International Journal of Computer Science and Business Informatics

IJCSBI.ORG

Onto Tagger: Ontology Focused Image Tagging System Incorporating Semantic Deviation Computing and Strategic Set Expansion

Gerard Deepak
Department of Computer Science and Engineering
University Visvesvaraya College of Engineering
Bangalore University, Bangalore, India.

Sheeba Priyadarshini J
Department of Computer Science
St. Josephs College (Autonomous)
Langford Road, Bangalore, India

ABSTRACT
Social Tagging of images uploaded to the Web is highly mandatory as tags serve as the entities for image retrieval. Manual Tagging of images makes the overall process tedious and moreover the tags when manually assigned become noisy. Several automatic tag recommendation systems are available but the background study proves that the tag relevance is not very high. In the era of Semantic Web, there is a need for a semantic driven tagger which would perform efficiently. Also, a system which bridges the gap between manual and automatic tag recommendation is required. An ontology driven semantic tagger for tagging images with social importance which tags the images based on limited reference tags is proposed. The proposed methodology combines ontology crawling using K-Means Clustering and Semantic Deviation Computation using Modified Normalized Google Distance Measured. The tag space is enhanced using Strategic Set Expansion incorporating a dynamic semantic deviation computation. An average precision percentage of 84.4 and an F-Measure percentage of 86.67 are achieved.

Keywords
Ontology Tagger, Semantic Deviation, Social Tagging, Strategic Set Expansion, Tag Recommendation.

1. INTRODUCTION
The World Wide Web is a repository of billions of images, video, text and other data. Owing to the increasing number of users of the World Wide Web, even the data on the Web is increasing exponentially. With the popularity of trends in Social Networking, the amount of social data is also

increasing almost every minute. It has been estimated that Social data mainly includes images and videos that are shared among the social actors (users and their friends) when compared to text and other kinds of data. Images are among the topmost in the list of social data that are the most shared.

The problem targeted is definitely not the increasing amount of social or web data or their means to handle it but retrieving the exact or highly similar items like images or video when searched is the main issue. Tags or Annotations become a very important means to retrieve web multimedia information. To facilitate easy searching or retrieval of exact images or items from the World Wide Web, Tagging of such social data items plays an important strategy as images without tags over the Web are ubiquitous[1].There can be two approaches which can be followed. The first approach is the traditional approach where the Search Engine or The Search Algorithm must be designed in a more efficient or intelligent manner to extract the required items. The second of the two approaches is by correctly tagging the social items like images and reducing the noise in such tags such that there is a quicker and faster convergence to retrieve the exact images or videos from the World Wide Web. However, there exists diversity in social tagging behaviors [2].

Tagging an image or a video or any such social items like blogs is like specifying the right address for the search engine to retrieve the tagged item. Tags bridge the gap between users' cognition and objective [3].Tagging that exists in the present day scenario varies from manual to automatic tag recommendation. Tagging an image on the social websites is a methodology of organizing the images and reliable tags must be specified for social photos [4]. Several automatic taggers are available but the best tag recommendation system must be semantically driven and must recommend them by re-organizing the tags such that the users may be convinced by the tags that are recommended by the search engine. Also the semantics of the tags assigned must correlate with that of the Web Search Engine such that the items that are browsed must have high relevance and must be significant.

Although, with the availability of several approaches for automatic tag recommendation, a semantic web approach which is actually budding for tagging will make it much efficient to comply with the web search engines. An ontological approach is proposed for tag construction and recommendation in order to minimize and remove noisy tags for images of community importance. Ontologies constitute important intrinsic structures of the Semantic Web. Due to a paradigm shift towards improved semantic web technologies, ontology focused approach for image tagging is a highly

commendable and feasible approach. The usage of ontologies towards tagging images not just enhances the relevance of the results but also increases the correctness and the quality of tags. The tags when included based on the ontologies, the probability of deviations and noise is minimized to a very large extent. Images with community importance are mainly concentrated because such images are searched more frequently by a large number of people and most of the times simultaneously. There are several social networking sites specifically for uploading of images like Flickr, Instagram, etc. Henceforth, a tagger that computes similarity of tags and automatically recommends by lowering tag noises is mandatory.

Motivation: The absence of a tagger which semantically tags images of community importance to without redundancy is the intrinsic motivation of developing such a system. A tagger which tags images by automatically recommending tags which similar images are also holding in order to reduce the tag redundancy. A tagging system which is semantically driven is in a high demand in the era of intelligent and semantic Web. The main motive of the proposed system is higher the quality of the tags better is the relevance of the images. Thus, there is a need for a system that automatically recommends tags which are just not straight forward but are of a very high quality by aggregating tags the tag space with relevant images. Most of the users who upload images feel that manual tagging is much better than automatic tagging. Thus, a gap is formed between the approach of manual tagging and automatic tagging that needs to be bridged. The Proposed approach fills this gap and is highly proficient when tagging images.

Contribution: A system that is highly efficient in tag recommendation which semantically computes for similar tags based on user driven reference tag is proposed. An ontological approach which uses tag level and image level ontologies for tagging is proposed. An innovative approach which incorporates K-Means Clustering for extracting similar images and eventually their tags is proposed for constructing the tag space. Furthermore, the ontologies in the tag space are re-ranked based on semantic similarity deviation computed using Modified Normalized Google Distance [5] [6] is implemented. Dynamic computation of semantic deviation is used for checking the tag level semantic similarity. Several Permutations of the tags from this tree are obtained and are recommended to the users for selection of tags.

Organization: The paper organization is as follows. The Section 2 provides a brief overview of related research work. Section 3 describes the Problem Definition. Section 4 presents the Proposed System Architecture. Section 5 discusses the implementation in brief. Results are presented in Section 6.

Performance is evaluated in Section 7.Finally, Section 8 concludes the paper.

2. RELATED WORKS

Zhou et al., [8] have proposed a hybrid collaborative model for image tagging which incorporates probabilistic techniques with a content based strategy for image tagging. The strategy uses an image tag associative matrix for and non-negative factorization of matrix for collaborative filtering. Correlations are estimated using a norm method. The approach is tested using three large databases with a large number of images and tags. Pantraki et al., [9] have proposed an image tagging and recommendation system which uses the strategy of Parallel Strategy Analysis 2 in which three matrices are used with visual features, user information and the tag information. This strategy was tested with ample data sets. The major drawback of this technique is that it uses three large data structures which may tend to increase the overall complexity of the solution.

Zhu et al., [10] have proposed an approach where the social information of the image is used as social clues for tagging of images. Along with social clues, the social group information and the tagging preferences of the image owner is used to predict and annotate tags. A strategy called as neighborhood voting on the Online Social Networks is used. This approach is quite good but the tag space doesn't predict numerous tags which becomes a disadvantage in this system. Qian et al., [11] have proposed a novel strategy of incorporating diversified semantics for image tagging and annotation over online social networks. The strategy uses considering the factor of relevance of improving the tag quality. The usage of semantic compensation onto the already determined top ranked tags is one of the major concepts on which this strategy is based.

Liu et al., [12] have proposed an innovative methodology of using more than a single for tagging of an image. This multi feature tagging approach learns several scores of features which are further converted to weights. A tag pair matrix of a predormintarily low rank is formulated such that several features are a significant in tagging of an image. Lin et al., [13] have proposed a strategy of image tagging on social websites by extracting sparse patterns of tags from tags that are already available in the web data. The tags which the users have contributed already are being used to predict and re-recommend tags in this approach. The problem of minimizing the quadratic loss is very well treated in this strategy by incorporating a bi-layer norm. Though there would be a good level of tag relevance, there will be a dependency on the already available tags rather than the image level compatibility.

Im et al., [14] have proposed a novel strategy where semantics is incorporated. The semantic relationship between a pair of tags is inferred in this approach through a semantic paradigm. The context of the word that is used to tag an image is strategically found out in this approach. The disadvantage is that the contexts might add a little amount of noise into the final predicted tags. Pliakos et al., [15] have proposed a unique methodology of image tagging as well as geological tag inference simultaneously. The underlying strategy is the usage of hypergraph incident matrix for geographical location prediction. A concept of group sparsity constraints enforcement is incorporated to this methodology to get better results.

Barmpoutis et al., [16] have proposed an approach in which the strategy of tensor structure decomposition of tags. This is purely a mathematical approach which aims at breaking the links between the multilink relations of the tag elements. This method involves the usage of four different unique matrices with four innovative mathematical strategies for tag decomposition. Pavlidis [17] has studied the interdependence of image tagging with that of the actual content of the image. This paper clearly points out that the complexity involved in the analysis of the contents of the image makes tagging and annotation of images also complicated. Furthermore, the paper proposed the non-pixel dependent strategies for image tagging rather than the conventional methodologies.

Huges et al., [18] have proposed an approach which incorporates machine learning techniques for computing and estimating the tag relevance of images that are tagged based on Geo-Spatial Information. The tags are classified using an SVM classifier and a semantic approach is used for computing the tag heterogeneity. This approach reduces the irrelevant tags and improves the overall tag quality and is applicable for geographically tagged images of community importance. The approach combines a statistical geographical distribution for spatial information estimation and Machine Learning Technique like SVM for tag classification. Wang et al., [19] has proposed a strategy which bridges the gap between the features extracted and the overall context. An approach of context regularization is implemented for the visual features which makes this strategy an effective one for image tagging over social networks.

Gong et al., [20] have proposed a robust model of image annotation using the strategic computation of semantic similarity between words. A language specific model is designed for estimating the semantic similarity in this approach. A cross media relevance model and a translation model is integrated with this strategy in order to enhance the relevance of the results in this approach. Sawant et al., [21] have proposed an innovative approach

of social image tagging based on social inputs. Online Social Websites are a repository for very large volumes of social data like the features and patterns of users, images and social tags. This approach mainly concentrates on the semantic analysis of these social inputs to formulate and predict tags for annotation.

3. PROBLEM DEFINITION

Tagging of images is of most importance as the tag names play a vital role when the images are retrieved in the search engine. The major objectives of our work which were tackled as individual problems are:

- To automatically recommend tags for an image uploaded driven by a limited input tag of users' reference.
- To implement an ontological approach for tag recommendation.
- To overcome and avoid noisy and redundant tags.
- To improve the overall quality of the tags for community important images.
- To improve the recall, precision, f-measure and accuracy of the proposed system.

4. PROPOSED SYSTEM ARCHITECTURE

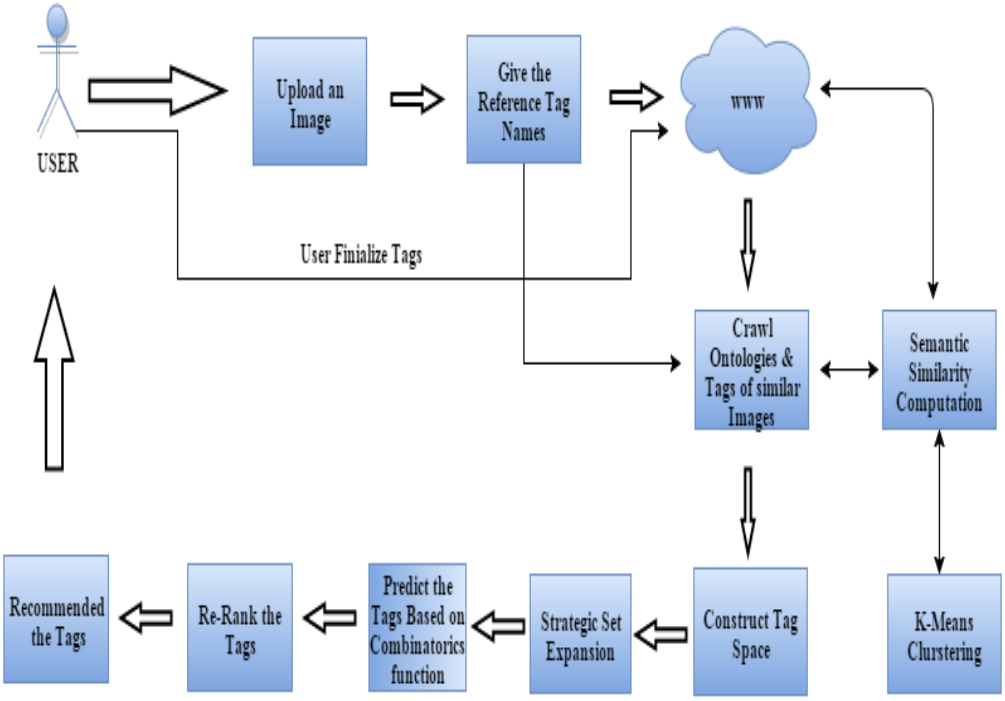

Figure 1. Proposed System Architecture

The overall architecture of the proposed system is depicted in Figure 1 where in a user initially uploads an image on the web through a social networking site or directly through the Application Programming Interface (API) of the system created. The API of the proposed system has its own Graphical User Interface (GUI) and the user can directly interact through the systems API with the World Wide Web. The system developed also has a facility to integrate its GUI with that of social networking sites like Instagram or Flicker where images that are of community interest are uploaded. The only restriction for integrating the systems' API with that of social networks is that the user must have an account in the social networking site to which the user wants to post his image through the proposed system.

As the user is uploading the image, it is a mandate that the user also specifies ten reference tag names which the user wishes to tag. The reference image tags are parsed and tokenized to remove redundancy and extract individual tag elements. The individual tag elements are further used as input query to cluster similar images. Furthermore, the tags of the resultant images are extracted and semantic similarity is computed to that of the reference tags. If semantic similarity is high, the tags are included in the tag space. The tag space is generally a vector which is dynamically constructed where in the tags which are highly similar are organized and can be used in several instances.

To enhance the relevance of the tags and to give many options, the ontologies similar to the reference tags are crawled using a focused crawler and once again the semantic similarity is computed with that of the reference tags and is included into the tag space. The focused crawler is integrated into the system which automatically crawls the ontologies of similar tags based on the input reference tags. Based on the weights of semantic similarity index, re-ranking of tags is done. Furthermore, using a combinatorial function, the tags are predicted to the user. The user can finalize the tags that can be used. The proposed system enhances the tag relevance and minimizes redundant and noisy tags. This ensures a very high tag quality. Moreover, since ontology is also considered as a parameter for tag formulation, the relevance of the tags is very high.

5. IMPLEMENTATION

The implementation of the proposed system is done using JAVA as a programming language and Netbeans as the IDE. The reference tag names are initially parsed and tokenized. Tokenization is performed by incorporating a JAVA based NLTK (Natural Language Tool Kit) Tokenizer

to split the individual tag elements. Also a redundancy checker which is custom written in JAVA is used for removing redundant tag items. Redundancy of tag elements is checked to minimize the complexity of the system and avoid unnecessary checks for duplicate items. This definitely enhances the performance of the system implemented.

The reference tag elements are used to cluster similar images using a standard nearest neighbor approach incorporating a basic K-Means Clustering. Furthermore, the semantic similarity computation using Modified Google Distance Measure is carried out. A Focused Ontology Crawler is integrated into the system which facilitates the phenomenon of extracting ontologies nearest to the reference tag items. As the ontologies are crawled, a fitness function incorporating Modified Normalized Google Distance which dynamically computes semantic similarity of the ontologies and checks with that of the reference tag items is also imbibed into the system. The motive behind this fitness function is to retain the ontologies which are correlative to the reference tags and enhance the overall relevance of the tags and accelerate the quality of tags. Also, the ontology information of the image uploaded is fed into the tag space for tag recommendation.

A tag space is formulated which is a vector containing all the semantically similar tags and as well as the best fit ontologies. The tag space is further enhanced by applying the strategic set expansion technique. Further, the re-ranking of the tags is done based on semantic weights of the tags. The combinatorial function is subject to the tags in order to predict the tags for finalization. The combinatorial function implemented is based on a neural network model with initial training such that the tags are predicted as per the user's choice. Parameters like the size of the tags, number of words needed to the tag, etc., serve as dynamic inputs to the adaptive artificial neural network for prediction.

Strategic Set Expansion refers to a set expansion paradigm where deviations are computed among the existing sets in the Tag Space. The standard threshold for the deviation is considered as average of the sum of initial deviations in the tag space before the set expansion is applied. During set expansion, the key procedure which takes place is when a new term is included; the semantic weight of the term is computed and is checked with the threshold deviation. If it's lesser than the threshold, then it is included in the set. Otherwise it is discarded if it exceeds the threshold. The immediate neighborhood with an order of maximum two hierarchies of the tags in the tag space is alone considered for set expansion. If TS be the current Tag Space, then the elements of the tag space are listed in (1). Deviation between a pair of tags in the tag space is computed and the Deviation set is depicted by (2).

$TS = \{ts1, ts2, ts3, ts4.....\}$ (1)

$D = \{d1, d2, d3, d4....\}$ (2)

Where $di = Semantic\ Difference\ (di, di+1)$, such that $i = \{1, 2, 3....\}$.

If Td is the threshold of the deviation, then Td = D/n where n is the number of elements in the tag space. In order to compute the Semantic Similarity Deviation or the Semantic Heterogeneity, a Modified Normalized Google Distance is considered which is given by the (3).x and y in (3) signifies the terms whose semantic heterogeneity has to be determined. The Google Distance Measure is modified owing to a reason that it yields better and more precise results increasing the overall confidence of the proposed approach.

$$ModfNGD(x, y) = \frac{\max\{\log f(x), \log f(y)\} - \log f(x, y) + \log(x - y) - \log(x + y)}{logN + log(x + y) - logx - logy - min\{log\ f(x), logf(y)\}} \quad (3)$$

5.1 EXPERIMENTATION

The experiment was carried out by using 1124 images which were of community importance. 350 images were collected from personal photographers while the rest were crawled from Bing and Google Image search engines. Several monuments which are popular and the ones which are not very popular are also included for experimentation. The amalgamation of less popular monuments with that of the popular ones actually makes the system ready for every kind photographs of community importance.

5.2 ALGORITHM

The Proposed Algorithm which inputs a reference tag pair and predicts tags is discussed in Table 1.

Table 1: The Proposed Tag Recommendation Algorithm

Begin
Step 1: Initialize the reference tag pair (rs1, rs2) for a specific image that is to be uploaded. The reference tags are incorporated as a discrete tag pair set RS. Ensure at least 1 tag pair is input for an image.
Step 2: The reference tag set RS is sent into a parser and then a Tokenizer to extract elements, remove redundancy and remove stem words and further a new set RS_{fin} is formulated.

Step 3: If RS_{fin} is non-empty, For each element do

a) Cluster Semantically Similar Images

b) Extract the individual tags from these Images and dynamically compute the semantic similarity of the tags using Modified Normalized Google Distance Measure. The tags are tokenized and stemmed to remove redundancies.

c) Construct a Tag Space vector with the new tags TS and apply strategic set expansion technique

Step 4: Crawl the essential ontologies O1, O2…. On corresponding to the reference tags and Check for assertions in the ontologies.

Step 5: While the ontologies are not assertions, compute the semantic similarity of the ontologies and check with that of the reference tags. Add them into the tag space TS.

Step 6: Using strategic set expansion, expand the Tag Space by Dynamically Computing the Semantic Heterogeneity.

Step 7: Re-rank the tags based on the Semantic Measure.

Step 8: Using Combinatorial Function, Predict and Recommend the tags to the user.

End

6. RESULTS

The tagging for several images that are uploaded are governed by the system implemented. Several images that are of community importance and are of interest to others are incorporated into the experiments. Various images that were crawled from the Google and Bing image search engines and were further uploaded through the proposed system for tagging. Table 2 depicts the images and their corresponding tags which were recommended by the system. The tags recommended were checked for correctness and relevance. The tags recommended by the system must be acceptable by the user and the user must finalize the tags based on the systems recommendation. The proposed system predicts and recommends tags of a high quality which will be evaluated in the performance evaluation

Table 2: Relevant Tags Recommended for images of social relevance

Images Uploaded	Reference Tags	Tags Recommended
	Amber Fort	Amber Fort, Jaipur Amber Fort, Jaipuri Amber Fort, Amer Fort, Amber Amer Fort, Amer Palace, Amer Fort Palace, Amber Palace Fort, Amber Royal Fort, Royal Amer Palace, Rajasthani Amber Fort, Rajasthan Fort Amber, Amber Lake Fort.
	Bom Jesus Church, Goa	Basilica of Bom Jesus, Parish Church of Bom Jesus, Borea Jezuchi Bajilika, Basílica do Bom Jesus,Francis Xavier Church, Goa Xavier Church, Goa Bom Jesus Church, Relic of St. Francis Xavier Church, Basilica of the Relic of Francis Xavier, Old Goa Church, Holy Jesus Church, Good Jesus Basilica.
	Philomena Church, Mysore.	Saint Philomena Church, Philomena,Church Mysore, Mysore Church, St. Josephs Cathedral Mysore, Mysore Philomena Josephs Church, Mysore Ashoka Road Parish Church.

	Velankanni Church	Our Lady of Good Health Basilica, Mother Mary's Shrine Velankanni, Velankanni Basilica, Annai Velankanni Aalayam, Arokiamarie Basilica Velankanni, Arokiya Matha Church.
	Fateh Prakash Palace	Fateh Palace, Prakash Palace, Royal Lake Palace, Lake Palace Uadipur, Prakash Lake Palace, Udaipur City Palace Complex, Lake Palace Complex, Pichola Lake Palace, Pichola Prakash Palace, Pichola Fateh Palace.
	Brihadeshwara Temple Tanjavur	Thanjai Periya Koyil, Thanjavur Periya Kovil, Brihadeshwara Temple Tanjore, Thanjavur, Brihadeeswarar Temple Thanjavoor, Thanjai Kalvettu Koil, Peruvudaiyar Kovil, RajaRajeswara Temple, Rajarajeswaram.

7. PERFORMANCE EVALUATION

The performance analysis of Onto Tagger is carried out by using four evaluation metrics namely the Recall, Precision, F-measure and Accuracy which is depicted in Equations (4), (5), (6) and (7) respectively. The percentage of the metrics used is considered in the approach proposed. The Precision or Positive Predictive Value referred to as a Fraction of Retrieved Instances [22] and is incorporated as the ratio of the number of relevant recommended tags to that of the tags formulated in the Tag Space. Recall also called as sensitivity is known as the fraction of relevant instances [22]

and is incorporated as the ratio of relevant recommended tags to that of the total number of relevant tags. F-Measure or F-Score relates Recall and Precision and is known as the harmonic mean [22] of Recall and Precision. Accuracy is formulated as the average of recall and precision.

$$Precision = \frac{Number\ of\ Relevant\ Tags\ Recommended}{Total\ number\ of\ Tags\ Formulated\ in\ Tag\ Space} \quad (4)$$

$$Recall = \frac{Number\ of\ relevant\ Tags\ Recommended}{Total\ number\ of\ Relevant\ Tags} \quad (5)$$

$$F\text{-}Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (6)$$

$$Accuracy = \frac{Precision + Recall}{2} \quad (7)$$

The images which were considered for performance evaluation are Amber Fort, Bom Jesus Church Goa, Philomenas Church Mysore, Velankanni Church and Qutab Minar. The images chosen range from the most popular to the less famous images. All images chosen have an importance to a specific community or a society in general. The percentage of Recall, Precision, and Accuracy for the images is individually depicted in Table 2. Also, the average measures of the depicted evaluation metrics is computed which is also lists in Table 2.

Table 2 : Performance Measurement of Onto Tagger

Image	**Recall %**	**Precision %**	**Accuracy %**
Amber Fort	88	83	85.5
Bom Jesus Chrurch,Goa	93	88	90.5
Philomenas Church, Mysore	86	75	80.5
Velankanni Church	92	86	89
Qutab Minar	87	90	88.5
Average	**89.2**	**84.4**	**86.8**

Onto Tagger is compared with Tag Relevance [7] which is used as a bench mark for performance comparison. The Precision Percentage of Onto Tagger was compared with that of Tag Relevance. Five arbitrary Precision values of Tag Relevance were considered which was then converted into a percentage of precision for comparison with the Precision Percentage of Onto Tagger. Figure 2 gives the graphical depiction of the Precision Percentage of Onto Tagger and Tag relevance. It is clearly evident from the line graph that the Performance of Onto Tagger is much better than that of Tag Relevance.

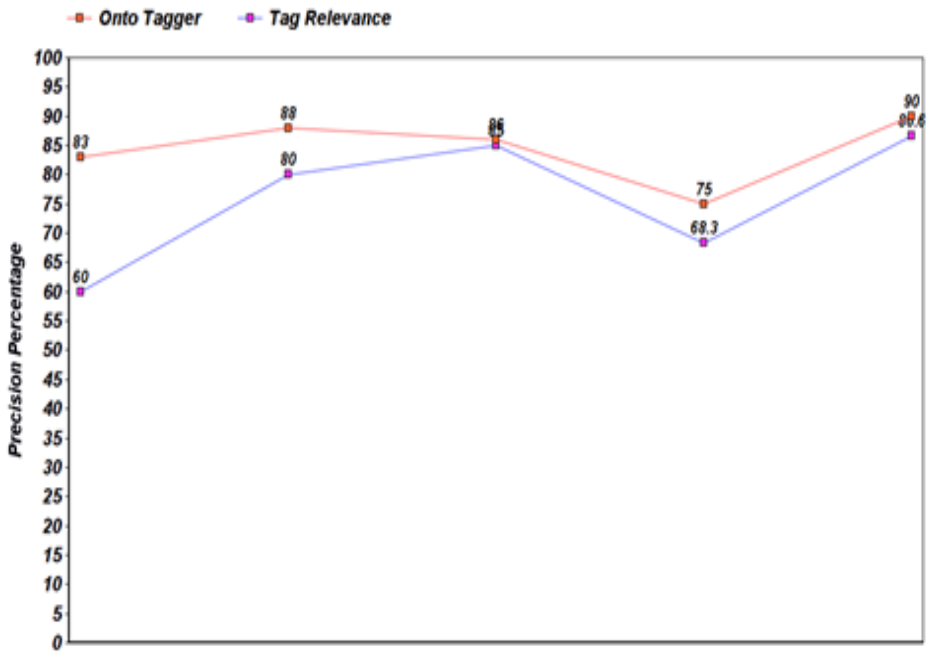

Figure 2: Comparison of the Precision Percentage of Onto Tagger with Tag Relevance

The graphical depiction of the F-Measure of the individual image elements considered for performance evaluation is depicted in Figure 3. The F-Measure of Amber Fort and Bom Jesus Church Goa is 85.43 % and 90.43 % respectively.Philomenas Church, Velankanni Church and Qutab Minar have F-Measure Value of 80.12 %, 88.89 % and 88.47 % respectively. The Comparison of the Average Precision Percentage of Onto Tagger and Tag Relevance is depicted in Figure 4 as a bar chart. It is clearly evident that Onto Tagger outperforms Tag Relevance by 1.9 %. From this we can infer that Onto Tagger is 1.9 percent more precise than Tag Relevance.

T

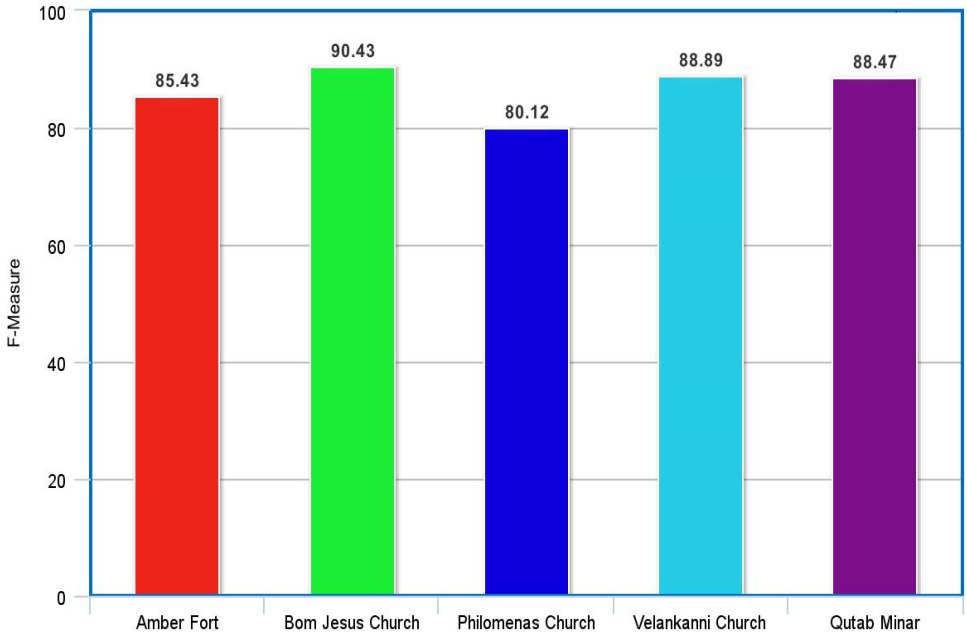

Figure 3: F-Measure in Percentage for various images

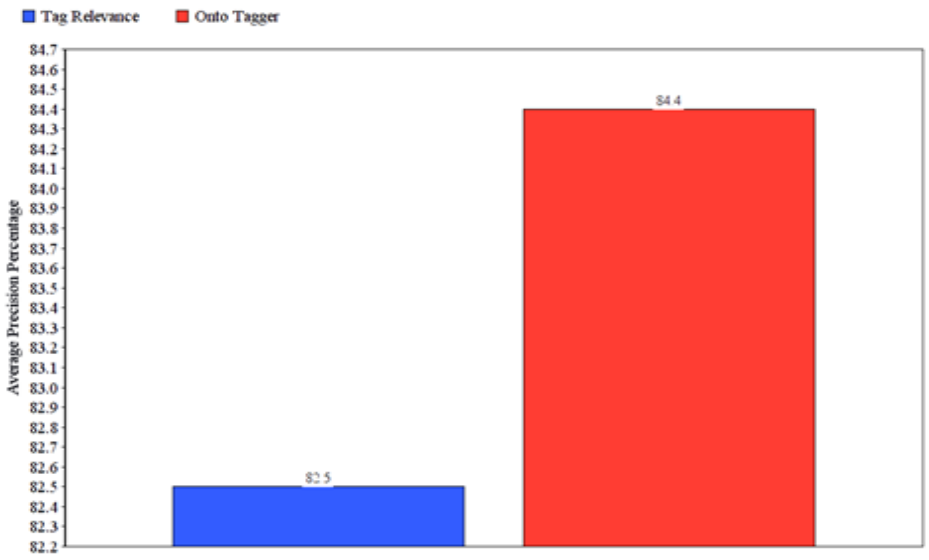

Figure 4 : Comparison of Average Percentage of Precision of Onto Tagger with that of Tag Relevance

8. CONCLUSIONS

An innovative approach for automatic tag recommendation which is based on limited reference tags is implemented successfully. The paradigm is driven by ontology extraction and incorporates a Modified Normalized Google Distance measure to compute semantic deviations between ontology entities and tags. A strategic set expansion strategy is incorporated to enhance the population of the tag space. The proposed methodology clusters similar ontologies using a basic K-Means Clustering Algorithm. Several community contributed images that are crawled from Flicker and Google Image Search are used for validating the approach. Experimental evidences show that the Proposed onto Tagger approach is the best-in-class method for tagging social images on the Web. Onto Tagger yields an average precision of 84.4 %, an average recall of 89.2 %, accuracy of 86.8 %. The average proposed F-measure of Onto Tagger is 86.67 %. Onto Tagger incorporates a strategic set expansion technique which increases the overall number of tags recommended. Onto Tagger extracts the ontology information of the similar tags. Owing to the reason of incorporating ontological information, the tags are of high relevance. Dynamic Semantic Deviation Computation that has been used in Onto Tagger eliminates the noisy and irrelevant tags. Onto Tagger is one of the image tag recommendation systems which is driven by the reference input tags which makes it a cognitive bridge between manual tagging and automating tagging systems.

ACKNOWLEDGMENTS

I whole heartedly thank my parents who stood by me and supported me during the implementation of this work. Above all, I thank God the Almighty and Everlasting Father and my Lord Jesus Christ who gave me strength, wisdom and health for the completion of this work.

REFERENCES

[1] Deng, Z.H., Yu, H. and Yang, Y., 2015, October. Image Tagging via Cross-Modal Semantic Mapping. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference* (pp. 1143-1146). ACM.

[2] Liu, Q. and Li, Z., 2016. Projective nonnegative matrix factorization for social image retrieval. *Neurocomputing*, *172*, pp.19-26.

[3] Jabeen, F., Khusro, S., Majid, A. and Rauf, A., 2016. Semantics discovery in social tagging systems: A review. *Multimedia Tools and Applications*,*75*(1), pp.573-605.

[4] Hong, Y., Chen, T., Zhang, K. and Sun, L., 2016, January. Personalized Annotation for Mobile Photos Based on User's Social Circle. In *MultiMedia Modeling* (pp. 76-87). Springer International Publishing.

[5] Wikipedia article *https://en.wikipedia.org/wiki/Normalized_Google_distanc*e last accessed on Feb 29th, 2016.

[6] R.L. Cilibrasi and P.M.B. Vitanyi, *The Google similarity distance* , IEEE Trans. Knowledge and Data Engineering, 19:3(2007), 370–383 or http://arxiv.org/abs/cs.CL/0412098

[7] Mishra, D., Singh, U.P. and Richhariya, V., 2014. Tag Relevance for Social Image Retrieval in Accordance with Neighbor Voting Algorithm. *IJCSNS*, *14*(7), p.50.

[8] Zhou, N., Cheung, W.K., Qiu, G. and Xue, X., 2011. A hybrid probabilistic model for unified collaborative and content-based image tagging. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *33*(7), pp.1281-1294.

[9] Pantraki, E. and Kotropoulos, C., 2015, September. Automatic image tagging and recommendation via PARAFAC2. In *Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on* (pp. 1-6). IEEE

[10] Zhu, S., Aloufi, S. and El Saddik, A., 2015, June. Utilizing image social clues for automated image tagging. In *Multimedia and Expo (ICME), 2015 IEEE International Conference on* (pp. 1-6). IEEE.

[11] Qian, X., Hua, X.S., Tang, Y.Y. and Mei, T., 2014. Social image tagging with diverse semantics. *Cybernetics, IEEE Transactions on*, *44*(12), pp.2493-2508.

[12] Liu, X., Liu, R., Cao, Q. and Li, F., 2013, November. Multi-feature Late Fusion for Image Tagging. In *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on* (pp. 34-37). IEEE.

[13] Lin, J., Yuan, J., Duan, L.Y., Luo, S. and Gao, W., 2012, July. Social Image Tagging by Mining Sparse Tag Patterns from Auxiliary Data. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on* (pp. 7-12). IEEE.

[14] Im, D.H. and Park, G.D., 2013, June. STAG: semantic image annotation using relationships between tags. In *Information Science and Applications (ICISA), 2013 International Conference on* (pp. 1-2). IEEE.

[15] Pliakos, K. and Kotropoulos, C., 2014, May. Simultaneous image tagging and geo-location prediction within hypergraph ranking framework. In*Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 6894-6898). IEEE.

[16] Barmpoutis, P., Kotropoulos, C. and Pliakos, K., 2015, September. Image tag recommendation based on novel tensor structures and their decompositions. In *Image and Signal Processing and Analysis (ISPA), 2015 9th International Symposium on* (pp. 7-12). IEEE.

[17] Pavlidis, T., 2009, June. Why meaningful automatic tagging of images is very hard. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on* (pp. 1432-1435). IEEE.

[18] Hughes, M., Connor, N.E. and Jones, G.J., 2012, May. A machine learning approach to determining tag relevance in geotagged Flickr imagery. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2012 13th International Workshop on* (pp. 1-4). IEEE.

[19] Wang, L., Zhao, Z., Su, F. and Sun, W., 2014, July. Content-based social image retrieval with context regularization. In *Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on* (pp. 1-6). IEEE.

[20] Gong, T., Li, S. and Tan, C.L., 2010, October. A semantic similarity language model to improve automatic image annotation. In *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on* (Vol. 1, pp. 197-203). IEEE.

[21] Sawant, N., Li, J. and Wang, J.Z., 2011. Automatic image semantic interpretation using social action and tagging data. *Multimedia Tools and Applications*, *51*(1), pp.213-246.

[22] Wikipedia article *https://en.wikipedia.org/wiki/Precision_and_recall* as accessed on Feb 29th, 2016.

This paper may be cited as:

Gerard, D. and Sheeba, P., J., 2016. Onto Tagger: Ontology Focused Image Tagging System Incorporating Semantic Deviation Computing and Strategic Set Expansion. International Journal of Computer Science and Business Informatics, Vol. 16, No. 1, pp. 44-61.