analytics and provided by every tool. The name thereby often is different and ranges from visitors and visits, over total visitors or unique visitors to sessions and cookies. It is important to understand how a tool really measured the specific metric. Most commonly the differentiation is the following:

Visits/Session: represents the number of visits (sessions) on a website during a specified time period. One visit thereby is an interaction by an individual with one or more pages of a website (Kaushik 2007, p.133; Web Analytics Association 2008).

Visit Duration/Time on Site/Length of Visit/ Time Spent per Visitor: It is counted as the difference between the timestamp of the first minus the last activity in a session and shows the length of time in a session. This time-based metric often is hard to evaluate unless it is known what the user did in between. Did he really viewed the page or was he doing something else? (Web Analytics Association 2008; Kaushik 2007, p.136). Furthermore, when a user stays on a page for e.g. five minutes that can mean that he is interested, that he is distracted, or cannot find what he was looking for (Sullivan 1997).

Visitors: counts the number of users of a website (Hassler 2010, p.98). It is important to notice that for example the number of unique visitors can be quite different from the total number of visitors and the question needs to be asked how the visitors were identified (Kaushik 2007, p.132).

Unique Visitors (unique browser) tries to count the number of distinct users within a given timeframe. Each individual is only counted once. To distinguish between individuals authentication, IP-addresses or cookies can be used. All have different limitations so that it is important to understand how the number was made up in the specific case and to which timeframe it is counting (Web Analytics Association 2008; Kaushik 2007, p.133).

New Visitors shows the number of unique visitors who viewed the site for the first time ever (Web Analytics Association 2008).

Return(ing) Visitor shows the number of unique visitors who visit the site, but also have visited it before (Web Analytics Association 2008).

Visits per Visitor: The number of visits divided by the number of unique visitors in a given time period (Web Analytics Association 2008).

Click-through: This number indicates how often a link was clicked by a user (Web Analytics Association 2008).

Referrer: Referrer can show where users came from and what they searched for. The referring website shows the page where the user came from before viewing the current page (Web Analytics Association 2008). Search key phrases show what users searched for in search engines before they got to the current site. But it is not always possible to get referrer information. Around 40 to 60% of referrer data will be empty. Reasons are, among others, that the page was bookmarked or the entry was through direct type in of an URL (Kaushik 2007, p.145).

Entry Page: It shows the first page of a visit (the URL). As usually many entry pages exist it is often displayed by a list of URLs with the number of visits (Web Analytics Association 2008).

Exit Page: The exit page represents the end of a visit. It is the last page accessed on a site before leaving it (Web Analytics Association 2008).

Exit Link: Within a page often external links are included. When a user clicks an external link and exits the own website the link is called an exit link.

Top Exit Page: The top exit page is the page where most visitors leave the site from. The exit rate is the percentage number of how many people left from a specific page.

However, how can these numbers be interpreted? Why do people leave at a specific site? The answer can be either because they found what they were looking for (easy to discover with a thank you page of an e-commerce site) or they left because they couldn't find it (Kaushik 2007, p.9).

Page Exit Ratio: It is the number of exits from a specific page divided by total number of page views of that same page (Web Analytics Association 2008).

Landing Page: A page viewed as a result of the beginning of a users experience resulting from a marketing effort is defined as landing page (Web Analytics Association 2008).

Event: An event is every recorded action with a specific date and time directed to the browser or the server. It can be either seen as a count (number of events) or as a dimension (how many visits can be associated with a specific event) (Web Analytics Association 2008) and often is associated with a specific page or session.

Completion/Order Conversion Rate: The percentage of visitors who purchase on the site (Ogle 2010; Hasan et al. 2009, p.703).

Bounce Rate: Depending on the author the bounce rate is defined differently. It is either the percentage of single page visits (Hasan et al. 2009, p.701) or the percentage of visits that only stayed on the site for a very short time (Kaushik 2007).

Assessment of Content Popularity: It is a ranking of the content of a website e.g. the most frequently showed texts (Ogle 2010).

Path Analysis: The path analysis shows how users click though a site (Ogle 2010). It is one kind of pattern mining as used within pattern discovery.

This list could be endless, as each tool is able to generate hundreds of different metrics. For example it could go on with the most popular pages, internal referrers and internal search phrases. The list above should only give an idea of what is possible.

Each Web Analytic tool measures the metrics a little bit differently and two different tools will never have the same numbers for one website. Around 10 to 20% differences

are quite common. But as long as the numbers generated over time are from the same tool this is regardless. Furthermore, absolute numbers do not really matter. For example, how high the real number of visitors is, if it is 5,000 or 6,000 is not important, as long as the number is stable or growing (Hassler 2010, p.104). However, to really understand the metric and to be able to estimate inaccuracies it is always important to know how the analytic tool which is used measures the metric.

Besides, Hassler (2010, p.104) points out four important issues to keep in mind when using metrics and their measurement inaccuracies:

- Look at ratios rather than at absolute numbers
- When trying to compare metrics the same analytic tool should be used
- Keep in mind the measurement inaccuracies when comparing between websites
- Benchmarking on own measurements from today and the past are relatively resistant against measurement inaccuracies

The different metrics and the way they were measured are very important in the Web Analytic process. Especially identified sessions and users are key factors when it comes to pattern discovery and analysis (Cooley 2003). The whole process is further outlined in Section 5.3.

All in all, the quality of data can be as good as it may be possible and the metrics as exact as possible, but "*The data is not telling me what I should do*" (Kaushik 2007, p.6). Therefore analysts are needed which further interpret the data and draw conclusions from it. Furthermore, metrics miss the reporting of usability problems such as a lack of privacy or security or inconsistent design. Therefore heuristic evaluations could be used (Hasan et al. 2009, p.705). Metrics are a good starting point for Web Analytics, but they shouldn't be seen alone. Other analysis methods and further investigations of analysts are always required.

5.3 Web Usage Mining Process

The Web Usage Mining process applies Data Mining techniques to web data to identify user browsing trends and patterns on a website (Grace et al. 2011, p.106). To get reliable results Data Mining needs high quality data. But the data gathered in Web Mining through the different collection methods all are of rather low quality. Beside the below outlined step of preprocessing, sophisticated Data Mining further requires

knowledge about the content of each page, the linkage topology structure between the pages and captured user data like log files or page tagging data (Chi 2002, p.65).

Within the literature the process of Web Usage Mining is often described a little bit differently concerning the numbers of steps and which parts are taken into account. However the process looks alike, as already expressed earlier the content and structure of the analysed website are always needed as inputs (Cooley 2003, p.94). Srivastava et al. (2000) and Suneetha & Krishnamoorthi (2009) for example describe the process in three steps namely preprocessing, pattern discovery and pattern analysis. These are the most common parts which are described. But others like Pierrakos et al. (2003) also include the data collection as one step. Figure 7 gives an overview of the whole Web Usage Mining process.

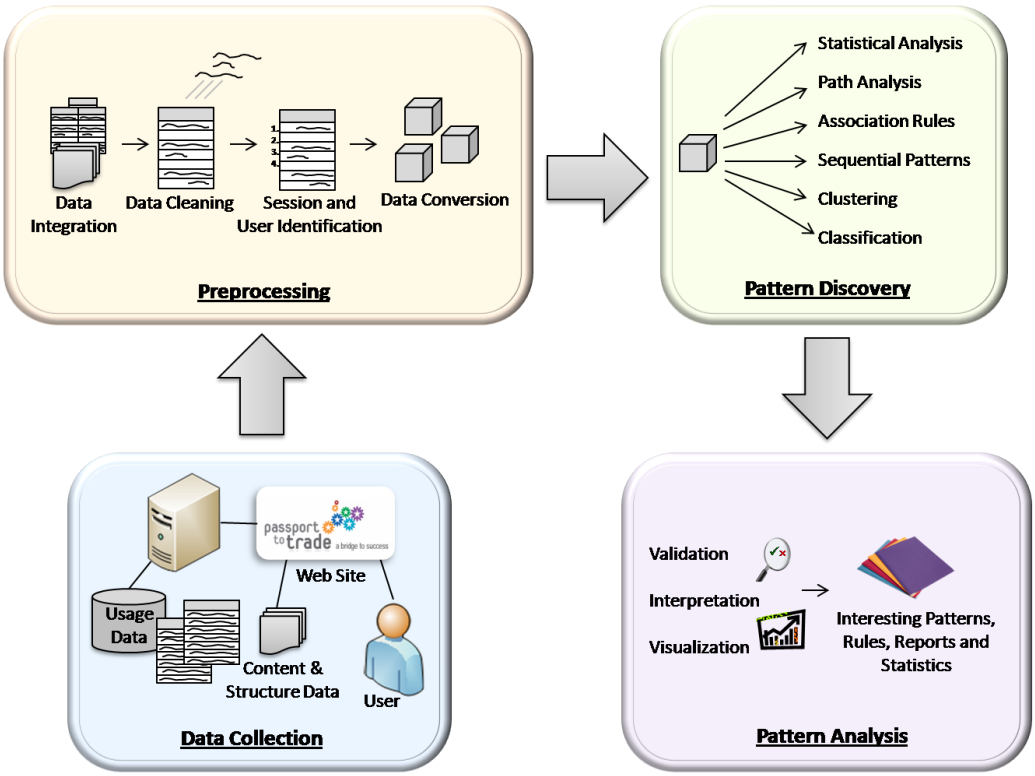


Figure 7: Web Usage Mining Process
(adapted from Cooley et al. 1997; Jaideep Srivastava et al. 2000; Pierrakos et al. 2003; Grace et al. 2011)

The different methods of **data collection** were already outlined in section 5.1. It includes the extraction of usage data from various sources (logs, tags, packet sniffer, etc.) and the identification of their content and structure (Pierrakos et al. 2003, p.320). This section will give an overview of the following steps: preprocessing, pattern discovery and pattern analysis and its sub stages.

5.3.1 Preprocessing

The purpose of the preprocessing or also called data preparation phase is to identify which collected data is useful for the mining process and the preparation of this data for further analysis and pattern discovery (Jaideep Srivastava et al. 2000). The previously collected data should restore the user's activities in different sequences (Spiliopoulou 2000, p.130). The content and structure of the website thereby is important in order to identify which data is relevant (Pierrakos et al. 2003, p.327). The overall task can be divided into several sub tasks. Same as in the overall Web Usage Mining Process, the different steps within preprocessing are divided differently by the various researchers. Overall the following tasks can be identified in order to generate good data that can be used as input for pattern discovery: Data Integration, Data Cleaning, Session and User Identification and Data Conversation.

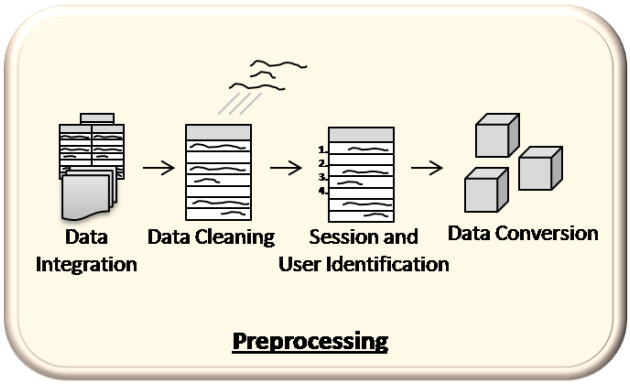


Figure 8: Steps within Preprocessing

Data Integration

If data were captured at different places or different data is available then these data need to come together. For example, where log files are used as the data capturing method and the website runs on different servers all log files from all servers need to be brought together in order to get the whole picture of web usage. Or, if there is the possibility to register on the website, the registration information should also be at hand for further processing and analysis (Cooley et al. 1997; Pani et al. 2011) if the permission of the user was given.

Data Cleaning

The collected data (raw web data) often cannot be used for Web Analytics as it is. It needs previous processing (Grace et al. 2011, p.106). Especially log files from web servers and proxy entries can be very noisy as all user interactions are recorded. Redundant or irrelevant data like image logs should be filtered out from the data set (Pierrakos et al. 2003, p.327). Undesirable entries within the collected data need to be removed (Pani et al. 2011, p.20). Entries with "error" or "failure" messages might be interesting to find technical problems and missing content, but should be removed for

the purpose of user pattern discovery. Following, the short list below gives an overview of what data should be removed (Pierrakos et al. 2003; Pani et al. 2011; Suneetha & Krishnamoorthi 2009; Cooley et al. 1997):

- Image (e.g. gif, jpg or png files) and multimedia logs
- Entries from search engine agents (robots)
- Error/Failure messages
- Logs of page style files

Session and User identification (Transaction Identification/Data Conversion)

This step is often seen as two separate activities distinguishing between finding users (user identification) and building sessions (session identification) (Pani et al. 2011, p.20). Within Data Mining this step is very important for the success as the identified users and sessions are used as the basic building blocks for pattern discovery (Pierrakos et al. 2003, p.327). There are different ways to identify different users and to separate them into categories. If every user needs to log in at a website in order to see/use it, the registration information can easily be used to distinguish between users. Otherwise, the IP-address or cookies are often used. When using IP-addresses different methods or rather definitions of who/what a new user is can be defined. Assigning a different user to each different existing IP-address might be the simplest approach. But due to the use of proxy servers (Pierrakos et al. 2003, p.328) and the fact that different users can use the same IP-address and one user can use different IP-addresses (Etminani et al. 2009, p.397) this method is quite inaccuracy. Other approaches go further and do not only look at the IP-address, but also on the used operating system and browsing software. If the IP-address is the same, but the used systems are different the assumption is made that each different agent type for an IP-address is a different user (Suneetha & Krishnamoorthi 2009, p.330). Another method to identify different users is the use of cookies. The pros and cons of cookies were already discussed in Section 5.1. Today they are an often used and reliable source for user identification. Even though they also have some issues they are still more accurate for user identification than IP-addresses (Hassler 2010, p.54).

Beside the identification of users, sessions need to be identified in order to understand the behaviour of single users. "*A "session" is a sequence of consecutive URL requests performed by the same visitor*" (Spiliopoulou & Pohle 2001, p.95). Thus, as a session

describes the activity/page references of one user at a given timeframe the user identification is required for session identification.

When users and sessions have been identified it is further possible (not required) to classify them into different groups. Berthon et al. (1996) for example distinguish between “short time visitors”, “active investigators” and “customers”. A short time user thereby only stays at a site for a short moment, an active investigator explores the site and a customer is a user who uses a service or buys a product. With this distinction Berthon et al. tried to separate the users with the help of the contact they had with the site. This might be useful in order to analyse reasons why different users interact differently because experienced customers might interact differently with the site than new visitors do (Spiliopoulou et al. 2000). Besides Pitkow (1997) following Novak and Hoffman (1997) differentiate between unidentified, session, tracked, and identified visitors. An unidentified visitor would be a user who visits a website and where no information would be available. Due to logs etc. this does not really exist within the web. A session visitor is a user who was identified through e.g. cookies or IP-address. A tracked visitor is uniquely and reliable identified across multiple visits to a site. And a tracked visitor where additional information is known is an identified visitor.

Data Conversion

Within the step of data conversation or also called formatting the data should be modified so that it has a suitable format for the next step of pattern discovery (Etminani et al. 2009, p.397).

5.3.2 Pattern Discovery

Pattern Discovery or Pattern Extraction is the key component of Web Usage Mining (Pani et al. 2011, p.22; Suneetha & Krishnamoorthi 2009, p.329). Navigation patterns can show how different user groups are using a website. They describe the routes users took along with statistics of, for example, how often a specific route was clicked (Spiliopoulou 2000, p.133). It is not possible to build a static site that fits all users but instead expected and unexpected patterns that show a user’s trace should be identified so that they can be supported in the right manner (Spiliopoulou 2000, p.134). In order to mine collected data analytic tools for user pattern discovery resort to different machine

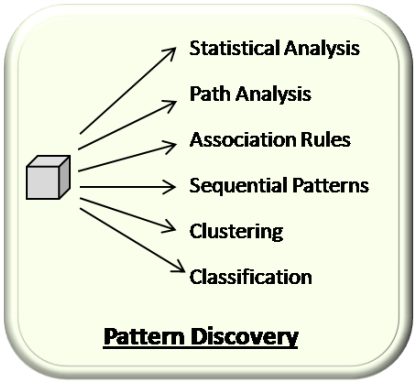


Figure 9: Steps within Pattern Discovery

learning and statistical techniques from AI, Data Mining, psychology, and information theory (Cooley et al. 1997; Pierrakos et al. 2003; Jaideep Srivastava et al. 2000). Thereby multiple kinds of pattern mining exist, including statistical analysis, path analysis, discovery of association rules and sequential patterns, clustering and classification (Cooley et al. 1997; Grace et al. 2011).

Statistical Analysis

Within the statistical analysis metrics such as frequency analysis, mean, median, etc. are used (Etminani et al. 2009, p.397). They build the most common method for knowledge extraction about visitors to a website. Therefore the different analyses are often performed on the metrics like page views as described in Section 5.2. In spite of the lack in depth which this method has, it can be useful for system performance improvements and support in marketing decisions (Jaideep Srivastava et al. 2000).

Path Analysis

The aim of path analysis is to reconstruct the sequence of pages a user visited (Pitkow 1997). Thereby it is possible to identify the initial start page, the steps a visitor took and the exit page (Ogle 2010, p.2606). Commonly used for visualization of path analysis are graph models which show the trace users click through a website. Thereby each node in the tree represents a web page and each edge a link between websites (Grace et al. 2011, p.107). Another way of visualization is to use a hierarchical tree. Parent-child relationships are done via a depth-first traversal of the pages. The click depth, most frequently visited path, and other metrics like entry and exit page are then also often used here for further analysis (Pitkow 1997; Cooley et al. 1997).

Association Rules

The different techniques of association rules are generally applied to large databases of transactions (Agrawal & Srikant 1994). However whichever algorithm for finding the rules is used, all techniques aim to predict the correlation of items (Grace et al. 2011, p.107). In Web Mining this means correlations among pages that were accessed together by a client are searched for (Etminani et al. 2009, p.397). An example of an association rule could look like this: "*x% of client who viewed the Web page x also looked at y"* (Cooley et al. 1997). The pages that were looked at together do not necessarily have to be directly connected to each other, but were looked at together (Jaideep Srivastava et al. 2000). Two correlation pages are called an itemset (Pierrakos et al. 2003).

Sequential Patterns

According to Kumar Jain et al. (2009, p.1) sequential access pattern mining in a sequence database was firstly introduced by Agrawal and Srikant (1995). Association rules do not look at times or take them into account. Within a database of sequences, where each sequence is a list of transactions with a specific time and each transaction is a set of items sequential patterns which also consists of a list of sets of items can be analysed (Srikant & Agrawal 1996). In other words, the search is for the existence of a set of items followed by another item in a specific time frame (inter-session). X% of clients who visited the page x had done a search in y, within the past week on keyword w could be one example for a sequential pattern (Cooley et al. 1997).

Classification

When establishing a classification rule a profile of items belonging to a particular group according to their attributes need to be found. With classification in Web Usage Mining a profile of a specific user group should be developed based on their access pattern (Cooley et al. 1997). In classification predefined classes exist and a data item is mapped to one of them (Etminani et al. 2009, p.397).

Clustering

"*Cluster analysis is the art of finding groups in data*" (Kaufman & Rousseeuw 1990, p.1). In the web area the aim of clustering can be to group users together, who have similar characteristics or to cluster pages which have similar content (Cooley et al. 1997; Jaideep Srivastava et al. 2000). A specified group is called a cluster. Thereby it is possible that different clusters overlap in specific parts thus one user can belong to more than one group (Paliouras et al. 2000). An overview of the different clustering methods can be found in the paper by Pierrakos et al. (2003).

Within the literature classification and clustering are often not differentiated (Kaufman & Rousseeuw 1990; Kaushik 2007). But when looking at the description of Cooley et al. (1997) the difference can be seen. Within classification, characteristics of a given group are searched for. With clustering the aim is to group users on the basis of their characteristics.

5.3.3 Pattern Analysis

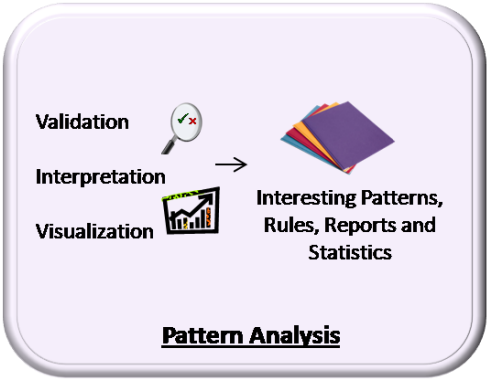


Figure 10: Steps in Pattern Analysis

Through the different ways of pattern discovery several types of patterns can be discovered. In order to find the interesting patterns and rules and generate reports and statistics the step of pattern analysis is needed. The pattern analysis phase helps to get a better understanding of the previously processed output (Cooley et al. 1997). Within pattern analysis the results of the pattern discovery need to be validated, interpreted and visualized (Etminani et al. 2009; Grace et al. 2011). When validating the identified pattern irrelevant patterns and rules need to be sorted out so that the interesting data can be identified for further processing. As the output of pattern discovery mainly has a mathematical form, the interesting patterns need to be converted for direct human interpretations.

Different forms of pattern analysis can be used, but the most common consists of knowledge query mechanisms such as SQL (Structural Query Language) or applying OLAP (OnLine Analytical Processing) operations to a data cube. Visualization of the end results then is often done via graphing or assigning colours to different values (Grace et al. 2011, p.108; Jaideep Srivastava et al. 2000). This can help to get a global picture of the results of the data analysis (Chi 2002, p.65). Today recent tools mostly include the development of many different reports and statistics as well as visualization techniques. Often the output of the tools is so large it is hard to keep track. More details about the tools and their possibilities can be found in Chapter 6.

5.4 Legal Issues

Whenever it comes to the storage of user information legal issues around privacy concerns arise. What makes it particularly difficult in the web domain is that websites mostly are available from more than one specific country. An American website for example, regardless if implicitly or explicitly, also addresses European users and thus needs to handle international legal regulations. A problem here is that different countries have different laws and regulations. The USA for example has soft laws concerning the protection of user information whereas the regulations in Europe are much more restrictive (Schubert et al. 2004; Hassler 2010). In the following some of the

main legal aspects are outlined. The discussion represents the German viewpoint but these issues are also similar for many other countries, especially countries within the European Union.

When looking at Web Analytics the area where it can get legally problematic is the data protection of user information. Conducting Web Analytics means saving user data. Impersonal or anonymous data of any kind are not problematic, but as soon as it comes to personal data that is clearly assignable to a natural person general privacy policies are affected (Hassler 2010, p.71). It needs to be distinguished between explicit profiles (identification profiles) where the user is recognised or deliberately adds data to a system and implicit profiles (interaction/transaction profiles) that are stored and generated in the background. Personal data via e.g. login can be saved and used, if the user grants permission. Permission can be asked for with a checkbox, but the checkbox must use the opt-in method. That means that the user must check the box by himself. It is not allowed that the box is already checked (Hassler 2010, p.72). But if the user does not need to login at the site it is problematic to ask him.

The storage and analysis of user data such as date, time, page views etc. are not legally forbidden, but when it comes to IP-addresses it becomes problematic. The German Telemedia Act (Telemediengesetz) for example states in §15, para.1 that the collection and usage of personal user data only is allowed in the extent which is necessary to use the telemedia (and billing purpose). It further defines personal user data as data that allows the identification of a user (Bundesministerium der Justiz 2007). From this the IP-address can be seen as within the gray area between personal and anonymous in Germany. Usually the IP-address can only be assigned to a specific person by the Internet service provider, but in some cases it is also possible for the website operator to do so. One example is that a private person has a fixed IP-address and has his or her own website running. The domain is registered by a registry and opens the way for identification. Different courts have made different decisions whether to see the IP-address as personal or not. One example is the judgment of the Landesgericht Berlin from the 6th September 2007 (23 S 3/07) which says that IP-addresses are personal information and are not allowed to be stored (Telemedicus 2007). A final decision made by a higher court is still awaited (Hassler 2010, p.73).

Due to the ambiguous legal situation the Düsseldorfer Kreis, the highest German authority for data protection in the non-public sector, advises not to save IP-addresses in connection with personal user data. One may indeed save user data using pseudonyms,

but IP-addresses are not pseudonyms in the sense of the Telemedia Act. However, this is only an advice and not binding (LDI NRW 2009).

In general it is not said that Web Analytic tool in general are illegal. But it might be safer to use those tools, which for example use cookies instead of IP-addresses when creating user profiles and do not save IP-addresses. As Schubert et al. (2004) quoting Briner 2002 and Beariswyl and Rudin 2002 say: "not everything that could be done can be done legally".

Therefore the suggestions given by Hassler (2010, p.73ff) provide good guidance:

- Include a data privacy statement on the website
- Use cookies instead of IP-addresses
- Do not combine user data with personal data
- Include an option to turn off the tracking
- Consider the location of the data storage

Within Web Analytics there are unique privacy issues that arise. The Web Analytics Association tries to outline and discuss the initiatives and issues as well as best practices on a dedicated webpage (Web Analytic Association 2011).

5.5 Summary of Key Challenges

Most challenges have already been described between the lines within this whole chapter. This section provides a short summary of all the challenges mentioned in order to draw them together and to stress their importance again. The most important points where problems and challenges occur are:

- The identification of visitors
- Filtering of bots
- Issues with caching
- Data overload
- Type of metrics
- Time on last page
- Appropriate tool
- Broad overview of whole site
- Interpretation of findings
- Privacy issues

The identification of visitors can be done though IP-addresses, caching or user profiles if the user needs to log in. Each method has its drawbacks and with all it needs to be

kept in mind, that they won't be 100% accurate and count each single individual user. The user profile might be the easiest and most accurate way, as the users identify themselves. But if they do not allow using this information for web analytic purposes, this data cannot be used. Furthermore, it must be taken care, that sensitive information like the password is not connected with the analytical data. When using caching it might be that we will not get the information of all users if people might protect themself from caching. Finally, when using the IP-address privacy issues might arise and same users might actually be different once and the other way around as already outlined above.

In particular, when using log files for Web Analytics spiders and bots need to be identified and removed before analysing the data. If not, they will dramatically blur human traces and distort the statistics. Another difficulty lies in caching. The only data collection method which is able to also measure cached websites which are viewed by users is JavaScript tagging. All other methods miss these pages which lead to too low results in the statistics.

The volumes of data to measure can be another problem area. Not all tools are capable of handling very large amounts of data (Sen et al. 2006, p.87). Especially when not analysing in real time, but trying to analyse heaps of data from the past that can cause problems.

As discussed earlier there are hundreds of metrics available within Web Analytic tools. It is not always easy to understand the differences between metrics and find the appropriate one for the right purpose (Sen et al. 2006, p.87). It furthermore needs to be clear, how the metrics were developed. The time which is spent on the last page before leaving for example cannot be measured accurately. As there is no incoming information from the next server most tools determine a session after 29 minutes of inactivity (Kaushik 2007, p.36).

The market for Web Analytic tools has risen within the last years. But which tool is the appropriate one and where are the differences? An answer to this question is given in Chapter 6. But beside the analytical part which is achieved with the tool there is a need to understand the website as a whole, including its content and structure (Cooley 2003, p.94). It is required to understand the metrics and to generate useful findings from the statistics. Often many analyses stop with the findings of how many people visited the website and similar. But really good Web Analytic investigations should go behind that and try to conclude with action for e.g. redesigning the website or getting ideas for marketing campaigns.

5.6 Conclusion

Aside from changes in the environment, like the introduction of a new product, understanding the user behaviour of a website is the main reason why websites are subject to changes today (Weischedel & Huizingh 2006, p.463). In the past consumers were mostly passive elements in the overall ecosystems. With the rise of the internet this has changed completely (Burby & Atchison 2007, p.6). Today they are very important participants and can provide much interesting data. Thus, it is desirable to understand how to support them best. With Web Analytics it is possible to go along the entire process and to see a website from the perspective of its users (Spiliopoulou 2000, p.133). Getting to know how users access a site and which path they are taking is critical for the development of effective marketing strategies as it allows organisations to predict user visits. It further helps to optimize the logical structure of a website (Cooley et al. 1997). "*improving Web communication is essential to better satisfy the objectives of both the Web site and its target audience"* (Norguet et al. 2006, p.430).

In the past WA has widely been used in the economic field (Wu et al. 2009, p.163) and most analytical investigations which are described within the literature were done for commercial websites. Nevertheless there is no reason why Web Analytics cannot also be helpful for other kinds of websites. There might be areas which are much easier to determine when looking at e-commerce websites like the successful visit of a user with a purchase, but there are lots of other areas that can also be used for the other kinds of websites.

It needs to be kept in mind that Web Analytics only is an analysis. It's not an exact science. No numbers will 100% show the reality. But "*It is better to be approximately right than precisely wrong*" (Hassler 2010, p.34f) because the approximately data can still be used in order to draw conclusions. It cannot give clear answers as no direct answers from users are given. With the help of analytic tools it might be easy to quantify a site, but the data needs to be interpreted by an analyst who takes action for improvements (Ogle 2010, p.2604).

"*In web analytics, "going wrong" often means just going halfway*." Very often vast amounts of money are invested into tools, but in the end only reports are produced and nothing more (Burby & Atchison 2007, p.43). The part of taking action after the analysis might be the most critical aspect in Web Analytics but is often neglected.

6 Review of WA Tools

The first generation of Web Analytic tools can already be found at the beginning of the World Wide Web. One of the first mining tools, called Analog was developed by Stephen Turner in 1995 (Wu et al. 2009). Kaushik calls this point the time of birth of Web Analytics (2007, p.2). Back then analytic tools were standalone systems which analysed web server log files. They were able to report basic metrics like hits or page views (Sen et al. 2006, p.89). With the diffusion of the Internet also new tools for Web Analytics emerged. One of the first commercial WA tools was WebTrends (Kaushik 2007, p.3). Today the complexity and diversity of technologies that support Web Analytics is enormous and a second generation of tools has emerged. Beside the basic metrics, tools nowadays collect clickstream data, track user pattern and present the data as meaningful information. Even though the broader use of Web Analytic tools is relatively new (Nakatani & Chuang 2011), the number of tools available on the market is extensive and diverse (Hassler 2010, p.39).

Section 6.1 of this chapter first gives an overview of the WA market. A discussion about tools with the primary focus for the usage in SMEs is then presented in Section 6.2. Section 6.3 summarises and concludes this chapter.

6.1 The Market

The market for Web Analytic tools is quite large. There are many products available which can be differentiated by the level of sophistication, the capabilities, the qualities and the reliability (Nakatani & Chuang 2011). Server logs which provided the raw data to be worked with were only the beginning of analysis. The amount of reports produced could and can be overwhelming (Weischedel & Huizingh 2006, p.464). Not only the data collection methods have changed since the beginnings, but also the kind of tools that are available. Before 2005 Web Analytic tools represented a market which was basically dominated by US products. With the release of Google's free tool Google Analytics in November 2005 the market grew (Hassler 2010, p.15). In 2007 four big commercial vendors could be found: Coremetrics, Omniture, WebTrends and WebSideStory, a number of mid-market vendors such as Unica, indexTools and ClickTracks and a range of open source products such as AWStats, Webalizer and StatCounter (Kaushik 2007, p.5). At this time, WebTrends was the most popular Web Analytic tool (Norguet et al. 2006, p.437). In particular, the introduction of very

powerful free tools increased the pressure on the market because of the price aspect (Lovett et al. 2009). Today the number of vendors that are very strong on the market is quite small when comparing with the total number. Only a few vendors have a total market share of 78% and the market penetration is around 58% (Lovett et al. 2009).

Adobe, Google, IBM and Webtrends currently represent the top-tier global vendors for WA (Gassman 2011, p.1). Adobe counts 6,000 WA customers which generate $500 million revenue annually (Gassman 2011, p.4) while Google Analytics is the most popular WA tool in the world (Stanhope et al. 2011) with over 200,000 organizations using the free tool Google Analytics Standard. With the acquisition of Coremetrics and Unica (NetInsight) in 2010 IBM stepped into the field of WA as well. With both products together they have around 3,000 customers. Webtrends represents the largest Web Analytics pure-play vendor with more than 3,500 customers.

Beside these big vendors new ones are emerging which have specialized offerings. Currently they do not have significant market share, but it is estimated that the market will become subject to change around 2016. By now the big vendors present barriers for new vendors in the general business. But in the long term it might come, that WA features will be offered through advertising agencies, web content management systems and e-commerce engines (Gassman 2011, p.3).

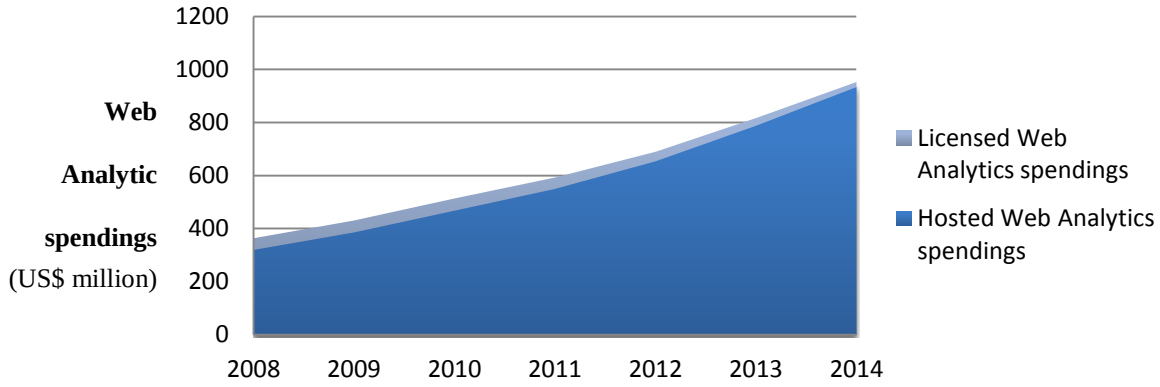
Reviewing the literature (e.g. Hassler 2010, p.341; Gassman 2011) and browsing through the web, many websites can be found which describe "The best tools", or the most used ones[2]. Reducing these to a common group it can be stated that the most used Web Analytic tools on the market right now are the following:

- Adobe SiteCatalyst (Omniture)*
- Analyzer NX (AT Internet)*
- Click Tale
- comScore Digital Analytix*
- Coremetrics Analytics (IBM)*
- Etracker
- Google Analytics*
- NetInsight (IBM)
- Netmind GO
- Piwik
- Sitestats
- StatCounter
- SiteCensus
- Webtrekk Q3
- Webtrends Analytics*
- Yahoo! Web Analytics*

[2] Examples are ttp://w3techs.com/technologies/history_overview/traffic _analysis, http://www.idealobserver.com/tools/web-analyse/, or http://www.kaushik.net/ avinash/web-analytics-tools-comparison-a-recommendation/.

Tools marked with an asterisk (*) are those tools most often used by enterprise-class clients with revenue of more than US$10 million.

More and more companies want to get insight of their web presence and start using Web Analytic tools (Gassman 2011). A study by Forrester Research estimated the worldwide market for Web Analytic tools to US$292 million in 2004 and a 27% growth rate for 2005 (Sen et al. 2006, p.86). A more recent study even estimates that US businesses alone will spend around US$953 million for Web Analytic software in 2014. Further, the average compound annual grows rate for the next five years within the US businesses will be 17% (Lovett et al. 2009). These predictions include expenditures for Web Analytic software, including licenses, service fees, installation and support from technology vendors and resell partners. Not included are spending for professional services from non-vendor consultants, agencies, or external third parties. As shown in Figure 11, it can be recognized, that the trend is clearly going towards hosted offerings.



								CAGR
Licensed Web analytics spending	$44	$45	$46	$43	$36	$29	$19	-16%
Hosted Web analytic spending	$320	$386	$468	$550	$654	$788	$934	20%
Total	**$364**	**$431**	**$514**	**$593**	**$690**	**$817**	**§953**	**17%**

Figure 11: Web Analytic Spending in the US
(adapted from Lovett et al. 2009)

In 2009 around 73% of US businesses used or piloted Web Analytic software. Thereby only 6% of these account for 38% of the total forecast revenue while operation websites with 2 million or more unique visitors per month. Because most heavily used websites already use Web Analytics the growth is estimated to come from new entrants with moderate traffic which means 500,000 to 2 million unique monthly visitors.

Considering a Garter report from 2010 (Herschel 2010) another source of growth can be found with the use of advanced functions. Even if 90% of the addressable market is already using WA tools in any manner, less than 50% use advanced functions.

The distribution of the Web Analytic spending is very interesting. Very large website owners which can record more than 10 million unique visitors a month spend more than $200,000 annually on Web Analytic technologies (Lovett et al. 2009). Within a recent Gartner report Gassman estimates a budget of at least $250,000 per annum to properly fund a WA programme and points out that some organisations even spend millions of dollars (Gassman 2009a, p.2). However, the industry average spending is around $15,000 per year (Lovett et al. 2009). The most common pricing method for WA tools thereby is counting the tags (page views or server calls) received by the analytics server with a combination of add-on charges (Gassman 2011, p.2).

One study which was conducted in 2011 and looked at the German Web Analytic usage asked members of the Web Analytic professional group on XING to get insights into the usage. The study by Zumstein et al. (2011) reports, that WA is mostly used within IT and knowledge-based industries as well as within the service sector. Furthermore, it appears that not only large companies, but also SMEs use Web Analytics. The five most used tools within this German study have been Google Analytics, Etracker, Omniture, Webtrends and Piwik. However, often companies not only use one, but more systems at the same time (see Figure 12).

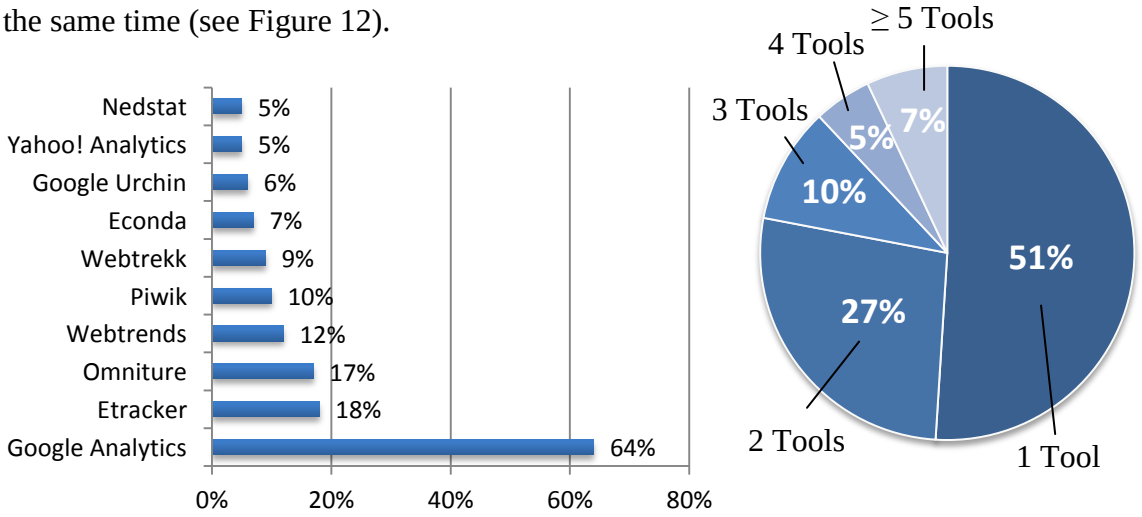


Figure 12: Tool Usage in Germany
(adapted from Zumstein et al. 2011)

6.2 Tool Comparison for SMEs

Even though the most used Web Analytic tools can be broken down to a manageable amount, a major challenge is the selection of an appropriate WA tool. Most tools are capable of working with almost every possible development scenario and track a lot of different things out of the box (Burby & Atchison 2007, p.46). To keep track of all the tools and to recognize the differences is not an easy task. Therefore it can help to classify the tools into different groups. One common practice is to differentiate on the basis of the data collection method – log files or page tagging. But as most major WA tools do not support log file analysis anymore and most are going towards the page tagging method (Gassman 2011, p.2) this differentiation often is not helpful anymore. Another way is to look at how the functions can be accessed. Is it software as a service (SaaS) through an application service provider (ASP) or in house installed software? Further it can be separated by the availability. Some tools can analyse and report in a near real-time manner. Others need 15 minutes or more to show results (Nakatani & Chuang 2011). Another way to distinguish is to look at the expenses. Especially for SMEs as well as non-profit organizations it might be interesting to use a free tool instead of a commercial one.

To assist in understanding the nature and scope of different tools an in-depth analysis and evaluation of major tools was undertaken on the basis of a feature analysis. This is quite a common method which is used when comparing information technology (IT) (Nakatani & Chuang 2011). Section 6.2.1 describes the selection of the analysed tools. The development of a criteria matrix which was used for tool analysis is outlined in Section 6.2.2 before the evaluation is presented (Section 6.2.3). At the end of this chapter the limitations of the used method are described (Section 6.2.4) and a conclusion is given (Section 6.3).

The analysis and evaluation presented in this chapter here was influenced by the Passport to Trade project outlined in Chapter 7. Therefore different restrictions and assumptions were made which are further outlined at the corresponding sections.

6.2.1 Selection of Tools

The focus of this thesis is on SMEs and non-profit organizations where financial resources and expertise are more limited. Thus the selection of tools which were analysed concerning their functionalities was guided by different factors like usage, costs and suitability concerning SMEs.

In order to identify appropriate tools different sources and criteria were used. Within different countries different tools can be found on the top used tool list. With the help of a review of current usage statistics from .uk, .de and .com websites a general overview was established in a first stage. This selection was guided by the P2T project. It is a project within the European Union Leonardo programme and is coordinated by a British University. Therefore usage statistics from .uk sites were taken into account and these were complemented by .de sites as the author and the project team working on the Web Analytics part of the P2T project are from Germany. Last, as the .com domain is the most common one it was also taken into account in order to get a general impression about the whole market.

In a second step the price for usage of those tools were checked. As mentioned previously most SMEs often have limited and small budgets and conducting WA is usually not their first priority. Therefore it is conceivable that a small amount of money is spent for Web Analytic software, but it should be kept within limits. In conjunction with this data about the tools was investigated – is it available for installation and/or in depth testing? If no trial versions were available or too little descriptive data was at hand, the tool could not be further analysed. A primary focus has thus been placed on freeware and less expensive tools.

Furthermore the kind of data which the tools are able to analyse was a criterion. At the old businesculture.org website from the P2T project for example only log files were saved. Even though this method no longer is the most common one it still is one way to do the analysis so that one requirement was to also include at least one log-file based Web Analytic tool.

Finally, it is further possible to distinguish the tools on the basis of their purpose. Some vendors explicitly describe that their tool is best for e-commerce website tracking. This is outside the scope of investigation for this thesis. Many studies have already been undertaken for e-commerce websites and the focus of this thesis is on informational websites.

Finally the following list of 15 tools was derived and the tools were analysed and compared:

- Analyzer NX
- AWStats
- Clicky
- Google Analytics
- Logaholic
- Open Web
- Site Stats
- StatCounter
- W3Statistics

- Deep Log Analyzer
- Etracker
- Analytics
- Piwik
- Saw Mill
- Webalizer
- Webtrekk Q3

Tools like Omniture, Webtrends or Yahoo! Web Analytics were excluded due to the above described reasons e.g. pricing or specialization.

6.2.2 Criteria Matrix

In order to make existing differences in tools clearly and quickly visible and understandable a criteria matrix was developed within an iterative process using themes identified in the literature review and an in-depth functional analysis of the above outlined WA tools.

The literature review conducted in Chapter 5 already contained issues such as, metrics and privacy which could be used in an analysis of the tools. They provided the basis for the first step in developing the criteria matrix. Moreover, when the descriptions of the different tools were analysed and the first experiences with the usage of the tools made, further criteria arose that were added to the matrix. Additionally earlier evaluations and their criteria[3] were considered. However, these internet sites were only used for getting ideas of possible criteria. No further content has been used from these sources, as the reliability is unknown. A consolidation of all these sources accompanied by an evaluation by the project team with regard to usefulness for SMEs, measurability and comparability lead to the final criteria which are included in the matrix (Figure 13).

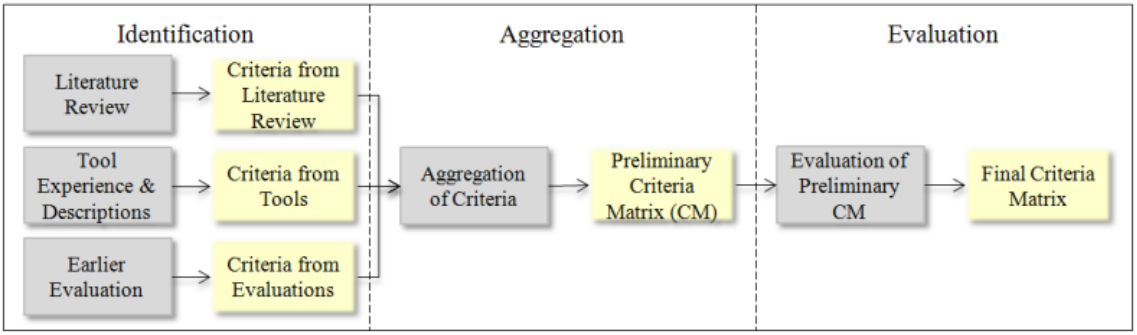


Figure 13: Research Steps for the Criteria Matrix Development

[3] Examples are http://www.idealobserver.com/tools/web-analyse/tabellarischer-tool-vergleich or http://web-analytics-review.toptenreviews.com

The resulting matrix comprises 69 criteria and information which addresses various aspects organised into 7 different criteria groups: general information, input, visitor details, traffic, monitoring & reports, events and data privacy. Following aspects are listed within the different groups:

General Information	▪ Tool name ▪ Manufacturer ▪ Version ▪ Keeping of data ▪ License model ▪ Price ▪ Admin/User concept ▪ Multi-language support ▪ Support ▪ E-Mail reports
Input	▪ Data source ▪ Technique ▪ Supported standards ▪ Import standards ▪ Export standards
Traffic	▪ Entry page ▪ Page views per user ▪ Unique visitor ▪ New/Returning visitor ▪ Events per page ▪ Events per session ▪ Time spent per page ▪ Time spent per session ▪ Exit page ▪ Exit link ▪ Average visits per day/month/year ▪ Total visits ▪ Most popular pages
Visitor Details	▪ IP address ▪ Country ▪ City ▪ Operation System ▪ Internet service provider ▪ Internet speed ▪ Browser ▪ Browser language ▪ Screen Resolution ▪ Cookies Enabled ▪ JavaScript Enabled ▪ Username
Data Privacy	▪ IP-masking ▪ Cookie deactivation ▪ Certificate
Monitoring & Reports	▪ Click path ▪ Bot/Spider filtering ▪ Bounce rates ▪ Live monitoring ▪ Visualization o Diagrams o Click map o Mouse tracking video o Heat map o Visitor origins map ▪ Multi-domain

Events	▪ Internal referrer ▪ Internal search keywords ▪ External referrer ▪ External search keywords ▪ Visit history ▪ Forms ▪ Downloads ▪ Flash events ▪ Multimedia streams ▪ Error pages ▪ RSS feed subscriptions		tracking ▪ Social media monitoring ▪ Mobile browser monitoring ▪ Mobile support ▪ Customizable dashboard ▪ Goal definition

Table 5: Criteria within the Criteria Matrix

A lot of aspects like name, price etc. are self-explanatory. Other terms like visits or sessions have been described earlier in the literature review chapters. The remaining terms are described in the following.

Within the "General Information" the issue of "Keeping of data" asks for the way, the data is stored. Is it in-house, at the ASP or on the website server? The "Support" aspect further asks which support e.g. a helpdesk or a hotline the vendor offers. Within the "Input" aspects it is established whether the user data comes from the server- or client-side ("Data source"), if the data was captured via log files, JavaScript etc. ("Technique") and which data formats are supported for import ("Supported standards"). Beside page views it is also possible to track event-based actions of a user. An event is any particular action of a user. This can be, for example, filling out a form or downloading a file. There is no particular definition of what an event is and there are nearly no limits what to count (Hassler 2010, p.264). The "Event" section within the criteria matrix includes 6 different events that the tools were tested for. In the category "Data Privacy" IP-masking, thus the anonymisation of the last octets within the IP-address was tested. The group "Monitoring & Reports" includes different aspects around the preparation and presentation of results. The "Click map" and "Heat map" thereby are two types of reports which help to visualize where the user clicked by showing it above the actual site. This can be done by showing tables with numbers on top of the site or by visualizing it with the help of different colours. It is also referred to

as site overly report or click density analysis. Kaushik (2007, p.157) refers to them as the "*more significant reports*" and Hassler (2010, p.84) justifies this by arguing that they help to put oneself in the position of the user more easily. Furthermore a "customizable dashboard" can help to quickly get an idea of the metrics which are important for the own site. Dashboards are set of views that can be seen as control panels (Hassler p. 80) and bring together the defined reports.

The criteria within the matrix are not weighted according to relevance or importance and do not include an assessment of usability. The matrix focuses on identifying the functionality and scope of tools and comparing between them. The next Section evaluates the selected tools with the help of the criteria matrix, but the final choice of a tool will depend on the requirement of the organization and their assessment of which criteria are important or not.

6.2.3 Evaluation

After the development of the different criteria outlined above, the 15 selected tools were analysed with the goal of understanding the scope and functionality of existing tools. This was done differently depending on the availability of the tools. Thus in a first step the tools were categorized according to their availability: software to install or online demo. In a second step the tools were either installed on a private system or the access to the online demo was established. Where both alternatives were possible both were used for testing. In addition vendor documentations were also used in the evaluation process. The extension of the criteria matrix with 2 separate columns for each tool, one for indication if the criterion was met/not met and one for additional comments/details if needed, represent step 3. The course of action in step 4 then was a successive testing of the criteria in each separate tool. Therefore each tool was tested for each criterion and the result was recorded in the criteria matrix. The whole process is shown in Figure 14 and Appendix A shows the filled in matrix.

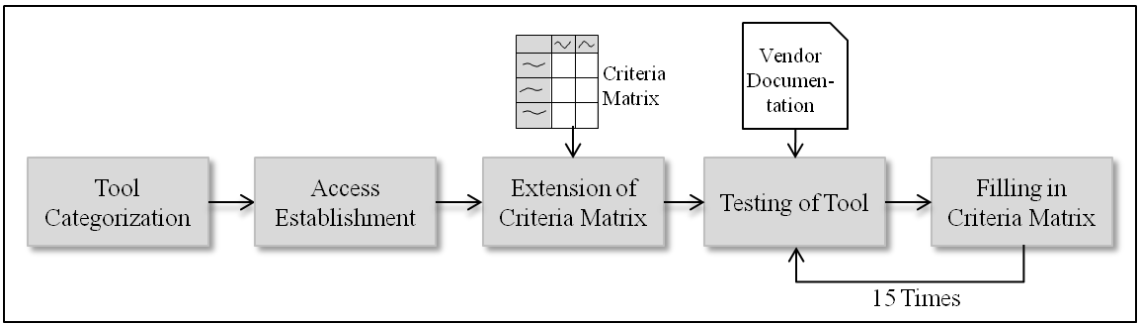


Figure 14: Research Steps within the Tools Evaluation Process

A recommendation for one specific tool on the basis of the criteria matrix cannot be given due to different reasons. First, it depends on the aims of the site owner and what he wants to get out of Web Analytics and which criteria are important or not. Second, the criteria were not weighted according to their relevance. Moreover, only because a criterion is met it doesn't mean that it is helpful. The criteria matrix does not include any statement regarding the usability of the tool. The matrix provides a profile of a tool to establish its scope and functionality and to see which criteria were met/not met. Besides the "General Information" and the "input" section 54 criteria are listed within the matrix. It was not possible to find each criterion within each tool. Table 6 gives an overview of the numbers for each tool. The column total measurable thereby indicated how many of the mentioned 54 criteria could be evaluated within each tool. It was not possibly to clearly state for each criterion if it is supported or not (column not supported and supported). Thus entries called 'unknown' are also listed within the matrix. An example why this could happen is that in the documentation a criterion was described as supported, but during the testing this statement could not be verified.

However, in order to make the numbers of supported and not supported criteria still comparable the last column indicated the ratio of how many percentage criteria are supported in one tool in relation to the total number which could be measured for this tool.

Tool	Not supported	Supported	Total measurable	% met (of measured)
Analyzer NX	0	50	50	100
Webtrekk Q3	5	44	49	89,80
Google Analytics	8	46	54	85,19
Etracker	8	45	53	84,91
Clicky	13	40	53	75,47
Open Web Analytics	13	37	50	74
Deep Log Analyzer	11	35	46	76,09
Piwik	15	37	52	71,15
STATCounter	16	32	48	66,67
W3 Statistics	19	30	49	61,22
Site Stats	21	32	53	60,38
Sawmill	25	24	49	48,98
AWStats	31	24	55	43,64
Webalizer	38	11	49	22,45
Logaholic	26	25	51	49.02

Table 6: Number of analysed functions

If one now tries to draw a ranking from this results the 5 systems from the sample which support the greatest functionality are (1) Analyzer NX, (2) Webtrekk Q3, (3) Google Analytics, (4) Etracker and (5) Deep Log Analyser. The result of Analyzer NX is not surprising. As already mentioned above, Analyzer NX is a tool which is often used of enterprise-class customers and no freeware. In return it is the most powerful tool of those which were tested. From the other four two can only be used for a fee (Webtrekk Q3 and Etracker) and two are freeware (Google Analytics and Deep Log Analyser (tested version)). A full evaluation of the 15 Web Analytic tools can be found in the final report of the Forschungspraktikum "Web Analytics" (winter semester 2011/2012) of the Business Software Research Group.

6.2.4 Limitations

As mentioned above, the conducted evaluation does not claim to be complete or valid. Feature analysis itself has a lot of limitations and is not always feasible (Nakatani & Chuang 2011). A few limitations were already outlined and are summarized in the following.

This whole chapter only considers tools which might be suitable for small and medium-sized enterprises. It further always had in mind requirements of the Passport to Trade project (cf. case study described in the next chapter). Thus the price of tools plays an important role. This in turn leads to the neglect of highly used and maybe very good tools like the Adobe SiteCalalyst, Webtrends Analytics or Yahoo! Web Analytics. Furthermore, because of human and time resources the number of analysed tools was set to 15.

In addition the developed criteria matrix should just be seen as a first study to understand the capabilities of the different tools, but cannot be used on its own in order to make the decision for one specific tool. It only contains a selection of possible metrics and reports. Most tools are able to analyse over 100 metrics and can produce dozens of reports. The selected ones for the criteria matrix are the most used and most important ones from the researchers' view. However, depending on the requirements of the website owner other metrics which are not listed here might be more important.

The analysis only assessed which criteria were met and which were not. No statements about how well the criterion were met, if additional functions exist and how usable the tools are were given. Table 6 lists the tools according to the percentage of how many criteria, from those which could be found out, were met. But it has to be kept in mind that the criteria were not weighted according to their possible relevance.

Despite these limitations, this analysis can be seen as a first step towards thinking about the question which tool is the most suitable for my organization. However as Nakatani and Chuang (2011) already outlined, a comparison only on the basis of features might not be sufficient. To clearly answer the question, requirements should be identified, the criteria matrix filled in and personal experience about usability of each tool should be made.

6.3 Conclusion

Finding the right Web Analytic tool for a given purpose is, because of the amount and diversity of tools available on the market, a relatively complex task (Nakatani & Chuang 2011). The value that a WA tool can bring to an organization is highly dependent on the organization's ability to understand the impact that Web Analytics can bring when taking actions on basis of results (Gassman 2009b). Gassman further points out that different kinds of websites like e-commerce and information websites have different needs and thus might need different tools according to their needs. The needs which are included in usage decisions should also reflect the needs and concerns of the whole organization. Results of Web Analytics as well as the tools itself can be used within different departments and different persons like marketing professionals, advertisers, content developers and web site operators (Gassman 2011, p.2). Their perspectives on what is needed might vary and be conflicting (Nakatani & Chuang 2011). However, they need to be brought together to a common denominator.

The above developed and outlined criteria matrix as well as the process of tool evaluation (Figure 14) should consider all these different needs and can assist in finding an appropriate WA tool. In his book Hassler (2010, p.39) describes a list of aspects that should be considered when preselecting a Web Analytics tool:

- Decide on the data collection technique (log file or page tag)
- Consider whether you want the solution to be hosted internally or externally (service provider)
- Be aware of privacy policies
- Are interfaces (e.g. API – Application Programming Interface) to other systems needed?
- Costs (including initial and running costs)

Some of these aspects are already included in the criteria matrix. However, these are not the actual features of the tools, but for example the way the tool is implemented within the organisation. Thus the criteria within the matrix which are listed under the category of general information are important as well and need to be considered, compared and evaluated alongside the feature evaluation.

Even though large amounts of money have been spent on Web Analytics tools and personnel in the past, only a few projects translated into positive returns. The problems encountered include for example, that the analytic tool is not delivering needed data and that thus the website cannot be tracked properly. Even though there are differences in tools these differences do not make or break the success. Burby & Atchison (2007, p.45) refer to a study by Forester Research from 2005 in which only 23% of all respondents stated, that they used the current (2005) Web Analytic tool for the last five years. Incredibly 48% had used three or more tools within these five years and 11% even tried five or more different tools. This continues today, as can be seen from the German study outlined above (Section 6.1), nearly half of all organizations used at least two tools or more.

Regardless of the tool it can be said that today nearly every website is trackable. Some need more and some less work to do it but not getting proper results is often just an excuse for poor work (Burby & Atchison 2007, p.44ff). Ogle (2010, p.2610) points this out by saying "*usability and optimization do not just happen. Organizations that want to constantly improve their web presence must dedicate staff time to use the various tools that they have purchased*". Also Lovett et al. (2009) supports this statement. They describe the pitfalls of Web Analytics as not within the tools, but within the investment, the staffing and the action chasm. Data can be seen as a commodity and the Web Analytic tool must be applied in the right manner to be productive (investment chasm). Furthermore, Web Analytic expertise is more valuable than the tool, thus good people are needed in order to conduct effective Web Analytics (staffing chasm). Last, actions are not taken because organizations have problems with understanding the results of Web Analytics.

"*The future success of Web analytics is dependent upon experts who translate raw data into insight*" (Lovett et al. 2009). Therefore Kaushik recommend that when planning for budget in Web Analytics only 10% should be invested into the tool. The other 90% is needed for people understanding the data (2007, p.81).

7 Case: Passport to Trade 2.0

To provide insights into the real-world challenges of developing a WA programme, an interpretive case study was conducted alongside the literature analysis and tool evaluation. The case study of the information website of the European project Passport to Trade 2.0 will be used. The project group can be seen as a non-profit group and therefore fits into the scope of this thesis. Their website must undergo a complete revision to improve, among other things, the information architecture and usability.

In order to understand the website as whole a qualitative analysis containing a visual design, content and structural analysis was conducted. These steps helped to identify the various information categories, content types and linkages between content and together built the overall information architecture. This information is also needed when conducting Web Analytics and provides the basis for understanding the outcomes of the quantitative log file analysis conducted here. The different steps which were undertaken are shown in Figure 15.

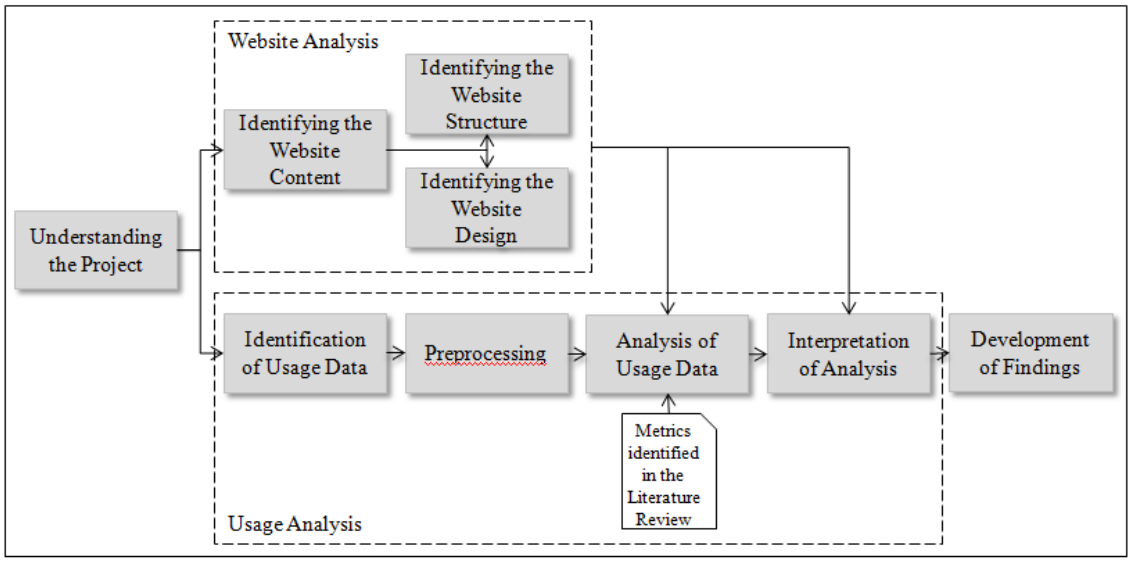


Figure 15: Research Steps of the P2T Project

This chapter was structured according to these research steps. Section 7.1 gives a brief overview of the history of the project and its current goals. It further outlines the involvement of the University of Koblenz-Landau within the project. The following sections then analyse the website developed by the previous project. They outline the content (Section 7.2), the structure (Section 7.3), the design (Section 7.4) and the user

interactions (Section 7.5). Furthermore, business needs which became obvious during working with the website and future capabilities are shortly discussed in Section 7.6.

7.1 The Project Background

The project Passport to Trade 2.0 is today in its third 'phase'. Between November 2003 and October 2005 the EMBER "Effective Marketing for Business in European Regions" initiative was carried out (http://www.ae.salford.ac.uk/extras/ember). This project was established in order to help SMEs who are dealing with organisations in other European countries. It was funded by the European Commission within their Leonardo da Vinci programme. Leonardo da Vinci is part of the Life Long Learning programme and funds projects in the field of vocational education and training especially when learning or working abroad (European Commission 2010). It pursues the following three main objectives: (1) to improve the skills and competencies of people in initial vocational training at all levels; (2) to improve the quality of, and access to, continuing vocational training and the lifelong acquisition of skills and competencies and (3) to promote and reinforce the contribution of vocational training to the process of innovation, with a view to improving competitiveness and entrepreneurship (European Commission 2006). Within the EMBER project an on-line information service was set up which outlined the differences concerning cross-national cultural differences to avoid misunderstandings when working in cross-national teams. At this time the project was restricted to the four countries United Kingdom, Czech Republic, Italy and Spain[4].

The EMBER project resulted in the Passport to Trade project which was carried out between November 2005 and October/November 2007 under the lead of the University of Salford (UK). According to its website it aimed to:

- "*Improve existing vocational training materials in business culture incorporating material from across the whole European Union;*
- *Widen and ease access to vocational training materials in business culture;*
- *Enable SMEs to benefit by improving sales potential through market expansion;*
- *Produce additional training materials in business culture;*
- *Implement e-learning facilities and support;*

[4] http://www.ae.salford.ac.uk/extras/ember

- *Improve the skill base of SMEs wishing to trade across the 25 EU member states.*"

Due to various reasons including technical problems the website is no longer accessible with its former content. However to assist the current project the old website was set up again by the Business Software Research Group of the University of Koblenz-Landau[5].

The new Passport to Trade 2.0 project commenced in October 2011. Still under the lead of the University of Salford (UK) but now with even more partners from Bulgaria, Czech Republic, Germany, Greece, Italy, and Romania the project is granted funding for 2 years. Based on various analyses of the old website, a new Website will be built within this project at the URL http://www.businessculture.org. Content of the old website will be retained and new information added. Within this process, among other things, following tasks will be performed:

- Increasing the number of covered European countries from 25 to 31
- Developing social media etiquette
- Including guidelines and training material for industrial placements
- Add the content in 6 additional languages
- Include search engine optimization and social media content

The new website will be developed to serve small and medium-sized enterprises as well as for students who are thinking about a placement within Europe. The aim for both is to develop easily accessible multimedia training material including the latest mobility needs.

The overall Passport to Trade 2.0 project is divided into 11 working packages. The title of the working package two, which is under the lead of the University of Koblenz-Landau, is "Website Analysis: Real Needs Identification". Among other things it includes the development of a methodology for website data analysis and evaluation and an analysis of user needs with the help of available website data (University of Salford 2011). Therefore the following sections as well as the whole thesis can be used as supporting material for this project working package.

[5] It is currently available at http://p2t-project.fgbas.iwvi.uni-koblenz.de/index.aspx?text=1

Note: Unless otherwise stated, all information described below has been tested while visiting the website with the settings of English as the language and the United Kingdom as the country to display information.

7.2 Content Analysis

As mentioned above the website deals with different topics to overcome cultural and social differences which arise when working with companies from different countries within Europe. In order to get an overview of the containing content the author analysed the website by clicking through the sites in a first step. The site consists of text, links, pictures, downloads and interactive elements. It soon became obvious that the headlines of the pages mostly represented the content which could be found on the page. Furthermore, one page called "Topics" gives an overview of all the country specific information which is available. In a second step it was further looked at content which goes beyond the country information. Thereby four different content groups could be identified: General Content, Interactive Content, Country Information and Project Information. Step 3 then was guided by analysing each category in order to discover the specific content of each category. Figure 16 gives an overview of the results.

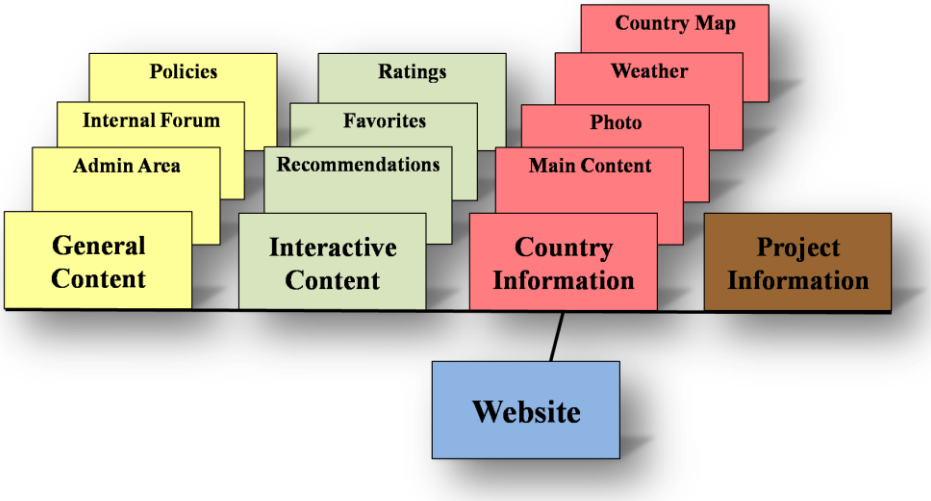


Figure 16: Website Content

The main language of the website is English. However it is possible to switch the language to French or German. Unfortunately not all content information is switching to the chosen language. Only the main text which is located in the middle of the page changes to the specific language. The rest of the website like the link menu, all project information or the interactive content remains English. In the following the different content areas are described in more detail.

7.2.1 Project Information

Always accessible from the top of the website are the four pages with general information about the project and how to use the website. The landing- and homepage "Home" welcomes users with a short introductory text and an additional drop-down menu to choose the language, country and topic of interest. A brief overview of the project background, its objectives, the research framework and addressors can be found on the "About" page. The project partners are briefly described at the "The Partners" page. Finally, the "Help" page can be seen as a kind of FAQ page. It answers questions about the use and the possibilities of the site.

7.2.2 Country Information

The middle part of the website is used to show information about the following 25 European countries: Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden and United Kingdom. This part represents the main and most important content of the site. The topics for each country are the same and are divided into 6 main categories. Most categories have further subcategories. An overview of all country content pages is given in Figure 17.

The content of these pages changes according to the selected country and provides specific information for this country. Only 7 pages could be identified which contain introductory text for the subcategories and are not country specific, these are:

- Political and Economic Environment
- General Business Environment
- Business Etiquette – Attitudes and Values
- Business Ethics
- Education and Training
- Business Practice
- How to Communicate

In addition to the text information a lot of external links are provided within the country information boxes which should further help to understand the topic. Beside the text and link information shown in the middle box of the website a map of the country, a picture of the selected country and the weather forecast are provided.

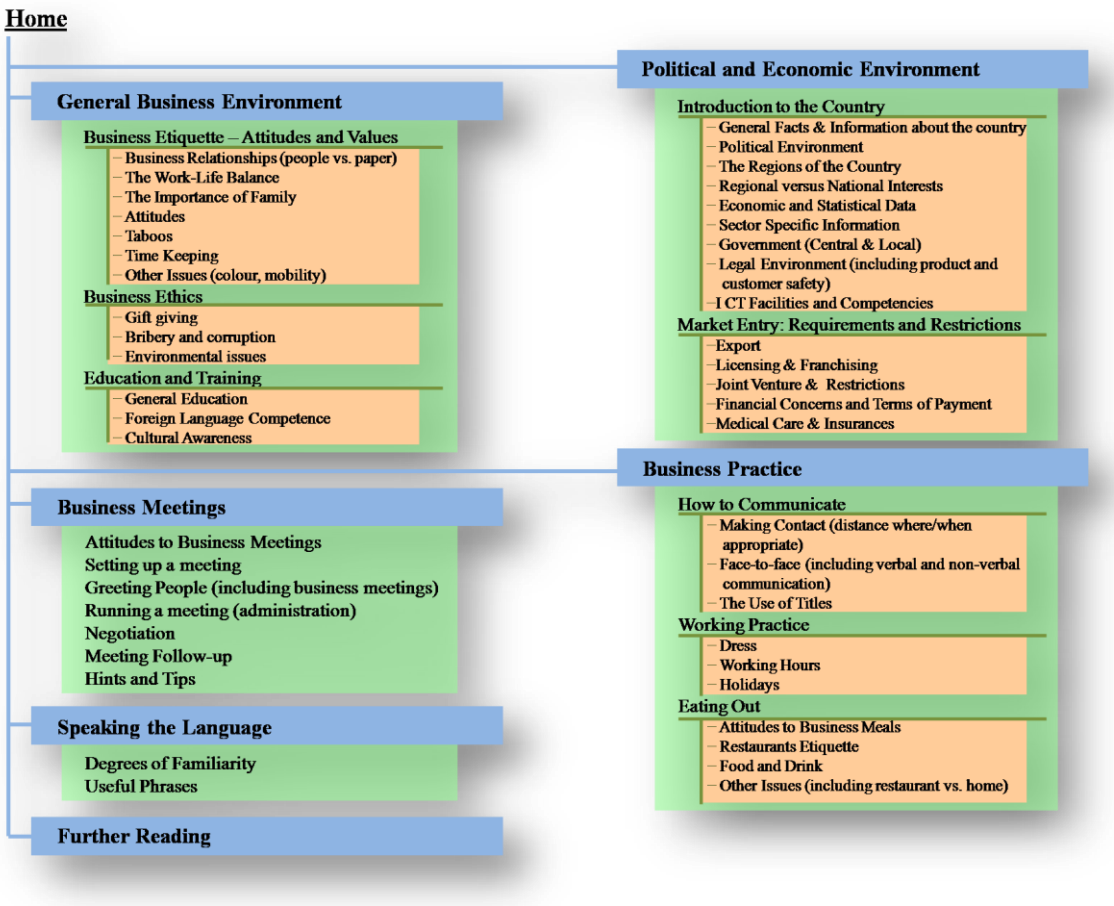


Figure 17: Sitemap of Country Information

7.2.3 Interactive Content

In addition to the main country content different interactive content can be found on the site where the user has the possibility to interact. Different boxes show someone's own favourites, give the possibility for ratings and provide recommending information.

Each country information page can be saved to a favourite list which then allows an easy access and a document export. Furthermore, these country content pages can be rated using a five star scale. All own rated pages are shown within a small box on the page. In addition two different areas on the website give recommendations. First, one box shows related articles to the one someone is looking at. Unfortunately, while testing the site it was noticed, that related articles here just mean article within the same subcategory. Is someone looking at a webpage from the lowest level the recommendation just shows the article which comes next within the listing of topics. Second, another box shows article other people have visited as they may also be relevant for the current user. In addition, the own page history is shown at the bottom of each page.

7.2.4 General Content

In addition to the above outlined content the website includes an admin area, an internal forum and copyright and privacy policies. The old forum is not accessible any more so that no statements about its content can be given.

An export and search function can also be found on the website. These are not really content information/content itself, but should be mentioned here, as they allow access to the content. The search functionality refers to different country specific information by using a keyword search. It can be searched for entries within a specific country and language as well as within a user's own favourites. Furthermore, at the homepage a tag cloud with the most popular searches is shown. The export functionality can be found at different places within the website. Each country-specific segment of information text within the selected language can be sent via e-mail saved as a favourite, a document (Word-document) or a pdf-file. The page "Topics" gives an overview of the topics of the country-information and the possibility for downloading a text-document or a pfd-file. Even though the downloadable information changes according to the country, here the information to download is only available in English.

7.3 Structural Analysis

Information Architectures are quite complex. While looking at the definitions of Wurman (1996) and Rosenfield and Morville (2006), Ding and Lin came up with their own broad definition of Information Architecture. They define it as being "*about organizing and simplifying information, designing, integrating and aggregating information spaces/systems; creating ways for people to find, understand, exchange and manage information; and, therefore, stay on top of information and make right decisions*" (Ding & Lin 2010, p.2). It is detached from graphical design and instead concerned with the structural design of information (Morville & Rosenfeld 2006). The purpose within this thesis in not to create a broad Information Architecture. Instead it will consider the general structure of the website as well as the internal and external linking structure. Therefore one of the topic categories was taken as an exemplary sample.

The site is structured in a way that mainly consists of boxes which are arranged around a text field in the middle. This text field is used for showing the main content. Three different kinds of links can be distinguished. First, the site includes internal links which open in the main content field and internal links which open a text above the boxes and the boxes then slip down. In addition, many external links can be found which open in a

new tab. Entering the Homepage, Figure 3 shows the link possibilities of the landing page.

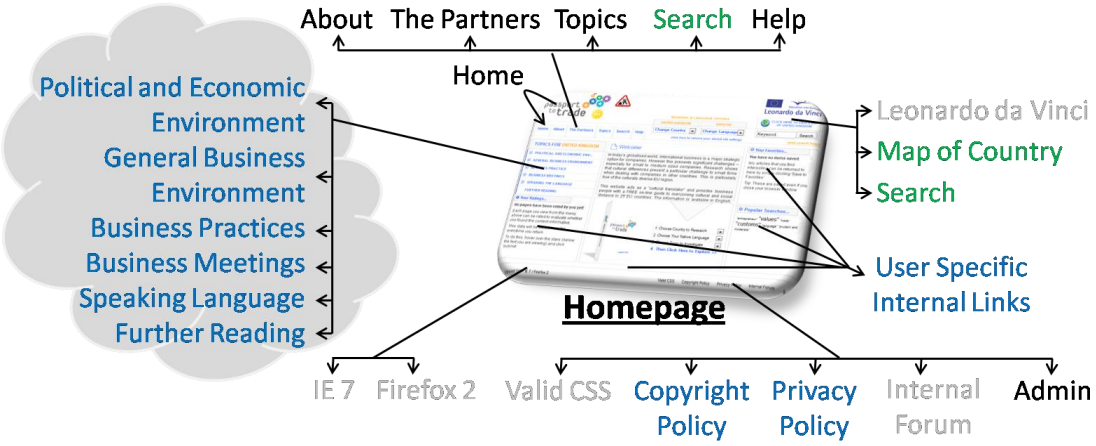


Figure 18: Link Structure at the Landing Page

Links which open above the boxes are marked black (only the admin link leads to a site without the overall design and structure). All links that change the text in the main content window are marked blue. Links which open a pop-up are marked green and external links gray. The cloud shows the main menu on the left hand side. The topics themselves have further links. Here, the headlines of the topics, thus the content, are the same as the link structure (see Figure 18). Not described by now are the external links included in the country information content. Due to the reconstructive setup of the analysed website (e.g. it is running on an asp.net) it was not possible to analyse the web graph/link structure with an online tool. Instead the authors examined the largest topic area “Political and Economic Environment” and its internal and external links. The result is shown in Appendix B.

Looking at the external links it becomes clear, that they are not named and structured very well. On the site each link has an explanatory text (listed in the Appendix B). Furthermore, the beginning of the linked URL can be seen directly. However, it sometimes happens, that on one page a link is shown several times but with different explanatory text. Or that the same text is used, but different external sites are linked. This leads to confusion as it is not exactly clear, what information can be found using a specific link.

The navigation possibilities on the website are thus using the menus (the top or the side) or using the direct links which are shown within the boxes around the content and emerge through user interaction and recommendations.

7.4 Website Design

When discussing website design not only the visual design can be described, but also the structural or technical design. Within their book Lawrence and Tavakol (2007) describe less about the technical aspects but summarise design aspects as described by different authors like Concepcion (2002), Cato (2001), Nierderst (2003) and Knuckles and Yuen (2004). For them good website design includes the consideration of different aspects like:

- Clear overall purpose
- Design and content should reflect target user needs and purpose
- Visual layout should be appropriate and create mood for the site
- Good Usability
- Appropriate navigation styles
- Relevant content

The purpose, content and structure of the Passport to Trade website including the navigation possibilities where already outlined above. Background information on the technical design is not available for the authors. The visual design is described in the following.

In general it can be said, that the overall site design is quite simple. There are not many colours used and beside the two logos in the top corners only one box is visible at country information pages includes a country specific picture. The site has a fixed predetermined width and is not scalable. Only the length changes with the pages.

Looking through pages of the site three different layouts of pages can be found. First, is the layout of the landing page which is discussed in Section 7.4.1. Second, is the layout of country information pages, which is close to the one of the landing page, but changes a little bit because of the displayed content. It is presented in Section 7.4.2. Finally, the project pages show their content on top of the landing page layout and are outlined in section 7.4.3.

7.4.1 Design of Landing Page

Figure 19 shows the landing page as it is visible when entering the site. The two logos in the top corners can be seen as the identification of the site. The left one (passport to trade beta) shows the logo of the project that developed the site. The logo in the right

corner (Leonardo da Vinci) shows the programme of the European Commission that funded the development. Beside these logos, 13 different areas can be identified and separated. Each are has a different purpose.

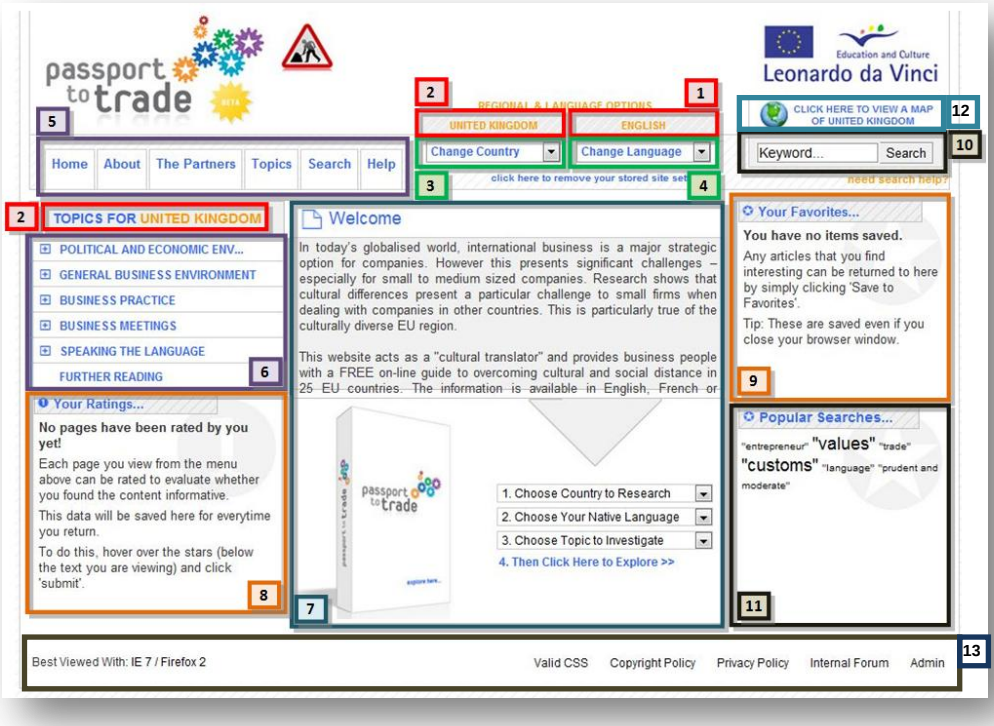


Figure 19: Design of Landing Page

As the website is available in different languages the current selected language for displaying the information is shown in (1). Furthermore the content is available for different countries. The selected country the information should be about is displayed in (2). It is possible to change both. The drop-down box (3) can be expanded to select the country and box (4) to change the language.

Apart from the language and country setting-menus there are two further main menus. One is located horizontally on the left site under the Passport to Trade Logo (5). The other one is presented vertically on the left side of the page (6). The horizontal navigation bar in the upper range (5) is the main menu about the project and the site itself. It allows access to project information, an overview of the content and the search function. Whereas the vertical menu on the left hand side (6) can be seen as the main navigation. It shows the different topics which are outlined on the website and expand if lower navigation paths are available. Through this navigation the access to the main content which is then shown in the box in the middle of the site (7) is available. As the

design of the page changes when selecting the country information a detailed description of the then visible design in outlined in the next section.

Beside the above mentioned areas two user-related content boxes can be found on the landing page. One on the bottom-left-side, which records the ratings of articles a user has made (8). The other box is on the right side of the website and shows the articles which the user saved as favourite article before (9). Both show an introducing text if no ratings or favourites are available.

The boxes numbered with (10) and (11) are search-related boxes. On the right side under the Leonardo logo is the general search-field (10). It is possible to search the website by entering a keyword. Moreover at the bottom-right is a box with popular search terms arranged in a search cloud (11).

Box number (12) links to a map of the selected country. Perhaps due to the reconstruction of the website this link is not working anymore and could not be tested. Thus no statements about the design of this page can be made. Last the footer (13) allows the access to internal and external links to main content independent information.

7.4.2 Design of Country Information Pages

As described above there are slight changes visible in the design of the page when displaying a country information page. These changes are due to the fact that the content is changing. Figure 20 shows an example of a country information page.

The main focus of this kind of web pages is the box in the centre containing the article – a specific section of information about a selected country (1). Apart from the title and the information itself, there is a returning link, which guides one back to the parent article in structure (1a) and the possibility to either send the article via email to someone, save the articles to the favourite box or to download it as doc- or pdf-file (1b). Then there is also the possibility to rate the article (1c) using a five-star-principle (one star = not so good article; five stars = great article).

On the left hand side the vertical menu is shown, which was also present on the main page (2). It opens and closes according to the article someone is viewing. In order to provide addition support in identifying the different levels a three colour system is used. The headlines of main topic pages are coloured in blue. Second level page titles are coloured green and third level titles in orange. Also already present on the main page are the boxes “Your Ratings” and “Your Favorites”. They are located at exactly the

same place, representing the ratings done and the articles saved as favourites by the user respectively.

Figure 20: Design of Country Information Page

On the right side under the favourite box more general information about the selected country is shown. One being a photo representing the country (5) and the other one being weather information (6).

Beneath the main article the user finds two boxes with link recommendations. The first one shows articles that are related to the article the user is reading at the moment (7), the other one recommends other articles, which were visited by other users before and thus could be also interesting for the current user (8).

At the bottom of the page the user now finds an additional horizontal bar recording the pages one visited before, called the "Page History" (9).

7.4.3 Design of Project Pages

Even though it is only a small change within the design it should shortly be mentioned within this section here. When clicking on one of the links of the horizontal navigation bar, the new content appears below the horizontal navigation, but above everything what is displayed on the landing page. The content which normally occupies this space simply slides down making space for the chosen content. Figure 21 shows an example of how it is displayed then. It is possible to close the new content-box by clicking on the link “click here to close”, which is in the top-right corner of the new box. An exception to this is the “Search”-link. Clicking on this link also opens a new box, but this opens in a separate popup over whatever someone is looking at.

Figure 21: Example of Inappropriate Fit of Text

7.5 Log File Analysis

Since the beginnings of the Passport 2 Trade project and the website www.businessculture.org, log files have been saved by the web server. The log files were not gathered in a systematic manner and in the four years that the site has been live

and no web usage analysis has been conducted. The original idea within the Passport 2 Trade 2.0 project was to analyse these log files using Web Analytic methods. Unfortunately it turned out that due to technical problems the data is no longer available. However, log files from the time between the 10th of October and 28th of November 2007 have been restored. Unfortunately, because of personnel changes within the responsible people for the IT for this project no further information about e.g. a formal record of why specific data types had been collected are at hand. Nevertheless, in the following an analysis was performed following the process outlined in Section 5.3 of this thesis. The procedure as well as the problems that arose and the results obtained are explained in the individual process steps below.

Data Collection

Through the steps described above the various information categories, content types, linkages between content and the overall information architecture have been identified. This knowledge is now used within the further steps of this analysis. A problem within the data collection step arose with the inadequate availability of the log files. As already mentioned, log data had been collected. However this data was less than optimal. A key challenge was to understand the nature and structure of the log files. The data the author obtained from the project team was a sample of 2 months of interaction in log files stored in a 22MB XML file which contained 33,250 lines. Looking at the second line of the XML code <?mso-application progid="Excel.Sheet"?> it could be recognised, that Spreadsheet XML, an XML schema for Microsoft Office Excel was used in order to open the data with Excel. It is not known, if the data was originally saved within this Excel file, or if it was transformed into it later. This led to considerable problems in the further processing. The tool analysis (Chapter 6.2) has identified several tools capable of analysing log files. However after reading the specification and trying some tools (Deep Log Analyse, AWStats, Analog, Webalizer) it had to be noted, that none of these tools were able to handle this data. Only tools like the so called Log Parser Lizard, which is a query software tool was able to import the data. Through the command-line interface it is possible to program your own queries with this tool. But this kind of tool use was out of the scope of this thesis and the results which should be achieved. The next idea of how to analyse the data was to find a possibility to convert the XML file so that it could be used with one of the standard Web Analytic log file tools. But, from comparing the fields and entries of the XML file with the different standard log formats it seems, that they used a custom format. Table 7 shows some example of saved logs from the XML Excel file.

User Agent	User Host Name	User Host Address	User URL Referrer	Log Date	Log Time	Page Title	Country	Lang-uage
Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.0; SLCC1; .NET CLR 2.0.50727; Media Center PC 5.0; .	82.27.235.161	82.27.235.161	-	10/10/2007	23:02:36.440232 6	1.0 Political and Economic Environment	Germany	English
Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; .NET CLR 1.0.3705; .NET CLR 1.1.4322; .NET CLR 2.	128.233.8.120	128.233.8.120	http://www.google.com/search?q=www.businessculture.org&rls=com.microsoft:en-us&ie=UTF-8&oe=UTF-8&sta	15/10/2007	19:40:09.972909	3.0 Business Practice	United Kingdom	English
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; Q312461; SV1; .NET CLR 1.0.3705; .NET CLR 1.1.432	146.87.255.18	146.87.255.18	http://www.manchestereveningnews.co.uk/news/business/s/1024941_how_to_do_business_abroad	20/11/2007	10:09:53.956223 3	1.1.1 General Facts & Information about the Country	United Kingdom	English

Table 7: Example Log Files from www.businesculture.com

The first column user agent shows entries like “Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.0; SLCC1; .NET CLR 2.0.50727; Media Center PC 5.0; .” It includes information about the software of the user PC. In the example, the browser Mozilla 4.0 was used running on a Windows Vista (Windows NT 6.0) operating system. As the website did not have a possibility to log on, the user host name and the user host address are both the same and represent the IP-address of the user. The user URL referrer shows the page, where user came from. It is only filled, if user clicked on a link to come to the businessculture.org website. If they for example typed in the URL directly no log entry could be made by the server. The log date and time show the exact moment, when the user accessed the page and the log was written. The page title shows which page was viewed. When saving logs within a standard form like the W3C format the here named page title would include the URL of the visited page and not only the page title. With the last two columns it is inferred, that they represent the country which was looked at (Country) and the language in which the page was viewed (Language). Thus no log information was saved which indicate from which country the user came from.

With all this information at hand and the problem of not being able to use a Web Analytic tool for the analysis the most fruitful approach was do a log file analysis by hand. It was clear that this would not bring the same effectiveness as with a WA tool, however it was undertaken to provide at least some insights into web usage. Therefore the next steps outline what was done.

Preprocessing

The main tasks within the preprocessing phase were to clean the data and identify which of the collected data was useful for the mining process. Further it was necessary to find users and to build sessions. Entries like image and error logs as well as visits from bots needed to be removed. Starting this it was noticed, that no image or error logs were in the table. This indicates that the XML file was already worked on before; however visits from bots were still within the file. From the 33,250 log file entries 18,847 bots could be identified by the captured user agent. Examples are the following:

- Google bot: Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)
- Yahoo bot: Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)
- Msn bot: msnbot-media/1.0 (+http://search.msn.com/msnbot.htm)

Filtering them out, left 14,403 remaining log file entries. In order to analyse these logs effectively single users and sessions needed to be identified. As no analytic tool could be used for this work and the researcher needed to do it by hand it was decided to only look at the first two weeks of the saved log files in order to keep the effort within a reasonable scope. Between the 10th of October and the 23rd of October 2007 708 log files were saved. As already outlined in Chapter 5 only looking at the IP-address to identify user is not the most appropriate way for user identification, but the method used here. Reasons for that include simplicity and time-effort relationship. Thus, 76 different users could be identified who have visited the website within these two weeks.

In a next step it was looked at different sessions. As within most Web Analytic tools one session was defined as a visit of one user where two consecutive log files were within the time frame of 30 minutes (30 minutes of inactivity). Using this definition 105 sessions could be identified. Table 8 gives a short overview about the numbers outlined in this sub-section:

Log files in total	33,250
Bots	18,847
Analyzed timeframe	10.October – 23.October 2007
Analyzed logs	708
Identified users	76
Identified sessions	105

Table 8: Log File Preprocessing Data

Pattern Discovery

The possibilities for pattern discovery, like statistical or path analysis were, due to the nature of the log files, quite limited. However, exemplarily a few statistics were discovered. 76 visitors could be identified. From this around 15% became returning visitors and came back to the website for a second time within the two weeks under analysis. One of the most important features of the website www.businessculture.org was that it was available in three different languages (English, French and German) and for 25 different countries. Nearly every user (74) used the website in English. It might be that when visiting the website for the first time English was always the default language as there were only two times that English wasn't used at all. The second most used language was German with 9 users and then French with 6. The total number (89) shows, that users switched within the language while visiting the site. The most used countries for getting information about where the following:

Country	Number of Users
United Kingdom	58
Italy	11
Ireland	9
Belgium	8
Czech Republic	5
France	5

Table 9: Country Usage

Latvia and Slovakia are the only countries no one looked at at all. More than 55% of all users only looked at information of one single country. The rest changed the country and viewed two or more different countries.

The most visited web pages as could be derived from the logs were the 6 main topics sites as showed in Table 10. A full overview of web page accesses can be found within Appendix C.

Web Page	Hits
1.0 Political and Economic Environment	145
4.0 Business Meetings	58
3.0 Business Practice	45
6.0 Further Reading	38
2.0 General Business Environment	37
5.0 Speaking the Language	32

Table 10: Web Page Usage

Within one session the average number of pages which were viewed was 3.9, ranging from the minimum of 1 page up the 29 pages within one session. While scanning the log file entries no paths could be identified. What can be stated is that it appears that people mostly didn't read a whole topic and all its sub sections, but switched within their reading between the main topics and a few sub sections. Within the log files the referrer URL, thus the URL of the website where a user came from to the own website was kept as well. 24 of the 105 sessions have a referrer URL kept. Within this, eight entries of users coming from the Google search engine like "http://www.google.ca/search?q=czech+republic++business+etiquette++what+to+wear &hl=en&start=10&sa=N" could be found. These log files not only show that people came to the website via the Google search engine, but they also show what search phrases user used. The following list shows search words which were used:

- Business culture
- Czech republic business etiquette what to wear
- Etiquette kissing on greeting people
- Etiquette in Portugal business personal space
- English meeting and greeting people

Beside the entrance through Google different websites which referred to the businesculture.org website could be identified. Included are the following:

- http://www.europublic.com/en/news640.html
- http://www.ae.salford.ac.uk/portfolio/viewall_cat.php?pageNum_Portfolio=1&totalRows_Portfolio=24&c=2
- http://www.iscapi.org/
- http://cor.europa.eu/en/activities/conferences_expo.htm
- http://www.europublic.com/en/links.html
- http://www.mondimpresa.it/Mondimpresa.aspx
- http://www.eic374.unioncamere.mondimpresa.org/Pagine.aspx?idMenu=14

The log files are from 2007 and thereby quite old thus an analysis of the used software might not be useful any more. Nearly all users (excepted one who used Opera) used Mozilla as their browser for viewing the page. This might for example be interesting to know for optimization reasons. Unfortunately due to the way the data had been collected it allowed little more than this very basic discovery.

Pattern Analysis

The interpretation and validation of the above outlined numbers will be discussed in this section. The reviewed log files of only two weeks cannot be seen as generalisable to the whole time the website was running. It only shows a snapshot and gives a first impression about the website usage. In order to get good insights of the usage behaviour a much bigger time frame needs to be observed. Furthermore, the whole analysis was only conducted by the author by hand. No tools could be used in order to get any results. Therefore only simple methods and statistics could be described. In-depth analysis could not be done due to missing information in the log files and capabilities of the author. For example no information about search words or the country of the visitor was available and no path analysis was conducted.

Nevertheless a few interpretations of the above outlined statistics can be made. Most users only visited the website once (about 85%) within these two weeks. This is an important discovery which should be observed for a longer period as it is not significant when only looking at two weeks. But it might indicate, one of two possibilities. The worse alternative would be, that people cannot find the information they are looking for and therefore do not come back. The better alternative is that they exactly found what they searched for and got all information. Different than it might be for an e-shop where people come back to buy another product this information site here tries to deliver

information which do not change that quickly and are for a quite specific purpose. If someone for example has a business partner in Spain he might look at the different topics for Spain, find and read through the information for Spain, probably download them for a later view and will not come back to the site. Even though a user only came to the site once the objective of the site might still be met.

As already described above it seems that the website was set to English and the United Kingdom by default. Thus the numbers of usage for this metrics need to be relativised. However, what can be seen from the numbers of how many users looked at the website within another language (9 German and 6 French) is that English still is the most used language for displaying the information. Furthermore, the most visited pages are the main topic entry pages. This can mean different things. On the one hand it might indicate, that these topics where well chosen and represent the information users where looking for. On the other hand it can indicate, that people didn't know what to expect from the headline and needed to have a look on the page to understand its purpose. To make a clear statement about this a path analysis is needed which shows if user went deeper into one topic after visiting the main topic page or if they left the topic. Further the time a user spends on the single page could indicate whether a user only shortly looked at the page or really read it. But to conduct such an analysis a Web Analytic tool is needed. Beside it was recognized that no logs of other pages than the country information pages were saved. Thus it is not possible to make any statements about the usage of those pages.

The search terms which are captured within the referrer URL show, how people found the website. These can give hints what people expect to find on a specific page and might give ideas for search engine optimization. Furthermore, only 22.9% (24 of 105) of sessions were started while coming from a referring website. As the return visitor number is not that high it's not because people captured the site via a bookmark but because they typed in the URL directly. This implies that the user either knew the website or they have read the URL somewhere else. The worst case is, that people do not know the site and cannot find it because too few external links are available or it cannot easily be found with the help of search engines.

7.6 Problems and Capabilities

The above conducted analyses have shown some aspects that need to be changed in order to conduct effective and actionable Web Analytics that can be used to help to understand user behaviour and improve the website. Furthermore, while conducting the

analysis different technical, functionality and design aspects became apparent and these are outlined below. In order to conduct Web Analytics in the future the method of data collection needs to be changed. The log files were not appropriate for Web Analytics. Further business requirements concerning the metrics which should be analysed need to be established and an appropriate WA tool chosen. These insights are used in the next chapter to develop a framework for the whole process of Web Analytics. Using this framework and an appropriate tool could then help to analyse the site over a longer timeframe and identify areas of improvements through the usage analysis.

A few recommendations for improvement of the website which emerged through the different analyses conducted above are described below:

- **Expand the content area**

  The content is the most important part of this website and only represents a small box surrounded by many other things. To put more attention on it and to make it easier to read this section should be made much larger.

- **Develop a complete translation of the site**

  Even though the content is available in different languages, the headlines of the different topics are only available in English. A consistent appearance where everything is in the chosen language might help users who cannot speak English.

- **Include the topic page in all languages**

  The topic pages give an overview of everything what is available on the site and gives the possibility for downloading it. The downloadable content changes with the country, but not with the language. At this page it is only available in English even though the same content is accessible in different languages through the menu.

- **Stay on the page when changing country or language**

  A user who changes the country or the languages always comes back to the main page of the first topic. As for example user might browse through the site for one specific topic within different countries they always needs to click through the whole menu again.

- **Provide more links and make them clearer**

  On the “Further Reading” page for example a lot of hints are given where to find more information. The suggested places where to find the information are listed,

but not linked. Also the project partners are only listed, but not linked. Besides, a lot of external links are not well described.

- **Anchor topics on the "Help" page**

  The "Help" page outlines answers to different topics but is a long page which must be scrolled down. An overview of the topics on the top with anchors to the answers might improve clarity and information design.

- **Include a dynamic scaling of the site**

  This would make the site more user-friendly when using big or small screens.

- **Provide different pictures and include a description of the picture**

  Each country only has one picture which is displayed on the right hand side. To make it more interesting this picture could change dynamically. Furthermore a description of what the picture shows might arouse more interest by the user.

- **Give space for topic suggestions or comments**

  A form where users can provide feedback or suggest new topics and ideas could help to further develop the site.

This case provides a good example of the shift that has taken place in recent years in both, websites and Web Analytics. The way the website was originally built and usage data was collected is no longer appropriate, making it difficult to analyse the site and its usage in any depth. However it shows the differences in technological possibilities with which websites are faced now. Besides the need to identify information needs for WA and to have a clear understanding of the purpose of the website and the Web Analytic requirements became obvious. For example, what are the desired outcomes of the Web Analytic process and what are the expected benefits of WA for the own organisation? Moreover the case provided the possibility to think about critical aspects involved in Web Analytics. Therefore conducting the different process steps by hand can also be seen positive as the author was forced to think deeply about each step and how to proceed in it. Even though only a few metrics were used the case study has provided the author with a sense of how carefully a metric needs to be discussed in order to interpret it properly. For example the significance of the number of returning visitors should be interpreted differently on an information website and an e-commerce platform.

8 A New Framework for Web Analytics

The preceding work and analysis conducted in the above chapters served to provide insights into the process and conducting of WA programmes. The following chapter will use these findings to assist in developing a framework to guide the overall WA process. Section 8.1 outlines the main gaps and issues that were identified. Further, the Gartner Maturity Model is described in Section 8.2. Based on this model and the discussion conducted in Section 8.3 the differences for information websites are explicitly shown. Section 8.4 then describes the broader Web Analytic process as presented by Araya et al. (2004) and outlines the general changes which need to be included in a new framework. Finally Section 8.5 presents the proposed new framework for developing a programme for Web Analytics.

8.1 Gap/Issue Identification

Within the different parts of this thesis different aspects of WA have been analysed and discussed. A number of recommendations about what should be done and what could be done have been identified. Furthermore, through the use of the case study to gather field experiences a number of further issues for organisations were identified. These recommendations, gaps, and issues are discussed in the following.

Until now no standards exist in Web Analytics that would unify the way tools analyse web data or report their metrics. This leads to a broad and diverse spectrum of what is possible. Organisations are not only faced with the decision of what data types are the appropriate ones to be collected but also with the question of which data collection method they should use. Thereby they need to understand the limitations of each method in order to get to know the boundaries. Furthermore they need to decide for a WA tool which is commensurate with the organisations requirements. At this point two problems arise at once. On the one hand the differences between tools often are not clearly obvious which makes it hard to decide for one tool against another. On the other hand business needs are not taken into account appropriately. A key issue for organisations is to clearly identify the purpose of the website and the corresponding WA requirements, and to understand what they need to measure and why. Only when these questions can be clearly answered a step towards a specific tool can be made.

Beside these tool and organisational issues the process itself as described in the literature includes some gaps. For instance, the actions that should be taken on the basis of the analysis are frequently mentioned, but when looking at the process as outlined in Figure 7 action taking is not part of the WA process itself. From the author's perspective this step is one of the most important since it is action taking that leads to improvements of the website. In addition the process is missing the connection to the Web Analytic tools. These tools need to be identified, selected, installed and configured. When these settings have been made correctly steps like the data cleaning are done by the tool itself. Furthermore, the process is often shown as a one time investigation. However it should rather be seen as a cycle: a continuous process with repeated stages of the same actions. WA is not a short project, but an ongoing programme.

The area around privacy concerns, whilst widely discussed has not been taken into account as part of a systematic WA methodology. Alongside addressing questions such as who is the owner of the captured data it is also important to ensure that the WA programme is legally compliant and consistent with privacy and data protection legislation.

Finally it was noted that most attention has been placed on WA for e-commerce websites but little attention has been given to other website types like information websites. In summary the issues range from macro-level issues such as understanding the business requirements for WA, selection of appropriate WA tools and compliance with legal requirements to more micro-level issues such as selection of data types, data analysis and data interpretation methods. It might not be possible to address all these issues at once with the establishment of a single process framework for Web Analytics, but the above outlined gaps and issues justify the motivation for developing a WA framework that can assist companies to deal with some of the identified challenges.

8.2 The Gartner Maturity Model

In order to answer the question of what organisations want to achieve with WA investigations Gassman outlines the Gartner Maturity Model for Web Analytics. He describes it as a "*model for assessing the state of an organization's Web analytics environment*" (Gassman 2008, p.1). It divides the degree of maturity into 5 different levels (Figure 22). Within each level it is outlined, what can be achieved and what needs to be done to reach the specific level of maturity in Web Analytics.

Leve1 1: Aware builds the first step into Web Analytics. Here, basic information about the site is known. The goal is to achieve general information about the site usage.

Standard metrics report different statistics, but these reports are not further analysed. They do not drive changes. Web Analytic tools are rather used to explore than to monitor. If attempts are being made to understand what people do when visiting the site and how they react to changes in content, structure and navigation the level of maturity moves to level 2: Tactical. Here insights into optimisation possibilities should be gathered and at least one web analyst is assigned to the tasks. In level 3: Strategic Web Analytics is used in order to optimise the web channel. It is part of the overall business operations and key performance indicators are defined and monitored and marketing campaigns planned. If the website is part of the sales cycle, level 4: Collaboration becomes critical. Here cross-channel activities are analysed and a 360° view of the customer is sought. Within the highest level, level 5: Pervasive, the insights gained with Web Analytics are used as parts of the corporate wide performance management. Metrics are used in budgeting, planning and forecasting.

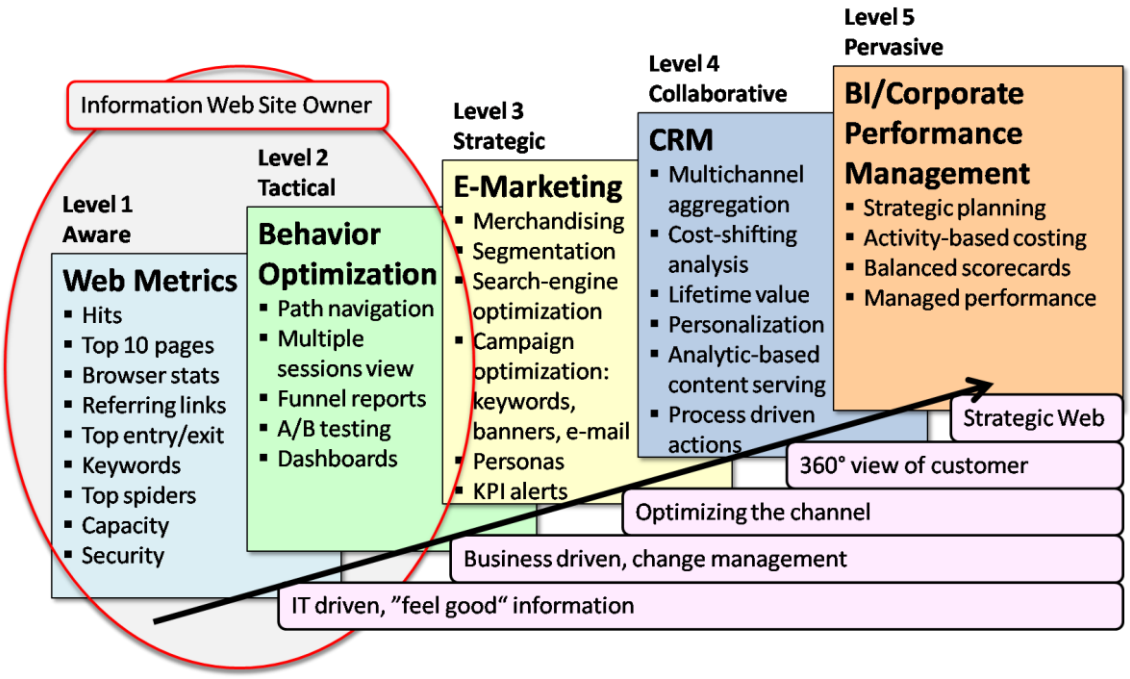


Figure 22: Maturity Model (adapted from Gassman 2008, p.3)

Not only because the costs and the complexity increase with the level of maturity (Gassman 2008, p.6), but also because of the description of the higher levels it falls into place, that not each Web Analytic programme should aim to reach level 5. Even though a many companies operate across multiple maturity levels, most state that they see themselves in level 1: Aware. Gassman argues that a move to level 2 should be pursued, if more than 5% of revenue is made through the web channel. This as well as the tasks that should be performed at different higher levels points out, that the model aims to

support e-commerce websites. This is supported by Gassman's statement that because of well defined paybacks which are easy to measure, retail websites exhibit the level 5.

The focus of this thesis work is on assisting organisations who are new to Web Analytics of beginning to develop a new programme. Thus, looking at the Gartner model from an information website owner new to WA it seems reasonable to try to reach level 2 with parts of level 3. Reaching level 2 help them to use WA to understand their users and to redesign the web site in order to support them better. However, search-engine optimisation to improve the site further is also essential. So moving a new organisation to at least level 3 should be targeted for information websites as well.

8.3 Differences for Information Websites

As within the Gartner maturity model, the different steps conducted in this thesis have identified distinct activities which need to receive attention when conducting Web Analytics for information websites. These differ slightly in nature from large e-commerce and transactional websites.

The major differences in particular can be found in the areas 'successful visit', 'use of metrics' and 'action taking'. The goal of an information website is to inform users about a special topic whereas the goal of an e-commerce website is to inform users about products and to sell products or services. With the action of an online purchase a visit of a user can be defined as a successful visit (even though no statements about the perceived quality are made). One may equate this with download possibilities of information on an information website. However not all information is downloaded and even if it is it does not mean that the information was what the user was looking for. Thus to determine a successful visit of an information website could be seen as more difficult compared to an e-commerce site. Also because of the differences in goals and thus the actions which can be taken on the different sites the metrics which are used to track and monitor a site are different with the different kind of websites. Of course not all are different. Metrics like visitors or entry and exit page stay the same, but when analysing e-commerce sites key performance indicators such as top-line revenue, conversion rate, gross online margin or increased revenue can be tracked as well and build important numbers for the whole business. In tracking information websites such KPI which look at the financial aspects cannot be used or are less useful. Expressed simply it can be said that the metrics which are used for information websites are the more basic ones such as where did the user click or which path did he/she take.

Also the ‘action taking’ or implementing the findings of Web Analytics may be different with information websites. For example an important factor to consider is the information refresh rate, how long does the information on the site stay unchanged, does it change frequently or not? The case study website for example contained information of the behaviour and culture of people in the different European countries. This does not shift that quickly. There may be more information which will be placed on the site on a specific topic but the overall content will stay similar for long periods of time. Information and functions on e-commerce websites change relatively rapidly. New product offerings or new marketing campaigns at least change the site visually. When starting with WA both might see huge possibilities for change which can be implemented, but an e-commerce site might have further possibilities for change with e.g. new discount offerings or other campaigns. Further, the resources that SMEs and non-profit organisations spend for information websites is likely to be significantly less than would be invested in large e-commerce sites.

8.4 The Broader Web Analytic Process

The aim of this thesis is to identify the specifics of information websites in Web Analytics. These were outlined above. However, the main goal is to establish a framework to guide SMEs through the whole process of Web Analytics. A basis for a process framework can be seen in the process as outlined in Chapter 5. Araya et al. (2004) took this process as described in the literature and their experience from projects and developed a methodology for general application of Web Analytics. They describe it as an iterative process which includes the following steps:

- identification of objectives and available data
- selection
- preprocessing
- transformation
- pattern discovery
- interpretation and evaluation of results
- business integration

Alongside the split of the preprocessing phase into preprocessing and transformation they add the first step: “identification of objectives and available data” and last one:

"business integration". With the identification phase they emphasise the importance of clearly defined objectives and the need to not only think about gathering web data, but to also look at business (data in traditional systems) and meta (content and structure) data. Then the process is performed as described in the steps outlined above. At the end, they include the (from the authors' perspective most important) step and one which is often neglected in the models found in the literature: the business integration phase. Findings from the previous steps are only useful, if they lead to actions in the respective business.

A structural process for Web Analytics is a crucial requirement for its success (Araya et al. 2004, p.139). However the processes and methodologies by now do not connect the theoretical view with the practical implementation. Each step described in the literature is important and should not be ignored, but there is more around it which also need to be looked at. Furthermore, when using Web Analytic tools some steps will be performed by the tool automatically or as a consequence of previous customisation. Therefore this thesis uses the three conducted input sources literature review, tool analysis and case study, which are all connected in order to find the gaps and to develop a framework which guides SMEs with information websites when conducting Web Analytic (Figure 23).

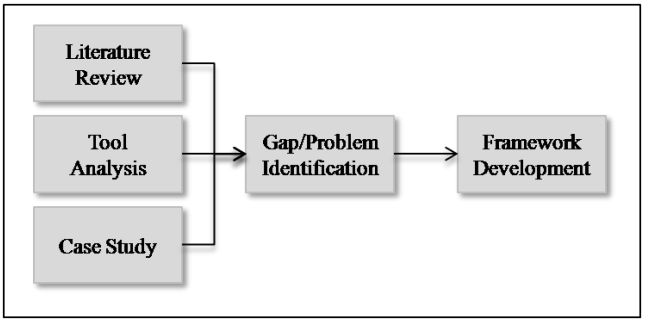


Figure 23: Framework Input Sources

8.5 The New Web Analytic Framework

Building on the findings from the literature, the tool analysis and the case study, the outcome of this research is a proposed WA framework. It is divided into two pictures. First Figure 24 shows the 5 steps within the high level WA process cycle. Following Figure 25 outlines each of these steps in more detail. The colours of the different steps within the framework are according to the corresponding steps within the process cycle.

Figure 24: High Level WA Process Cycle

The Web Analytic process can be divided into 5 phases. It starts with phase 1, *Business Requirements*. Here the needs for a clear understanding of the business requirements for Web Analytics are addressed by identifying the information needs and the information architecture of the website. Phase 2, *Planning for Web Analytics* build on the business requirements by conducting a detailed analysis of WA possibilities, identification of the organisation's own needs and the selection of a suitable WA tool. Phase 3, *Developing a Data Collection Capability* involves the implementation of the selected tools and the data processing. Phase 4, *Achieving Usable and Actionable Results* includes the steps of pattern analysis and identification and the action taking. Finally, phase 5, *Evaluation of Actions* questions if the WA process and actions for improvement result in positive outcomes.

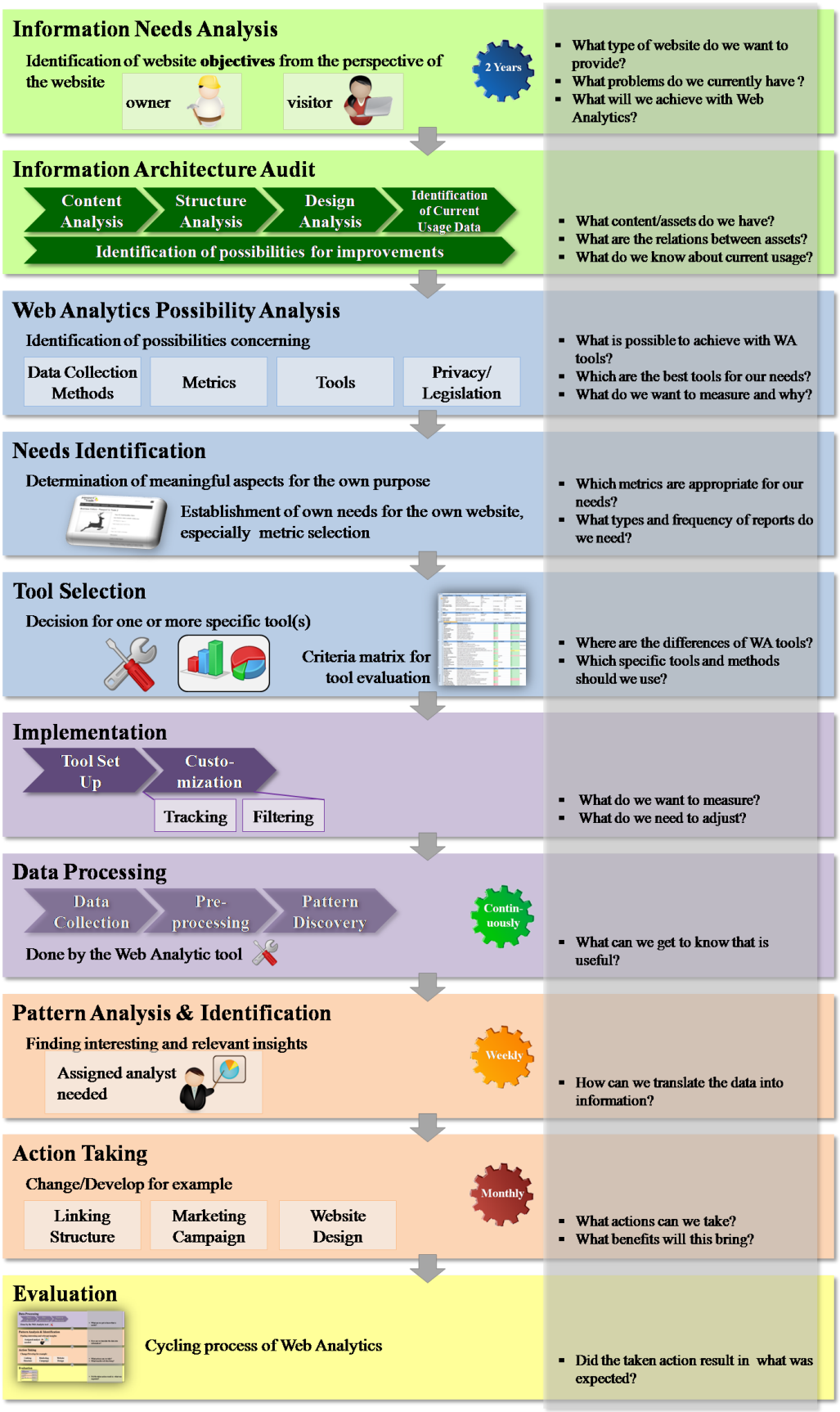


Figure 25: The New WA Process Framework

The WA process framework (Figure 25) builds upon the high level process cycle as shown in Figure 24. The 5 different steps of the process cycle form the broaden steps for undertaking effective Web Analytic programmes Each step of the cycle can thereby be assigned specific tasks and questions which guide the tasks and are further outlined in the process framework.

Phase 1 Business Requirements

Phase 1 comprises two stages: Information Needs Analysis and Information Architecture Audit. Within the Information Needs Analysis the objectives of the website owner, as well as the potential visitors should be identified. Therefore it is essential to answer fundamental questions about the website in general. In the following phase Information Architecture Audit the website then is analysed from different viewpoints: content, structure, design and usage data. This leads to a deeper understanding of the site itself and may help to already now detect possibilities for improvements.

Phase 2 Planning for Web Analytics

Phase 2 consists of three broader subtasks. Within the Web Analytics Possibility Analysis existing possibilities concerning the data collection method, metrics, tools and privacy and legal issues should be identified. The Needs Identification step builds upon the identified possibilities and determines the meaningful aspects for the organisation's own website before a specific tool which can satisfy these aspects is selected within the task of Tool Selection. The previously outlined criteria matrix (Section 6.2.2) could be used here to assist organisations to map their needs onto appropriate tools.

Phase 3 Developing a Data Collection Capability

Phase 3 contains two activities. First the Implementation of the Web Analytic tools and second the Data Processing. Implementation involves the setting up of the selected tool and its customization so that it captures and processes the relevant data types and the required metrics. Following is the activity of Data Processing. It deals with the data collection, preprocessing and pattern discovery itself. This step is the one that is most well developed in the current literature and was already outlined within the process as described in Figure 7. However, if the implementation is done correctly these sub steps of Data Processing are conducted automatically by the WA tool.

Phase 4 Achieving Usable and Actionable Results

Phase 4 includes the steps of Pattern Analysis & Identification and Action Taking. Based on insights that could be gathered through the usage analysis a clear pathway for action taking should be developed. In this phase a clearly assigned analyst is needed who interprets the findings, formulises improvements and develops metrics to further measure, e.g. increased downloads, fewer abrupt exits from the site, etc.

Phase 5 Evaluation of Actions

Finally, phase 5 illustrates the need for evaluation of results and the need for the cycling process. Through conducting the phases 3 to 5 on a regular basis it can be judged whether the WA programme is useful or not. The gear cogs within the framework are used to indicate cycles of action. It is necessary to establish the different action cycles and make decisions about how frequently analysis takes place. The data processing for example needs to be done continuously whereas the pattern analysis and identification should be conducted on a weekly basis. Here it needs to be monitored and checked that the right data is being collected. If new items are required then the tool must be further adjusted to capture these new data types. Monthly action taking involves a more detailed usage analysis and leads to, for example, fine tuning of the website, adjusting marketing strategies or fixing structural errors. Beyond this, the two-yearly action taking involves a major review of the website and its usage and may lead to, for example, major strategic re-positioning of the website (e.g. to move from an informational site to a transactional site), or a major re-design of the site to take advantage of new technologies or social/business trends (e.g. the addition of social media tools, taking advantage of the benefits of HTML5 etc.). The exact action cycle frequencies will be determined according to the organisation's specific requirements.

9 Summary and Conclusion

In this final chapter the researcher summarises the thesis and brings together the main conclusions drawn from the research findings. Therefore answers to the research questions are addressed and implications for future work are provided before a short conclusion to the thesis is given.

9.1 Research Questions

At the beginning of this thesis the objective of the work was defined by the introductory and overall research question

> **RQ: What is required in a framework to support small and medium-sized enterprises (SME) to provide an easy entry point for undertaking Web Analytic programmes?**

and a set of 8 subsidiary research questions which have been addressed in the various stages of this thesis. The efforts of the previous chapters including the literature analysis and synthesis, the review of WA tools and the conducted case study have provided evidence for answering these questions.

Through investigations of the relevant literature 3 subsidiary questions were addressed. First was **RQ1a: How can Web Mining be classified?** The researcher found one widely accepted and often used classification into the three areas Web Content Mining, Web Structure Mining and Web Usage Mining. Even if it is clearly delineated which subarea works with which data and tries to answer which questions they all are needed together for effective WA programmes. However, Web Usage Mining is the area which works with web usage data and makes the largest contribution to Web Analytics.

Thus the second sub question **RQ1b: What different methods for conducting Web Mining are described in the academic literature?** was focused specifically on methods for conducting Web Usage Mining. By investigating the process of Web Usage Mining it became obvious that most scientific authors use similar ways to describe the process as well as the methods used. In the step of pattern analysis they all list methods such as statistical analysis or path analysis. Only the algorithm for conducting these methods may differ. However, what became clear within the later step of tool analysis is that the different methods are not that important for the end user as the tools call for

understanding what metrics should be analysed, not for methods that should be conducted.

RQ1c: Which data capturing method is the best to use today? is the last question that was answered through the literature review. As was outlined in Chapter 5 a shift in the data capturing method could be observed in recent years. Whilst capturing and analysing web log files was the one and only method used at the beginnings of WA the most used method today is JavaScript tagging. Both have limitations which need to be kept in mind, but with JavaScript tagging it is possible to overcome some of the problems caused through log files and to capture even more data.

The second block of sub questions should be answered through the review of WA tools. In order to get an overview and to answer **RQ2a: What does the Web Analytic market look like?** a market analysis was conducted in Section 6.1. On average organisations spend $15,000 per year for Web Analytics and nearly 50% use more than one tool at the same time. Through this review it became clear that the four big vendors Adobe, Google, IBM and Webtrends dominate the market. The **RQ2b: What are the best Web Analytic tools to be used?** cannot be definitively answered. Within the market analysis the most used tools have been identified. However this does not indicate the quality of a tool. Furthermore is became clear, that it depends on the requirements of the organisation whether a tool is good and appropriate or not. Looking from the perspective of an SME operating an information website Section 6.2.3 gives a first hint of which tools might be interesting to analyse more deeply. In addition the developed criteria matrix (Section 6.2.2) provides the answer for **RQ2c: Which criteria can be used to describe and differentiate these tools?** Though, as already outlined above it need to be noted that a definitive statement about which tool is the best cannot be given through analysis with this matrix alone.

At various places within this thesis the differences when operating and analysing an information website compared to an e-commerce website have been described. Factors like measurement metrics and the website scope in particular make clear the differences. Thus **RQ3a: Is there a need for distinguishing the kind of website for what Web Analytics is used in order get clear results?** can be answered with ‘yes’. The same applies to **RQ3b: Are there differences between Web Mining for small or big businesses that need to be taken into account?** Especially because of a lack in financial resources as well as highly trained people (specialists) SMEs often do not operate WA programmes in the same deepness than big companies do.

Each of these questions supported the author to answer the overarching research question **What is required in a framework to support small and medium- sized enterprises (SME) to provide an easy entry point for undertaking Web Analytic projects?** The answer which is the framework itself was presented in Section 8.5.

9.2 Research Contribution and Future Work

By answering the above outlined research questions the author raised the consciousness of the characteristics that need to be taken into account in WA programmes for information websites of SMEs and to develop a framework that addresses these needs. Thus this research project has contributed to the development of theory in the area of Web Analytics by identifying special needs and building the awareness for problem areas. On the other hand it also contributed to WA practice by developing a process framework for WA which SMEs and non profit organizations can use for planning their WA programmes.

However, this thesis also opened up a number of areas for future research. Even though the presented framework was developed through intensive literature review and accompanied by a case study it is not further tested or validated. A future research project is needed in order to test its viability. Furthermore, it only shows the high level processes. Future research investigations might use the framework as a basis in order to dive deeper into the different areas outlined in the framework and describe them in more details. Guidelines could be developed that further help SMEs to conduct effective Web Analytics. In addition other research investigations could analyse in how far the framework can be transferred to other areas like big organisations or commercial websites.

In addition, it appears that the Web Analytic market will be subject to change in the future (Gassman 2011). Not only are new WA vendors entering the market who are specialising in different areas, but also the technological possibilities like Web 2.0, social media and smartphone or tablet tracking might become more important and open up new needs that require modified tools and processes.

As can be seen from the examples above the research in Web Analytics still is far from being at the end; may it be because of investigations in existing methods and techniques or because of technological development that require adoption. This thesis was scheduled for 6 months so that further or more detailed aspects could not be analysed. However it provides a solid basis and an opportunity for further research.

9.3 Concluding Remarks

In this thesis the current state of Web Analytic is examined through the use of a literature analysis, a tool analysis and a case study. Each approach identified a range of challenges and issues that need to be addressed when conducting Web Analytics. A special focus was placed on the needs for informational websites of SMEs and non-profit organisations in order to understand the challenges they are facing. A complex landscape of tools supporting WA was discovered. Whilst the academic literature described the analytical process from a data-centric view, the tasks that need to be undertaken to make effective use of the tools need to be more user- and business-focused.

Web Analytics is not just a project that can be undertaken. It is part of the continuous improvement process of an organisation (Gassman 2008) and should thus be seen as part of the whole. However the market is still without open standards for tools and metrics and each vendor has a unique approach (Herschel 2010).This is noting this thesis can or aimed to change, but what could be done was to establish a framework for developing a sustainable WA programme that widens the focus of the Web Analytic process to fully account for business requirements and to focus attention on achieving actionable outcomes.

References

Abraham, M., Meierhoefer, C. & Lipsman, A., 2007. *Cookie Deletion Whitepaper*, comScore.

Agrawal, R. & Srikant, R., 1994. Fast Algorithms for Mining Association Rules in Large Databases. In *VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases*. Conference on Very Large Data Bases. Santiago, Chile: Morgan Kaufmann, pp. 487–499.

Agrawal, R. & Srikant, R., 1995. Mining Sequential Patterns. In International Conference of Data Engineering. Taipei, Taiwan, pp. 3–14.

Aliaga, M. & Gunderson, B., 2002. *Interactive Statistics* 2nd ed., New Yersey: Prentice Hall.

Araya, S., Silva, M. & Weber, R., 2004. A methodology for web usage mining and its application to target group identification. *Fuzzy Sets and Systems*, 148, pp.39–152.

Benbasat, I., Goldstein, D.K. & Mead, M., 1987. The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, 11(3), pp.369–386.

Berners-Lee, T., 1989. Information Management: A Proposal. *The original proposal of the WWW*. Available at: http://www.w3.org/History/1989/proposal.html [Accessed October 23, 2011].

Berners-Lee, T. & Fischetti, M., 2000. *Weaving The Web: The Past, present and Future of the World Wide Web*, London: Texere.

Berry, M.J.A. & Linoff, G.S., 2011. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* 3rd ed., John Wiley & Sons.

Berthon, P., Pitt, L.F. & Watson, R.T., 1996. The World Wide Web as an Advertising Medium. *Journal of Advertising Research*, 36(01), pp.43–54.

Blaikie, N.W.H., 2000. *Designing Social Research*, Wiley-Blackwell.

Bruce, C.S., 1994. Research students' early experiences of the dissertation literature review. *Studies in Higher Education*, 19(2), pp.217–229.

Bundesministerium der Justiz, 2007. Telemediengesetzt - Einzelnorm. Available at: http://www.gesetze-im-internet.de/tmg/__15.html [Accessed April 1, 2012].

Burby, J. & Atchison, S., 2007. *Actionable Web Analytics: Using Data to Make smart Business Decisions*, Indianapolis, Indiana: Wiley Publishing.

Burks, A.W., 1946. Peirce's Theory of Abduction. *Philosophy of Science*, 13(4), pp.301–306.

Cavaye, A.L.M., 1996. Case study research: a multi-faceted research approach for IS. *Information Systems Journal*, 6(3), pp.227–242.

CERN, 2008. CERN - Where the web was born. *CERN - Where the web was born*. Available at: http://public.web.cern.ch/public/en/About/Web-en.html [Accessed October 23, 2011].

Checkland, P. & Holwell, S., 1998. Action Research: Its Nature and Validity. *Systemic Practice and Action Research*, 11(1), pp.9–21.

Chein, I., Cook, S.W. & Harding, J., 1948. The field of action research. *American Psychologist*, 3(2), pp.43–50.

Chen, W. & Hirschheim, R., 2004. A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal*, 14(3), pp.197–235.

Chi, H., 2002. Improving Web Usability Through Visualization. *IEEE Internet Computing*, 6(3), pp.64–71.

Chishom, R.F. & Elden, M., 1993. Features of Emerging Action Research. *Human*, 46(2), pp.275–298.

Connolly, D., 2000. A Little History of the World Wide Web. *W3C: A Little History of the World Wide Web*. Available at: http://www.w3.org/History.html [Accessed April 11, 2011].

Cooley, R., 2003. The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns. *ACM Transaction on Internet Technology*, 3(2), pp.93–116.

Cooley, R., Mobasher, B. & Srivastava, J., 1997. Web Mining: Information an Pattern Discovery on the World Wide Web. In *Proceedings of ICTAI' 97*. Newport Beach, California.

Crabtree, B.F. & Miller, W.L., 1999. *Doing Qualitative Research* 2nd ed., Thousand Oaks, California: Sage Publications.

Creswell, J.W., 2003. *Research Design: Qualitative, Quantitative, and Mixed Method Approaches* 2nd ed., Thousand Oaks, California: Sage Publications.

Crotty, M., 1998. *The Foundations of Social Research: Meaning and Perspective in the Research Process*, Sage Publications.

Crowder, P. & Crowder, D.A., 2008. *Creating Web Sites Bible* 3rd ed., Indianapolis, Indiana: Wiley Publishing.

Davidson, R.M., Martinsons, M.G. & Kock, N., 2004. Principles of canonical action research. *Information Systems Journal*, 14(1), pp.65–86.

Ding, W. & Lin, X., 2010. *Information Architecture: The Design and Integration of Information Spaces*, Morgan & Claypool.

Etminani, K., Akbarzadeh-T., M.-R. & Yanehsari, N.R., 2009. Web Usage Mining: users' navigational patterns extraction from web logs using Ant-based Clustering Method. *IFSA-EUSFLAT*, pp.396–401.

Etzioni, O., 1996. The World-Wide Web: quagmire or gold mine? *Communications of the ACM*, 39(11), pp.65–68.

European Commission, 2010. Leonardo da Vinci programme. *European Commission - Education & Training - lifelong learning programme - Leonardo da Vinci programme*. Available at: http://ec.europa.eu/education/lifelong-learning-programme/doc82_en.htm [Accessed December 26, 2011].

European Commission, 2006. Leonardo Da Vinci programme. Available at: http://ec.europa.eu/education/programmes/leonardo/objectives_en.html [Accessed December 26, 2011].

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

Fallman, D., 2003. Design-oriented human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. CHI '03. New York, NY, USA: ACM, pp. 225–232. Available at: http://doi.acm.org/10.1145/642611.642652.

Fink, A., 2009. *Conducting Research Literature Reviews: From the Internet to Paper* 3rd ed., Thousand Oaks, California: Sage Publications.

Friedman, K., 2003. Theory construction in design research: criteria: approaches, and methods. *Common Ground*, 24(6), pp.507–522.

Gassman, B., 2009a. *Five Best Practices for Web Analytics Initiatives*, g.

Gassman, B., 2008. *Introduction to the Gartner Maturity Model for Web Analytics*, Gartner Group.

Gassman, B., 2009b. *Key Challenges in Web Analytics, 2009*, Gartner Group.

Gassman, B., 2011. *Web Analytics Market Update, 2012*, Gartner Group.

Geertz, C., 1973. *The Interpretation of Cultures*, New: Bas.

Glaser, B.G. & Strauss, A.L., 1967. *The Discovery of Grounded Theory. Strategies for Qualitative Research*, Transaction Publishers.

Grace, L.K.J., Maheswari, V. & Nagamalai, D., 2011. Analysis of Web Logs and Web User in Web Mining. *International Journal of Network Security & Its Application (IJNSA)*, 3(1), pp.99–109.

Greening, D.R., 2002. Data Mining on the Web: There's gold in that Mountain of Data. *Data Mining on the Web*. Available at: http://drdobbs.com/architecture-and-design/184413868?pgno=3 [Accessed October 26, 2011].

Gregor, S., 2002. A Theory of Theories in Information Systems. *Information Systems Foundations: Building the Theoretical Base, Australian National University, Canberra*, pp.1–20.

Hannes, K. & Lockwood, C., 2011. *Synthesising Qualitative Research: Choosing the Right Approach* 2nd ed., jo.

Hasan, L., Morris, A. & Probets, S., 2009. Using Google Analytics to Evaluate the Usability of E-Commerce Sites. *Human Centered Design, HCII 2009, LNCS 5619, Springer*, pp.697–706.

Hassler, M., 2010. *Web Analytics: Metriken auswerten, Besucherverhalten verstehen, Webseiten optimieren* 2nd ed., mitp.

Hay, C., 2002. *Political Analysis. A Critical Introduction*, Basingstoke: Palgrave.

Heath, H. & Cowley, S., 2004. Developing a grounded theory approach: a comparison of Glaser and Strauss. *International Journal of Nursing Studies*, 41(2), pp.141 – 150.

Heaton, J., 2002. *Programming spiders, bots and aggregators in Java*, Sybex Book.

Herschel, G., 2010. *Hype Cycle for Analytic Applications, 2010*, Gartner Group.

Hevner, A.R. et al., 2004. Design Science in Information Systems Research. *MIS Quarterly*, 28(1), pp.75–105.

Hofacker, C.F., 2005. Using server log files and online experiments to enhance internet marketing. In S. Krishnamurthy, ed. *Contemporary research in e-marketing*. Pennsylvania: Idea Group, Hershey, pp. 226–249.

Holsti, O.R., 1969. *Content Analysis for the Social Sciences and Humanities*, Reading, MA: Addison Wesley.

Kaplan, B. & Duchon, D., 1988. Combining qualitative and quantitative methods information systems research: a case study". *Manage. Inf. Syst. Q.*, 12(4), pp.571–586.

Kaplan, B. & Maxwell, J.A., 2005. Qualitative Research Methods for Evaluation Computer Information Systems. In J. G. Anderson & C. E. Aydin, eds. *Evaluating the Organisational Impact of Healthcare Information Systems*. New York: Springer, pp. 30–55.

Kaufman, L. & Rousseeuw, P.J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*, Hoboken, New Jersey: John Wiley & Sons.

Kaushik, A., 2007. *Web Analytics: An Hour a Day*, Indianapolis, Indiana: Wiley Publishing.

Kohavi, R., Mason, Ll. & Zheng, Z., 2004. Lessons and Challenges from Mining Retail E-Commerce Data. *Machine Learning*, 57, pp.83–113.

Kosala, R. & Blockeel, H., 2000. Web Mining Research: A Survey. *SIGKDD Explorations*, 2(1), pp.1–15.

Kumar Jain, R., Kasana, R.S. & Jain, S., 2009. Efficient Web Log Mining using Doubly Linked Tree. *International Journal of Computer Science and Information Security (IJCSIS)*, 3(1).

Lawrence, D. & Tavakol, S., 2007. *Balanced Website Design: Optimising Aesthetics, Usability and Purpose*, Springer.

Lee, T.W., 1999. *Using Qualitative Methods in Organizational Research*, Thousand Oaks, California: Sage Publications.

Lewin, 1946. Action Research and Minority Problems. *Journal of Social*, 2, pp.34–46.

Lincoln, Y.S. & Guba, E.G., 1985. *Naturalistic Inquiry*, Beverly Hills: Sage Publications.

Lovett, J. et al., 2009. *US Web Analytics Forecast, 2008 to 2014*, Forster Research.

Madria, S.K. et al., 1999. Research Issues in Web Data Mining. In M. Mohania & A. Tjoa, eds. *Data Warehousing and Knowledge Discovery*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 303–312. Available at: http://dx.doi.org/10.1007/3-540-48298-9_32.

March, S.T. & Smith, G.F., 1995. Design and natural science research on information technology. *Decis. Support Syst.*, 15(4), pp.251–266.

Marshall, C. & Rossman, G.B., 2011. *Designing Qualitative Research* 5th ed., Thousand Oaks, California: Sage Publications.

Masand, B. & Spiliopoulou, M., 2000. *Web Usage Analysis and User Profiling* B. Masand & M. Spiliopoulou, eds., Springer Berlin / Heidelberg. Available at: http://dx.doi.org/10.1007/3-540-44934-5_1.

Merriam-Webster, 2001. Design - Definition and More. *Design - Definition and More*. Available at: http://www.merriam-webster.com/dictionary/design [Accessed September 12, 2011].

Miniwatts Marketing Group, 2011. Internet World Stats. *World Internet Usage Statistics News and World Population Stats*. Available at: http://www.internetworldstats.com/stats.htm [Accessed October 20, 2011].

Morville, P. & Rosenfeld, L., 2006. *Information Architecture for the World Wide Web* 3rd ed., Sebastopol, CA: O'Reilly.

Muijs, D., 2004. *Doing quantitative research in education with SPSS*, London: Sage Publications.

Myers, 2011. Qualitative Research in Information Systems. *Qualitative Research in Information Systems*. Available at: http://www.qual.auckland.ac.nz/ [Accessed May 12, 2011].

Myers, M.D., 1997. Qualitative Research in Information Systems. *MIS Quarterly*, 21(2), pp.241–242.

Nakatani, K. & Chuang, T.-T., 2011. A web analytics tool selection method: an analytical hierarchy process approach. *Internet Research*, 21(2), pp.171–186.

Negash, S., 2004. Business Intelligence. *Communications of the Association for Information Systems*, 13, pp.177–195.

Neuendorf, K.A., 2002. *The Content Analysis Guidebook*, Thousand Oaks, California: Sage Publications.

Neuman, W.L., 2003. *Social Research Methods: Qualitative and Quantitative Approaches* 5th ed., Boston: Allyn and Bacon.

Neuman, W.L., 2006. *Social Research Methods: Qualitative and Quantitative Approaches* 6th ed., Boston: Allyn and Bacon.

Nielsen, J. & Norman, D.A., 2000. Web-Site Usability: Usability on The Web Isn't A Luxury. *http://www.informationweek.com/773/web.htm*. Available at: http://www.informationweek.com/773/web.htm [Accessed October 20, 2011].

Norguet, J.-P., Zimányi, E. & Steinberger, R., 2006. Improving Web Sites with Web Usage Mining, Web Content Mining, and Semantic Analysis. *SOFSEM 2006, LNCS 3831, Springer*, pp.430–439.

Novak, T.P. & Hoffman, D.L., 1997. New metrics for new media: toward the development of Web measurement standards. *World Wide Web J.*, 2(1), pp.213–246.

LDI NRW, 2009. Datenschutzkonforme Ausgestaltung von Analyseverfahren zur Reichweitenmessung bei Internet-Angeboten. Available at: https://www.ldi.nrw.de/mainmenu_Service/submenu_Entschliessungsarchiv/Inhalt/Beschluesse_Duesseldorfer_Kreis/Inhalt/2009/Datenschutzkonforme_Ausgestaltung_von_Analyseverfahren_zur_Reichweitenmessung_bei_Internet-Angeboten/Analyse.pdf [Accessed March 1, 2012].

Nunamaker, J.F., Chen, M. & Purdin, T.D.M., 1990. Systems development in information systems research. *J. Manage. Inf. Syst.*, 7(3), pp.89–106.

Ogle, J.A., 2010. Improving Web Site Performance Using Commercially Available Analytical Tools. In Symposium: AJS Carl . Brighton Workshop on Health Informatics. Springer, p. 2604–261.

Orlikowski, W.J. & Baroundi, J.., 1991. Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research*, 2(1), pp.1–28.

Orlikowski, W.J. & Iacono, C.S., 2001. Research Commentary: Desperately Seeking the "IT" in IT Research - A Call to Theorizing the IT Artifact. *Information*, 12(22), pp.121–134.

Paliouras, G. et al., 2000. Clustering the Users of Large Web Sites into Communities. In *Proceedings of the Seventeenth International Conference on Machine Learning*. ICML '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 719–726. Available at: http://dl.acm.org/citation.cfm?id=645529.657810.

Pani, S.K. et al., 2011. Web Usage Mining: A Survey on Pattern Extraction from Web Logs. *International Journal of Instrumentation, Control & Automation (IJICA)*, 1(1), pp.15–23.

Peacock, D., 2002. Statistics, Structures & Satisfied Customers: Using Web Log Data to Improve Site Performance. *Museums and the Web 2002*, pp.2–14.

Pierrakos, D. et al., 2003. Web Usage Mining as a Tool for Personalization: A Survey. *User Modeling and User-Adapted Interaction*, 13, pp.311–372.

Pitkow, J., 1997. In search of reliable usage data on the WWW. In *Selected papers from the sixth international conference on World Wide Web*. Essex, UK: Elsevier Science Publishers Ltd., pp. 1343–1355. Available at: http://dl.acm.org/citation.cfm?id=283554.283441.

Piwik, 2011. JavaScript tracking - Piwik. *JavaScript tracking - Piwik*. Available at: http://piwik.org/docs/javascript-tracking/ [Accessed December 28, 2011].

Pope, C. & Mays, N., 1995. Reaching The Parts Other Methods Cannot Reach: An Introduction To Qualitative Methods In Health And Health Services Research. *BMJ: British Medical Journal*, 311(6996), pp.42–45.

Power, N. & Moynihan, T., 2003. A theory of requirements documentation situated in practice. In *Proceedings of the 21st annual international conference on Documentation*. SIGDOC '03. New York, NY, USA: ACM, pp. 86–92. Available at: http://doi.acm.org/10.1145/944868.944887.

Raporort, R.N., 1970. Three Dilemmas in Action Research: With Special Reference to the Tavistock Experience. *Human Relations*, 23(6), pp.499–513.

Remenyi, D., 2002. Research Strategies – Beyond the Differences (Dan Remenyi (2002)). *Research Strategies – Beyond the Differences*, 1(1), pp.38–41.

Ritchey, T., 2011. Wicked Problems: Structuring Social Messes with Morphological Analysis. In Swedish Morphological Society.

Rittel, H.W.J. & Webber, M.M., 1973. Dilemmas in a General Theory of Planning. *Policy Sciences*, 4, pp.155–169.

Rossi, M. & Sein, M.K., 2003. Design research workshop: a proactive research approach. *Presentation delivered at IRIS*, 26, pp.9–12.

Schubert, P., Kummer, M. & Leimstoll, U., 2004. Legal Issues of Personalized E-Commerce Applications. In *Proceedings of the European Conference on Information Systems (ECIS)*.

Sen, A., Dacin, P.A. & Pattichis, C., 2006. Current Trends in Web Data Analysis. *Communications*, 49(11), pp.85–91.

Souza, R. et al., 2000. *The Best Of Retail Site Design*, Cambridge, Massachusetts: Forster Research.

Spiliopoulou, M., 2000. Web Usage Mining for Web Site Evaluation. *Communications of the ACM*, 43(8), pp.127–134.

Spiliopoulou, M. & Pohle, C., 2001. Data Mining for Measuring and Improving the Success of Web Sites. *Data Mining and Knowledge Discovery*, 5(1-2), pp.85–114.

Spiliopoulou, M., Pohle, C. & Faulstich, L., 2000. Improving the Effectiveness of a Web Site with Web Usage Mining. In B. Masand & M. Spiliopoulou, eds. *Web Usage Analysis and User Profiling*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 142–162. Available at: http://dx.doi.org/10.1007/3-540-44934-5_9.

Srikant, R. & Agrawal, R., 1996. Mining Sequential Patterns: Generalizations and Performance Improvements. In *proceedings of the Fifth Int'l Conference on Extending Database Technology*. Conference on Extending Database Technology. Avignon, France.

Srivastava, Jaideep et al., 2000. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations*, 1(2), pp.12–23.

Stanhope, J., Frankland, D. & Dickson, M., 2011. *The Forrester Wave: Web Analytics, Q4 2011*, Forster Research.

Straub, D.W., Gefen, D. & Boudreua, M.-C., 2005. Quantitative Research. In D. Avison & J. Pries-Heje, eds. *Research in Information System: A Handbook for Research Supervisors and Their Students*. Amsterdam: Elsevier, pp. 221–238.

Sullivan, T., 1997. Reading Reader Reaction: A Proposal for Inferential Analysis of Web Server Log Files. Available at: http://www.pantos.org/ts/papers/rrr.html [Accessed April 10, 2011].

Suneetha, K.R. & Krishnamoorthi, R., 2009. Identifying User Behavior by Analyzing Web Sever Access Log File. *International Journal of Computer Science and Network Security (IJCSNS)*, 9(4), pp.327–332.

Susman, G.I. & Evered, R.D., 1978. An Assessment of the Scientific Merits of Action Research. *Administrative Science Quarterly*, 23(4), pp.pp. 582–603.

Takeda, H. et al., 1990. Modeling Design Processes. *Artificial Intelligence Magazine*, 11(4), pp.37–48.

Telemedicus, 2007. LG Berling: Tracking von IP-Adressen auf Webseiten, Urteil vom 06.06.2007 Az. 23 S 3/07. Available at: http://www.telemedicus.info/urteile/Allgemeines-Persoenlichkeitsrecht/Tracking-von-IP-Adressen/134-LG-Berlin-Az-23-S-307-Tracking-von-IP-Adressen-auf-Webseiten.html [Accessed April 1, 2012].

Trochim, W.M.K., 2006. Deduction & Induction. *Deduction & Induction*. Available at: http://www.socialresearchmethods.net/kb/dedind.php [Accessed October 12, 2011].

University of Salford, 2011. *Application Form: Leonardo da Vinci - Transfer of Innovation - Passport to Trade 2.0*,

Vaishnavi, V. & Kuechler, B., 2001. Design Science Research in Information Systems. *Design Science Research in Information Systems*. Available at: http://desrist.org/design-research-in-information-systems/ [Accessed September 11, 2011].

de Vaus, D., 2002. *Surveys in Social Research* 5th ed., London: Routledge.

W3perl, 2011. Logfile format. *Logfile format*. Available at: http://www.w3perl.com/docs/logformat.html [Accessed December 29, 2011].

Walsham, G., 1993. *Interpreting Information Systems in Organizations*, New York: John.

Web Analytics Association, 2008. Web Analytics Definition.

Web Analytics Association,2011. Privacy. *Privacy*. Available at: http://www.webanalyticsassociation.org/?page=privacy [Accessed September 11, 2011].

Weischedel, B. & Huizingh, E.K.R.E., 2006. Website Optimization with Web Metrics: A Case Study. In International Conference for Electronic Commerce. Fredericton, Canada: ACM, pp. 463–470.

Whitman, M.E. & Woszezynski, A.B., 2004. *The Handbook of Information System Research*, Hershey, PA: Idea Group.

Wu, J. et al., 2009. Using Web-Analytics to Optimize Education Website. *ICHL 2009, LNCS 5685, Springer*, pp.163–174.

Yin, R.K., 2009. *Case Study Research: Design and Methods* 4th ed., Sag.

Zhang, Q. & Segall, R.S., 2008. Web Mining: A Survey of Current Research, Techniques, and Software. *International Journal of Information Technology & Decision Making*, 7(4), pp.683–720.

Zumstein, D., Züger, D. & Meier, A., 2011. *Web Analytics in Unternehmen*, FELD M GmbH.

Appendix A – Criteria Matrix

The following pages show the criteria matrix and the results of the evaluation of the Web Analytic tools as described in Chapter 6. Therefore the next two pages give an overview of the used criteria and their description before the results of each tool are shown.

Note: At the time of finalising this thesis the completion of the evaluation was not yet completed. The matrix shown below is a work in process matrix.

	A	B
1	**General information**	**Description**
2	**Tool Name**	Name oft the tool
3	**Manufacturer**	Producer of the tool
4	**Version**	Tested version of the tool
5	**Keeping of data**	Where/ how data is stored
6	**Licence model**	Open Source / recurring / non-recurring etc.
7	**Price**	Price per licence/ per month etc.
8	**Admin / User concept**	User and policy management
9	**Multi-language Support**	Ability to view and print reports in multiple languages
10	**Support**	Helpdesk/ Hotline providede by the manufacturer
11	**Email Reports**	Deliver reports to an e-mail account on a regular basis
12		
13	**Input**	**Description**
14	**Data source**	Origin of the analyzed data (Server/ Client)
15	**Technique**	Log analysis/ Script/ Pixel etc.
16	**Supported standards**	Data formats supported for import and export
17	**Import standards**	Data formats supported for import
18	**Export standards**	Data formats supported for export

	A	B
1	**Visitor Details**	**Description**
2	IP Address	Track visitors IP address
3	Country	Which country the visitor is in
4	City	Which city the visitor is in
5	Operating System	The visitors operating system ()
6	Internet Service Provider	Identifies the visitors provider
7	Internet Speed	The visitors bandwidth
8	Browser	Which web browser the visitor uses
9	Browser Language	The visitors browser language settings
10	Screen Resolution	The visitors screen resolution
11	Cookies Enabled	Whether cookies are enabled or not
12	JavaScript Enabled	Wether JavaScript is enabled or not
13	Username	The visitors username if logged in
14		
15	**Traffic**	**Description**
16	Entry Pages	The first page of a visit
17	Page Views per User	The number of times a page was viewed
18	Unique Visitors	Each individual is counted only once in the unique visitor measure for the reporting period
19	New/ Returning visitor	Classification of each visit in a specified date range by its type
20	Events per Page	Tracks the number of events each visitor executes per page visited
21	Events per Session	Tracks the number of events each visitor executes over the duration of their visit
22	Time Spent on Page	The length of time spent on a single page
23	Time Spent per Session	The overall duration of a visit
24	Exit Pages	The last page on a site accessed during a visit
25	Exit Links	Links that take the visitor away from the site
26	Avg. Visits per Day/Month/Year	How many visitors per time period
27	Total Visits	Number of visitors since a defined date
28	Most Popular Pages	Which pages generate the most views
29	Internal Referrer	Measures from which page in your site visitors came from
30	Internal Search Keywords	Tracks which internal serch term referred to the visited page
31	External Referrer	Measures from which website visitors came the site
32	External Search Keywords	Tracks which external serch term referred to the visited page
33	Visit History	Tracks a visitor's history over an extended period of time
34		
35	**Monitoring & Reports**	**Description**
36	Click Path	Tracks the path through the site
37	Bot/Spider Filtering	Prevent Bots/ Spider from being analyzed
38	Bounce Rates	Percentage of visits in which the person left your site from the entrance
39	Live Monitoring	Real-time information about website traffic
40	Visualization	Ability to represent analysis and reports in graphics
41	*Diagrams*	Bar chart, graph, pie chart etc.
42	*click map*	Visualization of clicked links in percent
43	*mouse tracking video*	A video that shows the mouse movements an clicks of single users
44	*heat map*	graphical representation of click distribution and mouse position
45	*Visitor origins map*	A world map that shows the visitors country of origin
46	Multi-Domain Tracking	Collect visits to multiple domains in a single profile
47	Social Media Monitoring	Tracks what is being said about the site on social networking sites
48	Mobile Browser Monitoring	Tracks mobile visitors
49	Mobile Support	Monitor website performance using mobile devices
50	Customizable Dashboard	Rearrange the output in the main interface to personal needs
51	Goal definition	Definition of goals to measure conversion/ success rate
52		
53	**Events**	**Description**
54	Forms	Tracks forms filled out by visitors
55	Downloads	The number of times a file is downloaded
56	Flash Events	Track flash execution and clicks etc.
57	Multimedia Streams	The number of times a multimedia stream is clicked on
58	Error Pages	Tracks error pages shown to a visitor
59	RSS Feed Subscriptions	The number of visitors that sign up to receive RSS feeds from a website
60		
61	**Data privacy**	**Description**
62	IP-Masking	
63	Cookie deactivation	
64	Certificate	

Piwik

	A	B
1	**General information**	**Tool**
2	**Tool Name**	**Piwik**
3	**Manufacturer**	Mathieu Aubry et al.
4	**Version**	1.6
5	**Keeping of data**	inhouse
6	**Licence model**	open source
7	**Price**	free
8	**Admin / User concept**	yes
9	**Multi-language Support**	yes
10	**Support**	yes
11	**Email Reports**	yes
12		
13	**Input**	**Piwik**
14	**Data source**	client-based
15	**Technique**	JavaScript
16	**Supported standards**	
17	**Import standards**	Log-Files
18	**Export standards**	API, CSV

	A	B	C
1	**Visitor Details**	**Piwik**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	yes	
7	**Internet Speed**	no	
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	yes	
12	**JavaScript Enabled**	no	
13	**Username**	no	
14			
15	**Traffic**	**Piwik**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	yes	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	yes	
20	**Events per Page**	no	
21	**Events per Session**	yes	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	no	
24	**Exit Pages**	yes	
25	**Exit Links**	yes	exit path
26	**Avg. Visits per Day/Month/Year**	yes	day
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	unknown	
30	**Internal Search Keywords**	yes	via plugin
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	unknown	
34			
35	**Monitoring & Reports**	**Piwik**	**Tool details**
36	**Click Path**	yes	
37	**Bot/Spider Filtering**	unknown	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	yes	
40	**Visualization**	yes	
41	***Diagrams***	yes	graph
42	***click map***	no	
43	***mouse tracking video***	no	
44	***heat map***	no	
45	***Visitor origins map***	yes	
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	no	
49	**Mobile Support**	no	
50	**Customizable Dashboard**	yes	
51	**Goal definition**	yes	
53	**Events**	**Piwik**	**Tool details**
54	**Forms**	no	
55	**Downloads**	yes	
56	**Flash Events**	yes	
57	**Multimedia Streams**	no	
58	**Error Pages**	no	
59	**RSS Feed Subscriptions**	yes	
60			
61	**Data privacy**	**Piwik**	**Tool details**
62	**IP-Masking**	yes	full, via plugin
63	**Cookie deactivation**	yes	
64	**Certificate**	no	

Google Analytics

	A	D	E
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**Google Analytics**	
3	**Manufacturer**	Google Inc.	
4	**Version**	2012	
5	**Keeping of data**	ASP	Google Servers
6	**Licence model**	-	
7	**Price**	free	
8	**Admin / User concept**	yes	
9	**Multi-language Support**	yes	31 languages
10	**Support**	no	Online-Help only
11	**Email Reports**	yes	
12			
13	**Input**	**Google Analytics**	**Tool details**
14	**Data source**	client-based	
15	**Technique**	JavaScript, Pixel	
16	**Supported standards**	Google API, csv, tsv, pdf	Export only
17	**Import standards**		
18	**Export standards**		

	A	D	E
1	**Visitor Details**	**Google Analytics**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	yes	
7	**Internet Speed**	no	loading time only
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	no	
11	**Cookies Enabled**	yes	
12	**JavaScript Enabled**	yes	
13	**Username**	yes	with plugin or indivdual code
14			
15	**Traffic**	**Google Analytics**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	yes	
18	**Unique Visitors**	yes	Cookie based
19	**New/ Returning visitor**	yes	
20	**Events per Page**	yes	Most frequent events
21	**Events per Session**	unknown	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	no	can be added manually
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	yes	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	yes	
34			
35	**Monitoring & Reports**	**Google Analytics**	**Tool details**
36	**Click Path**	yes	
37	**Bot/Spider Filtering**	yes	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	yes	beta
40	**Visualization**	yes	
41	***Diagrams***	yes	
42	***click map***	yes	
43	***mouse tracking video***	no	
44	***heat map***	no	
45	***Visitor origins map***	yes	
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	yes	
48	**Mobile Browser Monitoring**	yes	
49	**Mobile Support**	no	but functions available in API
50	**Customizable Dashboard**	yes	
51	**Goal definition**	yes	
53	**Events**	**Google Analytics**	**Tool details**
54	**Forms**	no	
55	**Downloads**	yes	each link must be edited
56	**Flash Events**	yes	
57	**Multimedia Streams**	yes	
58	**Error Pages**	yes	
59	**RSS Feed Subscriptions**	yes	
60			
61	**Data privacy**	**Google Analytics**	**Tool details**
62	**IP-Masking**	yes	partly
63	**Cookie deactivation**	yes	Browser-addon
64	**Certificate**	no	

Open Web Analytics

	A	F	G
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**Open Web Analytics**	
3	**Manufacturer**	Peter Adams	
4	**Version**	1.5.0rc3	
5	**Keeping of data**	at the website	
6	**Licence model**	open source	
7	**Price**	free	
8	**Admin / User concept**	yes	
9	**Multi-language Support**	no	
10	**Support**	documentation, suppurt forums	
11	**Email Reports**	yes	
12			
13	**Input**	**Open Web Analytics**	**Tool details**
14	**Data source**	client-based	
15	**Technique**	JavaScript, PHP	
16	**Supported standards**		
17	**Import standards**		
18	**Export standards**	API	

	A	F	G
1	**Visitor Details**	**Open Web Analytics**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	yes	
7	**Internet Speed**	no	
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	no	
11	**Cookies Enabled**	unknown	
12	**JavaScript Enabled**	no	
13	**Username**	no	
14			
15	**Traffic**	**Open Web Analytics**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	yes	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	no	
20	**Events per Page**	yes	via custom variable
21	**Events per Session**	yes	
22	**Time Spent on Page**	yes	average time
23	**Time Spent per Session**	unknown	
24	**Exit Pages**	yes	
25	**Exit Links**	no	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	unknown	
31	**External Referrer**	yes	
32	**External Search Keywords**	unknown	
33	**Visit History**	yes	
34			
35	**Monitoring & Reports**	**Open Web Analytics**	**Tool details**
36	**Click Path**	yes	clickstream analysis
37	**Bot/Spider Filtering**	yes	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	yes	
40	**Visualization**	yes	
41	***Diagrams***	yes	graph, pie
42	***click map***	unknown	
43	***mouse tracking video***	yes	
44	***heat map***	yes	
45	***Visitor origins map***	yes	
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	no	
49	**Mobile Support**	no	
50	**Customizable Dashboard**	no	
51	**Goal definition**	yes	
53	**Events**	**Open Web Analytics**	**Tool details**
54	**Forms**	yes	action tracking
55	**Downloads**	yes	action tracking
56	**Flash Events**	yes	action tracking
57	**Multimedia Streams**	yes	action tracking
58	**Error Pages**	no	
59	**RSS Feed Subscriptions**	yes	
60			
61	**Data privacy**	**Open Web Analytics**	**Tool details**
62	**IP-Masking**	yes	
63	**Cookie deactivation**	no	
64	**Certificate**	no	

STATCounter

	A	H	I
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**STATCounter**	
3	**Manufacturer**	Aodhan Cullen	
4	**Version**	free version	
5	**Keeping of data**	STATCounter servers	
6	**Licence model**	per month	from 4€ - 95€
7	**Price**	free*	*in the basic version
8	**Admin / User concept**	yes	
9	**Multi-language Support**	no	
10	**Support**	yes	install guides, knowledge base, online support, forum, blog
11	**Email Reports**	yes	role-based
12			
13	**Input**	**STATCounter**	**Tool details**
14	**Data source**	client-based	
15	**Technique**	JavaScript, Cookies	
16	**Supported standards**	excel, csv	
17	**Import standards**		
18	**Export standards**		

	A	H	I
1	**Visitor Details**	**STATCounter**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	yes	
7	**Internet Speed**	no	
8	**Browser**	yes	
9	**Browser Language**	no	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	no	
12	**JavaScript Enabled**	yes	
13	**Username**	no	
14			
15	**Traffic**	**STATCounter**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	no	
18	**Unique Visitors**	yes	additional features
19	**New/ Returning visitor**	yes	
20	**Events per Page**	unknown	
21	**Events per Session**	unknown	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	yes	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	yes	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	yes	
34			
35	**Monitoring & Reports**	**STATCounter**	**Tool details**
36	**Click Path**	yes	
37	**Bot/Spider Filtering**	yes	image-based bots
38	**Bounce Rates**	no	
39	**Live Monitoring**	yes*	*hourly, not really real time?
40	**Visualization**	yes	also export posibilities for, e.g. excel
41	***Diagrams***	yes	
42	***click map***	no	
43	***mouse tracking video***	no	
44	***heat map***	no	
45	***Visitor origins map***	yes	
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	unknown	
49	**Mobile Support**	unknown	
50	**Customizable Dashboard**	yes	not really much customizing
51	**Goal definition**	no	
53	**Events**	**STATCounter**	**Tool details**
54	**Forms**	no	
55	**Downloads**	yes	
56	**Flash Events**	no	
57	**Multimedia Streams**	no	
58	**Error Pages**	no	
59	**RSS Feed Subscriptions**	no	
60			
61	**Data privacy**	**STATCounter**	**Tool details**
62	**IP-Masking**	unknown	
63	**Cookie deactivation**	yes	
64	**Certificate**	unknown	

Webalizer

	A	J	K
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**Webalizer**	
3	**Manufacturer**	B. Barrett	
4	**Version**	2.23-05	
5	**Keeping of data**	inhouse	
6	**Licence model**	free	
7	**Price**	free	
8	**Admin / User concept**	no	
9	**Multi-language Support**	yes	
10	**Support**	install guides	
11	**Email Reports**	no	
12			
13	**Input**	**Webalizer**	**Tool details**
14	**Data source**	Server-based	
15	**Technique**	Log Data	
16	**Supported standards**		
17	**Import standards**	clf, ncsa, wu-ftpd, w3c and more	
18	**Export standards**	clf, ncsa, wu-ftpd, w3c and more	

	A	J	K
1	**Visitor Details**	**Webalizer**	**Tool details**
2	**IP Address**	yes*	logged in data
3	**Country**	yes	
4	**City**	no	
5	**Operating System**	yes	
6	**Internet Service Provider**	no	
7	**Internet Speed**	no	
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	no	
11	**Cookies Enabled**	no	
12	**JavaScript Enabled**	no	
13	**Username**	no	
14			
15	**Traffic**	**Webalizer**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	no	
18	**Unique Visitors**	no	
19	**New/ Returning visitor**	no	
20	**Events per Page**	no	
21	**Events per Session**	no	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	no	
24	**Exit Pages**	yes	
25	**Exit Links**	no	
26	**Avg. Visits per Day/Month/Year**	yes	hour, day, motnth, year
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	no	*popular referrers
30	**Internal Search Keywords**	no	
31	**External Referrer**	no	*popular referrers
32	**External Search Keywords**	no	
33	**Visit History**	no	
34			
35	**Monitoring & Reports**	**Webalizer**	**Tool details**
36	**Click Path**	no	
37	**Bot/Spider Filtering**	no	
38	**Bounce Rates**	no	
39	**Live Monitoring**	no	
40	**Visualization**	yes*	
41	***Diagrams***	yes*	
42	***click map***	no	
43	***mouse tracking video***	no	
44	***heat map***	no	
45	***Visitor origins map***	no	
46	**Multi-Domain Tracking**	yes*	*filtering options
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	no	
49	**Mobile Support**	no	
50	**Customizable Dashboard**	yes*	user can decide which information to aggregate
51	**Goal definition**	no	
53	**Events**	**Webalizer**	**Tool details**
54	**Forms**	no	
55	**Downloads**	yes	*files
56	**Flash Events**	no	
57	**Multimedia Streams**	no	
58	**Error Pages**	no	
59	**RSS Feed Subscriptions**	no	
60			
61	**Data privacy**	**Webalizer**	**Tool details**
62	**IP-Masking**	unknown	
63	**Cookie deactivation**	no	
64	**Certificate**	no	

etracker

	A	L	M
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**etracker**	
3	**Manufacturer**	etracker GmbH	
4	**Version**	Basic	
5	**Keeping of data**	etracker Servers (Germany)	
6	**Licence model**	Pay per month	depends on page impressions (pi)
7	**Price**	9,90€/ 24,90€/ 49,90€	max. 100.000/ 500.000/ 1.500.000 pi
8	**Admin / User concept**	yes	unlimited number of users
9	**Multi-language Support**	yes	4 languages
10	**Support**	yes	
11	**Email Reports**	yes	
12			
13	**Input**	**etracker**	**Tool details**
14	**Data source**	client-based	
15	**Technique**	Javascript, Cookies, Pixel	
16	**Supported standards**	excel, csv, xml, API*, IntelliAd*, SEOlytics*	*unlimited version only!
17	**Import standards**		
18	**Export standards**		

	A	L	M
1	**Visitor Details**	**etracker**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	not in Basic version
6	**Internet Service Provider**	yes	
7	**Internet Speed**	yes	dsl/ modem/ cable/...
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	no	
12	**JavaScript Enabled**	yes	
13	**Username**	no	
14			
15	**Traffic**	**etracker**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	no	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	yes	
20	**Events per Page**	yes	has to be set up for each event
21	**Events per Session**	no	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	no	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	unknown	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	yes	
34			
35	**Monitoring & Reports**	**etracker**	**Tool details**
36	**Click Path**	yes	
37	**Bot/Spider Filtering**	unknown	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	yes	
40	**Visualization**	yes	
41	***Diagrams***	yes	
42	***click map***	yes	commercial add-on
43	***mouse tracking video***	yes	commercial add-on
44	***heat map***	yes	not in Basic version
45	***Visitor origins map***	yes	
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	yes	not in Basic version
49	**Mobile Support**	yes	iphone & ipad
50	**Customizable Dashboard**	yes	
51	**Goal definition**	yes	not in Basic version
53	**Events**	**etracker**	**Tool details**
54	**Forms**	yes	commercial add-on
55	**Downloads**	yes	
56	**Flash Events**	yes	
57	**Multimedia Streams**	yes	
58	**Error Pages**	no	
59	**RSS Feed Subscriptions**	no	
60			
61	**Data privacy**	**etracker**	**Tool details**
62	**IP-Masking**	yes	
63	**Cookie deactivation**	yes	
64	**Certificate**	yes	

Site Stats

	A	N	O
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**Site Stats**	
3	**Manufacturer**	Guanoo Inc.	
4	**Version**	Pro	
5	**Keeping of data**	Guanoo Servers (USA)	
6	**Licence model**	Pay per month	depends on page impressions (pi)
7	**Price**	20-120	0-1.000.000 pi
8	**Admin / User concept**	?	
9	**Multi-language Support**	no	
10	**Support**	yes	FAQ, installation help, email & phone support
11	**Email Reports**	yes	
12			
13	**Input**	**Site Stats**	**Tool details**
14	**Data source**	client-based	
15	**Technique**	Javascript, Cookies	
16	**Supported standards**	API, csv, alerts	
17	**Import standards**		
18	**Export standards**		

	A	N	O
1	**Visitor Details**	**Site Stats**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	yes	
7	**Internet Speed**	no	
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	no	
12	**JavaScript Enabled**	yes	
13	**Username**	no	
14			
15	**Traffic**	**Site Stats**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	yes	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	yes	
20	**Events per Page**	yes	# of clicks on a link
21	**Events per Session**	no	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	no	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	unknown	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	yes	in detail per visitor
34			
35	**Monitoring & Reports**	**Site Stats**	**Tool details**
36	**Click Path**	no	
37	**Bot/Spider Filtering**	yes	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	yes*	*hourly, not really real time?
40	**Visualization**	yes	
41	***Diagrams***	yes	
42	***click map***	no	
43	***mouse tracking video***	no	
44	***heat map***	no	
45	***Visitor origins map***	yes	
46	**Multi-Domain Tracking**	no	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	no	
49	**Mobile Support**	no	
50	**Customizable Dashboard**	no	
51	**Goal definition**	no	
53	**Events**	**Site Stats**	**Tool details**
54	**Forms**	no	
55	**Downloads**	yes	
56	**Flash Events**	no	
57	**Multimedia Streams**	no	
58	**Error Pages**	no	
59	**RSS Feed Subscriptions**	no	
60			
61	**Data privacy**	**Site Stats**	**Tool details**
62	**IP-Masking**	no	
63	**Cookie deactivation**	yes	
64	**Certificate**	yes	Privacy Bot Seal (unavailable!)

Deep Log Analyzer

	A	P
1	**General information**	**Tool**
2	**Tool Name**	**Deep Log Analyzer**
3	**Manufacturer**	Deep Software Inc.
4	**Version**	Version 4.0 (free Version)
5	**Keeping of data**	Server
6	**Licence model**	pay once
7	**Price**	free*
8	**Admin / User concept**	yes
9	**Multi-language Support**	yes
10	**Support**	yes
11	**Email Reports**	yes
12		
13	**Input**	**Deep Log Analyzer**
14	**Data source**	client-based, server-based
15	**Technique**	Javascript, Cookies, Log analysis
16	**Supported standards**	
17	**Import standards**	
18	**Export standards**	

	A	P	Q
1	**Visitor Details**	**Deep Log Analyzer**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	no	
5	**Operating System**	yes	
6	**Internet Service Provider**	no	
7	**Internet Speed**	no	
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	yes	
12	**JavaScript Enabled**	yes	
13	**Username**	yes	
14			
15	**Traffic**	**Deep Log Analyzer**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	no	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	yes	
20	**Events per Page**	no	
21	**Events per Session**	no	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	yes	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	unknown	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	yes	*aggregated
34			
35	**Monitoring & Reports**	**Deep Log Analyzer**	**Tool details**
36	**Click Path**	yes	
37	**Bot/Spider Filtering**	yes	
38	**Bounce Rates**	unknown	
39	**Live Monitoring**	yes	
40	**Visualization**	yes	
41	***Diagrams***	yes	
42	***click map***	unknown	
43	***mouse tracking video***	no	
44	***heat map***	unknown	
45	***Visitor origins map***	yes	
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	no	
49	**Mobile Support**	no	
50	**Customizable Dashboard**	partly yes	customizable reports
51	**Goal definition**	partly yes	planed schedule, calender,
53	**Events**	**Deep Log Analyzer**	**Tool details**
54	**Forms**		
55	**Downloads**	yes	
56	**Flash Events**	yes	
57	**Multimedia Streams**	yes	
58	**Error Pages**	yes	
59	**RSS Feed Subscriptions**	no	
60			
61	**Data privacy**	**Deep Log Analyzer**	**Tool details**
62	**IP-Masking**	unknown	
63	**Cookie deactivation**	yes	
64	**Certificate**	unknown	

Analyzer NX

	A	R	S
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**Analyzer NX**	
3	**Manufacturer**	AT Internet	
4	**Version**	Web edition	*four editions
5	**Keeping of data**	Server (Bordeaux, Frankfreich)	
6	**Licence model**	not available	
7	**Price**	individual pricing	
8	**Admin / User concept**	yes	
9	**Multi-language Support**	yes	
10	**Support**	yes	
11	**Email Reports**	yes	
12			
13	**Input**	**Analyzer NX**	**Tool details**
14	**Data source**	client-based	
15	**Technique**	cookies, traditional log analysis	
16	**Supported standards**	not available	
17	**Import standards**		
18	**Export standards**		

	A	R	S
1	**Visitor Details**	**Analyzer NX**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	unknown	
7	**Internet Speed**	unknown	
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	yes	
12	**JavaScript Enabled**	yes	
13	**Username**	yes	
14			
15	**Traffic**	**Analyzer NX**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	yes	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	yes	
20	**Events per Page**	yes	
21	**Events per Session**	yes	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	unknown	
24	**Exit Pages**	yes	
25	**Exit Links**	yes	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	*A/B Testing
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	yes	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	*long tail analysis
33	**Visit History**	yes	
34			
35	**Monitoring & Reports**	**Analyzer NX**	**Tool details**
36	**Click Path**	yes	*several kinds of (also Scroll View)
37	**Bot/Spider Filtering**	yes	
38	**Bounce Rates**	unknown	
39	**Live Monitoring**	yes	
40	**Visualization**	yes	
41	***Diagrams***	yes	
42	***click map***	yes	*ClickZone
43	***mouse tracking video***	yes	*ClickZone
44	***heat map***	yes	*ClickZone
45	***Visitor origins map***	yes	*segmentation
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	yes	*Rich Media, Web 2.0 Analysis
48	**Mobile Browser Monitoring**	yes	
49	**Mobile Support**	unknown	
50	**Customizable Dashboard**	yes	*very rich
51	**Goal definition**	yes	*long-term, short-term
53	**Events**	**Analyzer NX**	**Tool details**
54	**Forms**	yes	
55	**Downloads**	yes	
56	**Flash Events**	yes	*Event Log
57	**Multimedia Streams**	yes	*Web 2.0 Analysis
58	**Error Pages**	yes	*Event Log
59	**RSS Feed Subscriptions**	yes	
60			
61	**Data privacy**	**Analyzer NX**	**Tool details**
62	**IP-Masking**	yes	
63	**Cookie deactivation**	yes	
64	**Certificate**	yes	*OJD Certificate

Sawmill

	A	T	U
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**Sawmill**	
3	**Manufacturer**	Flowerfire	
4	**Version**	Enterprise (8)	
5	**Keeping of data**	inHouse	
6	**Licence model**	pay once	
7	**Price**	85-500€	
8	**Admin / User concept**	yes	
9	**Multi-language Support**	yes	English, German, Italian, Czech, Chinese (Traditional), Polish
10	**Support**	yes	can be bought additionally
11	**Email Reports**	?	
12			
13	**Input**	**Sawmill**	**Tool details**
14	**Data source**	server-based	
15	**Technique**	log-analysis	
16	**Supported standards**		
17	**Import standards**	884 supported log file formats, import of Sawmill Version 7 data	
18	**Export standards**		

	A	T	U
1	**Visitor Details**	**Sawmill**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	no	
7	**Internet Speed**	no	
8	**Browser**	yes	
9	**Browser Language**	no	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	no	
12	**JavaScript Enabled**	no	
13	**Username**	no	
14			
15	**Traffic**	**Sawmill**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	yes	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	no	
20	**Events per Page**	no	
21	**Events per Session**	yes	
22	**Time Spent on Page**	unknown	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	no	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	unknown	
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	no	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	unknown	
34			
35	**Monitoring & Reports**	**Sawmill**	**Tool details**
36	**Click Path**	no	
37	**Bot/Spider Filtering**	unknown	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	no	
40	**Visualization**	yes	graphs
41	***Diagrams***	no	
42	***click map***	no	
43	***mouse tracking video***	no	
44	***heat map***	no	
45	***Visitor origins map***	no	
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	no	
49	**Mobile Support**	no	
50	**Customizable Dashboard**	yes	
51	**Goal definition**	yes	
53	**Events**	**Sawmill**	**Tool details**
54	**Forms**	no	
55	**Downloads**	no	
56	**Flash Events**	no	
57	**Multimedia Streams**	no	
58	**Error Pages**	yes	
59	**RSS Feed Subscriptions**	yes	
60			
61	**Data privacy**	**Sawmill**	**Tool details**
62	**IP-Masking**	unknown	
63	**Cookie deactivation**	unknown	
64	**Certificate**	no	

W3 Statistics

	A	V	W
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**W3 Statistics**	
3	**Manufacturer**	W3 Solutions GmbH	
4	**Version**	free	
5	**Keeping of data**	?	
6	**Licence model**	open source	
7	**Price**	free	
8	**Admin / User concept**	no	
9	**Multi-language Support**	German, Englisch	
10	**Support**	yes	Forum
11	**Email Reports**	yes	
12			
13	**Input**	**W3 Statistics**	**Tool details**
14	**Data source**	client-based	
15	**Technique**	Pixel, (Log-Analysis starting at Professional Version)	
16	**Supported standards**		
17	**Import standards**		
18	**Export standards**	PDF, CSV, XLS, XML(via SOAP, starting at professional verion)	

	A	V	W
1	**Visitor Details**	**W3 Statistics**	**Tool details**
2	**IP Address**	no	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	no	
7	**Internet Speed**	yes	
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	unknown	
12	**JavaScript Enabled**	yes	
13	**Username**	no	
14			
15	**Traffic**	**W3 Statistics**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	unknown	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	unknown	
20	**Events per Page**	no	
21	**Events per Session**	no	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	no	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	yes	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	no	
34			
35	**Monitoring & Reports**	**W3 Statistics**	**Tool details**
36	**Click Path**	yes	
37	**Bot/Spider Filtering**	yes	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	unknown	
40	**Visualization**	yes	
41	***Diagrams***	yes	graph, pie, bar chart
42	***click map***	yes	
43	***mouse tracking video***	no	
44	***heat map***	no	
45	***Visitor origins map***	yes	
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	no	
49	**Mobile Support**	no	
50	**Customizable Dashboard**	yes	for each user
51	**Goal definition**	no	
53	**Events**	**W3 Statistics**	**Tool details**
54	**Forms**	no	
55	**Downloads**	yes	
56	**Flash Events**	no	
57	**Multimedia Streams**	no	
58	**Error Pages**	no	
59	**RSS Feed Subscriptions**	no	
60			
61	**Data privacy**	**W3 Statistics**	**Tool details**
62	**IP-Masking**	unknown	
63	**Cookie deactivation**	unknown	
64	**Certificate**	no	

Webtrekk Q3

	A	X	Y
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**Webtrekk Q3**	
3	**Manufacturer**	Webtrekk GmbH	
4	**Version**	4.0	
5	**Keeping of data**	in Europe	also in-House possible
6	**Licence model**	not available	
7	**Price**	not available	
8	**Admin / User concept**	Groups & Roles	
9	**Multi-language Support**	English, French	
10	**Support**	yes	
11	**Email Reports**	?	
12			
13	**Input**	**Webtrekk Q3**	**Tool details**
14	**Data source**	client-based	
15	**Technique**	Cookie	
16	**Supported standards**		
17	**Import standards**	via API	optional
18	**Export standards**	CSV	

	A	X	Y
1	**Visitor Details**	**Webtrekk Q3**	**Tool details**
2	**IP Address**	unknown	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	yes	
7	**Internet Speed**	yes	
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	yes	
12	**JavaScript Enabled**	yes	
13	**Username**	no	
14			
15	**Traffic**	**Webtrekk Q3**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	yes	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	yes	
20	**Events per Page**	unknown	
21	**Events per Session**	unknown	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	no	
24	**Exit Pages**	yes	
25	**Exit Links**	yes	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	yes	
30	**Internal Search Keywords**	yes	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	yes	
34			
35	**Monitoring & Reports**	**Webtrekk Q3**	**Tool details**
36	**Click Path**	yes	
37	**Bot/Spider Filtering**	unknown	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	yes	
40	**Visualization**	yes	
41	***Diagrams***	yes	netgraph
42	***click map***	no	
43	***mouse tracking video***	no	
44	***heat map***	yes	
45	***Visitor origins map***	no	
46	**Multi-Domain Tracking**	yes	
47	**Social Media Monitoring**	yes	
48	**Mobile Browser Monitoring**	yes	
49	**Mobile Support**	yes	
50	**Customizable Dashboard**	yes	
51	**Goal definition**	yes	
53	**Events**	**Webtrekk Q3**	**Tool details**
54	**Forms**	yes	
55	**Downloads**	yes	
56	**Flash Events**	yes	
57	**Multimedia Streams**	yes	
58	**Error Pages**	unknown	
59	**RSS Feed Subscriptions**	unknown	
60			
61	**Data privacy**	**Webtrekk Q3**	**Tool details**
62	**IP-Masking**	yes	
63	**Cookie deactivation**	yes	
64	**Certificate**	yes	TÜV

AWStats

	A	Z	AA
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**AWStats**	
3	**Manufacturer**	Laurent Destailleur	
4	**Version**	7.0	
5	**Keeping of data**	in-House	
6	**Licence model**	open source	
7	**Price**	free	
8	**Admin / User concept**	no	
9	**Multi-language Support**	yes	43 languages
10	**Support**	no	
11	**Email Reports**	no	
12			
13	**Input**	**AWStats**	**Tool details**
14	**Data source**	Server-Logfiles	javascript code for add. information available
15	**Technique**	Installation/ CGI-Script	externals cronjobs as scheduler needed
16	**Supported standards**	Apache, Webstar, IIS,...	XLF/ELF, CLF, personalized format, Web/FTP/Mail log files
17	**Import standards**		
18	**Export standards**		

	A	Z	AA
1	**Visitor Details**	**AWStats**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	From IP location/domain name
4	**City**	yes	need Maxmind database
5	**Operating System**	yes	need Maxmind database
6	**Internet Service Provider**	yes	need Maxmind database
7	**Internet Speed**	no	
8	**Browser**	yes	
9	**Browser Language**	no	
10	**Screen Resolution**	yes	with javascript only
11	**Cookies Enabled**	no	
12	**JavaScript Enabled**	yes	with javascript only
13	**Username**	no	
14			
15	**Traffic**	**AWStats**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	yes	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	yes	Caching irritates results
20	**Events per Page**	no	
21	**Events per Session**	no	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	no	
26	**Avg. Visits per Day/Month/Year**	yes	Daily, monthly, yearly
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	no	
30	**Internal Search Keywords**	no	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	no	
34			
35	**Monitoring & Reports**	**AWStats**	**Tool details**
36	**Click Path**	no	
37	**Bot/Spider Filtering**	yes	log file based only
38	**Bounce Rates**	no	
39	**Live Monitoring**	no	
40	**Visualization**	yes	
41	***Diagrams***	yes	
42	***click map***	no	
43	***mouse tracking video***	no	
44	***heat map***	no	
45	***Visitor origins map***	no	
46	**Multi-Domain Tracking**	no	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	no	
49	**Mobile Support**	no	
50	**Customizable Dashboard**	no	
51	**Goal definition**	no	
53	**Events**	**AWStats**	**Tool details**
54	**Forms**	no	
55	**Downloads**	no	
56	**Flash Events**	no	
57	**Multimedia Streams**	no	
58	**Error Pages**	yes	HTTP errors (e.g.error 404)
59	**RSS Feed Subscriptions**	no	
60			
61	**Data privacy**	**AWStats**	**Tool details**
62	**IP-Masking**	no	
63	**Cookie deactivation**	no	
64	**Certificate**	no	

clicky

	A	AB	AC
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**clicky**	
3	**Manufacturer**	Roxr Software, Ltd	
4	**Version**	2012	
5	**Keeping of data**	clicky Servers (USA)	
6	**Licence model**	Pay per month*	free for up to 3.000 pi
7	**Price**	$4.99/$9.99/$24.99	up to 10.000/30.000/100.000 pi per month
8	**Admin / User concept**	yes	premium feature
9	**Multi-language Support**	no	english only
10	**Support**	no	
11	**Email Reports**	yes*	premium feature
12			
13	**Input**	**clicky**	**Tool details**
14	**Data source**	client-based	
15	**Technique**	Javascript, Cookies	
16	**Supported standards**	API, csv*, pdf, alerts	*premium feature
17	**Import standards**		
18	**Export standards**		

	A	AB	AC
1	**Visitor Details**	**clicky**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	
4	**City**	yes	
5	**Operating System**	yes	
6	**Internet Service Provider**	yes	
7	**Internet Speed**	no	
8	**Browser**	yes	
9	**Browser Language**	yes	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	no	
12	**JavaScript Enabled**	yes	
13	**Username**	yes	
14			
15	**Traffic**	**clicky**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	yes	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	yes	
20	**Events per Page**	yes	
21	**Events per Session**	yes	
22	**Time Spent on Page**	yes	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	no	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	
29	**Internal Referrer**	no	
30	**Internal Search Keywords**	no	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	yes	
34			
35	**Monitoring & Reports**	**clicky**	**Tool details**
36	**Click Path**	yes	
37	**Bot/Spider Filtering**	yes	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	yes	
40	**Visualization**	yes	
41	***Diagrams***	yes	
42	***click map***	no	
43	***mouse tracking video***	no	
44	***heat map***	no	
45	***Visitor origins map***	yes	
46	**Multi-Domain Tracking**	unknown	
47	**Social Media Monitoring**	yes	only Twitter (premium feature)
48	**Mobile Browser Monitoring**	yes	mobile hardware only in premium version
49	**Mobile Support**	yes	third party applications
50	**Customizable Dashboard**	no	
51	**Goal definition**	yes	premium feature
53	**Events**	**clicky**	**Tool details**
54	**Forms**	no	
55	**Downloads**	yes	premium feature
56	**Flash Events**	yes	premium feature
57	**Multimedia Streams**	yes	yes
58	**Error Pages**	unknown	unknown
59	**RSS Feed Subscriptions**	yes	Feedburner subscribers
60			
61	**Data privacy**	**clicky**	**Tool details**
62	**IP-Masking**	no	
63	**Cookie deactivation**	no	
64	**Certificate**	no	

Logaholic

	A	AD	AE
1	**General information**	**Tool**	**Tool details**
2	**Tool Name**	**Logaholic**	
3	**Manufacturer**	Logaholic BV	
4	**Version**	Live Edition	also Web Analytics Edition
5	**Keeping of data**	Europe	ASP or inhouse
6	**Licence model**	Pay per month	restrictions
7	**Price**	6,99 $	30 day trial - unlimited storage
8	**Admin / User concept**	yes	
9	**Multi-language Support**	no	
10	**Support**	yes	direct contact via phone or email
11	**Email Reports**	yes	
12			
13	**Input**	**Logaholic**	**Tool details**
14	**Data source**	server-based	
15	**Technique**	JavaScript, Pixel-based	
16	**Supported standards**	csv, ncsa, iis, w3c	
17	**Import standards**		
18	**Export standards**		

	A	AD	AE
1	**Visitor Details**	**Logaholic**	**Tool details**
2	**IP Address**	yes	
3	**Country**	yes	*Top Contry
4	**City**	yes	*Top City
5	**Operating System**	yes	
6	**Internet Service Provider**	no	
7	**Internet Speed**	no	
8	**Browser**	yes	Version
9	**Browser Language**	no	
10	**Screen Resolution**	yes	
11	**Cookies Enabled**	no	
12	**JavaScript Enabled**	no	
13	**Username**	no	
14			
15	**Traffic**	**Logaholic**	**Tool details**
16	**Entry Pages**	yes	
17	**Page Views per User**	no	
18	**Unique Visitors**	yes	
19	**New/ Returning visitor**	yes	
20	**Events per Page**	no	
21	**Events per Session**	no	
22	**Time Spent on Page**	no	
23	**Time Spent per Session**	yes	
24	**Exit Pages**	yes	
25	**Exit Links**	no	
26	**Avg. Visits per Day/Month/Year**	yes	
27	**Total Visits**	yes	
28	**Most Popular Pages**	yes	*top: exit, entry, click path
29	**Internal Referrer**	no	
30	**Internal Search Keywords**	yes	
31	**External Referrer**	yes	
32	**External Search Keywords**	yes	
33	**Visit History**	yes	
34			
35	**Monitoring & Reports**	**Logaholic**	**Tool details**
36	**Click Path**	yes	
37	**Bot/Spider Filtering**	yes	
38	**Bounce Rates**	yes	
39	**Live Monitoring**	no	
40	**Visualization**	yes	
41	***Diagrams***	yes	
42	***click map***	unknown	conversion paths
43	***mouse tracking video***	unknown	
44	***heat map***	no	
45	***Visitor origins map***	no	
46	**Multi-Domain Tracking**	unknown	
47	**Social Media Monitoring**	no	
48	**Mobile Browser Monitoring**	no	
49	**Mobile Support**	no	
50	**Customizable Dashboard**	yes	
51	**Goal definition**		
53	**Events**	**Logaholic**	**Tool details**
54	**Forms**	no	
55	**Downloads**	no	
56	**Flash Events**	no	
57	**Multimedia Streams**	no	
58	**Error Pages**	yes	logfile-based
59	**RSS Feed Subscriptions**	no	
60			
61	**Data privacy**	**Logaholic**	**Tool details**
62	**IP-Masking**	no	
63	**Cookie deactivation**	no	
64	**Certificate**	no	

Appendix B – Exemplary Link Structure

- Political and Economic Environment
  - Introduction to the Country
    - General Facts & Information about the Country
      - World Fact book
      - Information about the UK
      - General Information about the UK
    - Political Environment
      - Politics of the UK
      - Political parties
      - Politics of the UK
      - British political Institutions
      - Political environment
    - The Regions of the Country
      - Regions of England
      - Regions overview
      - Regions of England
      - Regions of the UK
      - Regions of the UK
    - Regional versus National Interests
      - About the UK
    - Economic and Statistical Data
      - GDP
      - Inflation
      - Other economic indices
      - Information on trade
      - Economy of the UK
      - CIA Factbook: UK
      - National Statistics
      - Economy of the UK
      - Economic indicators

- Sector Specific Information
  - Economic and Social Council
  - Economic sector reports
  - Economic of the UK
  - Economic and Social Research Council
  - Economy of the UK
- Government (Central & Local)
  - System of government
  - Local authorities
  - Politics of the UK
  - Politics of the UK
  - Government systems of the UK
- Legal Environment (including product & customer safety)
  - Department of Trade and Industry
  - Health & Safety Executive
  - Product safety
  - Health and safety
- ICT Facilities and Competencies

Market Entry: Requirements and Restrictions

- Embassy
- UK Trade and Investment
- Trade Association Forum
- HM Customs & Excise
- Legal Services
- HM Revenue and Customs Trade Info
- HM Revenue and Customs
- London Chamber of Commerce
- UK Trade and Investment
- Department of Trade and Industry
- Business Link
- Start-up Business Advice
- London Chamber of Commerce

Franco-British Chamber of Commerce and Industry

German-British Chamber of Commerce and Industry

Export

HM Revenue & Customs

UK Trade and Investment

HM Revenue and Customs Trade Info

HM Revenue and Customs Trade Info

British commodities export to France

New British products available for export

English companies looking for German business partners

Licensing & Franchising

Department of Trade and Industry

British Franchise Association

British Association

Guide to UK Franchise Market

Product list requiring a license to be traded

Guide how to set up a business in the UK

Joint Ventures & Partnerships

Business Link

Import Regulations & Restrictions

HR Revenue and Customs

HR Revenue and Customs

UK Trade and Investment

Department of Trade and Industry

Business Link

British commodities imported to the UK from France

Financial Concerns and Terms of Payment

Financial Services Authority

IS4Profit Business Profit Organization

Start-up Business Advice

Bank of England

Medical Care & Insurance

Department of Work and Pensions

Health Advice for travelers

Information on the European Health Insurance card

(blue: links which open in the main content window; grey: external links)

Appendix C – Web Page Usage

1.0 Political and Economic Environment	145
1.1 Introduction to the Country	22
1.1.1 General Facts & Information about the Country	8
1.1.2 Political Environment	10
1.1.3 The Regions of the Country	2
1.1.4 Regional versus National Interests	4
1.1.5 Economic and Statistical Data	4
1.1.6 Sector Specific Information	1
1.1.7 Government (Central & Local)	7
1.1.8 Legal Environment (including product & customer safety)	9
1.1.9 ICT Facilities and Competencies	1
1.2 Market Entry: Requirements and Restrictions	15
1.2.1 Export	0
1.2.2 Licensing & Franchising	1
1.2.3 Joint Ventures & Partnerships	1
1.2.4 Import Regulations & Restrictions	2
1.2.5 Financial Concerns and Terms of Payment	3
2.0 General Business Environment	37
2.1 Business Etiquette - Attitudes and Values	12
2.1.1 Business Relationships (people vs. paper)	13
2.1.2 The Work-Life Balance	11
2.1.3 The Importance of the Family	8
2.1.4 Attitudes	11
2.1.5 Taboos	16
2.1.6 Time Keeping	5
2.1.7 Other Issues (colour, mobility)	3
2.2 Business Ethics	9
2.2.1 Gift giving	8
2.2.2 Bribery and corruption	3
2.2.3 Environmental issues	8
2.3 Education and Training	5
2.3.1 General Education	3
2.3.2 Foreign Language Competence	1
2.3.3 Cultural Awareness	5
3.0 Business Practice	45
3.1 How to Communicate	14
3.1.1 Making contact (distance where/when appropriate)	16
3.1.2 Face-to-face (including verbal and non-verbal communication)	21