DISCUSSION OF FUNCTIONAL DESIGN OPTIONS FOR ONLINE RATING SYSTEMS: A STATE-OF-THE-ART ANALYSIS

Journal:	*17th European Conference on Information Systems*
Manuscript ID:	ECIS2009-0607.R1
Submission Type:	Research Paper
Keyword:	E-commerce (B2B / B2C / B2G / G2C), Reputation, Web 2.0, User participation

DISCUSSION OF FUNCTIONAL DESIGN OPTIONS FOR ONLINE RATING SYSTEMS: A STATE-OF-THE-ART ANALYSIS

Winkelmann, Axel, University of Koblenz-Landau, Department of Computer Science, Institute for IS Research, Universitätsstraße 1, 56070 Koblenz, Germany, winkelmann@uni-koblenz.de

Herwig, Sebastian, University of Münster, European Research Center for Information Systems (ERCIS), Leonardo-Campus 3, 48149 Münster, Germany, sebastian.herwig@ercis.uni-muenster.de

Pöppelbuß, Jens, University of Münster, European Research Center for Information Systems (ERCIS), Leonardo-Campus 3, 48149 Münster, Germany, jens.poeppelbuss@ercis.uni-muenster.de

Tiebe, Daniel, University of Münster, European Research Center for Information Systems (ERCIS), Leonardo-Campus 3, 48149 Münster, Germany, daniel.tiebe@ercis.uni-muenster.de

Becker, Jörg, University of Münster, European Research Center for Information Systems (ERCIS), Leonardo-Campus 3, 48149 Münster, Germany, joerg.becker@ercis.uni-muenster.de

Abstract

Ratings are important for building up trust among different parties. Since the arrival of the internet era in the 1990s, countless online rating systems have emerged. For example, Amazon.com established a rating system for books and other products in 1995. Today online rating can be found everywhere, be it e-commerce sites, social networks, and information or recommendation platforms. In most cases, users provide the input to these systems which is then aggregated and directed to appropriate recipients. The increasing relevance of these rating systems forms a new distinct research field with a growing need for research on the design, effects, and validity of rating systems. Hence, we contribute to the body of knowledge by conducting a thorough analysis of the state of the art of online rating systems. We especially focus on the functional design options by analyzing 102 different systems with the help of a criteria catalog of 237 criteria. In this paper, we discuss an excerpt of our findings and present a morphological box that categorizes functional design options for online rating systems.

Keywords: E-Commerce, Reputation, Web 2.0, User Participation.

1 INTRODUCTION

Ratings are important to people, especially for building up trust among business partners. The traditional offline rating systems have always been and still are word of mouth and gossip (Cheung & Luo & Sia & Chen 2007, Dellarocas 2001). Nevertheless, since the internet era begun in the 1990s, new online rating systems emerged. As some of the first, Armstrong and Hagel III (1995) dealt with these systems and they found out that customers fall back on advices and recommendations of other customers when executing online transactions. Moreover, they discovered that online rating systems can even be a means to improve customer retention (Armstrong & Hagel III 1995). Being one of the rather progressive in this field, Amazon.com has enabled its users to rate books and other products since 1995. Many other online stores have also included rating systems on their websites and even pure rating web-portals have emerged; e.g. Epinions.com was launched in 1999 which – according to the Alexa Rating – is the most popular web-portal for reviews on products and services today. All this contributes to make electronic shopping a social experience because of an intensive sharing of recommendations and reviews (Vossen & Hagemann 2007).

The advantage of online – in contrast to offline – ratings is scalability and formalization (Bolton & Katok & Ockenfels 2004, Dellarocas 2003). Scalability means that ratings can be gathered from and communicated to a multitude of parties, independent from time and place (Resnick & Kuwabara & Zeckhauser & Friedman 2000). Users can access a huge number of ratings provided by other users in an easy and cost-efficient manner (Cheung et al. 2007). According to Töpfer, Silbermann and William (2008) especially the fast diffusion of up-to-date rating information is a main advantage (Conte & Paolucci 2002, Töpfer et al. 2008). Through unification of gathering, aggregation and presentation ratings become more comprehensible and their acceptance increases (Resnick & Zeckhauser 2001).

In fact, the main field of application of online rating systems is e-commerce which has gained increasing importance over the last two decades (Kapell 2007, Sebralla 2008). Nevertheless, one key problem of online transactions is the lack of trust among business partners. Users of online shops often have problems in estimating the quality and reliability of offers and their counterparts (Ward & Lee 1999). To address this issue, rating and reputation systems have become an integral part of many e-commerce sites and are especially supposed to facilitate the identification of trustworthy business partners. For instance, each transaction at Ebay.com is expected to be rated by the involved users in order to build up their rating profiles which are accessible to any other registered user. Analyses have shown that there is an impact of the seller's rating profile on the likeliness of a successful transaction as well as the achieved selling price (Armstrong & Hagel III 1995, Sebralla 2008).

Nevertheless, rating systems can also be found apart from e-commerce, e.g. in social networks that people use to manage their connections to other people. A typical example of a social network used for business purposes is LinkedIn, where users can write recommendations for colleagues, business partners or professional service providers. Similar to a traditional job reference, recommendations refer to their work and are listed in their profiles. Even more rating systems are found on non-business network sites like MySpace.com, Facebook.com or Skyrock.com where people are able to rate e.g. user profiles or videos.

The increasing relevance and spread of rating systems forms a new distinct research field. According to Peters and Reitzenstein (2008), there is an increasing need to do research on the forms, effects and validity of rating systems. For instance, relevant questions are how to aggregate rating input and represent valid rating results as well as the effects on (business) partner selection or trust building. So far, there is a lack of studies on the functional design of differing rating systems although it is the design of rating systems that determines the effect of their results to a large extent. In this paper, we will analyze the state of the art of existing rating systems and derive a consolidated set of design options which are discussed in respect to their influence on the entire rating system.

The remainder of this paper is structured as follows. The next section will provide a brief theoretical background on online rating systems. Thereafter, we will present the research method we applied for our state-of-the-art analysis. Then, we will describe excerpts from our results which we will subsequently discuss in order to provide a consolidated morphological box of design options. We will conclude with explaining the limitations of our study and providing an outlook on further research.

2 THEORETICAL BACKGROUND

The basic idea of online rating systems is to let users rate entities by means of web applications and hence to collect, aggregate and distribute reviews on entities (Resnick et al. 2000). The aggregated ratings about a given party or entity can be used to derive a score, e.g. a trust or reputation score, which is communicated to other parties. These scores can assist these parties in deciding whether or not to transact with certain other parties in future (Josang & Roslan & Boyd 2007). Typically, rating systems have a central authority that collects ratings and disseminates the rating results/scores.

The terms rating, review and reputation are closely linked. “Reputation is what is generally said or believed about a person’s or thing’s character or standing.” (Josang et al. 2007) It can be considered as a collective measure that is based on the reviews or ratings from members in a community. Web applications that gather reviews or ratings in order to compute and communicate reputation scores are called rating systems or reputation systems which we regard as synonyms throughout this paper. However, recommender systems must not be confused with reputation or rating systems. Recommender systems generate individualized recommendations to guide users in a personalized way by applying collaborative filtering or similar (Burke 2002); e.g. Amazon.com’s “Customers Who Bought This Item Also Bought” application. Rating systems can rather be described as collaborative sanctioning systems because they are based on the assumption that all users in a community should judge the performance or quality of an entity (e.g. transaction partner, product, information or multimedia contents). The aim is to sanction poor entities and to give an incentive for improvement (Josang et al. 2007).

A rating itself can be understood as a certain believe to which degree the rated entity is useful for a given objective (Miceli & Castelfranchi 2000). The criteria for this rating are usually determined by the objective whereas the entity determines the values that are assigned to each criteria (Schuler 2004). The appraisal usually happens subjectively and depends on the skills and experiences of the party that rates the entity. Obviously, a rating is only a single opinion which also refers to a specific situation and time (Sabater-Mir & Paolucci 2007). Furthermore, as an empirical analysis showed, users tend to rate positively and avoid negative statements (Resnick & Zeckhauser 2001).

The use of online rating systems requires an adequate design of the underlying mechanisms. According to Chen, Hogg and Wozny (2004), one of the main decisions refers to the gathering of information. Operators of such systems must determine which users are allowed to rate which entities. Especially, deliberate manipulations by single users must be avoided (Dellarocas 2003). To respect user privacy, users should have to opt-in before being subject of ratings (Ziegler 2008).

3 RESEARCH METHOD

Our main research objective was the analysis of the diverse functional designs of online rating systems leading to a consolidated set of design options. The visual design was not in scope of our analysis. Our research process included the following six steps:

- Selection of websites to be analyzed
- Identification of rating systems that are part of the online service
- Iterative design of the criteria catalog
- Appraisal of the online services according to the criteria
- Analysis of the appraisal data
- Consolidation of criteria into a morphological box of design options.

Selection of websites to be analyzed

Our initial literature review did not reveal any reference study that has analyzed rating systems in respect to their functional design. Consequently, we needed to decide on an appropriate sample of websites to be analyzed. The selection of the sample was supposed to be transparent and objective with enough rating mechanisms needed for a thorough analysis.

To meet these requirements, we selected the Top50 websites of the Alexa Traffic Ranking (April 2008; see Table 1). This popular ranking is basically derived from users that have added the Alexa Toolbar to their Browser. Although the traffic ranking is only an approximation, it provides a suitable means for the identification of the most popular websites. Due to adult contents and contents limited to languages apart from German or English, we excluded thirteen websites from the initial sample.

Rank	Website	Type of Service	Reason for Exclusion
1	Yahoo.com	Web portal	
2	Google.com	Search engine	
3	Youtube.com	Video portal	
4	Live.com	Search engine	
5	Msn.com	Web portal	
6	Myspace.com	Social network	
7	Wikipedia.org	Online encyclopedia	
8	Facebook.com	Social network	
9	Blogger.com	Blog community	
10	Yahoo.co.jp	Web portal	No English or German language
11	Orkut.com	Social network	
12	Rapidshare.com	Online storage service	
13	Baidu.com	Search engine	No English or German language
14	Microsoft.com	Enterprise homepage	
15	Google.co.in	Search engine	
16	Google.de	Search engine	
17	Qq.com	Web portal	No English or German language
18	Ebay.com	Auction portal	
19	Hi5.com	Social network	
20	Google.fr	Search engine	
21	Aol.com	Web portal	
22	Mail.ru	Web portal	No English or German language
23	Google.co.uk	Search engine	
24	Sina.com.cn	Web portal	No English or German language
25	Fc2.com	Online storage service	No English or German language
26	Photobucket.com	Picture portal	
27	Google.com.br	Search engine	
28	Amazon.com	E-Commerce portal	
29	Imdb.com	Film rating portal	
30	Vkontakte.ru	Social network	No English or German language
31	Google.it	Search engine	
32	Google.es	Search engine	
33	Google.cn	Search engine	No English or German language
34	Imageshack.us	Picture portal	
35	Youporn.com	Video portal	Pornographic content
36	Wordpress.com	Blog portal	
37	Google.co.jp	Search engine	
38	Yandex.ru	Web portal	No English or German language
39	Flickr.com	Picture portal	
40	Friendster.com	Social network	
41	Skyrock.com	Social network	
42	Adultfriendfinder.com	Social network	Pornographic content
43	Go.com	Web portal	
44	Odnoklassniki.ru	Social network	No English or German language
45	Goggle.com.mx	Search engine	
46	Bbc.co.uk	News portal	
47	Craigslist.org	Advertising portal	
48	Dailymotion.com	Video portal	
49	Redtube.com	Video portal	Pornographic content
50	Cnn.com	News portal	

Table 1. Overview of Alexa Top50 (April 2008)

Identification of rating systems that are part of the online service

Having selected the sample, the next step was to identify the rating systems on these websites. We had to consider that the Alexa Top50 includes portals like e.g. Yahoo.com which offer a huge range of separate online services. To avoid accessing every single page of such portals, we restricted our search to services that were linked on the home pages. Some services provided further sub-services which we also only included into our search when linked at the service home page. Where possible, we did at least one rating at each system and analyzed several ranked entities. Altogether, we identified 102 rating systems by searching the online services.

Iterative design of the criteria catalog

In order to gather the characteristics of each rating system systematically and to allow for comparisons, we iteratively elaborated a list of 237 criteria. The criteria are mainly related to functional aspects of rating systems but also cover the environment in which they are embedded. Lacking an existing list of criteria for this particular purpose, we started to derive relevant criteria by a literature review. We continuously extended the criteria catalog in the course of analyzing the broad range of existing online rating systems. We designed each criterion in a binary manner, i.e. a system either possesses the characteristic or it does not.

Appraisal of the online presences according to the criteria

We applied the list of criteria to each rating system and assigned the matching values. In some cases it was not possible to determine if a system possesses a certain characteristic. Therefore, we also assigned 'not applicable', and 'not testable'. A criteria is 'not applicable' if its prerequisites are not met, i.e. it is not possible to edit a review comment when the system just allows to rate an item. 'Not testable' was assigned when we could not determine the value without doubt, e.g. because of missing documentation.

Analysis of the appraisal data and consolidation of the criteria

Due to page restrictions we can only present exemplary excerpts of our appraisal results in the following section. We already ordered these excerpts according to the key dimensions of the morphological box we will present in section 5. The morphological box (Zwicky 1969) was developed through a critical discussion of the long list of criteria and cross consistency assessments by all authors in order to aggregate and structure the appraisal results (Bailey 1994).

4 ANALYSIS OF THE STATE-OF-THE-ART

The following presentation of our analysis results is structured along six key dimensions of the functional design of rating systems. The first dimension is the service provider who operates the rating system. Secondly, we identified which kinds of entities are rated. Furthermore, we had a closer look on how the rating is entered and aggregated. Finally, we analyzed the representation of the ratings as well as the incentives provided for reviewers in order to evaluate entities.

4.1 Service Providers

In our sample, the range of service providers that use rating systems is very broad. For example, we analyzed producers, retailers, review sites and broadcasters. The service provider diversity goes along with their diverging objectives. Consequently, rating systems can fulfill different functions. For our sample we were able to identify six different functions of rating systems as visualized in Table 2.

Functions of Rating Mechanisms	No. of Systems
Product evaluation at price / product portals	5
Sales promotion	8
Partner selection	9
Rating of information	16
Neutral rating mechanisms	27
Community rating	37

Table 2. Functions of Rating Systems

77 of our 102 investigated rating systems focused on the central entity of the service, i.e. it is possible to rate the entity that the website was actually intended for. Moreover, 12 of these 77 systems strongly depend on the rating systems to fulfill the intended website service. For example, Pixnay.bebo.com is a social community that is limited to people rating each other's pictures.

4.2 Entity

The entity can be seen as the central element of each rating system as it determines most of the design parameters. Concerning entity types we distinguished between subjects, abstract subjects and objects (see Figure 1). Subjects are active actors, in general other users. Objects are e.g. goods or pieces of information. The concept of an abstract subject means that an object or event is rated but the aggregation is assigned to a subject. For instance, at Ebay.com buyers rate a seller through transactions made and ratings are linked to the user profile of the buyer although actually an event (transaction process) was rated. The information behind this kind of ratings is limited to the specific situation and therefore not directly transferable to other contexts and situations.

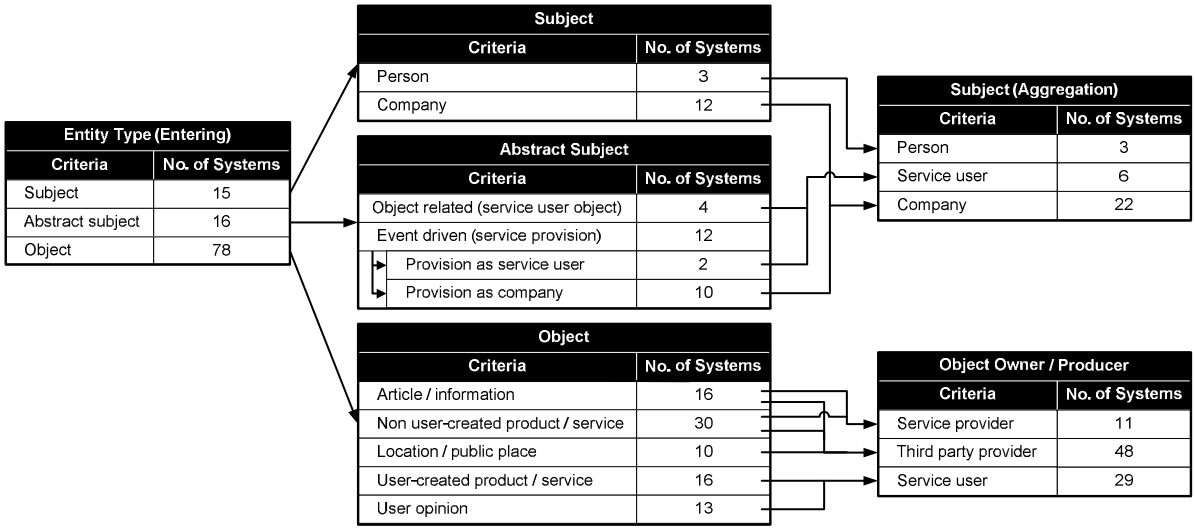

Figure 1. Classification of Entities

Some rating systems allow the rating of different entity types. For example, some city guide portals allow the rating of subjects (companies) and objects (points of interest). Most systems (78) offer the assessment of objects only. Ratings about products or services prevail. The object usually does not belong to the website owner (this is only the case for 11 systems) but is provided by users (29) or external third parties (48).

4.3 Gathering of Ratings

The gathering of ratings should be designed in a way that it allows users to enter ratings as easily as possible. However, 84 of the analyzed systems required an enrolment for the online service. About two third of these systems (62) asked for a valid email-address. At 11 systems, users had to additionally qualify for ratings, e.g. through membership (2) or through a specific member status (3; e.g. reaching a certain score at Answers.yahoo.com). In other cases (6) the accomplishment of a certain transaction is a pre-requisite for rating, e.g. at Ebay.com.

In 16 cases, the user (7 cases) or some other stakeholder is able to restrict the rating of entities. For example, at Pixnay.bebo.com a registered user is able to decide whether other users are allowed to rate his/her picture or not.

The rating entry forms of the service providers vary. Some are very detailed and therefore require a lot of input from the user. Furthermore, some allow a meta-rating such “X out of Y users found the rating very helpful”. However, most turn out to be very simple with just one or few rating options.

4.4 Aggregation of Ratings

93 out of 102 systems consolidate single ratings to one holistic score (see Table 3). This can be problematic as a consolidation reduces the rating information (e.g. it blurs the differences between positive and negative ratings). Some services such as ‘Karma’ at Orkut.com allow the entering of single values such as trusty, cool and sexy without consolidating the assessments. Furthermore, 8 systems display the number of ratings for each value (e.g. 17 positive, 3 neutral, and 23 negative ratings).

Criteria	No. of Systems				
	Yes	No	Not applicable	Not testable	Total
Aggretation per entity	93	9	0	0	102
Average value	78	22	1	1	102
Cumulative	***22***	78	1	1	102
→ Cumulation based on scale values	5	96	1	0	102
→ Relative values	1	100	1	0	102
→ Cumulation without negative ratings	4	96	1	1	102
→ Counting of votes	12	89	1	0	102

Table 3. Aggregation of Entities

78 rating systems compute the aggregated score as an average rating value (see Table 3). The other quarter of rating systems uses a cumulative approach such as counting of votes (12), cumulating based on scale values (5) or cumulating with subtraction of negative ratings (4) as well as relative values (1). Some websites such as Propeller.com or Epinions.com use further information for the calculation of the scores. At Propeller.com, the score is also based on comments and the number of users that have read the article. The exact calculation is not explicitly described to avoid manipulations. At Epinions.com, a score is additionally based on the number and the age of ratings.

4.5 Representation of Ratings

The representation determines the explanatory power of ratings (Sabater-Mir & Paolucci 2007) and relies on the data input (Dellarocas & Wood 2008). 89 systems work with ratings that are visible for all users (see Table 4). Of those 13 remaining, 4 are only visible for registered users and 9 rating systems require a qualification prior to accessing the ratings. In most cases (85), the total number of ratings is quoted on the profile page of the rated entity (e.g. in the user profile, product details, etc.). However, 28 rating systems use ratings for enriching search results on individual entities. 28 systems display the number of ratings for each scale values such as 17 users rated “very good”, 24 users rated “good”, etc.

Criteria	No. of Systems				
	Yes	No	Not applicable	Not testable	Total
Visibility of scoring values					
Visible for all users	89	***13***	0	0	102
→ Visible for registered users only	4	9	89	0	102
→ Qualification necessary	9	0	93	0	102
Total Number of Ratings					
Total number quoted	***85***	16	1	0	102
→ Quoted on profile page of entity	85	0	17	0	102
→ Quoted on search result / overview page	30	68	4	0	102
Specification of scoring value (on profile page)					
Number of ratings per scoring scale value	28	70	4	0	102
Percentage of positive ratings	7	94	0	1	102
Provision of additional Ratios	4	98	0	0	102
Activation					
Deferred visibility of ratings / activation by operator	***32***	69	0	1	102
→ Immediate visibility in spite of alleged deferment	12	14	69	7	102
Immediate visibility of ratings	69	32	0	1	102
Visibility of individual ratings					
Visibility of individual ratings	54	48	0	0	102
Collecting additional data for internal purposes	8	8	83	3	102
Standard order of individual ratings					
Latest rating first	38	13	48	3	102
Order based on algorithm (various input parameter)	3	48	48	3	102
Sorting / selection of ratings					
Visualisation of „TopRated“ / selectable	38	64	0	0	102
Sorting possible based on different criteria	28	26	48	0	102

Table 4. Representation of Ratings

The providers of 32 systems state that they approve each rating (see Table 5). E.g. Expedia.com claims to take up to 10 days for checking each evaluation. Our tests indicated that 12 out of these 32 systems did not approve the ratings but allowed an immediate visibility of our judgments. 69 systems allowed an immediate visibility of ratings without any editorial approval.

Some services provide a sorting or selection mechanisms. A good example is the filter mechanism of IMDb.org that provides extensive options for selecting movie ratings of specific rater groups (see Figure 2).

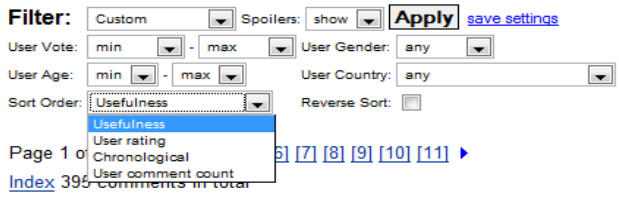

Figure 2. Selection of ratings at IMDb.org

4.6 Incentive Schemes

Incentive schemes can be used for encouraging users to rate specific entities. The reason behind this is the fact that a rating provides value for the consumer rather than for the evaluator. The visibility of login names or real names as an incentive to excel evaluators is only used by 51 systems (see Table 5). 26 systems allow a linkage to the individual profile page of the evaluator. 22 systems enable an aggregation of ratings within the own profile. Furthermore, a rating of evaluators and specific labels are used by a fraction of services. Only two online services, Epinions.com and Orbitz.com, offer monetary compensation for rating efforts. Refunding at Epinions.com is based on the number of users that read the rating. Orbitz.com offers coupons to each reviewer.

Criteria	No. of Systems				
	Yes	No	Not applicable	Not testable	Total
Visibility of evaluator per rating					
Visibility of login name	25	26	48	3	102
Visibility of real name	26	25	48	3	102
Profile creation in the context of ratings					
Visibility of given ratings within own profile	22	4	76	0	102
Ranking of evaluators	5	17	80	0	102
Usage of labels	***9***	72	21	0	102
→ Expert labels	3	6	93	0	102
→ Character labels	6	3	93	0	102
Monetary compensation of rating effort	2	100	0	0	102

Table 5. Incentive Schemes

5 DISCUSSION OF RESEARCH RESULTS

Our analysis – based on 237 criteria applied to 102 different rating systems – shows a wide variety of characteristics among rating systems. In our sample, a quarter of all systems are used as a means to gather and provide user information for specific entities e.g. a certain type of car. The support of sales activities and product comparison are common areas of use. There are some elements that are similar in nearly all rating systems. However, it is not possible to identify a best practice approach that is suitable for every situation. Every type of entity and business model has its own requirements on rating systems.

Hence, we will discuss the key parameters and parameter values we derived from our findings. We included these key parameters into a morphological box (Zwicky 1969). They were determined

through a critical discussion of all identified criteria as well as cross consistency assessments by all authors (Bailey 1994). The goal was to aggregate and structure the results of our analysis

	Parameter	Parameter Values			
Provider	Service	Commercial		Non-commercial	
Entity	Type	Subject		Abstract subject	Object
	Reference to service	Specific		Unspecific	
	Integrity (Encapsulation)	Service area only	Domain	Domain-spanning	Provider-spanning
Gathering	User group	Open		Closed	
	Precondition	None		Nonrecurring qualification	Qualification per evaluation
	Rating type	Direct		Relative	
	Direction of rating	Unilateral		Bilateral	
	Level of judgment	Overall Assessment		Section	
	Multiple ratings	Not possible		Conditional	Possible
Aggrega-tion	Input scoring value	User assessments only		Diverse input values	
	Calculation method	None		Cumulative	Average
	Weighting	Equal	Time-dependent	Qualitiy-dependent	Depending on Evaluator
Represen-tation	Consumer of ratings	Unlimited		Limited	
	Visibility	Immediately		Deferred	After Editorial
	Exposure of rating	Not exposed		Limited	Full
	Annotation of ratings	Not possible	Open user	Closed user	Indirect
Incentives	Naming of Evaluator	Not named	Anonymous	Name	Profile
	Incentives	None		Indirect	Direct

Figure 3. Conceptualization of Design Options

The searched online services are of either commercial or non-commercial interest. It turned out that 100% of all analyzed rating systems have a commercial background. A further analysis of non-commercial services may add to our framework, although we believe that their rating systems will deal with the same entities and objectives.

The rated entities turned out to be either subjects such as individual people, abstract subjects or objects. In most cases (77%) the rated entities had been objects such as hotels, products or services. Rating systems for objects are usually not restricted to certain brands. Only at YahooShopping.com we found that it is possible to rate some shoe brand, while it is not possible to rate other brands.

Furthermore, providers use rating systems not only for their specific service but also in other service areas or even across various providers. However, there is no standardized mechanism for exchanging rating information between platforms. With the emergence of Open Social (OpenSocial.org) at the end of 2007, which is a standardized API for social applications across multiple websites, we expect the interaction among different platforms will increase within the next years.

Linking the permission for rating to certain prerequisites can improve the rating quality (Dellarocas 2000). In most cases, the user group is restricted (82%). Hence, it is possible to automatically validate whether a user is authorized to conduct a rating or not. However, there are also some rating systems which do not require a user registration. The decision whether to provide an user-restricted or unrestricted rating system very much depends on the acceptance of people, the general business model behind the website and the importance of the rating. If, for instance, the user has to be registered in order to use services anyway, a restricted rating system is preferable. But if the rating system is the main part of the service (e.g. Pixnay.com's picture rating), a unrestricted rating system might limit the acceptance of such a platform. In some contexts it may even be sensible to set up further prerequisites before a rating is admitted. For example, at Ebay.com people have to conduct a purchase first before they are able to rate the seller of the product. Qualification can also be useful in areas of high interest but low general knowledge. Hence, a qualification mechanism is sensible for websites that need specific expert knowledge or for social networks where only friends are allowed to rate other friends.

We found that most ratings are direct ratings that express an opinion directly ('I like Mandy', 'This hotel receives 5/5 stars', etc.). For the provider, this sort of rating is very easy to aggregate and calculate. However, there are academic discussions and studies about problems with fraud (Gregg & Scott 2006). In addition, a bilateral direct rating may lead to revenge ratings. Even at Ebay.com, the old bilateral direct rating system, which was restricted to transactions, turned out to be very problematic and has been changed to a more or less unilateral rating system recently. In few cases we also found some relative rating systems on the internet where ratings are made based on the comparison with other entities (Dellarocas 2003). Unfortunately, such an approach is only possible in social applications with a high degree of cross linking (Botsch & Luckner 2008). In general, relative rating systems can help overcome some (but not all) fraud issues since a rating is only possible in comparison to other subjects or objects. Nevertheless, this kind of rating systems is not applicable to all services. For instance, relative ratings may not be helpful in situations where the entity cannot be compared to other objects (e.g. other services) or the evaluator only knows a limited number of comparable entities (e.g. comparing a hotel to other hotels in this region).

Providers of rating systems tend to aggregate individual ratings in order to receive an overall score although they risk losing relevant individual information. For example, an average mark does not give any information about the number of positive and negative ratings. Hence, it is sensible to display ratings for each score value (e.g. for positive, neutral, and negative ratings, see Google.com/products as an example). Since many ratings systems calculate scores out of single ratings, the method for this calculation is one of the key design options.

Most aggregation approaches for ratings have proven to be very simple. Usually, all ratings counted equally. Simple Systems compute an average of all ratings or a total score as the sum of positive scores minus the sum of negative scores. Advanced methods compute a weighted average of the ratings, where the rating weight can be determined by different factors, e.g. trustworthiness of the rating participant, age of the rating or distance between rating and current score. Furthermore, even more sophisticated statistical methods can be applied. In almost any case (98%), the aggregated score is only based on individual user ratings but there are also other factors that might influence ratings (e.g. at Epinions.com).

Retrieving ratings or scores is usually open to everybody that accesses the according online service. The high number of generally visible ratings indicates the importance of ratings for individual business models. E.g. e-commerce websites heavily rely on user ratings in order to offer additional value to potential buyers.

The need for incentives to attract ratings largely depends on the general activity and involvement of users on the platform and hence the business model itself. For example, pure rating platforms like Pixnay.com do not need any form of incentive for rating at all. In contrast, platforms that sell services that have to be used before rating them (e.g. hotel stays, craftsmen services, etc.) will need incentives in order to persuade users to return again to their websites for rating.

The social aspect of profile building and reputation is a basic incentive that is very popular among online services. These services either use the real name of a person or a pseudonym. In case of using the real name there will be a connection to the real world. Depending on the service, this can be an advantage (e.g. the user can prove that he is an expert in his profession) or a disadvantage (users fear a loss of privacy). Advanced incentives can be of monetary or other beneficial nature. For example, Epinions.com offers money for reviews. Some services offer indirect incentives such as the opportunity to participate in a raffle as a reward for entity rating. Hence, they provide additional value to the user with relatively low effort.

6 LIMITATIONS AND FUTURE RESEARCH

Our analysis is based on a large sample of websites and their rating systems. However, due to language barriers we were only able to analyze German and English websites. Furthermore, we only considered the most popular ones. There may be a possibility that there will be other rating systems in other cultural regions (e.g. Asia) that would additionally contribute to our body of knowledge. At the end, our criteria catalog was made out of 237 criteria and we believe it to be fairly comprehensive. Nevertheless, the analysis of additional rating systems might add even further criteria.

Unfortunately, we were only able to test the rating algorithms from the user perspective and could not look behind the curtains. However, we are convinced that our observations are quite accurate. Future research should also include less popular websites in order to understand differences between rating systems on high traffic websites and low traffic websites.

Today, most rating systems are on objects rather than subjects. We believe that the emergence of social networks and social applications will change the focus of rating systems towards users. Social networks may offer new rating systems within the next years and will raise additional questions regarding data privacy, self-determination, etc. Furthermore, new APIs such as Open Social may change the way of interaction among websites and hence the design of cross-platform rating systems. We believe our recent "snapshot" of the state-of-the-art of rating systems to be a valuable contribution to these and further emerging questions that have to be addressed by future research.

References

Armstrong, A. and Hagel III, J. (1995). Real profits from virtual communities. The McKinsey Quarterly (3), pp. 126-141.

Bailey, K. (1994). Typologies and Taxonomies - An Introduction to Classification Techniques. Sage University Papers: Sage Publications, Thousand Oaks.

Bolton, G., Katok, E. and Ockenfels, A. (2004). How Effective are Online Reputation Mechanisms? An experimental investigation. Management Science, Volume 50 (11), pp. 1587-1602.

Botsch, J. and Luckner, S. (2008). Empirische Analyse von Bewertungskommentaren des Reputationssystems von eBay. In Proceedings of the Multikonferenz Wirtschaftsinformatik (Bichler, M., Hess, T., Krcmar, H., Lechner, U., Matthes, F., Picot, A., Speitkamp, B. and Wolf, P. Eds.), pp. 1199-1210, München.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, 12 (4), pp. 331-370.

Chen, K.-Y., Hogg, T. and Wozny, N. (2004). Experimental study of market reputation mechanisms. ACM, New York, USA.

Cheung, C. M. Y., Luo, C., Sia, C. L. and Chen, H. P. (2007). How do People Evaluate Electronic Word-of-Mouth? Informational and Normative Based Determinants of Perceived Credibility of Online Consumer Recommendations in China, Auckland, Neuseeland.

Conte, R. and Paolucci, M. (2002). Reputation in artificial societies: social beliefs for social order. In Multiagent systems, artificial societies, and simulated organizations (Weiss, G. Ed.). Kluwer Academic Publishers, Boston u. a.

Dellarocas, C. (2000). Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior. In Proceedings of the 2nd ACM Conference on Electronic Commerce (Jhingran, A., Mason, J. M. and Tygar, D. Eds.), pp. 150-157, Minneapolis.

Dellarocas, C. (2001). Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms. ACM, Tampa, Florida, USA.

Dellarocas, C. (2003). The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. Management Science, 49 (10), pp. 1407-1424.

Dellarocas, C. and Wood, C. A. (2008). The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. Management Science, 54 (3), pp. 460-476.
Gregg, D. G. and Scott, J. E. (2006). The Role of Reputation Systems in Reducing On-Line Auction Fraud. International Journal of Electronic Commerce, 10 (3), pp. 95-120.
Josang, A., Roslan, I. and Boyd, C. (2007). A Survey of Trust and Reputation Systems for Online Service Provision. Decision Support Systems, 43 (2), pp. 618-644.
Kapell, E. (2007). Deutschland und UK führen bei B2C. Lebensmittel Zeitung 2007-08-17, p. 28.
Miceli, M. and Castelfranchi, C. (2000). The role of evaluation in cognition and social interaction. In Human cognition and agent technology (Dautenhahn, K. Ed.), pp. 225-261, John Benjamins, Amsterdam
Peters, R. and Reitzenstein, I. (2008). Reputationssysteme im eCommerce - Funktionsweise, Anwendung und Nutzenpotenziale. HMD - Praxis der Wirtschaftsinformatik, 45 (261), pp. 43-50
Resnick, P., Kuwabara, K., Zeckhauser, R. and Friedman, E. (2000). Reputation systems. Communications of the ACM, 43 (12), pp. 45-48.
Resnick, P. and Zeckhauser, R. (2001). Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. Working Paper for the NBER workshop on empirical studies of electronic commerce.
Sabater-Mir, J. and Paolucci, M. (2007). On representation and aggregation of social evaluations in computational trust and reputation models. International Journal of Approximate Reasoning, 46 (3), pp. 458-483.
Schuler, H. (2004). Der Prozess der Urteilsbildung und die Qualität der Beurteilungen. In Beurteilung und Förderung beruflicher Leistung. 2 Edition (Schuler, H. H. Ed.), pp. 33-60, Hogrefe-Verlag, Göttingen u.a.
Sebralla, M.-L. (2008). Über 875 Millionen Menschen shoppen online. 40 Prozent mehr als noch vor zwei Jahren. Downloaded from http://de.nielsen.com/news/pr20080206.shtml on 2008-07-29.
Töpfer, A., Silbermann, S. and William, R. (2008). Die Rolle des Web 2.0 im CRM. In Handbuch Kundenmanagement (Töpfer, A. Ed.), pp. 651-675, Springer, Berlin, Heidelberg.
Vossen, G. and Hagemann, S. (2007). Unleashing Web 2.0. Morgan Kaufmann Publishers, Burlington.
Ward, M. R. and Lee, M. J. (1999). Internet shopping, consumer search and product branding. Journal of Product and Brand Management.
Ziegler, P.-M. (2008). Datenschützer verhängt Bußgeld gegen Bewertungsportal meinprof.de. Downloaded from http://www.heise.de/newsticker/Datenschuetzer-verhaengt-Bussgeld-gegen-Bewertungsportal-meinprof-de--/meldung/107123 on 2008-06-23.
Zwicky, F. (1969). Discovery, Invention, Research - Through the Morphological Approach. The Macmillian Company Toronto.