COMPLIANCE OF PRIVACY POLICIES WITH LEGAL REGULATIONS

Compliance of Privacy Policies with Canadian PIPEDA

Nolan Zhang, Peter Bodorik
Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada

Dawn Jutla
*Dept. of Finance, Information Systems, and Mgmt. Science, Sobey School of Business, Saint Mary's University
Halifax, Nova Scotia, Canada*

Keywords: Privacy Technologies, Platform for Privacy Preferences (P3P), P3P Agent, P3P Privacy Policy, Natural Language Privacy Policy, Compliance of Privacy Policy with Legal Requirements, Classification of Privacy Policy, Support Vector Machine, Decision Tree Analysis, Principal Components Analysis.

Abstract: The W3C's Platform for Privacy Preferences (P3P) is a set of standards that provides for representation of web-sites' privacy policies using XML so that a privacy policy can be automatically retrieved and inspected by a user's agent. The agent can compare the site's policy with the user's preferences on collection and use of his/her private data. If the site's privacy policy is incompatible with the user's preferences, the agent informs the user on the privacy policy's shortcomings. The P3P specification defines XML tags, schema for data, set of uses, recipients, and other disclosures for expressing web-sites' privacy policies. It is important for the user's agent to determine whether the site's privacy policy actually satisfies privacy regulations that are applicable to the user's current transaction. We show that the P3P specification is not sufficiently expressive to capture all of the legal requirements that may apply to a transaction. Consequently, to determine whether or not a site's privacy policy satisfies the requirements of a particular law in question, the site's privacy policy expressed in the natural language must also be retrieved and examined. To determine which legal requirements of a particular law are satisfied by the site's P3P privacy policy, which is an XML document, we examine the document's XML tags - a relatively straight-forward task. To determine whether legal requirements, which cannot be satisfied by using P3P XML tags, are present in the site's privacy policy expressed in the natural language, we use standard classification algorithms. As a proof of concept, we apply our approach to the Canadian PIPEDA privacy law and show up to 88% accuracy in identifying the legal privacy clauses concerning the Safeguard principle in privacy statements.

1 INTRODUCTION

Web-sites frequently request personal identifiable information (PII) for various reasons, such as personalizing customer experiences or conducting financial transactions. Hence it is no surprise that privacy has received a great deal of attention and that laws, legislations, regulations, and standards have been recently created/enacted to safeguard and manage the privacy of personal information in the digital world. Research and development on privacy has led to supporting technologies and emerging standards. Although most web-sites post their privacy policies written in natural language, research has shown that such policies tend to be written in "legalese" and that people find them hard to read and understand (Cranor, 2003). Furthermore, they are hard to "digest" by automated agents. Yet, it has also been shown that the issue of privacy is dear to the hearts of web-users wherein privacy plays an important role in fostering customers' trust and managing privacy well may lead to an increasing use of the web in conducting business (Adams, 2000; Ackerman and Cranor, 1999; Chellappa and Pavlou, 2000).

Zhang N., Bodorik P. and Jutla D. (2007).
COMPLIANCE OF PRIVACY POLICIES WITH LEGAL REGULATIONS - Compliance of Privacy Policies with Canadian PIPEDA.
In *Proceedings of the Second International Conference on e-Business*, pages 277-284
DOI: 10.5220/0002113202770284

It has been opined that there are two motivations for companies to be serious about privacy issues – the carrot and the stick (Chan et al, 2005). The carrot is improving the company's image and strategic differentiation by providing visible privacy protection to users since studies show (e.g., (Adams, 2000)) that privacy is important to users as they perform activities on the web. The stick is in the form of privacy laws, regulations, business associations' standards and possible retributions if a company does not provide privacy protection according to applicable privacy regulations, e.g. the Health Insurance Portability and Accountability Act (HIPAA), Gramm-Leach-Bliley Act Regulation P, and Canada's Personal Information Protection and Electronic Documents Act (PIPEDA).

1.1 Semantic Web's Support for Privacy

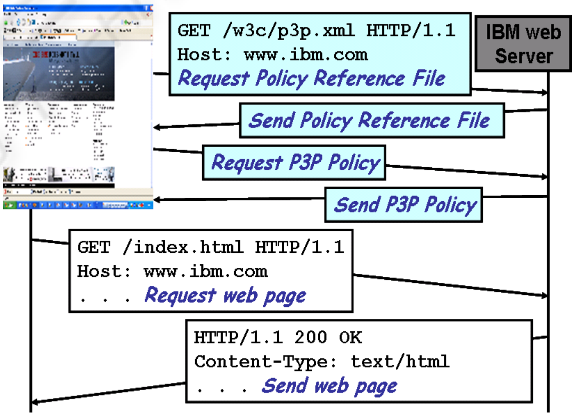
Currently, the W3C's Platform for Privacy Preferences (P3P) recommendation is the most mature to support privacy on the semantic web. P3P specifies how web-sites can use XML with namespaces to express their privacy practices related to collection of personal data about users and usage and distribution of such data. We standardly refer to any web-site's privacy policy expressed using the P3P specification as a P3P policy. The specification includes a schema for data, set of uses, recipients, and other disclosures, and XML format for expressing policies. It specifies also where privacy policies are posted, so that they could be found by automated agents, and how they can be retrieved using HTTP.

In addition to a P3P privacy policy, the site must also include, in the discuri XML element, the URL of its privacy policy expressed in natural language (which we assume to be English). Agents can thus retrieve not only a site's P3P policy but also the applicable privacy policy expressed in natural language, simply referred to as a natural policy.

A simple interaction of a P3P enabled web-site and a user's agent (web-browser) is depicted in Figure 1. A P3P user agent uses the protocol defined in the P3P specification to retrieve the privacy policy from a web server (IBM's web server in the figure). Initially, the user agent sends a standard HTTP request to fetch the P3P policy reference file at www.ibm.com/w3c/p3p.xml. With the policy reference file sent back by the web server, the user agent is able to locate and download the P3P policy file and compare it with the user's privacy preferences. If the retrieved policy satisfies the user's preferences, the web-page is retrieved; otherwise the user is informed – the policy either does not satisfy the preferences or the agent is not certain and seeks further guidance from the user. Preferences are expressed in a language, such as the P3P Preference Exchange Language (APPEL), Xpref, or Semantic Web Rule Language (SWRL), languages that are not considered to be user-friendly (Hogben, 2003; Cranor, 2003).

1.2 Objectives

As mentioned previously, many privacy regulations have been enacted in various countries. Users are certainly interested whether a web-site's privacy policy satisfies not only the user's preferences but also requirements of applicable privacy laws. This leads to a question: Is current P3P expressive enough to represent requirements of applicable privacy laws? Section 2 of this paper shows that it is not and hence the user agent must retrieve and analyze both of the site's privacy policies, the P3P policy and the privacy policy expressed in natural language, and *determine whether the site's privacy policy complies with legal requirements of a specific law/regulation in question* – how this problem can be tackled is the objective of this paper.



```
GET /w3c/p3p.xml HTTP/1.1
Host: www.ibm.com
```

```
GET /index.html HTTP/1.1
Host: www.ibm.com
. . .
```

```
HTTP/1.1 200 OK
Content-Type: text/html
. . .
```

Figure 1: Retrieving a P3P policy.

The tags of the P3P policies are examined in order to determine which portions of the particular legal act under consideration are addressed by the P3P privacy policy. The natural language policy is also analyzed to determine whether it addresses legal requirements that cannot or were not expressed using P3P. We have experimented with this approach by classifying privacy policies posted by various organizations in order to determine whether they comply with the legal requirements of the Canadian Protection of Privacy Preference and

Electronic Documents Act (PIPEDA). This paper describes the approach and the results of experimentation in detail.

1.3 Outline of Further Sections

Section 2 examines the PIPEDA (Privacy Commissioner of Canada, 2000) in relation to the P3P-defined format and determines that XML tags are insufficient to describe all of the PIPEDA's requirements. Thus to determine whether a web-site's privacy policy is compliant to PIPEDA principles, its natural language privacy policy must be retrieved and analyzed in addition to the P3P privacy policy. Section 3 describes how a classification technique can be applied to a site's natural privacy policy in order to determine whether it complies with PIPEDA's legal requirements. Results of experimentation are described in Section 4, while section 5 offers summary and conclusions.

2 PIPEDA PRINCIPLES AND P3P XML TAGS

PIPEDA, which came in force in 2004, specifies 10 privacy principles to which any Canadian business must comply when storing and managing private data. As a consequence, any privacy policy a business posts on its website must also satisfy, or comply with, these principles. We have analyzed these principles and compared them to the P3P-defined XML tags in order to determine which tags, if any at all, can be used to express any of the PIPEDA principles. As Table 1 shows, P3P defines corresponding tags for seven of PIPEDA's principles. For instance, the PIPEDA's principle dealing with identification of the purposes for which personal data is collected can be expressed in a P3P privacy policy using the tag <PURPOSE>. Similarly, the PIPEDA's principle of *Consent* can be expressed in a P3P policy using XML tags <REQUIRED> and <opt-in>.

There are three PIPEDA principles, namely, Accountability, Accuracy, and Safeguards, for which P3P does not define XML tags and, hence, cannot be expressed in a P3P policy. Consequently, the natural language privacy policy must be retrieved and examined in order to determine whether or not the site's privacy policy complies with PIPEDA.

Table 1: PIPEDA Principles and P3P Tags.

Ref. in PIPEDA	Privacy Principles	Corresponding P3P Tags
4.1	Accountability	**NONE**
4.2	Identifying Purposes	<PURPOSE>
4.3	Consent	<REQUIRED>, <opt-in>
4.4	Limiting Collection	<PURPOSE>
4.5	Limiting Use, Disclosure, and Retention	<PURPOSE>, <RETENTION>, <RECIPIENT>, <POLICY>, <opturi>
4.6	Accuracy	**NONE**
4.7	Safeguards	**NONE**
4.8	Openness	<POLICY>, <discuri>
4.9	Individual Access	<ACCESS>
4.10	Challenging Compliance	<DISPUTES-GROUP>

3 CLASSIFICATION OF NATURAL PRIVACY POLICY

The objective is to classify a web-site's privacy policy to determine whether it satisfies the PIPEDA's principles. The user agent first examines the site's P3P policy, expressed using XML. It examines the presence and content of any tags that correspond to the seven PIPEDA principles – a relatively straightforward task.

To determine whether the site's stated privacy practices comply with the PIPEDA's principles of Accountability, Accuracy, and Safeguards, which do not have any corresponding P3P tags, in our current implementation of the user's privacy agent, the site's natural language privacy policy is analyzed by a standard classification algorithm, which first needs to be trained.

For training, a data set for each privacy principle is collected, labelled, and classified by a human trainer as either addressing a PIPEDA principle or not. In order to improve the information gain of the selected feature privacy principle, a privacy policy file is divided into smaller files, which are labelled independently. Then standard classification algorithms are trained by the labelled training data set, and thus the predicting model learns through the training process.

To test the coverage of a privacy policy file, the file is divided into smaller files, and then the predicting models are used to label each of the

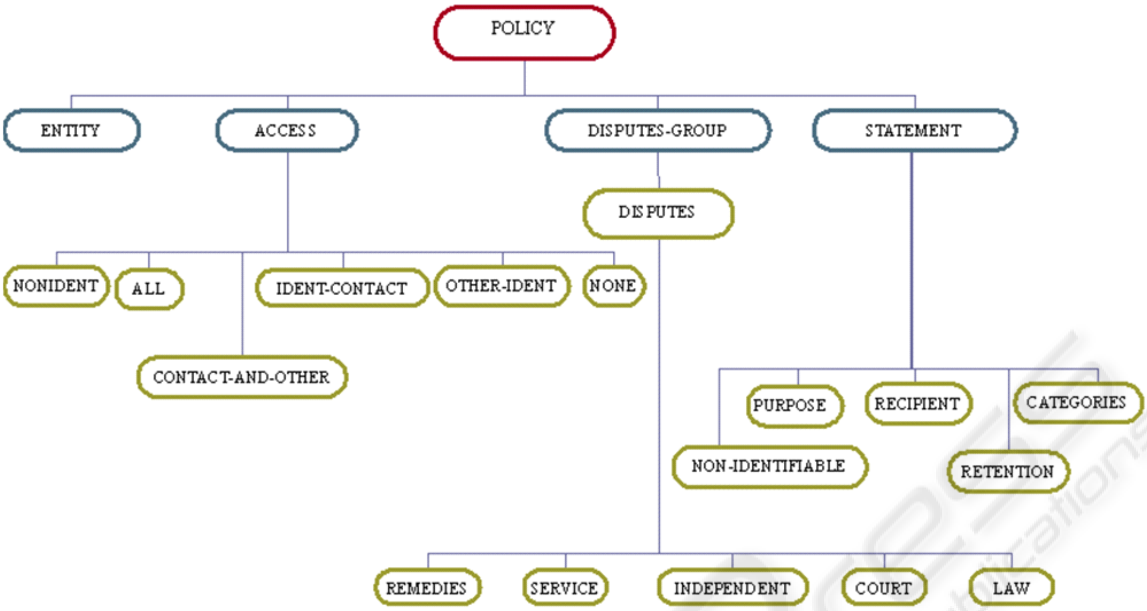


Figure 2: P3P XML Tags Decision Tree.

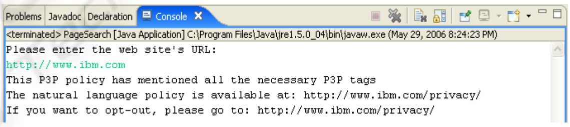
smaller files. Conceptually, if the smaller files are labelled by all three PIPEDA principles, and the seven P3P tags were found in a separate pre-process, then the privacy policy addresses all ten PIPEDA principles.

3.1 Checking P3P Policy

To check that the P3P policy satisfies the seven P3P principles that have corresponding P3P-defined XML tags, the tags were stored in a tree structure as defined in the P3P tag hierarchy.

The input to the agent is a web site's URL, and the P3P policy file is automatically retrieved based on the P3P specification that the Policy Reference File is located at "well-known" location. The agent constructs the Policy Reference File's URL from the web site's URL, i.e., appending "/w3c/p3p.xml" to the end of the web site's URL. The agent fetches both the natural language policy and the P3P policy. The natural language policy is retrieved at the URL specified by the <policy> tag, and converted into a plain text file without any HTML tags, images and hyperlinks. While parsing the P3P policy file, every tag is checked against a decision tree representing the structure of P3P XML tags shown in Figure 2.

This is used to determine whether appropriate tags that correspond to the PIPEDA privacy principles are present and with appropriate content, that is with further appropriate embedded tags. For instance, checking the P3P policy located at http://www.ibm.com determines that it has appropriate tags for 7 of the PIPEDA principles and the output is shown in Figure 3.



```
<terminated> PageSearch [Java Application] C:\Program Files\Java\jre1.5.0_04\bin\javaw.exe (May 29, 2006 8:24:23 PM)
Please enter the web site's URL:
http://www.ibm.com
This P3P policy has mentioned all the necessary P3P tags
The natural language policy is available at: http://www.ibm.com/privacy/
If you want to opt-out, please go to: http://www.ibm.com/privacy/
```

Figure 3: Checking PIPEDA Principles using CML Tags.

3.2 Classifying Natural Language Privacy Policy

To check that the fetched natural language policy addresses the PIPEDA's principles of Accountability, Accuracy, and Safeguards, two well known classification algorithms were chosen, Support Vector Machine (SVM) and decision tree. We report here on applying the method only for the Safeguards principle. Here we explore only the proof of concept and determine the baseline performance rather then finding the best classification algorithm for this particular application. SVM was chosen because for some classification applications, such as described in (Chen et al, 2004), it produced the best results as compared to a number of popular algorithms. Decision tree analysis was chosen for comparison purposes to a middle-of-the-road classification algorithm.

3.3 Data Collection and Labelling

Our data set consists of privacy policy files from Canadian organizations' web sites. We examined them and manually classified them to either having satisfied or not having satisfied the Safeguard principle. We then selected an equal number of positive (satisfying the Safeguard principle) and negative (not satisfying the Safeguard principle) privacy files, at 110 each. For labelling purposes, the following aspects were considered:

- Personal information should be protected by proper security safeguards to prevent loss, illegal access, unstated disclosure, or modification.
- When personal information is recorded on paper, organizations should have physical security methods, such as locking filing cabinets, controlling access to offices.
- Internally, an organization should have organizational security measurement, such as educating employees on privacy awareness and restricting access on a "need-to-know" basis.
- Organizations should utilize secure web technology, such as HTTPS, SSL, encryption and password authentication.
- When the retention period has expired, organizations should have appropriate procedures to dispose or destroy the collected personal information.

The files were pre-processed in the usual manner by converting all text to lower case, filtering with a list of stop words and stemming using Porter's stemmer algorithm (Porter, 1997). See (Zhang, 2006) for details.

3.4 Matrix Dimension Reduction

Both decision tree and SVM algorithms take a term matrix as input. The matrix is composed of rows representing data files in the datasets and columns representing word stems in all the positive training data files. Presence of a word stem in each data file is recorded by a binary value. Since the term matrix is the co-occurrences between word stems and data files, "the context for each word becomes the data file in which it appears" (Bellegarda, 1998). An important property of this term matrix is that two words with similar meaning are expected to appear in the same class documents (Bellegarda, 1998). Therefore, the term matrix was used as input to represent the training data files. In one test run, a 200 * 737 matrix was generated, where 200 indicates the number of the training data files, and 737 indicates the number of word stems in all of the positive training data files.

Principle Component Analysis (PCA) (e.g., (Partridge and Jabri, 2000)) was used to transform a large set of uncorrelated variables into a smaller set of correlated variables. It was used to identify any data patterns and reduce the input matrix dimensions. Since there are 737 (the number of word stems) columns in the matrix, in the worse case, 737 principal components could be generated by using the PCA algorithm. To visually identify the data set pattern, we only kept the first three principal components, which have the greatest contribution to variance and thus reducing the number of columns in the matrix from 737 to 3. That is, these three principal components explain the most important features related to the Safeguard principle in each data file. Projection of the matrix into three dimensions is shown in Figure 4.

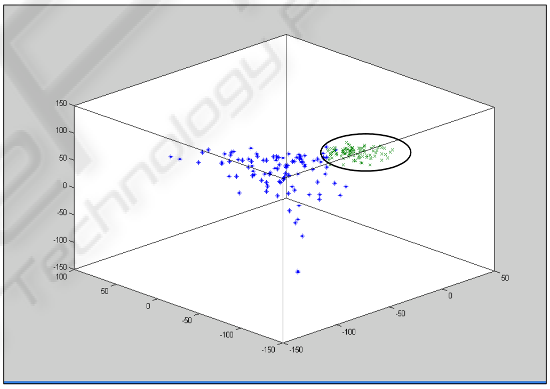


Figure 4: Projected dataset for 3 principal components.

There are two distinguishable clouds in the figure that represent the respective positive and the negative classes. (Since these clouds are clearly visible only when the figure is viewed in colour (clouds are in blue and green colours), an ellipsis has been added to the figure to identify one of the clouds.) It is obvious that the two classes are well defined by their own features. By further using a classification method, such as the SVM-based methods or decision tree based methods, these two classes can be identified.

3.5 Applying Classification Algorithms

We used SVM non-linear classification with RBF (Radial Basis Function) kernel function to separate our training data points into two classes. The SVM version of the program has two parts, i.e., training

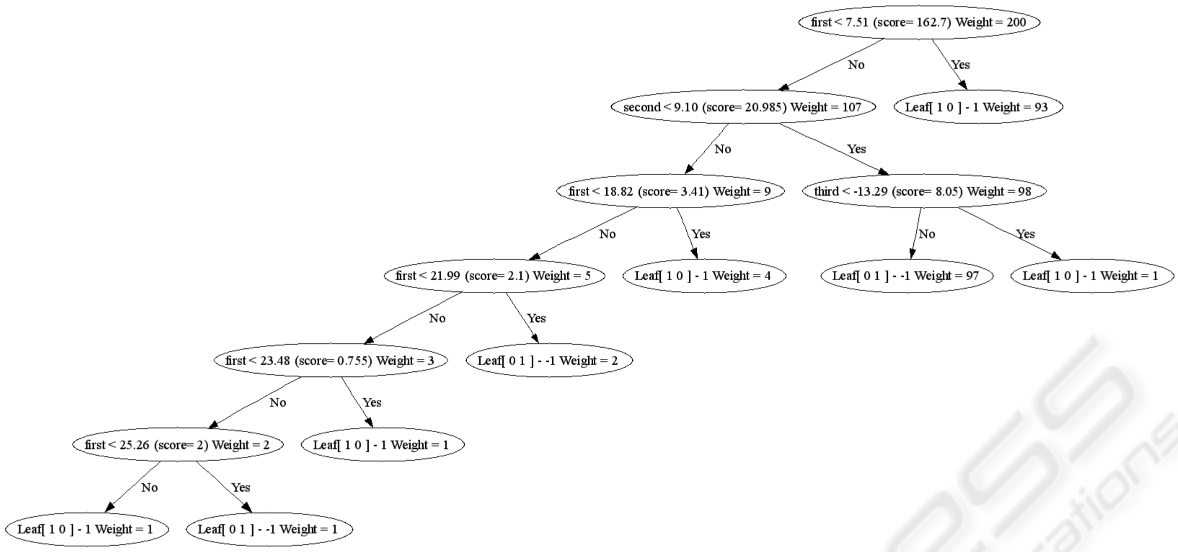

first < 7.51 (score= 162.7) Weight = 200
No
Yes
second < 9.10 (score= 20.985) Weight = 107
Leaf[1 0] - 1 Weight = 93
No
Yes
first < 18.82 (score= 3.41) Weight = 9
third < -13.29 (score= 8.05) Weight = 98
No
Yes
No
Yes
first < 21.99 (score= 2.1) Weight = 5
Leaf[1 0] - 1 Weight = 4
Leaf[0 1] - -1 Weight = 97
Leaf[1 0] - 1 Weight = 1
No
Yes
first < 23.48 (score= 0.755) Weight = 3
Leaf[0 1] - -1 Weight = 2
No
Yes
first < 25.26 (score= 2) Weight = 2
Leaf[1 0] - 1 Weight = 1
No
Yes
Leaf[1 0] - 1 Weight = 1
Leaf[0 1] - -1 Weight = 1

Figure 5: Decision Tree.

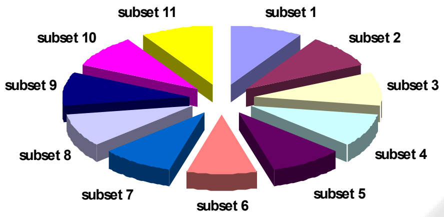


Figure 6: 11 Fold Cross Validation.

and predicting parts. In the training part, the program takes a matrix generated from the training dataset as input, and outputs an SVM model. In the predicting part, the program predicts unknown files using the SVM model. Training inputs a matrix of test data files and outputs (classification) labels for the test data files.

Our decision-tree version of the agent implementation takes a training data file matrix and a test data file matrix. To shrink the tree size, both matrices' columns are reduced to three dimensions using the PCA algorithm discussed above. Consequently, the time spent on building the tree and predicting unknown files is much shorter. While training the decision tree algorithm, the decision tree is built. The test data files are labelled by using the constructed decision tree. The outputs are the labels for the corresponding test data files. The decision tree built by the agent program is shown in Figure 5.

3.6 Validation

We used the K-fold cross validation technique (Leisch, Jain, and Hornik 1998), which is an improved version of the Holdout method, for both SVM and the decision tree versions of the classification algorithm. The method separates the dataset into K subsets, and the classification algorithm gets trained and tested K times. At each training and testing cycle, K-1 subsets together are used as training dataset, and the one subset left is used as testing dataset.

We chose K = 11 to make each slice contain exactly ten files. That is, the dataset is divided into 11 subsets, and the program executed with different training and testing files 11 times. At each execution, both versions of the program are trained with 100 positive data files and 100 negative data files, and tested with 10 positive data files and 10 negative data files, i.e., with a total of 20 files. Both SVM and the decision tree versions of the program were trained and tested by the same dataset for each test run.

Table 2: Example of Stem Words Frequency.

Appearance Frequency	Stem Words
414	inform
220	person
215	secur
173	protect
156	access
65	unauthor
64	encrypt
59	employe
52	provid
50	measur

At each round of training and test evaluation, the frequency of each stem word's presence in the positive training dataset is produced so that the input matrices can be generated. Table 2 shows the first 10 of the most frequent stem words for one instance of the execution.

The semantic meanings of these most frequent stem words are related to privacy and safeguard. For example, "access" and "unauthor" can be used in sentences restricting the access to the collected personal information. In addition, most of these stem words are the key words which were used while collecting and classifying the training and testing dataset.

The results of classification are shown graphically in Figure 7, with the overall average of 87.7% for SVM and 80% for the decision tree. As expected, the SVM algorithm performed better. As can be seen, however, both versions of the classification algorithm did not perform well at the 7th test run, at which the SVM version had 80% accuracy, while the decision tree version had 65% accuracy. In the case of this run, test files overlapped with each other and it was difficult to distinguish between them. The files were also located far from the concentrated area, which means that these files do not have as many features as other files located in the concentrated area.

4 RELATED WORK

The RDF-Group (2007) provides several complementary ontology classes for the legal domain: Actor (individuals and groups), Drama (events – both discrete and open-ended), Prop (Products and legal properties), Scene (place and time), Role, Script (document type), and Theme (topics of script or drama). OntoPrivacy (Cappelli et al, 2007) builds on Legal-RDF, reusing some of its classes, to model Italian privacy legislation.

A prototype implementation of one layer of the model for a privacy ontology for Canadian legislation was developed in (Jutla and Xu, 2004) using OntoEdit ver. 2.6.5, and Sesame. The prototype contains concepts and relationships pertinent to PIPEDA. To show proof-of-concept, the authors developed and successfully tested commonly used queries, such as "Does PIPEDA address privacy concerns about user monitoring?"

Gandon and Sadeh (2004), Jutla et al (2006), and Rao et al (2006) explore the use of semantic web technologies to support privacy and context awareness in e-commerce and m-commerce. They choose ontology languages, e.g. OWL and ROWL, to represent contextual information including privacy preferences. To achieve privacy compliance, a privacy compliant architecture, called Enterprise Privacy Architecture (EPA) (Karjoth and Schunter, 2002) has been proposed and extended (Karjoth et al, 2003). The privacy management framework proposed in (Anton et al., 2004) addresses privacy management problems which firms face, and which are not solved only by languages such as P3P and IBM's Enterprise Privacy Authorization Language (EPAL) (Backes et al, 2004) to support privacy policy enforcement within an enterprise.

5 SUMMARY / CONCLUSIONS

We show that the P3P privacy policy language cannot express all of the requirements of a privacy legislation, such as PIPEDA and, consequently, the privacy policies expressed in the natural language need to be examined. We also show that standard classification algorithms are useful in assisting the user to determine whether or not a privacy policy, expressed in natural language, satisfies particular requirements stipulated by a privacy law or regulation.

There are a number of laws/regulations that may be applicable to any of the users activity on web in which exchange of personal information occurs – for instance, privacy laws at the federal level and then at the state/provincial level, and yet, possibly, standards for a particular vertical business domain created by, say, a business association, may be applicable. To use our approach, for each law/standard, a mapping to P3P specification would need to be performed. Furthermore, for any privacy requirements, of a particular law/standard, which cannot be mapped to the semantic web's P3P specification, training of classification algorithms would have to be performed. Such training is a substantial task with a resulting classification algorithm's prediction accuracy that cannot be guaranteed. On the positive side, training activities need to be done only once while their results can be used by any user agents.

REFERENCES

Ackerman, M. S., Cranor, L., 1999. Privacy Critics: UI components to safeguard users' privacy. Computer Human Interaction, 1999.

Adams, A., 2000. Multimedia information changes the whole privacy ballgame, In *Proceedings of the Tenth Conference on Computers, Freedom, and Privacy: Challenging the assumptions.* Toronto, Canada, April 2000.

Anton, A. I., Bertino, E., Li, N., Yu, T., 2004. A roadmap for comprehensive online privacy policy. Technical report, CERIAS, Purdue University, West Lafayette, CERIAS-2004-47.

Backes, M., Bagga, W., Karjoth, G., Schunter, M., 2004. Efficient comparison of enterprise privacy policies. In *Proceedings of the 19th ACM Symposium on Applied Computing (SAC'04)*, pages 375–382, Nicosia, Cyprus.

Barth, A., Mitchell, J., 2005. Enterprise privacy promises and enforcement. *ACM 1-58113-980-2* Pages: 58 - 66

Bellegarda, J.R., 1998. A multispan language modelling framework for large vocabulary speech recognition. In *IEEE Transactions on Speech and Audio Processing*, 6, 5, 456-467.

Cappelli, A., Bartalesi Lenzi, V., Sprugnoli, R., Biagioli, C., 2007. Modelization of Domain Concepts Extracted from the Italian Privacy Legislation. In *Proceedings of 7th International Workshop on Computational Semantics*, 10-12 January 2007, Tilburg, Netherland

Chan, Y., Culnan, M., Greenaway, K., Laden, G., Levin, T., Smith, J., 2005. Information Privacy: Management, Marketplace, and Legal Challenges. In *Communications of the Association for Information Systems* (Volume 16, 2005) 270-298.

Chen, H., Zhou, H., Hu, X., Yoo, I., 2004. Classification comparison of prediction of solvent accessibility from protein sequences. In *Proceedings of the Second Conference on Asia-Pacific Bioinformatics, Australia*, 333-338.

Chellappa, R., Pavlou, P.A., 2002. Perceived information security, financial liability, and consumer trust in electronic commerce transactions. *Journal of Logistics Information Management,* Special Issue on Information Security.

Cranor L.F., 2003. P3P: Making Privacy Policies More Useful. In *IEEE Security & Privacy*, November-December 2003 (Vol. 1, No. 6), pp. 50-55.

Gandon, F., Sadeh, N., 2004. Semantic Web Technologies to Reconcile Privacy and Context Awareness. In *Web Semantics Journal*, Vol. 1, No. 3, 2004.

Hogben, G., 2003. A technical analysis of problems with P3P v1.0 and possible solutions. *Position paper for "Future of P3P" Workshop*, Dulles, Virginia, USA.

Jutla D.N**.**, Bodorik P, Zhang Y., 2006. PeCAN: An Architecture for Privacy-aware Electronic Commerce User Contexts, *Information Systems*, (Elsevier), June 2006, Vol. 31, Issue 4-5, pp. 295-320.

Jutla D.N., Xu L., 2004. Regulatory Privacy Agents and Ontology for the Semantic Web. In *SIGABIS Minitrack, Americas Conference on Information Systems, AMCIS 2004*, Special Interest Group on Agent-based Information Systems, 8 pages.

Karjoth, G. and Schunter, M., 2002. A privacy policy model for enterprises. In *Proceedings of the 15th IEEE Computer Security Foundations Workshop*.

Karjoth, G., Schunter, M., and Herreweghen, E. V., 2003. Translating privacy practices into privacy promises - how to promise what you can keep. In *Proceedings of the 4th IEEE International Workshop on Policies for Distributed Systems and Networks (POLIC'03)*, Lake Como.

Leisch, F., Jain, L.C., and Hornik, K., 1998. Cross-validation with active pattern selection for neural-network classifiers. *IEEE Transactions on Neural Networks,* 9, 1, 35-41.

Partridge, M., Jabri, M., 2000. Robust principal component analysis. In. *Proceedings of the 2000 IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing*, 289-298.

Privacy Commissioner of Canada, 2000. Personal Information Protection and Electronic Documents Act (PIPEDA). Available from http://laws.justice.gc.ca/en/P-8.6/text.html; accessed 20 March 2007.

Porter, M. F., 1997. An algorithm for suffix stripping. In *Readings in information Retrieval*, K. Spark Jones and P. Willett, Eds. Morgan Kaufmann Multimedia Information And Systems Series. Morgan Kaufmann Publishers, San Francisco, CA, 313-316.

Rao, J., Dimitrov, D., Hofmann, P., Sadeh, N., 2006. A Mixed Initiative Framework for Semantic Web Service Discovery and Composition. In *Proceedings of the IEEE International Conference on Web Services (ICWS 2006)*, Sept. 2006.

RDF-Group, 2007. Accessed March 30, 2007; http://www.hypergrove.com/legalrdf.org/index.html

Zhang, Z., 2006. Compliance of Privacy Policies with PIPEDA. M. Comp. Sci. thesis, Dalhousie University, Halifax, Nova Scotia, Canada