© 2016 IJSRSET | Volume 2 | Issue 2 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Sharing of Images in Content Sharing Sites Based on User Profile Inferences

S. Pavithra, K. Saranya
P. A. College of Engineering and Technology, Coimbatore, Tamil Nadu, India

ABSTRACT

Usage of social media are increased considerably in today world it enables the user to share images with one another. Sharing the images may leads to performance violation. Web mining is use of data mining technique to discover and extract the information from the web. Web content mining is the extraction and integration of data, information and knowledge from web page. In the web, one can mine the images and find association between various images. Privacy techniques needed to adapt in order to improve the satisfaction level of user, by means of automated privacy policy generation. Adaptive Privacy Policy Prediction system helps the user to compose customized privacy settings. A Two level framework is proposed with Speeded Up Robust Feature for identifying the feature points of an images, by demonstrating in MATLAB tool. The region of selected points is an effective method for identifying the extracted features in the images based on the image feature extraction.

Keywords: Adaptive Privacy Policy Prediction, Scale Invariant Feature Transform, Speeded Up Robust Feature.

I. INTRODUCTION

One of the key enablers of users' connectivity is images. Sharing May takes place between previously established groups of known people or social circles like Google+, Flickr or Picasa. However, semantically rich images may reveal content sensitive information. Consider a photo of a 2014 college annual day celebration, for example. It could be shared within a Facebook or Flickr group, but may unnecessarily expose the student's family members and other friends. Sharing the images within online social media, may lead to unwanted disclosure and privacy violations [2],[15]. The nature of social media makes it possible for other users to collect profile information about the owner of the published content [3].

The aggregated information can result in unexpected exposure of one's location and may leads to abuse of one's personal information. Users are allowed to enter their privacy preferences in most websites. Unfortunately, users struggle to maintain such a privacy settings that was shown in recent studies [1], [12]. Therefore, we go for policy recommendation systems which can help the users to easily and properly configure privacy settings [7], [13], [19], [11]. Existing proposals for automating privacy settings appear to be insufficient to address the privacy needs of images [10], [5],[14] due to the amount of information implicitly carried within images.

II. METHODS AND MATERIAL

A3P Core

A3P System is the combination of A3P core and A3P social shown in Figure 1. A3P core consist of two major components namely Image classification and Adaptive policy prediction. For each user particular images are first classified based on content and metadata. The privacy policies of each category of images are analyzed for the policy prediction. To mine both image features and policies together a two stage approach is more suitable for policy recommendation than applying the common one stage data mining approaches [17]. The user is waiting for a recommended policy if he uploads a new image.

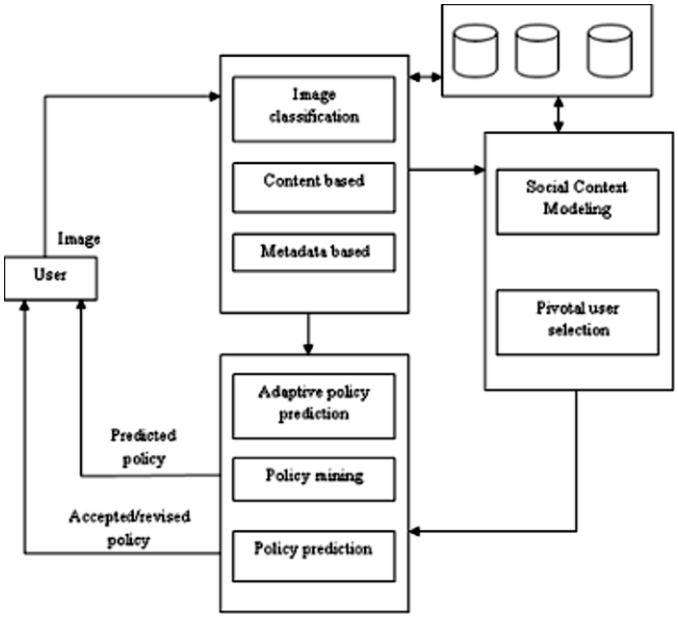

Figure 1. A3P Framework System Architecture

To find the candidate sets of images for the subsequent policy recommendation the two stage approach allows the system to employ the first stage to classify the new image. The one stage mining approach cannot be able to locate the right class of the new image because its classification criteria need both image features and policies of the new image are not available [16].

Single classifier can lead a system very dependent to the specific syntax of the policy by combining both image features and policies. The whole learning model would need to change if a change in the supported policies were to be introduced.

A. Image Classification

Image classification classifies images first based on the contents and then refines each category into subcategories based on the metadata. Images that do not have metadata will be grouped only by content. It gives a higher priority to image content and minimizes the influence of missing tags. It is possible that some images are included in multiple categories as long as it contains the typical content features or metadata of those categories. Content based classification is based on an efficient and accurate image similarity approach. Based on quantified and sanitized version of Haar wavelet transformation classification algorithm compares image signatures defined [4]. The wavelet transform encodes frequency and spatial information related to image color, size, invariant transform, shape, texture, symmetry for each image. A small number of coefficients are selected to form the signature of the image. The content similarity through images is then determined by the distance between the image signatures. The selected similarity criteria include texture, symmetry, shape and SIFT. Set the system to start from five generic image classes: (a) adults, (b) kids, (c) scenery, (d) animals. The preprocessing step is to populate the five baseline classes by manually assigning to each class a number of images crawled from Google images, resulting in about 1,000 images per class. Large image data set beforehand reduces the chance of misclassification.

Then signatures of all the images are generated and stored in the database. Upon adjusting the settings of content classifier, conducted some preliminary test to evaluate its accuracy.

The accuracy of the classifier is verified and discussed the use of the context of A3P core. User uploads an image and it is handled as an input query image [6]. A signature of images in the current image database is compared with the signature of the newly uploaded image. To determine the class of the uploaded image, first find its *m* closest matches. The class of the uploaded image is then calculated as the class to the majority of m images belongs. If no predominant class is found, a new class is created for the image. The image will be inserted into the corresponding image category in the image database, to help refine future policy prediction if the predicted policy for this new image turns out correct,. In the prototype, *m* is set to 25 that are obtained using a small training data set.

The metadata based classification groups the images into subcategories under aforementioned baseline categories. Metadata classification has three main steps. The first step is to extract keywords from the metadata associated with an image. The metadata considered in the work are tags, captions and comments. In this step all the nouns, verbs and adjectives in the metadata are identified and stored as metadata vectors. The second step is to derive a representative hypernym from each metadata vector. First retrieve the hypernym for each metadata vector based on the Wordnet classification and list a hypernym. A metadata vector t is cousin, first steps and baby boy. The cousin and baby boy have the same hypernym kid and first steps have a hypernym initiative. The hypernym list h are (kid, 2) and (initiative, 1). Select the

hypernym with the highest frequency as a representative hypernym. If there are more than one hypernyms with the same frequency the most relevant baseline class to be the representative hypernym. If a hypernym list h are (kid, 2), (cousin, 2) and (initiative, 1), select kid to be the representative hypernym and it is closest to the baseline class kids. The third step is to find a subcategory of an image belongs to. This is an incremental procedure. The first image forms a subcategory as itself and the representative hypernyms of the image becomes the subcategories representative hypernyms. Then, compute the distance between representative hypernyms of a new incoming image and each existing subcategory.

B. Adaptive Policy Prediction

For newly uploaded image the policy prediction algorithm provides a predicted policy to the user for particular reference. For users privacy concerns the predicted policy will reflect the possible changes. [18].

The policy normalization is a simple decomposition process to convert a user policy into a set of atomic rules and the data component is a single element set. Policy mining is carried out within the same category of the new image because images in the same category are under the similar level of privacy protection. For an image, a user decides appropriate group to access the image and thinks about the specific access rights should be given and finally refine the access conditions like setting the expiration date. The hierarchical mining first look for popular subjects defined by the user, and actions in the policies containing the popular subjects and finally for popular conditions in the policies containing both popular subjects and conditions.

The policy mining phase may generate several candidate policies and the goal of the system is to return the most promising one to the user. An approach is chosen to the best candidate policy that follows the users privacy tendency. To model the users privacy tendency, define a notion of strictness level. The strictness of a policy is described by the strictness level. In particular, a strictness level L is an integer with minimum value in zero and the lower the value, the higher the strictness level. It is generated by two metrics: major level (denoted as l) and coverage rate (α), l is determined by the combination of subject and action in a policy and α is determined by the system using the condition component. l is obtained by all combinations of common subject and common actions that are enumerated and assigned an integer value according to the strictness of the corresponding subjects and actions. In this view action is considered more restricted than tag action. Given a policy, its l value can be looked up by matching its subject and action.

If the policies have multiple subjects or actions and results in multiple l values, consider the lowest one. The computation of the coverage rate α is designed to provide fine grained strictness level. α is a value ranging from 0 to 1 and is adjusted but still not obtain the previously major level. In particular, define α as the percentage of people in the specified subject category that satisfy the condition in the policy. If users have five family members documented in the system and two are kids. And specifies a policy with the condition age > 18, only three family members will satisfy this condition. The corresponding α is then 3/5 = 0.6. The larger the value of α, the more people are allowed to access the image and the policy is less restricted is given in Equation 1.

$$L = l - (1 - \alpha) \qquad (1)$$

C. A3P Social

The A3P social provides mechanism that generates representative policies by strengthening the information related to the users social attributes and general attitude about once privacy [20]. A3P social will be invoked by the A3P core in two scenarios. If the user is a newbie to a site and does not have enough information stored for the A3P core to infer meaningful and customized policies. The system notices significant changes of privacy trend in the users social circle, it may be of interest for the user to possibly adjust the privacy settings accordingly. Users with similar social attributes may tend to have similar privacy inferences and also availability of collected data. This observation inspires to develop a social context modeling algorithm that can capture the common social elements of users and identify communities formed by the users with similar privacy concerns. The recommendation system acts as a base when it has rich set of images. The social context modeling algorithm consists of two major steps. The first step is to collect and provide once profile

information which is informative about their privacy settings. The second step form a group of users based on the identified social attributes. First, model each users social context as a list of attributes as sc1, sc2; . . . ; scn and in this sci denote a social context attribute and n is the total number of collected attributes in the social media. Social context attributes are extracted from users profiles is given in Equation 2.

$$\text{Conn:} \quad \frac{N^{u}_{R1}}{\sum^{n}_{i=1} N^{u}_{Ri}}, \ldots, \frac{N^{u}_{Rn}}{\sum^{n}_{i=1} N^{u}_{Rn}} \qquad (2)$$

There are four types of relationships being used by users in the system: R1 is family, R2 is colleague, R3 is friend and R4 is others. If Bob have 20 contacts with 10 family members, five colleagues and five friends. Social connection is represented as 10/20; 5/20; 5/20; 0/20. The number of social context attributes may grow with more rich information is collected by social networking sites in the future and the algorithm has a capability of dealing with many numbers social attributes being considered. The second step is to identify groups of user who have similar social background and privacy settings. In social context, it rarely happens that users share the same values of all social context attributes. In some cases a group of users have common values for a subset of social context attributes. Such subset information can be different for different groups of users and it makes grouping of users a challenging task.

If there are five users u1, u2, ..., u5 in a social networking site. Each one is associated with five social context attributes: gender, hobbies, occupation, location and social connection.

u1: [Female, movie, accountant, NY, {0.6, 0.1, 0.2, 0.1}]
u2: [Female, movie, teacher, IL, {0.7, 0.1, 0.1, 0.1}]
u3: [Male, ski, student, CO, {0.3, 0.1, 0.5, 0.1}]
u4: [Male, ski, student, KS, {0.6, 0.15, 0.15, 0.1}]
u5: [Male, ski, student, MO, {0.2, 0.1, 0.6, 0.1}]

From the users profile and social connection, two natural social groups can be formed. G1 = {u1; u2} – female who love movies and share data frequently with family members.

G2 = {u3; u4; u5} - all male students love sports.

First social group is formed based on social attributes: gender, hobbies and social connection and the second social group is formed based on a different set of attributes gender, hobbies and occupation. Dynamic social groups, employ an a priori based data mining algorithm. The original a priori this requires exact matches of items in different transactions. This is important for matching the social connection attribute just slightly different in the same social group. Define the matching of social connection attribute as the values of this attribute within a small threshold.

The collected social groups have not taken into account for privacy preferences. It is possible for users within the same social group maintain various privacy preferences. In order to join social groups to privacy preferences, again divide the social groups into sub groups according to the closeness of the privacy preferences. Sort the users in the same social group in an ascending order based on the privacy strictness levels.

D. Identifying The Social Group

The policy recommendation process done based on the social groups. User U uploaded a new image and the A3P core invoked the A3P social for policy recommendation. The A3P social will find the social group and is most similar to user U and then choose the exemplary user in the social group along with users images to be sent to the A3P Core policy prediction system to generate the suitable policy for user U. Given that users of social sites is large and one user may have account with multiple social network, it would be very time consuming to compare the new users social context attributes against the frequent pattern of each social group [9]. In order to speed up the group identification process and ensure reasonable response time, leverage the inverted file structure to organize the social group information. The inverted file maps keywords occurring in the frequent patterns to the social groups that contain the keywords. First sort the keywords in the frequent patterns in an alphabetical order. Each keyword is associated with a link list and stores social group ID and pointers to the detailed information of the social group.

Three social groups G1, G2, G3 are formed based on the following frequent keywords.

G1: {female, movie, {0.6, 0.1, 0.2, 0.1}}
G2: {male, ski, student}
G3: {male, movie, IL}

Select the frequent attribute values except the social connection and build an inverted file as follows.

female: {G1}
IL: {G3}
hiking: {G3}
male: {G2;G3}
movie: {G1;G3}
student: {G2}

Given a new user, search users attribute values in the inverted file and obtain a set of candidate social groups. Then count the number of occurrence of the candidate groups during the search. Select the candidate group with the highest occurrence as the social group for the new user. User social context attributes are: {female, movie, teacher, NY, {0.65, 0.1, 0.15, 0.1}}, find that only the keywords female and movie appear in the inverted file. The social group related to female is G1 and the social groups related to movie are G1 and G3. Observe that G1 occurs twice in the search and G2 only once.

That means the new user have more matching keywords with G1 than G2 and other social groups and accordingly G1 is a better group for the new user. In the identified social group, further examine its subgroups by comparing the strictness levels of the sub groups with the new users preferred privacy strictness level if provided. Select the sub group based on the strictness level matches. If the new user did not specify privacy preference, select the sub group with the largest members. Selected sub group, look for the new user. Then there is a need to compare the new users and the group members remaining attributes that are not included in the previous pattern. The selected user and particular images and policies are sent to the A3P Core module to generate the recommended policy for the new user. Update the social group information by including the new user as a probation member. The probation member will not be chosen by A3P Social module to until the user uploaded sufficient images and becomes a regular member.

III. RESULTS AND DISCUSSION

Scale Invariant Feature Transform

SIFT key points are extracted from a set training images stored in the database. When a user uploads a new image, each individual feature is compared with images present in the database and finds the matching features by calculating the Euclidean distance of feature vectors. From the feature matches, subsets of key points which agree on the object and its location, scale and orientation in the uploaded image are identified to filter out the best matches. Hash table implementation of the generalized Hough transform is used to perform the determination of consistent clusters. by using an efficient feature extraction gives the accuracy to fit and number of false matches. Object matches which pass the tests can be identified as correct.

Lower method for image feature identification transforms an image into a large collection of feature vectors; each of it is invariant to image translation, scaling and rotation, partially invariant to illumination changes and robust to local geometric distortion. Object reorganization in primate vision uses the neurons in inferior temporal cortex for identifying features with similar characteristics.

Key locations are said to be as maxima and minima of the result of difference of Gaussians function which is applied in scale space to a series of smoothed and resample images. Dominant orientations are applied to localized key points to confirm that the key points are more stable for matching and recognition. SIFT descriptors robust to local affine distortion are then obtained by considering pixels around a radius of the key location, blurring and resembling of local image orientation planes.

A. Speeded Up Robust Feature Transform

SURF uses Hessian blob detector to detect interest points, it is computed with 3 integer operations using a precompiled integral image. Its feature descriptor works based on the sum of the Haar wavelet response around the point of interest.

It can also be computed with the help of the integral image. SURF descriptors can be used to locate and

recognize objects, people or faces, to make 3D scenes, to track objects and to extract points of interest shown in Figure 2.

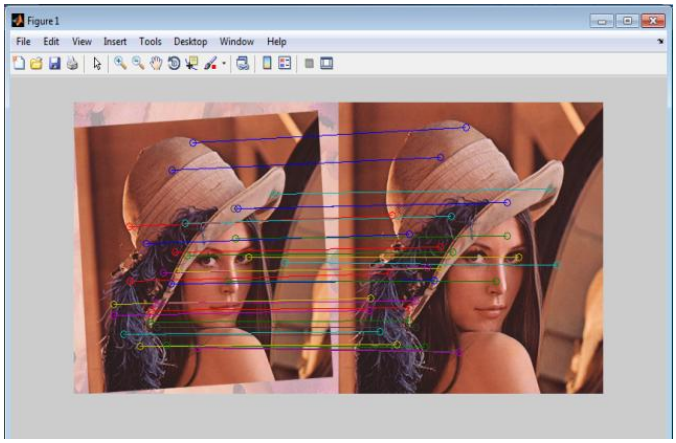

Figure 2. Speeded Up Robust Feature

SURF acts as a detector and a descriptor for identifying the points of interest in images in this method the image is first transformed into coordinates by using the multi resolution pyramid technique.

This method makes a copy of the image with Pyramidal Gaussian or Laplacian Pyramid shape and obtain image with reduced bandwidth and the same size. Scale Space is achieved for the original image which is a special blurring effect. This technique guards that the points of interest are scale invariant.

B. Result

An Adaptive and Intelligent photo sharing approach for an image is implemented using the MATLAB (MATrix LABoratory) software. Collection of images and their appropriate tags are stored in the databases. The database has four set of image classes. For each image class the first half data set is taken as training dataset and next half of the dataset is taken for the classification. The newly uploaed image is taken as the testing image shown in Figure 3.

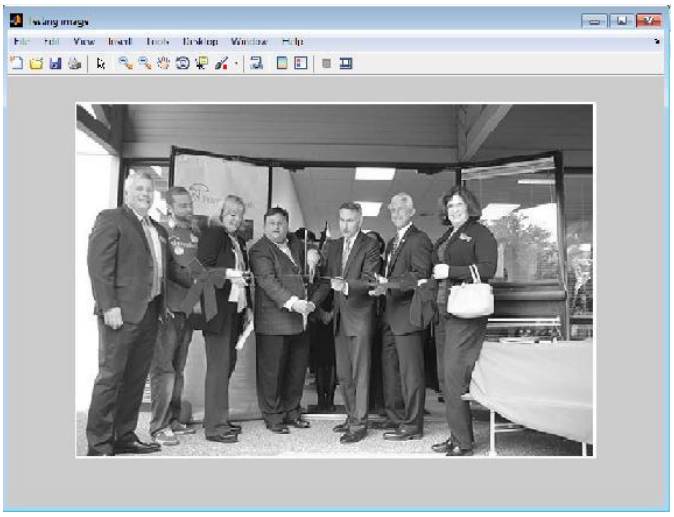

Figure 3. Testing Image

```
Tag_sequences =

    'PN'    'VB'    'PP'    'VB'    'DD'    'VB'

Selected image is Adult
Selected policy is
     1     1     1     0


        Family  : YES

        Friends  : YES

        Coworker  : YES

        Others  : No
fx >>
```

Figure 4. Predicted Policy

Finally, the newly uploaded testing image is displayed and automatic policy prediction system predicts policy for the given image shown in Figure 4. If, the user is satisfied with the predicted policy he/she just accept it. Otherwise, the image will be uploaded again.

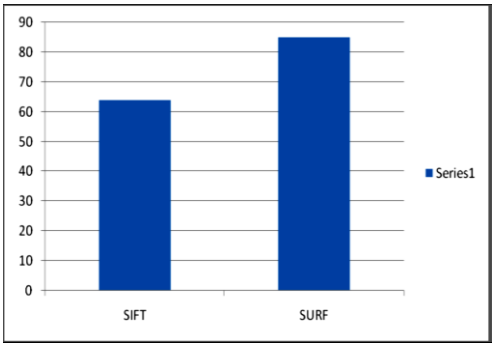

Figure 5. Comparison Chart

The comparison chart based on SIFT and SURF performance is shown in Figure 5.

IV. CONCLUSION

Images are one of the key enablers of users connectivity. Sharing takes place both among previously established groups of known people or social circle. Most content sharing websites allow users to enter their privacy preferences. The A3P system helps users automate the privacy policy settings for their uploaded images. The A3P system provides a comprehensive framework to infer privacy preferences based on the information available for a given user. The SURF effectively tackled the issue of interest point detection.

Misclassification of the images is one of the issues in A3P system. As future work, elaborate the existing system with the tampered image detection to identify the forgery images from the given image.

V. REFERENCES

[1] Acquisti, Aand Gross,R.(2006) 'Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook'Privacy Enhancing Technologies.

[2] Ahern S., Eckles D., Good NS., King S., Naaman M., and Nair R., (2007)'Over-exposed?: Privacy patterns and considerations in online and mobile photo sharing,' in Proceeding Conference Human Factors Computation System,, pp357–366.

[3] Ames Mand Naaman M., (2007) 'Why We Tag: Motivations or Annotation in Mobile and Online Media,' in Proceeding SIGCHI CHI'07.

[4] Bonneau J., Anderson J., and Danezis G., (2009) 'Prying data out of a social network,' in Proceeding International Conference Advances Social Network Analytical Miningpp.249–254.

[5] Chen H.-M., Chang M.-H., Chang P.-C., Tien M.-C., Hsu WH., and Wu J.-L., (2008) 'Sheepdog: Group and tag recommendation for flickr photos by automatic search-based learning,' in Proceeding 16th ACM International Conference Multimedia,, pp737–740.

[6] Choudhury MD., Sundaram H., Lin Y.R., John Aand Seligmann DD., (2009) 'Connecting content to community in social media via image content, user tags and user communication,' in Proceeding IEEE International Conference Multimedia Expo, pp.1238–1241.

[7] Datta R., Joshi D., Li J., and Wang J., (2008) 'Image retrieval : Ideas, influences and trends of new age', ACM Computation survey, vol40, no.2, p5.

[8] Jones Sand O'Neill E., (2011) 'Contextual dynamics of groupbased sharing decisions,' in Proceeding Conference Human Factors Computation System, pp1777–1786 9Kapadia A., Adu-Oppong F., Gardiner CK., and Tsang PP., (2008) 'Social circles: Tackling privacy in social networks,' in Proceeding Symposium Usable Privacy Security

[9] Klemperer P., Liang Y., Mazurek M., Sleeper M., Ur B., Bauer L., Cranor LF., Gupta N., and Reiter M., (2012) 'Tag, you can see it!: Using tags for access control in photo sharing,' in Proceeding ACM Annual Conference Human Factors Computation System, pp377– 386.

[10] Lerman K., Plangprasopchok Aand Wong C., (2007) 'Personalizing image search results on flickr', CoRR, volabs/0704.1676.

[11] Liu Y., Gummadi KP., Krishnamurthy Band Mislove A., (2011), 'Analysing facebook privacy settings: User expectations vsreality', in proceeding ACM SIGCOMM Conference Internet Measuring Conference, pp61-70.

[12] Lipford H., Besmer A., and Watson J., (2008) 'Understanding privacy settings in facebook with an audience view,' in Proceeding Conference Usability, Psychology, Security

[13] Loy Gand Zelinsky A., (2003) 'Fast radial symmetry for detecting points of interest', IEEE Transactions Pattern Analysis Machine Intelligence, vol25, no8, pp959-973

[14] Maximilien EM., Grandison T., Sun T., Richardson D., Guo S., and Liu K., (2009) 'Privacy-as-a-service: Models, algorithms, and results on the Facebook platform,' in Proceeding Web 2.0 Security Privacy Workshop.

[15] Mazzia A., LeFevre K., and E A..,, (2012) 'The PViz comprehension tool for social network privacy settings,' in Proceeding Symposium Usable Privacy Security.

[16] Ravichandran R., Benisch M., Kelley P., and Sadeh N., (2009) 'Capturing social networking privacy preferences,' in Proceeding Symposium Usable Privacy Security

[17] Wagner RAand Fischer MJ., (1974) 'The string-to-string correction problem,' JACM, vol21, no1, pp168–173

[18] Yeh C.-H., Ho Y.-C., Barsky BA., and Ouhyoung M., (2010) 'Personalized photograph ranking and selection system,' in Proceeding International Conference Multimedia, pp211–220.

[19] Yeung CA., Kagal L., Gibbins N., and Shadbolt N., (2009) 'Providing access control to online photo albums based on tags and linked data,' in Proceeding Social Semantic Web: Where Web 2.0 Meets Web 3.0 at the AAAI Symposium, pp9–14.