

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Financial news content publishing on Youtube.com

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© IEEE

VERSION

VoR (Version of Record)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Sykora, Martin D., and Marek Panek. 2019. "Financial News Content Publishing on Youtube.com". figshare.
<https://hdl.handle.net/2134/6420>.

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Financial News Content Publishing on Youtube.com

Martin D. Sykora, Marek Panek

Research School of Informatics, Computer Science Department

Loughborough University

Loughborough, LE11 3TU, United Kingdom

E-mail: m.d.sykora@lboro.ac.uk, marek@lh.pl

Abstract- Recently a number of academic publications have investigated various properties and dynamics of author's contributors on online Web 2.0 communities. The most intensively examined communities have until now been discussion boards and blogs. In this paper we look at, and identify revealing patterns of content publishers on YouTube. This is the most popular community based service on the world wide web as ranked by the frequency of visited webpages. The research is motivated in terms of gaining better understanding of content publishing habits within financial news topic specifically and we investigate the potential for trend detection in financial markets area. A number of side issues, such as publisher bias and various video properties are also examined. It turns out, the number of video submissions by professional authors has increased considerably in recent months. There is significant quantity to allow statistical analysis of events. As we show, the potential value of the data on YouTube cannot be ignored any longer. We present recent results and work on a project that investigated a completely new dataset, not really considered in previous literature.

I. INTRODUCTION

Youtube is the most popular community based website, based on web traffic which has been constantly growing (since its founding in February 2005), earning it ranking in top 3 most frequently visited websites in the world, according to alexa.com. It reaches about 5% of internet users in a day and generates 20% pageviews on Internet. A number of companies seem to have realised this and there now seems to be a trend of traditional media companies starting to make heavy use of media sharing websites, youtube in particular. This seems to be motivated by a number of factors, among others; receiving user community exposure and acknowledgement through providing quality video content and direct advertising by pushing products or services onto youtube community and visitors. In this paper we investigate identities of YouTube video uploaders in financial news and the properties characteristic of uploaded content. It is becoming important to understand user properties and dynamics in Web 2.0 systems. Rodrigues [23] for example analysed response habits of online discussion board members and found interesting yet different network collaboration structures for various discussion topics. Choudhury et al, [8] analysed response times and magnitude of activity on commenting blogs by categorising posters of blog comments into various information roles, such as early responders, loyals, late trailers and outliers.

The rest of this paper is organized as follows. In section 2,

background and past literature is discussed, section 3 presents some system considerations, with results being demonstrated in section 4, followed by conclusions drawn and further work highlighted in section 5.

II. BACKGROUND

Web 2.0 is not a new version of the World Wide Web, but it is a different approach to user interaction with web. It breaks the old paradigm of communication between webmasters and users by allowing web-users to create and upload web content through an interaction with user interface [31]. Such an evolution of principles of using the web, stimulated rapid explosion of many community-based portals like myspace.com, facebook.com or youtube.com [31, 28, 22, 25].

A. Becoming serious

Most people have heard of twitter, facebook or blogs. These web 2.0 systems started off for "fun". They provided means for a novel way of keeping in touch and sharing information with other users. In recent US presidential elections twitter and facebook were heavily employed and blogs are now used by companies as synonymous communication channels with customers. Web 2.0 media seems to have gained considerable significance in the society. This can be backed up by research [13, 1, 11, 12, 21, 32, 26].

Youtube among with other media sharing websites such as Revver.com and others, have become popular mainly for entertainment. Most videos on youtube in particular, used to be on music, comedy and general entertainment. With its rising popularity however, many users now upload every and any kind of imaginable videos, ranging from personal videos, to mobile phone videos of s US soldiers in Iraq. It seems video sharing is now increasingly used for serious purposes. A number of companies have for example launched advertising campaigns via youtube (e.g. Nike)¹ and youtube was even used as an advertising channel for tobacco companies [14]. In this paper we investigate whether financial news are well represented on youtube. Our results show much promise.

B. Previous Work

A number of academics have investigated web 2.0 applications, such as wikis [15, 5, 3, 2], blogs [20, 5, 27, 17, 8], podcasts [5], community media sharing [7, 6, 16, 19, 18, 14] or discussion boards [29, 30, 4, 10, 9, 24, 23]. We are

¹ <http://www.youtube.com/blog> is a good information source about most campaigns and promotion projects.

now starting to better understand these new media, by analysing user behaviour and habits. It must be pointed out that in relation to finance, most work, up to this point was done on blogs or discussion boards. There was no investigation into media sharing applications in conjunction with finance. Therefore, this paper presents a valuable contribution to the field.

A brief rundown of some existing literature on youtube follows. Havley and Keane in “Exploring Social Dynamics in Online Media Sharing” [19] found that youtube users prefer browsing rather than uploading their own videos. On average, they view 966 clips against 11 submitted files. Moreover, community facilities available after signing-in are not widely exploited. Most users are anonymous and do not participate in various web 2.0 activities like commenting, video responding or rating. Furthermore, only small percentage of users has subscriptions for favorite clips or channels.

In the paper “I tube, You Tube, Everyone Tubes: Analyzing the World’s Largest User Generated Content Video System” [6], it was found that Youtube on average experienced 65’000 new video submissions per day. Similarly, to “Exploring Social Dynamics in Online Media Sharing” this research concludes that there is very little web 2.0 activity, most users stay anonymous and even if they register, the level of their participation is low. Authors also draw attention to very important aspect, which is content aliasing. As anyone can upload nearly any content on Youtube, there exist multiple copies of videos relating to a single event. Hence there are many identical videos submitted by different users and this probably dilutes popularity of a corresponding video. “Analysis of Online Video Search and Sharing” written also by Havley and Keane [18], reveals that in general all of the videos that receive greater number of hits, have more descriptive meta information. It means that more textual information in the form of tags, title, and description makes these pages more popular than others. This is probably a consequence of the fact that the search algorithm picks these videos up with more likelihood and matches them with search terms. [14] investigated whether youtube videos promote smoking. It was found that search term “smoking” had returned 29’325 results on YouTube. After in depth analysis of the content of top 50 clips, it turned out that YouTube is used as a channel for advertising tobacco.

It is clear that with 65’000 new video submission every day [6], and as we show in section IV, still rising, even though [19, 6] point out low web 2.0 activity in relation to the entire user base, youtube is clearly a powerful web 2.0 phenomenon. [14] presents an interesting aspect of one of many ways in which advertisers make use of youtube.

III. CONSIDERATIONS

A. Input data

Youtube is a free community-driven website through which registered users can upload unlimited number of videos and share them with other users. Each video must be given a title and be assigned to a specified category (e.g. News, Music, Entertainment). A publisher can optionally provide

further details.

Every uploaded video on youtube is in the form of a video file and a set of related meta data describing the file. Such meta data contains video title, description, category, date of submission, view count, duration and author. Since youtube is a social website it also allows users to comment, rate (1 out of 5) and submit response videos. Videos can also be tagged with arbitrary tags that might help identify a video better. The meta data attached to a file is shown in Fig. 1.



Fig. 1. Video meta data associated with a youtube file. Highlighted fields represent three main streams of textual data

In Fig 1. Title, Description, Category and Tags provide basic information as to the content of a video clip. Author, Date, communicate who and when submitted the file. Ratings and Duration tell us a little bit more about the video. Viewcount and Comments can be quite important for an analysis, the former can be important in judging the popularity of a video and the latter also provides us with collective opinion about a video contribution in textual form. Related videos are video recommendations that might be of similar content to the target video. This is done by an algorithm that is kept secret². Response videos are actual file responses to the original clip, and are usually used to create a so called “video-debate“.

B. Data Retrieval

In November 2006 youtube was acquired by Google, Inc. and after a few months Google implemented their API called GDATA, enabling developers to integrate systems with youtube platform. We made use of this API to extract as much financial market video meta data as possible for the entire available time period. Videos based on search keywords FTSE, DOW JONES, NASDAQ, NIKKEI, CAC, DAX³, and also related and response videos were retrieved. Since each of these videos has a lot of meta data associated, we extracted altogether about, 90’400 videos, 89’000 tags and 3’749’000 comments on submissions related to finance news.

A number of issues with API were encountered, some errors were discovered and some limitations imposed by the youtube terms and conditions. It was ensured that terms and conditions were complied with by an appropriate implementation of our scripts. Main bulk of data extraction process took over 7 days. After this we run extraction scripts daily to ensure database was kept up to date with recent video

² According to youtube support section, <http://help.youtube.com/support/youtube/bin/answer.py?hl=iw&answer=95612> (Consulted on 2 April 2009), this algorithm is kept secret.

³ UK, USA, Japanese, French and German stock indices respectively

submissions. Since our system required manipulation of large amounts of data, a powerful server set up; DELL PowerEdge 1950 III with two CPUs Intel Quad Core Xeon 2.66GHz, 8GB of RAM memory and SAS hard drives running Linux Debian OS and MySQL was used in experiments.

C. Querying Youtube

Retrieval of youtube videos can be challenging for at least two reasons. First, there are a number of limiting conditions enforced on the amount of retrievable clips on particular topics and secondly ensuring that videos contain the target content requires some filtering. To avoid problems with the first issue we introduced coping mechanisms into our scripts. In order to deal with the second issue, it was ensured that searches were constrained. We found that best way to constrain a search was to impose restrictions on the tags that could be associated with a video submission. It was found that filtering content by a combination of tags was very effective. Queries that were only filtered by keywords or topic (i.e. News) often returned too much unrelated content.

IV. RESULTS

First and foremost, the question of how much market news is really submitted on youtube arises. That is, it is very important to acknowledge that maybe there is too little video submissions on financial news within youtube. Contrary to this however we found that youtube contains a large number of file submissions. Samples of files have been inspected, most of the videos are of relatively high quality, often reporting on financial events throughout the day or analysing possible strategies for the next day or week(s). As can be appreciated in Fig. 2., especially in the period after September 2008 there was a rapid increase in video submissions. This could be associated to the crisis. During this time, awareness of crisis and risk of recession became widespread throughout (for example. bankruptcy of Lehman Brothers at this time dragged attention of many reporters to financial collapse and economic instability). However it could just be the rapid increase in overall popularity of youtube. Therefore benchmark data was needed to associate this trend with one or the other reason. We hence retrieved similar quantity of videos from 3 independent categories, namely; music, entertainment and sport. A comparison of monthly time series data for each category showed that only financial video submissions experienced a rapid increase (statistically significant) in fourth quarter of 2008, see Fig.2 below.

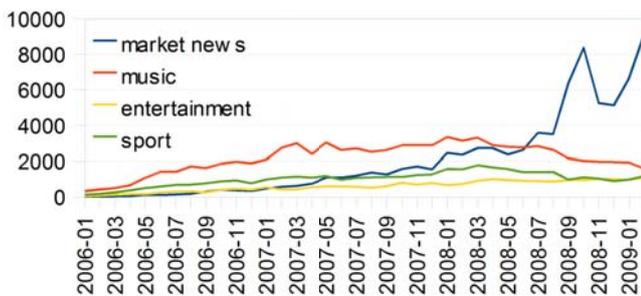


Fig. 2. Comparison of video submissions over various categories

Given this rapid increase in financial news video submissions, we were interested into who actually uploads videos and how these video uploads are distributed over time.

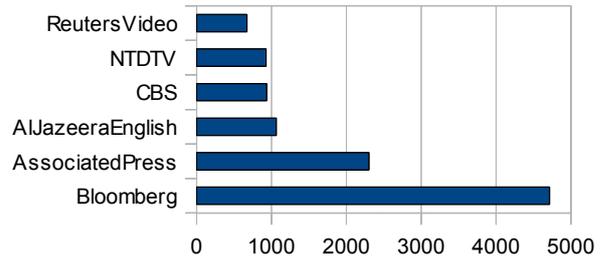


Fig. 3. Ranking of authors (top 5 and Reuters)

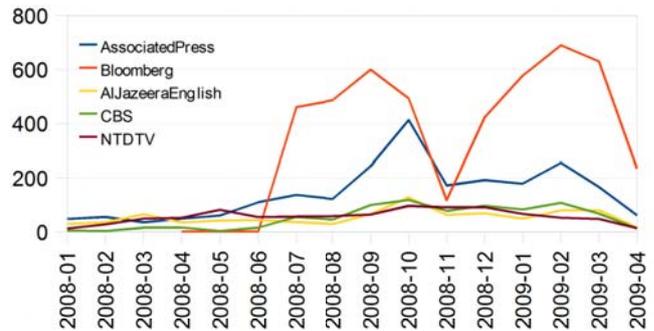


Fig. 4. Video submissions of selected popular authors

Figures Fig. 3 and Fig. 4 show that not only passionate users upload videos but apparently, most active users turned to be worldwide press agencies like Bloomberg, Associated Press, Al Jazeera or CBS. They became much more active in second half of 2008 and maintain this trend in 2009.

TABLE I
AUTHOR STATISTICS

Author	Video submission properties					
	Avg. Title Length	Avg. Desc. Length	Avg. View Count	Avg. Rating	Avg. no. of Raters	Avg. Duration (secs)
Bloomberg	51	126	308	1.81	1	268
Associated Press	39	289	10'082	3.79	27	85
AlJazeera English	50	284	13'090	4.62	41	416
CBS	26	172	34'363	3.98	100	192
NTDTV	36	1'065	2'776	4.39	8	122
Reuters Video	28	272	5146	3.51	8	115

Table I shows quite interesting data about these authors, i.e. different attributes of their videos and how users perceive these clips. For example, Bloomberg tends to submit videos with very short descriptions of 126 characters and duration of 268 seconds on average. Associated Press's videos average description length is more than twice greater (289 characters)

however videos themselves have shorter duration (84 seconds on average). The community appreciates Associated Press videos more and manifests that by much higher rating (3.8 against 1.8) and intensity of rating (27 against 1 rates per video). AP videos have average view count of 10'089 whereas the same figure for Bloomberg equals 308. It would confirm the findings of [18] mentioned in section II. As AP has longer descriptions, it therefore has more keywords to match by search engine. This may be the reason of such disproportion of view counts in contrast to Bloomberg.

Inspecting these attributes over a total of 1'000 publishers we found following statistically significant (all beyond 95% significance level) correlations:

- Avg. View Count – Avg. no. of Raters 0.852
- Avg. Title Length – Avg. Desc. Length; 0.315
- Avg. Title Length – Avg. Rating; 0.187
- Avg. View Count – Avg. Rating; 0.133
- Avg. Rating – Avg no. of Raters 0.174
- Avg Rating – Avg Duration 0.149

Most of these are self explanatory, such as the relationship between number of people who saw a clip and the number of people who also rated a clip is clearly correlated. An interesting insight is provided by the second correlation, that is that most publishers who use descriptive (longer) titles also tend to use more descriptive video descriptions or summaries. The correlation of 0.133 between view count and increasing average rating, indicates that users generally appreciate videos on youtube. This does not go against intuition that since there is a recession, ratings should be negatively correlated. This is because ratings do not usually address the content of the video messages itself. Instead, as we discovered ourselves, ratings are generally only relevant towards quality, accuracy and stylistic factors of a video.

We attempted to built a number of regression models to describe some of the attributes in Table 1. Models for different attributes as dependent variables were optimised on 1'000 instances of publishers. No significant or interesting model other than the bi-variate correlation relationships described above were found using simple linear regression. As an example a linear regression model for the number of raters (y), is of the form, see equation 1.

$$y = 7.79 - 0.957 * t + 19.59 * r + 0.003 * v + \epsilon \quad (1)$$

where t , r and v stand for title length, rating and view count, respectively. This is not quite so interesting and can be summarised as; number of raters depend on increasing number of viewers (standardised beta = 0.842), to some degree on average rating and a smaller title length.

A. News Topics

We looked at numerous financial news publishers that actively report on the youtube platform, however what is it that they report on? It turns out the most accurate way to categorise videos by sub-topics is to filter them over associated tags (see III subsection C.). A number of recent

topics that received tremendous attention over last few months were selected and aggregated over months.

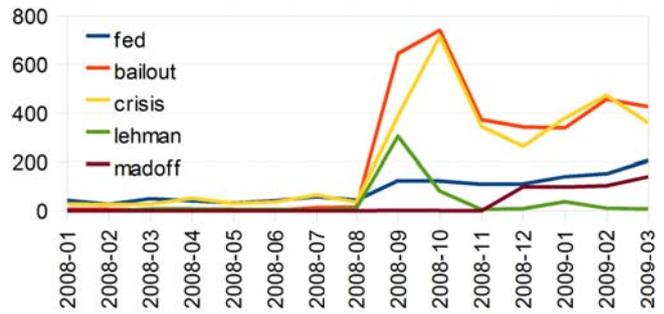


Fig. 5. Video submissions related to significant events / subjects

In Fig. 5. video submissions related to significant events in the world, are presented. When Lehman Brothers collapsed, governments of different countries tried to rescue banks from bankruptcies through bailouts. It was one of many signs that global economic slowdown and a recession was approaching. Traditional media devoted much attention to this problem. There is hence strong increase of videos containing tags 'bailout' and 'crisis' in September 2008 to above 700 videos in the month. Likewise, fall of Lehman brothers was reflected on youtube by about 300 of videos in the same month. The fraud case of Madoff's financial pyramid was revealed in December and resulted in loud, public arresting of Bernard Madoff. More than 100 videos in December and in consecutive months refer to the issue.

B. Publishers in other Categories

We inspected financial news publishers on youtube in quite some detail. However, what about other categories of video topics on youtube, who is responsible for their submissions. Table II presents top 10 publishers of content in music, entertainment and the sports categories.

TABLE II
TOP 10 PUBLISHERS IN DIFFERENT CATEGORIES

Rank	Market news	Music	Entertainment	Sports
1	Bloomberg	SonyBMG	Machinima	NBA
2	Associated Press	Hoover4000	Booklvr1256	Expertvillage
3	AlJazeera English	Andromeda 881	Ginaya	NHLVideo
4	CBS	Jwcfree	CBS	RandomPokr
5	NTDTV	Rpoland	Uluvshane	TNAwrestlmg
6	Thermal1	Somedia	Ichglotzutube	TeamFlight
7	Expertvillage	Parlophone	JPizzle1122	Pennyccw
8	Ehowfinance	Kinagrannis	QuickUploadr	Jon747
9	Reuters Video	Sori1004jy	LaaDida3	Ialousse24
10	FXbootcamp	SilvaGunner	Newscribe	TrueWrestlmg

As can be appreciated from top 10 publishers in Table II, most content within Music and Entertainment seems to be submitted by individual users and / or artists, except the odd big label company, such as “SonyBMG” in Music. Within sports and market news on contrary, professional publishers seem to provide bulk of the content. NBA or NHL, are the official channels (publishers) for the U.S. Basketball and Hockey leagues respectively.

IV. CONCLUSIONS

Individual content publishers within financial news domain, on the most popular web 2.0 media sharing platform, youtube.com, were investigated in this paper. A unique and relatively novel dataset was analysed. We looked at who publishes content and in what quantities. Various properties and community feedback was analysed, and a number of interesting relationships discovered. An investigation of publisher dynamics over time clearly showed that numerous reputable (financial) news agencies, now actively submit video content. The strong presence of established agencies on youtube seems to have a good effect on reliability and quality of submitted content.

Most of all this paper highlights the significance web 2.0 communities have gained. In fact it seems that for Bloomberg and others, youtube is becoming an information publishing channel of importance. The crisis has pushed financial news reporting on youtube to before unprecedented levels. There is no reason why this should change and in fact, media sharing services are set to continue in this trend of growth.

REFERENCES

- [1] L. A. Adamic and N. Glance. The political blogosphere and the 2004 u.s. election: divided they blog. In *International Conference on Knowledge Discovery and Data Mining*, 2005.
- [2] B. Adler, K. Chatterjee, L. de Alfaro, M. Faella, and V. R. I. Pye. Assigning trust to wikipedia content. In *WikiSym 2008: International Symposium on Wikis*, 2008.
- [3] B. Adler, L. de Alfaro, I. Pye, and V. Raman. Measuring author contributions to the wikipedia. In *WikiSym 2008 - International Symposium on Wikis*, 2008.
- [4] W. Antweiler and M. Z. Frank. Is all that talk just noise? the information content of internet stock message boards. *Journal of Finance*, pp. 59-1259, 2004.
- [5] M. N. K. Boulos, I. Maramba, and S. Wheeler. Wikis, blogs and podcasts: a new generation of web-based tools for virtual collaborative clinical practice and education. In *BMC Medical Education* 2006.
- [6] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: Analyzing the worlds largest user generated content video system. In *Proceedings of ACM Internet Measurement Conference (IMC)*, San Diego, USA, 2007.
- [7] X. Cheng, C. Dale, and J. Liu. Understanding the characteristics of internet short video sharing: Youtube as a case study. In *Procs of the 7th ACM SIGCOMM Conference on Internet Measurement*, pp. 15-28, 2007.
- [8] M. D. Choudhury, H. Sundaram, A. John, and D. D. Seligmann. Can blog communication dynamics be correlated with stock market activity? In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, 2008.
- [9] P. M. Clarkson, D. Joyce, and I. Tutticci. Market reaction to takeover rumour in internet discussion sites. *Accounting and Finance*, Vol. 46, pp. 1-31, 2006.
- [10] S. Das, A. Martinez-Jerez, and P. Tufano. einformation: A clinical study of investor discussion and sentiment. *Financial Management*, Vol 34, pp. 103-137, 2005.
- [11] P. Dwyer. Building trust with corporate blogs. In *Proceedings of International Conference on Weblogs and Social Media (ICWSM07)*, 2007.
- [12] L. Emova and J. Grudin. Crossing boundaries: A case study of employee blogging. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, 2007.
- [13] H. Farrell and D. W. Drezner. The power and politics of blogs. *Public Choice*, Vol. 134, pp. 15-30, 2008.
- [14] B. Freeman and S. Chapman. Is youtube telling or selling you something? tobacco content on the youtube video-sharing website. Technical report, tobaccocontrol, 2007.
- [15] J. Giles. Internet encyclopaedias go head to head. *Nature*, Vol. 438, pp. 900-901, 2005.
- [16] P. Gill, M. Arlitt, Z. Li, and A. Mahantix. Youtube trac characterization: a view from the edge. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pp 15-28, 2007.
- [17] P. Gloor, J. S. Krauss, S. Nann, K. Fischbach, and D. Schoder. Web science 2.0: Identifying trends through semantic social network analysis. Technical report, Social Science Research Network, 2008.
- [18] M. J. Halvey and M. T. Keane. Analysis of online video search and sharing. In *Proceedings of the eighteenth conference on Hypertext and hypermedia*, pp. 217-226, 2007.
- [19] M. J. Halvey and M. T. Keane. Exploring social dynamics in online media sharing. *Proceedings of the 16th international conference on World Wide Web*, pp. 1273-1274, 2007.
- [20] B. A. Huberman and L. A. Adamic. Information dynamics in the networked world. *Lecture Notes in Physics*, Vol. 650, pp. 371-398, 2004.
- [21] J. Huh, L. Jones, T. Erickson, W. Kellogg, R. Bellamy, and J. Thomas. Blogcentral: the role of internal blogs at work. *CHI 07 extended abstracts on Human factors in computing systems*, pp. 2447-2452, San Jose, USA, 2007.
- [22] C. Li and J. Berno. *Groundswell: Winning in a World Transformed by Social Technologies*. Harvard Business School Press, 2008.
- [23] E. M. Rodrigues. Mining online communities. In *Proceedings of the UK Symposium on Knowledge Discovery and Data Mining*, 2009.
- [24] S. Sabherwal, S. K. Sarker, and Y. Zhang. Online talk: does it matter? *Managerial Finance*, Vol. 34 pp. 423-436, 2008.
- [25] C. Shirky. *Here Comes Everybody: The Power of Organizing Without Organizations*. Penguin Press, reprint edition edition, 2009.
- [26] A. Stocker and K. Tochtermann. Investigating Weblogs in Small and Medium Enterprises: An Exploratory Case Study, *Bis 2008 workshops proceedings*. pp. 95-107, 2008
- [27] A. Sun, M. Hu, and E. Lim. Searching blogs and news: A study on popular queries. *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 729-730, 2008.
- [28] D. Tapscott and A. D. Williams. *Wikinomics - How mass collaboration changes everything*. Atlantic Books, 2008.
- [29] J. D. Thomas and K. Sycara. Integrating genetic algorithms and text learning for financial prediction. *Proceedings of Data Mining with Evolutionary Algorithms*, pp. 72-75, Las Vegas, USA, 2000.
- [30] R. Tumarkin and R. F. Whitelaw. News or noise? internet postings and stock prices. *Financial Analysts Journal*, Vol. 57, pp. 41-51, 2001.
- [31] G. Vossen and S. Hagemann. *Unleashing Web 2.0: From Concepts to Creativity*. Morgan Kaufmann, 2007.
- [32] S. Yardi, S. Golder, and M. Brzozowski. The pulse of the corporate blogosphere. *Conference Supplement of CSCW*, 2008.

