

2 User-Generated Content

*Realizing the value of social media
requires innovative computing research.*

Shneiderman, Preece & Pirolli (2011, p. 34)

Social media, especially user-generated content, is a relatively young field of research. A research community is just forming around the topic. There is no common understanding of the subject, yet. Therefore, the following Chapter begins with an introduction of the notion of user-generated content and related terms for the work at hand in Section 2.1. I shall also give a summary of the range of understandings existing among other scientists. Furthermore, the classification of user-generated content is established that this work is based on.

User-generated content has effects on society, economy and the individual user himself. The fundamental changes and the potential that user-generated content is accompanied by are addressed in Section 2.2. But, the possibility for every user to publish content also causes a huge amount of information. The navigation through and classification of this information becomes a growing challenge for users. The user's possibilities to classify user-generated content today are presented in Section 2.3.

There are several works that aim at social media analysis or ranking. Some of them exploit every detail of information given for a specific kind of user-generated content allowing interesting computer-generated insights. The more an approach is specialized for a specific

platform, the more it tends to exploit platform specifics, which are difficult to transfer to other platforms. The specialization causes a lack of cross-platform comparability. Existing approaches and their compatibility with other platforms are subject of Section 2.4.

2.1 What User-Generated Content Is and What It Is Not

There are several different terms that evolved around the subject of user-generated content and the technology that enables its creation. *Web 2.0*, *social media*, *social Web*, *read/write Web*, *social computing*, *social software*, *collective action tools*, *socio-technical systems*, *computer-mediated communication*, *consumer-generated media*, *user-generated content*, *virtual communities*, *online communities* are some of them. For each term there is a variety of notions. In the following, the terminology used in the work at hand is introduced. Additionally, an overview of the variety of notions in related works is provided.

2.1.1 Terminology for Key Concepts Used in this Work

The democratization of information, the shift from one-to-many to many-to-many communication, and the transformation from consuming to publishing users of the Internet are essential aspects of social media (e.g., Lessig, 2001; Munker, 2009; Hansen, Shneiderman & Smith, 2010; Solis, 2010; Anderson, 2012). The user's ability to make content available to many people forms the basis of these aspects. Making content available to a large group once required printing and a publisher. Through social media it has become possible for almost¹

¹Of course, there are still basic requirements such as online access. It is important to be aware of potential regional differences. Inequalities in terms of access to information and communication technologies between countries or regions are known as **digital divide**. The world share of internet users grew from 18 percent in 2006 to 35 percent in 2011. Still, regional differences remain.

everyone to make information publicly available without the need of a publisher.

According to the *Encyclopædia Britannica* (2007, p. 415), publishing is “the activity that involves the selection, preparation, and distribution of written and visual matter.” Examples for published matter are books, magazines and newspapers. The traditional view of publishing regards only printed matters as published matters. But, since “electronic books and online newspapers” are also reckoned to be types of publishing, the printed form is no longer a requirement for publishing (*Encyclopædia Britannica*, 2007, p. 416). The *Encyclopædia Britannica* itself—after a history of almost 240 years of printed editions²—is no longer printed, but is exclusively accessible online.³ Although, in many cases of digital publishing the principles of printed publishing still apply, it becomes difficult to mark the boundaries of publishing clearly. The technical progress has changed the traditional notion of publishing.

To apply the concept of publishing in the context of social media, it has to be rendered more precisely. In the context of this work, **to publish** means making information publicly available. To adapt the concept to social media, several levels of public are distinguished. **General public** means that no receiver is specified by the contributing user. The content is available for everyone. This means that the audience is potentially unlimited. **Limited public** means that no receiver is explicitly specified by the contributing user, but the audience is limited. The limitation can be caused by platforms that require registration prior to reading. This is the case, for example, if a platform presents its content only to registered users; although anyone might be admitted to register, the audience is limited to the re-

Average Internet penetration levels in developing regions rose to 26 percent by 2011. In sub-Saharan Africa however, they remain below 15 percent. (United Nations, 2012)

²The first edition of the *Encyclopædia Britannica* was issued in 1768 (*Encyclopædia Britannica*, 2007). The last printed version was published in 2007.

³<http://www.britannica.com/>

gistered users. Limited public can be subdivided into known-limited public and unknown-limited public. **Known-limited public** comprises the cases when no receiver is specified but the audience is limited to known people. An example is content shared with a group, such as *friends* in social networks.⁴ This example illustrates that known-limited public is similar to private. **Unknown-limited public** describes the case when the audience is limited but not exclusively to known people. An example is when content is shared with a closed community such as *friends of friends* in social networks.

If the audience is limited to specified receivers, it is called **private**. In the context of this work, private communication is not user-generated content. This means, telephone calls, written letters, faxes, emails, SMS, instant messages, and so on, do not fall under the notion of user-generated content as it is used in this thesis. However, private communication can be part of a social media platform. Social networks for instance, usually allow to publish content to the general public, to a limited public as well as private messages.

In social media, the user who contributes a piece of content does not need to define his audience, but he can limit the audience. **Reach** is the number of people who receive a message. The less the audience is limited by a contributor, the more potential reach the message has. The degree of intimacy increases with limitation of the audience. Figure 2.1 shows the private and public levels of communication. Their characteristics are illustrated in relation to reach and intimacy.

This notion of *public*, adapted for user-generated content, replaces the concept of *sender and receiver* by *contributor and audience*. The **contributor** is the user who published a message. The contributor does not necessarily have to be the creator of the content. Whether or not the contributor is the creator of the content, in social media he is usually⁵ displayed as *author* and will therefore be referred to as **author**.

⁴For more information about social networks refer to Subsection 2.1.3

⁵In some cases, platforms display citations.

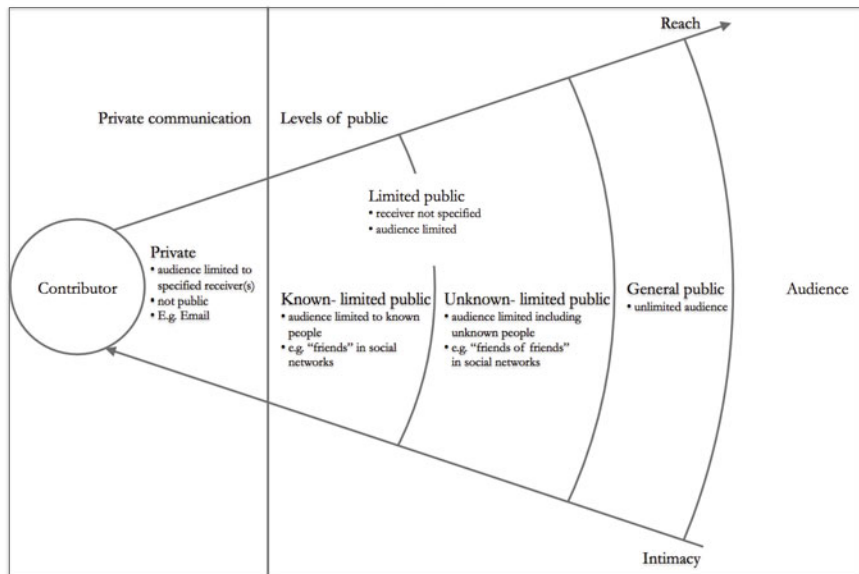


Figure 2.1: Reach-Intimacy-Model.

The Reach-Intimacy-Model illustrates private communication and the levels of public for user-generated content in relation to reach and intimacy. Private communication is not regarded as user-generated content.

In the context of this work, the central characteristic for user-generated content is the possibility for users to publish content to others.

User-generated content is content published on an online platform by users. The term **social media** comprises platforms that contain user-generated content. Users do not need programming skills to publish content on a social media platform.

Whether content contributed by a company on a social media platform is considered user-generated content, depends on the notion of *user*. *User* can refer to the *user of a social media platform*. In this case, the content contributed by a company on a social media plat-

form would be regarded as user-generated content. *User* can also refer to *private individual* as opposed to *professional* or *business person*. In this case, the content contributed by a company on a social media platform would not be considered user-generated content. In the context of this work, **user** refers to the user of a social media platform.

For search engines the smallest unit is a Web page with a URL as identifier. For user-generated content this view does not sufficiently apply. One Web page, one URL usually contains several social media entries from different authors. For social media the smallest unit is the **user-generated content unit**. A user-generated content unit is one single contribution by one author at a given time. Collaboratively created content usually has more than one author. This case is not covered in this thesis.

A user-generated content unit consists of core data and metadata. The given piece of information—the content—is the **core data**. **Metadata** is information about a given piece of information (Baeza-Yates & Ribeiro-Neto, 2003). Examples for metadata about user-generated content are *date of publication*, *status of the author in the community*, and *number of views*.

In this work, opinions that are expressed by one click are also referred to as **one-click-opinions**. Examples are Facebook’s *likes*, Google’s *+1*, Youtube’s *thumbs up*, and so on. Ratings of user-generated content units by other users are **peer-ratings**.

2.1.2 The Variety of Terminologies in Related Works

Aside from the notion of social media and user-generated content used in this work, there is a variety of other notions used in other authors’ works. It is important to be aware of this whenever key terms are used without further explanation. There is no consensus about which platforms and services belong to social media and which do not.

Many works about social media do not explicitly address the question what is meant by *social media*. This implies that it is often assumed that there is a common understanding about what the term *social media* refers to. But, a closer look at the uses of the term reveals that there is actually a variety of notions.

One might suspect that because there is not a uniform conception, there are many descriptions in various publications about the topic. But as a matter of fact, attempts to define the terms are rare. To many it is not exactly clear what social media is and what it is not (Lovett, 2011). Many existing descriptions for social media are incomplete, imprecise, or contradicting. Anderson (2012, p. 1) speculates that the reasons for this shortcoming might be the relative novelty of the subject or the “slippery character” of the subject “that’s hard to pin down.”

This is also the case for related terms such as *Web 2.0* and *user-generated content*. Some authors use the terms synonymously to social media, whereas others make a difference between them. Grabs & Bannour (2011) for example, describe Web 2.0 as the possibility for every user to create content and to share them via different channels among each other. They do not differentiate between the terms *Web 2.0* and *user-generated content* and use them interchangeably. Another example is Munker’s interpretation of Web 2.0, which is also very similar to what others describe as social media. He describes Web 2.0 as the trend to design a Web page in a way that it is considerably co-designed by its users. The degree of participation may vary from commenting or rating like it is practice at Amazon⁶ to platforms where the content is exclusively created by users (Munker, 2009).

Others use the term *Web 2.0* to stress technical aspects, such as AJAX and open APIs, whereas *social Web* or *social media* are used to stress social aspects (e.g., Anderson, 2012; Lessig, 2001).

The term *Web 2.0* has been coined with a different meaning. It can be tracked back to Knorr (2003) quoting Dietzen who “calls the Web 2.0,

⁶<http://www.amazon.com>

where the Web becomes a universal, standards-based integration platform.” It became popular when O’Reilly (2005) published an article titled “What is Web 2.0” about the ideas and changes behind the Web 2.0 concept.

Despite the expectations the title of the article might evoke, O’Reilly does not describe *the* Web 2.0 as form or enhancement of the World Wide Web. Instead he uses the term as an adjective and describes principles that he considers as the core competencies of companies that *are* Web 2.0. Examples are “companies that provide services with cost-effective scalability instead of packaged software are Web 2.0”, “Web 2.0 companies trust users as co-developers”, and “companies that are Web 2.0 have lightweight user interfaces, development models and business models” (O’Reilly, 2005). O’Reilly also provides “Web 2.0 Design Patterns” with recommendations for companies that want to become Web 2.0. These examples illustrate O’Reilly’s business driven point of view when he uses the term *Web 2.0*.

Kaplan & Haenlein explicitly address the question of what social media is. Their attempt to define social media is often referenced (e.g., Fischer & Reuber, 2011; Foster, Francescucci & West, 2010; Nack, 2010; Wikipedia, 2013a,b). Kaplan & Haenlein agree that there seems to be confusion as to what should be included in the term *social media* and how it differs from *Web 2.0*. They dedicate a section to the question “What is Social Media—And what is not?” Herein, they characterize social media as “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content.” Meanwhile, Web 2.0 is described as “the platform for the evolution of Social Media” and “Web 2.0 represents the ideological and technological foundation.” User-generated content is “the sum of all ways in which people make use of Social Media.” (Kaplan & Haenlein, 2010, p. 60 ff)

Rephrasing their circular reference, the three terms are not synonymous, but related and are somehow a foundation for each other. Social media refers to applications, Web 2.0 refers to underlying technological aspects, whereas user-generated content is specified as “ways of use.” It is a suggestion for how to arrange the concepts towards each other. But Kaplan and Haenlein do not offer a definition for what social media is and what it is not. They do not specify “people” in context of “user-generated content.” Considering blogs, it remains unclear whether Kaplan & Haenlein would consider blogs published by professionals and journalists as “user-generated content.” Companies use microblogs such as Twitter and social networks such as Facebook for communication, too. It is unspecified whether they regard this kind of content “user-generated content.”

There are also inconsistent uses of terms and very broad conceptions that describe much more than online communication and user participation. Safko (2010, p. 3) claims: “Social media is the media we use to be social. That’s it.” This would also include telephones, emails, letters and so on, none of which are addressed in his book. Hansen et al. (2010, p. 12) see social media as “a set of online tools that supports social interaction between users. The term is often used to contrast with more traditional media such as television and books that deliver content to mass populations but do not facilitate the creation or sharing of content by users.” But then they include also corporate Web sites (e.g., www.ford.com) in their taxonomy of social media, which—as television and books—only deliver content and do not facilitate the creation or sharing of content by users (Hansen et al., 2010).

Summing up, there is no terminology that is universally agreed upon. A variety of other notions is used in other authors’ works. Recent works which broach the issue of terminology show no clear tendency that the different understandings are converging towards a consensus. Consequently, it is important to pay attention to the specific notions that underlie key terms in other works.

2.1.3 Social Media Categories Used in this Work

Resulting from the divergent comprehension there is a series of different opinions about which platforms belong to social media and how they can be categorized. Most works contain enumerations of platforms with an exemplary character instead of methodically built taxonomies. Whenever social media categories are mentioned without further elaboration, there is a range of interpretations.

There are numerous platforms that allow users to publish content and belong to social media. Constantly more are emerging. A proper categorization should include all platforms that meet the underlying definition of social media. Categories should be methodically derived from a feasible criterion.

In the context of this work, platforms are allocated to categories by the type of metadata provided by the platforms.⁷ This way to categorize social media platforms is chosen because it is especially suited for query-independent ranking of user-generated content that utilizes metadata to evaluate content units.

The following classification will serve as basis of understanding for this work:

1. Blogs
2. Forums
3. Location sharing and annotation platforms
4. Media sharing platforms
5. Microblogs
6. Question and answer platforms
7. Rating and review platforms
8. Social networks

⁷For the detailed analysis of metadata of user-generated content from different platforms refer to Chapter 3.

Social Media

Fundamentals, Models, and Ranking of User-Generated
Content

Wyrwoll, C.

2014, XVII, 167 p. 28 illus., Softcover

ISBN: 978-3-658-06983-4