

On the Architecture of Social News Services

What is Behind the Story?

Christian Matt and Thomas Hess

Ludwig-Maximilians-Universität München,
Institute for Information Systems and New Media,
Ludwigstr. 28, 80539 Munich, Germany
{matt, thess}@bw1.lmu.de

Abstract. Due to the development of social news services users now have the possibility to influence the news offer they can choose from. Finally it is not a small number of editors but rather the voices of many that is responsible for creating, selecting and ranking news content. Despite the large potential of social news services, it is still not clear what the underlying technical foundations are and how these services function. Based on the empirical analysis of 37 social news services we identified an application architecture which maps the general functions and data pools that all services share. Building upon this common architecture, social news providers still have to take several functional design decisions which are described in detail. The identified framework serves as a basis for classifying social news services and for building further research upon.

Keywords: Social news services, framework, application architecture, design decisions, functionality.

1 Introduction

In the past, news production and selection was only exposed to human control, merely in form of a single or a small group of editors. At this, it is the editors' task to decide not only which news is published, but also which emphasis is devoted to every single news item, which will in turn have an effect on how the public rates the importance of the different stories. In this context editors, respectively mass media in general, fulfill "gatekeeper" and "agenda setter" functions [13,20]. Despite a vast number of news papers on the market, this role allocation may lead to a biased news selection.

Due to the new possibilities information and communication technology (ICT) offers, the power of editors and traditional mass media companies may shrink. Automatic news aggregators, such as Google News, combine news without any editorial interaction [17]. In this case, it is not the editors, but some artificial algorithm that decides which news is published. However, even if the potential for human abuse is much lower in this setting, the readers still do not have any control on the news selection process.

This has changed with the development of social news services. Social news services are news websites where users are asked to create and to evaluate news articles. The evaluation data is then taken as an input for a ranking algorithm that automatically calculates the news ranking. This ranking algorithm decides which news is displayed, where it is displayed and for how long it remains on the website. Therefore, it is not a small number of editors but rather the voices of many (“the wisdom of the crowds” [19]), that has now taken over the editorial power and has the task of identifying the best news stories [2]. For this, social news services rely on a large community of news consumers, who produce and evaluate content, as well as communicating actively with other community members. Thus social news services have finally put users in the middle of the news selection and creation process.

Whereas the conceptual forerunners “Slashdot” and “Fark” were already founded in 1997, resp. 1999, “Digg”, which is today arguably the most well-known and most widely used social news service, was founded in 2004 [6,9,15]. If the social news concept proves to be successful, this could have a significant impact on print newspapers as well as on online news portals, i.e. social news services could become a serious threat for these players. However, some online news portals may also decide to integrate social news functions into their website. Due to the modified gatekeeper and agenda setter functions, a success of social news services could even lead to fundamental changes in the process of opinion making on a societal level. Thus it is important to find out how social news service work and what their fundamental technical elements are. Although significant potential exists, the amount of research being performed on social news services is still limited. This is particularly evident for any kind of literature addressing the fundamental and structural basics of social news services.

Therefore, based on the empirical analysis of 37 popular social news services the paper at hand presents a common framework for illustrating the main characteristics of this new concept of news processing. Furthermore, we point out possible deviations from this framework and what kind of options social news providers have in order to set their service apart from others.

A literature review focusing on findings from related areas is given in the next section, followed by a description of the study design and a delimitation of the research scope. Afterwards, the application architecture, illustrating the general functions and data pools of social news services, is first presented and described and subsequently replenished by an outline of the different design decisions social news providers have to face. Finally the findings from this study are summarized, limitations are pointed out and future research steps are highlighted.

2 Related Work

Social news services emerged as part of the Web 2.0 development, which has been coined by O’Reilly in 2005 and which has become a buzz word since then [14]. Web 2.0 has reshaped online communication and collaboration patterns as well as the way information is produced and consumed [21]. In traditional publish-

ing models content creation has been executed by editors. Users had to accept the given content or to look after a different source of information. Due to the emergence of Web 2.0, individuals are now at the centre of the content creation process. This phenomenon has been referred to as “user-generated content”, i.e. content created by individuals instead of paid professionals [7].

User-generated content involves two challenges for Web 2.0 providers, which have been subject to substantial research efforts recently. The first research line tries to identify why users are willing to contribute content voluntarily. Based on large scale surveys, several studies have identified various reasons for and against producing user-generated content and how Web 2.0 providers can benefit from this [4,18]. The second research question focuses on how Web 2.0 providers can separate high and low quality user-generated content. Due to the missing editors’ control, the quality of user-generated content varies significantly from excellent pro-founded content to abusive content and spam. Successful Web 2.0 communities accumulate big amounts of user-generated content every day, thus manually supervising the content is in many cases extremely complex and expensive. Research has provided first strategies for identifying and filtering high quality content automatically [1].

These two research questions are essential for social news services as well. The development of social news services and their diffusion in the market has taken place over the past 5-10 years and based on the observations conducted in this study, the market seems yet to stabilize. Finding sophisticated answers to the aforementioned questions might even decide on the future of social news services. From a research perspective, so far most studies have focused on the social news service “Digg”. Lerman has made some empirically based contributions related to Digg and its news rating system, its user participation and on emerging voting patterns [9,10,11]. For instance, she has shown that users tend to like stories that were submitted by their friends. The evaluation and ranking of news is an essential part of social news services. Thus several studies have created algorithms and frameworks for an automatic ranking of social news articles [8,23].

Focusing on a media-centric perspective, Rölver and Alpar try to identify technologically driven changes in communication and media and what kind of role social news services play in this progress [16]. Furthermore, they perform an empirical analysis of several social news services in order to answer various research questions related to top ranked news items on social news websites.

Despite these valuable findings, to the best of our knowledge, there has been no previous work attempting to provide a general architectural and functional framework that serves as a common basis for classifying and categorizing social news services. Therefore, we aim to create an application architecture which covers all different forms of social news services and illustrates the main functions and data pools in order to provide a common ground for future efforts in this research area.

3 Study Design

Due to the dynamic, fast moving environment it is impossible to cover every single social news service worldwide. In order to identify a large number of social news services, Google has been used to look for several social news related search terms. Additionally, five freely accessible websites containing popularity rankings of English and German social news services have been identified. In most cases, these rankings were based on website traffic data. All news services found through Google or listed in the different website rankings have been aggregated and checked for availability as well as for meeting the following social news criteria:

- The news service in question is accessible through a website.
- Users can submit, consume and evaluate news stories.
- News stories meeting the basic requirements of the service need to be alike a regular news story, i.e. they need to have a substantial length (a link is not enough) and they need to add something “new”, i.e. containing a time-related attribute whereas the value of the information decreases over time.
- The evaluation process performed by the social news community has an effect on the ranking of the news items.

Finally 37 services have been identified and considered for further empirical research. In order to acquire and to evaluate information on the social news services in questions, a content analysis has been performed in June 2009. The analysis aimed to cover all steps a social news user performs when exploiting the underlying social news assumptions mentioned above. Based on the collected information, common elements have been identified and clustered into several groups. Potential relations between different clusters have been recognized and mapped in an application architecture. Possible design decisions and their characteristics have been outlined in the homonymous section of this paper. As part of this study the authors did not conduct any kind of performance measurement.

4 Application Architecture

All 37 social news services have been analyzed using a content analysis of the websites as well as extensive testing of the main functionalities. Based on this empirical analysis, one common architectural layout that is valid for all services from the sample has been identified. Thus apart from some small deviations, all social news services share one common basis to build their service upon. The application architecture mainly comprises the functionalities “sourcing”, “community”, “ranking algorithm”, the data pools “news pool”, “evaluations”, and the website sections “queue”, “news space”, “recycle bin” (cp. figure 1).

From a process view, a news item is first generated through user interaction and then automatically transferred to the news pool, which is a data pool storing all news items of a social news service. From now on, the news article is exposed to a consumption and evaluation process through the social news community,

which further leads to the storage of specific evaluation data for each news. A ranking algorithm combines the news-specific data stored in “news pool” and “evaluations” and automatically calculates a ranking. The ranking determines in which section of the page the news is to be displayed, resp. when the news is to be removed from the page. As every freshly created news item starts with zero evaluations, each news article is first placed in the queue and is only transferred to the regular news space or to the recycle bin after meeting specific criteria related to the popularity of the news item. The main elements of the application architecture are described in greater detail subsequently:

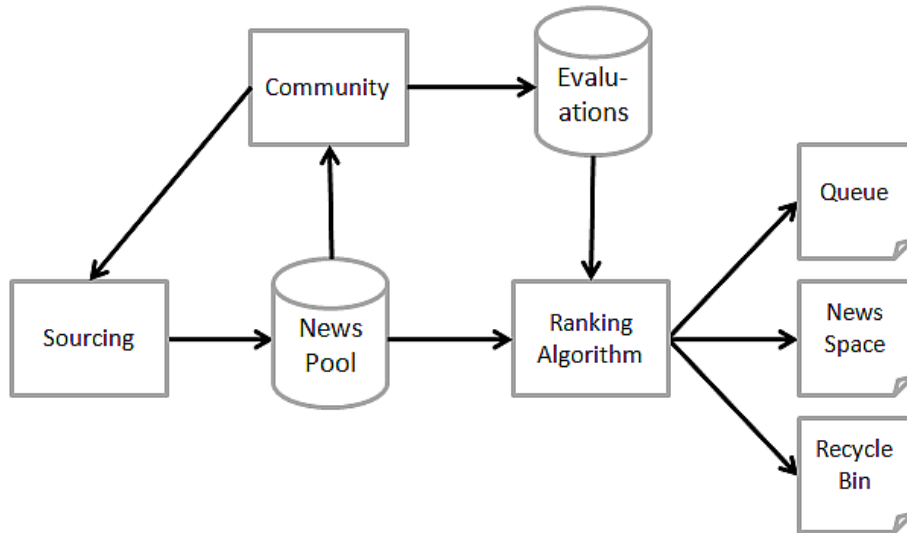


Fig. 1. Social news services architecture.

4.1 Sourcing

The sourcing function determines where the news items come from and in which way they are created. In contrast to traditional news websites, users of social news services have the possibility to create news stories themselves. In the simplest form regular website forms are used to submit news items. Some services provide tools and gadgets to assist users in directly linking and transferring content from the original source website to the social news service. However, the crucial point is that the news story is created through a manual interaction; the exact characteristic of the news transmission is of less importance. If services use scouts to create news items (cp. section “community composition”), the submission is still performed through manual interaction, although the subjects are

paid for their activities. For all social news services from the sample, users need to have a user account in order to create news items. Besides preventing spam, facilitating the community idea may possibly be the primary reason for this.

4.2 News Pool

The news pool contains all news items of a social news service. Novel news items consist of a headline, a message text, and a hyperlink to the original news source. Generally social news services deliver only a moderate amount of new content, which is mainly found in the message text. The form of the message text depends on the social news service in question, usually it is a short summary or an abstract of the original article. Some social news services allow for attaching pictures or videos in order to visualize the content.

Along with this, additional metadata is stored for each news item. The meta-data contains the time of the submission, the author of the recent news item, and an optional tag, which helps to retrieve the news item at a later date. In contrast to classical mass media communication, in the case of social news services, the author of the original source and the author of the social news article are not necessarily one and the same; in most cases they are actually different. Hereby authors of social news articles are only slightly responsible for the content they provide [16].

4.3 Community

The original intention of social news services is that social news providers are not to interfere with the news production and selection process, apart from preventing spam or illegal content. In this case, every step of the social news process is done through user interaction, from the submission of news items to the following consumption and evaluation process. Hereby the users of social news services take on a dual role, they become both consumer and producer, which is also called a “prosumer” [12]. Prosumers can be seen as ideal users as they both consume and produce content voluntarily. Successful prosumers can finally become an “opinion leader” [3] - the second category of users within a social news community. Based on their social status, opinion leaders possess an increased influence within the social news community, i.e. many users follow their contributions. Apart from prosumers and opinion leaders, social news communities usually also contain a number of “lurkers” [22]. Based on the idea of social news, lurkers can be seen as free riders as they simply consume content without actively contributing. However, user contributions are necessary to keep a social news service alive. There is no fixed target ratio for the distribution of these three different user categories within a social news community. Generally speaking, it is desirable to have a quite large share of prosumers, whereas a high share of lurkers may hurt the development of the social news service significantly.

Within the social news community, social interaction among members takes place and social ties between users are established [5]. Users can interrelate with each other by means of sending messages, commenting on news, recommending

news and becoming a follower of someone. A follower is notified on every activity of the user he/she is interested in. However, this is not a symmetric relationship, i.e. having a follower does not mean that a user needs to be a fan of his follower as well. A functioning submission and evaluation process is crucial to the dynamic development of the community as it is responsible for the final selection of news articles. The selection of news in turn has an effect on the selection of users that are attracted to a social news service.

Feedback from users is provided with votes or comments, which can be viewed as one form of intrapersonal communication. Since votes and comments are visible to the whole community this can be seen as one form of mass media communication. However, users have the possibility to join the discussion at any time, thus transforming it into intrapersonal communication again. Hence, mass media communication can easily become a form of intrapersonal communication and vice versa [16]. With fulfilling creation, consumption and evaluation functions, the community is the most essential part of a social news service.

4.4 Evaluations

The news story as a whole, i.e. including message text and headline, is part of an evaluation process. The question of whether or not the article meets the users' needs proceeds to be decided by the community. At this, social news services rely on democratic principles, so each vote, no matter from which user, has an equally weighted influence. Submitted votes are counted immediately, so users directly see the effect of their vote in form of a change in the score of the news item and potentially even a change in the ranking. At this, social news providers have to make sure that users can vote on each news item only once. In most cases users can change their opinion on a news item, but hereby the old vote is cancelled. Letting users vote several times on an item leads to potential abuse, e.g. heavily voting in favor of news submitted by friends. In order to prevent this, users have to be logged-in before submitting a vote.

4.5 Ranking Algorithm

The resulting score from the evaluation process serves as input data for the ranking algorithm. The algorithm continuously calculates the current position of each news story within the queue, and decides whether the story fulfills the criteria to enter the news space, whether it needs to remain in the queue or whether it should be removed from the website. If the story manages to enter the news space, the same process is applied again as the news item is still subject to user consumption and evaluation. Thus the algorithm permanently determines the ranking of the news story, and therefore the story can rise or fall based on further votes.

Once a news story has left the queue, either to the recycle bin or to the news space, it cannot be returned to queue. The queue contains only news items that have been created recently. If the story has collected a threshold number of votes it is transferred to the news space and remains there until it does not

receive a specific number of new votes anymore and thus proceeds to the recycle bin. News items in the recycle bin are not visible to users any longer and there is no possibility for the news to be transferred to the news space again. These articles may remain in the system for archiving purposes, but they are not part of the consumption and evaluation process anymore. Spam and abusive content is transferred to the recycle bin as well.

Articles listed on the news space are not necessarily the ones with the most all-time votes since the start of the social news service; the news space represents just the news with the highest amount of positive votes during a specific period of time. This consideration is necessary for serving up-to-date news. If the ranking was solely based on overall popularity, some formerly top ranked news items would remain on the news space for a long time due to their high score balance, regardless of becoming irrelevant and out-of-date.

4.6 Website Sections

Social news services consist of the three website sections queue, news space and recycle bin. As mentioned before, recently created news is first directed to the queue section of the website in order to receive its initial feedback. Successful news items proceed to the news space whereas less successful news stories and spam are directed to the recycle bin.

5 Design Decisions

The analysis of 37 social news services has provided evidence that based on the common architecture that all services share, social news providers have the possibility to customize their product offer in all of the three common functions that are mapped in the application architecture (cp. figure 1). Therefore the potential deviations can also be attributed to “sourcing”, “community” and “ranking algorithm” (cp. table 1):

Social News Function	Characteristic	Design Options		
Sourcing	Acquisition Method	User-based		User-based + Automatic
Community	Community Composition	Non-paid Users Only		Scout Support
	Evaluation Method	Point Scale	Positive/Negative	Positive Only
Ranking Algorithm	News Order	Algorithm Only		Algorithm + Editorial Review

Table 1. Design decisions for social news providers.

Since the different design decisions for each social news function are independent of each other, social news providers can implement any possible combina-

tion of parameters. Subsequently, the different design decisions are illustrated in further detail:

Acquisition Method. Apart from purely user-based sourcing, social news services may also implement automatic news sourcing in addition. In the automatic form, news stories are transmitted from the original source to the social news service without any user interaction. For this purpose, social news providers often have agreements with news agencies (e.g. “Associated Press”) or other large newspapers. Social news providers may want to use automatic sourcing additionally, if pure manual sourcing does not lead to a sufficient number of submitted novel news stories due to a lack of user participation. As user-generated content is one key element of social news services, using automatic sourcing only is not sufficient to fulfill the social news criteria.

Community Composition. The social news community, usually consisting of prosumers, opinion leaders and lurkers, is responsible for ensuring the quality of the news selection process. However, since spam and other inadequate content can be easily promoted, some services have integrated paid scouts to supervise the quality control process. Furthermore, some services also implement scouts, if the level of user interaction is not sufficient. At this, scouts shall help to foster communication and to mobilize the community with frequently posting news items and comments. However, using scouts for opinion making and generating new articles can lead to criticism regarding objectivity and censorship. Thus the presence of a vital community is essential for the sustainable success of any social news service.

Evaluation Method. Various schemas and scales are in use to evaluate news stories. Some social news services apply “thumb up/down” functions, others use point scales (usually 5 to 10 points at a maximum), where the more points a story receives the more popular it is. In some cases, users can only submit positive votes in form of a “thumb up” vote, but they do not have the possibility to negatively vote on a news story, apart from reporting the news item as abusive.

News Order. The criteria for leaving the queue and entering the news space or being removed from the website differ from service to service. Furthermore, the specific criteria remain confidential in almost all cases from the sample. This is because a well functioning news ranking algorithm can serve as a competitive advantage. Moreover, a public ranking algorithm could give users the opportunity to systematically exploit certain characteristics of the algorithm. Thus there are only few open source social news services for which the code can be analyzed and the main determinants of the ranking can be identified. Anyway, the amount of positive votes is presumably the most influential variable in the ranking process for all social news services.

As mentioned before, in the pure form of social news, there is no editorial selection and post-processing; the news selection and ranking is based only on user interactions and evaluations. Letting users decide which articles to publish

also implies the question whether regular users, without any profession in journalism, have the ability to distinguish between satisfactory and unsatisfactory articles. Therefore, some social news services hire editors that can alter the automatically calculated ranking of the algorithm in an arbitrary way in order to ensure a certain quality in the news selection process. However, using editors potentially leads to a lack of transparency as it is not clear any more what share of influence in the ranking process users still possess.

6 Conclusion

Based on the empirical analysis of 37 social news services, the study has identified the main functions and data pools that all services share and integrated these elements in a common application architecture. The identified model has revealed that social news services consist of the functions “sourcing”, “community” and “ranking algorithm”, the data pools “news pool” and “evaluations” and the website sections “queue”, “news space” and “recycle bin”. Social news services provide a selection of user-generated news items, which is then consumed and evaluated by the social news community. The evaluation process leads to a ranking of news, which is automatically calculated by an algorithm. Despite a common architectural foundation, social news providers still have to face some design decisions, attributable to and having effects on the identified main functions and data pools mentions above. These design decisions give social news service providers the possibility to differentiate their offering from those of their competitors. We have presented an application architecture that accounts for any of these possible deviations and provides a solid basis for further research. The study at hand shall also help to distinguish social news services from other related concepts with providing an overview of the essential elements and criteria. We also hope that the findings are beneficial both for social news providers as well as for traditional online news portals, who can now better evaluate the ease of integration of the social news concept into their current website.

However, some limitations apply to these findings. We aimed to cover the most popular services, as it seems impossible to capture every single social news service. Due to the fast moving environment, the popularity, and even the existence of some services may change quickly. Moreover, all findings result from an external user-centric perspective. Having access to the internal processes of the different services may help to reveal additional insights.

We encourage further research in the field of social news services, primarily in form of a deeper analysis of aforementioned aspects. A special focus should then be on understanding and enhancing the social news algorithms as they are crucial for the news selection process and therefore for social news services as a whole. Further empirical evidence is necessary to evaluate the success of social news services on a large scale and to identify which design decisions work best in praxis. We believe it is further worthwhile to expand the scope of the study in form of an extended sample of social news services and to control for other factors, such as cultural or country-related influences.

With the study at hand we have integrated the main functions and data pools of social news services as well as the possible design decisions social news providers have to face in one common framework. With this contribution we hope to have set a common ground that smoothes the way for further research on this exciting and possibly groundbreaking new way of news selection.

References

1. Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G.: Finding High-Quality Content in Social Media. In: *Proceeding of the First ACM International Conference on Web Search and Data Mining*, February 11-12, Stanford, CA, USA. pp. 183–194 (2008)
2. Bonabeau, E.: Decisions 2.0: The Power of Collective Intelligence, *MIT Sloan Management Review*, 50, 2, pp. 44–53 (2009)
3. Burt, R.S.: The social capital of opinion leaders, *The Annals of the American Academy of Political and Social Science*, 566, 1, pp. 37-54 (1999)
4. Daugherty, T., Eastin, M. and Bright, L.: Exploring Consumer Motivations for Creating User-Generated Content, *Journal of Interactive Advertising*, 8, 2, pp. 1–24 (2008)
5. Ganley, D. and Lampe, C.: The ties that bind: Social network principles in online communities, *Decision Support Systems*, 47, 3, pp. 266–274 (2009)
6. Gómez, V., Kaltenbrunner, A. and López, V.: Statistical analysis of the social network and discussion threads in slashdot. In: *ACM*, pp. 645-654 (2008)
7. Knight, A.: Who Is A Journalist?, *Journalism Studies*, 9, 1, pp. 117-131 (2008)
8. Ko, J., Kim, K., Kweon, O., Kim, J., Kim, Y. and Han, S.: Open Editing Algorithm: A Collaborative News Promotion Algorithm based on Users' Voting History. In: *International Conference on Computational Science and Engineering 2009*, August 29-31, Vancouver, Canada. pp. 653–658 (2009)
9. Lerman, K.: Social Networks and Social Information Filtering on Digg. In: *Proceedings of the International Conference on Weblogs and Social Media*, March 26-28, Boulder, CO, USA. pp. 1–8 (2007a)
10. Lerman, K.: User Participation in Social Media: Digg Study. In: *Proceedings of the International Workshop on Social Media Analysis (SMA07)*, *Conference on Web Intelligence and Intelligent Agent Technology*, November 2-5, Silicon Valley, CA, USA. pp. 255–258 (2007b)
11. Lerman, K. and Galstyan, A.: Analysis of Social Voting Patterns on Digg In: *Proceedings of the ACM SIGCOMM Workshop on Online Social Networks*, August 18, Seattle, WA, USA. pp. 7–12 (2008)
12. Lister, M., Dovey, J., Giddings, S., Grant, I. and Kelly, K.: *New Media: A Critical Introduction*, Routledge, London (2003)
13. McCombs, M.E. and Shaw, D.L.: The agenda-setting function of mass media, *Public Opinion Quarterly*, 36, 2, pp. 176–187 (1972)
14. O'Reilly, T., <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
15. Rangwala, H. and Jamali, S.: Co-Participation Networks Using Comment Information. In: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC, USA. AAAI Press, Menlo Park, CA, pp. 315–318 (2010)

16. Rölver, M. and Alpar, P.: Social News, die neue Form der Nachrichtenverteilung? In: M. Rölver and S. Blaschke (Eds.) Web 2.0-eine empirische Bestandsaufnahme, 295–330. Vieweg Teubner, Wiesbaden (2008)
17. Schroeder, R. and Kralemann, M.: Journalism Ex Machina - Google News Germany and its news selection processes, *Journalism Studies*, 6, 2, pp. 245–247 (2005)
18. Stöckl, R., Rohrmeier, P. and Hess, T.: Why Customers Produce User Generated Content. In: B. H. Hass, G. Walsh and T. Kilian (Eds.) Web 2.0 neue Perspektiven für Marketing und Medien, 272–287. Springer, Berlin/Heidelberg (2008)
19. Surowiecki, J.: The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations, Doubleday Books, New York (2004)
20. White, D.M.: The Gatekeeper. A Case Study in the Selection of News, *Journalism Quarterly*, 27, 3, pp. 383–390 (1950)
21. Wijaya, S., Spruit, M. and Scheper, W.: Webstrategy Formulation: Benefiting from Web 2.0 Concepts to Deliver Business Values. In: Miltiadis D. Lytras, Ernesto Damiani and Patricia Ordóñez de Pablos (Eds.) Web 2.0: The Business Model, 373–384. Springer, New York (2008)
22. Xia, M., Huang, Y., Duan, W. and Whinston, A.: Ballot box communication in online communities, *Communications of the ACM*, 52, 9, pp. 138–142 (2009)
23. Xu, G.-L., Liu, N.-Z., Feng, J.-L. and Liu, Y.-C.: A Voting Algorithm Based on Attribute Theory for Social News Sites, *CAAI Transactions on Intelligent Systems*, 4, 2, pp. 118–121 (2009)