Master’s Thesis

Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

Submitted by:
Livia Hächler
Baumackerstrasse 9
8050 Zurich
lihaechler@gmx.ch
Matriculation number: 04-450-458

Advised by:
Edy Portmann

Supervised by:
Prof. Dr. Louis Bosshart
Prof. Dr. Andreas Meier

Submitted on:
August 31st 2010
Abstract

The Internet has influenced the nature of public communication, thus it is directly influencing (online) journalism. As a medium overcoming the barriers of time and space, the Web is a mixed blessing. On the one hand it facilitates fast information exchange, at the same time it erodes the traditional means of existence, leading to sparse resources and drastically increased time pressure. Especially online journalists are confronted with enormous time constraints for two different reasons. Firstly, online news publishing is a matter of minutes. Secondly, the Web has changed the readers’ behavior causing a 24 hour news cycle that is constantly consumed in small bits.

In such an environment it has become difficult to meet the professional standards expected of a journalist. Achieving the professional objective concerning research and uncovering the truth requires time, and time has become very limited in an online setting. Consequently, to do research quickly the Web has become to be an important tool, simply because it is fast.

This thesis tries to reveal the potential of the Internet for research that meets professional journalistic standards. As a point of reference, the momentary situation of the use of the Web for research purposes among online journalists in Swiss online news outlets has been examined. Swiss online journalists seem to vary in how they use the Web depending on their own IT-knowledge or on the media organization they are working for.

Taking a deeper look into the nature of the Web 2.0 and its tools as well as that of the Web 3.0, it can be concluded that the main problem in using online sources is their trustworthiness. The advantage of the Web, namely its openness and participative nature, is at the same time its drawback because the information is never completely trustworthy in regard to today’s ways of ascribing trust. The Web is a communication platform flattening hierarchies and forming public communication in previously unknown ways. To use the Web to its fullest the author concludes that these developments might entail some fundamental paradigm shifts in society as well as in the field of journalism.

Keywords: journalism, research, professionalism, Internet, Web 2.0, Web 3.0, semantic Web, social media, blogs, twitter, social networks, folksonomies, search engines
List of Contents

1. Introduction ........................................................................................................................................... 1

2. The Function of Journalism in Society ................................................................................................. 4
   2.1. The Meaning of Public Communication in Society ......................................................... 4
   2.2. The Debate on Journalistic Professionalism ................................................................. 6
   2.3. The Meaning of Research within the Journalistic Process ........................................... 10

3. The Internet as a New Mass Medium .............................................................................................. 15
   3.1. Structural Changes in the Media System Caused by the Internet .................................. 15
   3.2. The Role of the Online Journalist .................................................................................. 20
   3.3. The Meaning of Research in the Online Environment .................................................... 23

4. The Use of the Internet as a Research Tool in Swiss Online Editorial Offices .......... 26
   4.1. The Research Questions ...................................................................................................... 26
   4.2. The Method ....................................................................................................................... 27
   4.3. The Results ....................................................................................................................... 30
       4.3.1. Where Online Journalists Search ............................................................................. 30
       4.3.2. The Art of Using Google .......................................................................................... 32
       4.3.3. Searching Through the Web 2.0 .............................................................................. 34
           4.3.3.1. Social Networks ................................................................................................. 34
           4.3.3.2. Microblogs .......................................................................................................... 35
           4.3.3.3. Blogs .................................................................................................................. 36
           4.3.3.4. Wikis .................................................................................................................. 37
           4.3.3.5. Tagging .............................................................................................................. 38
       4.3.4. Investigating Credibility and Trustworthiness Online ............................................ 39
       4.3.5. Giving Machines the Power of Decision ................................................................. 41
       4.3.6. Visualizing the Future ............................................................................................... 41
   4.4. Short summary ............................................................................................................................ 43

5. The Characteristics of the Web 2.0 and 3.0 .................................................................................... 46
   5.1. The Web 2.0 .......................................................................................................................... 46
       5.1.1. The Users’ Web .......................................................................................................... 48
       5.1.2. A Web of Interactions ............................................................................................... 50
       5.1.3. Limits to the Web 2.0 ............................................................................................... 55
   5.2. The Web 3.0 .......................................................................................................................... 56
       5.2.1. The Web of Data ........................................................................................................ 57
           5.2.1.1. Exploring Meanings through Machine-Generated Languages ..................... 58
       5.2.2. Enhancing Personalization ......................................................................................... 60
       5.2.3. The Importance of Trust ............................................................................................ 62
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

6. The Research Possibilities on the Internet ................................................................. 64
   6.1. Blogs – Start Discussing! ......................................................................................... 64
   6.2. Microblogging – Start Contributing! ................................................................. 69
   6.3. Social Networks – Start Connecting! ................................................................. 72
   6.4. Wikis – Start Editing! ......................................................................................... 76
   6.5. Folksonomies – Start Organizing! ................................................................. 79
   6.6. Search Engines – Start Finding! ......................................................................... 83

7. The Responsible and Desirable Use of the Internet for Journalistic Research ......87
   7.1. The Requirements of Professional Journalistic Research................................. 87
   7.2. The Momentary Use of the Internet in Swiss Online Editorial Offices .......... 88
   7.3. The Internet’s Possibilities for Credible and Efficient Research.................... 88
       7.3.1. Answering the Wh-questions ....................................................................... 90
       7.3.2. The Problem of Trust ................................................................................. 93

8. Conclusion and Outlook .................................................................................................. 97

9. Bibliography ..................................................................................................................... 101

Appendix .......................................................................................................................... 110

Appendix 1: Interview Questions for Online Journalists .............................................. I
Appendix 2: Interview Questions for Media Coaches ...................................................... II
List of Figures

Figure 1: “The Long Tail” (source: Anderson [2005]) ................................................................. 21
Figure 2: The evolution of the Web (source: Murugesan [2010]) ................................................. 47
Figure 3: A RDF-Triple (source: Klyne / Carroll [2004]) ........................................................... 59
Figure 4: Examples of tweets on Twitter (29th July 2010) ............................................................ 69
1. **Introduction**

“The Web was not a physical ‘thing’ that existed in a certain ‘place’. It was a ‘space’ in which information could exist.”

These words stem from the founder of the Internet Tim Berners-Lee and describe the original idea of the Internet. It nicely depicts how the Web is what it is solely through information and how that information is being shared. The information is actually forming the Web. Taking this into account, it is clear that the Web is greatly relevant for journalism, as it is the profession dealing with information and its dissemination.

Online journalism has especially been influenced by this new medium in three main ways. Firstly and obviously, the Web has actually created online journalism by offering a new and fast channel for publications. Secondly, the Internet has changed the audience’s behavior in regard to news consumption and advertising resulting in new economic situations for the publishers. Traditional ways of earning money have started to vanish, directly affecting the resources. Thirdly, the Internet has opened new ways of gathering information, thus changing traditional workflows and, especially, accelerating the whole process of publishing.

Considering the importance and the role of public information in a democratic society, these changes are not marginal, but rather fundamental. The Web is changing how people interact and how organizations or politicians are interacting with people. The communication streams are no longer one-way streets, but much more a two-way freeway consisting of numerous fast lanes. Even though the traditional gatekeeper might have become irrelevant in this new type of environment, professional mediators are still needed to provide credible and trustworthy information. However, these professional mediators are the ones being affected the most by the structural changes caused by the Web. Professionalism is sought, but now needs to be achieved in a different environment, on a different basis. Developments in technology are occurring quickly, making it hard to follow as a layman. With resources being sparse, it is especially online journalists who are lacking the time to study information technology and the constantly changing Web tools.

This thesis tries to suggest solutions to this problem by analyzing the online journalists’ demands of professional research and the Web’s possibilities in meeting these demands. The focus lies on online journalists for the following reasons: Online journalists are

---

confronted with time pressure to a much greater extent than other journalists. Being the first to publish something online is usually a matter of minutes. Time then becomes one of the most central factors in the working process, making the procedure of finding relevant information quickly a key capability. Additionally, one can conclude that, being under such a tight time schedule, the time left for online journalists to get to know the Web and its possibilities is very limited.

Knowing the research tools is essential in journalism and a lack of such knowledge can result in decreasing quality. Considering the importance of professional public communication in a democratic society, such a decrease is not desirable. By combining the two fields of media sciences and information technology this thesis tries to not only describe the momentary situation of online research, but also to take a look at expected future developments in technology and their effect on online research possibilities. Media sciences tend to focus on events and social occurrences at present or in the past. Seldom do they illustrate the needs of current and future behaviors. This paper is an attempt, for once, to turn one’s attention towards the existing and especially future possibilities and needs. As will be shown, the Web is by no means a finished product, but a medium that is in need of being formed. In analyzing it in retrospect, it is justifiable to fear that the science of media communication and journalists might miss some great chances of actually forming it in their best interests.

With this in the back of one’s mind, the thesis was written to address online journalists or people with the equivalent knowledge. The technological aspects and information technology presented here do not go into any irrelevant technical details. Considering the fact that one Web tool alone could be the topic of a whole thesis, it was difficult to set limits on the presented information. The core notion of the Web, however, was identified and connected to the professional journalist’s needs.

The structure of the thesis is as follows:

Chapter 2 and 3 present the current situation of online journalism from the perspective of the media and communication science. In chapter 2 general theories about journalism, its importance and function in society as well as its standards of professionalism and research are explained. Chapter 3 highlights the changes that the Internet as a new mass medium has brought and identifies the new demands towards a journalist working on the Web.

To gain insight into how knowledgeable online journalists are today about the Web as a research tool and how they implement it, interviews were held with journalists of Swiss online news outlets. Chapter 4 describes this research and its results.
Chapter 5 starts to examine information technology presenting the characteristics of the existing Web 2.0 and a future Web 3.0. The different (typical) tools and social media existing on today’s Web are reflected in more detail in chapter 6.

Chapter 7 combines the knowledge collected to answer the initial question of what possibilities the Web offers online journalists to fulfill professional expectations. The final answer to the thesis’s research question of how and where to find credible and relevant information online, is answered in the conclusion in chapter 8.
2. The Function of Journalism in Society

2.1. The Meaning of Public Communication in Society

There have been many different attempts to describe the structure and the characteristics of today’s society. One theory, common in Anglo-American studies, is the concept of an information society. Five different dimensions outline the characteristics of this kind of society: technology, economy, occupation, space and culture. It is argued that all of these dimensions have been profoundly changed through the growth of informational activities, information networks or simply through the quantity and quality of information.\(^2\)

“…[Q]uantitative changes in information are bringing into being a qualitatively new sort of social system, the information society.”\(^3\) However, it is criticized that the idea of a new information society overly emphasizes a primacy of change over continuity. William Martins states that it is questioned to what extent an information society can really be differentiated from an industrial society.\(^4\) He maintains that nothing has changed. “All that had happened was that the dominant forces in capitalist society had found new avenues of exploitation and new technological means by which to pursue them.”\(^5\)

Secondly, in German speaking countries the theory is of a media society, which has gained more influence than the theory of an information society.\(^6\) The media is seen as the basic element and primary institution of that society. Media communication has penetrated all spheres of social life and affects all social communication. The main traits of a media society are the disengagement of the media from political or social actors, the acceleration of communication and the linkage of global and local communication. Critics here proclaim the dominance of the role of the media. The logic of the media is seen to have an absolute control over social life.

In summary, as Bonfadelli and Meier have argued, neither concept is complete in itself.\(^7\) However, what is of crucial importance in both theories is the information, respectively the transmission of information. Then considering society as functionally differentiated, it is

---

\(^3\) Ibid. p. 9.
\(^5\) Ibid. p. 5.
\(^7\) Ibid. p. 35.
clear that the different social systems interact through communication. Martin even calls information the lifeblood of society and points out the active role of information in shaping society. “Information is not just affected by its environment, but is itself an actor affecting other elements in its environment. Information is not just embedded within a social structure but creates that structure itself.” Manfred Rühl contributes to the same theory when claiming that information is produced in temporal, spatial and social interdependency. It is the primary and crucial function of journalism to produce and allocate issues and information for public communication. “[…] der Mangel an kontinuierlicher journalistischer Information beschränkt die Wahrmöglichenkeiten für Entscheidungen in vielen Lebenslagen und das Erleben innerhalb einer entwickelten Gesellschaft ohne journalistische Informationen büsst an Stabilität ein.”

Modern society is characterized by many interdependently linked subsystems, which leads to great instability. Journalism, as a social subsystem, helps solving this problem by structuring and reducing the complexity of the world. It has developed programs to conquer the fast growing information flow, to structure it and to make it manageable for the public. It is of vital importance for the different social subsystems to be able to observe and follow the developments of the other systems. There is an intrinsic need for constant observation of the specific subsystem’s environment. Journalism is the subsystem, which has made this problem its own task. Journalism communicates issues that are relevant for at least one other subsystem.

Following Luhmann, the mass media – within which journalism acts as a program – conducts the introspection of society. The mass media guarantees all social subsystems a publicly accepted and known presence, which acts as the basis for any following

10 Ibid. p. 21.
12 Ibid. p. 238.
communication and action. He summarizes that the mass media produces the memory of society. And this memory supports the constant connecting of the past with the future. Linking to the theories of constructivism, it is clear that the media cannot depict an absolute reality in an objective way. “Journalismus erscheint insgesamt als eine prinzipiell wandelbare, vielfältig bedingte, eigengesetzlich operierende, fakten- und aktualitätsbezogene Wirklichkeitskonstruktion in den Medien.” The media creates a reality, which is determined by the programs and means in which journalists select what is newsworthy and of public interest. Rühl states that abnormal and extraordinary matters are favored and, that through the relationship between journalism and its audience, a ‘symbolic reality of the extraordinary’ is constructed. He maintains that the public is not able to recognize that the published issues are a result of journalistic selection, thus claiming that these issues are often taken as the reality. The increasing importance of the media as the generators of reality implies the significance of the journalist himself and of the practice of his profession.

2.2. The Debate on Journalistic Professionalism

It is difficult, if not impossible, to define the concept of a professional journalist. Journalism has differentiated itself to react to the complexity of the world to the extent that there are many varying images of the journalist’s occupation. At this point, what can be found as common ground in all theories is the conviction that journalists need to act in a professional manner in order to protect their freedom and independence, which is vital for the communication processes in a democracy. Pamela J. Shoemaker and Stephen D. Reese define a profession by the following characteristics:

1) It is a full-time occupation.

2) Its practitioners are deeply committed to the goals of the profession.

---

3) Entrance to and continuance in the profession are governed by a formal organization that has established professional standards.

4) Its practitioners are admitted to the profession following prescribed formal schooling and the acquisition of a specialized body of knowledge.

5) It must serve society.

6) Its members must have a high degree of autonomy.

Barbie Zelizer quotes sociologists when saying that the subsequent traits identify professions in general: a certain level of skill, autonomy, service orientation, licensing procedures, a testing of competence, an organization, codes of conduct, training and educational programs.\(^\text{19}\) It appears that journalists meet only few of these criteria. There is no special licensing organization controlling the entrance to journalism. Based on the constitutional freedom of speech – at least in democracies – everybody can be a journalist. Even though the number of journalism programs offered at universities has increased, there is no strict requirement for a journalist to have a degree in journalism. Another important factor is the autonomy. Shoemaker and Reese argue that “journalists as a group are subject to a wide range of organizational constraints that dictate what they do and when they do it.”\(^\text{20}\)

Is one forced to conclude that journalism is not a profession? It is not quite that simple. John Hohenberg states that journalism has its own disciplines and its share of necessary routine like any other profession.\(^\text{21}\) Although he admits that there is no complete response to the simple question of “What is a journalist?”, he sees an essential element that is implicit to the journalist and to the assessment of his profession: “It is a love of his work and a fierce and uncompromising belief in its importance.”\(^\text{22}\) For Michael Haller, the basic qualification of a journalist is the skilled and public communication of current events.\(^\text{23}\) Zelizer refers to a set of practices by which one qualifies as a “journalist”.\(^\text{24}\)

Many authors have tried to make a more differentiated attempt in describing journalistic professionalism by naming two relevant fields. John Herbert calls for knowledge and


\(^{22}\) Ibid. p. 14.


ability. Rühl describes professionalization as a process much more than a state, which is defined by the two aspects of journalistic education and journalistic socialization. The first mentioned, journalistic education, corresponds to the cognitive structures of journalistic behavior. It refers to the knowledge and the knowhow of journalistic practices and decisions. By journalistic socialization Rühl introduces the more informal side of journalism. It is the learning of generalized ethical norms and attitudes in the daily routines. He then identifies journalistic professionalization as the complex process of incorporating the individual into the occupation. Parallel to Herbert and Rühl, there is Siegfried Weischenberg’s dichotomy of journalistic expertise, which is defined by profound knowledge of the matter (“Sachkompetenz”) and professional competence (“Fachkompetenz”). By “Sachkompetenz” Weischenberg indicates the importance of knowledge about the subject to be published itself. This may include specialized knowledge within the field or about an issue. The “Fachkompetenz” is comparable to Rühl’s journalistic education. It implies the instrumental knowledge and know-how, such as research, selection of issues, editing, organization and technology. It also includes the knowledge about the media and the ability to reflect on the influence of one’s action, on the process of news production and on the economic, political, technical and organizational conditions and constraints.

Michael Krzeminski combines many of the previously stated points and states that journalistic action is based on five competencies and profound knowledge in: technology, the matter itself, expertise, mediation and in social questions and relations. Furthermore, he indicates three characteristics that account for professional work. It is the systematization of the required knowledge, the orientation along socially accepted norms and a self-organization meaning the creation of a social system. “Mit Professionalität lässt sich eine durch Organisation, Ethik, Wissenschaft und Berufskultur in besonderer Weise qualifizierte Form der Berufsausübung bezeichnen, deren Standards durch Ausbildung und berufliche Sozialisation vermittelt und im Berufsleben habituell

---

27 Ibid. p. 428.
30 Ibid. p. 15.
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

ausgeprägt werden.”

The important aspect of this definition is that one can and one needs to learn to work professionally.

In a more generalized view, the process of professionalization has emerged through the vertical and horizontal processes of differentiation, which were essential to the growth of journalism. Specific roles and techniques were created and the technical knowledge of journalistic expertise became of greater importance than the political or moral attitudes of a journalist. According to Blöbaum, journalism has created five programs to meet its social task. The first program consists of sorting and organizing. This may refer to the different editorial offices or even the different editorial sections within a newspaper, such as regional or international news. The second program indicates the presentation referring to the different media types. The techniques of this program can change through the introduction of new media technologies. The collection of information constitutes the third program and involves the techniques of journalistic research. Number four is the selection program. This program states the criteria according to which information becomes relevant news. The last program corresponds to the act of verification. All information within the journalistic system can potentially be investigated and confirmed. To Blöbaum the possibility of the investigation of every fact has a threatening effect on the informants.

The journalistic structures and programs decouple the individual from its occupational role and stabilize the whole system. According to Rühl, instability would make journalism incapable of communicating. “Es ist ein zeitlich vorab ausgewähltes Repertoire an Entscheidungsprämissen, das das journalistische Leben strukturiert.” These programs of journalistic selection include specific expectations of norms, roles, positions and techniques, which in the end set up a reliable pattern to conquer the complexity of the world and the events to be described. The journalist then communicates not (solely) as an individual but through his role within the journalistic system. This formal responsibility reduces the pressure on the journalist as an individual without removing the personal responsibility in working accurately and professionally.

31 Ibid. p. 19 et seq. In English: “Professionalism is the specially qualified form of executing one's profession through organization, ethics, science and occupational culture, whose standards are conveyed through education and occupational socialization and are developed and shaped through the habits of the daily professional life.”
33 Ibid. p. 209 et seq.
In conclusion the concept of the professional journalist is highly pluralistic. To be able to refer to a set of criteria when arguing on professionalism in Internet research, an excerpt of the definitions of journalists’ responsibilities by the “Schweizer Presserat” is used. The “Schweizer Presserat” is a foundation of different Swiss associations of media professionals and is the authority, which can be contacted by the public in case of complaints concerning media ethics.\textsuperscript{35} The following points are considered the responsibilities of a journalist:\textsuperscript{36}

1) Journalists seek the truth no matter what the consequences entail. They let themselves be led by the right of the public to know the truth.

2) They defend the freedom of information and the hereto-connected rights, the freedom of commentary and of criticism and the independence and reputation of their occupation.

3) They publish only information, documents, pictures and audio from sources that are known to them. They do not suppress any important elements of information nor do they misrepresent any facts, documents, pictures, sounds or any from others expressed opinions. They mark unconfirmed announcements, pictures or audio bites as such.

4) They do not collect information, audio bites, pictures or documents in a dishonest way. They do not edit any pictures to falsify the original. They do not plagiarize.

5) They correct published information of which the content has become partially or totally wrong.

6) They protect the secrecy of the editorial office and do not divulge the sources of confidential information.

2.3. The Meaning of Research within the Journalistic Process

“Journalism is a simple profession. It is all about asking questions.”\textsuperscript{37} Herbert continues that the profession of journalism is focused on knowing the right questions to ask. Research then is the skill with which journalists do all their questioning. As Ambros


\textsuperscript{37} Herbert (2000): p. 5.
Kindel points out, journalism without research would become empty and far from reality. He identifies research as one of the most significant processes within the journalistic working process. Thomas Leif even names research the instrument to accomplish journalistic quality. For its utmost impact this instrument constantly needs to be cultivated, questioned and developed.

Many authors in journalistic literature emphasize that research is not just a skill but a craft one can and needs to learn. Leif even proclaims that research education should become as self-evident to a journalist as one’s driver’s license. Johannes Ludwig calls research the retrospective reconstructing of events and relationships. He stresses the fact that journalists are constructors of reality and that they need to have the relevant know-how. For Haller, research is the systematic collecting and evaluating of empirical issues with the goal of reconstructing real procedures. It is the purpose of every research to understand the progression and meaning of the event and to publish these results. Research then is the technique to check and verify information and its sources. According to Kindel, research means to ask the right questions of the right person at the right time. It is the method of taking nothing for granted or for obvious, but rather to collect as much information as possible concerning a certain issue. Hohenberg as well constitutes the first rule in journalism as taking nothing for granted. He sees the hallmark of journalism in skepticism. Volker Wolff agrees by claiming that in journalism one should constantly be inquiring. It is research that uncovers the new and important aspects of any information. He maintains that it is research that differentiates journalism from simple announcements and makes it journalism in terms of its public task. “Alle, die in den Medien auf das Recherchieren verzichten, versagen bei ihrer Arbeit.”

---

40 Ibid. p. 10.
According to Haller, journalistic research has emerged as a professional procedure since the end of the 19th century. There are three different types of research to be defined:

1) Research connected to a certain event: Core tasks of this research type are found in the checking and completing of the information, as well as the reconstructing and describing of the events including the actors.

2) Research on a certain topic or issue: A certain topic or issue of present importance is the main subject of this research, which is conducted through thesis and experience.

3) Investigative research: This kind of research is of great importance, especially in Anglo-American countries. The journalist here is seen as a watchdog uncovering grievances and injustice so as to make them known to the public.

All types of research and the research methods have gone through different developments, which are dependent on the social and political structures the journalists are embedded in. Haller claims that the more a society emphasizes the importance of free access to information, the more distinct the motivation of journalists to do good research will be.46

In response to the question if journalists have always done research in the same way, Haller answers with yes and no. Yes, because there are certain well-proven methods that make for good research and good journalism. They can be learned and are dependent on the abilities of the journalist himself. No, because research is always linked to the function of journalism and the mass media in society. Social orders and the conception of the journalist’s role have changed and will keep on changing. However, journalists know that research is their most important activity.47

Remembering the responsibilities of a professional journalist stated at the end of the previous chapter, research is the method to fulfill the following:

1) Through research journalists try to seek the truth.
2) Through inexhaustible search for the relevant information and its publication, journalists defend the right to information and the freedom of commentary and criticism.
3) Through the checking of sources they ensure publishing only the information that is known to them. Through a comprehensive and extensive research all the important and relevant elements pertaining to an issue shall be considered and expressed.

Haller identifies three main claims accounting for research: Relevance, validity, and comprehensibility. Each claim requires research in itself, concluding that the objective of

47 Cp. ibid. p. 16 et seq.
any research lies within the tackling of these claims. If this goal is achieved, the journalist will be able to present the facts and circumstances clearly, he will be able to reconstruct the chronology of the events and name the responsible actors within, he will be able to sketch most possible causes and consequences as well as showing their importance for the public. The first claim, relevance, is described as being like a filter. It ensures that the debated issue is of great enough importance to be newsworthy. Haller states that the initial information needs to enclose generally important and interesting aspects concerning the target reader as well. The decision of relevancy is not only based on the question of who the active actors within the issue are, but also on the profile of the journalist’s media organization. Its audience and their interests can vary depending on the media type as well as the topical and geographical range of coverage. Ele Schöfthaler writes that especially when the journalist is personally very interested in a topic he needs to check thoroughly to whom else the story would apply.

The second claim of validity is the most central.

“Jedes journalistische Thema basiert auf überprüfmbaren Aussagen (= Informationen) über Vorgänge und Ereignisse, deren Gültigkeit als Erstes abgeklärt werden muss. Jede Recherche beginnt also mit der Überprüfung der Informationen, die den Ausgangspunkt lieferten (Basisrecherche).”

This claim deals with the reliability of the facts and sources. In his work the journalist should always be led by skepticism, be it towards the source or the information itself. Especially sources need to be treated with reservation. The common rule here is that an informant can be trusted the most when he is most unbiased towards the issue in question. If the source represents specific interests or is even involved in the events, it cannot be viewed as neutral and reliable. Wolff writes that information from one source needs to be confirmed by at least one other source, especially since journalists are often subject to exploitation by companies and politicians for the sake of their own interests. He identifies the controlling of sources and their motives as the first and most important step within the research process. To reach the full truth it is a must for a journalist to include all attainable and relevant sources in his investigation. Kindel remarks that a source might

---

49 Cp. Ibid. p. 55.
50 Schöfthaler, Ele (1997): Recherche praktisch. Ein Handbuch für Ausbildung und Praxis. München: List Verlag. P. 61. In English: “Every journalistic topic is based on verifiable statements (=information) – pertaining to incidents and events – whose validity has to be clarified. Every investigation, therefore, begins with the checking of the information that provided the starting point (Basic research).”
become the subject of research itself.\textsuperscript{53} Knowing the source and checking its reliability is the basis for secured and trusted facts. Furthermore, Haller states that one always needs to be aware of the ties and relationship of the informant towards the issue. He maintains that the more openly a source talks about its own motives the more it can be trusted. In addition, the more authentic and competent an informant is (eyewitness or expert) the more reliable the information will be.\textsuperscript{54} There are different alternative resources to reduce one’s dependence on one specific source. Independent experts, such as professors, associations or public authorities, serve as trustworthy suppliers of information.\textsuperscript{55}

Besides considering the sources, collecting and checking facts is vital for the claim of validity. The famous wh-questions have become institutionalized in the researcher’s work and function as a methodical framework. Kindel calls them the fundamental devices for a journalist and include the following:\textsuperscript{56} Who is the subject of the event? What happened? This question constitutes the actual story. When did it happen? Where did it happen? In which way / how did it happen? Why did it happen? The different authors added the seventh ‘W’ as the one of ‘what source is the informant’?\textsuperscript{57}

The first four ‘Whs’ refer to the basic facts and circumstances of an event. They are the primary questions to be answered. Through the collecting and comparing of information from many different sources the basic truth on who did what, where and when should be answered. The matter of how and why is more complex and it is important that testable facts have been checked in advance. For Haller, the questions of how and why are located on the level of analysis, whereas the first four ‘Whs’ refer to the level of facts and circumstances.\textsuperscript{58} Kindel remarks that the question of why can most often not be answered thoroughly, but serves as additional explanations, connections and backgrounds.\textsuperscript{59}

The inquiry on how and why is the link to the third claim, comprehensibility. Often the basic facts of an event are enough for publication. Answering the questions on how and why analyzes the happenings, reflects on the connections and causes, and tries to interpret the actions, responsibilities and motives. The objective of this claim lies in making the issue understandable and comprehensible.

\textsuperscript{54} Haller (2004b): p. 89.
\textsuperscript{55} Wolff (2006): p. 31 et seq.
\textsuperscript{56} Kindel (2004): p.117.
\textsuperscript{58} Haller (2004b): p. 61.
\textsuperscript{59} Kindel (2004): p. 117.
Wolff states that the research of a news item is then complete when all wh-questions have been answered, when the facts and chronology as well as the actors and responsibilities of the event can be precisely described and presented in detail.60 To achieve this master thesis’ objective of finding tools and online sources for a professional and reliable research, these wh-questions will constitute the main point of reference.

It is clear that media organizations and journalists work within certain constraints such as financial or time resources making a perfect and complete research in daily work practically impossible. Elizabeth Hart claims that “finding information for media is an exercise in targeting the appropriate sources so as to make best use of your time and energy.”61 The process of information gathering needs to be fast and expedient. The Internet is a new research tool that has solved many problems linked to efficient research such as finding information quickly and without geographical borders. Then again, the Internet is the cause of many structural changes within the media system that entail financial consequences, which logically affect journalists’ work in general and specifically in their research.

3. The Internet as a New Mass Medium

3.1. Structural Changes in the Media System Caused by the Internet

According to Christoph Neuberger, technically spoken, media enable the generation, the reproduction, the storage, the diffusion and the reception of messages.62 The Internet has fundamentally changed the structures within this mediation process. In comparison to the traditional mass media the Internet provides a much greater set of possibilities. Its technical potential affects the following three dimensions:

First of all, the social dimension: The Internet integrates the different types of communicators. Journalists are not solely facing a one-to-many form of communication. Now readers may also become active communicators themselves. Communication on the Web can be one-to-one, one-to-many, many-to-many. It can be private or public, one- or two-sided, participatory and interactive. This phenomenon is what Neuberger calls the

---

flexible switching between different roles. Also Jürgen Habermas claims that Internet communication has compensated for the asymmetric and anonymous character of the traditional mass media by enabling interactivity between readers and writers. Jane Singer emphasizes how, in addition to the flexibility of changing roles, all communication and communicators are also interlinked and connected:

“Message producers and message consumers are interchangeable and inextricably linked. You may be a producer one minute, a consumer the next – or if you’re a good multi-tasker, both simultaneously. Moreover, you are always connected to others who also are occupying both roles.”

Amy Schmitz Weiss and Vanessa de Macedo Higgins Joyce conclude that the interactive characteristic of the medium facilitates the blurring of the roles of the producer and receiver, therefore shortening the social space between communicators, between journalists and their audiences specifically.

As the second dimension, which is effected by the technical potential of the Internet, Neuberger names the “channel dimension” or the dimension of signs and symbols. This dimension is affected by the Internet’s ability to converge the different hitherto existing media forms into one medium. Text, photography, audio, video and animations can all be combined on the World Wide Web’s platforms. So far the outputs of the different traditional media types such as articles or TV shows had been clearly separable from each other. The Internet overcomes these boundaries and connects them through the technical convergence.

The third effect concerns the dimension of time and space. Information research and production has been drastically accelerated through the new technical possibilities and have become decentrally organized. According to Neuberger the Internet combines a fast and constant dissemination in respect to the possibilities of data storage. The old can be connected with the new and, in addition, it is accessible within seconds from anywhere on the globe. Especially with the new generation of mobile phones, which are developing to be technical all-rounders, information can be distributed and received at any place, at any time. For online-journalism this means the evolution of a 24-hour news cycle and constant updating. Elizabeth Bird indicates that the news habit, especially among young

---

63 Ibid. p. 23.
people, has completely changed. “Today, research reveals that younger people (and many older too) essentially consume news in a steady stream of information bites.”

Neuberger states that the Internet is still a medium evolving and especially through its multitude of options it is necessary for it to be formed. The decentralization and the possibility of participation lead to the fact that many actors at once are involved in this process, which results in a great need for coordination. In the Web self-monitoring becomes vital and, according to Neuberger, more important than any other external control, be it political, judicial or economical.

Based on this fact one needs to distinguish between the technical possibilities and the present-day practices that have developed in the use of the Internet.

Natalie Fenton also remarks that the Internet brings new ways of collecting and reporting information. The new medium has changed the nature of news productivity. “This new journalism is open to novices, lacks editorial control, can stem from anywhere (not just the newsroom), involves new writing techniques, functions in a network with fragmented audiences, is delivered at great speed, and is open and iterative.”

The following characteristics together created a new brand of journalism in the Internet-Age: Speed and space, multiplicity and polycentrality, interactivity and participation. Speed may mean more news. However, its downside is to be found in the increasing emphasis on immediacy. Fenton indicates that news organizations are encouraged to release and update stories before the usual checks for journalistic integrity have taken place. In addition, the theoretical possibility of more news through speed and also through participation including a multiplicity of views and voices, does not always translate into diversity. The increased quantity leads to more of the same.

In their study Joanna Redden and Tamara Witschge seem to have found the same results. They detected a common approach of updating and the reusing of material, which leads to homogeneity of online news within and between mainstream news sites.

---


72 Ibid. p. 6.

73 Cp. Ibid. p. 7 et seq.

It is important to remark that these new developments in the field of journalism do not solely arise from the technological innovations themselves. Fenton emphasizes the fact that the ethos and vocation of journalism is embedded in a relationship with democracy and its practice, that is, in a history of commercial practice, regulatory control and technological innovation. Journalism then depends on the working environment, which is shaped by economic, social, political and technological factors. The latter form a dense inter-meshing of commercial, ethical, regulatory and cultural components. “News is also what journalists make it.”

The nature of change is not to be attributed to technology alone but rather to the convergence of many forces. “New media emerge by merging existing socio-material infrastructures with novel technical capabilities […] this evolution is influenced by a combination of historical conditions, local contingencies and process dynamics”. One must speak of a technical novelty rather than of technical determinism. Eugenia Mitchelstein and Pablo Boczkowski also conclude that research has highlighted the importance of organizational and institutional contexts that shape the dynamics of innovations in different journalistic settings.

As much as the Internet is a decisive driver of change in the news environment, the economic and labor dynamics are also a critical part in the process of technological change. The Internet and the economic circumstances of a news organization are linked in a symbiotic way. The new medium then becomes a mixed blessing. It has simplified many procedures such as information gathering, sharing and distributing through its ability to overcome the boundaries of time and space. Costs may have been cut through the free information flow on the Web. However, these exact features have set the traditional business model for delivering news in crisis. Neuberger summarizes that although the Internet technically offers many new possibilities, economically it erodes a professional journalism’s means of existence.

Des Freedman remarks that established news organizations see their audiences decline. Younger audiences are deserting them for immediacy and interactivity on the Internet. He proclaims a tremendous growth in the number of news outlets, which heightens competition and the pressure on the single news

organization. The emergence of a 24-hour news cycle and the popularization of online platforms have contributed to a volatile and unstable environment, which journalists now see themselves confronted with. Advertisers are increasingly attracted by the possibilities of targeting their audiences more accurately on the World Wide Web. However, online advertising does not guarantee substantial profits for the news organization. In conclusion: The economic conditions of online journalism are precarious. This fact is not only fortified by the global financial crisis starting in 2008, but also through the common approach among the audience of being able to consume everything on the Web for free. The vast majority of news is freely available. For Freedman the central economic factor about online news is the reluctance to pay for news content. Also Mitchelstein and Boczkowski mention different studies revealing that users do not seem prepared to pay for content. A business model for online news remains very much at an experimental stage. These circumstances and conditions have influenced the development of online journalism.

“The internet has the potential to expand the diversity of news sources, to improve the quality and breadth of news coverage, and to deepen the interaction between news providers and their audiences. Yet, given today’s harsh economic circumstances, the internet has instead contributed to a possibility that the news of the future is going to be sustained by a declining number of specialist news organizations, a growing band of generalist news and information businesses, and a handful of parasitical aggregators supplemented by an army of contributors working for free. Market logic, in this scenario, is set to prevail over news logic.”

Elisia Cohen remarks how in a market-driven journalism viewers and readers are transformed into customers and news into products. On a meso-level the market logic then becomes the mold for routines of discovering newsworthy events and of selection. Where professional norms encourage news directors and newspaper editors to produce independent, fair, and complete coverage of public issues, coverage would center on the least expensively gathered information likely to generate the largest audience if purely economic norms were to prevail. Neuberger states that, among journalists, it is often criticized that the publishers lack the willingness to invest, but try to raise the number of clicks by any means. Organizations are confronted with cutting costs but increasing

82 Cp. Ibid. p. 37.
84 Freedman (2010): p. 44.
87 Ibid. p. 50.
89 Cp. Ibid. p. 535 et seq.
productivity simultaneously. Cross-media integration has heightened the journalists’ workloads, thus again boosting the rapid repackaging of largely unchecked second-hand material.\footnote{Cp. Freedman (2010): p. 41.} Freedman concludes that the threat to journalism lies in under-investment. Bird’s example underlines this fear. She writes that the publisher of the online Pasadena Now fired his entire Pasadena staff, and “now outsources news coverage to writers in India who write news and features, using email, press releases, the Web and live video streaming from a cell phone at City Hall.”\footnote{Bird (2009): p. 294.} Freedman’s conclusion at this point seems more than viable: “There are no short cuts: the future of news […] depends on imagination and independence but, above all, on investment – in technology, in resources and, especially, in journalists themselves.”\footnote{Freedman (2010): p. 50.}

\section*{3.2. The Role of the Online Journalist}

In pre-Internet times the media environment was a one-way information flow. At that time journalists delivered news about what was happening in ‘the world out there’; the public occupied merely a passive role in the information process. The World Wide Web has fundamentally changed this situation. Participation and interaction have become the heartbeat of a new form of public communication, which can be described as a complex network of information. A network is not linear and Singer concludes: “Information in this environment flows not only through the journalists but also, continuously, around them.”\footnote{Singer (2009): p. 64.}

There is no external space or role apart from the journalist anymore. All participants and communicators function as citizens of the network and they all contribute to it. “Because in a networked world, there no longer is the ‘journalist’, ‘audience’, and ‘source’. There is only ‘us’.”\footnote{Ibid. p. 75.}

Neuberger also remarks on the fact that now all single individuals and laymen can connect through the Internet and that it is this feature of the Internet that is mainly being referred to when the Web 2.0 is spoken of.\footnote{Neuberger (2009): p. 39 et seq.} Communication following publications can now be readdressed to the journalists themselves. Neuberger describes this figuratively saying that now there is opposing traffic on a former one-way street.\footnote{Ibid. p. 40.} Based on this notion of a network, Chris Anderson’s theory of “The Long Tail” contradicts all theories of a
fragmented public, saying that it is the ability of the Internet to connect and integrate the different public spheres. Smaller and specialized publics have existed before the broad usage of the World Wide Web, but now are able to be seen and heard.98 “The Long Tail” describes a graph showing how a few popular items and distributors on the Internet receive the most attention and how a long tail of other products and information line up to be noticed but only acquire little attention.

Although the popularity rests within a small number of dominating items, “The Long Tail” shows how a major part of the population definitely adds to the Web content. This means that many niche products might not attract the most people but can easily be found on the Web. User’s self-monitoring and interests now define what is to be found as information. It is no longer the journalist acting as a gatekeeper. Through effective networking and cross-linking smaller and unknown publics can now attract attention in an easier way.99 Anderson assumes that consumers will start dislocating into niches because one might be more content with what he finds there and, therefore, will consume more of it.100 Moreover, Peter Lee-Wright indicates the new motions in public communication and the emphasis on the user himself as well. He claims that the Internet with its endless options and direct sources all give the consumer the choices previously made by the news editor.101 The pressure for journalists to involve the public has grown. Also Fenton remarks on the

---

99 Cp. Ibid.
growth of the impact of the readers and the resulting demystification of journalism.\textsuperscript{102} Journalists have lost their privileged position as a gatekeeper. In the universe of traditional media, journalists were central to the flow of information by not letting misinformation or disinformation pass through the gate. This role is based on ethics and needs to be redefined in linked and participative communication.\textsuperscript{103} The line between news producers and news audiences has become blurred. Bird maintains that “[t]he new digital environment has jolted traditional journalism out of its conservative complacency; news operations are much more responsive to their empowered and engaged audiences.”\textsuperscript{104} There seems to be a new type of journalism developing in which traditional roles of journalism such as gatekeeping and agenda-setting are becoming less important. Speed and immediacy, hypertext and multimedia are the new premises according to Mitchelstein and Boczkowski. News workers are seen to be engaged in ‘gate opening’ through practices that foster user participation rather than the kind of content selection associated with the traditional gatekeeping tasks.\textsuperscript{105} Despite the open and participative online communication, it is argued that the role of a gatekeeper remains viable, but in a different form. Singer describes this new type of gatekeeper as one that has to do with sense-making by helping people understand, interpret, and use information.\textsuperscript{106} The Internet communication entails that the recipients need to sort and interpret the found information themselves. Neuberger states that the openness, the flexibility and “revisability” of Web content overly challenges the users both in a quantitative and qualitative way.\textsuperscript{107} The lack of time and competence of Web users make it hard for any online communicator to attract attention. Participation inherits the paradox that the more people that take part in online communication the harder it is for a single actor to be taken notice of and maybe be responded to. It has become difficult to generate attention and credibility on the World Wide Web. Neuberger concludes that it is hard to anticipate and recognize the effects of one’s contributions and topics in this open and global medium. Organizations have seen themselves forced to systematically observe what is being said and done in the vast information network.\textsuperscript{108} The consequence lies in the Web users need of orientation and of mediation. Lee-Wright confirms Neuberger’s theory

\textsuperscript{105} Mitchelstein / Boczowski (2009): p. 571.
\textsuperscript{107} Neuberger (2009): p. 49 et seq.
\textsuperscript{108} Cp. ibid. p. 50 et seq.
indicating that “everyone needs some guidance, some authoritative aid in reading their world.”

In this new situation Singer sees the fundamental ethical principles of journalists remaining vital, but for different reasons than in the past. Journalists’ primary loyalty is still to the public and their goal is still to provide credible information. However, they need to recognize that they are not the only ones providing it. In the environment of online communication journalists are no longer faceless professionals, but real human beings within the network. Singer warns that trying to stand apart from an external public in order to achieve the idea of journalistic objectivity implies the danger of isolation and irrelevance. The relationship between the journalist and the people is no longer solely based on the delivering of information to a mass, anonymous civic entity, but on relating to other individuals in a more human-to-human context. In traditional media environment the audience was asked to trust the journalists more or less blindly. Now, accountability and credibility need to be demonstrated. As before, telling the truth remains vital, but not solely because of wanting the public to know the truth, but rather because it is fundamental to any social relationship. “Thus we are moving toward a situation where authenticity or credibility becomes more a matter of the relationship that an individual establishes with his or her readers than with the institutional role of the media organization.”

Being able to disseminate trustworthy information implies finding it as well. The technological innovations have naturally changed the possibilities of journalistic research. The following pages shall reveal the nature of online research.

### 3.3. The Meaning of Research in the Online Environment

According to Mitchelstein and Boczkowski, there are four aspects of changes in journalistic practices, through the emergence of the Internet, that have been central to the scholarly research of online news: Modification in editorial workflow, alterations in news-gathering practices, acceleration of temporal patterns of content production, and the convergence of print, broadcast, and online operations. They see increased pressures on journalists to carry out multiple tasks at once and to combine news-gathering and storytelling techniques in different media formats. In her study Angela Phillips interviewed the

---

111 Ibid. p. 72.
112 Ibid. p. 67.
authors of 89 news stories in a range of British elite, daily and national papers about their sourcing of each story. She identified the same developments as Mitchelstein and Boczkowski, adding that training seemed to be concentrating more on the techniques of delivery rather than on routinely used techniques of research. This is quite discomforting considering the fact that through the Internet journalists are faced with a great deal more information that needs to be tackled. “Being a […] journalist is rather like standing in the middle of a hurricane trying to pick out twigs. You’re […] constantly looking around to see what’s significant.” This quote of an interviewed specialist reporter shows how journalists in the Web environment are trying to deal with a blizzard of information. Mitchelstein and Boczkowski also use the term ‘news cyclone’ to describe what has happened to the former news cycle. This development is mainly rooted in the acceleration of temporal patterns and workflows. To them, acceleration has become a key feature of the contemporary news industry. Christoph Neuberger, Christian Nuernbergk and Melanie Rischke name two possible ways for a journalist to receive information: actively and passively. Research then is the active acquisition of information. They as well note that the time given to do active research has shrunk through the economical and technical influences. In this situation online research has quickly become of great value within the field of journalism.

The Internet offers new possibilities in doing research. Neuberger et al. mention the following: It has become easier to access information of foreign countries, the amount and the variety of sources has augmented, participative sourcing of information has increased and has become easier for citizens and unknown interest groups to publish information and to be heard as a source. Ralf Blittkowsky describes the Internet as a mingle-mangle of information and products that are offered. He continues by saying that the essential part in online research is to report novelties and to do this quickly.

According to Phillips the Web has massively increased the efficiency of fact-finding. “Much of the material that in the past would have come via photocopies in brown envelopes, can now be found

115 Ibid. p. 94.
118 Ibid. p. 297.
published on, or hinted at, in forums or Web sites of official or semi-official organizations.”

Aside from all the useful and informative sources, a great part of online information is also redundant. Following Blittkowsky, all publications on the Web need to be treated with reservation in regard to their credibility. He claims that it is the central task of an online researcher to compose a set of trustworthy online sources. Especially user-generated content is found to be too untrustworthy and time-consuming. And, according to Phillips’ study, journalists feel that it has rarely produced genuinely new information. Online research has often been accused of using unreliable sources and can only be seen as a useful help for pre-research work. However, considering the fact that it is the goal of any type of research to find new knowledge and information, offline means as well as online means can be useful to accomplish this task.

Neuberger et al. differentiate between the use of online sources and the use of online search aids. Where sources can be actors having relevant information or documents providing such information, search aids direct the journalist to the relevant and needed source. To find the right sources and to efficiently use search aids journalists must have a sound knowledge of IT-technologies and their possibilities. However, according to Phillips, very few journalists really know how to use Web-searching tools.

The Internet might be creating better opportunities for checking material, finding alternative sources and improving the independence and, therefore, the democratic and cultural relevance of news outlets. Yet the acceleration in the news production process caused by the technological innovations result in a lack of time for research and also for further education on Web technologies and their usage. Every story goes online within minutes, mainly because of the intense pressure to not miss anything that has appeared on a rival Web site. Online journalists constantly check other news outlets to make sure they themselves have not fallen behind. These constraints lead to much homogeneity throughout the different news sites and create agency dependent and secondhand journalism. Many stories are solely being slightly reangled with minimal verification. Phillips detects a ‘cannibalization’ of the information and states that the Internet is actually

124 Ibid., p. 31.
126 Cp. ibid. p. 96.
narrowing the perspective of many reporters.\textsuperscript{127} The speed of online news entails the danger of undermining the very point of reporting and the individuality of the different outlets. “Online it is difficult to maintain ‘difference’ because stories will simply be stolen by every other outlet.”\textsuperscript{128}

The little time left for verification and individual research has also forced journalists to use safe sources, meaning known sources. The Internet might offer a blizzard of information, but when it comes to finding credible information quickly, the reporters prioritize known and, therefore, ‘safe’ organizations in order to control the flow. Phillips concludes that, in the end, this narrowing of source relationships results in a growing power of journalists versus the power of other citizens, even though citizens now have an easier access to the public themselves.

Time and technological knowledge seem to be vital to any online research. One may deduce that especially because of the little time available, technological knowledge becomes even more important. The next chapter reveals how journalists in Swiss online news outlets use the Web for their research.

4. The Use of the Internet as a Research Tool in Swiss Online Editorial Offices

4.1. The Research Questions

To an online journalist, the most important and profound effect of Web communication lies within the acceleration of the news production. The Internet has come to be a mixed blessing, making news gathering easier, also by overcoming geographical boundaries, but eroding the professional journalism’s means of existence. The vast majority of online news is free of charge; a functioning business model to cover the production costs is still in its infancy.

The acceleration and simplification of the flow of information means the news outlet has to produce on a 24 hour news cycle. Constant comparing with other news sites, ensuring itself not to have fallen behind, sets journalists under an enormous time pressure. Together with the economic constraints leading to increased workloads, online journalists face difficult conditions to produce qualitative good journalism.

\textsuperscript{127} Cp. ibid. p. 99.  
\textsuperscript{128} Ibid. p. 101.
The Internet as a research tool can be of great help in such an environment, considering the vastness of information available and the speed in finding it. Yet, technology is fast growing and new online features and tools seem to be developing in short time periods. Considering the fact that online journalists already have a full-time job, it is questionable if they find enough time to deal with all the possibilities of the Web. Research is vital to any professional journalism. Naturally, knowing one’s research tools should be a must. Therefore, it was of interest to analyze today’s use of the Internet as a research tool in online editorial offices. The main research question to be answered was the following: What is the current situation in the use of the Internet for research in Swiss online editorial offices? More detailed that is: What is the common knowledge about the research possibilities of the Internet? How is the Internet being used for research purposes? What tools of the Web 2.0 are being used and how? What are the desires and expectations towards the Internet and upcoming technological innovations in regard to journalistic research?

The next pages introduce the method, which was used to answer these questions.

4.2. The Method

The objective of this research done in the Swiss online news outlets was to acquire as much information as possible to be able to construct a clear picture describing the momentary use of the Internet as a research tool. To achieve this goal, qualitative interviews were chosen as the method. It is obvious that the samples and the resulting information cannot be seen as representative of the whole journalistic field, but as an indicator of the common and also different ways in which online journalists use the Web for their research.

Even though online research may be relevant to all sorts of journalists, the sample for this work has been limited to online journalists for the following reasons: Online journalists are confronted with the 24 hour news cycle as well as with the acceleration in the whole news production processes to a much greater degree than journalists of any other medium. They need to publish as fast as possible, yet credibly and truthfully. As mentioned, the economic constraints within journalism in general, but especially in online journalism, result in higher workloads and fewer resources. Online journalists need to work very efficiently, yet wanting to meet the standards of journalistic professionalism. This would entail knowing
one’s research tools. The Internet, being a huge pool of information that it is, may help in finding information quickly, which is of vital importance to any online journalist.

In accordance with the theory on qualitative interviews, written by Jan Kruse, the sample of the different online editorial offices was a consciously made decision based on the objective of a maximal variation. This means that the sample was to consist of online news outlets of different sizes, different regions, different journalism types and different media organizations. Geographically the sample was limited to the German part of Switzerland. This decision was simply due to language. Maximal variation here was sought through supra-regional and regional outlets in the different parts of the German speaking part of the country. To integrate the different journalism types, members working for different media outlets, such as boulevard media, dominant and defining media as well as private or publicly owned media, were interviewed. Another variation to be covered was to combine the different media types, such as newspaper, free press, radio, television, news agency or purely online news, which operate an online news portal. All outlets needed to offer daily news of national relevance. The person interviewed ideally needed to be a journalist working for the national news and had to have a number of years of journalistic online experience. News of national relevance was chosen for the following reasons: Firstly, national news is of interest to many people and to many news organizations, resulting in competition between the different portals. Secondly, information of national relevance might not be found right outside on the doorstep, making the Internet a required tool for journalism in this field.

To find the online news organizations needed, lists of all the existing news media in the German speaking part of Switzerland were used to filter out the ones fulfilling the requirements of this research. The lists of radio and television stations were easily found on the government’s Web page. To have all the daily newspapers a list of WEMF AG, a Swiss institution for research on printed media, was used. This directory consisted of daily and weekly newspapers, of which the weekly newspapers were manually deleted. To

identify purely online news portals, a list of NET-Metrix, a neutral organization publishing statistics and data about the use of the Internet in Switzerland, was used.132 The online media landscape in Switzerland is rather small and relatively new, making it quite difficult to ideally fulfill the requirements. In the end, nine editorial offices were chosen, of which mainly online journalists or news desk coordinators with experience in doing the journalistic online work were interviewed. Aside from journalists, two interviews were also done with media coaches working in the field of multimedia journalism and in the field of online research. The objective of these interviews was to detect what ongoing online journalists and newcomers to the job already know regarding online research.

All of the interviews lasted between twenty-five and forty-five minutes. They were recorded and transcribed to be evaluated. As Jochen Gläser and Grit Laudel mention in their theory on expert interviews and qualitative content analysis, the rules and the precision of the transcript depend on the objective of the research.133 In this case the interviewee’s behavior was not of relevance, as the main intent was to gather the most information about the journalists’ use of the Internet as possible. For this reason it was not considered important to mark all the different hesitations or pauses in speech of the person being interviewed, but to note everything said in regard to content. The interrogations were held in Swiss German; the transcripts naturally were written in German. The interviews were done following an outline of questions set up according to the research questions and according to the theories about journalistic research processed in chapter 2.134 All interviewees received this set of questions a few days before the interview to have a rough guideline of the topics to be discussed. In this way, all the participants were able to think ahead about these issues and were able to respond with greater knowledge and also with more precision to the questions asked.

The evaluation of the transcripts was done by means of categories set up in relevance to the research questions. Gläser and Laudel call this type of analysis extraction, meaning that all the relevant information found in the transcripts is being extracted into the system of categories to be interpreted.135 The next chapter will present the results of this procedure.

---

134 See Appendix 1: Interview Questions Online Journalists. P. I. And: See appendix 2: Interview Questions Media Coaches. P. II.
135 Ibid. p. 199 et seq.
4.3. The Results

Based on the communication theories about research and on the questions asked, five main categories, some consisting of sub-categories, were implemented for the analysis: Wh-questions, tools of the Web 2.0, search engines, reliability as well as credibility, and lastly technology.

The category of ‘Wh-questions’ involves all the replies received as to where online journalists seek their information in answer to the questions concerning the ‘who’, ‘when’, ‘what’ and ‘where’ of a story. Although in the interview as well as in the evaluation the ‘who’ was separated from the other more factual matters of ‘when’, ‘what’ and ‘where’, it appeared to be difficult to split these issues. The second category includes the different tools of the Web 2.0, such as social networks, microblogs, blogs, wikis and tagging. Search engines define the third group including both the use of and suggestions of a different form of visualization than the current lists. The last category, named technology, was again divided into two sub-categories; one describing the attitudes towards an increased application of machine-generated decisions on relevance, the other covering the wishes of the online journalists for future innovations that would be useful in their research work.

4.3.1. Where Online Journalists Search

The input of an article in online news is mostly the news agencies’ releases. For many, using the Internet is mainly to gain background information or additional information, which goes beyond the facts already known through the agency’s news feed. Only a few journalists working in specialized fields can generate stories themselves, choosing the topic and doing the research on their own. Mostly the time pressure in the online news business is too immense and the resources too meager to have more independent reporters. News agencies seem to have a much more conservative attitude towards the Internet and its use than the other news outlets. Their way of research is much more conventional. The Web is used to attain types of information that have been used in pre-Internet times already, such as press releases, governmental or party information, but that now can easily and quickly be found online. Otherwise, the Internet is described as an accompanying medium, which is used daily to inform oneself, to receive background information, to get a picture of something and to be able to put issues into context. Seldom is it used to find the hard facts or even quotes. There is only one exception one can make: economic and market exchange information, of course, is being followed over the Web.
Then, for many, there is normally no need to search for information concerning the subject of an article online. These facts are usually already included in the input, be it from a news agency or from somewhere else. The news agency itself has received all the information about the ‘who’ of a story, mostly by being on-site of the happening or by being in direct telephone or email contact with the involved person. If indeed there is a need to find out more about a person or an organization, different sites and portals, such as the following, were named depending on the field of work. Databases of one’s own organization, especially of organizations that have already been established over many years’ time, were often mentioned as being a very important tool in research. The second database used by most institutions was the Swiss Media Database (SMD), which was founded in 1996 and archives articles of most newspapers in Switzerland. Not a database but online, other newspaper articles, including international ones, are regularly used to do research. They can also be found through Google News. About politicians or information around government issues the government’s Web site www.admin.ch or the site www.parlament.ch are the first places to be searched. For people in the private sector the Web page www.moneyhouse.ch was mentioned. If organizations are sought, the commercial register is where to look. If the subject really is one sole person about whom not much is known, getting into direct contact with him or her is attempted. This may happen by finding his or her Web page or finding his name and phone number through Google and calling the person. Doing telephone research was repeatedly mentioned and does not seem to vanish in times of Internet communication. The Web is then of help to find the relevant telephone numbers via Google or online telephone books such as www.tel.search.ch. Finally, not often, but still stated, was Facebook. The role of social networks will be discussed in chapter 4.3.3, together with Wikipedia that was also named frequently as a good starting point in one’s research process.

In regard to the factual matters of ‘what’, ‘when’ and ‘where’, mainly the same online sources were relevant to the online journalist’s work. One input seemed interesting, stating that Web sites which are local to the event to be reported were used. For example, if the issue to be researched on happened in New York, Web sites of New Yorker newspapers and other New Yorker sites would be looked at. However, what is remarkable is the outstanding and dominant role of Google for all sorts of questions. As one journalist answered the question about how search engines are being used: “Wie? Also wir brauchen sie vor allem dauernd. Also es ist so der erste Reflex, wenn wir etwas nicht wissen, geben
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

wir es mal bei Google ein.” To all, Google is the access and starting point of almost any research done on the Web. The media coaches, although mentioning that many journalists do not know how to use it properly, also recommend gaining access to information through a search engine. In addition to the already listed options of online sources, Facebook and Xing, Twitter, YouTube, specialist and biographical sites, blogs, Google Scholar, Google Books, virtual libraries and discussion forums were also named by the coaches. The variety of given possibilities seems to be even bigger. However, the access and the credibility of the different sources differ and journalists are warned not to solely use the information found on one of these portals.

4.3.2. The Art of Using Google

Google is central to online research and is a helpful tool. However, it is also necessary that one knows how to use it. As one coach remarked: “Im Google hat man einfach das riesige Universum und findet dann vielleicht nicht immer gleich das, was man wissen will, aber findet dann eben auch viel Schrott.” In addition, ninety percent of the people using Google mainly use the search bar and forget that there are other possibilities for advanced searches. Another coach maintains that the average journalist does not know more than ten percent of all the possible tricks that could be used in doing a Google query. One trick though that does seem to be somewhat common is searching a Web site via Google by narrowing the search with the word ‘site:’. Journalists and media coaches have stated that this is often more successful than using the Web site’s own search bar. It is important to know the syntax of a search engine and to think of a set of relevant keywords before starting the query. The more exact, the more specific and the better combined one uses the keywords, the more relevant and qualitatively better the search results will be. Journalists also stated that the keywords used are very essential to the outcome of a search. The more one knows what he is searching for, the easier it will be to find it through the search engine. It is hard to find relevant information if there are only few hints concerning the relevant topic or if the topic is really complex.

---

136 See Appendix 3: Confidential Interviews. P. III. Quote in English: „How? Well, above all, we use them constantly. It’s our first reflex; if we don’t know something, we google it.”
137 See Appendix 3: Confidential Interviews. P. III. Quote in English: „In Google you find a huge universe, and yet you might not immediately find what you actually wanted to know, but you do also find a lot of junk.”
138 See Appendix 3: Confidential Interviews. P. III.
Knowing that commonly only the first few pages of a query’s results are actually looked at, both media coaches recommend using more than one search engine, or even using a meta search engine in order to heighten the chance of really finding the needed and relevant information. However, in the field, meta search engines do not seem to be of great relevance. One journalist said that his hits in Google are more useful than the ones of a meta search engine. This is because, according to him, meta search engines also explore sites and show results that are dubious and suspicious. Therefore, he feels they are not of great use and subsequently cause him more work in selecting the relevant hits. Using more than one search engine – meaning using any other search engine other than Google – is seldom the case. Only a few stated that they do use Yahoo! or Bing for their research, even though they might have found some good results through Google. The only reason to use a second search engine is mainly in the case of not finding the needed information on Google in the first place. In any other case, the time pressure is too big to be able to check other search engines so as to possibly get a different access to the issue of relevance. In some cases, if the query has not been successful, no more search engines are used for further research, but other ways are sought to find the needed information. This would mainly entail a Web site where it is found feasible to find the necessary material. Only one journalist remarked that sometimes one needs to have the patience to also check the results up to hit 39 in order to succeed. Current events can especially influence the ranking of hits on an issue and may force the journalist to check the results a few pages further down the line.  

The fact that Google is the only search engine used is generally not seen as problematic, because Google is either one way of many within the research process or the needed information is viewed as superficial. Some state that Google is the first step in doing pre-research and in gaining a rough overview of the relevant information. Still, all sorts of information are searched for through Google, even if not solely. Often Google can be seen as a link to all the different Web sites and online sources mentioned above, that are used to answer the central Wh-questions on an issue.

139 See Appendix 3: Confidential Interviews. P. III.
4.3.3. Searching Through the Web 2.0

4.3.3.1. Social Networks
The use and integration of information found on social networks vary. To many, social networks and especially Facebook are irrelevant for journalistic research because of the lack of reliability and credibility. Social networks are viewed as prone to manipulation. Using information published on Facebook or other social networks would entail further research to verify the found facts. If this verification through other sources has taken place, some journalists then might integrate Facebook-information into their article. It is noted that stories entailing such additional information are often very popular and well read.

Having interviewed mainly journalists working for the national news this skeptic attitude might not be surprising. It is more in the field of boulevard media that Facebook is used as a source of information about a person and about a person’s ties and interests. Especially pictures are easily found on Facebook and are free to use, since the person him or herself has published them on the Web. The different Facebook entries vary because of the different privacy settings, which is an additional barrier in using this tool as a source. One journalist states that on profiles of celebrities, that are mostly accessible and managed by professionals, he seldom finds genuinely new and unknown information.\(^{140}\) The fact that professionals are cultivating their profile in a professional manner then again shows that the importance of Facebook is increasing. Swiss International Airlines, during the air space ban over Europe due to the cloud of volcanic ash coming from Iceland, updated its customers very effectively over Facebook, which possibly indicates that the social media is becoming more important in public communication. Even for the journalists not quoting or using Facebook-material, social networks do function as a source of input for ideas or stories. The Web in general, including social networks, is monitored to know what is being talked about, what is central in the different communities. Created groups may evolve to being almost a type of non-profit organization in a very non-institutional way. One journalist remarked that in such a case it could be possible to try to contact someone via Facebook in order to talk to that person directly over the telephone later on. Generally speaking, topics of the virtual world of communication may find their way into journalism if they are of relevance and interest.

Neither Xing nor MySpace seem to be used in any way. None the less, the media coaches see a great potential in Xing if one is registered, simply because Xing is more serious and

---

\(^{140}\) See Appendix 3: Confidential Interviews. P. III.
reliable than Facebook and can reveal interesting connections and information. Facebook is seen as a possible additional source, but with little impact due to the lack of reliability and the problem of accessing private information.

4.3.3.2. Microblogs
The prevailing example of microblogging is by way of Twitter. Twitter seems to be becoming more and more relevant, yet still quite unknown to some journalists and used with varying intensity among all the online journalists. It can be said that the post-election conflicts in Iran in 2009 helped Twitter to a type of break-through within the media. Many journalists have mentioned this event as causing them to really be confronted with the new tool for the first time. Twitter was the only open channel left to access information from that Middle Eastern country. As with social networks, here too, the problem seen in using Twitter is the lack of reliability and the potential for manipulation. Some news organizations do not integrate Twitter information because of a lack of time and resources to verify the facts. Other journalists do not use Twitter because the information seems to be superficial and irrelevant. Nevertheless, mentioning the fact that a Tweet is often only a link to more profound information, this tool did start to sound interesting.

If Twitter is used, it is mainly when catastrophic incidents happen such as the oil spill in the gulf, the ditching of an airbus on the Hudson River, the bomb attack in Moscow or the riots in Iran. Information, quotes and pictures are available in no time and can be directly integrated into the article. Speed is a vital imperative in online journalism and Twitter is a tool enabling the dissemination of really recent and up-to-date information from anywhere in the world. In situations when the state of knowledge of what happened is still vague and chaotic, some journalists find it acceptable to integrate Twitter information and pictures when doing a journalistic check of plausibility. Again the field of boulevard favors Twitter because of its possibility to follow celebrities and benefit from information and pictures. Still, many are cautious when using Twitter as it is seen as a great grapevine. One of the media coaches feels Twitter is the wrong tool to find hard facts for serious journalism although recognizing the opportunity of obtaining inputs. The second media coach, on the other hand, sees great potential in Twitter especially for specialists of a certain field, nevertheless mentioning the necessity of already having certain knowledge to be able to estimate the quality of the information found. She maintained that, for example during the
protests in Iran, only an Iran-specialist would have been truly capable enough to do so.141 Still, having some knowledge and experience and following the right people in a special field of interest, Twitter is then seen as an excellent tool by this media coach.

4.3.3.3. Blogs
There is a great difference among the individual habits of the journalists in using blogs as a source. To some, blogs play no role at all because of the fact that blogs display opinions and can be manipulated. Taking information from blogs is seen to be really insecure. One journalist even said that blogs, to her, are a second or even third class source.142 For another journalist, blogs seem very interesting and she maintained that blogs do get read by some of her colleagues for background information and input. Yet, information found on blogs needs to be double-checked concerning its seriousness. The required resources are missing, which means that found information cannot be integrated into the actual article.143 As an input though, blogs do seem to have some influence even for the very cautious and conservative journalists. Statements or pictures can reveal different aspects of official stories and force journalists to question what has happened.

Journalists using blogs point out their potential for special areas of expertise. The blogosphere is not just scanned to find relevant information, but specific blogs of specific authors are read. In the end the blog can be a basis of a story or just providing some quotes. However, they are integrated more or less intensively in the news item. The media coaches also see a great potential in blogs if they are used for a specialist’s field. Because blogs consist of personal opinions, thoughts and reflections, it is important not just to double-check them but also to find the relevant and trust-worthy blogs within the whole blogosphere. As one coach stated, of the thousands of existing blogs, maybe a few dozen are professionally being taken care of. The art in using blogs is to find these reliable blogs. There is always the danger of them being abused for PR-purposes. One journalist, who uses blogs intensively, confirms this by describing the Internet as a huge lake, in which one needs to find the interesting drops.144 Through time one can feel and figure out which authors can be trusted and which are not as reliable. Many bloggers have built up a reputation they would not want to destroy by spreading unqualified information or even lies. To know the different authors it is important to follow them and read their posts so as

---

141 See Appendix 3: Confidential Interviews. P. III.
142 See Appendix 3: Confidential Interviews. P. III.
143 See Appendix 3: Confidential Interviews. P. III.
144 See Appendix 3: Confidential Interviews. P. III.
to obtain a good picture of who is to be trusted. Some bloggers are also well-known people with solid reputations and great influence in the real world using the blog as an additional communication channel. These are viewed as very good sources and seem to be used much too little in journalism in Switzerland.

What some journalists also mentioned was the possibility of finding other sources through blogs. Following the marked links one can discover information one would never have found otherwise. One journalist maintains he acquires more knowledge than before and continues by saying that freedom of press has received a whole new meaning through blogs, thus being a fascinating development, although it is not to be idealized.145

4.3.3.4. Wikis

Asking about the importance of Wikis, and for example Wikipedia, one journalist answered: “Alle benutzen es. Nur schon deswegen ist es relevant.”146 Wikipedia is used by all journalists to gain general information on an issue. Only one journalist said that she does not use it in the first place, but only as a reassurance for what she has already written. This is mainly due to her education, where she was taught not to use Wikipedia as a source. To all other journalists, Wikipedia is quite an important tool to find basic information, to have a first general scheme of a topic and also to find more information using the listed links and sources. Only the news agency seemed to be more reserved and conservative towards Wikipedia, reasoning that they need their own known and reliable source for any sort of information and Wikipedia is then too insecure.147

The way information found on Wikipedia is used varies greatly among the interviewees. Very basic information, such as the size of a country, may be integrated directly and without any further checking. Some even go further saying that factual, objective, undisputable and insensitive information is taken from Wikipedia. It is even expressed that in fifty percent of the cases the time is missing to do a double-checking of such information and that it is used directly for the article. Another journalist mentioned that the danger and temptation of doing copy-paste from a Wikipedia-site exists, especially when working under time pressure.148 Still, many journalists state skepticism in using Wikipedia.

145 See Appendix 3: Confidential Interviews. P. III.
146 See Appendix 3: Confidential Interviews. P. III. Quote in English: “Everyone uses it. That fact alone makes it relevant.”
147 See Appendix 3: Confidential Interviews. P. III.
148 See Appendix 3: Confidential Interviews. P. III.
and would never quote it. Wikipedia is seen as a starting point or complement, never as the one and only source.

Also the media coaches see great potential in Wikipedia if used as a type of dictionary serving as a starting point for all ongoing research. According to them, studies have shown that Wikipedia is of not much lesser quality than other hard copy encyclopedias and that the controlling on Wikipedia has been greatly increased. However, hard facts need to be checked by either a second source or by experts.

As with user-generated content in general, using Wikipedia is a question of reliability. This seems to be a greater problem when mentioning the example of WikiLeaks. This wiki was discovered mainly through the published US-bombing in Iraq this past spring. The lack of reliability and trustworthiness keeps journalists from using this tool as anonymous information always could entail individual interests. The time and resources needed, to be able to seriously check and use these facts, would exceed the given capacities.

4.3.3.5. Tagging

The last tool connected to the Web 2.0, taken into account for finding the right information therein, is social bookmarking and tagging, which would corresponded to the journalists’ own way of work and organization of his or her known sources. However, tagging is not being used at all among journalists for various reasons. Some simply do not know about the tool and have never tried it. Mentioning to them that on social bookmarking sites it is possible to see what other users have bookmarked, using the same or similar keywords and tags, this tool did sound interesting to them. Others do know about the tool, but have heard that it is time-consuming and not really helpful. To some of them the links, tags and cross-references are valuable and viewed as a good possibility for research. Still, they do not use social bookmarking to organize their preferred sites, but have the relevant Web pages saved and organized in their own browser. One journalist remarks, that he knows the important and relevant sites for his work and sees no need in organizing them on such a portal. Another interviewee actually sees the benefit of organizing ones’ bookmarked Web sites through tags, but has simply avoided the effort of doing so due to the time needed. Time is also an important factor for one journalist using social bookmarking in private, but not at work. He finds it to be too time-consuming and to be able to use it would be luxurious. Another journalist confirms that starting to use new tools entails having the

149 See Appendix 3: Confidential Interviews. P. III.
150 See Appendix 3: Confidential Interviews. P. III.
time to learn new skills. He explains that the lack of time forces one to stick to the known and tried old habits. If it were possible, one would spend time on new tools.\(^{151}\)

In general, the possibility of tagging and of the social bookmarking sites is not known well enough to be an institutionalized tool in one’s own organization and structuring. However, many seemed to be interested in it, and as one interviewee said: “[…] Sachen, die arbeitserleichternd sind, werden sich zumindest zu einem gewissen Prozentsatz durchsetzen.”\(^{152}\) At the moment, tagging is not central to a journalist’s work.

4.3.4. Investigating Credibility and Trustworthiness Online

Talking about the different sources and tools on the Web, one main issue always arose when considering their value and usefulness: the question of credibility, trustworthiness or reliability. Being able to assign credibility to a source was the central indicator to the possibility of using the found information. According to a media coach, there is a whole list of checks to be done so as to be able to ascribe reliability to a source: one needs to check who the Web master of a Web site is and to investigate that person. Then, are there any references to other sources and are these testable? New, and possibly disputed, information needs to be verifiable at other places. If this is not the case, one needs to be very cautious. The next check is to find out if the information is a plagiarism or an original and to detect any interests that could be behind it. In addition, it is important to analyze how exhaustive the new-found information is and to see if there are any links to ongoing research. Web sites of dubious firms often have only one or two links connecting to another suspicious site. And, as a last check it was stated, it is important to identify how current and up-to-date the published information is. With the new content managing systems this has become more difficult. Yet, if there is no indication as to when that information was written and uploaded, one needs to be careful.\(^{153}\)

It is clear, to her as well, that in practice all these checks are only done for larger investigations and stories. She maintains that in general maybe a third, or even only ten percent, of this list is really done. Journalists seem to be confirming this statement, considering the fact that except for this one media coach none have mentioned this whole list of checks to be done. Instead, the common way to test the trustworthiness of a site is to

\(^{151}\) See Appendix 3: Confidential Interviews. P. III.
\(^{152}\) See Appendix 3: Confidential Interviews. P. III. Quote in English: “[…]things that prove to be efficient, will, at least to a minimal degree, become used.”
\(^{153}\) See Appendix 3: Confidential Interviews. P. III.
follow the journalistic rule of always having more than one source. This principle is not particular to online research but a general imperative in journalism. Offline and online the different sources need to be evaluated concerning their credibility. On the Web it is mainly a question of who the author of published information is. Official Web sites of the government and the parties are seen as reliable and useable, just as former and current articles of credible national and international newspapers found online or in the databases are. For the Web sites of organizations it depends on the general credibility of this organization in society. Knowing that the publications are always biased, it is mostly implemented stating the source. If the site or the relevant organization is unknown, many journalists indicate using the imprint to get in contact with the people managing the site and to find out more information about it. Although, again, if they are too pressed for time, this step sometimes is left out causing the test of credibility in the end to be a check of plausibility according to one’s own journalistic experience and knowledge. The trust, knowledge and experience of a journalist in general influence the decision-making process of the source’s credibility.

If, at last, a site is not considered trustworthy, it is not used. If a piece of information is used or not, again, also depends on how sensitive and interest-bound the relevant information is and of what interest it would be to use it in a manipulative way. However, this, too, is not a fact particularly connected to online sources, but a general characteristic of journalistic work. Furthermore, the way the sources are used also depends on the news organization. Smaller and more regionally working news outlets with few resources available seem to be more generous when it comes to attributing credibility than news agencies, where knowing and trusting a source is essential. A news agency’s imperative is to distribute trustworthy news items that their media clients can rely on for having been checked and being the truth. The Internet then is used with great reservation and more traditional ways of research and information gathering are sought.

Many journalists indicate that essential facts are not found through the Internet and that the Web is mainly a medium to acquire additional information and knowledge. Some even maintain that online sources in general are not reliable sources and that information is seldom taken from it on a one-to-one basis. Still, official governmental or even organizational sites are used as well as basic information from Wikipedia, indicating that the way of how to use Internet resources differs greatly and the rules in doing so are very vague and are still being explored.
4.3.5. Giving Machines the Power of Decision

As anticipated and seen through the answers concerning the use of online information, being able to attribute credibility and trustworthiness is vital. It was then of interest to find out if journalists were in favor of technology making this decision by machine-generated learning. Would tools that on a basis of collected data indicate the relevance and credibility of a site be helpful and wanted? In general it can be said, that the journalists want to be able to make decisions of this sort themselves. At first sight, such tools might seem to alleviate one’s work load and seem quite interesting, especially considering the fact that it is practically impossible to maintain an overview of all the information available on the Web. On second thought, the essential question is: what are the criteria the decisions are based on? What are the parameters? For example, a government’s Web site cannot be viewed as reliable simply because it is an official site. It depends on the country and its political system. Many answered, that in the end, they prefer making those decisions on their own. It might be of help at certain times as an additional indicator among others. However, this is seen as a possible procedure of the far future, more so than of the present day.

Although it is also mentioned that machine-generated decisions, especially concerning the relevance, is already a feature of current technological developments, it would be considered a shame more than as a great help. Some, especially when working in a specialist field, fear that non-mainstream topics will then be left aside and would be harder to find than today. The research efforts would automatically be greater again. If such technology should be employed to answer questions of reliability and credibility of an online source, it is seen as essential to not solely rely on these machine-generated results. “Der Journalismus wird das nie sein und darf das auch nicht sein, dass man sich einfach auf eine Suche in einer Suchmaschine verlässt.”

4.3.6. Visualizing the Future

Looking at what the technological future will bring, possibilities of different visualization forms of the search engine’s results were discussed. It is remarkable how many journalists seem to be content with today’s way of listing the results, even though this means searching on many pages and sometimes still not finding the needed information. It seems

---

154 See Appendix 3: Confidential Interviews. P. III. Quote in English: „Journalism will never be such and must never become such that it simply relies on a search machine for research.”
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

to be a simple form of visualizing the results that people have gotten used to and also have adjusted to when dealing with these algorithms. On the other hand, it has been stated that search queries are mostly successful when one can search very specifically knowing exactly what he or she needs to find. In this case it is easy to search with the right keywords. In other cases using many keywords helps narrow down the list of results and leads to finding the relevant data. If the topic is not known well enough and keywords are missing, searching on search engines can end up being very frustrating.

A new visualization more like a type of spider Web was suggested, which would show relevancies, for example, by size or position and enable the user to zoom in and out of the topics, meaning going into detailed information or moving between larger contexts. Some journalists did not seem too interested in such an improvement; others felt they could see quite some potential and help in such a new technology. Many remarked that it was hard to imagine working with such a tool simply because they have never used anything similar to it so far. Still, the likelihood of increasing the chance of finding the relevant information is recognized, especially when one does not know exactly what one is looking for. Despite that, in general the answers remained skeptical, not towards the idea but towards the implementation of such a tool.

Again, central are the criteria by which the results are being organized and displayed. What are the criteria by which the search engines assign relevancy? Will it focus on mainstream topics and will it make it harder to find specific and specialist niches? The awareness of the complexity of such a visualization form makes journalists hesitate even though the idea as such seems to meet approval. Moving away from a banal way of listing the results to a more complex type of visualization causes journalists to think that they lose a certain control over their own selection of primary sources.

In spite of their hesitation, when talking about their wishes in regard to future technology for journalistic research, very often the fine-tuning and the further development of search engines are mentioned. Finding the relevant and needed information quicker is only one of many wishes stated. Moreover, many different inputs were given in respect to being able to search more specifically and more interconnectedly. For example, one journalist would want a database including all the different formats pertaining to one topic, such as text, photography and video. This way, she would not have to search numerous pages for that one single topic. Similarly another journalist wishes for a meta search engine combining all sources, such as the media archive, Google, expert panels, and so on. Even more ambitious is the vision of having a database linking all the information about a person or
institution found on the different Web sites and search engines to date, and visualizing them within a type of network so that the temporal and topical interconnectedness is visible.

A media coach hopes for a search engine enabling a more local search. Many newcomers start in their own local news scenes when entering journalism. So far the Internet has helped do research on topics that are more national or international. She states having a hard time helping novice journalists because the Web is not that useful for regional information. Having search engines focusing on local information would be of great help.155 Another wish asking for more precise queries entails a search engine that only rifles through pictures that are allowed to be published and are not copyrighted, or a search engine that enables one to explore only certain fields such as merely quality newspapers.

One journalist sees room for improvement in organizing one’s own online research. She mentions that after scanning through dozens or hundreds of pages, one forgets where to have found some interesting information bits. Being able to link and mark the different pieces of data would be very helpful, adding that probably such technology has already been developed but that she simply does not know of it.156 In the same direction is the remark of another journalist stating that computers are actually too slow. Working on three different monitors at the same time, the human being links the different information of the different windows much quicker and more intuitively than the computer, ending up doing all the connections manually. He would hope for a machine being able to react as intuitively as mankind.157

One interviewee stated what could be seen as a summary of all the more specific wishes mentioned. The solution to today’s problems in research lies in getting better control over all the information available online. That is a great challenge and soon provokes the fear of being limited in one’s own decisions.158

4.4. Short summary

At the beginning of this chapter the research questions were stated and aimed to be answered by interviewing online journalists. The main objective was to find out about the current situation in the use of the Internet for research matters in online news outlets with

155 See Appendix 3: Confidential Interviews. P. III.
156 See Appendix 3: Confidential Interviews. P. III.
157 See Appendix 3: Confidential Interviews. P. III.
158 See Appendix 3: Confidential Interviews. P. III.
questions such as “What is the common knowledge about research possibilities?” and “How is the Internet used for research?”

It can be said that the way the different journalists use the Web as a research tool varies greatly. Various reasons lead to this situation. Firstly, depending on the field a journalist works in, the different tools can be of diverse relevance. Looking at the field of boulevard journalism, tools such as Facebook or Twitter seem to be of great value. In contrast, a news agency is much more conservative and cautious in using Internet items because of the possibility of it having been manipulated. On the other hand, journalists in specialist fields do find other experts via the Internet, for example via blogs, and enjoy being able to read international papers as an additional input. The only tools used in a similar manner by all interviewees seemed to be Google and Wikipedia. These online features are taken as a starting point and as an access to research, which help gain a general overview of the issue. What differs once again is the way the information found on Wikipedia is integrated into the story to be written. Where to some journalists it is simply too unsure and too prone to manipulation, others do incorporate such data simply by doing a check of plausibility. This is mainly the case when the information is considered as very factual and non-sensitive. A double-check is then left out mainly for reasons of time.

Time and time pressure seem to be of great influence and a determining factor in the use of online sources. Thus for two reasons: Firstly, it effects the decisions on accounted credibility and reliability of a source. Many sites, such as Facebook, Twitter, Blogs and unknown Web sites, are viewed as insecure; and information found there would need to be double-checked. Often the time for such checks is missing resulting in either dropping the source or using it by simply checking its plausibility. This can be observed very well when a catastrophic incident has happened and pictures or statements are quickly being sent via Twitter.

Secondly, it influences the journalist’s possibility of even getting to know the new technological tools. Especially tagging was unknown among the media workers mainly because of the lack of time and motivation to even experience it. Those that know of it do not seem to view it as important. Most journalists, though, have only heard of it or do not know of it at all. Social bookmarking was not known as an alternative way of finding interesting online sources. The same could be detected with Twitter and even with Blogs. These tools are time-consuming if one really wants to get to know them and find the qualitatively good authors. In the journalistic day-to-day life it is not possible to find such an amount of time to do so.
However, time alone is not the only reason why some online sources and the Internet in general are not being trawled. The Internet holds a vast amount of information, much of it irrelevant to a journalist’s work. This principle is the same in the real world. Everyone can give journalists inputs, hints and background information, but for a journalist it is vital to ask the right and relevant person. Any information, be it online or offline, needs to stem from a relevant and especially from a credible and reliable source. The difficulty in finding a trustworthy source seems to be harder on the Web because much is prone to manipulation. To know what can be viewed as the truth and trustworthy, journalists seek to know the sender of the information. Official governmental sites are then looked upon as reliable. The trustworthiness of the sites of private companies is dependent on their generally conceived credibility. Data might be taken, always stating the source and mentioning it as one side of the story. Time pressure usually prevents journalists from doing all the checks ideally done before using a source. In the end, the general journalistic principle of having more than one credible source and checking its credibility by way of experience and also by way of direct telephone contact, leads all journalists to conduct online research as much as it does offline research. The Internet is often defined as an additional tool in doing research, especially for background information. Of course, this medium has also greatly accelerated traditional ways of information gathering. The examples here are of finding studies and press releases online, of having live streams of press conferences or of finding telephone numbers of the wanted contact.

When talking about desired future developments, many inputs connected to better and more specific search engines were stated. These wishes either entailed search engines increasingly connecting different sources and databases to reduce the need of searching through many different portals or involved being able to do a more specific search, which would lead to only non-copyrighted pictures, qualitative local searches or specific branches, such as only looking through quality newspapers. In general, the interviewees mentioned finding relevant results better and quicker, especially when only having few keywords, would be of great help. Google’s lists seem to be easy to handle, but much information is missed in the end because it was maybe not on the first three pages. Different visualizations of the results could help in conquering this problem. However, giving the machines the power of making the decisions on relevance and importance of information does not suit many journalists. Even though journalists, as all Web users, need technological help in keeping control over the great sea of information and finding the
relevant drops, they do not like the feeling of losing their say in deciding where to search and what is relevant.

The next chapters are to reveal the nature of the Web 2.0 and also of a future Web 3.0 and their potential in regard to a professional journalistic research, accounting for the journalists’ special needs in retaining their own freedom and capability of finding their primary sources.

5. The Characteristics of the Web 2.0 and 3.0

5.1. The Web 2.0

“Web 2.0 represents more than just a change in Internet technology. It is a global change in how we engage with one another.” The concept of Web 2.0 was born at an O’Reilly Media brainstorming session in 2004. Although there is no single commonly accepted definition, Tim O’Reilly defined the term as “the business revolution in the computer industry caused by the move to the Internet as a platform, and an attempt to understand the rules for success on that new platform.” Seeing the Web as a platform is probably the most central key to understanding the Web 2.0. It involves considering the Internet as a medium of interaction, participation and collaboration. What seems to be simple and easy to explain includes such a huge phenomenon “that it resists any attempt to pin it down.” As San Murugesan notes, the Web 2.0 has made people change how they gather information, do their work, buy goods and services, how they connect with friends, family and how they spend their leisure time. It has even influenced the face of politics and governance. This is not the effect of a sudden change, but rather a slow evolution. According to James Governor, Dion Hinchcliffe and Duane Nickull, the Web 2.0 is a constant, gradual change taking place as older devices are retired and new applications are connected on the platform.

162 Ibid. p. 1.
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

The term Web 2.0 was coined in order to visualize the Web’s evolution in the different stages. Then, the Web 1.0 can be seen as the Web of the 1990s, providing and connecting information. It was a one-way publishing medium where only a small number of providers with specialized skills created Web sites. In contrast, on the platform of the Web 2.0 it is all about connecting people. It is also called the democratic Web because of the social revolution, which is inherent to it. Thus, the Internet now allows all users to create, share and distribute information and images through technologies known as social software. Hence, it is the people-centric Web, because now nearly everyone can contribute content online. Although technology is seen as important, it is secondary to achieving the greater goal of promoting free and open access to knowledge.\footnote{Cp. Murugesan (2010): p. 2 et seq.} Murugesan pictures the Web’s evolution in an onion-like way.

![Figure 2: The evolution of the Web (source: Murugesan [2010])](image)

This figure is supposed to illustrate that the different stages do not replace one another, but build on each other to achieve different objectives. Taking a quick look towards the future, the Web 3.0 is supposed to enable a more productive and intuitive user experience through machine-facilitated understanding of information. Chapter 5.2. will go into more detail on the notions of the Web 3.0. According to Murugesan, even the concept of a Web 4.0 exists, indicating higher levels of intelligence, identifying it also as the intelligent or smart Web.

Continuing with our look at Web 2.0, Murugesan indicates seven key characteristics of it: participation, standards, decentralization, openness, modularity, user control and
Especially the users play a vital role in this generation of the Web. They actually create it themselves. This phenomenon will be explained adjacently in more detail.

5.1.1. The Users’ Web

In 2006, the magazine *Time* selected you as the person of the year on the basis of what was happening on the Internet in and for society. The key to the functioning of the Web 2.0 is the user himself. Governor et al. and many other authors indicate that consumers are now becoming producers and that the Web is putting you in charge of just about anything you can imagine. The users, formerly solely consumers, can now add and create content as well as react to other already published content on the Web. In his article titled “What Is Web 2.0” O’Reilly identifies the architecture of participation and the harnessing of collective intelligence as two core competencies of today’s Web. “There’s an implicit ‘architecture of participation’, a built-in ethic of cooperation, in which the service acts primarily as an intelligent broker, connecting the edges to each other and harnessing the power of the users themselves.” He continues by saying that companies succeeding in the Web 2.0 era are the ones that have embraced the power of the Web to harness collective intelligence. This means that Web applications are to be seen as services and no longer as software products being released on scheduled dates. These services are not sold, nor do they involve any licensing. There is just continuous improvement through usage, through the users’ feedback.

O’Reilly names three good examples here: Google, Amazon and eBay. Google’s success is based on the principle of PageRank, which rates the relevance of a Web page according to the used links. EBay is a service that grows organically in response to the user’s activity and Amazon has successfully integrated the users through their recommendations and ratings. The success of these services can also be seen in understanding Chris Anderson’s theory of “The Long Tail”. As already explained previously, it is the collective power of the small sites that make up the bulk of the Web’s content. Especially Google has taken this into account when starting Google AdSense. This service allows people to sell advertising space on one’s homepage. According to O’Reilly, Google as well as Overture, Yahoo!’s provider of such a service, eschewed using ad-agency friendly

---

165 Ibid. p. 5.
168 Cp. ibid.
advertising formats in favor of using context-sensitive, consumer-friendly text advertising. In cases like these, Google, or any software service, then act as a middleman between the user and his or her online experience. The key principle is that the service automatically gets better the more people use it. For commercial companies this means that, following O’Reilly, network effects from user contributions are the key to market dominance.

In this environment it is clear, that data and database management play a vital role. “Without the data, the tools are useless; without the software, the data is unmanageable.”

Looking at Google’s effective search results, Amazon’s recommendations, blogs, comments and links, they all involve data that needs to be managed in order to generate a positive effect in their application and, in the end for the user. The software tool is only as good as the way the dynamic data has been administered. By using the tools, each Web user adds certain data. This can happen through uploaded content such as posts, comments or ratings, but likewise through ways, as Napster had done it, that transform every downloader also into a server connected to the Web and thus making the (music) network grow.171 To O’Reilly, this is another key lesson: Users add value. By pursuing their own “selfish” interests, users build collective value as an automatic byproduct. It is the harnessing of this collective intelligence that is significantly important in the Web 2.0 era. Database management is a core competency of Web 2.0 companies. The race to own certain classes of core data is on. Recent debates have shown how big computer technology companies, such as Google or Apple, have been criticized for collecting data secretly. The newest article is about Apple collecting information for already as long as two years from iPhone and Mac users without them knowing it.172 And on the same day it was reported that 38 US-states have allied to clarify whether Google has broken the law by simultaneously collecting data from open networks while photographing the streets for the service Google Street View.173 It is the great challenge of today’s society to figure out ways to collect needed data without conflicting with any laws concerning data privacy.

In a participative environment shaped by services and their users, it is important to no longer consider the users solely as customers, but rather as co-developers. The Web 2.0 entails open source development practices, where the motto “release early and release

171 Cp. ibid.
often” becomes central. New tools are also being implemented and developed through feedback from the users working with it. Users can effectively be a part of the development of Web 2.0 applications by identifying the set of required features and validating the already implemented ones. This phenomenon reduces the lifecycles of the applications and improves their usability; an approach known to be described by the term “the perpetual beta”. Now, the users shape the Web not only by adding content and data, but also by actually evaluating and forming the applications used. O’Reilly summarizes this as the key principle of innovation in assembly. He calls for innovations that are designed for “hackability” and remixability to minimize the barriers in re-using the different applications in the process of generating new ones. The Web 2.0 then is not only a Web of participation in regard to communication but also in regard to technical improvements. Users can be influencers.

Governor et al. conclude that the Web 2.0 nudges the balance of power toward the average person, balancing the unequal powers between large corporations and individual rights. The Internet is seen as the largest single marketplace known to mankind. The architecture of participation is seen as the framework the Web 2.0 provides its users to interact. Interactivity through social software and applications are central to the Web 2.0 and will be reflected on the following pages.

5.1.2. A Web of Interactions

In a participative Web interactivity has an important and central constructive role. “Interactivity in the electronic marketplace is the degree to which computer mediated communication is perceived by each of the communicating entities to be (a) bidirectional, (b) timely, (c) mutually controllable, and (d) responsive.” This definition by Yadav and Varadarajan is one of many presented by Sachdev, Nerur and Teng. Summarizing the different aspects, the authors conclude that there are three dominant perspectives to the concept of interactivity: a) as the user’s perception, b) as a function of the properties of the medium, and c) as a process of message exchange or interaction with the message or

---

175 O’Reilly (2005).
medium.\textsuperscript{177} Interactivity then can be viewed as a multi-dimensional construction with the core dimensions being control, responsiveness and reciprocal communication.

First of all, control as a dimension entails being able to manage the navigability, the content or the pace of an action on the Web. It refers to having control over the communication process, over the discourse and over the ability to exchange roles. The range of interactivity then is determined by the number of attributes that can be manipulated and the amount of variation possible within each attribute. The higher the number of attributes the user can modify the greater the degree of perceived control. In the end it depends on the perception of the user, rather than on objective evaluations of control attributed to an application.

Secondly, responsiveness refers to the possibility of responding as well as it includes the element of time. The quick transitions between one’s actions and the consequent outcome make a Web site interactive. Users are no longer just browsing through the Web consuming information passively, but are interacting actively with the medium by changing its content and by interacting with others. The speed and the possibility of a response, as well as the probability of a response, play vital roles.

Thirdly, providing two-way communication is a central competence of an interactive medium. This dimension of reciprocal communication indicates facilitating interpersonal communication. Sachdev et al. differentiate between two types of reciprocal communication: User to user and user to Web site. The former can vary in forms of one to one, one to many or many to many. The latter involves easy options on a site to provide feedback to the site administrators.\textsuperscript{178}

Adding to these three core dimensions, the authors see three further dimensions as essential to interactivity on the Web: social presence, self-presentation and deep profiling. These aspects embrace the importance of identity communication, which is salient in social computing. The notion of social presence refers to the user’s perception of a human or artificial presence. It can be described as the degree to which users sense the physical existence of others and the extent of interaction with other users.

Self-presentation entails the way one portraits oneself to the other user. Sachdev et al. describe this action as helping others form a more sophisticated and accurate understanding of “Who am I?”\textsuperscript{179} According to Goffman, communication of identity is the

\textsuperscript{177} Ibid. p. 590.  
\textsuperscript{178} Ibid. p. 593.  
\textsuperscript{179} Ibid. p. 594.
first step in any new interaction. On the Web self-presentation enables people to judge the personality and identity of others and, furthermore, helps them to find people with shared interests or tastes.

Deep profiling directs the attention not to what people have said about themselves, but to what people have done. Such records could be rankings, feedbacks, archives of user contributions that provide a context and information about another user who is the subject of a possible interaction.

In summary, control, responsiveness, reciprocal communication, social presence, self-presentation and deep profiling are the central dimensions to interactivity. It is the so-called social software allowing this new type of social computing to happen. The IBM Social Computing Group defined social computing as follows: “Social computing refers to systems which support the gathering, representation, processing and dissemination of social information, that is, information which is distributed across social collectivities such as teams, communities, organizations, cohorts and markets.” Social software then has emerged as a driving force of the Web 2.0. According to Lorna Uden and Alan Eardley, the Web 2.0 is not a single development, but rather a heterogeneous mix of new and emergent technologies. It is the social software enabling the users to actually communicate, collaborate, contribute, connect and share their interests. Examples of these new technologies are blogs, wikis, social networks and folksonomies. Sachdev et al. proclaim that in 2007 48% of the adult online consumers in the US participated in such activities. Social software tools enable users to create, publish and distribute contents such as videos, pictures and documents. People can coordinate their activities and even create reliable, robust and complex products, such as open software applications like Linux.

Uden and Eardley make out three characteristics that are commonly attributed to social software: conversational interaction, social feedback and social networks. In the end it is the community that gains by using social software, because it enables the communication between many people and between groups, it provides the gathering and sharing of

181 IBM quoted in ibid. p. 587.
183 See chapter 6.
resources, it delivers collaborative collecting and indexing of information, it allows syndication and it assists in the personalization of priorities. The authors see social software engaging people as active participants in online social networks and communities, achieving new and exciting effects through distributed collaboration, co-production and sharing. Governor et al. allude that the long-term effect of this connectedness will no doubt be as unpredictable as it will be significant, due to the fact that control over information and content becomes relentlessly decentralized. “The Web is a system without an owner, a platform that’s under no one’s control, though anyone is free to build a new platform on top of it.”

Uden and Eardley identify four key attributes of social software: functionality, usability, interactivity and naturalness. An application that is difficult to use, will not be adopted by the users, making usability evaluation necessary. Usability of an application aims for effectiveness, efficiency, safety, good utility, learning ease and an easy retention of how to use it. Here again, it can be seen how the end-user is part of the development process of any new application and service on the Web.

While using social software, people express themselves and produce so-called User-Generated Content (UGC), including blogs, wikis, digital videos, Internet broadcasting, photography sharing, mobile phone photography and so on. Casoto et al. define UGC as any kind of published content that is the result of a nonprofessional activity with a creative effort. Table 1 shows the different classifications of UGC and specific providers.

<table>
<thead>
<tr>
<th>Blogs, Message boards and forums</th>
<th>WordPress, Technorati, Blogger</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Review / rating sites</strong></td>
<td>Amazon, Tripadvisor, Epinions, Yelp, Ebay</td>
</tr>
<tr>
<td><strong>Clubs or groups, Photo and Video sharing</strong></td>
<td>Flickr, YouTube, GoogleVideo, DailyMotion, MetaCafé, PodZinger</td>
</tr>
<tr>
<td><strong>Social networking</strong></td>
<td>LinkedIn, MySpace, Friendster, Facebook, SecondLife</td>
</tr>
<tr>
<td><strong>Collaborative Authoring</strong></td>
<td>Wikipedia, Google Docs, PBWiki, SlideShare</td>
</tr>
<tr>
<td><strong>Social bookmarking and knowledge sharing</strong></td>
<td>CiteULike, Connotea, CiteSeerX, Del.ici.ous, SharingPapers</td>
</tr>
</tbody>
</table>

Table 1: Classifications of UGC applications (source: Casoto et al. [2010])

---

186 Ibid. p. 575.
Blogs are seen as the most powerful UGC media. O’Reilly refers to the blogosphere as the equivalent of a constant mental chatter in the forebrain. Imagining the whole Web as a kind of global brain, the blogosphere as the forebrain is than described as the conscious thoughts and thus has begun to have a powerful effect.\textsuperscript{189}

Where blogs are usually first-person narratives, message boards and forums are based on a set of users acting as authors, mapping their own reputation. Review and rating sites allow users to share their own experiences or evaluate the goodness and usefulness of the previously published contents provided by other users. Clubs or groups are mainly specialized and highly focused sites accessed by a limited number of participants. Photo and video sharing applications enable people to publish their multimedia content. By tagging, voting and aggregating uploads with further content, users add value by enriching the data with a sort of collective intelligence. Wikipedia, Google Docs and other examples are listed in the category of collaborative authoring. These tools allow users to add content in a collaborative way. Worldwide people can share, edit and store a set of documents simultaneously. Where Google Docs can be used by only a self-defined number of people, Wikipedia is open to all. To O’Reilly, Wikipedia is a radical experiment in trust.\textsuperscript{190} The last classification of social bookmarking and knowledge sharing refers to applications that help create, organize and share different, more complex, UGC. These different Web 2.0 applications in particular will be analyzed in more detail in Chapter 6.

It is also the UGC that has had an economic effect on the traditional media, forcing them to employ UGC themselves. Even though the business models are several and diverse, there are still only few sites with viable business models.\textsuperscript{191} Yet, one of the most important features of UGC is the possibility to access and analyze the spontaneous conversations of the users and derive new strategic knowledge, which is of value for one’s organization. Making it easy to contribute and publish causes a vastness of information that makes it hard to find and deduce the relevant and important information. In this sense the Web 2.0 is subject to several severe limitations related in particular to the retrieval and organization of UGC.

\textsuperscript{189} O’Reilly (2005).
\textsuperscript{190} Ibid.
5.1.3. Limits to the Web 2.0

The Web 2.0 enables users to interact and share, to produce and provide content as well as metadata. The side effect is visible in an information explosion and overload and in the lacking accuracy of the retrieval tools and their difficulty to create adaptive filtering mechanisms with respect to the user’s informational needs. Casoto et al. have divulged several limitations of the Web 2.0 connected to information retrieval within UGC. Retrieval mechanisms, so far, do not provide a full understanding of the content meaning and do not look at the contents in relation to other available shared contents. Users can easily be overwhelmed by the amount of retrieved contents. In addition, rich media worsens the problem of information retrieval. According to the authors, rich contents should be treated on the basis of their real contents and only by using traditional methods, such as text descriptions or tags.

UGC generates an open corpus of documents that do not share a common ontology. Information in the form of comments and conversations are often characterized by informal language and no specific structure. It can constantly change and expand what makes knowledge management increasingly complex. Furthermore, the rising number of producers and contents makes it constantly more difficult to measure the trust and quality of the information. The accuracy of search engines becomes even more significant and pressing. Not only the accuracy but also the personalization is seen as becoming more necessary since the sea of information is getting larger and larger, making it harder to find the relevant information needed. Personalized systems are aimed to overcome this problem of information overload and, according to the authors, have proven successful to a finite corpus of documents. It still needs to be developed for an open corpus of documents as on the Web 2.0.

The last limitation refers to the uncontrolled vocabulary used in folksonomies for tagging and social bookmarking, which hinders effective classification and information management. Although social bookmarking represents an attempt to improve Web research and to solve the information overload, the problem of vocabulary and the fact that social bookmarking sites are too limited in size prevent them from gaining significant impact.

193 Cp. ibid. p. 316 et seq.
In conclusion, current technologies are not adequate to solve fundamental problems, which are present in the Web 2.0. Research to solve these problems has resulted in ideas and notions for a future Web 3.0, which are examined in the next chapter.

5.2. The Web 3.0

“[…] Web 3.0 technologies will help filter the ‘wisdom of the crowd’ so that it doesn’t become the ‘madness of the mob.’”[194] The vision of a Web 3.0 seems to include many different technological evolutions creating a new generation of the Web. Terms like Semantic Web, 3D Web, media centric Web as well as pervasive and ubiquitous Web are labels defining the different fields of research all contributing to a Web 3.0.[195] The 3D Web refers to the ability to create virtual worlds, in which users can live the way they could never do in the real world. They create avatars and make them live on their behalf. Research on touch and haptics allow new dimensions for interaction.

Central to the media centric Web is the approach to use media objects as the main element in searching for other media objects. In this case user interfaces need to evolve into intelligent, natural and multimodal interaction systems that are supposed to consist of the ability to recognize hand gestures, voices and even people’s faces and their moods, causing a response in a multimodal fashion. Again, haptics research will allow for new forms of interacting and socializing.

The concept of the pervasive and ubiquitous Web refers to people being permanently connected to the Web without even thinking about it. Virtually the Web is everywhere: on every computer, every mobile device and even in common everyday objects such as our clothes, furniture, appliances et cetera. They all seek information on the Web to provide a specific service or information update. For example, windows and curtains could check the weather report and would open and close accordingly.[196] The one concept receiving the most attention, however, is the concept of a Semantic Web. It was outlined by Tim Berners-Lee in 2001 and is strongly supported by the World Wide Web Consortium. Berners-Lee is the creator and father of the World Wide Web. He is a

professor at the MIT and also directs the World Wide Web Consortium (W3C). The Semantic Web is a revolutionary technological approach for organizing and exchanging information in a dimension of cross-application. It can provide a rich and powerful technical infrastructure for any kind of Web application. What this entails and how it should become reality is described in the next chapter.

5.2.1. The Web of Data

According to the W3C, “[t]he Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.” (Original emphasis). The Web 2.0 is characterized through the different applications and platforms enabling users to become producers as well as being consumers. In comparison to its predecessor, the Web 1.0, the Web 2.0 refers to a different way the Internet technology is being used. People interact. The Web has actually become a Web of people connecting with each other and producing content.

The Semantic Web, on the other hand, is an evolutionary step on a technical basis. It is designed to master the vastness of information produced and help find the needed information more specifically and personally by actually understanding it. One can say that the Web 2.0 and the Semantic Web are complementary and together constitute what is termed Web 3.0.

To understand the notion of the Semantic Web, one needs to imagine it not as a Web of documents, such as the Web 2.0, but as a Web of data. The information that can be found on the Internet, then is not seen in whole documents or Web pages anymore, but in the different pieces of data available that actually build the Web pages. There seems to be a paradigm shift caused by a new era in integrating and aggregating information. “If the current Web is like a giant text file – which you can search for instances of particular words – the Semantic Web would be like a database, where every item of information is categorized, and new queries can combine categories in any imaginable way.”

---

201 Hardesty, Larry (2010).
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

can be seen as a giant networked database. What is fundamentally new is that this data can be inquired through technology that actually understands the data in a human-like way. Ontologies and the Resource Description Framework (RDF) are tools to process and interpret the data found, to categorize it and set it into relationships so that it makes sense to the user and helps him find the specific and needed information. In Berners-Lee’s words, the data merely displayed in today’s Web will actually be processed and understood by machines. Therefore, this data then serves as a source and can be used for different kinds of applications, depending on the user’s needs.

In this environment the essential property of the World Wide Web is its universality. Using the Web as one big data source means to create a common framework that allows data to be shared and reused across applications. The data needs to be processed automatically by tools that reveal new relationships among the pieces of data. To achieve this interoperability uniform and open standards are needed to create data that is machine-readable. W3C has set these standards in RDF, Web Ontology Language (OWL) and the query language, SPARQL.

5.2.1.1. Exploring Meanings through Machine-Generated Languages

Linked data is the heart of the Semantic Web. There is a collection of Semantic Web technologies providing an environment where applications can query that data and draw inferences. Making new relationships between different data is central in creating the Web of data. RDF is one of the fundamental building blocks of the Semantic Web. It is a standard model, allowing data exchange between the different applications. RDF actually ascribes meaning to a certain piece of data by encoding it into a set of triples. Each triple consists likewise of a subject, verb and object and can be written using eXtensible Markup Language (XML) tags.

---

203 Ibid.
205 W3C: Linked Data. URL: http://www.w3.org/standards/semanticweb/data.html (21st July 2010).
If a document is written in an RDF format, the machine can read that particular things (subject, i.e. people) have properties (such as “is sister of”) with certain values (another person, another Web page). Berners-Lee et al. identify this procedure as the natural way to describe the vast majority of the data processed by machines. In this sense the triples of RDF form webs of information about related things. Through Universal Resource Identifiers (URI), subjects and objects are each identified exactly to ensure that concepts are not just words in a document but are tied to a unique definition that everyone can find on the Web. As Universal Resource Locators (URL) stand for links that interconnect different Web pages, URI connect between things, giving the relationships names. This linking structure extends the Web of linked documents to a Web of linked data and relationships.

It is clear that in an environment as complex as the one of languages, ambiguities are bound to exist. Ontologies help to integrate ambiguous data by describing and representing areas of knowledge. They classify the terms used in a particular application and characterize their relationships. According to Berners-Lee et al., the most typical kind of ontology for the Web has taxonomy and a set of inference rules. The taxonomy defines classes of object and the relations among them. To express these sets of relationships languages like RDF Schema or OWL have been set as a standard for the ontologies. This example given by the W3C may help understand the function of ontologies:

“A bookseller may want to integrate data coming from different publishers. The data can be imported into a common RDF model, eg, by using converters to the publishers’ databases. However, one database may use the term ‘author’, whereas the other may use the term ‘creator’. To make the integration complete, and extra ‘glue’ should be added to the RDF data, describing the fact that the relationship described as ‘author’ is the same as ‘creator’. This extra piece of information is, in fact, an ontology, albeit an extremely simple one.”

206 Berners-Lee et al. (2001).
207 Ibid.
Ontologies can be very simple or really complex. However, they can improve the accuracy of Web searches by relating the information on a page to the associated knowledge structures and inference rules.

Inference rules offer ways to express constraints on certain relationships. For example, a rule could state, that if two people have similar names and addresses as well as homepages and email addresses, then they are identical. Or if databases work with different columns for names, first names, middle names and surnames, rules could unify the data by saying: “Put everything up to the first space character in ‘first name,’ anything after the last space character in ‘last name,’ and anything else in ‘middle.’”

Rule Interchange Format (RIF) is the newly released standard for inference rules for Web programmers.

Inference, in the Semantic Web context, can be defined as the discovering of new relationships. Based on data and additional information in form of an ontology or rules, automatic procedures can generate new relationships. For example, if one data set proclaims that a pochard is a type of duck and an ontology declares that all ducks are birds, the Semantic Web program then concludes that a pochard is also a bird, even though the original data did not entail this information. In this way, a new relationship has been discovered.

In conclusion, RDF, OWL and RIF are the core technologies creating the Semantic Web. They provide the knowledge, which can be read by the machines. To extract the information from this semantic source a specialized query tool needed to be developed which would be able to read the semantic information. SPARQL is the set query standard that is able to mine the information of the data. Together these technologies allow a much more precise and personalized Web research.

5.2.2. Enhancing Personalization

With these technological conditions and possibilities the Web is no longer a network where one can find whatever has been produced as such by others. Moreover, every user now is able to aggregate the existing data in a personal way, so that he can find the information or the service he needs. “[…] while an ordinary database has categories selected in advance by a programmer, the Semantic Web is ‘a database where each person controls their own

209 Hardesty, Larry (2010).
210 Ibid.
data”. To the W3C the Semantic Web technologies can be used in a variety of application areas. They all form a better and different user experience of the Web than so far. Data integration, resource discovery and classification, cataloging to describe the content and the content relations, content rating and intelligent software agents are all means of treating the found information and data in a way to enhance the user’s success. By using data from different sources at different times and actually understanding it, the computer or any device constantly learns and adds knowledge to its own database upon which future actions and applications take place. In this sense the computers learn from the user’s behavior as well as from the data found; and they combine this knowledge by providing more specific and intuitive search results than before. Applications can be adjusted to the user’s needs by using specific data as input.

As a technology-based improvement, the Semantic Web may act behind the scenes, rather than directly influencing the “look” on the browser, which results in a better user experience. However, according to Silva et al., the user interface is vital to the adoption of the Web 3.0. They maintain that new technologies will succeed or fail depending on how they present themselves to the user. The adoption should be as natural as possible with user interfaces that are very cognitive.

Applications, which function based on data, are one part of the Web experience. To help users find and interpret the data needed, so-called software agents process the information available according to the user’s needs. Following Murugesan, smart software agents will help users manage the complexity of their digital lives. Berners-Lee et al. note that the effectiveness of software agents will increase exponentially to the availability of more machine-readable Web content and to the automated services. “The real power of the Semantic Web will be realized when people create many programs that collect Web content from diverse sources, process the information and exchange the results with other programs.” Through these agents the Semantic Web allows more flexibility than the Web so far, because the different agents share understandings by exchanging their ontology that provides the vocabulary needed for discussion. To illustrate the dimension of this new way of using the Web, the example and vision of Berners-Lee et al. of the

---

212 Hardesty, Larry (2010).
216 Berners-Lee et al. (2001).
217 Ibid.
possibilities of the Semantic Web is quoted below as stated in the paper on the Semantic Web.

“The entertainment system was belting out the Beatles’ ‘We Can Work It Out’ when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other local devices that had a volume control. His sister, Lucy, was on the line from the doctor’s office: ‘Mom needs to see a specialist and then has to have series of physical therapy sessions. Biweekly or something. I’m going to have my agent set up the appointments.’ Pete immediately agreed to share the chauffeuring.

At the doctor’s office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom’s prescribed treatment from the doctor’s agent, looked up several lists of providers, and checked for the ones in-plan for Mom’s insurance within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services. It then began trying to find a match between available appointment times (supplied by the agents of individual providers through their Web sites) and Pete’s and Lucy’s busy schedules. (The emphasized keywords indicate terms whose semantics, or meaning, were defined for the agent through the Semantic Web).

In a few minutes the agent presented them with a plan. Pete didn’t like it – University Hospital was all the way across town from Mom’s place, and he’d be driving back in the middle of rush hour. He set his own agent to redo the search with stricter preferences about location and time. Lucy’s agent, having complete trust in Pete’s agent in the context of the present task, automatically assisted by supplying access certificates and shortcuts to the data it had already sorted through. Almost instantly the new plan was presented: a much closer clinic and earlier times – but there were two warning notes. First, Pete would have to reschedule a couple of his less important appointments. He checked what they were – not a problem. The other was something about the insurance company’s list failing to include this provider under physical therapists: ‘Service type and insurance plan status securely verified by other means,’ the agent reassured him. ‘(Detail?)’

This procedure, of course, is still a dream and futuristic. Even Berners-Lee called it a little “too sci-fi”219. However, it does show the possibilities and the power behind the principle of a Semantic Web. It illustrates the vision for which the developments are taking place at the moment, even though they still are dreams of the future.

5.2.3. The Importance of Trust

The Semantic Web does have one problematic issue that Ulrich Ultes-Nitsche even gives equal attention to as compared to the other central aspects of ontologies and social applications: this is the issue of security.220 He states that the Semantic Web, unfortunately,

---

218 Ibid.


does provide a multitude of ways to abuse the technology. As machines act and react to data resources, these inputs are central and need to be correct in order for the applications and agents to function truthfully.Cryptographic procedures are necessary to ensure that information is being acquired through a trustworthy resource and not through some fake copy. Murugesan calls this new type of security threat the semantic attack, noting that falsifying input into a computer process can be much more devastating than in the processes today.\textsuperscript{221}

Thus, trust becomes vital in a Web 3.0 context. He identifies different ways to obtain trust: If there is a history of interactions between systems in which the latter have behaved correctly, trust may grow. A second possibility to achieve trust is through structures of recommendations. If a trustworthy system recommends another system, the trust of the latter is heightened. According to a system’s interactions and recommendations, a numeric value is given. Then, a system will only interact with another system if its value exceeds a certain threshold level. Systems not attaining this level are put on a black list.\textsuperscript{222}

Another big issue arising is the one of privacy. A Web of data needs a huge amount of data to function. For Silva et al. there is an increasing concern over the personal information and the control of its digital distribution. They proclaim that people need to know and control what information gets shared on the Web. “There is the need to establish the appropriate balance between automation, human behavior patterns, and those human interventions necessary to mediate the environment.”\textsuperscript{223} This is a difficult goal to achieve. And as already mentioned in chapter 5.1.1., the race on data is on and the discussions about it as well.

Having presented the general notions of the Web 2.0 and the Web 3.0, the next chapter introduces the specific tools and applications in more detail and in connection with their possibilities for journalistic research.

\textsuperscript{222} Cp ibid. p. 10 et seq.
6. The Research Possibilities on the Internet

6.1. Blogs – Start Discussing!

The expression ‘weblog’ – or short: blog – stems from the words ‘Web’ and ‘log’ and describes a kind of online diary. The author of the blog posts stories in the form of short publications. These texts, called posts, are sorted chronologically, starting with the most recent. The people publishing blogs are called bloggers. The whole arena of blog-communication is the blogosphere.

To O’Reilly, blogging is one of the most highly touted features of the Web 2.0.224 Also Casoto et al. regard blogging as one of the most common activities introduced by the Web 2.0 and identify blogs as the most powerful UGC media.225 Compared to Web sites, blogs are easier to handle and are more flexible in their use. Through the so-called Content Management System226 information can easily be added and administrated. This simplicity is the basis for the vast and rapid global spread of the blog-technology and its usage.

The dialogue-based communication form, which is inherent to blogs, is also a result of the large number of links. Posts are connected and referred to through a function called trackback. Trackbacks at the end of a post lead to other blog entries that have been written about the same topic. Often these trackbacks are actually reactions to the post they are linked to, but have been published in another blog instead of in the form of a comment. In order for trackbacks to work, permalinks are necessary. This technology enables each publication to have its own address. In other words, each post or comment has its own explicit Web address through which it is retrievable at any time or place within the Web.227 O’Reilly sees great significance in permalinks. By building bridges between the blogs, permalinks turned them from an ease-of-publishing phenomenon into a conversational mess of overlapping communities.228

The notion of the Web 2.0 is turning the Web into a kind of global brain. Taking this as a basis, O’Reilly then describes the blogosphere as the equivalent of a constant mental chatter in the forebrain and as an equivalent to conscious thought, while the deep brain is often unconscious. As a conscious thought the blogosphere has begun to have a powerful

227 Ibid. p. 38.
228 O’Reilly (2005).
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

effect. According to Technorati’s report, State of the Blogosphere 2009, the influence of the blogosphere is rising and was well perceived during the presidential elections of 2008 in the US or during the protests of the presidential election in Iran. As a marketplace of ideas the blogosphere can become a key forum, where in complex issues its impact is considered immediate and noticeable. Thus, blogs contribute to the rapid identification of situations by encouraging discussion and collaboration.

“More generally, blogging is the next step in a process of advancing communication from radio to TV to internet messaging. The breadth and depth of the blogosphere allows sophisticated information – and special expertise – enhanced range. Comments and follow-up posts allow for original ideas to be refined and perfected even as they are spreading around the world.”

Paul Levinson states the two main characteristics of blogs: Firstly, anyone can blog about anything. Secondly, the actual impact of a blog, as well as the time of its maximum impact, are unpredictable. Blogs can and do have impact. However, it is not to be idealized. As Greg Myers notes, blogs can mobilize society to bring down politicians, hold a corporation to account, popularize a book or spread a video, but blogs can also endlessly repeat a vicious lie, long after it has been debunked in the press.

Blog posts are not only immediate and universally accessible, but they also last forever. In principle, it is impossible to completely delete a post once it has been published. The freedom of each blogger to be in control of his own blog and the possibility of constantly revising it is a deceiving control mechanism which is very limited due to the capacity of anyone being able to copy and save whatever is on the blog for further dissemination. Statements can continue to circulate even though they have been revised. Myers concludes that the provisionary nature of blogs has its own dangers.

Blogs do not underlie a certain authority or hierarchy as known in the fields of the press, the academia, medicine or law, which forms the interpretations of credibility. The central characteristic of the blogosphere is that it flattens the various hierarchies; it equalizes the relationships through the fact that anyone can blog about anything. According to Myers, the users create a social world; and also Axel Bruns and Joanne Jacobs mention that

---

229 Ibid.
235 Ibid. p. 21.
interactivity in the right setting becomes intercreativity. “By personalizing content, blogs go beyond a purely informative role and provide a platform for debate, deliberation, and the expression of personal identity in relation to the rest of the (blogging) world.”

Jim Macnamara connects to Chris Anderson’s theory of “The Long Tail” and names the blogosphere the greatest vector for new voices ever created.

Looking once more at Technorati’s State of the Blogosphere 2009 report, the main motivation for bloggers to publish content is self-expression and sharing expertise. Furthermore, the percentage of professional bloggers is growing. 28% of all bloggers enter Technorati’s category of professional bloggers. Also Macnamara states that an increasing number of blogs are written by professional journalists and experts, such as scientists, IT specialists, engineers, health professionals and so on. In Technorati’s report 27% of all bloggers still work for and 35% have worked for the traditional media. The main source for blogs is still the mainstream media, even though they mock them. Bloggers do not replace the reporters, they need the media to sort out what is news. The blogosphere and the mainstream media increasingly seem to overlap.

In general, it can be said that the blog population is widely disparate. Bloggers are generally highly educated, come from different jobs, speak different languages and come from different places. Myers describes the community of bloggers not so much as a public sphere but as a set of little sphericules. This term refers to “multiple publics that pursue their own discussions without reference to a single unified national or global ‘public’”. The blogosphere is breaking up into different blogospheres such as the conservative blogosphere in the US, the Farsi blogosphere in Iran, a scientific blogosphere, Christian blogospheres, et cetera. Different types of blogs are emerging such as warblogs,

237 Ibid. p. 5.
mommyblogs, lawblogs and so on. Myers states that a successful blogger does not write for the world at large, but for the people alike, wherever they may be.246 However narrow the specific group or blogosphere may be, they are all linked in complex ways. Blogs are hard to pin down because it is typically full of links to other texts. Bruns interprets it nicely by using the author and blogger Neil Gaiman’s words: “the blogosphere is not organized, but it’s really well disorganized.”247 For Myers, linking is the currency of the blogosphere, mentioning that nearly all bloggers link to other blogs. “We could think of each blog as one point, with a number of arrows going out to other blogs and to the various readers who comment back.”248 Blogs are designed for the Internet allowing instant access to quoted sources. This affects the way they are read. Blogs afford a reading style involving moving to another text, and maybe back again, striving to figure out what the blogger is trying to say and what stance he is taking on the new text.
Participating in the blogosphere also involves commenting. Levinson indicates that comments are the most frequent form of sustained written discourse and attribute two other functions inherent to them. Firstly, comments can be an effective promotion for one’s own blog. Secondly, and more important, comments do not only serve as the voice of the people, but also as the conveyors of the truth. They correct, if necessary, a post. “The whole world, in principle, is not only reading what you write when you blog but is waiting there as a potential safety net and source of correction for any mistakes you might make.”249 In this sense, the social group acts as a guarantor of truth or accuracy. Levinson remarks that this factor is also found in phenomenon like Wikipedia or Twitter.250 Knowing how to ascribe reliability, trustworthiness and accuracy to a blog is becoming more and more crucial and is especially important in journalism. Seeing as blogs are unedited content, the easiest way is to not believe any of the bloggers at all. However, as Will Richardson mentions, one would be ignoring some very smart and relevant voices that are gaining more and more prestige. He states, that in the end it is a matter of trust, because there is no way of telling for sure if something you read on the Web in general is true or not. One needs to determine him- or herself who to trust about what.251 However,
there are certain actions to take to find out about the trustworthiness of a blog: find out about the author and his or her reputation, evaluate the sites linking to the blog, read through other posts and look at the links, check who is quoting the source and search for corroboration from a different source.\textsuperscript{252} This, of course, requires time. Richardson advises becoming a blogger oneself in order to “navigate through the murkiness”\textsuperscript{253}. He maintains that there is no better way to understand the impact of the Read/Write Web better than by becoming a part of it.

In the blogosphere traditional hierarchies have vanished and interactivity is central to it. Other ways of control seem to be emerging. Bloggers keep an eye on their audience; and their main task is to manage the interaction and to keep it going. Evaluating the importance of the different bloggers is mainly done by analyzing their received attention. Links, hits, trackbacks and being mentioned in other top blogs are indicators of value. Wanting to be spread and respected in the blogosphere makes bloggers careful when marking their own opinions. Myers observes that most bloggers adjust the ways they express their opinions to better interact with the audience. They convey the complexity, interest and novelty of their views. Just asserting one’s point of view could sound bullying and egotistical, which would make it hard for the blog to survive since it has to live in an interactive environment.\textsuperscript{254} When traditional media – while emphasizing that they check the facts and aim to present the truth – fail to present accurate information, and even though it was the citizen journalists who ‘outed’ the falsity, that should not be idealized. As Myers states, bloggers do not worry much about getting the facts right because they only present them as provisional, time-bound and limited.\textsuperscript{255}

One last aspect to be discussed is: making money through blogs. Although, according to Technorati, more bloggers than ever seem to be generating money through their blogs, it is still not a majority.\textsuperscript{256} Levinson describes five ways of earning money through blogs: using Google AdSense, Amazon Associates, PayPerPost, PayPal donations, or placing ads on one’s blog on your own.\textsuperscript{257} Where Google and Amazon act as middlemen for advertising, the service of PayPerPost entails people writing blogs on a given subject and receiving money for it. This has been viewed critically and has not functioned within the blog

\textsuperscript{255} Ibid., p. 127.
\textsuperscript{257} Levinson (2009): p. 28.
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

community if the text revealed any advertising intent. Another common way to earn money seems to be holding speeches on topics the blogger posts about. In this case, it all comes back to expertise and reputation.

To gain status and to become well-known, bloggers use Twitter much more than the general population does. They use this tool to promote their blogs, bring interesting links to light and to understand what people are buzzing about. The next chapter will introduce this service known generally as microblogging.

6.2. Microblogging – Start Contributing!

Figure 4: Examples of tweets on Twitter (29th July 2010)

These are examples of so-called tweets posted by twitterers on the social platform Twitter. Twitter is a microblogging service. Kevin Makice describes the term microblogging as the publication of short messages reporting on the details of one’s life. Twitter asks, “what’s happening?” and allows users to answer in no more than 140 characters. “The throwaway answers include messages of context, invitation, social statements, inquiries and answers, massively shared experiences, device state updates, news broadcasts, and announcements.” People twitter to share interesting Web links, to report witnessed local news, to rebroadcast fresh information, to philosophize, to make brief direct commentaries to another person, to emote and vent, to pose questions, to get advice, to crowd-source, to meet new people, to keep informed, to stay informed of breaking current events, to build new applications, to build business, to promote business and also to talk to influential

258 Sussman, Matt (2009c).
259 Sussman, Matt (2009d).
261 Ibid. p. 3.
people and celebrities. Tim O’Reilly and Sarah Milstein summarize this diversity of possible topics by identifying Twitter as “a goldmine of ideas, feelings and conversations.”

The service was founded in 2006 by the creator Jack Dorsey and early funder Evan Williams. They describe the newest social media as “a real-time information network powered by people all around the world that lets you share and discover what’s happening now.” They identify Twitter as a platform that allows the user to influence what is being talked about around the world. Twitter is the fastest growing social media. While counting six million users in February 2009, Twitter registered 32 million by May 2009. Although there are smaller and unknown microblogging services, Twitter is, as Levinson states, the solitary titan in microblogging.

Two key design strategies were instrumental to Twitter’s success: its simplicity, and its granting developers access to the membership and content mechanisms. First of all, the strategy of simplicity refers to the fact that the core of Twitter’s service is very simple; and Twitter never tried to be more than that. All it does is to make sure that the short messages written by the members are distributed in a self-organizing network. Twitter’s revolution lies in its instantaneousness. The Sending and receiving of the brief messages is nearly as instant as their writing. Richardson describes it as the running river of conversation and ideas, blending the personal and the professional. Levinson adds by saying that Twitter is not only the most immediate written medium in history, but also the most integrated combination of interpersonal and mass communication. He describes Twitter as taking the classroom to a global level, where the student asking a question might be in an interpersonal conversation with the teacher, while the whole class is listening as a mass audience. Twitter can be depicted as a chat room, a classroom or a gathering that goes on 24 hours a day and seven days a week.

Unless someone protects his or her tweets, the posts can be seen by anyone. This is one of the big differences to social networks such as Facebook, where interactions can normally

264 Twitter: Twitter is a real-time information network powered by people all around the world that lets you share and discover what’s happening now. URL: http://twitter.com/about (29th July 2010).
only take place when the two people are friends. Twitter allows politicians or celebrities to talk to fans without being friends with them. Reagan even sees the advantage of now being able to “stalk” the people one admires.\textsuperscript{270}

The second strategy is that Twitter provides the technical information needed for third-party applications to exchange information. This way personal investment and creativity is the key to Twitter’s success. Twitter can and is being formed by developers and by its users. According to Twitter more than 50’000 third-party Internet and mobile applications exist.\textsuperscript{271} For example, one can update his Google calendar by sending a direct message via Twitter to @gcal. Another example is the service updating people about the traffic conditions in Austin, Texas. One simply needs to follow the Twitter user @AusTraffic and knows all about the most recent traffic situation. Many other applications have been developed especially to get a better grasp on the “twitosphere” as a whole. Such applications include sites showing the trending topics on Twitter (whatthetrend.com, twitscoop.com), or sites making it possible to see if one’s Web site has been tweeted (backtweets.com), or sites making picture uploads possible (twitpic.com) or the very important application for shortening URLs (i.e. http://bit.ly) so that they fit in a 140-character tweet including a statement accompanying the link.

Twitter’s strength is that it does not limit what people can do with it, nor does it presume to know exactly how everyone should interact with it. As Makice nicely concludes, Twitter is about emergence. Users compose their own information stream and tailor their own experience to their own wants and needs.\textsuperscript{272} Twitter itself asks all the members to do so: “Just remember, how you use Twitter is completely up to you. Follow hundreds of people. Follow a dozen. Post every hour. Post never. Search for your favorite topics and create lists. Or not. You are in control on Twitter.”\textsuperscript{273}

Different features, such as @replies or hashtags, have emerged and been institutionalized because of the users. Many twitterers started their tweet with @ followed by the user’s name they were answering to. It had gotten so common that Twitter linked the @[name] to that twitterer’s account. Hashtags refer to the #-sign which users had started typing together with the topic the tweet was about. This way it is easier to find all tweets related to that topic. For example: all members of a convention would include #[convention’s

\textsuperscript{271} Twitter: Twitter is a real-time information network powered by people all around the world that lets you share and discover what’s happening now.
\textsuperscript{273} Twitter: Twitter is a real-time information network powered by people all around the world that lets you share and discover what’s happening now.
name] in their tweet and, therefore, be able to find all tweets referring to that convention quite easily.

Twitter has various advantages. It is very recent, if not to say real-time. Inputs, even pictures of breaking events can be found very quickly. It is even harder to find tweets a day old than the ones a second old. Especially in situations like the demonstrations in Iran in 2009, Twitter can be called a powerful enabler of democratic expression, as people from everywhere on the streets were able to communicate to the world through their mobile phones.274 Another advantage of Twitter is that your followers often point to new sources and valuable information published elsewhere on the Web. O’Reilly and Milstein also indicate the importance of following not only friends and family, but also people you don’t know, because they help you develop a sense for another region, industry or social sphere.275 Non-mainstream information and specialist knowledge can be found by following the right people on Twitter. Another fact about Twitter is very interesting and might be the one that distinguishes it from other social networks, like Facebook or MySpace. According to O’Reilly and Milstein, Twitter is not so much a broadcast channel as it is a discussion channel. “Indeed, the secret of social media is that it’s not about you, your product or your story. It’s about how you can add value to the communities that happen to include you. If you want to make a positive impact, forget about what you can get out of social media and start thinking about what you can contribute.”276

6.3. Social Networks – Start Connecting!

And start presenting yourself! Social media and especially social networks have become an important avenue for self-expression and identity formation. Lemi Baruh and Levent Soysal view social networks in relation to the current notions of today’s society. To them an important characteristic of the current culture is the elevation of individualism and the rise of the individual experience as a guarantor of truth. The individual has increasingly come to the center stage of social, economic and technological order.277 In addition, society seems to become more and more defined by an expressive culture blurring the distinction between private and public. They maintain that gradually sociality is taking place in the

virtual world as well as intimacies are being carried into the virtual world. In such an environment it then becomes a crucial component as to how individuals develop and negotiate their own identity. Each person is responsible for creating a suitable persona of him- or herself. Identifying this task as the new individual’s labor of self-presentation, they take note, that for many individuals this has become a full-time job.278 In conclusion, the resulting person is as intimate as its public.

They define social media as “an all-encompassing term that describes loosely organized online applications through which individuals can create personas and communicate with each other.”279 A more detailed definition of social network sites stems from Lee Humphreys, who calls them “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.”280 For Levinson this species of social media has the very purpose of creating and developing social networks, thus enabling people to connect for whatever intention.281 Social networks give their members a single platform from which they can engage in a wide variety of activities such as private messaging, bulletins or group messaging, blogging, posting of photos, videos or music, as well as instant messaging and groups devoted to common interests. It is this social medium’s designation to bring people together. Groups and similar online activities such as forums or message boards are then a fundamental component of online life. Groups share and discuss links, texts, photos and videos. This interconnectedness implies that the most significant feature of any social network is the ‘friend’.282 These friends can be known solely virtually or be known in real life. Facebook and MySpace are the two dominators in the realm of social media.283

MySpace was launched in January 2004 and acquired by Rupert Murdoch’s News Corp. in October 2005. According to its own description, MySpace sees its commitment in providing “our global community with tools to discover, share, connect and communicate in every way they choose.”284 More than 122 million active users around the globe are

278 Ibid. p. 395.
279 Ibid. p. 393.
282 Cp. ibid. p. 123 et seq.
283 Ibid. p. 110.
participating in the community. MySpace’s music pages are especially revolutionary. Initiated in 2005, anyone can set up an account for a special kind of page showcasing uploaded mp3-files. Anyone can invite anyone to listen to his or her music. Where in former times one could either go play in a club or send demo tapes to record companies, one can now spread his or her music and present it to a recording company after having already generated some excitement in the public. MySpace, therefore, eliminates several middlemen or experts. In 2006 MySpace even started cooperating with SNOCAP, an online jukebox that offers free samples of one’s music together with a sales option for the full recording, where the musician can set the price him or herself. Kate Nash and Lily Allen are two examples of now internationally known artists who started out on MySpace. Kate Nash joined MySpace in 2006. Her first single had a limited 1000-copy release. By February 2009 her MySpace profile registered twelve million views; her album “Made of Bricks” sold one million times and became a No. 1 hit in the United Kingdom in 2007. Lily Allen started posting demos on MySpace in 2005 attracting fans and mainstream press. Her single “smile” was No. 1 in the United Kingdom and her album “Alright, Still” sold more than 3.3 million copies. MySpace enhances getting in contact with other people of the same interest or fans, not only in the music field, but also in fields such as poetry or comedy. Thus, it does encourage being friends with people unknown to one in the real world.

Facebook, on the other hand, according to Levinson, has a much higher ratio of real-life friends. Developing from being a student network, helping to link easier throughout the campus, this real-world grounding continued even after opening up to the whole world. Mark Zuckerberg and the co-founders Dustin Moskovitz, Christ Hughes and Eduardo Saverin launched Facebook in February 2004 from their Harvard dorm room as a tool for students. By December of that year the network already counted one million active users. In September 2006 Facebook expanded letting anyone join. Since then, Facebook has grown at great speed, is provided in 25 different languages and was able to count its 500 millionth active user on July 21st 2010. The company’s own description identifies Facebook as “a social utility that helps people communicate more efficiently with their friends, family and co-workers. The company develops technologies that facilitate the

286 Ibid. p. 114 et seq.
287 Ibid. p. 116 et seq.
sharing of information through the social graph, the digital mapping of people’s real-world social connections.”

With its headquarters in California, the company has more than 1400 employees. According to their statistics, people spend over 500 billion minutes per month on Facebook. The average user is connected to 60 pages, groups and events, and creates seventy pieces of content each month. In summary, more than 25 billion pieces of content, such as Web links, news stories, blog posts, notes or photo albums, are shared each month.

To Levinson, Facebook is not just an interactive knowledge base, but also a live, real-time knowledge resource. Questions can be asked and answered in real-time and, in addition, small groups and their discussions take on the notion of an online colonial pub, where time and distance have no relevance and meetings can take place over days, weeks or even years. Obtaining information has become easier than at any time in history. Compared to MySpace, he sees Facebook’s potential in becoming a vehicle for social causes. This could be seen in February 2009 as twelve million people in 190 cities around the world were out on the streets protesting against Columbia’s terrorist group FARC.

There is one hidden dimension to all the self-productions through social media and that is the endurance of everything once it has been written or created and posted. Publications can have an effect which might have been unintended or unforeseen at the time of its posting. Users leave traces of data, which is being continuously collected. This collection happens automated, invisibly and mostly involuntarily. Levinson notes that most members see their MySpace or Facebook profile as personal and not professional. Users might feel that the social media is theirs, because of the extraordinary power of production and self-projection it provides. In the end the medium Facebook is Facebook’s, MySpace is MySpace’s. The automated collection and interpretation of the data result in the individual’s uncertainty and, as previously stated, the debate about privacy in the virtual world is gaining momentum.

---

292 Ibid. p. 125 et seq.
293 Ibid. p. 130.
6.4. Wikis – Start Editing!

“A wiki is a system that allows one or more people to build up a corpus of knowledge in a set of interlinked Web pages, using a process of creating and editing pages.”296 (Original emphasis). According to Macnamara, wikis are another key application of the Web 2.0. There are two claims about the origin of the term wiki. One identifies it as a Hawaiian term meaning quick or fast. The other definition maintains that wiki stands for “what I know is”.297 This second description points out the sense and objective of a wiki. With this technology, anybody can contribute equally to a joint online publication. Wikis are based on the convention that contributors can directly “post whatever they know about a subject for others to confirm, clarify, add to, or correct rather than all content being approved by a ‘Web Master’ as occurs with traditional Web sites.”298 Centralized production and top-down methods of knowledge sharing are being pushed aside by the belief, and the new concept, that everyone together is smarter than one alone. Macnamara detects a fundamental paradigm shift being underway. The times when knowledge needed to be vouched for, authorized and approved by experts before it could reach the wide audience seem to be passing. Wikis are characterized by emergent properties in the media and public communication.

Wikis’ objective is to produce collective knowledge. Therefore, they emphasize the participation, the contribution and collaboration of the users. According to Richardson, the key word for wikis is “easy”. Anyone can edit and contribute to a wiki, requiring it to be easy to handle. Wikis normally have a link named ‘edit’, through which anyone can start writing and adding to the content making oneself the editor in chief.299 They closely follow the open-source software ideal, which entails that the quality of the collectively produced product is more important than owning the idea or the code. For Richardson, “wikis can play havoc with the traditional ideas of copy-right and intellectual property.”300

Whereas blogs are good for discussions, wikis are not as ideal for carrying out a conversation about ideas. Most discussions, concerning which article and what entry needs to be changed in what way, may happen parallel to it. However, this procedure shows how the collaboration is made transparent and coincides with the ideal of open-source knowledge gathering. For Macnamara wikis can revolutionize science by sharing

298 Ibid. p. 42.
300 Ibid. p. 59.
information more quickly and openly or by allowing for feedback, comments and even collaboration without having to first publish any books or articles in journals.\textsuperscript{301} For Myers, using wikis is not simply creating a new genre of publication and knowledge sharing, but creating a social world.\textsuperscript{302}

The most common example of a wiki is Wikipedia. Referring to Richardson, wikipedia.org is the most important site on the Web.\textsuperscript{303} And Levinson identifies it as the most thorough-going, consistently user-driven system on the Internet.\textsuperscript{304} Wikipedia was founded in 2001 and has grown to attract 78 million visitors monthly as of January 2010.\textsuperscript{305} The founder Jimmy Wales describes the service as following: “Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That’s what we’re doing.”\textsuperscript{306} Wikipedia encourages and maximizes participation, making the sense community its essential component. Anyone can contribute to the online encyclopedia. Age, gender, education, location: none of these factors play a role. An impressing example is given by Richardson when looking at the entry made for the horrific event of the tsunami in December 2004 in Asia. The first 76-word entry was created about nine hours after the happening. Only 24 hours later, the post had been edited more than 400 times and had grown to about 3000 words including photographs, charts documenting the dead and injured as well as other graphics showing how the tsunami had spawned. 48 hours after the first posting, the entry entailed more than 6500 words and had been edited 1200 times containing more than a dozen graphics including a video of the wave. Six months later, after 7000 changes, the post settled at around 7200 words.\textsuperscript{307}

Wikipedia itself indicates that older articles tend to be more comprehensive and balanced, while newer articles more frequently contain significant misinformation or even vandalism. “As a wiki, articles are never considered complete and may be continually edited and improved. Over time this generally results in an upward trend of quality and a growing consensus over a neutral representation of information.”\textsuperscript{308} Levinson states that there is an ever-waging war on Wikipedia between those who attempt to make and keep entries truthful and those who seek to disrupt this process. He describes how he fought for

\textsuperscript{301} Macnamara (2010): p. 43 et seq.
a few hours against someone constantly trying to change the name Perikles into Pickles.  

The openness and the dependence on the community then become the strength as well as the vulnerability of Wikipedia. Myers maintains that Wikipedia depends on the ability to fight off vandals, and this again depends on the willingness of a community to share its goals. “Were this willingness to fade, the most contentious articles on Wikipedia, the articles on abortion and Islam and evolution, would be gone within hours, and it’s unlikely the whole enterprise would survive a week.” (Original emphasis).

Although Wikipedia flattens the hierarchies by letting everyone contribute, the community does build its own kinds of authorities in a self-organizing way. An example of this is the Wikipedia administrators that have been chosen by a public discussion and a consensus of editors. They have access to some significant administrative tools and can delete articles, block accounts or IP addresses and edit fully protected articles, which can no longer be edited by the public. Then there are the bureaucrats, a super-kind of administrator determining if the consensus has been reached to promote an editor to an admin and providing him or her with the necessary tools to do so. According to Wikipedia, there are not very many bureaucrats.

Although these hierarchies and types of controls as well as defined standards have become institutionalized, Wikipedia depends greatly on the community to function successfully. Wikipedia can set up the standards, but then again it is up to the contributors and readers to apply them. In this sense Macnamara defines Wikipedia as a “self-correcting adhocracy”. Also Richardson indicates the importance of the group as a community watchdog providing security. In this process the hyper-transparency of Wikipedia is important. The complete history of each entry is easily available for everyone to see all the changes that have been made previously to the article. This is one of the main differences to blogs or social networks, where changes cannot be detected. The open-source generating of knowledge seems to work fairly well when believing a study done by Nature in 2005, claiming that in Wikipedia articles the average number of inaccuracies was four compared to the average of three in Britannica articles.

However, the success of Wikipedia triggered what Richardson calls a wiki revolution. Wikis seem to be accessing many areas of life. For example, anyone can add his or her favorite recipes to wikirecipes.net or their best vacation bargain to wikitravel.org. Even one of Switzerland’s biggest supermarkets, Migros, recently launched its own wiki called Migipedia, where people can exchange experiences made with Migros and Migros products and can also rate the products.

Another wiki, which could be of interest to journalism, is WikiLeaks. Founded by Julian Assange in 2007, WikiLeaks tries to give whistleblowers, journalists and activists with sensitive information a platform to publish it on. Here, not like on Wikipedia, articles and source documents are kept pristine. The document summaries are usually written by WikiLeaks staff, sometimes in collaboration with the submitter. WikiLeaks is a project of The Sunshine Press following the idea of improving transparency in institutions and especially in governments. “We think better transparency is at the heart of less corruption and better democracies. By definition spy agencies want to hoard information. We want to get it out to the public.” The latest hit by WikiLeaks happened on July 25th 2010, when it released over 91’000 reports written by soldiers and intelligent officers about the war in Afghanistan. WikiLeaks worked together with Spiegel Online, The New York Times and The Guardian, which analyzed the raft of the mostly classified documents and published articles about them on July 25th as well, creating political pressure.

6.5. Folksonomies – Start Organizing!

The interactive and participative possibilities of the Web 2.0 also have their effect on the way people can organize and share their own online sources. Tagging and the emerging folksonomies are the result of people describing or labeling Web content. Smith defines folksonomy as the popular term describing the bottom-up classification systems that emerge from social tagging. Tags can be identified as metadata about a resource,
meaning information about information. According to the National Information Standards Organization (NISO) metadata is “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource.”

Tags can be used in different situations. For example, the service Flickr allows users to upload their photos, sort and organize them through tags. Social bookmarking sites entail being able to add tags to one’s bookmarks. By adding whatever keywords suit the user best, each user creates his own collection of links that are being categorized through these keywords. Del.icio.us was the first social bookmarking site and started its service in 2003. Whenever a Web user finds a Web site that is important to him to save or mark, he can do so by describing it through keywords, aka tags. This way he adds metadata to the online source, creating multiple and especially personal ways of finding it again. In addition, the tags and sources are being shared with the whole community using Delicious, thus making it possible to find new resources containing the same or similar tags. Tagging improves the find-ability of resources by using one’s own vocabulary and by enabling everyone to organize their link collection their own way. In tagging all keywords can be chosen freely and are treated equally with no hierarchical background. As an example, a student could label a source with the tag “toread” and thus organize all the assignments needing to be done. Or, one could tag something found in an online store with “momsgift”, making it easy to find it again, when the gift is needed.

These user-added keywords are the fundamental organizational elements for the emerging folksonomies. The term folksonomy originates from the term “taxonomy”, which means to classify things according to certain relationships and categories. Add the word “folk”, meaning the common people or the greater masses, and it then refers to the concept that categorizations of resources are not predefined but evolve from the people’s usage of resource descriptions through tags. While traditionally, metadata was created by professionals following strict taxonomies and preset vocabularies, the categories based on user-created metadata are more flexible. As Milan Stankovic and Jelena Jovanovic mention, taxonomies are hierarchical and exclusive. An object can only belong to one unambiguous category. Folder trees, known from filing documents on one’s own

320 Ibid. p. 5.
computer, depict this feature. Tags, on the other hand, are neither exclusive nor hierarchical. The three entities – tags, users and resources – constitute what is called a folksonomy. The term was initially used by Thomas Vander Wal and expresses what he calls the organic system of organization. Aggregating the tags and inferring relationships from them create patterns and categories. Adam Mathes indicates that folksonomies are more like a synthesis of similarity than a systematic arrangement. It is a grassroots community classification of digital assets. Different authors have illustrated folksonomies as foot-worn paths, which start to emerge through use and are paved after originating, ensuring optimal utility.

As with wikis, the openness and user-generated data are the advantage and the disadvantage of folksonomies at the same time. Its simplicity and low entry barriers encourage people to actively participate in tagging and thus adding meta-data to the Web. It is a very easy process that takes no additional skills because each user can use his or her own vocabulary. While the traditional and professional creation of metadata is time and effort consuming, folksonomies can keep up with the vast amounts of new content continuously being created on the Web. They allow quick adaption of new terms when traditional vocabulary is missing. As Smith notes, folksonomies are “simply a serendipitous spin-off benefit of people tagging bookmarks for their own personal use.” Mathes specifies that browsing the interlinked related tags is wonderful for finding things unexpectedly in a general area. He claims that there is a fundamental difference in the activities of browsing to find interesting content, as opposed to direct searching to find relevant documents in a query. The immediate feedback of the tagging systems is a fundamental reason for this. When tagging an item, a cluster of items carrying the same tag is immediately shown and the real power of such systems emerge when including all items from all users that match your tag.

However, there are limitations and drawbacks resulting from the democratic way of labeling Web content. Tagging’s nature, and therefore that of folksonomies, is

324 Ibid. p. 2 et seq.
330 Ibid. p. 9.
fundamentally chaotic, often resulting in problems of imprecision and ambiguity because there is no predefined vocabulary to be used. Tags like ‘apple’ can refer to the computer company or the fruit. Acronyms can similarly be a problem, such as ‘ant’ being the insect or the acronym for Actor Network Theory, stemming from social sciences. Using more than one tag can help in finding and labeling the resource more precisely. Additionally, there is no control over synonyms. The same resources can be described in different ways. So far, tagging systems store no information about two tags relating to the same concept. This is even more obvious when trying to tag with multiple words and spaces. Since tags are designed to be one word, users usually try to solve the problem by connecting two words with different strategies. A tag then could be written “likeThis” or “likethis” or “like-this” or “like_this”, all referring to the same source, but not being recognized as the same description. The semantics are still missing, leaving only limited information about the context. For example, do the tags “Madonna” and “music” point to a review or to the actual music?

All in all, folksonomies are transforming the creation of metadata for resources from an isolated professional activity into a shared, communicative activity by the users. This shift is of dual nature and also causes some difficulty. On one hand, tagging is a great system for individual organization, at the same time there is an inherent compulsion to share in order to generate folksonomies and to reveal the full and useful power of the system for the user. Smith identifies four different tension points intrinsic to tagging systems: One, he mentions personal versus social tagging, which refers to the objective one follows in tagging. Who do people tag for? Two, the debate about idiosyncrasy versus standardization includes the question of whether tags should be completely unique or follow a standard to be used for search. Three, the tension between freedom and control is contained in the discussions about how much the user should be influenced and controlled when tagging. And four, weighing amateurism against expertise deals with the question of how qualified people need to be in order to tag. However the dualisms may be weighed, folksonomies are aggregated through huge amounts of metadata created by the users. The fundamental difference to traditional classification schemes lies in the reduced complexity. Stewart

---

331 Ibid. p. 5 et seq.
Butterfield, one of the creators of Flickr, concludes adequately: “Aside: I think the lack of hierarchy, synonym control and semantic precision are precisely why it works.”

6.6. Search Engines – Start Finding!

“The web is where society keeps the sum total of human knowledge.” As stated by Ian Witten, most, perhaps even all, of what people need to know is on the World Wide Web, and if not today, then it will be tomorrow. The Web is growing fast. According to Sebastian Erlhofer, experts make a cautious guess that the amount of Web sites worldwide doubles within six months. With this current situation, finding relevant information soon becomes central for every Web user. To Elizabeth Van Couvering, search engines are the obvious place to begin the analysis of Web content. Search engines have become a new gatekeeper and mediate between society’s treasure trove of information and the consumer. Search engines determine how the information stored on the Web is being accessed.

All search engines consist of doing three main activities: First, finding the newest information and Web sites within the Internet; secondly, indexing them on their servers and thirdly, finding and ranking the relevant results to a person’s query. Firstly, a search engine can only find Web sites that are already registered in their index on their own server. Special software tools continuously search the Web to find the newest information and Web sites. These tools are called crawlers, spiders or robots. They “crawl” from page to page searching for links to file in the search engine’s index.

Secondly, the Information Retrieval System then structures the crawler’s findings and analyzes the Web sites for their content. This procedure is called Document Preprocessing and includes the analysis of the words and their order on the specific homepage. The content analysis and the structuring of the results together constitute the whole process of

---

337 Ibid.
indexing, which is the core of a search engine. According to Monika Henzinger, search engines were indexing tens of billions of Web pages per day in 2007. The third part of each search engine, the query module, processes a Web user’s query. It searches the index and shows the keyword-relevant results according to the algorithm of the specific search engine. The algorithm is the key to the order by which the results are sorted.

In the early stages of search engine technologies, the ranking of the results mainly occurred through large-scale full-text searching. These techniques worked by counting words, weighing them and measuring how well each document matched the user’s query. This attempt functioned for detached information sources, but the sites online are hyperlinked. So today’s search engines try to acknowledge this fact by counting and weighing the links as well. Searching the Web is combining full-text searching with link analysis. Link analysis follows the simple principle of judging a page by what other pages say about it through their linkages. Not only does the number of links pointing to a page count, but also the prestige of the page linking to it. A more “prestigious” site, therefore, weighs more in its recommendations and links to other sites. This concept was initially implemented by the Google founders Larry Page and Sergey Brin, giving their ranking algorithm the name PageRank. PageRank’s success was responsible for elevating Google to the position of the world’s preeminent search engine.

The use of a site’s hyperlink structure, to rank it within the result lists, together with the full-text search are the main pillars on which search queries are being dealt with. Of course, each search engine’s exact algorithm goes into a lot more detail and is a well-kept secret. In the end, it is their commercial trade secret. Witten describes the search engine business as extremely volatile. And Van Couvering adds that only a few large players dominate the industry. comScore reports that in the United States as of May 2010 63.7% of all searches happened through Google. Yahoo! followed with 18.3%, Microsoft sites

345 Ibid. p. 142.
346 Ibid. p. 137.
347 Ibid. p. 108.
treated 12.1% of all queries, Ask Network and AOL also belong to the top five, but only registered 3.6% and 2.3% of all queries done.349

Using search engines entails conquering several hurdles, as Erlhofer detects.350 One of them is that the results found are not live or real-time, seeing as Web sites need to go through the whole indexing process. Furthermore, one finds a number of so-called dead links, which do not exist anymore or are temporarily not available. A third difficulty is the vast number of results that are usually listed to one query. The user is forced to rely on and trust the search engine’s perceived relevance. To search more exactly, there are different symbols and signs that can be used to limit the number of search results. Google lists different tricks, such as searching for exact phrases by using quote symbols or searching for a specific site by adding ‘site:’ and the Web page’s URL to the query. One can exclude terms by using the minus sign or ask Google to fill in blanks by adding the *-symbol.351 There are many more tricks in order to limit the results to the relevant ones in answer to the query.

The fourth difficulty lies in the commercialization of the search engines’ service. Van Couvering also emphasizes that the search results are not objective results but shaped at every stage by commercial processes.352 The most common type of advertising is the new kind of search-specific advertising, which involves linking the query terms to the ads shown. Van Couvering calls this a new and very targeted ad vehicle that search engines have developed; and it is commonly known as paid performance.353 The concept was criticized since search engines are supposed to be objective in their results, causing paid results to now usually be marked as such. These ads are sold on a cost-per-click basis, the cost depending on an automated auction system. The price then depends on how many advertisers are bidding on the same search term. More specific search terms can be cheaper than more general terms. According to Van Couvering, there are two major advertising networks: Google’s AdWords and Yahoo!’s Overture.354 The different hurdles and the commercial aspect have shown that the end user’s evaluation of the results is important. However, Witten detects that most users trust their own ability

353 Ibid. p. 113.
354 Ibid. p. 113.
as Web searchers. He notes that the less Internet experience users have, the more successful they regard their own searches. He warns against a dangerous vicious circle when users believe they are capable searchers precisely because they are uncritical towards the results. He appeals to users to recognize Plato’s dilemma of knowledge, which entails that you cannot tell when you have arrived at the truth when you do not know what the truth is. There is a need for users to observe the results critically in consideration of the commercialism and also, as he mentions, of the search engine’s ability to track someone’s search behavior.

The collective memory of the world, that is the Web, is in constant flux. The information found through search engines should no longer be viewed as objective, but as having a social character. Optimizing the results and generating a better search engine in this social environment requires that search engines get to know the users’ needs. Machine learning seems to be one way of tackling this problem. “Learning techniques have the potential to continually improve their performance as new data are gathered from the users of a search engine.” Information is gleaned from the user’s behavior after being confronted with the results. Using full-text search followed by link analysis might make the integrating of user-feedback to be the next wave in search engines’ development. To search engines the query log is a mine of valuable information and it is the search engines that see what the world wants to know. They silently observe the requests for information and hold the monopoly on this information. Collecting data for search engines’ improvements shows the observable trend towards the personalization and the notions of the Web 3.0 as discussed earlier. As Witten states, referring to search engines as Web dragons: “If the outcome improves the results of your searches, you might well be prepared to share this information with the dragons. After all, you’re sharing your queries.”

---

356 Ibid. p. 23.
357 Ibid. p. 126.
358 Ibid. p. 126.
7. **The Responsible and Desirable Use of the Internet for Journalistic Research**

The objective of this chapter is to answer the question of the possibilities and limitations of the Internet in accomplishing a professional research in journalism. The first two subchapters will summarize what has been elaborated on concerning professional standards and the momentary use of the Web as a research tool in Swiss Online news outlets. The third section ought to combine this information with the discussed current features of the Web and examine what tools are useful in what way so as to meet the professional journalist’s needs.

7.1. **The Requirements of Professional Journalistic Research**

Research in journalism is done to meet certain professional demands and standards. Even though professionalism in journalism is a highly pluralistic construct, certain values have been set as an orientation to what it means to be a professional journalist. Three of these are directly connected to research, or more precisely, are to be achieved by doing research: Firstly, journalists seek the truth. They are led by the conviction that the public has the right to know the truth. Secondly, journalists try to defend the freedom of information and the right to criticize and therefore, search inexhaustibly for the relevant information and its publication. However, they only publish information that is known to be true by checking the sources. This third value entails the basic but vital conception of treating every source and item of information with a sense of skepticism and reservation. Knowing the source and checking its reliability is the basis for secured and trusted facts. If an announcement is unconfirmed it must be marked as such.

To accomplish these values for professional work, journalism has created a methodical framework guiding the research efforts. The famous wh-questions compose what exhaustive research needs to be able to answer, especially the first four – who, what, when, where – being essential and basic. The related and interpretative questions of why and in which way, or how, an event has occurred might not always be definable and their answers try to heighten the comprehensibility for the target reader. In contrast, it is imperative to successfully deal with the last ‘Wh-question’, that of “What source is the informant?”.

---

359 Cp. chapter 2.2.
7.2. The Momentary Use of the Internet in Swiss Online Editorial Offices

The use of the Internet for research purposes varies among the different news outlets, depending on their size and type. There are three main issues influencing the online research being done, which are not mutually exclusive but rather interconnected: time, technological knowledge and attested reliability.

Time and technological knowledge coincide to the degree that journalists do not often have the time to actually get to know some of the Web tools available, much less their actual possibilities and limitations. This can be observed with tools such as Twitter and especially social bookmarking or even blogs. The lack of knowledge and time does, in some cases, cause journalists to abandon a source rather than to look at it more closely.

The factor of time also influences the way acquired information is treated. Especially online journalists find themselves constrained by the little amount of time on hand to publish articles. The ideal checking of sources may be suppressed by a sole check of plausibility. This can be seen when using very recent information coming from a source like Twitter, but also for basic information stemming from Wikipedia or other expert Web sites – or at least Web sites perceived as such. Some news outlets in these cases act more cautiously and conservatively than others. Still, the Internet is used for research as it saves time as well.

The core reason for hesitation towards an online source seems to lie in its reliability. Knowing the source and the informant is – as discussed – essential to a professional journalist’s work. The Web technologies have opened public communication to everyone, making it harder for a journalist to find the relevant and trustworthy facts. By means of the knowledge collected about the Web 2.0 and the different currently used Web tools, the following pages attempt to reveal which tools can be of use and of help to an online journalist trying to achieve a high level of professionalism as well as which limitations define the momentary situation online.

7.3. The Internet’s Possibilities for Credible and Efficient Research

Today’s Web is the user’s Web. It increasingly emphasizes participation and collaboration through the different tools and applications it offers. It has become the new communication platform, which flattens hierarchies and connects people with no constraints of time or space. The various social media stress the importance of the community and most applications work better the more people there are using them. Through the Internet people
have found new ways of collaborating, contributing, sharing and discussing, thus making the Web gain more and more significance over time.

One of the reasons for the dynamic increase of the Web’s meaning in social communication is its emergent nature. Many technologies that have been introduced to the field of social media have been formed into what they are by the users and their way of using the tool. The Web’s possibilities in regard to communication are constantly being explored and formed by the users and their feedback to the developers. Twitter as well as Wikipedia are very good examples of this development. Web applications have become never-ending stories requiring the user to write their next chapter. When a new tool is introduced, nobody can really say what it is actually capable of doing. What are its advantages? What are the dangers? How trustworthy is it?

The virtual world of communication that is evolving seems to be quite similar to the real world in many ways. Anybody can say anything, anytime, anywhere. Although simultaneously there is one big difference: On the Web anyone, from any place in the world, can read what that other person has said, and this at any point in time. Web publications, ranging from professional articles to short comments or tags, last and are traceable. Endurance is an inherent characteristic of this new medium, including all the possible advantages and disadvantages that go with it. The storage of all that has been said results in a great information overflow and demands the competence of knowing how to sort through this sea of collective chatter. A new requirement is demanded of the user to not just consume and to not just contribute but to also evaluate and navigate through the muddle of the Web by him- or herself.

The core characteristic of today’s Web is its openness. And it is this feature that is its strength and its weakness at the same time. The participation of all communicators and readers involve getting hints and knowledge one would not have found otherwise. Being open to everyone then again makes one vulnerable for abuse. However, interactivity has gained immense importance in public communication like it has never had before, simply because the means to let everyone interact had been missing. It is questionable whether the traditional view of the journalist, standing aside the whole mass of people chattering and trying to get an “objective” picture of the events, is still realistic in this new form of public communication or not. Seeking to be detached is contrary to the notion of the Web 2.0. Everyone, starting from John Doe to the expert in a specialized field, is using the Internet. Chris Anderson’s theory has shown how niches have gained importance on the Internet. The phenomenon is too complex to condemn it as not trustworthy enough or irrelevant to a
journalist’s needs. As in the real world, it’s all about finding the right people to follow, to trust and to listen to in addition to waiting for good leads to new sources with interesting information. Considering the emergent nature of the social media tools and the speed of their development, the only way for journalists to really find their jewels within the vast information network is to participate in it themselves, just as they do in the real world. Of course, this takes time. And time, especially in online journalism, is rare. The next pages strive to reveal some inputs on what tools are more or less valuable to answer a journalist’s famous wh-questions integral to professional research.

7.3.1. Answering the Wh-questions

The four Wh-questions referring to the basic facts of a story are who, what, when and where. Some tools might be of greater help in doing these examinations, others less. Twitter, for instance, is a great tool for finding very recent information. Events that just happened a second ago can be read on Twitter at almost the same time they have occurred. Pictures are sent throughout the whole world within seconds. In addition, needing only an Internet access, anyone can publish information, making censorship all the more difficult. However, its promptness is also its disadvantage. Information published as quickly as on Twitter can hardly be checked. Consequently for a journalist, using information from Twitter either entails trusting the sender fully or doing the checking him- or herself. Twitter is described as a running river of ideas and conversation. Everyone is chattering, and they twitter especially when something unusual has happened. For that sake, Twitter might give insights or flashes on what the incident must have looked like. The little bits of information might hint to facts, like where something has happened or when. In the end, despite that, it is not a river of facts flowing by for journalists to pick from. Social networks, on the other hand could become a tool for information referring to the who of a story. Especially the music pages on MySpace have been created and are being used by musicians and, most of all, newcomers to present themselves to the (music) world. Thus, it is in their own interest to be seen in a true context. Of course, wanting to attract attention could entice one to present oneself in a different or better light than what is real. Still, for basic information about a musician, especially the ones that have succeeded and have professionals taking care of their online profile, this platform could become an institutionalized source.
Facebook is perceived as being more personal than MySpace. Therefore, most users are not interested in presenting themselves objectively correct, but rather in the way they want to be seen. It is a platform of self-expression and identity formation on which anyone can create their persona and engage in social discussions. Especially celebrities or even politicians have discovered the importance of embracing this platform for self-representation and engagement. Taking information about publicly-known people from Facebook then will hardly differ from the information retrieved by contacting the person’s management. Taking personal facts from profiles of unknown people, however, could be more critical since they have not been worried about their perception in public and may have joked around. Furthermore, many people protect their profile so that it can only be seen by friends, hindering journalists from taking information from it.

Wikis are usable depending on how much one trusts the wisdom of the crowd. Wikipedia’s advantage lies in its hyper-transparency. All changes that have ever been made to an article are traceable. One can see how many people have contributed, added, changed and corrected what has been written. Discussions about the article can be followed. As Wikipedia itself states, older articles are usually more balanced and comprehensive than newer articles, which might include misinformation or are a target of vandalism. When trusted, Wikipedia is a great source of information to answer questions such as who, what, when and where. However, it remains a matter of trust. The inquiry whether the community really is smarter than one alone cannot be answered generally. On the one hand, studies show that it is quite trustworthy compared to other encyclopedias as the willingness of the community has formed it into the Internet’s most important site. Yet on the other hand, all the standards that have been set up to make it a reliable source are dependent on the user abiding by them.

Compared to Wikis, blogs have become more a platform of discussion and debate. They can be seen as a forum and marketplace of ideas contributing to the identification of the current situation. This can happen in any field of interest, be it political, scientific or about topics of everyday life. Being a discussion tool, blogs might not just answer the questions about the who, what, when and where but could actually be used as a good information tool to learn about the background and to analyze the why and how. Blogs are interlinked and thus present great amounts of information. The difficult part, again, is to find the trustworthy blogs. Who are the expert bloggers? Who are the people that do know about the topic and are using blogs as a new way to communicate their knowledge?
According to Technorati, the percentage of professional and expert blogs is rising; and many bloggers have mentioned self-expression and the sharing of expertise as their main reasons for blogging. Even then, their primary objective might not be getting the facts right, but expressing their position. In a sphere of discussion it’s all about taking part in the debate and presenting one’s knowledge and attitude. While some casual bloggers might do this for their own pleasure and personal gain, others have built up a reputation in the blogosphere that needs to be cultivated and that could quickly be destroyed if they were to publish misinformation. Especially in specialist fields an ignorant blogger would not be welcomed by the community and be reproved. Moreover, there are well-known and respected personalities in the real world, who are using blogs as a different medium to communicate. Here too, it would be false to ignore these messages simply because they stem from a blog. So all in all, blogs can answer the wh-questions acceptably, but one needs to identify them, because the great majority of blogs do not.

Locating the right sources, in general, is essential in using today’s Web. It was shown, that most journalists, if not to say all, mainly use Google to find their way through the Web. This fact emphasizes how search engines have become a new gatekeeper of information by determining the user’s access to all Web content. However precise and good they may work, it still remains important that the user should definitely evaluate the results because of the various constraints that influence a search engine’s ratings. Commercialization and the volatile business field lead to the fact that the results are affected by economic processes and are not truly objective. In addition, the Web as a collective memory of human knowledge is, naturally, continuously in flux. Its information and the selection of information then is a social process, which is constantly under way. And, the Web is growing too quickly for the machines indexing the Web sites to keep up. Exceptionally new information might not be found through a search engine.

On top of all that, search engines work well when the user knows quite exactly what he or she is looking for. Open facts concerning the wh-questions will not be answered directly through the search engine, because it is only a tool pointing to the right source. However, one needs to know somewhat about the topic to be able to search successfully for the needed information. If not, search engines’ results can easily overwhelm the user and it will take time to click through the links and the further links on the found pages to fully comprehend the information found.

---

When only few keywords are known and the information is very recent, folksonomies can be of help. By using the user’s vocabulary instead of predefined taxonomies one can find Web content through common expressions more easily. In addition, social bookmarking allows a user to discover many more sites also related to the query term without further knowledge in advance. Through social bookmarking a user does not only open multiple and personal ways of finding information, but the shared metadata also results in finding sources unexpectedly. With tags and folksonomies one can start browsing the Web without searching explicitly.

These two search methods – search engines and folksonomies – can be of help for any of the wh-questions because they point to the relevant information, be it facts or background information. However, the problem with both tools, again, is the vastness of the Web content. Search engines cannot keep up with the fast growth of the Web and users only pay attention to the first few hits in the thousands of results shown. Social bookmarking sites are not complete either, because they are dependent on the user to continually mark and describe the Web pages found. European sites, and especially Swiss sites, are not greatly represented. The tool is not known well enough, and as mentioned earlier, Web 2.0 tools work best the more people there are using them.

It appears that the potential of the different tools for research vary, but there is one common notion underlying them all: They are open to anyone and everyone. That is what constitutes them and makes them work in the first place. So for a journalist, using facts from a source – be it offline or online – greatly depends on knowing the informant. It depends on who says it, and not as much on through which medium. The Internet has accelerated journalistic practices, forcing online-journalists to find these reliable sources more quickly. As of today though, the problem of trust in online sources is the main drawback for journalists, preventing them from using these tools more extensively.

7.3.2. The Problem of Trust

As pointed out, knowing and trusting the source is vital in (online) journalism. Being able to answer the wh-question of what source? is one of the pillars for professional work and distinguishes journalism from the simple announcing of unconfirmed information. The Web might have helped to overcome barriers of time and space when doing research. However, in being an open platform it has not helped in finding the trustworthy sources; actually it has even made it more difficult to identify them. Yet, concluding to ignore all
Web content would be wrong, because just as the untrustworthy, the trustworthy and valuable information is online as well. The Internet itself might be over twenty years old, but the times when users themselves are creating the content have only just begun. The platform is still gaining momentum and experts just as laymen are discovering all the different possibilities, advantages and disadvantages. The virtual world of communication is getting more and more important, thus meaning for online journalists that it has become essential know how to find the credible sources.

At the moment this is still quite difficult to do. The advantage of the Web being open to all makes it prone to manipulation. There are two fields of research connected to the semantic Web trying to improve this situation: personalization as well as recommendation and reputation systems.

Approaching a Web 3.0, personalization has become the new buzzword. Through semantics the Web will turn into a Web of data, leaving each person to combine and aggregate that data through applications and agents in a personal way. Based on the personal history of habits and interests the machines will start learning from the user’s behavior. They will add new data to the existing, thus striving to achieve more specific and more accurate search results. Witten describes how individuals only use a tiny fraction of the Web when searching, but global ranking techniques are applied.\footnote{Witten (2007): p. 217.} The integration of personal data would narrow and specify the results essentially. Integrating personal data in the field of online journalism could then entail excluding certain sources that over time have been evaluated as untrustworthy and focusing on experts’ or politician’s contributions. Personalization here would mean moving and searching in the field of the voices also seen as trustworthy in the real world, but now making themselves heard on the virtual communication platform.

Personalization could also become important for journalists working in a specialist field by generating inputs of new information within that community. Machines could help find new information by knowing the preferences of the user. Thus, the journalist does not just search the Web when researching on a specific topic but could be guided to new interesting information through automatically generated hints. The more the applications learn about what the journalist evaluates as useful and trustworthy, the more it will provide the journalist with similar information. This would mean that the time needed to evaluate the new information would decrease in the long run. The question then is: is the system open
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

enough to find new valuable content and not focus on the same authors and sources over and over again? And, how does it go about finding new sources, without having all unusable information on the same topic integrated as well?

In the end, personalization does not completely solve the problem of trust. The second attempt to conquer it lies in the recommendation and reputation systems. Be it virtual or not, reputation is what makes people trust. These systems also work with the metadata created by the users judging content or people. Again, the more people that ‘vote’ the more accurate the outcome and the judgment are going to be. According to Witten, reputation evaluation methods differ in respect to what is being measured, ranging from the history of uploads, the satisfaction, the level of user participation and the popularity of the content.\textsuperscript{362}

In the end, a reputation will be built through the history of interactions and recommendations made by the community. People’s trust will be determined by the prestige and the perceived integrity of the Web site, the institution or the person.

For these methods to work reliably, they are dependent on user participation. Even if the technology in years to come will be developed well enough to promote a wide use of such recommendation tools generating online reputation, in the end it still remains a question of trusting the crowds. Is the crowd more intelligent and reliable than one person alone? This notion stands opposite society’s traditional views on reliable sources and knowledge. Trusting the experts and relying on hierarchical structures has influenced and defined public communication, public reputation and thus the way journalism has functioned.

Journalistic research is seen as a craft including the following certain rules and habits of professional work. Referring to experts or other trusted sources is like a security net in a journalist’s work. Reputation and trust is attributed mainly to an individual by means of his or her history, position or knowledge. Trusting a crowd is something fundamentally new, not just to journalists but also to society in general. Having a communication platform is not just opening new possibilities in sharing information. It seems to be shifting a fundamental paradigm in public communication. Hierarchies are vanishing. Organizations have entered, or are beginning to enter, this new virtual world of communication and thus becoming a node in the network just as any other private person is. Public communication consists no longer of sole communication streams, well organized and thought-through by big institutions in a one-way direction. It seems to have become more of a constant buzz and chatter by everyone at the same time – in some corners louder, in some corners more

\textsuperscript{362} Ibid. p. 228.
often and in others less. Now, it is the user together with the community making out what is important, what is real and what is trustworthy. The Internet is not like any former medium that can solely be observed and from which information can be taken when it is interesting or important. Media like television, radio or newspapers can be consumed in that way and, if necessary, cited or taken as an input for further investigation on a topic. This is not the way the Internet works. This medium is participatory and engaging. Standing aside observing will not expose the valuable information. Finding the right information means being a part of the network, means being linked and guided to unknown sources. This goes against the natural instincts of journalists, who do not like to be guided and helped when looking for information. They like to judge for themselves what is trustworthy and what not, in order to meet the standards of objectivity.

The Web is a great information source, but using it to its full potential will entail including the user’s voice and becoming a part of the community. To do this in a professional way demands finding the right nodes in the whole network. And this especially takes time. Thus, developing a good reputation online, as offline, is a long-winded construction, even though the Web allows communication to be faster than never before.
8. Conclusion and Outlook

The Web is a fast developing and emergent medium. Many features have just evolved throughout the past few years. Within just a short amount of time they have had a serious impact on the nature of public communication and specifically on journalism, which today includes online journalism.

In accord with the title “Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information”, this thesis tried to take a deeper look into all the new developments and possibilities the current Web offers for online journalists to do professional research. The reason behind this effort was the fact that especially online journalists are suffering from the acceleration the Internet has brought to their profession, forcing them to do their work very quickly, yet complying with the standards of professional research. The acceleration and the constant publishing around the clock have led to a homogenization of the content throughout the online news outlets. In addition, due to the time pressure, online journalists do not find enough opportunities to really explore and discover the new developments and tools that are available and could be used on the so-called Web 2.0. One adheres to the known sources and procedures because they have proven themselves to be reliable.

With this background, the objective of this thesis was then to give online journalists an insight into, and maybe even a type of recommendation for, where to search reliably on today’s Web, so that valuable information can be sought quickly and online journalism could profit from more profound news and a greater variation in the texts. Online news is definitely gaining importance, but to publishers they are simply too expensive because no really successful business model has been found to sustain it. This situation has led to limited resources, personnel-wise and time-wise. So while the Web is gaining importance in everyday life and also in the knowledge gathering done by the people, online journalism is still fighting to find a way to satisfy their needs in an economic and professional way.

With all the knowledge processed throughout the last ninety some pages, the outcome at first sight is actually quite disappointing: The Web has become too complex of a structure to just be able to point to various sites and tools, which could simply be perceived as reliable. The essential characteristic of the Web 2.0 is its openness, making the consumer equally a producer. The Internet cannot just be seen as a medium, but as a connecting platform. There appears to be a virtual world running parallel and interconnected to the real world. The difference then is that it overcomes time and space as well as hierarchies.
Social media are accessible to all, connecting experts with laymen, fans with their idols, newcomers with professionals, and so on. Interest communities are being created and traditional gatekeepers no longer exist.

A closer look at the different tools shows that they do offer different types of information. Twitter is good for very recent information; blogs cultivate debates; wikis try to combine all knowledge and bring out facts; and social networks show information about people. However, they all entail the problem of trust and reliability, which is why a general recommendation of the different Web 2.0 sources cannot be made as such. The core to all the tools – resulting from being the core of the Web 2.0 – is the emphasis on the user, or more specifically on the crowd. Today’s Web functions through the participation and the feedback of the users. Tools are formed to what they are after being introduced. Communities are created and are essentially the ‘glue’ to them all, as can be seen on Wikipedia, in the blogosphere as well as with search engines, when considering the fact that all of three of them establish their results by links. It seems comprehensible that the Web 2.0 is called the democratic Web, when realizing the fact that it offers the means to implement the ideal of everyone contributing to a common good.

Using information from such sources then becomes a question of trusting the crowd. There are two problems inherent to this idea. Firstly, this way of thinking is contrary to the traditional and more familiar habits of trusting experts and hierarchically high positioned people within the field of interest. Ascribing reliability to a crowd, instead of ascribing it to a sole expert, would mean a fundamental paradigm shift in public communication and publicly ascribed reputation. It almost seems as if everyone would be voting on everyone and everything through their contributions and behaviors on the Web.

Secondly, for this paradigm shift to really happen, the new platform-like communication with its Web 2.0-tools still needs to become thoroughly institutionalized in society. Many features are not known adequately or are being viewed skeptically. The interviews done with the online journalists show this quite clearly. The use of the different tools varies significantly and some tools are not being used at all simply because one does not know them sufficiently. The new ways of communicating online are not used intrinsically and have not as yet found their place in society, at least not in Europe and not in professional journalism. One gets the impression that many journalists want to wait and see how the tools develop and start using them when they have proven themselves.

There is one tragic flaw to this line of thinking. The Web has introduced technologies that are by nature never finished. The way applications are and can be used depends greatly, if
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

not solely, on the user and his feedback. The “perpetual beta” needs the user’s information and the innovative developers to continually form the tool. The technologies are very young and people are probably just standing at the beginning of all there is that is possible to do with them. Consequently, for these tools to become relevant and interesting for online journalists, it is vital that they also be formed by them. This might require online journalists – or journalists in general – to no longer just observe and consume information on the Web, but actually dive in and become a part of the network simultaneously letting go of the hierarchical distance between him and the audience. This would entail another big paradigm shift; a shift in what is viewed as the professional journalist. It is clear at this point in time that it is too early for this shift to happen because of the ignorance towards the possibilities and the non-institutionalized state of the Web 2.0 features. The Web is too young and the technology in many ways not far enough along. However, it is claimed here, that it is time for publishers, employers and journalists alike as well as the public to rethink the online journalist’s role and way of work in a communication environment that differs fundamentally from traditional public communication known in pre-Web 2.0-times. Public communication is now happening within a network, and it is a network where organizations and private people have the same status. Journalism as professional communication will remain important, but, naturally, it cannot happen outside the network. This is a new situation with which journalism has never been confronted and its natural result will be questioning the traditional ways of working.

The different wishes stated by online journalists concerning further technological developments that would help them in doing their work, ranged from more specific searches all the way to interlinked databases. What has mostly been seen as some wishful thinking might not be that far off when the semantic Web really starts to enrich and permeate today’s Web. Using not only a Web site but the actual pieces of data, Web applications will combine the information on the Web in new innovational ways. The objective of these technological developments is to generate more specific and personalized search results. Wishes like finding only pictures with no copyright on them, or databases that combine all the different media (text, video, pictures), or databases visualizing the networks of a certain person’s history, then might become quite realistic. Of course, this is still far from today’s reality, but semantics are already slowly penetrating the Internet. For them to function in the best way possible, however, it is necessary to have enough data available, which explains why big companies, like Google or Apple but also Facebook, are collecting so much data and arousing discussions on privacy and online
security. It seems to be a vicious circle. If someone in the long run wants to have Web applications supporting his or her needs in the most ideal way, then data is needed. But, it is actually needed now in order to build them, before people even know what can be done with them. The same applies to online journalists. By using the tools and sharing data, new technologies could be formed to better suit their demands. If they stay skeptical, which is inherent to their idea of professionalism, they might hinder the developments and miss their chance to form the communication platform to meet their own needs. The Web is a user-formed sphere of communication. If online journalists want it to be useful to them, they need to start using it in all its ways.
9. Bibliography


IBM: What is social computin? URL:

Impressum (2006): Pflichten und Rechte. Erklärung der Pflichten und Rechte der Journalistinnen und Journalisten. URL:


Knobel, Reto (2010): Apples heimliche Datensammlerei. URL:


Kruse, Jan (2009, October): Reader „Einführung in die Qualitative Interviewforschung“, Freiburg (Taken from: http://www.sozioologie.uni-freiburg.de/kruse).


Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information


Twitter: Twitter is a real-time information network powered by people all around the world that lets you share and discover what’s happening now. URL: http://twitter.com/about (29th July 2010).


W3C: Linked Data. URL: http://www.w3.org/standards/semanticweb/data.html (21st July 2010).


Figures


Appendix
Appendix 1: Interview Questions for Online Journalists

- Wo suchen Sie nach persönlichen Angaben und Hintergründe? (Antworten zur Recherchefrage wer?)
- Wo suchen Sie nach zeitlichen, sachsenpezifischen Angaben? (Antworten zur Recherchefrage wann?)
- Wo suchen Sie nach den Informationen zum Sachverhalt / zum Geschehnis selbst? (Antworten zur Recherchefrage was?)

- Wonach / Welche Art Information suchen Sie auf Suchmaschinen?
- Nutzen Sie jeweils mehr als eine für eine bestimmte Suche? (Obwohl eine möglicherweise schon die Informationen geliefert hätte)
- Wie zuverlässig empfinden Sie die Resultate?
- Würden Sie eine andere Visualisierung der Daten und Kontextdaten vorziehen als Listenform? (Bsp. Thematische Visualisierung → eventuelle Möglichkeiten des zukünftigen Web 3.0)

- Welche Datenbanken nutzen Sie? Wofür?
- Spielen soziale Netzwerke eine Rolle in der Recherche?
- Wie intensiv werden Microblogs / Twitter in die Recherche miteingebunden?
- Wie sehr spielen Blogs eine Rolle in der Recherche?
- Wie sehr spielen Wikis (Bsp. Wikileaks) eine Rolle in der Recherche?

- Ist Ihnen Tagging ein Begriff?
- Wenn ja, nutzen Sie dieses Tool? (Wenn ja, wie?)
- Benutzen Sie Social Bookmarking Seiten wie bspw. Delicious (http://delicious.com/), um online Inhalte zu verwalten?

- Wie überprüfen Sie die Zuverlässigkeit und Glaubwürdigkeit der Online-Quelle?
- Würde ein Empfehlungssystem dem Problem der Glaubwürdigkeit Abhilfe schaffen?

- Was wünschen Sie sich von zukünftigen Online-Technologien für Möglichkeiten?
- Was wäre für Ihre Recherche von extremem Nutzen?
- Wie stehen Sie dazu, den Maschinen die Entscheidung der Relevanz ganz oder teilweise zu geben und dem Menschen die Bewertung der Relevanz ganz oder teilweise zu entziehen? Bspw. mittels maschinellem Lernen, Mustererkennung, verbesserten Rankingalgorithmen, usw.?
Appendix 2: Interview Questions for Media Coaches

- Where do you look for personal details and backgrounds? (Answers to the research question who?)
- Where do you look for time-specific, specific details? (Answers to the research question when?)
- Where do you look for information about the event / the event itself? (Answers to the research question what?)
- What kind of information do you look for in search engines?
- How do you handle search engines?
- Where are the advantages / strengths, where are the weaknesses of this source?
- Do you see potential in social networks regarding a journalistic research?
- What kind of information do you look for in social networks?
- What kind of information do you look for at microblogs like Twitter?
- What strengths and weaknesses do you see at Twitter?
- Can microblogs be used well for a journalistic research?
- What potential do you see for blogs?
- What potential do you see in wikis (e.g. Wikileaks) for research?
- How are guidelines and procedures taught concerning the credibility of online sources?
- What do you wish for from future online technologies for possibilities?
- What do you think about giving machines the decision of relevance completely or partially and removing the judgment of relevance completely or partially from the people? E.g. via machine learning, pattern recognition, improved ranking algorithms, etc.?
Appendix 3: Confidential Interviews

The interviews conducted for this survey were held in Switzerland, seeing as the author of this thesis lives and works in Switzerland as well as attends a Swiss university. The interviews thus were held in German, the mother tongue of all interviewees, making it possible for all respondents to talk in their first language. Besides lifting the language barrier, it was also necessary to guarantee the interview partners total confidentiality of their answers in order to acquire the needed and interesting information. For this reason, the transcripts, which contain the original information for the chapters 4, 7 and 8 of this thesis, are not printed in this public copy.
Web 2.0 and 3.0: How Online Journalists Find Relevant and Credible Information

ERKLÄERUNG

Ich versichere, dass ich die vorstehende Arbeit selbständig angefertigt und entsprechend den Grundsätzen wissenschaftlicher Ehrlichkeit abgefasst habe.

Es ist mir bekannt, dass andernfalls die Abteilung gemäss dem Fakultätsrat vom 09.11.2004 das Recht hat, den auf Grund dieser Arbeit verliehenen Titel zu entziehen.

…………………………………, den ……………………20……

………………………………………………
.

(Unterschrift)