

PRIVACY PRESERVING AND OBSCURE DELICATE DATA WITH COLLABORATIVE TAGGING

A Divya¹, S Madhu Sudhanan², M Poornima³

¹PG scholar, Computer science and engineering, Prathyusha institute of technology and management, Chennai, Tamil Nadu

²Assistant professor, Computer science and engineering, Prathyusha institute of technology and management, Chennai, Tamil Nadu

³Assistant professor, Computer science and engineering, Prathyusha institute of technology and management, Chennai, Tamil Nadu

Abstract

Tagging system is one of the most diffused and popular services available online. This system allows users to add free text labels generally referred as tags to the Internet resources for example web pages, images, video, music and even blogs. Web metadata have a potential to improve search, retrieval and to protect end user from possible harmful content. The Organization updates their Company portal with public sharing data along with Sensitive data. The query is processed based on the User Profile Analysis. In actual system provide taxonomy of tagging system and system web technologies help to specify labels and rate for that labels which assess the trustworthiness of resources to enforce web access personalization. To enhance the efficiency of tag suppression the privacy ensured skim with Support Vector Machine along with Privacy Enhancing Technology is implemented. SVM is used for extraction of data and obscure delicate data. PET is achieved by using the technique Tag Suppression which has the role of providing the privacy for information. Web user will search using a keyword. The keyword may be the location, feedback or cost to analyze the data. The authentication of the portal is done by the management. Management classified as two roles they are Department Head Role and Admin. Department Head Role is to update their part of portal and retrieve only the corresponding data. Final authentication and approval is done by the admin. Through the analysis efficiency guarantees of proposed scheme is achieved.

Keywords: Collaborative tagging; Privacy enhancing technology; tag suppression; Support vector machine; unsupervised duplicate detection.

1. INTRODUCTION

A system of segregating derived from the tradition and method of collaboratively constitutes and transcribes tags to annotate and categorize content notorious as collaborative tagging. A social Bookmarking service is centralized online services which facilitate users to add, annotate and share bookmarks of web documents. Tagging is a compelled feature of social bookmarking systems, empowering users to organize their bookmark in malleable ways and develop shared vocabularies noted as folksonomy [3]. The availability of metadata describing web resources has been considered as a controversy in a public information space. Taxonomy is the designations according to pre regulate system, whose catalogue is to provide a conceptual framework for analysis. The development of taxonomy brings into account of separating aspect of group into sub groups that includes all possibilities. Folksonomy has little do with taxonomy which influence the key to developing a semantic web, in which every web page contains machine readable metadata that describes its content which improve the precision in retrieval list.

A perception to use metadata is to protect users from inappropriate content [5]. Despite collaborative tagging is

vital handling to support tag based resource discovery and browsing data, also exploited for other objective. The tags are compiled by social bookmarking services can be exploited to enhance web access functionalities like trustworthiness based on preferences specified by user. These issues can be achieving by collaborative environment and semantics web technologies [5]. The devise mechanism is to assess the trustworthiness of web metadata with the availability of WBSN (Web Based Social Networks) providing their availability of specifying and sharing metadata. Extending collaborative tagging by incorporating a policy layer to protect the information which available in the social services. This can be prior by analyzing multi-layer incorporate with collaborative tagging.

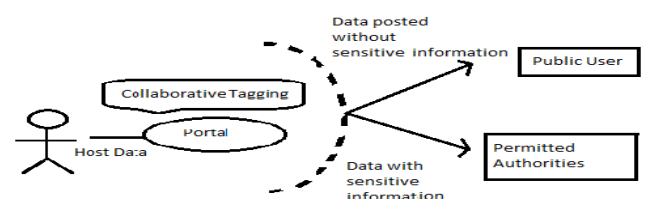


Fig 1 Technique enables to protect the privacy by refraining From Tag suppression to separate the normal data from Sensitive, Can be visible only to permitted authorities.

The area of social tagging system portrays two organizational taxonomies flourished by analyzing and comparing design and feature of the system. System design and attributes greatly touch the nature and distribution of tags and the system collects the attributes of information. User incentives are the form of contribution will touch the characteristics [4]. Privacy protection in social tagging is an issue which is measured by Shannon entropy and model of their apparent user profile as PMF over categories of interest [26]. Privacy ensured skim with Support Vector Machine (SVM) for extraction of data and obscure delicate data with Privacy Enhancing Technology (PET), namely Tag suppression used to protect the privacy separate the normal data from sensitive data (see Fig. 1). The exploratory of privacy protection can assess the impact of PET (Privacy Preserving Technologies) namely Tag suppression. Collaborative Tagging is extended as effective keywords for fetching the link and authority can add keywords to get the suggestion from users for accessing the sensitive data.

The approach protects user privacy to a certain extent, by dropping those tags that make a user profile views toward certain category of interest. The keyword based on category can be skim by support vector machine where classified data are achieved. Along with classified data, duplicate data are also available. To evacuate the duplicate data in the classified data can be achieved by UDD. Unsupervised data are not assumed the labels where twin data are removed using the Unsupervised Duplicate Detection.

More precisely, architecture is built which consists of additional services. The user specifies their resources of interest based on query are processed before that authorization has been performed. Using collaborative tagging for the Query analysis model for ease of data retrieval and extended as effective keywords for fetching the URL link. The tag suppression that preserves the user privacy in the semantic web. Along with that Support Vector Machine is used to classify their data based on the keywords and Unsupervised Duplicate Detection is also implemented. The combination of these services allows broadening the functionality of collaborative tagging and concurrently provides user and organization with a mechanism to preserve their privacy while processing.

2. OVERVIEW OF THE PROPOSED APPORACH

A social Bookmarking service is centralized online services which facilitate users to add, annotate and share bookmarks of web document and support to collaborative tagging can be considered as valuable knowledge as online resources as concerned [3]. Collaborative tagging which it support tag-based resources search, despite the system can enhance the architecture with additional services address the issue available in the service. The tagging is used to protect the privacy by refraining from privacy Enhancing Technology to separate the normal data from the sensitive data. Only the permitted authorities can access the sensitive data can be achieved by collaborative tagging with some services as tag suppression used to suppress the data.

The architecture of the proposed system can be achieved by using the additional services are Support Vector Machine where used to classify their data based on the keywords. The keywords can be feedback, location or cost. In previous the location can be inclined manually enforce the simple mechanism and concerning. For this reason, an android mobile client is an application that access a service made available by a server. The server is often but neither invariably on another computer, in which case the client access the service by way of a network. The terms applied to devices could interact with remote computers via a network. Before sending the request user has to be registered in the server and the information stored in the database.

The user endorsed into the server, the user is requested option to search their data on Query based or location based that are displayed by using mobile GPS the location can be determined.

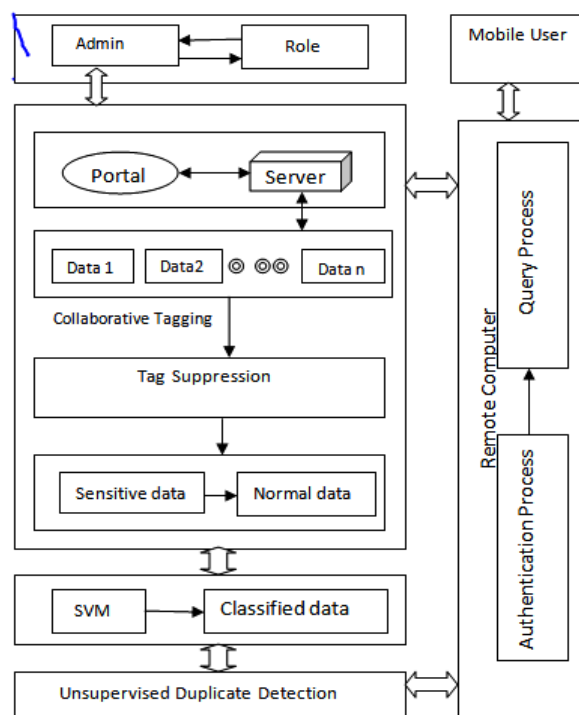


Fig 2 Architecture of the proposed enhanced tagging services

Using tag suppression the end user privacy can be protected and it's a suitable strategy for the enhancement of privacy in the scenario of collaborative tagging. Tag suppression is used to classify the sensitive data from the normal data. The Normal data can be viewed by the end users and the sensitive data about the organization can be viewed only by the permitted authorities. By using the SVM algorithm can provide feedback for each and every category like product accessories, display, volume, user friendly and so on. Based on the keywords classification are performed and ratings will be displayed. Classified data from the SVM carries duplex data that are evacuated by Unsupervised Duplicate Detection. Equivalent data can be construe as the ratio of number of duplicates generated to the number of record pairs taken from the data source and extract those data and display to the user.

2.1 Tag Suppression

Collaborative tagging requires the enforcement of mechanism that enable users to protect their privacy by allowing them to hide certain data without making that useless for the purposes they have been provided. Users tag resources on the web according to their personal preferences. Therefore contribute to describe and classify those resources is inevitably revealing their profile. In general the information can be trapped by aggressor; users may endorse a privacy Enhancing technology based on data perturbation is considered as tag suppression. It is a technique that has the purpose of preventing privacy assailant from profiling user's interest on the footing of the tags what user stated. Tag suppression as a suitable strategy for the enforcement of user privacy in the scenario of tagging because inferior on the impact of linguistic functionality.

2.2 Support Vector Machine

The user is to provide the feedback for the product or service. Here also using Support Vector Machine Algorithm to get the feedback. By using this algorithm can provide feedback for each and every category like Products Accessories, display, volume and user friendly etc and extract the keywords to get the feedback from the user. The keywords are like Good, Fair, Awesome, Magnificent, Bad and Poor etc.

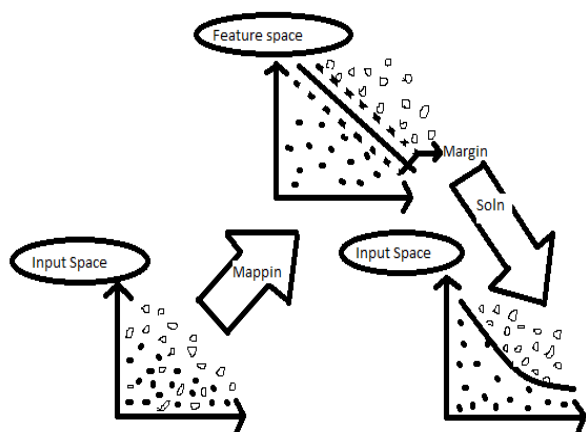


Fig 3 Support Vector Machine

Based on the keywords, feature space are performed and ratings where displayed to get the good product. The margin is used to rift the data based on the keywords are assigned as origin value and to efficient these can be build up by expanding the boundary space. A margin is a p -dimensional real vector and wants to find the maximum-margin hyper plane that divides the points having boundary.

The boundary value can be of two points are mapped as shortest distance to the closest positive point and shortest distance to the closest negative point. Any hyper plane can be written as the set of points are available in plane. The parameter determines the offset of the hyper plane from the origin along the normal vector. Based on the keywords, feature space are performed and ratings where displayed to get the good product.

2.3 Unsupervised Duplicate Detection

Duplicate detection is used interchangeably with record linkage. There have been many efforts in finding solution to the problem of identifying duplicates records. Most of the research for identifying duplicate records is based on predefined matching rules hand-coded by domain experts or matching rules learned offline by some learning. Such approaches work well in a traditional database environment, where all instances of the target databases can be readily accessed. In a Web database scenario, the records to match are highly query-dependent and they can only be obtained through online queries. When using a traditional database the data to be retrieved is known before hand which can be used to train and identify duplicates beforehand. This is in contrast to Web databases where results from multiple sources have to be combined with no pre-determined training data. Duplicate ratio can be defined as the ratio of number of duplicates generated to the number of record pairs taken from the data source.

3. RELATED WORKS

Collaborative tagging as a challenging research topic in early days is most popular services available in online. Assorted paper considered its specific characteristics, the similarity and difference with traditional annotation techniques along text labels. The tagging and bookmarking may contribute various incentives [2]. In prior that are contrasted peer-to-peer knowledge management with tagging approaches [4][18]. Measuring the relationships among tags or tagged resources is an active research area. It provides a model of semantic-social networks for extracting lightweight ontology from del.icio.us [1].

Then exploring the privacy paradox in the content of online social networking further understanding divulge private information in an online social network which are posed some research question about the common characteristics, motivating factors and control to protect their private information[5]. A second hypothesis (H2) is that individuals who communicate through virtual social networks feel they have control over their own private information and (H3) has the majority of individuals who communicate through virtual social networks will confirm that they did not read the privacy policy before becoming a member. The privacy gain in this work bears certain similarity with the concept of co privacy, brought in for a P2P in [6] [13][18]. Co privacy was defined as a situation where the best strategy for a peer to preserve privacy is to help another peer in preserving privacy. The advantage is that they make privacy preservation of each specific individual.

A user profile as a histogram of relative frequencies of tags across categories of interest and quantify the degree of privacy attained by modification as its Shannon entropy [6] without empirical evaluation. An issue concern with the trustworthiness has been focused on services. Social tagging services are supported virtually by any online application which increases the risk of cross referencing. The perturbative technique to the collaborative filtering algorithm based on singular value decomposition [21] technique might not be able to conserve privacy

4. CONCLUSIONS

Collaborative tagging system has the potential to improve traditional solution where the information available in online services which are extremely popular. Although it is basically used to support resource search, probable is still to be exploited. One of these potential is the provision of web access functionalities such as content filtering and hole up the delicate data. However it would be necessary to extend the architecture of the current collaborative tagging services so as to include a policy layer that supports the enforcement of user preferences. A collaborative tagging is one of the most diffused and gaining a popularity, it has become more evident the need for privacy protection.

Motivated by all this, contribution is an architecture that incorporates on support of enhanced and private collaborative tagging. More specifically the proposed architecture consists of two additional services built on it. The support vector machine and unsupervised duplicate detection to classify the data based on the feedback, location or cost. The location can be identifying by using GPS that are performed by connecting mobile user to the remote computer and extract the duplex from the classified data.

The combination of these two services allows broadening the functionality of collaborative tagging system and provides users with a mechanism to preserve the privacy while tagging can be achieved by using tag suppression showing its effectiveness in terms of privacy and analyze the data. Considered that what reported in this paper can be useful to evaluate the further future development in the field. Future work includes the development of a complete prototype and understanding of the constraints and affordance of tag based information systems.

REFERENCES

- [1]. B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and S. Gerd, "Evaluating Similarity Measures for Emergent Semantics of social Tagging," Proc. 18th Int'l Conf. World Wide Web(WWW),pp. 641 -650,2009.
- [2]. C. Marlow,M. Naaman,D. Boyd, and M. Davis, "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article to Read,"Proc.17th Conf. Hypertext and Hypermedia(HYPertext),pp. 31-40,2006.
- [3]. B.Carminati, E. Ferrari, and A.Perego, "Combining Social Networks and Semantic Web Technologies For Personalizing Web Access," Proc. Fourth Int'l Conf. Collaborative Computing:Networking,Applications and Worksharring,pp. 126-144,2008.
- [4]. R. Gross and A. Acquisti, "Information Revelation and Privacy in Online Social Networks," Proc. ACM Workshop Privacy Electronics Soc.(WPES),pp.71-80,2005
- [5]. S. B. Barnes, "A Privacy Paradox:Social Networking in the United States," First Monday, vol.11,no. 9,Sept.2006.
- [6]. J. Parra-Arnau, D. Rebollo- Monedero, and J. Frone, "A Privacy Preserving Architecture for the Semantic Web Based on Tag Suppression," Proc. Seventh Int'l Conf. Trust, Privacy, Security, Digital Business (TrustBus), pp. 58-68, Aug.2010.
- [7]. J. VoB, " Tagging Folksonomy & Con – Renaissance of Manual Indexing?" Computer Reaserch Repository, vol.. abs/cs/0701072, 2007.
- [8]. J. Castella `-Roca, A. Viejo, and J. Herrera-Joancomartí , "Preserving User's Privacy in Web Search Engines," Computer Comm., vol. 32, nos. 13/14, pp. 1541-1551, 2009.
- [9]. D. Rebollo-Monedero, J. Forne ´, and J. Domingo-Ferrer, "Copriate Query Profile Obfuscation by Means of Optimal Query Exchange between Users," IEEE Trans. Dependable and Secure Computing, vol. 9, no. 5, pp. 641-654, Sept.-Oct. 2012.
- [10]. C. Schmitz, A. Hotho, R. Jäschke, and G. Stumme. Mining association rules in folksonomies. In Data Science and Classification: Proc. of the 10th IFCS Conf., pages 261–270, Berlin, Heidelberg, 2006. Springer.
- [11]. P. Schmitz. Inducing ontology from Flickr tags. In WWW Collaborative Web Tagging Workshop, May 2006.
- [12]. G. Mishne. Autotag: a collaborative approach to automated tag assignment for weblog posts. In Prox. 15th Intl. Conf. on World Wide Web, pages 953–954. ACM Press, 2006.
- [13]. J. Parra-Arnau, D. Rebollo-Monedero, and J. Forne ´, "A Privacy- Protecting Architecture for Collaborative Filtering via Forgery and Suppression of Ratings," Proc. Int'l Workshop Data Privacy Manage- ment, Autonomous Spontaneous Security (DPM), pp. 42-57, Sept. 2011.
- [14]. Z. Yun and F. Boqin, "Tag-Based User Modeling Using Formal Concept Analysis," Proc. IEEE Eighth Int'l Conf. Computer Informa- tion Technology (CIT), pp. 485-490, 2008.
- [15]. A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke, "Persona- lized Recommendation in Social Tagging Systems Using Hier- archical Clustering," Proc. ACM Conf. Recommender Systems (RecSys), pp. 259-266, 2008.
- [16]. K. Bischoff, C.S. Firan, W. Nejdl, and R. Paiu, "Can All Tags Be Used for Search?" Proc. 17th ACM Conf. Information and Knowledge Management (CIKM), pp. 193-202, 2008.
- [17]. X. Li, C.G.M. Snoek, and M. Worring, "Learning Social Tag Relevance by Neighbor Voting," IEEE Trans. Multimedia, vol. 11, no. 7, pp. 1310-1322, Nov. 2009.
- [18]. S. Marti and H. Garcia-Molina, "Taxonomy of Trust: Categorizing P2P Reputation Systems," Computer Networks, vol. 50, pp. 472-484, Mar. 2006.
- [19]. L. Specia and E. Motta, "Integrating Folksonomies with the Semantic Web," Proc. Int'l Semantic Web Conf., pp. 624-639, 2007.
- [20]. M. Bundschuh, S. Yu, V. Tresp, A. Rettinger, M. Dejori, and H.-P. Kriegel, "Hierarchical Bayesian Models for Collaborative Tagging Systems," Proc. IEEE Int'l Conf. Data Mining (ICDM), pp. 728-733, 2009.
- [21]. H. Polat and W. Du, "SVD-Based Collaborative Filtering with Privacy," Proc. ACM Int'l Symp. Applied Computing (SASC), pp. 791-795, 2005.