

Chapter 13

From Increased Availability to Increased Productivity: How Researchers Benefit from Online Resources

Joe Strathern, Samer Awadh, Samir Chokshi, Omar Addam, Omar Zarour, M. Ozair Shafiq, Orkun Öztürk, Omair Shafiq, Jamal Jida, and Reda Alhajj

Abstract With the rapid growth of data on World Wide Web, this paper presents researchers with an analysis that summarizes data collection and evolution features in the World Wide Web. We have reviewed a large corpus of published work to extract the research features supported by advanced Web features, i.e., blogs, microblogging services, video-on-demand web sites and social networks. We have presented summary of our review and analysis. This will help in further analysis of the evolution of communities using and supporting the features of the continuously evolving World Wide Web.

Introduction

The World Wide Web (WWW) continues to grow at an unpredictable rate. The first decade of the new millennium has seen the birth of so-called “Web 2.0” services through which massive amounts of data are continuously generated at high rate. Much of this data are user generated. The once-complex task of identifying and categorizing web communities is even more complex.

In this seemingly endless flood of data, researchers interested in studying a specific phenomenon first need to start by identifying a suitable data source. For

J. Strathern · S. Awadh · S. Chokshi · O. Addam (✉) · O. Zarour · M.O. Shafiq · O. Öztürk · O. Shafiq
Department of Computer Science, University of Calgary, Calgary, AB, Canada
e-mail: omaddam@gmail.com

J. Jida
Department of Informatics, College of Science, Lebanese University, Tripoli, Lebanon

R. Alhajj
Department of Computer Science, University of Calgary, Calgary, AB, Canada

Department of Computer Science, Global University, Beirut, Lebanon

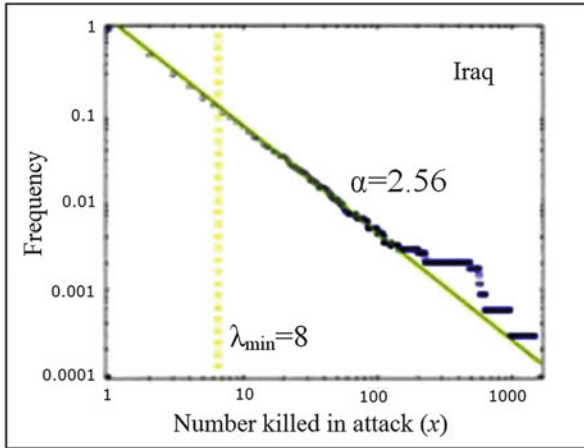


Fig. 13.1 Statistical data gathered from news articles

instance, researchers studying the patterns of how breaking news spread first need to identify a data source (such as a social network) and ensure that source is suitable for the purpose of their study. The data related to the identified source is then analyzed for effective knowledge discovery that covers the domain to be investigated.

In this paper we undertake the task of trying to present the research community with a categorization of some Web 2.0 services and features found in each service. Our hope is that the categorization will help researchers jump-start their work by providing a research-guided categorization. This work is in no way exhaustive, and is only intended as a starting point. As the evolution of the World Wide Web continues, continuous future work is needed to maintain the relevance of the results of this study.

Next we present our motivation for conducting this research, followed by survey of relevant work, and present our findings followed by conclusions.

Motivation

The WWW appears to be an ever-growing repository of information. Data is flowing in from variety of sources and need to be analyzed for maximizing the benefit to users who are interested in figuring out the best usage of the data to at least compensate for the effort of collecting the data. This wealth of data can aide every group of domain experts in a specific fiend covered by the data, e.g., computer scientists, sociologists, psychologists, and even security forces analyze, understand, and theorize about the real world. One example is the work of Bohorques et al. [2] in which they used raw statistical data gathered from news articles, NGO's, and cable

news and resulted in the graph shown in Fig. 13.1. To the authors, this seemingly ordered distribution of the number of casualties killed in attack (x) to the frequency of attacks was surprising. When they repeated their work in other conflicts (such as Afghanistan, Senegal, and Colombia), the same distribution appeared. This allowed the authors to present a formula that could help answer questions such as why the conflict in Iraq continues until today.

Such power of raw data motivated us to ask what can Web 2.0 services tell us about the world. The main research question this paper tries to answer is what types of raw data can micro-blogging, blogging, social networks, and video web sites support research in, and how we can use the information collected from these information resources in a meaningful and useful way to increase the productivity of the raw data while having increase in availability.

Survey and Analysis

This section describes in detail our analysis of the features and formation of communities over the World Wide Web from Blog, Microblogging, Video sharing and Social networks.

Blog

Blogs are a popular source of a large amount of information that is posted and accessed daily. “As the amount of available information and its ‘dispersion’ among many related blogs increases, it becomes more difficult to get a complete picture of the public opinion on a particular topic” [18]. This is why we present some of the applications of blog mining in research and to solve some practical problems.

In [6], De Choudhury et al. explore the possibility of whether blog communications can be correlated to stock market movement. If such a correlation exists, corporations can identify the moods of clients after the release of a new product, for example. Furthermore, such a correlation, if it exists, can help create targeted advertising for clients. The authors develop a model to study and analyze such correlation. The model was then applied activity in the blogosphere and the stock market. The stocks of Apple, Google, Microsoft, and Nokia were used in the study and the results were encouraging. This is not the only study pointing at the blogosphere as a source for data on the movement in the stock market to (see [20] for another example).

Blogs can also be used as an accurate measure of public opinion with relation to movies, music, and so on. The authors of [18] took movies as a case study, with possibility to easily modify the approach for music. In the case of movies, the authors were able to show a significant correlation between “the amount of buzz a movie generates and its critical or financial success” (Fig. 13.2).

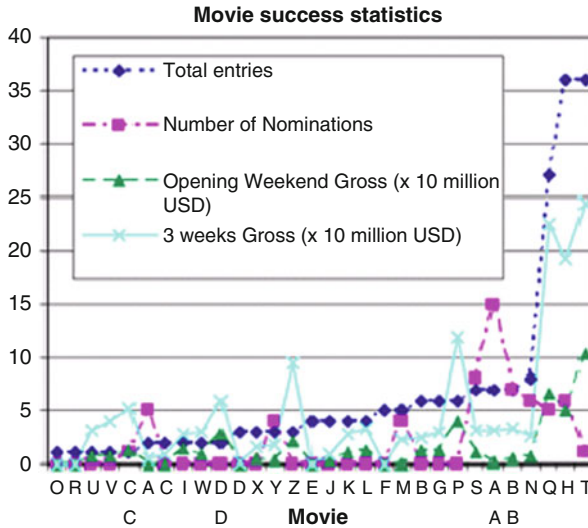


Fig. 13.2 Entry, nomination and gross statistics for each movie

The role the blogosphere can play in education does not seem to be obvious. Shaohui and Lihua [16] for instance, hypothesizes about the potential role blogs can play in education, but the results are not clear. The other question to be asked is whether blogs have been used in education. Also, the answer to that question seems unknown (Fig. 13.3).

Microblogging Services

Microblogging services, represented by Twitter for example, focus on sharing very short updates. The phenomenal success of Twitter and the hundreds of millions of users around the world make the service an attractive source of data for research.

The types of research that can be done in microblogging sites are similar to those that can be done with blogs (see [1, 19] for example), but microblogging sites provide a few added advantages. The first is that those shared “updates” are usually time-sensitive, which potentially makes microblogging communities a source to study how breaking news, as an example of a time-sensitive topic, spread in populations. In [14], the authors present Buzzer, a real-time news recommendation system. Combining Twitter chatter and user’s RSS feeds preferences, Buzzer can recommend news articles of interest to the user, in real-time, as the articles are published. The approach the authors use to rank pages in the recommendation engine can be modified, of course, but the relevance of this work is that it provides evidence that Twitter, with input about users’ preferences (i.e. what kind of news they are following and/or interested in), can be used to bring news articles from around the web, in real-time, to readers.

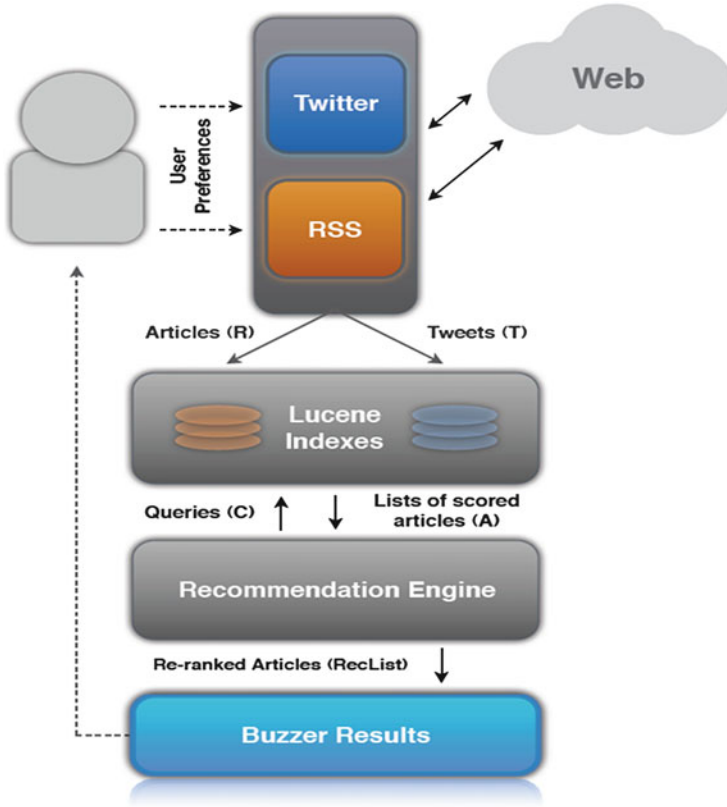


Fig. 13.3 Buzzer back-end architecture

Another feature that researchers can exploit is user-generated, geographically-referenced (georeferenced) updates. Services such as Twitter have been shown to be a useful source of georeferenced data for geographically relevant research. There are two major ways for users to geo-tag their updates. One is manually mention where they are when they publish the update. The other is the use of a GPS-enabled mobile device (e.g. smartphone or laptop) to geo-tag the update. In [7], the authors successfully demonstrate the effectiveness of using *tweets* to track the spread of wild fire.

The paper has covered the application of Twitter as a source of spatial-temporal information for crisis events, and used a major wild fire in France as a case study. The analysis of the temporal dimension showed that user-generated content was accurate and synchronized to actual events. This was not true, however, for when the forest fire started. The media broke the story before users, which makes the use of Twitter as a way to break news questionable. After that stage, Twitter proved effective following the progress of fire over time. It should be noted that only 20 % of content was user-generated original content. The rest was referencing other URLs.

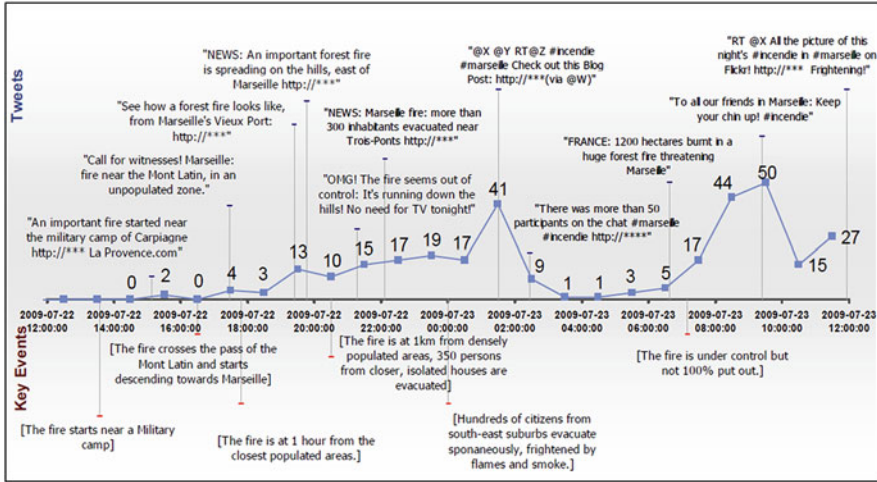


Fig. 13.4 Chronology of the Marseille Fire, number of related tweets per hour and selected tweets' contents

The following two figures from [7] are included below to demonstrate the accuracy the data gathered from microblogging sites.

Video Sharing

Video sharing is another feature for which behavior of communities has been surveyed and analyzed in terms of Self-Expression, Education, Marketing, Geotracking, Medicine, Financial, Counter-Terrorism and Politics.

Self-Expression/Art

In [11] researchers explored the influence of user generated input upon the online video site, YouTube. Classifying each studied video in a range between either professional or user-generated, they worked towards a goal of isolating the methods required to support access to these materials once they are publicly released. They created a research agenda to find unique authoring mechanics to aid in a user having more self-expression in the creation of their videos, and worked towards finding new ways to present this media. Taking a random sampling of 100 videos uploaded in 2008, they analyzed the audio, video, and popularity and found some interesting trends (Figs. 13.4–13.6).

Previously, the researchers had felt user-generated content was the least viewed due to poor editing and a short length. Their research found some user-generated

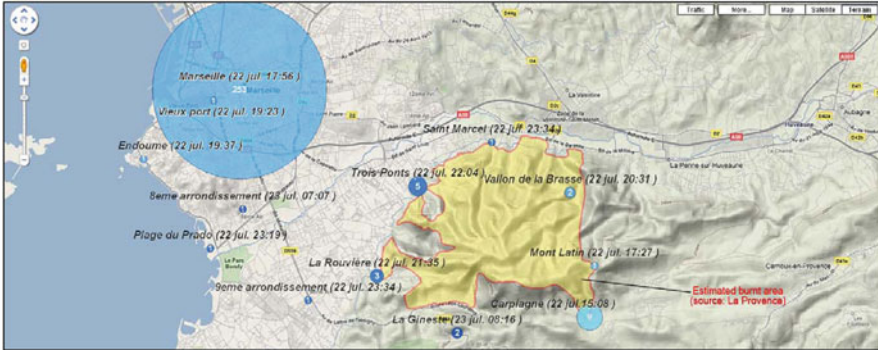


Fig. 13.5 Location, frequency and time of the first citation of place names cited in tweets, and estimated total burnt

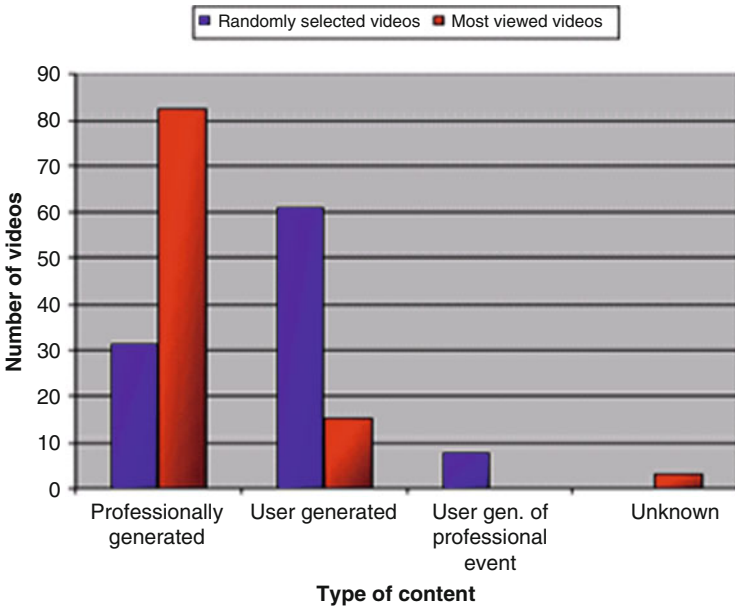
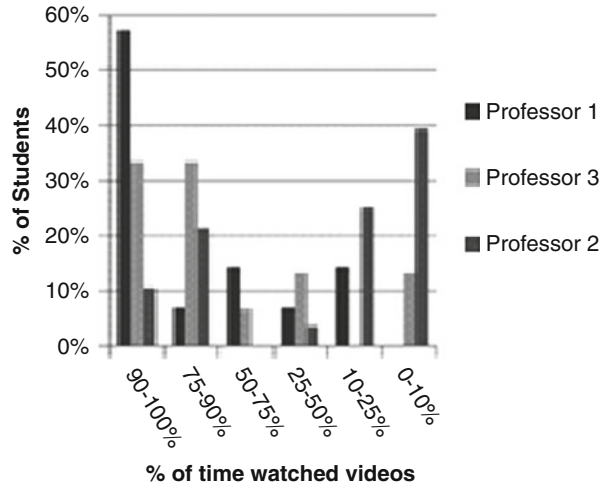


Fig. 13.6 Type of movie content

videos that did extremely well to disprove this. A father’s recording of his daughter gained more than five million views. However, this seemed to be an exception to the rule, and the vast amount of popular content on YouTube comes from a professional realm as opposed to user generated.

From this research paper, we can draw our own conclusions about using YouTube as a research source for studying forms of self-expression. The previously mentioned research paper states that user generated videos on YouTube were not very popular with the online community as a whole, and because of this we can

Fig. 13.7 Videos watched vs. students



expect the vast majority of user generated videos to never be viewed. If these videos are never viewed, the web community will slowly find little reason to post their videos online, knowing that unless they get lucky, no one will watch their videos. The paper states that it is possible people upload these videos simply to share with a smaller sub-network of the web, including their friends and associates, yet they explained several factors that would make it very difficult for researchers to use YouTube to find out about self-expression in online videos. One primary reason is the difficulty they had in finding these videos. Other than by popularity of a broad search, YouTube does not have a simple way to browse these videos for examples of user-generated content, making it much more difficult to examine any sort of self-expression in them. Additionally, since many users upload for only a small sub-network, they can give out exclusive rights to these videos, blocking access to them for not only the general public, but researchers in the topic as well. Due to these factors, it seems safe for the time being to conclude that YouTube is not an ideal medium for the research of self-expression/art in web communities (Fig. 13.7).

Education

In [3], researchers give students 21 YouTube videos on programming in Java, and study how it affected their work. Noticing only about 20–30% of students do their assigned readings, the researchers noticed that the student demographic spend much more time online watching videos instead of reading, and therefore decided to try to present their material in a new manner. Making short videos to keep their attention, the researchers recorded audio and video for the video lectures, and posted them on YouTube for the students as an alternative to doing readings from a textbook. They found that out of three professors, the one that did the least lecturing and most in class hands-on support had his students watch the videos and learn the most from them.

The example of YouTube videos as an education tool is one of many that can prove YouTube can be effectively researched for its educational value. Through the posting of informational videos, and studying the education received by watching them, researchers should be able to draw results on the effect of these videos as opposed to other mediums. This example explicitly showed how videos could be used as an alternative to reading as an educational tool, but researchers could draw many more results from similar studies. Sites like YouTube give public access to this knowledge, and by letting users access it in their web browsing, researchers can study the spread of information. One further example mentioned in the paper was that when they posted these videos on YouTube, it was visited by several members outside their class, especially in the 13–17 year old demographic. Another element researchers would be required to consider is the simple effect of visual cues on video sharing sites to assist in the education of their viewers. Researchers could use this information to study why people may choose to use videos as an educational tool, and what effect this would have upon video based web communities.

Marketing/Ads

The work described in [8] outlines a new method for pop-up advertising in videos. It suggests that instead of YouTube's method of overlay ads during the video, you detect a set of positions before streaming the video and associate ads with those positions. This would give relative ads during that video that could interest the viewer, and maximize the effect of the advertisement.

With the vast amounts of videos one can obtain from the web, and viewing how many times the most popular ones have been watched, it is clear that online videos, especially in web communities such as YouTube, are extremely popular. Web video popularity adverts an opportunity by providing a new medium for advertisers to reach their audience. The development of AdOn and other alternate methods of advertising in these videos can demonstrate to researchers how important of a medium web videos is becoming. By analyzing how advertisements are done in the online video community, researchers can find many of the latest and unique techniques advertisers will choose to implement in order to sell their products. Programs like Ad-On are even able to adapt and change in order to optimize the product being advertised to the user watching the video. This is a relatively new idea, that's effectiveness can be researched so that the method can be fine-tuned and improved. Just like movies, television, and radio, web-videos are becoming a very important medium for advertising, and with that comes the possibility for a great amount of research to improve our methods and designs.

Geotracking

YouTube is used as a resource in the paper [4] in order to improve a user's video browsing based upon their geographical interests. These researchers developed GeoTracker, a tool that uses geospatial representation and a temporal presentation

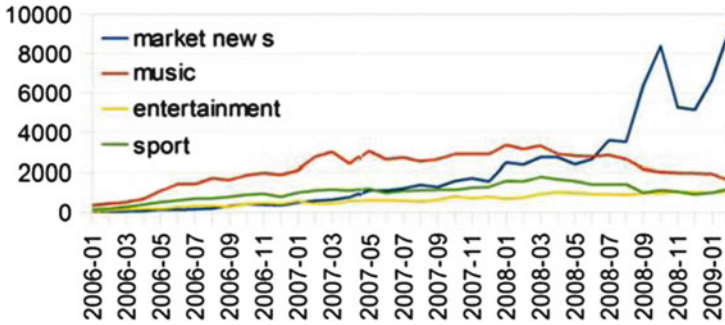


Fig. 13.8 Recent geographical trends

in order to help users spot relevant RSS feed updates as quickly as possible. Once it finds these relevant RSS updates, they are pin pointed to a map and are given key words and tags. GeoTracker then uses the RSS feed tags to search popular video sites, such as YouTube and Google Video, to find related clips and help in any sort of geo-analysis the user intends to do, even streaming it to their mobile device (Fig. 13.8).

By giving visual feedback to important and recent geographical trends, tools like Geotracker can use video streaming sites to enhance their ability to geo-analyze. This method will allow research into geo-analysis to now use video-streaming communities as a prime tool, and as such, it can be researched to find out more about the geotracking topic. Researches can view these RSS feeds, see what videos are being sent to geo-analyst, and find trends and patterns in how they are able to use this data, and how useful this data ends up being to them. For instance, researchers may be able to see that whenever a natural disaster occurs around the world, a geo-analyst can receive a streaming video from it, notice a key element in its appearance, and use that information to warn nearby area of the disasters path. Similarly, by receiving videos related to a specific place at a certain time, geo-analysis can be done anywhere in the world instead of having to go to that spot to examine it. If researchers could prove these means as effective for geo-analysis, it could greatly cut down the costs required to geo-analyze and otherwise optimize the field, increasing its contributions to society.

Medicine

In a medical letter titled “YouTube as a source of information on Immunization: A Content Analysis” [10], a researcher studied YouTube videos that related to either immunization or vaccines. Extracting information from these videos based upon comments and ratings, the researchers found that negative videos, or ones that portrayed immunization negatively, were more highly rated and commented upon. Many of these negative videos even contradicted the reference standard.

Based upon this researcher's work on immunization and how it is presented on video sites like YouTube, it can be safe to conclude it is not a good medium to visit or research for information on medicine, unless you desire to study the perceived opinion of the medicine. The researcher, in the letter, warns Clinicians that patients may be ill-informed about certain immunization from sources like YouTube, and that these video-steaming sites will often be quite critical of immunization. Since content on video web communities is often not regulated, it can be quite difficult for users to discern fact from fiction, and in the field of medicine this could have harsh repercussions. Some videos do seem to report the facts, and the previous research claims these videos were often public service announcements and never contradicted the reference standard, these videos seemed infrequent and unpopular in the web community. These facts will all work together to prove that, for the time being, the video web community is not an effective way for researcher's to study things in the field of medicine due to misinterpreted information.

Financial

In [17] researchers examine content publishing habits of financial news topics and determine the trends in the financial market area. Through various analysis of financially related YouTube videos, the researchers were able to conclude that numerous reputable financial news agencies are now using YouTube as a medium and submitting video content. This helps to prove the reliability and quality of these financial videos, as the sources that upload them are themselves reliable. They also notice how this medium was barely used years ago, and is growing quite strong in the coming years, with no sign of its progress slowing down.

This paper puts a vast amount of study, research and time into proving the reliability of the sources that upload financial information to YouTube. With their results proving that these resources are indeed reliable, stemming from the fact they come from reliable financial companies like Bloomberg, Associated Press and CBS. Research proved that these companies are constantly posting financial videos on sites such as YouTube, and the amount of videos posted seems to only increase as time goes on. Additionally, by studying the trends of stock market and financial news posts on YouTube over a period of time, the researchers were able to prove a sharp incline in the number of videos being uploaded. With more and more of this reliable information being posted on YouTube, it becomes an even better web community to analyze in order to research in the topics of finance. The only issue that could arise in researching finance with YouTube comes from the difficulty of finding valid videos, which can be countered by introducing a search by a series of tags, simple to do over the long term but a fair bit of overhead work for a short quick analysis.

Counter-Terrorism

Hsinchun et al. [9] is a research paper that aims to study terrorist activities on the web, especially in sites like YouTube. Their previous research found that certain terrorist extremist videos are tagged on YouTube, and can be classified in a variety of severities. Their research states that although YouTube aims to filter out inappropriate videos, it can often be days before these videos are tagged and removed, due to the sheer volume of videos uploaded. Through both videos linked you YouTube from other pages, and the videos directly found on YouTube, the researchers were able to find several videos related to Jihadist activities, ranging from explosive devices to Al-Qaeda recruitments.

The vast majority of these videos were linked from blogs that had known affiliations with terrorist activities. Yet this research describes various ways to find and isolate these videos, which can be of extreme importance to anyone researching counter-terrorism and stopping terrorist activities. By researching where these blogs and sites are coming from, and who is uploading these videos onto sites like YouTube, as well as finding more efficient ways of removing them once uploaded, researches can work hard to eliminate the web, especially video web communities, from being popular gathering places for terrorist activities. Research and analysis of certain videos can even determine additional methods and strategies used by terrorist groups to isolate and attack their targets. The researchers themselves suggested YouTube find more efficient ways to monitor uploaded content, and stated that since the Jihadist information is displayed visually, their videos have a much greater impact upon the viewer and need to be blocked from public viewing as quickly as possible. While it is important to note that much of this information may only be coming from Jihadist supporters as opposed to the group itself, they are still working to spread around Jihadist information and do a similar role as some Jihads. By examining these videos to find out more about terrorism, and finding ways to block these videos from being viewed by the public, YouTube and other video streaming sites can be effectively used by researchers to counter terrorism and other extremist Jihad groups in the global sphere.

Political

Mustafaraj et al. [13] explores the impact YouTube and other streaming video websites had upon the result of the 2008 election. Researchers analyzed things like video submission date, view count, ranking, keywords, political messages in the video, and comments to discover the degree of impact the web had upon the election. Through YouTube, BarackObama.com and JohnMcCain.com, they found that political groups were the most common to post videos, as opposed to simply supporters of the political group. Most advertisement videos were negative against their opponent, and that upload timing seemed to always be right before the election, and within weeks these rose in popularity.

A vast number of people visit the internet daily, and because it is quickly becoming a large medium for information, it seems only natural the politics are discussed in it as well. With the popularity of the US election, research has shown us that dozens of videos were uploaded to sites like YouTube, and ended up being viewed by millions in a short time frame. With this large amount of traffic, it would be very easy for researchers to study YouTube and other sites in the video-streaming web community in order to analyze its effect on politics. For example, now users can watch political debates and see politicians speak without having to wait for the news; they can simply log onto their computer and watch a pre-recorded video of it. This is bound to have a profound effect upon the voting population, and can easily be researched to find common trends and patterns. Additionally, it could be worth researching the differences between how politicians act online vs. otherwise, and see if they decide to present themselves in a different manner. However, the previously mentioned researchers noticed that although certain videos ranked higher, trends were tough to notice because most viewers would only watch the already top-ranked videos. In respect to the 2008 US election, it is important to notice that every major politician knew the role of the internet on their voters, and were all online to some degree. One factor research does notice, is certain sites, like YouTube, seemed to have a political bias in the videos that were posted, and like any bias, must be kept in mind while doing research to keep things objective. Research also suggests that due to the vast amount of views political videos received in this time, it is likely they had a strong effect on the outcomes of the election, proving such videos are reliable sources, and can deserve research in proving their effect on the viewer population.

Social Networks

Formation of social network is another important behavior of the formation and evolution of communities over the Web. The analysis for this has been described in terms of activities like Sports, Education, Music and Health.

Sports

In [12], the authors try to show how student athletes at a large university use Facebook differently than fellow students. These differences include the size of their networks, their usage of the website as a whole, and perceptions about their audience. They are often considered on-campus celebrities or the representatives of the university. Because of these differences, the on-campus celebrity athletes have a higher percentage of being watched compared to other students. Hence, they have a higher status and will receive more comments than other students on Facebook.

The researchers did a basic content analysis of the student athletes' pages to get a good idea what goes on their page. As well, they collected data from a large U.S. university and conducted two surveys, one for the athletes and the other for

the students. The response rate for the athletes survey was 27 % (202 students), while the response rate for student one was 21 % (419 students). Both surveys were conducted 8 months apart and judging by previous work, the time difference had no strong effect on the results. The final results were given to the coaches and the student support staff.

The results showed that most student athletes rated their overall Facebook experience as positive. Even though they did see risks and rewards in using it, they felt that they represented themselves well through their profiles. The athletes use the site to either coordinate their activities or keep in touch with fans. This is completely based on group norms and the popularity of the sport though. As predicted before, their network size is large. The researchers showed their results in tables. The table highlighted student athlete experiences on Facebook.

Education

In [5], the researchers are trying to show that Computer Science lecturers in South Africa are using Facebook as an academic tool to enhance their teaching. The way students engage and interact with their instructors has changed tremendously over the years because of technology advances. Initially used only for social interaction, students soon started forming study groups on Facebook for academic purposes. Members in these groups create posts to share knowledge with other students. Hence, peer-to-peer learning and the level of engagement increase.

The researchers distributed an online questionnaire to professors in both the Information Systems and Computer Science departments, and 45 of the questionnaires were completed. The study was limited. Meaning, the questions were unknown to the professor beforehand and no demographic data was collected. There were eight questions in the questionnaire.

The questionnaire asked the professors what they do on Facebook and how they felt about the social networking website. The results showed that many instructors do not have a Facebook account and the majority who do have one only use it for social purposes, not academic. Hence, they do not participate in a group related to their teaching/research areas. One reason for this is that they may want to maintain their professional image and not gain strong relations with other students. Even though most instructors have not applied for online social-networking tool in their teaching, most do believe that they (Facebook for example) can be used as an academic tool for teaching. However, most professors would not use one because they already have dedicated “secure” website to interact with students.

With the successful research study of the impact of using a social network like Facebook as an education tool, we can conclude that social networks can be accurately studied for their educational value. The study itself showed that by giving the students a social network to communicate with, they had a tendency to co-operate and work with each other a little more. Researchers would be able to view this co-operation, and study on how different the interaction is from typical classroom interaction, as well as factoring in the ease of access to it. They could also

research on how, if they give instructors access to the same community, they could evaluate student interaction and even moderate it, preventing cheating, plagiarism, and more. A final example of the topic that could be studied using this research into Facebook as an educational tool is how a student would view their teacher through a social networking site. They could study if the student began to view the professor as less professional, or if it would simply be a great way to ask the professor questions outside of class time.

Music

By analyzing on a dataset from the social-networking site Last.fm [15], the researchers want to determine the music preferences of people and how they change over time. They proceed by extracting data from each of the users and building/clustering a graph of similar users. By doing this, they obtain groups of people with similar music preferences. Once this is done, they label all these clusters and show how these clusters evolve over time. This includes how new clusters emerge, die, merge and split from other clusters.

From the Last.fm website, the researchers were able to obtain the user listening behaviour over the next 4 proceeding months. The site provides for each user and interval a list of the top artists whose songs are played the most. Once this was done, they conducted experiments using data from 16 intervals to determine the parameters e and n for the clustering procedure. Finally, they apply the incremental DENGGRAPH algorithm to obtain the results.

Their final analysis showed that DENGGRAPH detects and obtains groups of users with similar music taste by clustering their profiles. Some of these groups overlap and some of them are separated. If they overlap, they are considered more similar because they have tags in common. If they are separated, the groups have no similarity. In addition, they were able to observe with their incremental procedure approach the growth, decline, creation and removal of new clusters, and the merge and splits of other clusters as well.

With the research proving direct links between music and social network sites, it opens up many possible research opportunities. The one suggested in this paper is important, as by following and studying user's music preferences over a long range of time, they can determine and discern patterns and trends certain people, or even types of people, may follow. Since music is considered a very personal topic by many, there is no more important source to query than the listeners themselves, and this research paper helps to prove that through a social network site like Facebook, users are willing to publicly share their tastes and opinions of music. Further research could delve into proving things like links between user age and music genre, through a query or data mining of a social network web site. Lastly, another topic researchers would be able to study with the use of social networks are how relationships are formed and strengthened based upon similar, or even possibly different, musical tastes.

Health

The researchers for this area of study are trying to explore whether the main features from a social networking site that can be used to help ease interventions for Health Behaviour Change. Using qualitative methods, this helps increase social support to interventions. For this study, five Norwegian citizens (three men and two women) were used and the average age was 60 years old. The method used to collect data was computer-assisted interviews, which each participant had to complete. An interview guide was created per person consisting of open-ended questions about social networks. The questions were mainly focused on the participant's on views and experience navigating through network sites such as Facebook and MySpace. During these interviews, the responses by the people were constantly analyzed.

These results show that staying in touch family and friends in social networks seems more important than staying in touch with distant friends/strangers. In essence, a network with a few close relationships gives more social support than a network of more distant relations. Social networks are an easy way to keep in touch with close friends and their whereabouts, especially if you don't see them often. Because of all these benefits, the features on these sites could be used to ease social support in health behaviour change interventions.

No graphs or visuals were shown to support or show these results. The researchers even state that more quantitative research is needed to show if social-networking features can ease interventions. Social support networks are vital to a person wellbeing and health, and this research paper proves that through a social network site like Facebook, users can build up these social support networks. This would open up many opportunities for research into social support networks, particularly how they are formed inside the constructs of a web community. Additionally, researchers could find very interesting information about the formation, dissolution, strengthening and weakening of social support networks through online communities like Facebook as opposed to real life situations. Another topic that this research proves Facebook can be scientifically studied would be how much emotional and structural support an individual could gain from an online friend in Facebook, communicating through mainly text, and a real life friends, with many more sources available to them such as phone, in person, and more. Researchers could study the differences in willingness to support an online friend in Facebook versus a real life friends, and see if there is a greater physical connection to meeting them in person.

Results

From the survey of works above, we can summarize our findings in the following table (Table 13.1). Note that the table is not exhaustive, nor complete. The undertaking of classifying features of web services will require future continuous work. The following table is meant to provide an idea of how such work can be useful for researchers.

Table 13.1 Summary of results

Feature	Microblogging	Blogosphere	Video	Social networks
Georeferencing	Y	Y	Y	?
Breaking news	Y	?	?	?
Politics/current affairs	Y	Y	Y	Y
Gaming	?	?	?	Y
Movies	?	?	?	Y
Music	?	?	?	Y
Research/education	?	?	Y	Y
Genealogy/personal history	?	?	?	?
Counter-terrorism/attack	?	Y	Y	?
Jobs/workplace	?	?	?	Y
Food/health	?	?	N	Y
Sports/athlets	?	?	?	Y
Travel/tourism	?	Y	?	Y
Marketing/ads	Y	?	Y	Y
Self-expression/art	?	?	N	?

The results summary table is an easy to use visualization for researchers, where there is a ‘Y’ if the web service could be used for such research in the same row, ‘N’ if it could not, and we put ‘?’ for ones where we could not find published research supporting that. This table can be used in one of two ways:

1. Using a more complete table, a researcher interested in studying self-expression can avoid services such as YouTube, a result we think is counter-intuitive but can save valuable research time and resources. On the other hand, a researcher who is interested in how studying the spatiotemporal spread of news can confidently use a Twitter dataset, knowing that the results they will obtain are relevant.
2. Researchers can easily identify areas where research is needed. Where there is ‘?’ can continue to indicate a gap in our knowledge where research could be done.

Conclusions

In this paper we attempted to analyze a corpus of published researched that uses some of the most recent advances in the World Wide Web (WWW), to extract information and provide insights in different aspects of its evolution. With the WWW growing rapidly, we provide researchers with a visualization of what features some of the most recent advances can support research in.

Some questions arise from this paper. Which advancements are better for the features supported by all services of the Web? Whether there is a way to rank those services or not is interesting, as YouTube may be the best source for raw data when it comes to politics and current affairs, even though all types of technology we looked

support the same feature. Also, where there is overlap, as in the case of politics, would the use of a combination of services be better than using only one? All these questions are open ended. However, our analysis takes the first step in analyzing and categorizing the evolving features.

References

1. Admati A, Pfleiderer PC (2001) Disclosing information on the internet: is it noise or is it news? Technical report, Graduate School of Business, Stanford University, . . .
2. Bohorquez JC, Gourley S, Dixon AR, Spagat M, Johnson NF (2009) Common ecology quantifies human insurgency. *Nature* 462(7275):911–914
3. Carlisle MC (2010) Using YouTube to enhance student class preparation in an introductory Java course. In: Proceedings of the 41st ACM technical symposium on computer science education (SIGCSE'10), Milwaukee, 10–13 Mar 2010. ACM, New York
4. Chen YR, Di Fabbriozio G, Gibbon D, Jora S, Renger B, Wei B (2007) Geotracker: geospatial and temporal RSS navigation. In: Proceedings of the 16th international conference on world wide web (WWW'07), Banff, 08–12 May 2007. ACM, New York
5. Colete C, de Villiers C, Roodt S (2009) Facebook as an academic tool for ICT lecturers. In: Proceedings of the 2009 annual conference of the Southern African computer lecturers' association (SACLA'09), Eastern Cape, 29 June–01 July 2009. ACM, New York
6. De Choudhury M, Sundaram H, John A, Seligmann DD (2008) Can blog communication dynamics be correlated with stock market activity? In: Proceedings of the nineteenth ACM conference on hypertext and Hypermedia (HT'08), Pittsburgh. New York, p 55
7. De Longueville B, Smith RS, Luraschi G (2009) OMG, from here, I can see the flames!: a use case of mining location based social. In: Proceedings of international workshop on location based social networks, Seattle
8. Guo J, Mei T, Liu F, Hua X (2009) AdOn: an intelligent overlay video advertising system. In: Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval (SIGIR'09), Boston, 19–23 July 2009. ACM, New York
9. Hsinchun C, Thoms S, Tianjun F (2008) Cyber extremism in Web 2.0: an exploratory study of international Jihadist groups. In: IEEE international conference on intelligence and security informatics (ISI 2008), Taipei, 17–20 June 2008, pp 98–103
10. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K (2007) YouTube as a source of information on immunization: a content analysis. *J Am Med Assoc (JAMA)* 298(21):2482–2484. doi:10.1001/jama.298.21.2482. Retrieved 29 Apr 2010
11. Kruitbosch G, Nack F (2008) Broadcast yourself on YouTube: really? In: Proceeding of the 3rd ACM international workshop on human-centered computing (HCC'08), Vancouver, 31–31 Oct 2008. New York
12. Lampe C, Ellison NB (2010) Student athletes on facebook. In: Proceedings of the 2010 ACM conference on computer supported cooperative work (CSCW'10), Savannah, 06–10 Feb 2010. ACM, New York
13. Mustafaraj E, Metaxas PT, Grevet C (2009) The use of online videos in the 2008 U.S. congressional elections. In: International conference on computational science and engineering (CSE'09), Vancouver, 29–31 Aug 2009, vol 4, pp 320–325
14. Phelan O, McCarthy K, Smyth B (2009) Using twitter to recommend real-time topical news. In: Proceedings of the third ACM conference on recommender systems (RecSys'09), New York, p 385
15. Schlitter N, Falkowski T (2009) Mining the dynamics of music preferences from a social networking site. In: Social network analysis and mining (ASONAM 2009), Athens

16. Shaohui W, Lihua M (2008) The application of blog in modern education. In: 2008 international conference on computer science and software engineering, Wuhan, pp 1083–1085
17. Sykora MD, Panek M (2009) Financial news content publishing on youtube.com. In: 3rd international workshop on soft computing applications (SOFA'09), Szeged, 29–30 July 2009, pp 99–104
18. Tirapat T, Espiritu C, Stroulia E (2006) Taking the community's pulse, one blog at a time. In: Proceedings of the 6th international conference on web engineering, Palo Alto. ACM, p 176
19. Tumarkin R, Whitelaw RF (2001) News or noise? Internet postings and stock prices. *Financ Anal J* 57:41–51
20. Wuthrich B, Permuntilleke D, Leung S, Cho V, Zhang J, Lam W (1998) Daily prediction of major stock indices from textual WWW data. In: Proceedings of the 4th international conference on knowledge discovery and data mining, New York, pp 364–368