Machine Learning Research
2017; 2(3): 78-85
http://www.sciencepublishinggroup.com/j/mlr
doi: 10.11648/j.mlr.20170203.11

Research/Technical Note

Secrecy of Uploaded Images on Locates Using A3P Algorithm

Sivanantham Sivakumar, Musrathasleema Abdul Rahman, Nandhini Balu, Nivetha Nainar

Department of Information Technology, Adhiyamaan College of Engineering, Hosur, Tamil Nadu, India

Email address:

sivananthan.it@adhiyamaan.in (S. Sivakumar), thasee1995@gmail.com (M. A. Rahman), nandhininandhinib@gmail.com (N. Balu), nivethanrs95@gmail.com (N. Nainar)

To cite this article:

Sivanantham Sivakumar, Musrathasleema Abdul Rahman, Nandhini Balu, Nivetha Nainar. Secrecy of Uploaded Images on Locates Using A3P Algorithm. *Machine Learning Research.* Vol. 2, No. 3, 2017, pp. 78-85. doi: 10.11648/j.mlr.20170203.11

Received: February 13, 2017; **Accepted**: March 15, 2017, 2017; **Published**: March 29, 2017

Abstract: With increasing volume of images users share through social sites, maintaining privacy has become a major problem, as demonstrated by a recent wave of publicized incidents where users individually shared personal information. In light of these incidents, the need of tools to help users control access to their shared content is apparent. Toward addressing this need, we propose an Adaptive Privacy Policy Prediction (A3P) system to help users compose privacy settings for their images. We examine the role of social context, image content and metadata as possible indicators of user's privacy preferences. We propose a two-level framework which according to the user's available history on the site, determines the best available privacy policy for the user's images being uploaded. Our solution relies on an image classification framework for images categories which may be associated with similar policies on a policy prediction algorithm to automatically generate a policy for each newly uploaded image and also according to user's social features. Over time, generated policies will follow the evolution of user's privacy attitude.

Keywords: Content Sharing Sites, Privacy, Metadata, A3P, Social Media

1. Introduction

Images are now one of the key enablers of user's connectivity. Sharing takes place both among previously established groups of known people or social circles. (e.g., Google+, Flickr or Picasa), and also increasingly with people or social circles, for purposes of social discovery-to help them identify new peers and learn about interests and social surroundings. However, semantically rich images may reveal content sensitive information [1] [2].

Most content sharing websites allow users to enter their privacy preferences. Unfortunately, recent studies have shown that users struggle to set up and maintain such privacy setting [3] [4] [5] [6]. One of the main reasons provided is that given the amount of shared information this process can be tedious and error-prone. Therefore, many have acknowledged the need of policy recommendation systems which can assist users to easily and properly configure privacy settings [7] [8] [9]. However, existing proposals for automating privacy settings appear to be inadequate to address the unique privacy needs of images due to the amount of information implicitly carried within images and their relationship with the online environment wherein they are exposed.

In this paper, we propose an Adaptive Policy Prediction (A3P) system which aims to provide users a hassle free privacy settings experience by automatically generating personalized policies. The A3P system handles user uploaded images and factors in the following criteria that influence one's privacy settings of images:

The impact of social environment and Personal characteristics. Social context of users such as their profile information and relationships with others may provide useful information regarding users' privacy preferences. For example, users interested in photography may like to share their photos with other amateur photographers. Users who have several family members among their social contacts may share with them pictures related to family events.

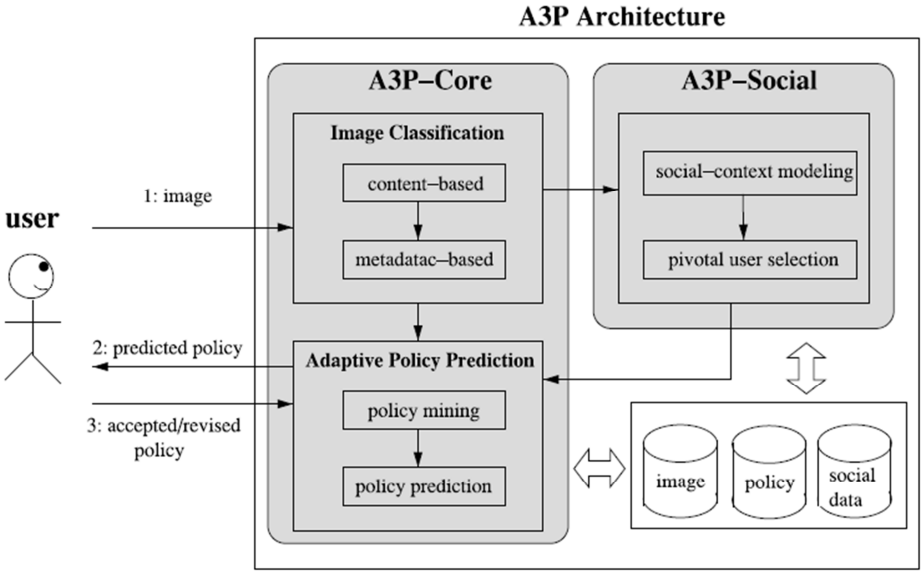

Figure 1. *System Overview.*

The role of images content and metadata. In general, similar images often privacy preferences, especially when people appear in the images. For example, one may upload several photos of his kids and specify that only has family members are allowed to see these photos. He/she may upload some other photo of landscapes which he took as a hobby and for these photos, he may set privacy preference allowing anyone to view and comment the photos. While a more conservation person may just want to share personal images with his family members. In light of these considerations, it is important to find the balancing point between the impact of social environment and user's individual characteristics in order to predict the policies that match each individuals may change their overall attitude towards privacy as time passes. In order to develop a personalized policy recommendation system, such changes on privacy opinions should be carefully considered.

Role of Image's Content and Metadata

In general, similar images often incur similar privacy preferences, especially when people appear in the images. For example, one may upload several photos of his kids and specify that only his family members are allowed to see these photos. He may upload some other photos of landscapes which he took as a hobby and for these photos, he may set privacy preference allowing anyone to view and comment the photos.

Analyzing the visual content may not be sufficient to capture users' privacy preferences. Tags and other metadata are indicative of the social context of the images, including where it was taken and why [10] and also provide a synthetic description of images, complementing the information obtained from visual content analysis.

Corresponding to the aforementioned two criteria, the proposed A3P system is comprised of two main buildings blocks are A3P Social and A3P Core

The A3P core focus on analyzing each individual user's own images and metadata, while the A3P-Social offers a community perspective of privacy setting recommendations for a user's potential privacy improvement. We design the interaction flows between the two building blocks to balance the benefits from meeting personal characteristics and obtaining community advice.

In this work, we present an overhauled version A3P which includes an extended policy prediction algorithm in A3P-Core (that is now parameterized based on user groups and also factors in possible outliers) and new A3P- Social module that develops the notion of social context to refine and extend the prediction power of our system.

2. A3P System Overview

The A3P System consists of two components are

(i) A3P Core
(ii) A3P Social

When a user uploads an image, the image will be first sent to the A3P Core. The A3P-Core classifies the images and determines whether there is need to invoke the A3P-Social. In most cases, the A3P-Core predicts policies for the users directly based on their historical behavior. If one the following two cases is verified true, A3P- Core will invoke A3P Social:

(i) The user does not have enough data for the type of uploaded image to conduct policy prediction.
(ii) The A3P- Core detects the recent major changes among the user's community about their privacy practices along with user's increase of social networking activities.

The A3P- Social groups users into social communities with similar social context and privacy preferences and continuously monitors the social group for user and sends back the information about the group to the A3P-core for policy prediction. At the end, the predicted policy will be displayed to the user. If the user is fully satisfied by the

predicted policy, he or she can just accept it. Otherwise, the user can choose to revise the policy. The actual policy will be stored in the policy repository of the system for the policy prediction of future uploads.

3. A3P Core

There are two major components in A3P- Core

(i) Image Classification

(ii) Adaptive policy prediction

For each user, his/her images are first classified based on content and metadata. Then, privacy policies of each category of images are analyzed for the policy prediction.

Adopting a two- staged approach is more suitable for policy recommendation than applying the common one-stage data mining approaches to mine both image features and policies together. Recall that when a user uploads a new image, the user is waiting for a recommended policy. The two-stage approach allows the system to employ the first stage to classify the new image and find the candidate sets of images for subsequent policy recommendation. As for the one- stage mining approach, it would not be able to locate the right class of the new image because its classification criteria needs both image features and policies whereas the new image are not available yet. Moreover, combining both

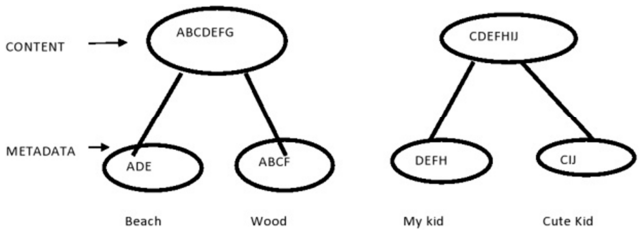

Figure 2. *Two- level image classification.*

Image features and policies into a single classifier would lead to a system which is very dependent to the specific syntax of the policy, if a change in the supported policies were to be introduced, the whole learning model would need to change.

3.1. Image Classification

To obtain groups of images that may be associated with similar privacy preferences, we propose a hierarchical image classification which classifies image first based on their metadata. Images that do not have metadata will be grouped only by content. Such a hierarchical classification gives a higher priority to image content and minimizes the influence of missing tags. Note that it is possible that some images are included in multiple categories as long as they contain the typical content features or metadata of those categories.

Moreover, Fig. 2 shown an example of image classification for ten images named on A, B, C, D, E, F, G, H, I, J respectively. The content based classification creates two categories: “landscape” and “kid”. Images C, D, E and F are included in both categories as they show kids playing outdoor which satisfy the two themes: “landscape” and “kid”. These two categories are further divided into sub categories based on tags associated with the images. As a result, we obtain two subcategories under each theme respectively. Notice that image G is not shown in any sub category it does not have any tag; image a shows up in both sub categories because it has tags indicating both beach and food.

3.1.1. Content Based Classification

Our approach to content based classification is based on an efficient and yet accurate image similarity approach. Specifically, our classification algorithm compares image signatures defined based an quantified and sanitized version of Haar wavelet transformation. For each image, the wavelet transform encodes frequency and spatial information related to image color, size invariant transform, shape, texture, symmetry, etc., Then, small number of co efficient are selected to form the signature of the image. The content similarity among images is then determine by the distance among their image signature.

Our selected similarity criteria include texture, symmetry, shape (radial symmetry and phase congruency) [11] and SIFT [12]. We also account for color and size. We set the system to start from 5 generic image classes: (a) explicit (e.g. nudity, violence, drinking, etc.,), (b) adults, (c) kids, (d) scenery (e.g beach, mountains), (e) animals. As a preprocessing step, we popular the 5 baseline classes for manually assigning to each class a number of images from Google images, resulting in about 1,000 images for class. Having large image data set beforehand reduces the chance of misclassification. Then, we generate signature of the all the images and store them in the database.

3.1.2. Metadata Based Classification

The metadata based classification groups images into sub categories under aforementioned based in categories. The process consist of three main steps.

The first step is to extract keyword from the metadata associated within image. The metadata consider in our work are tags, captions and commons. We identified all the nouns, verbs and objectives in the metadata and store them as metadata vectors t_{noun}1/4 $ft_1;t_2;\ldots; t_i$g, t_{verb}1/4 $ft_1;ft_2;\ldots; t_j$g and t_{adj}1/4 $ft_1;ft_2;\ldots;t_k$g, where i, j and k are the total number of nouns, verbs and adjectives respectively.

The second step is to derive a representative hypernym (denoted as h) from each metadata vector. We first retrieve the hypernym for each t_i in a metadata vector based on the word net classification obtain a list of hypernym h ¼ fðv$_1$, f$_1$Þ, ðv$_2$, f$_2$Þ, g, where v denotes hypernym and f denotes its frequency. For example, consider a metadata vector t ¼ f ‘cousin’, “first steps”, “baby boy”g. we find that “cousin” and “baby boy” have the same hypernym “kid” and “first steps” has a hypernym “initiation”. Correspondingly, we obtain the hypernym list h ¼ f(kid, 2), (initiative, 1)g. in this list, we select the hypernym with the third step is to find a subcategory that an image belongs to. This is an incremental procedure. At the beginning, the first image forms a subcategory as itself and the representative hypernyms of the image becomes the subcategory’s representative hypernyms. Then, we compute the distance between representative hypernyms of a new

incoming image and each existing subcategory. Given an image, let h_n, h_a and h_v denote its representative hypernyms in a metadata vectors corresponding to nouns, adjectives and verbs respectively. For a subcategory c, let h^c_n, h^c_a and h^c_v denotes its representative hypernyms of nouns, adjectives and verbs respectively. The distance between the image and the subcategory is computed as a weighted sum of the edit distance [13] between corresponding pair of representative hypernyms as shown in equation (1), where W denotes the weight and D denotes the edit distance $Dist_m$1/4 $w_nDh_n;Dh_n;h^c_n$ þ w_a $Dh_a;h_a{}^c$ þ $w_vDh_v;h_v{}^c$.

3.2. Adaptive Policy Prediction

The policy prediction algorithm provides a predicted policy of newly uploaded image to the user for his/her reference. More importantly, the predicted policy will reflect the possible changes of a user's privacy concern. The prediction process consist of three main phases: (i) policy normalization; (ii) policy mining; and (iii) policy prediction. The policy normalization is a simple decomposition to convert a user policy into a set of atomic rules in which the data component is a single element set.

3.2.1. Policy Mining

We propose a hierarchical mining approach for policy mining. Our approach leverages association rule mining techniques to discover popular patterns in policies. Policy mining is carried out within the same category of the new image because images in the same category are more likely under the similar level of privacy prediction. The basic idea of hierarchical mining is to follow a natural order in which user defines a policy. Correspondingly, the hierarchical mining first look for popular subjects define by the user, then look for popular actions in the policies containing the popular subjects and finally for popular conditions in the policies containing both popular subjects and conditions.

Step 1: In the same category of new image, conduct association rule mining on the subject component of policies. Let S1 and S2 denotes the subjects occurring in policies. Each resultant rule is an implication of the form X. Y where X, Y fS1, S2g and X\Y ¼; Among the obtained rules, we select the best rules according to one of the interestingness measures, i.e. The generality of the rule defined using support and confidence as introduce in [14]. The selected rules indicate the most popular subjects combination in policies. In the subsequent steps, we consider policies which contain at least one subject in the selected rules. For clarity, we denote the set of such policies as G^{sub} corresponding to a selected rule R^{Ssub}i.

3.2.2. Policy Prediction

The policy mining phase may generate several candidate policies while the goal of our system is to return the most promising one to the user. Thus, we present an approach to choose the best candidate policy that follows the user's privacy tendency.

To model the user's privacy tendency, we define a notion of strictness level. The strictness level is a quantitative metric that describes how "strict" a policy is. In particular, a strictness level L is an integer with minimum value in zero, where in the lower the value, the higher the strictness level. It is generated by two metrics: major level (denoted as 1) and coverage rate (a), where 1 is determined by the combination of subject and action policy, and isdetermined by the system using the condition component.

4. A3P Social

The A3P Social employee multi criteria inference mechanism that generates representative policies by leveraging key information related to the user social contexts and his general attitude towards privacy as mentioned earlier, A3P social will be invoked by the A3P core into scenarios. When the user is newbie of a site and does not have enough images stored for the A3P Core to infer meaningful and customized policies. The other is when the system notifies significant chances of privacy trend in the user social circles, which may be of interest for the user possibly adjust his/her privacy settings accordingly.

4.1. Modelling Social Context

We observe that users with similar background tend to have similar privacy concerns has seen in previous research studies and also confirmed by are collected data. This observation inspires us to develop a social context modelling algorithm that can capture the common social elements of users and identified formed by the users similar privacy concerns. The identified communities who have a reach set of images can then serve has the base of subsequent policies recommendation. The Social context modelling algorithm consist of two major steps is to identified and formalized potentially important factors that may be informative of one's privacy settings the second step is to group users based on the identified factors.

First, we model each users social context has a list of attributes; {Sc_1, Sc_2;…; Sc_n}, where Sc_i denote a social context attribute and n is the total number of distinct attributes in the social networking sites. These social context attributes are extracted from user's profile. Besides basic elements in users' profile, many social sites also allow users to group their contact based on relationships. If search grouping functionality is available, we will consider it influence on privacy settings too in a social site some users may only have their family members as contacts, while some users may have contacts including different kinds of people that they met offline are on the internet. The distribution of contact may shed light on the user behavior of privacy settings. We assume that users who mainly share images among family members may not want to disclose personal information publicly, while users having large group of friends may be willing to share more images with large audience [15]. Formally, we model the ratio of each type of relationship among all contacts of a user a social connection. Let R1, R2…Rn denote the n types of relationship observed among all the users. Let $N_R{}^U{}_i$ denote the number of users U's contacts belonging to relationship type R_i the connection distribution is represented as below

$$\text{Conn: } \left(P_{ni}N_{1/_{41\ Ru}}N_{1\ R_u i};\ \ldots;\ P_{ni}N_{1/_{41RUn\ Un}}\right) : N_R$$

4.2. Identifying Social Groups

We now introduced the policy recommendation process based on the social groups obtain from the previous step.

Suppose that a 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 along with his images to be sent to the A3P- Core policy prediction module to generate the recommended policy for user U. Given that the number of users in social network may be huge and that users may join a large number of social group, it would be very time consuming the compare the new users social contacts attributes again the frequent pattern of each social group in order to speedup the group identification process and ensure reasonable response time, we leverage the inverted file structure organize the social group information. The inverted file maps keywords (values of social context attribute) occurring in the frequent patterns to the social groups contain the keywords. Specifically we first sort the keywords in the frequent patterns in an alphabetical order each keyword is associate with a link list which store social group ID and pointers to the detailed information of the social group.

5. Screenshots

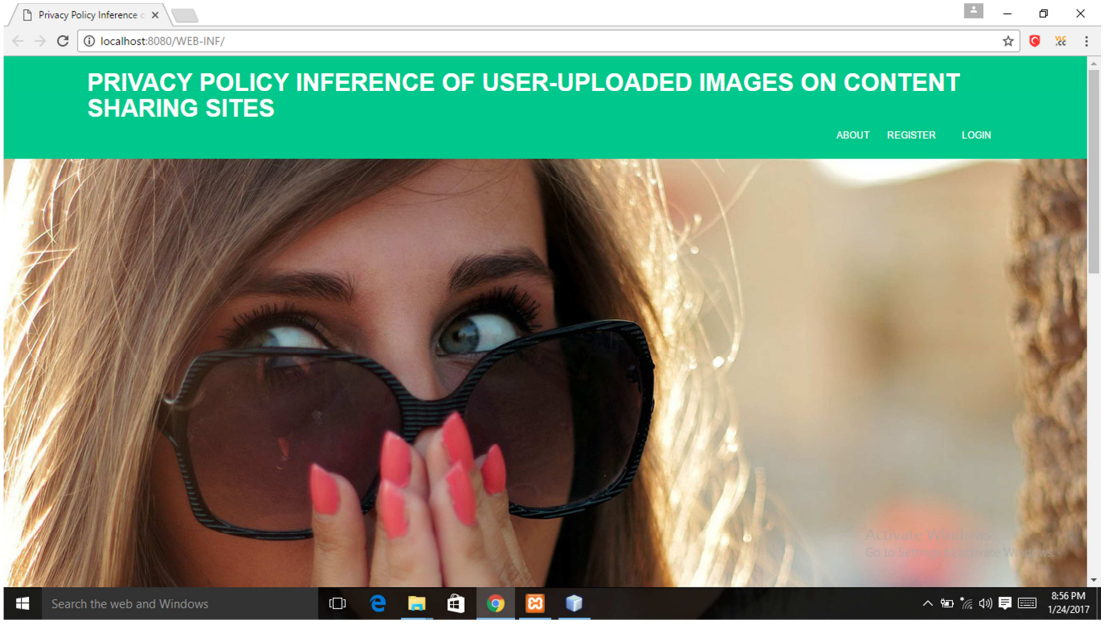

Figure 3. *Home Page.*

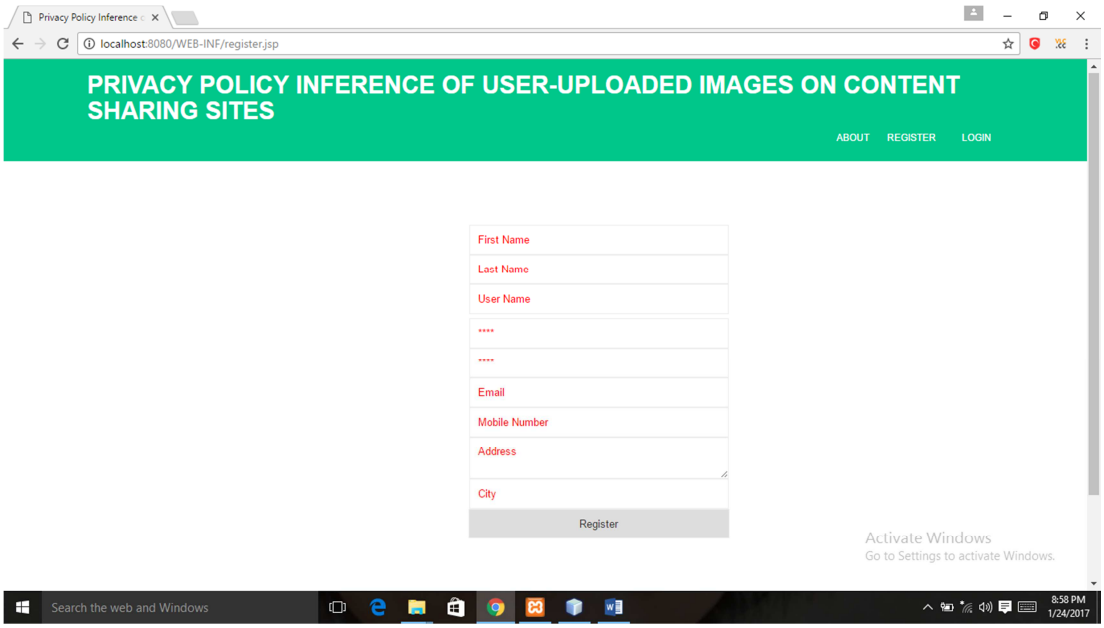

Figure 4. *Register Page.*

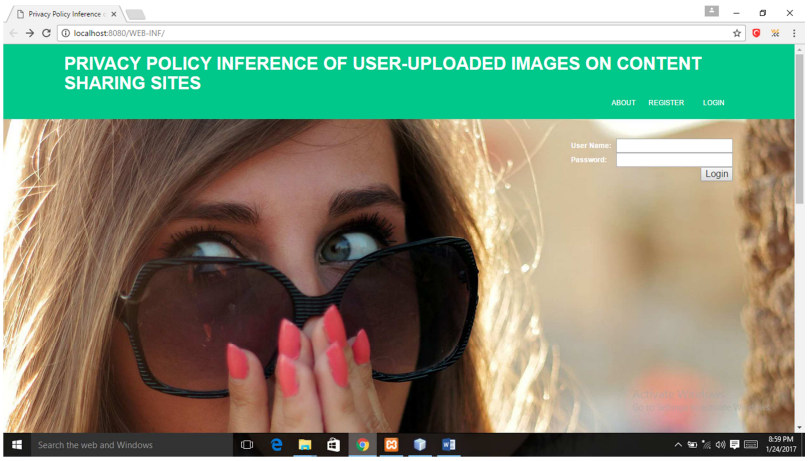

Figure 5. *Login Page.*

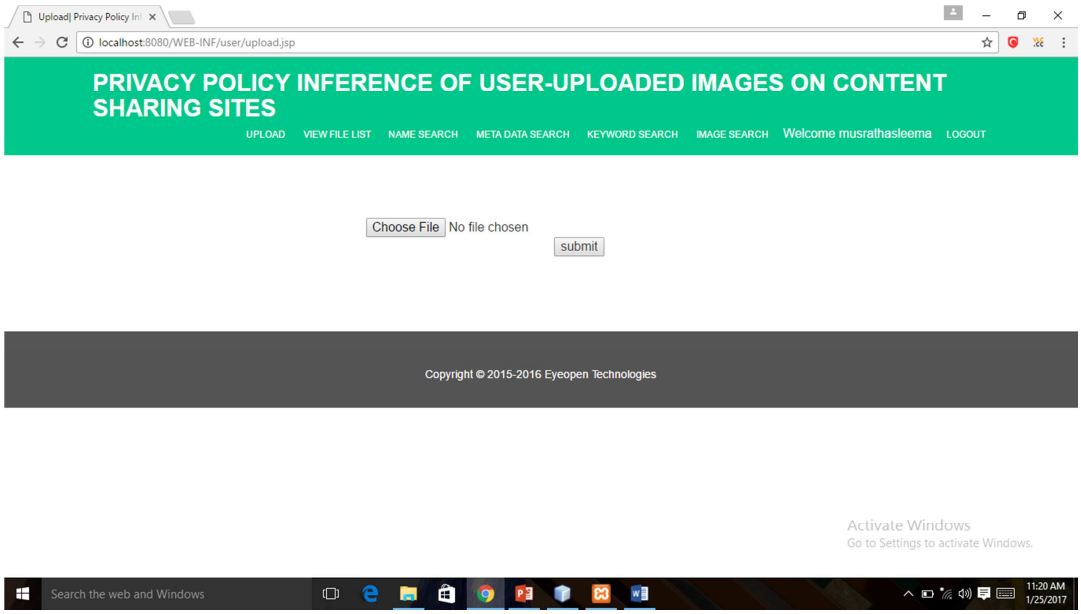

Figure 6. *Upload Page(a).*

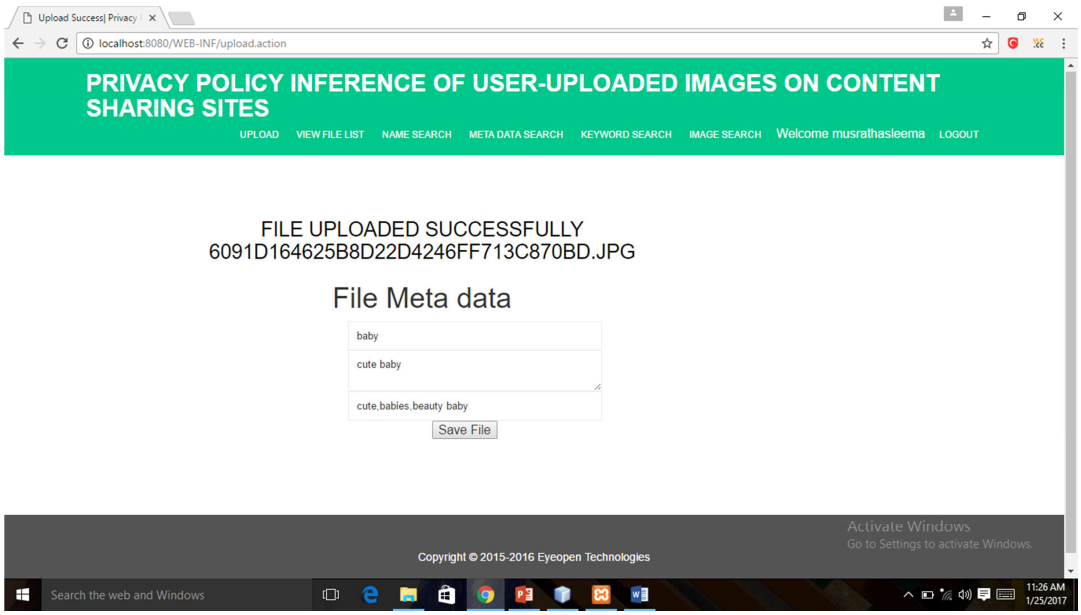

Figure 7. *Upload Page (b).*

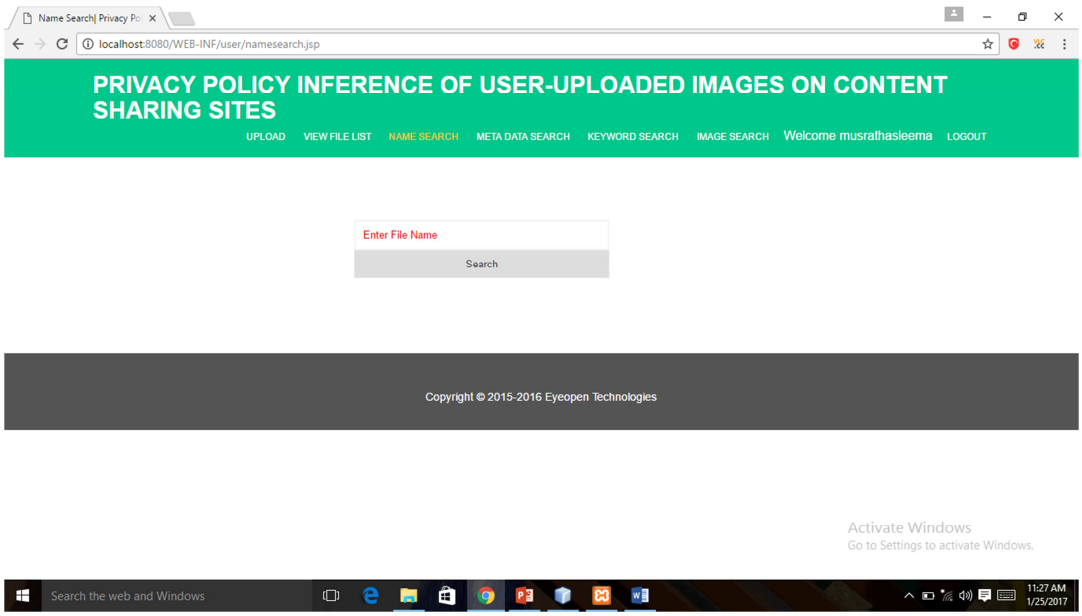

Figure 8. *Name search.*

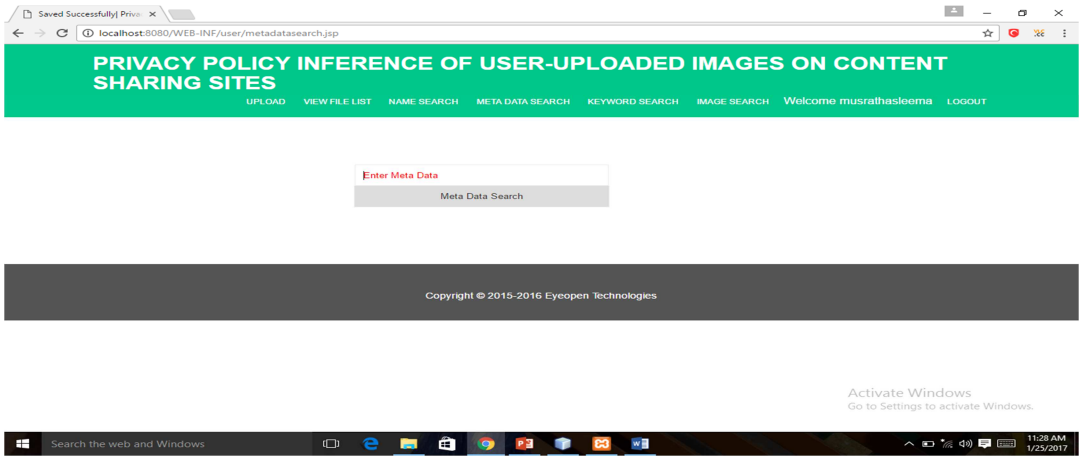

Figure 9. *Metadata Search.*

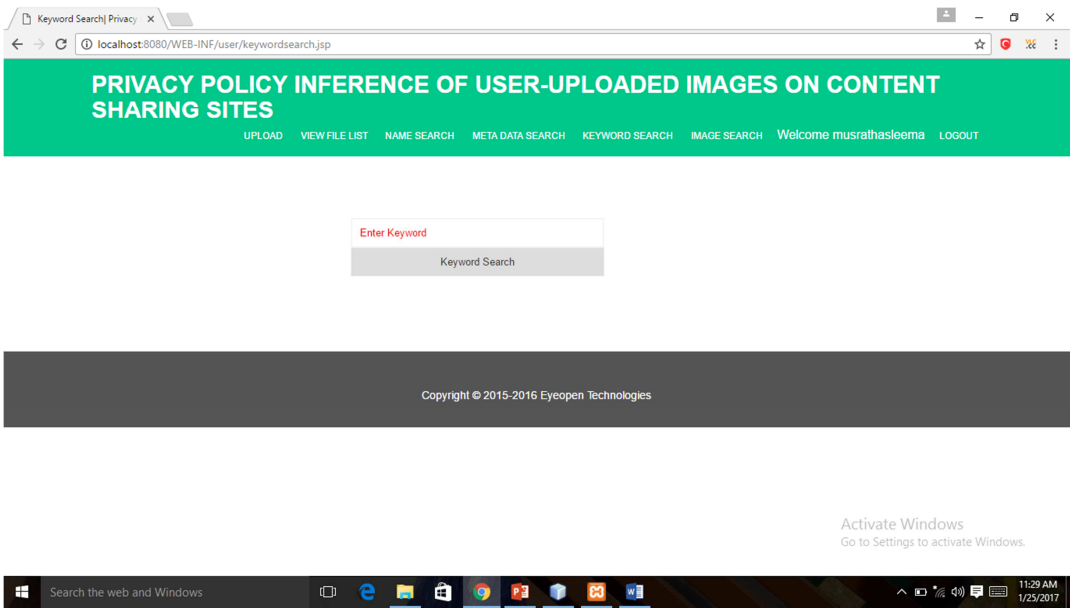

Figure 10. *Keyword Search.*

6. Conclusion

We have used an adaptive privacy policy prediction (A3P) system that helps user automate the privacy policy settings for the uploaded images. The A3P system provides a comprehensive framework to infer privacy preferences based on the information available for given user. We also effectively tackled the issue of cold-start, leveraging social context information. Our experimental study proves that our A3P is a practical tool. That offers significant improvements over current approaches to privacy.

References

[1] S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair, "Over-exposed? Privacy patterns and considerations in online and mobile photo sharing," in Proc. Conf. Human Factors Comput. Syst., 2007, pp. 357–366.

[2] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Analyzing facebook privacy settings: User expectations vs. reality," inProc. ACMSIGCOMM Conf. Internet Meas. Conf., 2011, pp. 61–70.

[3] S. Jones and E. O'Neill, "Contextual dynamics of group-based sharing decisions," in Proc. Conf. Human Factors Comput. Syst., 2011, pp. 1777–1786. [Online]. Available: http://doi.acm.org/ 10.1145/1978942.1979200.

[4] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the facebook," in Proc. 6th Int. Conf. Privacy Enhancing Technol. Workshop, 2006, pp. 36–58.

[5] L. Church, J. Anderson, J. Bonneau, and F. Stajano, "Privacy stories: Confidence on privacy behaviors through end user programming," in Proc, 2009, 5th Symp. Usable Privacy Security.

[6] H. Lipford, A. Besmer, and J. Watson, "Understanding privacy settings in facebook with an audience view," in Proc. Conf. Usability, Psychol., Security, 2008.

[7] K. Strater and H. Lipford, "Strategies and struggles with privacy in an online social networking community," in Proc. Brit. Comput. Soc. Conf. Human-Comput. Interact, 2008, pp. 111–119.

[8] J. Bonneau, J. Anderson, and L. Church, "Privacy suites: Shared privacy for social networks," in Proc. Symp. Usable Privacy Security, 2009.

[9] A. Mazzia, K. LeFevre, and A. E., "The PViz comprehension tool for social network privacy settings," in Proc. Symp. Usable Privacy Security, 2012.

[10] R. Ravichandran, M. Benisch, P. Kelley, and N. Sadeh, "Capturing social networking privacy preferences," in Proc. Symp. Usable Privacy Security, 2009.

[11] A. Besmer and H. Lipford, "Tagged photos: Concerns, perceptions, and protections," in Proc. 27th Int. Conf. Extended Abstracts Human Factors Comput. Syst., 2009, pp. 4585–4590.

[12] C. A. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt, "Providing access control to online photo albums based on tags and linked data," in Proc. Soc. Semantic Web: Where Web 2.0 Meets Web 3.0 at the AAAI Symp., 2009, pp. 9–14.

[13] M. Ames and M. Naaman, "Why we tag: Motivations for annotation in mobile and online media," in Proc. Conf. Human Factors Comput. Syst., 2007, pp. 971–980.

[14] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," IEEE Trans. Pattern Anal. Mach. Intell. 2003, vol. 25, no. 8, pp. 959–973, Aug.

[15] D. G. Lowe, (2004, Nov.). Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. [Online]. 60 (2), pp. 91–110. Available: http://dx.doi.org/10.1023/B: VISI.0000029664.99615.