

# An Application Model of Semantic Web Based Social Web Mining

Reshma. P. K

Department of Computer Science  
Mahatma Gandhi College, Iritty,  
Kannur, Kerala

Lajish V.L

Department of Computer Science  
University of Calicut,  
Kerala-673635

**Abstract:-** Social network has gained remarkable attention in recent years. Accessing social network sites such as Twitter, Facebook, and LinkedIn through the Internet has become more feasible and affordable. People are become more interested in social networks for information, news and opinion of others on different subjects and to post their own messages. The popularity of social network sites causes the generation of huge amount of multimedia data with three computational issues namely size, noise and dynamism. These often make social network data very complex to analyze them manually. Data mining provides many techniques for detecting useful knowledge from massive datasets. Multimedia mining is a recent but challenging sub field in Data Mining. There are no unified conclusions in the concept, content and methods of multimedia mining, architecture and framework. This paper discourses the practicality of using semantic web framework for mining the multimedia data form the social web.

**Keywords-** *Multimedia, Social Multimedia, Social Networks, Data Mining, Social Networks, Web Mining, Semantic Web, Ontology*

## I. INTRODUCTION

The term Social network is used to describe web based services that allow individuals to create a public/semi-public profile within a domain such that they can communicatively connect with other users within the network [1]. Social media services allow people to share multimedia content in a massive scale. The multimedia content production became simple; the cost of publishing that content became lower and wide potential reach result in a significant amount of content available on the web. Social networks play a major role in making the world a global village through the universally accepted communication means. They are popular for information broadcasting, personal activities posting, product reviews, online pictures and video sharing, professional profiling, news alerts, political debates, opinion or sentiment expression and a lot more [2].

However social media sites provide data which are huge, noisy, dynamic and distributed. Above all they cause a tremendous increase in the multimedia data on the web. Because of the huge size of data, data structure complexity, patterns and diverse characteristics, the study of multimedia database has been a difficult one. Based on the characteristics of multimedia database, multimedia data mining has attracted wide attention but is still at the start up stage.

## II. SOCIAL MULTIMEDIA

The term social multimedia refers to multimedia resources available through social networks or online sources of multimedia content posted in settings that promote significant individual participation and that promote discussion and re-use of content [5]. Social multimedia captures and leverages community activity around multimedia data, using explicit user input like tags and comments [6, 7] as well as implicit input from users like mass viewing patterns in item and sub-item levels [8]. Apart from the scale of available content, such services make new context information and metadata about the content widely available. These may include textual descriptors, information about location, camera properties, user information and social network data.

Social multimedia also offers the opportunity to design interactive systems that extract new explicit and implicit metadata from user interaction. Such interaction and user input is often driven by social motivations [9, 10] and can improve the data available for multimedia applications. Thus, social multimedia offers several opportunities that go beyond and above other "Web multimedia" sources where many of these opportunities are not available [11].

The social media have its own significant limits and challenges. The above mentioned context and available metadata are noisy and often inaccurate, wrong or misleading [12]. As a result, there is very little "ground truth" for social media data. The noise and lack of semantics make even the simplest of metadata such as user-provided tags, difficult to use.

Social multimedia search and mining demands a shift of focus from traditional multimedia applications. The generalized approach to social multimedia applications is described here as a series of steps [11], including:

*Step I:* Identify topic and application domain and use simple context-based tools to identify relevant content items.

*Step II:* Use application-specific, constrained and "knowledge-free" (unsupervised) content analysis techniques to improve precision, representation and selection of items.

*Step III:* Use the content analysis output to further improve metadata for aggregate multimedia items.

*Step IV:* Leverage user interaction for improving relevance and representation

Thus the social multimedia offers different opportunities for research in multimedia domain, like

analyzing community activity around multimedia resources and pooling of content in social settings.

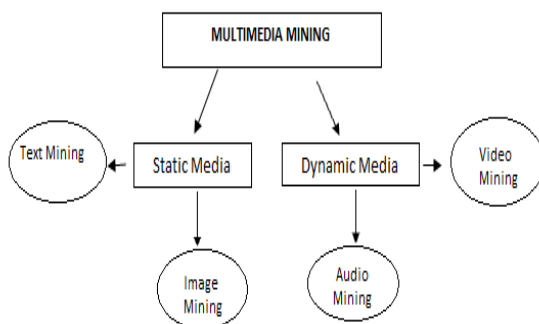
One notable benefit of social multimedia is the opportunity to aggregate data or analyze activities around the individual resources to better reason about their content or to understand the interesting areas of certain groups. Social network platforms enable rapid information exchange between users regardless of the location. Many organizations, individuals and even government now follow the activities on social network. The network enables big organizations, celebrities, government official and government bodies to obtain knowledge on how their audience reacts to postings that concerns them out of the enormous data generated on social network. The application of efficient data mining techniques has made it possible for users to discover valuable, accurate and useful knowledge from social network data [2].

The data mining techniques should handle the dominant features of social network data such as size, noise and dynamism to effectively use the data available through these networks.

### III. MULTIMEDIA MINING

Multimedia mining is a subfield of data mining which is used to find interesting information of implicit knowledge from multimedia databases. A multimedia database management system (MM-DBMS) manages and provides support for storing, manipulating and retrieving multimedia data from a large collection of multimedia objects such as video, image, audio and hypertext data. Multimedia data are classified into five types; they are (i) text data, (ii) image data (iii) audio data (iv) video data and (v) electronic and digital ink [3]. Figure 1 shows different categories of multimedia mining.

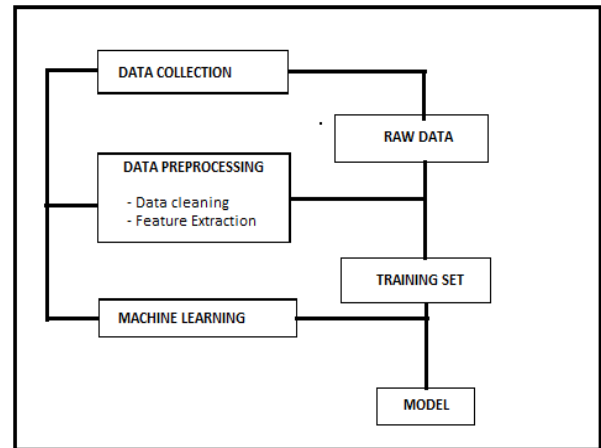
FIGURE 1. CATEGORIES OF MULTIMEDIA MINING



The process of applying multimedia mining consists of different steps. Data collection is the first point of a learning system, as the quality of raw data is the factor which determines the overall achievable performance. The main goal of data pre-processing is to discover the important patterns from the raw data, which includes the concepts of data cleaning, normalization, transformation, feature selection etc. Learning can be simple, if informative features can be identified at pre-processing stage. Detailed procedure depends highly on the nature of raw data and problem's domain. The product of data pre-processing is the training

set. Given a training set, a learning model has to be chosen to learn from it and make multimedia mining model more iterative. The process of Multimedia mining is shown in figure2 [4].

FIGURE 2. MULTIMEDIA MINING PROCESS



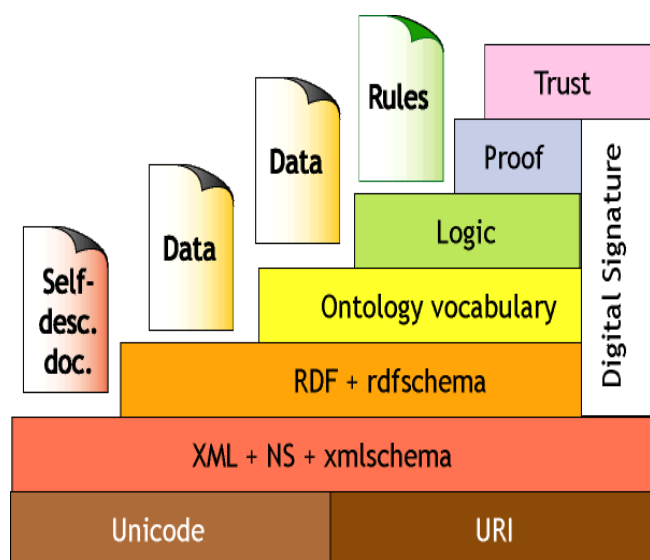
The multimedia files from a database are first pre-processed to improve their quality and followed by feature extraction. With the help of generated features, information models can be devised using data mining techniques such as pattern discovery, rule extraction and knowledge acquisition to discover significant patterns from multimedia database [13].

### IV. SEMANTIC WEB FRAMEWORK

Semantic Web mining is the result of combining two fast growing areas such as semantic web and web mining. The tools of semantic web can be used to improve web mining and vice versa. As mentioned earlier the web contains a huge amount of unstructured data which is in a human understandable form. The goal of the semantic web is to provide machine interpretable semantics to offer greater machine support for the user. Thus semantic structures can improve the mining task by allowing the algorithms to operate on certain semantic levels or choose appropriate levels of abstraction.

The semantic web has a layered structure that defines the level of abstraction applied to the web as given in figure 3. This is the structure for the semantic web suggested by Tim Berners Lee [14].

FIGURE 3. THE SEMANTIC WEB LAYER AS PRESENTED BY TIM BERNERS-LEE



The detailed description of each layer in the semantic web framework is given below.

#### Layer 1- UNICODE and URI:

A common syntax is provided for the first two layers. Uniform Resource Identifiers (URI) provides a standard way to refer to resources whereas Unicode provides a standard for exchanging symbols of different languages. With these protocols one can transmit the web pages over the Internet. At this level one does not deal with syntax or the semantics of the documents

#### Layer 2- XML, XML Scheme and Namespace:

The Extensible Markup Language (XML) is a language used to represent data in a structural way. It describes what is in the document, not what the documents looks like, while XML Schema provides grammars for valid XML documents which can refer to different Namespaces (NS) to make explicit context of different tags. Namespaces allow the combination of different vocabularies. The formalizations on these two layers are widely accepted nowadays.

#### Layer 3 – RDF and RDF Schema:

The Resource Description Framework (RDF) can be the first layer where information becomes machine understandable. According to W3C, RDF is a foundation for processing metadata; it provides interoperability between applications that exchange machine understandable information on the Web. RDF documents consist of three types of entities: Resources, properties, and statements. Resources may be Web pages, parts or collections of Web pages, or any (real-world) objects which are not directly part of the WWW. In RDF, resources are always addressed by URIs. Properties are specific attributes, characteristics, or relations describing resources. Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different application. Statements can be considered as object-attribute-value triplets. A value is either a literal, a resource or another statement. RDF Schema

extends RDF and is a vocabulary for describing properties and classes of RDF-based resources, with semantics for generalized-hierarchies of such properties and classes.

#### Layer 4 – Ontology Vocabulary:

The next layer is the ontology vocabulary. Ontology is a formal explicit description of concepts in a domain of discourse, properties of each concept describing various features and instances of the concept. Ontology with a set of instances of classes constitutes a knowledge base.

#### Layer 5 – Logic:

Logic the next layer in this architecture. Nowadays, logic and Ontology are used in an integrated fashion because most ontologies allow logical axioms. By applying logical deduction, new knowledge can be inferred from the information which is stated explicitly.

#### Layer 6 – Proof:

Proof is the layer placed above the Logic layer. It is assumed to be a language used in a manner that describes for agents why they should believe the results. This will be a useful semantic web service.

#### Layer 7 – Trust:

A lot of efforts have been exerted to reach the trusted web, but this is very complicated and difficult task and has not become a reality. Trust has many meanings in the semantic web. Trust is the final layer in the semantic web architecture. It depends on the source of information as well as the policies available on the information source which can prevent unwanted applications or user from access to these sources.

#### Digital signature:

Digital Signature is the only vertical layer in the semantic web architecture. It begins from layer 3 and ends at layer 6. Digital Signature is a step towards a web of trust. By using XML digital signature, any digital information can be signed [15]. There are specific elements in XML syntax used for this process such as Signed Info, Reference and Digest Value [16]. The final layers are logic, proof and trust. The idea here is, how the information on the web is trusted? Obviously it depends on whom it comes from. To carry out trust negotiation, interested parties have to communicate with each other and determine how to trust each other and how to trust the information obtained on the web. This semantic web framework can be effectively used for mining the multimedia data from the social web.

## V. CONCLUSION

In recent years, several web based sharing and community services have made a rapid change in the size and type of multimedia content, features and depth of metadata available online. This is the right time for multimedia research as this is the real goal behind the development of the web by Tim Berners-Lee, to help people work together and to support and improve our web like existence in the world. Facebook, Twitter and LinkedIn generate a tremendous amount of

valuable social data, but these should be analysed in a robust manner to ensure the propagation of the right information to the right people. There are different approaches used. The effort behind the Semantic Web is to add semantic annotations to web documents in order to access knowledge instead of unstructured material allowing knowledge to be managed in an automatic way. This paper gives a pathway for the future research by providing the basic knowledge of multimedia mining from the social media based on the semantic web architecture.

#### ACKNOWLEDGMENT

We thank the staff members and research scholars of the Department of Computer Science, University of Calicut and the staff members at Department of Computer Science, Mahatma Gandhi College, Iritty for fruitful discussions and support.

#### REFERENCES

- [1] Chen, Z. S., Kalashnikov, D. V. and Mehrotra, S. Exploiting context analysis for combining multiple entity resolution systems. In Proceedings of the 2009 ACM International Conference on Management of Data (SIGMOD'09), 2009 D.
- [2] Mariam Adedoyin-Olowe , Mohamed Medhat Gaber and Frederic Stahl, A survey of Data Mining Techniques for social Network analysis,.
- [3] Sarla More, Durgesh Kumar Mishra, Multimedia Data Mining: A Survey. Pratibha: International Journal of science, spirituality, business and technology (ijssbt), vol. 1, no.1, march 2012 issn (print) 2277—7261.
- [4] Bhavani Thuraisingham, Managing and Mining Multimedia Data-bases. at International Journal on Artificial Intelligence Tools Vol. 13, No. 3 739-759, 2004.
- [5] Susanne Boll. Multitube—where web 2.0 and multimedia could meet. Multimedia, IEEE, 14(1):9–13, Jan.-March 2007.
- [6] Munmun De Choudhury, Hari Sundaram, Ajita John, and Dore Duncan Seligmann. What makes conversations interesting? themes, participants and consequences of conversations in online social media. In WWW '09: Proceeding of the 18th international conference on World Wide Web, New York, NY, USA, 2009. ACM..
- [7] David A. Shamma, Lyndon Kennedy, and Elizabeth Churchill. Statler: Summarizing media through short-message services. In CSCW '10: Proceedings of the 2010 ACM conference on Computer supported cooperative work, New York, NY, USA, 2010. ACM.
- [8] David A. Shamma, Ryan Shaw, Peter L. Shafon, and Yiming Liu. Watch what i watch: using community activity to understand content. In MIR '07: Proceedings of the international workshop on Workshop on multimedia information retrieval, pages 275–284, New York, NY, USA, 2007. ACM
- [9] Morgan Ames and Mor Naaman. Why we tag: Motivations for annotation in mobile and online media. In CHI '07: Proceedings of the SIGCHI conference on Human Factors in computing systems, New York, NY, USA, 2007. ACM Press.
- [10] Oded Nov, Mor Naaman, and Chen Ye. What drives content tagging: the case of photos on flickr. In CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, pages 1097–1100, New York, NY, USA, 2008. ACM
- [11] Mor Naaman. Social Multimedia: Highlighting Opportunities for Search and Mining of Multimedia Data in Social Media Applications, Multimedia Tools and Applications, Springer, pages 56: 9-342012,
- [12] Dick C.A. Bulterman. Is it time for a moratorium on metadata? IEEE MultiMedia, 11(4):10–17, October 2004.
- [13] Dianhui Wang, Yong-Soo Kim, Seok Cheon Park, Chul Soo Lee and Yoon Kyung Han, “Learning Based Neural Similarity Metrics for Multimedia Data Mining” Soft Computing, Vol. 11, Number 4, , pp. 335- 340, 2007.
- [14] D. Fensel, 2002. Layering the semantic web: Problems and Directions. In the Proceeding of 1st International semantic web Conference (ISWC, 2002). Sardinia, Italy, 9-12 June, pp: 476. ISBN: 3540437606, 9783540437604.
- [15] R. Cloran and B. Irwin, 2005. XML Digital Signature and RDF, [http://icsa.cs.up.ac.za/issa/2005/Proceedings/Poster/026\\_Article.pdf](http://icsa.cs.up.ac.za/issa/2005/Proceedings/Poster/026_Article.pdf)
- [16] T. Haytam, Al-Feel, M. Koutb and H. Suoror, semantic web on Scope: A New Architectural Model for the semantic web, Journal of Computer Science 4 (7): 613-624, 2008