Chapter 2
Community-Contributed Media Collections: Knowledge at Our Fingertips

Tania Cerquitelli, Alessandro Fiori, and Alberto Grand

Abstract The widespread popularity of the Web has supported collaborative efforts to build large collections of community-contributed media. For example, social video-sharing communities like YouTube are incorporating ever-increasing amounts of user-contributed media, or photo-sharing communities like Flickr are managing a huge photographic database at a large scale. The variegated abundance of multimodal, user-generated material opens new and exciting research perspectives and contextually introduces novel challenges. This chapter reviews different collections of user-contributed media, such as YouTube, Flickr, and Wikipedia, by presenting the main features of their online social networking sites. Different research efforts related to community-contributed media collections are presented and discussed. The works described in this chapter aim to (a) improve the automatic understanding of this multimedia data and (b) enhance the document classification task and the user searching activity on media collections.

2.1 Introduction

The past few years have witnessed the steady growth of Web-based communities such as social networking sites, blogs, and media-sharing communities. This recent trend in the WWW technology, commonly known as Web 2.0, finds its natural outcome in the emergence of well-known online media-sharing communities. For example, licensed broadcasters and production companies were historically the only publishers of video content in online Video-on-Demand (VoD) systems. However, the advent of video-sharing communities, partially supported by the popularity of affordable hand-held video recording devices (e.g., digital cameras, camera phones), has reshaped roles. Nowadays, hundreds of millions of Internet users are self-publishing consumers. This has resulted in user-generated content (UGC) becoming a popular and everyday part of the Internet culture, thus creating

T. Cerquitelli (✉)
Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy
e-mail: tania.cerquitelli@polito.it

E. Pardede (ed.), *Community-Built Databases*,

DOI 10.1007/978-3-642-19047-6_2,

new viewing patterns and novel forms of social interaction. Accordingly, an increasing research effort has been devoted to analyzing and modeling user behavior on social networking sites.

The abundance of contents generated by media-sharing communities could potentially enable a comprehensive and deeper multimedia coverage of events. Unfortunately, this potential is hindered by issues of relevance, findability, and redundancy. Automated systems are largely incapable of understanding the semantic content of multimedia resources (e.g., photos, videos, documents). Queries on multimedia data are thus extensively dependent on metadata and information provided by the users who upload the media content, for example, in the form of tags. However, this information is often missing, ambiguous, inaccurate, or erroneous, which makes the task of querying and mining multimedia collections a nontrivial one. Hence, user-contributed media collections present new opportunities and novel challenges to mine large amounts of multimedia data and efficiently extract knowledge useful to improve the access, the querying, and the exploration of multimedia resources.

The aim of the chapter is to present how information retrieval and data mining approaches can be used to extract and manage the content generated by media communities. The chapter is organized as follows. Section 2.2 reviews different collections of user-contributed media, such as YouTube, Flickr, and Wikipedia, by presenting their main features. Section 2.3 discusses several aspects of the reviewed media collections and their associated user communities. In particular, the structure of the underlying social networks is analyzed. Different research efforts aimed at studying media content distribution and user behavior are then presented. Finally, the semantics of content found in these collections are addressed. Section 2.4 describes three taxonomies, one for each media collection, proposed to sum up and classify the main research issues being addressed. An overview of results achieved in the media annotation domain is presented in Sect. 2.5, whereas Sect. 2.6 describes diverse research efforts targeted at developing novel and efficient data mining techniques to (a) extract relevant semantics from image tags, (b) train concept-based classifiers and automatically organize a set of video clips relative to a given event, and (c) efficiently categorize a huge amount of documents exploiting Wikipedia knowledge. Finally, Sect. 2.7 draws conclusions and suggests research directions offered by the community-built media collections.

2.2 Social Media-Sharing Communities

The past few years have witnessed the rapid proliferation of social networking sites, wikis, blogs, and media-sharing communities. The advent of media-sharing communities, partially spurred by the popularity of affordable hand-held image and video recording devices (e.g., digital cameras, camera phones), has favored the growth of social networks. Nowadays, hundreds of millions of Internet users are self-publishing consumers. This has resulted in user-generated content (UGC)

becoming a popular and everyday part of the Internet culture, thus establishing new viewing patterns and novel forms of social behavior.

This section reviews different collections of user-contributed media like YouTube [1] (see Sect. 2.2.1), Flickr [2] (see Sect. 2.2.2), and Wikipedia [3] (see Sect. 2.2.3) by discussing their main features. We discuss these media collections as representatives of the rapid growth witnessed in the Internet-based multimedia domain. In addition, the interest of Web users in these Web services has significantly grown, attracting an increasing attention from the scientific community.

2.2.1 YouTube

YouTube [1] is one of the largest and most successful online services allowing users to upload, share, and watch video material freely and easily. Since its establishment in early 2005, YouTube has become one of the fastest-growing Web sites and ranks fourth in Alexa's [4] top popular site list, with 24 h of new video content uploaded every minute, as of March 2010.

YouTube's primary features include the ability to upload and play back video clips. Any user with a Web browser can view YouTube videos, but users are required to create an account to publish their own content and interact with each other.

Registered users are assigned a profile page, named a "channel", which serves as an index to the user's uploaded material. Users may easily customize the looks of their channel page by selecting an available graphical theme or by creating a new one. They may optionally disclose their personal details, subscribe to other users' videos, or "make friends" with them. Comments can be posted by registered users on another user's profile page or on a specific video's page. Viewers can additionally rate videos or join groups that focus on particular interests. YouTube thus shows its strong community nature: it is a social networking site, with the added feature of hosting video content [5].

Videos can be uploaded in most existing container formats and are automatically converted into the Adobe Flash Video format (FLV). The Adobe Flash Player browser plug-in, required to play the FLV format, is one of the most common pieces of software installed on personal computers.

Currently, video clips uploaded by standard users are limited to 10 min in length and a file size of 2 GB. In YouTube's early days, it was possible for users to upload longer videos, but a time restriction was introduced in March 2006, when it was found that the majority of videos exceeding this length were unauthorized uploads of copyrighted materials. Since November 2008, YouTube has been making an ongoing effort to improve picture quality. Videos were thus made available in HD format and currently use the H.264/MPEG-4 AVC codec, with stereo AAC audio.

YouTube assigns each video a distinct 11-digit identifier, composed of digits and (uppercase and lowercase) letters. Associated with the videos are some metadata, including the name of the uploader, the date of upload, and the number of views,

ratings, and comments. A title, a category, a set of keywords (tags), and a textual description are also provided by the user who added the video. A list of related videos is generated by YouTube, which consists of links to other videos that have a similar title, description, or tags, and are thus supposedly similar in content. Uploaders may also specify a set of related videos of their choice.

YouTube provides a dedicated feature to share videos on other online communities (e.g., Facebook [6], Twitter [7], MySpace [8]) and makes it easy to embed them in an external Web page by automatically generating the required HTML code.

2.2.2 *Flickr*

Flickr [2] is a popular photo-sharing Web site that allows users to store, search, and share their photos with family, friends, and the online community. Flickr, hosted on 62 databases across 124 servers, manages about 800,000 user accounts per pair of servers and contained more than 5 billion images as of August 2009.

Launched in February 2004 by a Vancouver-based company, Flickr was originally a multiuser chat room, called "FlickrLive", with the capability of exchanging photos in real time. In March 2005, Flickr was acquired by Yahoo! and became a photo-sharing Web site allowing users to upload, share, and view photos freely. In March 2009, the ability to upload and view HD videos was also added to Flickr's features.

Users can join the Flickr network by means of two account types. Free accounts allow users to upload 100 MB of images in a month and at most 2 videos. However, free account users can manage at most 200 photos in their photostream. A photo can be added to at most ten groups, and statistics about photos are not accessible. A free account is automatically deleted when it has been inactive for 90 consecutive days. "Pro accounts", characterized by unlimited storage, allow users to upload any number of images and videos every month and to access account statistics. Furthermore, each image can be added to up to 60 groups. However, pro account users are charged a yearly fee of about 25 dollars.

Flickr, as a photo-sharing Web site, provides both private and public image storage. Private photos are visible by default only to the photo owner. By contrast, public photos are viewable by all Flickr users. In addition, each user maintains a list of "favorite photos" on the Flickr Web site, which is publicly visible to the user's contacts when they log into Flickr. However, they can also be marked as viewable by friends and/or family. Furthermore, for each uploaded image, belonging to either the private or the public category, the user (photo owner) can define a "contact list" to control image access for a specific set of users.

Flickr users, characterized by common interests, can gather to form self-organized communities, referred to as groups. The main purpose of groups is to facilitate the sharing of user photos in the group pool, i.e., a collection of photos shared by any member with the group. There are three types of groups: (a) public, where anyone can see the group photo pool, and anyone can join it, (b) public, where anyone can

see the group but only "invited" users can join it, and (c) private, where the group is hidden from the community and an invitation is required both to see and to join it.

Browsing techniques and diverse sophisticated search functionalities are also available on Flickr. For example, it is possible to filter the search results according to geographical locations, time intervals and media type.

Since community-built photo collections are typically very conspicuous in size, the efficient browsing of these collections is still an open research issue. Different approaches aimed at improving the browsing of large image collections have been proposed in the literature [9, 10].

2.2.3 Wikipedia

Wikipedia [3] is a free, collaborative, multilingual encyclopedia. Any Web user can collaborate to create or edit a Wikipedia page. Since 2001, this effort of the Web community has produced over 15 million articles. The corpus is composed of several collections of articles in different languages. For example, the English version contains over 3.2 million articles at March 2010, whereas the German version contains more than 1 million.

In recent years, Wikipedia has been considered the most powerful, accurate, and complete available free encyclopedia. Due to its nature, Wikipedia is one of the most dynamic and fastest-growing resources over the Web. For instance, articles about events are often added within few days from their occurrence. Moreover, since everyone can edit the information, errors are usually removed after few revisions by other Web users. Unlike other encyclopedias, Wikipedia is freely available and covers a huge number of topics. For example, Encyclopedia Britannica, one of the oldest encyclopedias which is considered a reference book for the English language, with articles typically contributed by experts, contains only around 120,000 articles in the last release. Indeed in [11] Wikipedia was found to be very similar, in terms of accuracy, to Encyclopedia Britannica, but the coverage of topics is greater.

The main advantage of Wikipedia is its community, which improves and controls the content of the encyclopedia. Except for a few pages, every article may be edited anonymously or with a user account, while only registered users may create a new article. Once a new article has been added, the page is owned by the community, which can modify the content. Even though every Web user can contribute to the growth of the encyclopedia, the Wikipedia community has established "a bureaucracy of sorts", including a clear power structure that gives volunteer administrators the authority to exercise editorial control. The administrators are a group of privileged users who have the ability to delete pages, lock articles from being changed in case of vandalism or editorial disputes, and block users from editing.

Since many Web users may not be proficient at creating and managing Web content, the editing model of Wikipedia is based on "wiki". This technology allows

even inexpert users to create and/or edit complex Web pages with structured information, such as internal and external links, tables, images, and videos. Furthermore, the wiki model makes changes to an article immediately available, even if they contain errors. The German edition of Wikipedia is an exception to this rule. It has been testing a system of maintaining stable versions of articles to permit readers access only to versions of articles that have passed certain reviews.

Many features have been implemented to assist contributors. For example, the "History" page attached to each article records every single past revision of the article. This feature makes it easy to compare old and new versions, undo changes that an editor considers undesirable, or restore lost content. The "Discussion" pages associated with each article are used to coordinate work among multiple editors. Pieces of software such as Internet bots (e.g., Vandal Fighter) are in wide use to remove vandalism as soon as it is committed, to correct common misspellings and stylistic issues, or to ensure that new articles comply with a standard format [12].

Since Wikipedia grows very dynamically and is human contributed and mainly composed of free text, the structure of the media collection is very complex. Dumps of articles are generated automatically every week and can be downloaded to apply offline analyses of the content.

The basic entry in Wikipedia is an article (or page), which defines and describes an entity or an event and consists of a hypertext document with hyperlinks to other pages, within or outside Wikipedia. The role of the hyperlinks is to guide the reader to pages that provide additional information about the entities or events mentioned in an article. Each Wikipedia article is uniquely referenced by an identifier, which consists of one or more words separated by spaces or underscores, and occasionally a parenthetical explanation. Some articles can contain an Infobox, a table which sums up key information about the article. The community provides a collection of templates for different categories (e.g., City, Company) to avoid ambiguous annotations and present information with a uniform layout.

The hyperlinks within Wikipedia are created using the articles' unique identifiers. Since every article can be edited by any user, one critical issue is the consistency with respect to these identifiers. "Redirect" pages, which contain only a redirect link, exist for each alternative name of a concept and point readers to the one preferred by Wikipedia.

Another issue is the different meanings that words can assume according to the context. Disambiguation pages are specifically created for these ambiguous entities and are identified by the parenthetical explanation "(disambiguation)". These pages consist of links to articles defining the different meanings of the entity.

2.3 Community Nature and Media Collection Features

Each community-contributed media collection has different features due to both the structure of the Web service and the managed media type. Many research efforts have been devoted to studying the structure of social networks [13], proposing

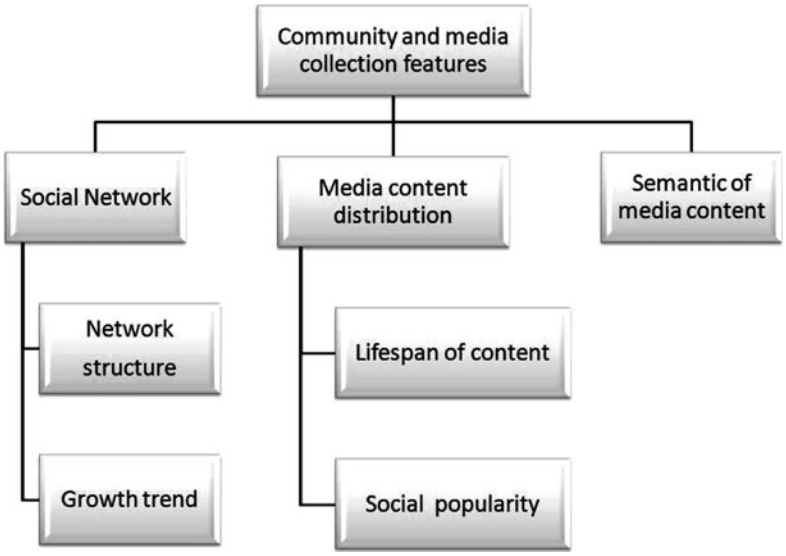

Fig. 2.1 Taxonomy of user-contributed media collection features

folksonomy models [14], and analyzing the properties of information spreading through communities [15].

In this section, we discuss diverse works analyzing the community nature and the media collection features of Flickr, YouTube, and Wikipedia. We propose a taxonomy, shown in Fig. 2.1, to categorize the discussed works in three main topics: (a) social networks (see Sect. 2.3.1), (b) media content distribution (see Sect. 2.3.2), and (c) semantics of media content (see Sect. 2.3.3). The social network topic includes two main issues, network structure [5, 13, 16] and growth trend [17], while the media content distribution topic includes lifespan of content [18] and social popularity [19, 20].

2.3.1 Social Networks

Social networks have been studied from two different points of views: network structure [5, 13, 16] and growth trend [17], as shown in the taxonomy reported in Fig. 2.1. From the structure point of view, social networks are modeled by graphs where nodes represent the users and links the friendships among users. Different graph mining algorithms have been exploited to extract relevant and useful information on social community structure [13].

Online social networking sites also represent a unique opportunity to study the dynamics of social networks and the ways they grow. The growth trend (see taxonomy in Fig. 2.1) in the Flickr social network has been investigated in [17].

In [5], the social network structure of YouTube has been analyzed as a first step toward understanding the kind of media and social space that it represents. The analysis draws upon a crawl of YouTube's user profiles, characterized by the aggregate of all tags used by the authors to annotate the uploaded videos. Clusters of authors and associated keywords were identified through vector-space projection and hierarchical cluster analysis. Nine author clusters were thus found, corresponding to distinct genres of popular Internet video. Similar clustering was identified by analyzing the authors' networks of friends, represented as graphs. These results show that socially coherent activity (friendship among users) is strongly characterized by semantic coherence (similar descriptive tags).

A different research issue has been addressed in [16]. The work is founded on the observation that Web surfers are not required to register or upload videos in order to view the existing materials; a large proportion of YouTube's audience is in fact expected to fall into this category of users. As a consequence, the network among (registered) users does not necessarily reflect that among the videos. The social networking among videos has thus been studied, and relationships between pairs of related videos have been modeled by means of a directed graph. Measurements on the graph topology revealed definite small-world characteristics. This phenomenon, also known as six degrees of separation, refers to the principle that items (e.g., people, files, URL links) within a given environment are linked to all others by short chains of "acquaintances". Similar results were achieved on other real-world user-generated graphs. However, compared with, for example, the graph formed by URL links in the World Wide Web, the YouTube network of videos exhibits a much shorter characteristic path length, thus implying a much more closely related group.

Authors in [17] analyze the growth trend in the Flickr social network in order to understand the link formation process. Flickr can be modeled as a directed network, where users represent nodes and directed edges model links between a pair of users. The presence of a link from a user to another does not imply the presence of the reverse link. As shown in [17] the creation of the first link affects the second, since users tend to rapidly respond to the incoming link by creating a link in the reverse direction. Since users explore the network by visiting their neighbors, users tend to connect to nearby users in the network. Furthermore, the number of links created and received by users is directly proportional to their current number of links.

2.3.2 Media Content Distribution

The enormous, steady growth of the well-known online media-sharing communities, such as YouTube, Flickr, Facebook, MySpace, and the emergence of numerous similar Web sites, confirm the mass market interest. While similar on the surface to standard commercial media distribution systems, UGC collections follow in fact

much less predictable evolution trends. Furthermore, the popularity of UGC presents rather diverse and complex dynamics, thus making traditional content popularity predictions unsuitable.

Works devoted to studying media content distribution address either (a) the lifespan of the media content or (b) social popularity as shown in Fig. 2.1. The lifespan of the media content is highly dependent on the user behavior and interests. Several studies have been investigating the main reasons for the popularity and diffusion of videos and photos.

For example, in [18] the analysis of the popularity evolution of user-produced videos in YouTube and other similar UGC Web sites is presented. The key observation is that understanding the popularity characteristics can prove pivotal in discovering weaknesses and bottlenecks in the system and suggest policies to improve it. The study was conducted on several datasets containing video meta information crawled from YouTube and Daum [21], a popular search engine and UGC service in Korea. Their analysis reveals that the popularity distribution of videos exhibits power-law behavior with a steep truncated tail (exponential cut-off), suggesting that requests for videos are highly skewed toward popular files. Rather than this skewness being a natural phenomenon due to the low level of interest in many UGC videos, filtering effects in search engines, which typically favor a small number of popular items, seem most likely responsible for the significant imbalance in the video popularity distribution. Popular videos thus tend to gain more and more views, while niche videos reach a much smaller audience than expected. Proper leverage of the latter could increase the total number of views by as much as 45% and reveal the latent demand created by the search engine bottleneck.

Other studies consider also the influence of social contacts in the popularity of media content (see taxonomy in Fig. 2.1). For example, the characterization of the Flickr media collection has been studied in [20]. An analysis has been performed along three dimensions: (a) the temporal dimension, which allows tracking the user interest in a photo over time, (b) the social dimension, aimed at discovering the social incentives of users in viewing a photo, and (c) the spatial dimension, which analyzes the geographic distribution of user interest in a photo. Experimental results reported in [20] show that users discover new photos within 3 h of their upload. Furthermore, for the most popular photos (i.e., photo with high view frequency) almost 45% of the new photo views are generated within the first two days, while for infrequent images (i.e., photo with low view frequency), this ratio increases to 82%. Moreover, the following two factors affect photo popularity: (a) the social network behavior of users and (b) photo polling. In fact, people with a large social network within Flickr have their photos viewed many times, while people with a poor social network have their images accessed only few times. Finally, the geographic distribution of user interest is also dependent on the photo popularity. In fact, the geographic interest in a photo is worldwide when the photo has many views, while for infrequently viewed photos the geographic distribution is around a given geographic location.

In [19] the analysis performed is more focused on the influence of social contacts on the bookmarking of favorite photos to estimate the potential spreading capability of 1,000 favorite photos, over 15,000 unique fans and 35,000 favorite markings. As shown in [19] the information dissemination through social links can be modeled as a slight variation of the model exploited to study the spread of infectious diseases throughout human populations [22]. A social cascade begins when the first user includes the photo in his/her list of favorites. Then, the cascade continues along social links. The following aspects of Flickr in the social behavior of users have been observed: (a) users maintain their favorite photos indefinitely (e.g., until photos are removed from the list) and (b) the higher the number of neighbor users, the higher the dissemination rate. Furthermore, the time required by the dissemination is in inverse relation to the number of neighbors. Hence, social links are an effective mechanism for disseminating information in online social networks.

2.3.3 Semantics of Media Content

In the last few years, an increasing effort has been devoted by the researcher community to drawing a general picture of the semantic distribution of UGC. Since this aspect is fundamental to understanding the interests and the predominant uses, we included it in our proposed taxonomy, shown in Fig. 2.1.

In [16], an in-depth measurement study of the statistics of YouTube videos is provided. Based on a dataset including metadata about more than 3 million videos, content distribution among categories has first been analyzed, showing that music videos are prevalent (22.9%), followed by entertainment videos (17.8%) and comedy (12.1%). Accordingly, the video length distribution indicates that the majority of videos does not exceed 3–4 min in length.

Similar works have analyzed the distribution of the encyclopedic content available on Wikipedia. Wikipedia articles are classified into categories and subcategories exploiting a predefined taxonomy, enriched by article authors. When an article is uploaded, authors can freely choose one or more categories for their new encyclopedia entry, or create a new one. Due to this uncontrolled mechanism, categories associated with a given article may not always optimally suit its content. The distribution and the growth of Wikipedia topics were studied in [23]. The approach employed for this analysis is based on the evaluation of the semantic relatedness of each article with taxonomy categories. The topic coverage of Wikipedia article is unfair. “Culture and arts”, “People”, and “Geography” cover more than 50% of all the articles, while society and social sciences covers 12%. However, some of the other topics are rapidly growing. The results of the analysis, reported by the Wikipedia Foundation, show that the geographic distribution of articles is highly uneven. Most articles are written about North America, Europe, and East Asia, while only few about Africa and large parts of the developing world.

2.4 Taxonomies on Research Issues

A lot of investigation has been carried out on community-contributed media collections to improve the user searching activity and document classification of media collections. To provide a clear overview of the research efforts, we propose three taxonomies (see Figs. 2.2–2.4), one for each media collection, to sum up the main issues addressed in the research.

Figure 2.2 summarizes the works addressing video-sharing communities. A first class of approaches is aimed at studying whether and how meaningful, coherent annotations can be derived from the collaborative tagging effort of community users. The second group of research activities is instead focused on the exploitation of these video collections from a data mining perspective. In the works reviewed, video clips and their associated tags are exploited to (a) train concept-based classifiers and improve query representations in a media retrieval framework or (b) generate relevant metadata about captured events and enhance the user viewing activity.

Figure 2.3 depicts the taxonomy proposed to classify the works focused on photo-sharing communities. Research efforts can be grouped according to the addressed topic, like tag recommendation to effectively support photo annotation, and automatic extraction of semantics. Among the tag annotation works, two main approaches have been proposed. The first one resorts to basic techniques that consider tag frequencies in the past to suggest useful and relevant tags, while the second one exploits collective knowledge, residing in the Flickr community, to support photo annotations. Furthermore, numerous works have been devoted to exploiting data mining techniques to (a) discover location and event semantics from

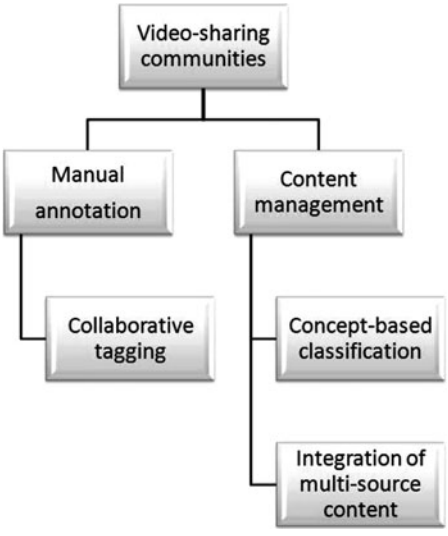

Fig. 2.2 YouTube: taxonomy of discussed approaches

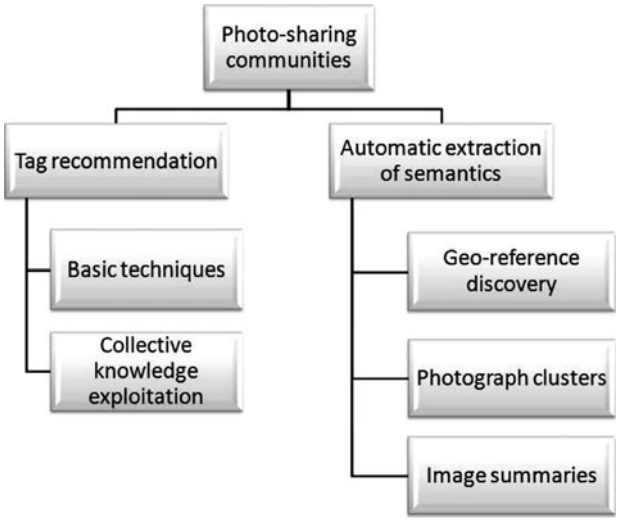

Fig. 2.3 Flickr: taxonomy of discussed approaches

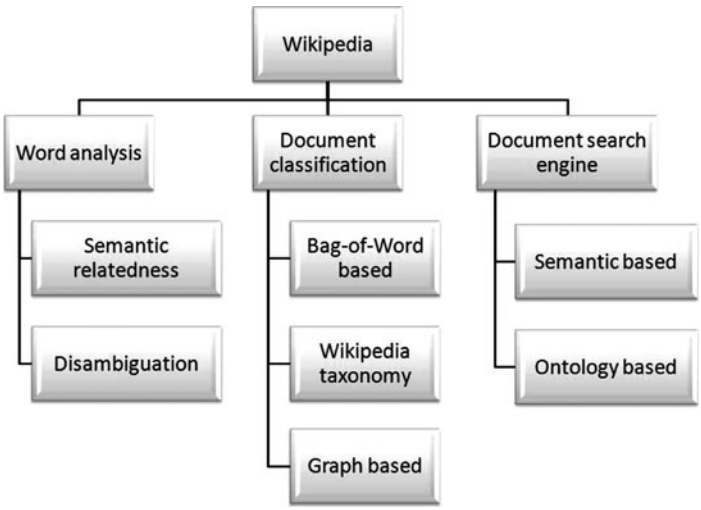

Fig. 2.4 Wikipedia: taxonomy of discussed approaches

photo tags, (b) identify clusters of similar photographs, or (c) summarize a large set of images.

Figure 2.4 shows a possible taxonomy to categorize the discussed works relating to the analysis and the employment of Wikipedia data. Three main topics can be identified according to the use of Wikipedia information and the purpose of the work: (a) word analysis, (b) document classification, and (c) document search engine. The word analysis addresses the evaluation of the semantic relatedness

among Wikipedia topics and the disambiguation of terms in documents, according to the Wikipedia categories. Wikipedia information can also be used to build more efficient text representations in terms of classification performance. Different approaches have been proposed, based on (a) the bag-of-word representation, (b) the analysis of Wikipedia taxonomies, and (c) the analysis of the Wikipedia graph structure. Moreover, some works have been devoted to developing search engines which retrieve documents according to (a) the semantic analysis of terms in the documents, based on Wikipedia taxonomies, or (b) the employment of ontologies extracted by Wikipedia infoboxes.

The proposed taxonomies are useful for categorizing the works discussed in Sects. 2.5 and 2.6.

2.5 Media Annotation

An interesting challenge when dealing with knowledge collected on a large scale is that of making it searchable and thus usable. Despite the growing level of interest in multimedia Web search, most major Web search engines still offer limited search functionality and exploit keywords as the only means of media retrieval [24]. In the context of media (e.g., video, images, documents), this requires content to be annotated, which can be done manually or automatically.

In the first case, the process is an extremely time-consuming, and hence costly, one. As pointed out in [25], a potential drawback of manual annotation is its subjective nature as an indicator of content. The same media may produce rather disparate reactions from different users or groups of users, who may also have varying motivations for annotating it. This would result in the media being annotated very differently. However, automatic annotation of media content may require content analysis algorithms to extract descriptions from media data.

Community-built media collections are typically designed in such a way as to enable user queries on the content, and thus provide varying levels of media annotation. Tagging is the most popular form of annotation and has proved successful over the past years, as shown in [26–31]. In addition, it is available at virtually no cost, because the annotation task is spread across the entire community.

2.5.1 Photo Annotation

Photos uploaded on Flickr can be enriched by different kinds of metadata, in the form of tags, notes, number of views, comments, number of people who mark the photo as their favorite, and even geographical location data.

The analysis of “how users tag photos” and “what kind of tags they provide” is presented in [31]. By analyzing 52 million photos collected between February 2004 and June 2007, authors show that the tag frequency distribution can be modeled by

a power law [32], and the probability of a tag having tag frequency x is proportional to $x^{-1.15}$. The head of the power law fit contains tags that would be too generic to be useful as a suggestion, while the tail contains the infrequent tags that often correspond either to wrong words or to highly specific tags. Furthermore, the distribution of the number of tags per photo also follows a power law distribution. The probability of having x tags per photo is proportional to $x^{-0.33}$. The head of the power law contains photos annotated with more than 50 tags and the tail contains more than 15 million photos with only a single tag, while almost 17 million photos have only two to three tags. The majority of photos is thus annotated with only a few tags, which describe where the photo is taken, who or what appears in the photo, and when the photo was taken.

Different works addressed the research issues on photo annotation. To support automatic photo annotations, diverse tag recommendation techniques have been studied. As shown in the taxonomy, depicted in Fig. 2.3, the proposed approaches can be classified into basic techniques and collective knowledge supporting.

Flickr offers a service to suggest tags when a user wants to tag a picture. Suggested tags, sorted lexicographically, include recently used tags and those most frequently employed by the user in the past. However, this service is rather limited. One step further toward personalized tag suggestion for Flickr was presented in [27]. Three algorithms have been proposed to suggest a ranked list of tags to the user. Proposed algorithms receive as input parameters the identity of the user, an initial set of tags (if available), and the corresponding tagging history of all users. The recommendation is based on the tags that the user or other people have exploited in the past. Furthermore, the suggested tags are dynamically updated with every additional tag entered by the user. The first two methods consider only those tags exploited by the user in the past and suggest a ranked list of tags, sorted by considering both tag frequency and past inserted tags. The last method considers both the set of tags exploited by other people in the past and the tags similar to those entered by the user for the same picture in the past. In particular, a set of promising groups is first identified by analyzing both the user and the group profiles. Then, for each of these groups, a ranked list of suggested tags is generated according to tag frequency and past inserted tags.

To validate the methods proposed in [27], different pictures have been downloaded and divided into two groups: (a) 200 pictures with four to eight given tags and (b) 200 pictures with more than ten given tags. The method which considers tags exploited also by other people is more effective in suggesting relevant tags. Furthermore, the results obtained for the second set of pictures are better than those obtained for the first set because users who add more tags to an individual picture usually have a better tagging history. In fact, the methods proposed in [27] yielded better accuracy on pictures with a large number of given tags.

Different and more effective approaches have been proposed to effectively support photo annotations [27, 31].

Authors in [31] proposed to exploit the collective knowledge that resides in the Flickr community to support tag recommendation. Given a photo, the proposed recommendation system selects a list of relevant tags. The proposed system operates

in two steps. From a photo with user-defined tags, an ordered list of candidate tags is derived for each of the user-defined tag, based on co-occurrence. Then, the lists of candidate tags are aggregated and classified to generate a ranked list of recommended tags. The co-occurrence between two tags is computed as the number of photos where both tags are used in the annotation. The obtained value is normalized with respect to the overall frequency of the two tags individually. Two measures have been proposed to normalize the tag co-occurrence: symmetric and asymmetric measures. The first, according to the Jaccard coefficient [33], is defined as the size of the intersection (co-occurrence of the two tags) divided by the size of the union of the two tags (sum of the frequencies of two tags). This measure can be exploited to identify equivalent tags (i.e., tags with similar meaning). By contrast, the asymmetric measure evaluates the probability of finding tag t_j in annotations, under the condition that these annotations also contain tag t_i. For each user-defined tag, an ordered list of candidate tags is derived from the collective knowledge (i.e., user-generated content created by Flickr users). The larger the collective knowledge, the more relevant and useful the list of candidate tags. Given diverse lists of candidate tags, they are merged in a single ranked list by means of two strategies: voting and summing. The first one computes a score for each candidate tag, and the ranked list of recommended tags is obtained by sorting the tags according to the number of votes. On the other hand, the summing strategy computes for each candidate tag the sum of all co-occurrence values between the considered tag and the user-defined tags.

To evaluate the recommendation system proposed in [31], 331 photos with at least one user-defined tag have been analyzed. 131 photos were used as a training set, while the remaining 200 photos were the actual test set. Experimental results show the effectiveness of the proposed recommendation system in selecting relevant tags. For almost 70% of the photos, the system suggests a good recommendation at the first position of the ranked list, and for 94% a good recommendation is provided among the top 5 ranked tags.

2.5.2 Video Annotation

To extend the accessibility of video materials and enhance video querying, manual or automatic annotation is needed. Since manual annotations often reflect personal perspective, videos may be tagged very differently by different users. However, the study reported in [25] suggested that user interaction with multimedia resources within social networks could help generate more consistent and less ambiguous tagging semantics for video content. In particular, it has been observed that when multiple users are allowed to label content over a long period of time, stable tags tend to emerge [34]. This can be thought of as a form of user consensus built by letting users interactively correct tags, similarly to wiki pages, thus providing more reliable metadata. This form of "collaborative tagging" (see taxonomy in Fig. 2.2) has been investigated by inferring semantics for the content from user behavior, both explicitly (i.e., through direct user input) and implicitly (i.e., by monitoring

user activity). To this aim, Facebook's public APIs were exploited to build a social network application, Tag!t. The application makes it possible to share and interact with video content in a broader way than allowed by Facebook's features. Besides sharing videos and aggregating them in collections, users can tag specific timestamps of their friends' videoclips and link additional materials (e.g., videoclips, images, Web pages) to them. The application also keeps track of user interaction with the media itself (e.g., play/pause/seek events). Experiments with the Tag!t application showed that users within a selected group tend to tag in a similar manner. In addition, the semantics suggested by a user were found to be scantly biased by other users' tagging of the same content, thus indicating that collaborative tagging leads to coherent semantics.

2.5.3 Document Annotation

Wikipedia articles can also be used to analyze words in order to improve keyword extraction from documents and disambiguation algorithms, as shown in the taxonomy, depicted in Fig. 2.4. For example, Semantic MediaWiki [35] is an extension of the MediaWiki software to annotate the wiki contents in the articles. The aim of this tool is to improve consistency in Wikipedia articles by reusing the information stored in the encyclopedia.

Some approaches have attempted to detect the semantic relatedness between terms in documents to identify possible document topics. In [36] the authors introduced "Explicit Semantic Analysis" (ESA) which computes the semantic relatedness between fragments of natural language text using a concept space. The method employs machine learning techniques to build a semantic interpreter which maps fragments of natural language to a weighted sequence, named "interpretation vector", and built of Wikipedia concepts ordered by their relevance. The relatedness between different interpretation vectors is evaluated by means of cosine similarity.

In [37], the Wikipedia Link-Based Measure is described. The approach identifies a set of candidate articles which represent the analyzed concepts and measures the relatedness between these articles using a similarity measure which can be a tf-idf-based measure, the Normalized Google Distance, or a combination of both. Experimental results show that the ESA approach is effective in identifying the relatedness between terms.

The Wikify! system [38] supports both algorithms for keyword extraction from documents and word sense disambiguation to assign to each extracted keyword a link to the correct Wikipedia article. The keyword extraction algorithm is based on two steps: (a) candidate extraction, which extracts all possible n-grams that are also present in a controlled dictionary, and (b) keyword ranking, which is based on tf-idf statistics, χ^2 independence test or Keyphraseness (i.e., the probability that a term be selected as a keyword for a document). Three different disambiguation algorithms are integrated in the system. The first one is based on the overlap between the terms in the document and a set of ambiguous terms stored in a dictionary. The second one

is based on a Naïve Bayes classifier whose model is built on feature vectors of correlated Wikipedia articles. Finally, a voting system which takes care of disambiguation results obtained by previous techniques was also employed. Independent evaluations carried out for each of the two tasks showed that both system components produce accurate annotations. The best performance for the keyword extraction task is achieved by the Keyphraseness statistics with accuracy, recall, and F-measure results of 53.37%, 55.90%, and 54.63%, respectively. The disambiguation procedure reaches an accuracy of 94% at best.

2.6 Mining Community-Contributed Media

An overview of different works focused on mining huge collections of community-contributed media is presented in the following. Section 2.6.1 describes different approaches for the extraction of semantics from photo tags available on Flickr, while Sect. 2.6.2 presents how Wikipedia articles can be used as a knowledge base to achieve an automatic classification over electronic documents. Finally, Sect. 2.6.3 describes two research efforts aimed at categorizing and automatically organizing large sets of video clips.

2.6.1 Semantics Extraction from Photo Tags

Photo tags, in the form of unstructured knowledge without a priori semantics, can be efficiently mined to automatically extract interesting and relevant semantics. Many works have been devoted to these issues, which can be classified according to the taxonomy shown in Fig. 2.3.

A lot of research effort has been devoted to jointly analyzing Flickr tags with photo location and time metadata. Approaches proposed in [36, 39] analyze inter-tag frequencies to discover relevant and recurrent tags within a given period of time [39] or space [40]. However, semantics of specific tags were not discovered.

One step further toward the automatic extraction of semantics from Flickr tags was based on analyzing temporal and spatial distributions of each tag's usage [41]. The proposed approach extracts place and event semantics by analyzing the usage in the space and time dimensions of the user-contributed tags assigned to photos on Flickr. Based on temporal and spatial tag usage distributions, a *scale-structure identification* (SSI) approach is employed, which clusters usage distributions at multiple scales and measures the degree of similarity to a single cluster at each scale. Tags can ultimately be identified as *places* and/or *events*. The proposed technique is based on the intuition that an event refers to a specific segment of time, while a place refers to a specific location. Hence, relevant patterns for event and place tags "burst" in specific segments of time and regions in space, respectively. In particular, the number of usage occurrences for an event tag should be much higher in a small segment of time than the number of usage occurrences of

that tag outside the segment. However, the segment size and the number of usage occurrences inside and outside the segments significantly affect the analysis.

The evaluation has been focused on photos from the San Francisco Bay area. Experimental analysis has been performed on a dataset including both photos and tags. The dataset consists of 49,897 photos with an average of 3.74 tags per photo. Each photo is also characterized by a location and a time. The location represents the latitude–longitude coordinates either of the place where the photo was taken or of the photographed object. The time represents either the photo capture time or the time the photo was uploaded to Flickr. These photos cover a temporal range of 1,015 days, starting from 1 January 2004. The average number of photos per day was 49.16, with a minimum of zero and a maximum of 643. From these photos, 803 unique tags were extracted. The maximum number of photos associated with a single tag was 34,325 for the San Francisco Bay area, and the mean was 232.26. The method described in [41] achieves good precision in classifying tags as either a place or event.

A parallel effort has been devoted to enhancing the approach proposed in [41] to efficiently mine the huge photographic dataset managed by Flickr. The approach proposed in [42] is organized in three steps. First, the issue of generating representative tags for arbitrary areas in the world is addressed using a location-driven approach. Georeferences associated with the uploaded photographs are initially exploited to cluster photographs. Candidate tags within each cluster are then ranked to select the best representative ones. The extracted tags often correspond to landmarks within the selected area. Second, the method to identify tag semantics proposed in [41] has been exploited. The method allows the automatic identification of tags as places and/or events based on temporal and spatial tag usage distributions. Lastly, tag-location-driven analysis is combined with computer vision techniques to achieve the automatic selection of representative photographs of some landmark or geographic feature. Tags that represent landmarks and places are initially selected by the aforementioned location-driven approach. For each tag, the corresponding images are clustered by the k-Means [33] clustering algorithm according to their visual content to discover varying views of the landmark in question. For this purpose, range of complementary visual features is extracted from the images. Clusters are subsequently ranked by applying four distinct methods so as to identify the ones which best represent the various views associated with a given tag or location. Finally, images within each cluster are also ranked according to how well they represent the cluster.

The proposed techniques have been evaluated on a set of over 110,000 georeferenced photos from the San Francisco area by manually selecting ten landmarks of interest. Results showed that the tag-location-visual-based approach is able to select representative images for a given landmark with an increase in precision of more than 45% over the best nonvisual technique (tag-location based). Across most of the locations, all of the selected images were representative. For some geographical features, the visual-based methods still do not provide perfect precision in image summaries, mainly due to the complex variety of scenarios and ambiguities connected with the notion of *representativeness*.

2.6.2 *Wikipedia*

The huge amount of data available with Wikipedia articles represents an interesting media collection that can be used to improve automatic document understanding and retrieval as shown in Fig. 2.4. An overview of results achieved in the document categorization domain is presented in Sect. 2.6.2.1, while Sect. 2.6.2.2 discusses some research efforts targeted at developing novel and efficient search engines.

2.6.2.1 Document Classification

The categorization task is usually performed by building models based on the statistical properties of small document collections. The variety of Wikipedia articles, the hyperlink graph structure, and the taxonomy of categories have been employed in different studies to build automatic categorization approaches or improve the performance of existing models. According to the representation, we can divide the discussed works into three categories as depicted in the taxonomy in Fig. 2.4: (a) bag-of-word, (b) Wikipedia taxonomy analysis, and (c) Wikipedia graph analysis.

In [43], Gabrilovich and Markovitch presented one of the first works employing Wikipedia as an external resource for the document categorization task. The idea is to improve document representation by using the knowledge stored in the encyclopedia. A feature generator identifies the most relevant encyclopedia articles for each document. Then, the titles of the articles are used as new features to augment the bag-of-words (BOW) representation of the document. In the BOW representation, a document or a sentence is represented as an unordered collection of words, disregarding the structure of the text. This representation is usually associated with statistical measures such as tf-idf (term frequency-inverse document frequency). The tf-idf is used to evaluate how important a word is with respect to a document in a collection: the higher this value, the more representative the word. Empirical evaluation shows that, using background knowledge stored in Wikipedia, classification performance on short and long documents drawn from different datasets can be improved with respect to traditional classification approaches based only on the BOW representation.

A similar idea was presented in [44], where the authors automatically constructed a thesaurus of concepts from Wikipedia. The thesaurus was extracted using redirect and disambiguation pages and the hyperlink graph of Wikipedia articles. Similarly to the previous approach, the authors search candidate concepts mentioned in each document, but then they add synonyms, hyponyms, and correlated concepts of these candidate concepts, used as new features to enrich the BOW representation. This extended knowledge can be leveraged to relate documents which did not originally share common terms. Therefore, such documents are shifted closer to each other in the new representation. The effectiveness of this approach was empirically demonstrated by means of a linear Support Vector Machine (SVM)

[33] to classify documents from different datasets. The micro-averaged and the macro-averaged of the precision–recall break-even point (BEP) were used to compare classification performance with respect to the baseline approach. Compared with the baseline method, the proposed approach yielded an improvement of 2–5%.

A new classification model based on Wikipedia information was proposed by Schönhofen in [45]. The relatedness between a document and a category of the Wikipedia taxonomy was computed by evaluating the similarity between that document and the titles of articles classified under that category. Since each article can belong to different categories, relevance statistics are used to rank the categories. The method was tested on the Wikipedia article body, which is not used to build the model, and on news datasets. The best results are achieved by combining Wikipedia categorization with the top terms identified by tf-idf. For example, the accuracy achieved on the news dataset is around 89%.

An improvement over Schönhofen's approach was suggested in [46]. The authors propose to exploit both the words appearing in the article titles and in the hyperlinks. In fact, the hyperlinks better characterize the content of the article. Empirical results on a subset of Wikipedia articles show an improvement in precision and recall with respect to Schönhofen's method. Using only the top-3 Wikipedia categories returned by the method, the improvement in precision and recall is around 15% and 35%, respectively.

A different text categorization approach based on an RDF ontology extracted from Wikipedia Infoboxes was presented in [47]. The method focuses on news documents of varying themes. For each document, the authors manually selected the Wikipedia category which best relates to its topic. A text document is then converted into a "thematic graph" of entities occurring in the document. Since the thematic graph can include uncorrelated entities, a selection of the most dominant component is applied. Finally, the text is classified according to the best coverage class of the entities belonging to the graph. The accuracy achieved by this approach on two different document collections is worse than that of a Naïve Bayes classifier [33] based on BOW representation. One of the reasons for misclassifications may be the manual mapping of Wikipedia categories to the document topics. Moreover, news documents – unlike encyclopedia content – may be biased to reflect the interest of the readers. Yet, an interesting highlight of the ontology-based categorization approach is that it does not require a training phase, since all information about categories is stored in the ontology.

2.6.2.2 Search Engine

Several search engines based on Wikipedia have been developed to retrieve documents which are highly correlated with the keywords typed by the user. As shown in the taxonomy depicted in Fig. 2.4, the proposed approaches can be classified according to the Wikipedia information employed in the query analysis.

Some approaches have been addressing the user interactivity in an effort to improve the relevance of the results. For example, the Koru search engine [48] allows the user to automatically expand queries with semantically related terms through an interactive interface to extract relevant documents. The interface is composed of three panels: (a) the *query topic panel*, which provides users with a summary based on a ranking of significant topics extracted from the query, (b) the *query results*, which presents the outcome of the query in the form of a series of document surrogates, and (c) the *document tray*, which allows users to collect multiple documents they wish to peruse. Real users were asked to experiment with the system to identify the improvements offered over traditional keyword search. The main advantages reported by testing users were the capability of lending assistance to almost every query and the improved relevance of the documents returned.

Other approaches have been devoted to improve efficiency, in terms of query processing scalability, and proficiency in entity recognition. Bast et al. [49] present the ESTER modular system for highly efficient combined full-text and ontology search. It is based on graph-pattern queries, expressed in the SPARQL language, and on an entity recognizer. The entity recognizer combines a supervised technique with a disambiguation step to identify concepts in the query and in the documents. In addition, the system includes a user interface which suggests a semantic completion based on the typed keywords, and the display of properties of a desired entity. The interface is designed in such a way as to offer all the features of a SPARQL-based query engine, with the added benefit of being intuitive for inexpert users. For example, when a user has typed "Beatles musician", the system will give instant feedback that there is semantic information on musicians, and it will execute, in addition to an ordinary full-text query, a query searching for instances of that class (in the context of the other parts of the query), showing the best hits for either query. Good performance in terms of scalability and entity recognition has been achieved by the proposed system.

2.6.3 Video Content Interpretation

User-contributed video collections like YouTube present new opportunities and novel challenges to mine large amounts of videos and extract knowledge useful to categorize and organize their content. Following the taxonomy depicted in Fig. 2.2, Sect. 2.6.3.1 presents a classification technique based on the visual features of video clips, while Sect. 2.6.3.2 describes a novel approach to synchronize and organize a set of video clips related to a concert event.

2.6.3.1 Concept-Based Classification

Automatic indexing of video content has already received significant interest as an alternative to manual annotation and aims at deriving meaningful descriptors from the video data itself. Since such data is of sensory origin (image, sound, video), techniques from digital signal processing and computer vision are employed to

extract relevant descriptions. The role of visual content in machine-driven labeling has been long investigated and has resulted in a variety of content-based image and video retrieval systems. Such systems commonly depend on low-level visual and spatiotemporal features and are based on the query-by-example paradigm. As a consequence, they are not effective if proper examples are unavailable. Furthermore, similarities in terms of low-level features do not easily translate to the high-level, semantic similarity expected by users.

Concept-based video retrieval [50] tries to bridge this semantic gap and has evolved over the last decade as a promising research field. It enables textual queries to be carried out on multimedia databases by substituting manual indexing with automatic detectors that mine media collections for semantic (visual) concepts. This approach has proven effective and, when a large set of concept detectors are available, its performance can be comparable with that of standard Web search [51]. Concept detection relies on machine learning techniques and requires, to be effective, that vast training sets be available to build large-scale concept dictionaries and semantic relations. So far, the standard approach has been to employ manually, expert-labeled training examples for concept learning. This solution is costly and gives rise to additional inconveniences: the number of learned concepts is limited, the insufficient scale of training data causes overfitting, and adapting to changes (like new concepts of interest) remains difficult.

In [52], the huge video repository offered by YouTube is utilized as a novel kind of knowledge base for the machine interpretation of multimedia data. Web videos are exploited for two distinct purposes. On the one hand, result video clips of a YouTube search for a given concept are employed as positive examples to train the corresponding detector. Negative examples are drawn from other videos not tagged with that concept. Frames are sampled from the videos and their visual descriptors are fed to several statistical classifiers (Support Vector Machines, Passive-Aggressive Online Learning, Maximum Entropy), whose performance has been compared. On the other hand, tag co-occurrences in video annotations are used to link concepts. For each concept, a bag-of-words representation is extracted from the tags of the associated video clips. The process is then repeated with user queries, which are thus mapped to the best matching learned concepts. The approach has been evaluated on a large dataset (1,200 hours of video) by manually selecting 233 concepts. Precision in detecting concepts rapidly increases with the number of videoclips included in the training set and stabilizes when 100–150 videos are used. Results show that the average achieved precision, although largely dependent on the concepts, is promising (32.2%), suggesting that Web-based video collections indeed have the potential to support unsupervised visual and semantic learning.

2.6.3.2 Automated Synchronization of Video Clips

The abundance of video material found in user-generated video collections could enable a broad coverage of captured events. However, the lack of detailed semantic and time-based metadata associated with video content makes the task of identifying and synchronizing a set of video clips relative to a given event a nontrivial one.

Kennedy and Naaman [53] propose the novel application of existing audio fingerprinting techniques to the problem of synchronizing video clips taken at the same event, particularly concert events. Synchronization allows the generation of important metadata about the clips and the event itself and thus enhances the user browsing and watching experience.

A set of video clips crawled from the Web and related to the same event is assumed to be initially available. Fingerprints are generated for each of them by spectral analysis of the audio tracks. The results of this process are then compared for any two clips to identify matches. Both the fingerprinting techniques and the matching algorithm are quite robust against noisy sources – as is often the case with user-contributed media. Audio fingerprinting matches are exploited to build an undirected graph, where each node represents a single clip and edges indicate temporal overlapping between pairs of clips. Such a graph typically includes a few connected components, or clusters, each one corresponding to a different portion of the captured event. Based on the clip overlap graph, information about the level of interest of each cluster is extracted and highly interesting segments of the event are identified. In addition, cluster analysis is employed to aid the selection of the highest quality audio tracks.

Textual information provided by the users and associated with the video clips is also mined by a tf-idf strategy to gather descriptive tags for each cluster so as to improve the accuracy of search tasks, as well as suggest metadata for unannotated video clips.

This system has been applied to a set of real user-contributed videos from three music concerts. A set of initial experiments enabled the fine-tuning of the audio fingerprinting and matching algorithms to achieve the best matching precision. Manual inspection showed that a large fraction of the clips left out by the system were very short or of abnormally low quality, and thus intrinsically uninteresting. Proficiency in identifying important concert segments (typically hit songs) has then been assessed by comparison with rankings found on the music-sharing Web site Last.fm, with a positive outcome. A study with human subjects has also been conducted which was able to validate the system's selection of high-quality audio. Finally, the approach proposed to extract textual descriptive information proved successful in many cases, with failure cases being mostly related to poorly annotated clips and small clusters.

2.7 Perspective

This chapter presented and discussed different community-contributed media collections by highlighting the main challenging research issues opened by the online social networking sites.

In the last few years, an increasing research effort has been devoted to studying and understanding the growth trend in social networks and the media content distribution. Furthermore, the abundance of multimodal and user-generated media has opened novel research perspectives and thus introduced novel challenges to mine large collections of multimedia data and effectively extract relevant knowledge. Much research has been focused on improving the automatic understanding

of user-contributed media, enhancing the user searching activity on media collection, and automatically classifying Wikipedia articles.

However, less attention has been paid to the integration of heterogeneous data available on different online social networking sites to tailor personalized multimedia services. For example, users can explore and gain a broader understanding of a context by integrating the huge amount of data stored in Wikipedia, YouTube, and Flickr. Digital libraries which cover a specific field (e.g., geography, economy) can be enriched by extracting related information from Wikipedia. Similarly, news videos can be integrated and contextualized with information provided by Wikipedia articles. Only few approaches have been proposed to address these issues [54, 55].

The problem of integrating Wikipedia and a geographic digital library is presented in [54]. The integration approach consists of identifying relevant articles correlated to geographical entries in the digital library. The identification is carried out by analyzing *List_of_<G>* Wikipedia pages, where *G* is a geographical entity (e.g., region, country, city). Finally, additional information is extracted by parsing infoboxes content of selected Wikipedia pages. Experimental evaluation on the extraction of relevant articles and metadata information show good performance in precision and recall.

The integration of news videos and Wikipedia articles about news events is addressed in [55]. The method aims to automatically label news videos with Wikipedia entries in order to provide more detailed explanations about the context of the video content. Using "Wikinews," a sister project of Wikipedia, all news-related articles are extracted from Wikipedia, while information about videos is extracted from the content caption (CC), which is usually provided. The CCs of news stories are labeled with Wikipedia entries by evaluating date information. Experimental results show that Japanese news videos broadcast over a year were accurately labeled with Wikipedia entries with a precision of 86% and a recall of 79%.

A host of technical challenges remain for better exploiting the community-contributed media into personalized applications able to tailor multimedia services according to the context in which the user is currently involved. Multimedia services for mobile devices could be personalized by exploiting both the current context of the user and the huge amount of media content available on the online social networking sites. For example, WikEye [56], a system for mobile technology, is able to retrieve interesting information on touristic places from Wikipedia spatial and temporal data. Future research directions might also exploit relevant knowledge available in different media collections.

References

1. Youtube website. http://www.youtube.com/
2. Flickr website. http://www.flickr.com/
3. Wikipedia website. http://www.wikipedia.com/
4. Alexa the web information company. http://www.alexa.com/
5. Paolillo, J.: Structure and network in the youtube core. In: hicss, p. 156 (2008)

6. Facebook website. http://www.facebook.com/
7. Twitter website. http://twitter.com/
8. Myspace website. http://www.myspace.com/
9. Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries. In: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, pp. 326–335 (2002)
10. Heesch, D.: A survey of browsing models for content based image retrieval. Multimedia Tools and Applications **40**(2), 261–284 (2008)
11. Giles, J.: Special report: Internet encyclopedias go head to head. Nature **438**(15), 900–901 (2005)
12. Smets, K., Goethals, B., Verdonk, B.: Automatic vandalism detection in Wikipedia: Towards a machine learning approach. In: AAAI Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy, pp. 43–48 (2008)
13. Tang, L., Liu, H.: Graph Mining Applications to Social Network Analysis. Managing and Mining Graph Data pp. 487–513 (2010)
14. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. The Semantic Web: Research and Applications pp. 411–426 (2006)
15. Lerman, K., Galstyan, A.: Analysis of social voting patterns on digg. In: Proceedings of the first workshop on Online social networks, pp. 7–12 (2008)
16. Cheng, X., Dale, C., Liu, J.: Statistics and social network of youtube videos. In: Proceedings of 16th International Workshop on Quality of Service, pp. 229–238 (2008)
17. Mislove, A., Koppula, H.S., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Growth of the flickr social network. In: Proceedings of the first workshop on Online social networks, pp. 25–30 (2008)
18. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y., Moon, S.: I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, p. 14 (2007)
19. Cha, M., Mislove, A., Adams, B., Gummadi, K.P.: Characterizing social cascades in flickr. In: Proceedings of the first workshop on Online social networks, pp. 13–18 (2008)
20. Van Zwol, R.: Flickr: Who is looking? In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, pp. 184–190 (2007)
21. Daum website. http://www.daum.net/
22. Pastor-Satorras, R., Vespignani, A.: Epidemics and immunization in scale-free networks. In: Handbook of Graphs and Networks: From the Genome to the Internet (2002)
23. Kittur, A., Chi, E., Suh, B.: What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In: Proceedings of the 27th international conference on Human factors in computing systems, pp. 1509–1512 (2009)
24. Tjondronegoro, D., Spink, A.: Web search engine multimedia functionality. Information Processing & Management **44**(1), 340–357 (2008)
25. Davis, S., Burnett, I., Ritz, C.: Using Social Networking and Collections to Enable Video Semantics Acquisition. IEEE MultiMedia (2009)
26. Ames, M., Naaman, M.: Why we tag: motivations for annotation in mobile and online media. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 971–980 (2007)
27. Garg, N., Weber, I.: Personalized tag suggestion for flickr. In: Proceeding of the 17th international conference on World Wide Web, pp. 1063–1064 (2008)
28. Golder, S., Huberman, B.: Usage patterns of collaborative tagging systems. Journal of Information Science **32**(2), 198 (2006)
29. Marlow, C., Naaman, M., Boyd, D., Davis, M.: HT06, tagging paper, taxonomy, Flickr, academic article, to read. In: Proceedings of the seventeenth conference on Hypertext and hypermedia, p. 40 (2006)
30. Noll, M., Meinel, C.: Authors vs. readers: a comparative study of document metadata and content in the www. In: Proceedings of the 2007 ACM symposium on Document engineering, p. 186 (2007)

31. Sigurbjörnsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proceeding of the 17th international conference on World Wide Web, pp. 327–336 (2008)
32. Reed, W.: The pareto, zipf and other power laws. Economics Letters **1**, 15–19 (2001)
33. Tan, P., Steinbach, M., Kumar, V.: Introduction to data mining (2006)
34. Mikroyannidis, A.: Toward a Social Semantic Web. COMPUTER pp. 113–115 (2007)
35. Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., Studer, R.: Semantic wikipedia. In: Proceedings of the 15th international conference on World Wide Web, p. 594. ACM (2006)
36. Gabrilovich, E., Markovitch, S.: Computing semantic relatedness using wikipedia-based explicit semantic analysis. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence, pp. 6–12 (2007)
37. Witten, D., Milne, D.: An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. In: Proceeding of AAAI Workshop on Wikipedia and Artificial Intelligence: an Evolving Synergy, AAAI Press, Chicago, USA, pp. 25–30 (2008)
38. Mihalcea, R., Csomai, A.: Wikify!: linking documents to encyclopedic knowledge. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, pp. 233–242 (2007)
39. Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., Tomkins, A.: Visualizing tags over time. In: Proceedings of the 15th international conference on World Wide Web, pp. 193–202 (2006)
40. Jaffe, A., Naaman, M., Tassa, T., Davis, M.: Generating summaries and visualization for large collections of geo-referenced photographs. In: Proceedings of the 8th ACM international workshop on Multimedia information retrieval, pp. 89–98 (2006)
41. Rattenbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from flickr tags. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 103–110 (2007)
42. Kennedy, L., Naaman, M., Ahern, S., Nair, R., Rattenbury, T.: How flickr helps us make sense of the world: context and content in community-contributed media collections. In: Proceedings of the 15th international conference on Multimedia, pp. 631–640 (2007)
43. Gabrilovich, E., Markovitch, S.: Overcoming the brittleness bottleneck using Wikipedia: enhancing text categorization with encyclopedic knowledge. In: Proceedings of the National Conference on Artificial Intelligence, vol 21, p 1301 (2006)
44. Wang, P., Hu, J., Zeng, H., Chen, Z.: Using Wikipedia knowledge to improve text classification. Knowledge and Information Systems **19**(3), 265–281 (2009)
45. Schönhofen, P.: Identifying document topics using the Wikipedia category network. Web Intelligence and Agent Systems **7**(2), 195–207 (2009)
46. Huynh, D., Cao, T., Pham, P., Hoang, T.: Using hyperlink texts to improve quality of identifying document topics based on wikipedia. International Conference on Knowledge and Systems Engineering pp. 249 –254 (2009)
47. Janik, M., Kochut, K.: Wikipedia in action: Ontological knowledge in text categorization. In: International Conference on Semantic Computing, pp. 268–275 (2008)
48. Milne, D., Witten, I., Nichols, D.: A knowledge-based search engine powered by wikipedia. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, pp. 445–454 (2007)
49. Bast, H., Chitea, A., Suchanek, F., Weber, I.: ESTER: efficient search on text, entities, and relations. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, p. 678 (2007)
50. Snoek, C., Worring, M.: Concept-based video retrieval (2009)
51. Hauptmann, A., Yan, R., Lin, W., Christel, M., Wactlar, H.: Can high-level concepts fill the semantic gap in video retrieval? a case study with broadcast news. IEEE Transactions on Multimedia **9**(5), 958–966 (2007)
52. Ulges, A., Koch, M., Borth, D., Breuel, T.: TubeTagger-YouTube-based Concept Detection. In: IEEE International Conference on Data Mining Workshops, pp. 190–195 (2009)

53. Kennedy, L., Naaman, M.: Less talk, more rock: automated organization of community-contributed collections of concert videos. In: Proceedings of the 18th international conference on World wide web, pp. 311–320 (2009)
54. Lim, E., Wang, Z., Sadeli, D., Li, Y., Chang, C., Chatterjea, K., Goh, D., Theng, Y., Zhang, J., Sun, A., et al.: Integration of Wikipedia and a geography digital library. Lecture Notes in Computer Science 4312, 449 (2006)
55. Okuoka, T., Takahashi, T., Deguchi, D., Ide, I., Murase, H.: Labeling News Topic Threads with Wikipedia Entries. In: 11th IEEE International Symposium on Multimedia, pp. 501–504 (2009)
56. Hecht, B., Rohs, M., Schöning, J., Krüger, A.: Wikeye - using magic lenses to explore georeferenced Wikipedia content. In: Proceedings of the 3rd International Workshop on Pervasive Mobile Interaction Devices (2007)