2011 Sixth International Conference on Availability, Reliability and Security

Learning Privacy Preferences

Inger Anne Tøndel, Åsmund Ahlmann Nyre, Karin Bernsmed
SINTEF ICT
Trondheim, Norway
{*inger.a.tondel, asmund.a.nyre, karin.bernsmed*}*@sintef.no*

***Abstract*—This paper suggests a machine learning approach to preference generation in the context of privacy agents. With this solution, users are relieved from the complex task of specifying their preferences beforehand, disconnected from actual situations. Instead, historical privacy decisions are used as a basis for providing privacy recommendations to users in new situations. The solution also takes into account the reasons why users act as they do, and allows users to benefit from information on the privacy trade-offs made by others.**

I. Introduction

The Platform for Privacy Preferences (P3P) [1], accompanied by A Privacy Preference Exchange Language (APPEL) [2], allowed users to have their privacy preferences automatically matched against a web site's privacy policy. With APPEL it was envisioned that privacy agents, i.e. software applications acting on behalf of the user, could perform automatic matching of preferences against policies. The privacy agent would then subsequently inform the user of the matching result or potentially even block access to the site in the event of a mismatch. Privacy agents could e.g. be implemented as core browser functionality, or as an add-on. Several privacy agents have since then been developed [3][4][5], but despite the efforts, none of these agents have gained any major impact in the user community. Although there are several reasons for this, we will in this paper focus on the importance of privacy preferences. We argue that previous solutions are unable to correctly represent the users' real intentions and that this is rooted in the users' difficulties to specify their privacy preferences. If the pre-specified preferences held by the agent do not match the users' real intentions regarding sharing of personal data, the foundation of the privacy agent is invalid. The users will perceive the warnings and recommendations received from the agent as meaningless, or even annoying, and will most likely end up turning it off.

Privacy is in many ways a subjective concept. Users have quite different opinions on what is acceptable and what is not [6], and thus privacy enhancing technologies that are able to adapt to varying privacy preferences are more likely to be useful to a larger group of users [7]. Identifying the preferences and making them available in machine-readable form is however challenging [8], as privacy decisions are complex and commonly based on a trade-off between the perceived benefit and perceived cost of sharing personal information. Hence, opinions will not only vary between users, but also depend on the context in which the trade-off assessment is being made. Specification of preferences is further complicated by

- *the privacy paradox*; users may claim to be worried about online privacy, but this is rarely reflected in their actual behaviour [9], and
- *the difficulty of understanding privacy technology*; experiments [10][11] show that misconceptions are common, even among educated users.

It is also important to note that the way people are expected to manage their privacy online is quite different from the way privacy decisions are made in real life [12]. Users are not accustomed to explicitly stating their privacy attitudes and catering for all the possible privacy trade-offs that they may encounter, and therefore have difficulties specifying their privacy preferences [13]. Furthermore, if users neglect to maintain their preferences over time, any recommendations made based on them can be difficult to understand [6].

Existing solutions to privacy preference specification can be divided into five categories:

- *General:* The specified privacy preferences apply regardless of the situation. This approach is common for browsers' privacy settings (e.g. for cookie management). Another example is the privacy agent AT&T Privacy Bird [3] where preferences can be specified either directly in APPEL or via the user interface.
- *Roles:* Preferences are defined for each role users may take on in their interaction with service providers. An example is the role-centred paradigm [5] considered in PRIME (Privacy and Identity Management for Europe)[1] where users state their current preferences by selecting a role.
- *Sites:* Preferences are specified for different types of sites in order to be able to account for variations in the users' expectations for e.g. web shops compared to social networks. An example is the approach suggested by Kolter and Pernul [4] where users can specify preferences for twelve service types.
- *Relationships:* For users it may be easier to state their preferences regarding individual sites as opposed to types of sites, as the latter are more abstract and

[1] http://www.prime-project.eu

978-0-7695-4485-4/11 $26.00 © 2011 IEEE
DOI 10.1109/ARES.2011.96

therefore potentially more challenging. This also allows users to take into account issues such as reputation and trust. Within PRIME, both the relationship-centred and the TownMap-based paradigms [5] utilise relationships with sites. The relationship-centred paradigm extends the role-centred paradigm so that roles can be associated with sites, while the TownMap-based paradigm visualises relationship-based preferences graphically.

- *Communities:* Communities can serve as a basis for providing users with suggestions and recommendations regarding privacy. An example of such a system is Acumen [14] where other users' behaviour is used to advise users to accept or block cookies. Another example is the suggestions of Kolter et al. [15] on using communities for exchanging privacy-related information on service providers, and also to share privacy preferences.

As privacy decisions are context sensitive and take the form of trade-off assessments, it is very difficult to specify general privacy preferences that correspond to the actual privacy attitudes in real-life situations. Though the role-, site- and relationship-based approaches are more adaptable to the situation, they fail in their expectations to the users. Users need to understand the abstract privacy preferences associated with roles or sites so that they can specify upfront what are their privacy attitudes in various roles or relationship. Alternatively, they must be privacy conscious enough to remember to update their roles or otherwise change their preferences when they visit new sites. Communities can support users in their privacy decisions, and provide prespecified privacy preferences that users can reuse, but these must be understood. The quality of the preferences and the recommendations shared via such a community is also an issue, as users' competence regarding privacy is in general low [10][11].

In this paper we present a new approach to privacy preferences that is based on machine learning. In Section II we provide an overview of the architecture of our solution. Section III describes how the solution works as seen from the user. The machine learning approach is discussed in Section IV, and Section V concludes the paper.

II. A MACHINE LEARNING APPROACH TO PRIVACY PREFERENCES

Imagine that privacy agents are able to infer users' privacy preferences, without explicit statements by the users upfront. In this section we present our solution for utilising machine learning algorithms [16] to efficiently capture user preferences.

A. CBR and collaborative filtering

Before we move on to describing our solution, we give a very brief introduction to the concept of Case-Based Reasoning (CBR) [17] and collaborative filtering [18]. CBR emerged as an alternative to the rule-based expert systems in the late 1980's to cater for the situations where rules were hard to define or changed rapidly. Instead of using prespecified rules, CBR resembles a form of human reasoning where previous experienced situations (cases) are used to solve new ones. The key idea is to find a stored case that closely resembles the problem at hand, and then adapt the solution of that problem. Often, CBR is used in combination with a rule base such that whatever domain knowledge is available a priori can strengthen the reasoning process, especially when the case base is scarce. Another important feature of CBR systems is their ability to provide explicit information on the reasoning underlying the output. Some systems even let users manually tune the reasoning steps to correct any mistakes made in the argumentation [19] [20]. Bayesian networks have proven successful in e.g. spam filtering [21], but does not have these explanatory capabilities. This is why we presume CBR to be a better alternative for reasoning about privacy preferences.

Collaborative filtering is common in recommender systems and are popular in order to exploit community knowledge and experience. The idea is to utilise similar users' ratings to predict the current user's rating of a subject. The algorithm has been popularised by Amazon[2] through the phrase *"people who bought this item also bought..."*. The key here is not to recommend what most users have done, but what most *similar* users have done.

B. Preference learning architecture

Our solution consists of a local privacy agent running on the client platform, and a community portal at some external location as depicted in Figure 1. When the user visits a website, the machine-readable policy of this website is used together with the current context as a basis for providing the user with recommendations as to whether or not to share personal information with this site. The reasoning engine evaluates the current case (location, current role, time of day, site/service, provider and privacy policy) towards the historical cases in order to identify the cases that are most similar to the current situation. These most similar cases are then used to come to a conclusion on what recommendation to give the user. Being able to identify similar cases is critical to the success of our solution. If the cases selected as a basis for the user recommendation are not relevant, the advice the user gets will likely be irrelevant and the agent will be useless. The prevailing retrieval algorithm for CBR systems is K-Nearest Neighbour (KNN) [16], which requires a definition of what is consider the nearest case. Expert knowledge is used to provide such definitions through similarity metrics and domain knowledge [22].

In the event that the reasoning engine's assurance falls below a configurable threshold it requests a recommendation from the community portal. This portal is assumed to be

[2]http://www.amazon.com

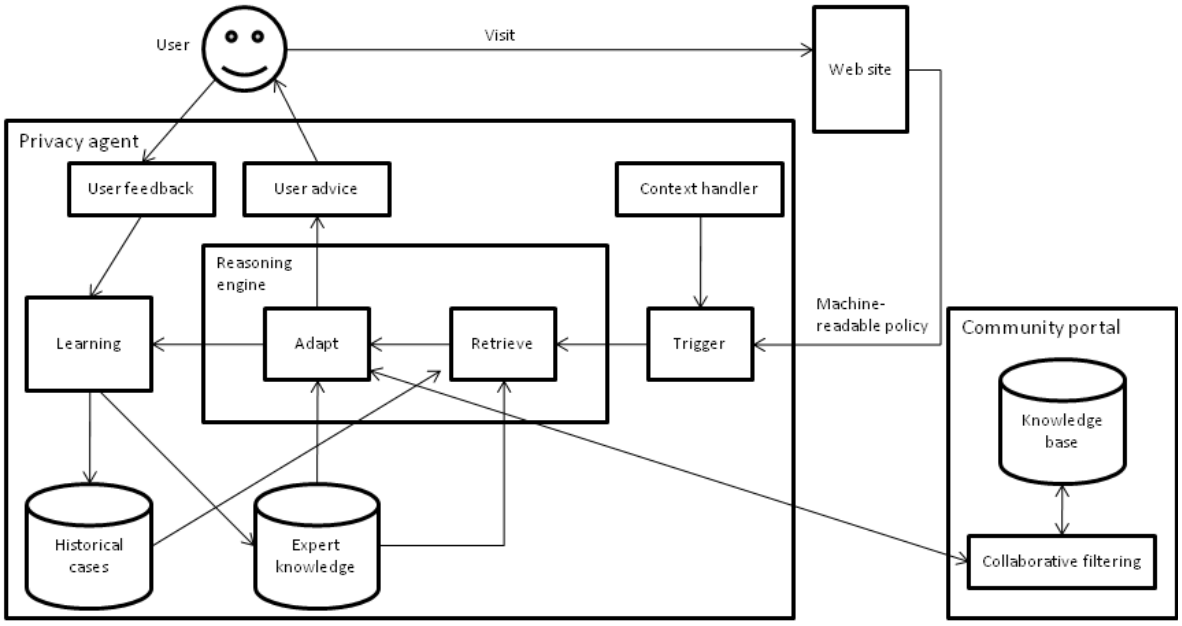

Figure 1. Preference learning architecture

a stand-alone service that provides additional input to the privacy agent's reasoning process whenever required. In order to receive a recommendation from the portal the privacy agent must input a set of experienced cases that can help determine what kind of user it is. The cases are given anonymously and do not contain privacy sensitive information (e.g. exact location) that can be used to identify the user. All cases contributed to the community portal help grow the community knowledge base and will potentially increase the precision of future requests.

A key property of our solution is the agent's ability to explain its reasoning and receive corrections from the user. That is, the user can request an explanation from the agent and for example see that the agent bases its decision on the fact that the user previously has rejected a similar service with a somewhat similar privacy policy. The user can then correct this by adding that the service is meant for professional use such that the re-evaluation can utilise the knowledge that privacy protection is less important during working hours and arrive at a new conclusion. It is this white-box interaction between the user and the agent that constitutes the actual preference learning and can be utilised to create new rules, exceptions, and update the expert knowledge that contribute to the privacy decision.

C. Case validity

A fundamental problem of all machine learning approaches is to allow for dynamic knowledge. What was true yesterday, may not be true today. In a privacy setting, alterations to the services, their privacy policies or how they are used may falsify the stored cases both locally and at the community portal. Thus, the knowledge base may rather quickly be filled with invalid cases, potentially reducing the benefit of the tool. Therefore, our solution requires cases to be re-evaluated whenever the basis for the decision is changed. For example, if the privacy policy of a service changes, the recommendation made by the privacy agent is re-evaluated at next visit to ensure that the decision always is made on the basis of the most current information. Similarly, if the user suddenly changes attitude and becomes stricter on privacy requirements, all cases in the knowledge base need to be reassessed. The actual reassessments are made the next time the user decides to interact with the site or service in question. It is optional whether re-evaluations cause deletion of the original case, since the history of decisions may be utilised later to solve new problems. However, to prevent over-filling the knowledge base such that the matching process is slowed down, we believe that deleting cases is important. For the community portal, deletion is the only option since cases are contributed anonymously. Hence, the portal deletes stored cases after a predefined time interval (e.g. one year).

III. Example usage scenarios

To illustrate how preference learning is experienced by users, we now describe a few usage scenarios. Note that not everything included in the scenarios is supported by our solution which only concerns generation of preferences.

Anna visits a website she has not visited before while using the privacy agent. The agent tries to retrieve various information on the website, like its machine-readable privacy policy, the provider and the type of service it offers. The

agent also has an overview of Anna's current context, e.g. the device used, her location and the time of day. Then the agent compares its knowledge of the website and the context, with its knowledge of previous user behaviour, expert advices and the historical behaviour of similar users.

Preference generation happens in the context of a privacy agent, and historical cases are used together with other factors to determine what privacy advices to make. The intended users are privacy conscious enough to install a privacy agent, and can thus be expected to be willing to take part in preference generation.

In this case, the agent warns Anna that the privacy policy of the website allows wider sharing than what Anna has been known to accept in the past. Anna explains to the agent that she will accept the policy since the service offered is very important to her. The agent subsequently records the decision and explanation to be used for future reference.

Users are allowed to specify preferences in a very concrete way. The need for abstract thinking is reduced as users make decisions regarding actual sites and actual privacy policies. They also build preferences during normal interaction with the system, as opposed to upfront. Preference specification is however, also with this solution, dependent on user interaction and assumes that users understand the choices they make. If not, the solution runs the risk of only learning common user behaviour and not the actual privacy preferences. To account for this risk, users can inform the agent about the adequacy of its recommendations and the reasons for the users' privacy decisions in a user-friendly way. Note also that machine learning techniques are only used for generating preferences, and not for automatic sharing of personal information. The users stay in control of their information sharing.

Anna has a few sites she visits regularly and that she has already shared data with, and she generally accepts their privacy policy. Now one of these sites has changed its policy in a way that makes it substantially worse privacy-wise. When Anna re-visits this website, the privacy agent detects that the policy has been changed and warns Anna.

There are a number of reasons why users may accept a privacy policy they determine as bad. Users may e.g. find that there are no other good options available, or that characteristics of the current situation (e.g. in a hurry) make it necessary to make an exception. In addition users may not understand a given privacy policy or may not have time to think it through properly. In this case Anna may find that she wants to continue using this site because she has become accustomed to it, although she does not approve the policy changes. If this is the case, the agent need to take this into account and not use this decision to automatically accept all policies of this type in the future, as this will not comply with Anna's intentions. Our solution provides the necessary functionality for assuring this.

Anna is a frequent user of social network sites, but has lately become increasingly concerned about the privacy implications of using them. She is now considering changing her behaviour, but the privacy agent is not aware of this change in attitude and will not warn Anna about the privacy policies of social network sites as she has already accepted a large amount of bad policies from such sites. Anna is not aware of this "rule". However, next time she visits a social network site, her new privacy concerns make her utilise the opportunity offered by the agent to tell her about the privacy policy of the site and the reason why the agent expects her to accept the policy. She corrects the agents reasoning by stating that social network sites should not be automatically accepted from now on. The agent then computes a new recommendation for Anna.

Users can check the reasoning behind all recommendations made by the agent, and make corrections. In the case that users become more restrictive in what they are willing to accept, this change in attitude needs to be communicated to the agent as it will be almost impossible to detect automatically. The suggestions in the previous section regarding deletion of historical cases will also make the agent better able to cope with changing user preferences.

IV. Discussion

Our solution only addresses preference generation, and does not intend to solve all problems related to privacy agents. Thus we for instance do not go into details about how to assure user-friendliness in the interactions with the agent. Instead we build on existing user-friendly solutions for communicating privacy information to users, e.g. using icons [23] or tabular forms [24] for presentation of privacy policies.

We are not aware of any previous work on privacy preference generation that utilise machine learning of user behaviour in a privacy agent context. Berendt et al. [25] have however suggested that future privacy enhancing technologies should go in this direction, and studies performed by Sadeh et al. [8] show that machine learning techniques have the potential of generating more accurate preferences than users themselves. The most closely related work that we are aware of is that of Kelley et al. [26] on preferences for a mobile social network application and that of Bufett and Fleming [27] on preference modelling for eliciting preferences. The work of Kelley et al. is limited to preferences for sharing location with friends. The preferences are less complex than we envision, and machine learning techniques are used to suggest incremental changes to user-specified preferences. The work of Bufett and Fleming is in many ways similar to specifying general or site type specific preferences upfront, only that reasoning techniques are used to extract preferences.

Compared to the existing approaches to preference specification (outlined in the Introduction), the machine learning

approach reduces the need for users to understand the preferences as these are generated automatically. It is however of paramount importance that users understand the privacy implications of their behaviour, and thus users need to understand the privacy policies and the alternative actions they may take based on policies. In addition, users need to interact with the agent to inform it of the reason for their actions. The solution is however not dependent on users themselves remembering to make the necessary updates, and the input required is concrete and meet Lederer's recommendation that users should "practice privacy as a natural consequence of their normal engagement with the system." [13]

Making privacy understandable is a key to success for all privacy preference approaches. Understandability may be improved by offering better support for the cost-benefit trade-offs of users when it comes to privacy. Current solutions for preference specification only focus on the costs, i.e. what users do not want to share and how they do not want their information to be treated. Agents that also consider benefits will be able to make recommendations based on what users are given in return for their information, thus offering more relevant user advice. Cost-benefit can be said to be indirectly supported in e.g. the role- and relationship-centred approaches, as the benefits offered are likely to influence the roles selected for various sites. In the machine learning approach cost-benefit aspects are included by users providing reasons for their actions. Users are also informed of what other users have done, indirectly indicating whether similar users found that the benefits were worth the costs.

Solutions that utilise machine learning techniques for privacy preference generation face the following main challenges: The problem with black box configuration [8] and the potential mismatch between users' actions and their intentions [25]. The first challenge emphasize that users need to know and understand what the agent considers to be their current preferences. This is important for trust in the agent, and also for assuring that users are still in control of their privacy [8]. The second challenge is more fundamental. Learning preferences from user behaviour can result in not learning users' actual preferences but rather their common behaviour - something which is not the intention. Our strategies in order to cope with these challenges are to:

- Make agents able to provide reasons for their actions, and make users able to correct them.
- Provide users with easily understood descriptions of privacy policies and concrete privacy advices, so that they understand the implications of their actions.
- Have users share the reasons for their privacy decisions with the agent. As users provide a reason, they are more likely to reflect over why they act as they do, something that may result in behaviour more in line with their principles.

A key dilemma is finding the right level of user involvement. It is important to involve users in the learning process, but if users receive a lot of requests for feedback on preference modifications, this may be considered to be annoying interruptions and will result in the machine learning approach increasing user effort instead of reducing it. Though the solution targets users that are willing to invest some time and effort in their privacy, their willingness should not be over-estimated. To further reduce the risk of not learning preferences, it will be interesting to explore how machine learning can be used together with role and relationship, e.g. to suggest or perform role changes and modifications to relationships. More research is also needed on how to utilise the community in this respect, e.g. by including the role of experts.

As pointed out previously, there are fundamental problems with basing privacy decisions on community support. We recommend using the community only as one of the factors considered, and to use it to say something about similar users' behaviour, not the content of privacy policies. It is also important to fully address the privacy implications of the community approach. Increased privacy based on sharing information on user behaviour is in many ways a contradiction. Therefore, to prevent the community portal from becoming a privacy liability, we only store anonymous information in the knowledge base and limit the retention time.

We will continue our research on using machine learning techniques for privacy preference specification, with a focus on evaluating how the resulting privacy advices are perceived by the users. We are in the process of implementing a prototype of our solution, and plan to use this prototype in user evaluations of the approach.

V. Conclusion

As users in general differ in their privacy expectations, privacy agents that take users' privacy preferences into account are more likely to be useful for a larger user group. Users however find it difficult to express their privacy preferences in a way that makes them available to such agents. Current approaches to privacy preferences rely on users being privacy aware and able to reason about privacy in the abstract, if the resulting preferences it to correctly represent the users' privacy expectations in varying situations. In this paper we have suggested a new approach where the privacy agent is able to learn user preferences based on the privacy decisions users make in their normal interactions on the web. We argue that learning of privacy preferences has the potential to increase the accuracy of preferences, without putting too high requirements on the users' privacy knowledge, awareness and willingness to invest time and effort in their privacy.

ACKNOWLEDGMENTS

We would like to thank our colleague Martin Gilje Jaatun for contributing to the work that has led to the ideas presented in this paper.

REFERENCES

[1] "W3C. Platform for Privacy Preferences. http://www.w3.org/P3P/."

[2] L. Cranor, M. Langheinrich, and M. Marchiori, *A P3P Preference Exchange Language 1.0 (APPEL1.0)*. World Wide Web Consortium, 2002. [Online]. Available: http://www.w3.org/TR/P3P-preferences/

[3] L. F. Cranor, P. Guduru, and M. Arjula, "User interfaces for privacy agents," *ACM Trans. Comput.-Hum. Interact.*, vol. 13, no. 2, pp. 135–178, 2006.

[4] J. Kolter and G. Pernul, "Generating user-understandable privacy preferences," in *Proceedings of the Fourth International Conference on Availability, Reliability and Security (ARES 2009)*, 2009, pp. 299–306.

[5] J. S. Pettersson, S. Fischer-Hübner, N. Danielsson, J. Nilsson, M. Bergmann, S. Clauss, T. Kriegelstein, and H. Krasemann, "Making PRIME usable," in *SOUPS '05: Proceedings of the 2005 symposium on Usable privacy and security*, 2005, pp. 53–64.

[6] S. A. Bagüés, L. A. R. Surutusa, M. Arias, C. Fernández-Valdivelso, and I. R. Matías, "Personal privacy management for common users," *International Journal of Smart Home*, vol. 3, no. 2, pp. 89–106, 2009.

[7] M. S. Ackerman, L. F. Cranor, and J. Reagle, "Privacy in e-commerce: examining user scenarios and privacy preferences," in *EC '99: Proceedings of the 1st ACM conference on Electronic commerce*, 1999, pp. 1–8.

[8] N. Sadeh, J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabaker, and J. Rao, "Understanding and capturing people's privacy policies in a mobile social networking application," *Personal Ubiquitous Comput.*, vol. 13, no. 6, pp. 401–412, 2009.

[9] S. Spiekermann, J. Grossklags, and B. Berendt, "E-privacy in 2nd generation e-commerce: privacy preferences versus actual behavior," in *EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce*, 2001, pp. 38–47.

[10] S. Flinn and J. Lumsden, "User perceptions of privacy and security on the web," in *Proceedings of the Third Annual Conference on Privacy, Security and Trust (PST 2005)*, 2005.

[11] R. W. Proctor and K.-P. L. Vu, "A multimethod approach to examining usability of Web privacy polices and user agents for specifying privacy preferences," *Behavior Research Methods*, vol. 39, no. 2, pp. 205–211, 2007.

[12] M. Ackerman, "The intellectual challenge of CSCW: The gap between social requirements and technical feasibility," *Human-computer interaction*, vol. 15, no. 2, pp. 179–203, 2000.

[13] S. Lederer, I. Hong, K. Dey, and A. Landay, "Personal privacy through understanding and action: five pitfalls for designers," *Personal Ubiquitous Comput.*, vol. 8, no. 6, pp. 440–454, 2004.

[14] J. Goecks and E. Mynatt, "Supporting privacy management via community experience and expertise," *Communities and Technologies 2005*, pp. 397–417, 2005.

[15] J. Kolter, T. Kernchen, and G. Pernul, "Collaborative privacy management," *Computers & Security*, vol. 29, no. 5, pp. 580–591, 2010.

[16] T. Mitchell, *Machine Learning*. McGraw Hill, 1997.

[17] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI Communications*, vol. 7, no. 1, pp. 39–59, March 1994.

[18] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web (WWW '01)*, 2001, pp. 285–295.

[19] B. Porter, R. Bareiss, and R. Holte, "Concept learning and heuristic classification in weak theory domains," *Artificial Intelligence*, vol. 45, no. 1-2, pp. 229–263, September 1990.

[20] A. Aamodt, "Knowledge-intensive case-based reasoning in Creek," in *Proceedings of the 7th European Conference on Advances in case-based reasoning*, vol. LNAI 3155, 2004, pp. 1–15.

[21] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A bayesian approach to filtering junk e-mail," in *AAAI Workshop on Learning for Text Categorization*, July 1998.

[22] I. A. Tøndel, Å. A. Nyre, and K. Bernsmed, "Towards a similarity metric for comparing machine-readable privacy policies," in *iNetSec 2011: Open Problems in Network Security (to be published)*, 2011.

[23] M. Hansen, "Putting privacy pictograms into practice - a european perspective," *INFORMATIK 2009 - Im Focus das Leben*, 2009.

[24] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder, "A "nutrition label" for privacy," in *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*, 2009, pp. 4:1–4:12.

[25] B. Berendt, O. Günther, and S. Spiekermann, "Privacy in e-commerce: stated preferences vs. actual behavior," *Commun. ACM*, vol. 48, no. 4, pp. 101–106, 2005.

[26] P. G. Kelley, P. H. Drielsma, N. Sadeh, and L. F. Cranor, "User-controllable learning of security and privacy policies," in *Proceedings of the 1st ACM workshop on AISec*, 2008, pp. 11–18.

[27] S. Buffett and M. Fleming, "Applying a preference modeling structure to user privacy," in *Workshop on Sustaining Privacy in Autonomous Collaborative Environments (SPACE 2007)*, 2007.