Privacy concerns around recommendation systems

Juan Bernabé Moreno
University of Granada, Department of Computer Science and Artificial Intelligence

Abstract— **This work introduces the privacy concerns around the recommendation systems and the risks about systems that rely on personal information from their users. The privacy preserving principles are presented as well as a set of guidelines on how to comply with them, and the technical challenges around those guidelines.**

***Index Terms*— Recommendation systems, privacy**

I. INTRODUCTION

The ultimate aim of a recommendation system, unlike a query based information retrieval is providing users with the information they want or need, without expecting from them to ask for it explicitly [1]. To serve to this purpose, the user to be given a recommendation needs to be categorized and classified according to similarities to other users and statistical information collated by tracking mechanisms that log the user interaction with the system. Thus, the information provided by a recommendation system could somehow compromise the privacy of its users.

The problem of privacy protection in recommendation systems takes up the contents of the catalogue of items or services to be recommended have an intrinsic commercial value and should be protected from competitors or bargainers who might be masquerading as potential consumer and therefore user of the recommendation system.

The privacy risks associated to a recommendation system depends very much on the method the information about the user is derived. The following section will describe the recommendation process model and will take up the main issues to address when designing a recommendation system, establishing a framework where we will explain the privacy concerns around recommendation systems.

II. RECOMMENDATION SYSTEMS TYPOLOGY

Figure 1 depicts the general model of recommendation process that will be taken as reference.

The recommendation seeker may actively ask for a recommendation or the recommendation system will be generating recommendations periodically or triggered by a given event without having to prompt the seeker. Seeker preferences are conveyed to the system. The recommender recommends an item the seeker will like, based on a set of known preferences – his/her own, the seeker's, and those of other people-.

Additionally, the recommender might identify people with similar interests the seeker may want to communicate with, and generalize recommendations for the group

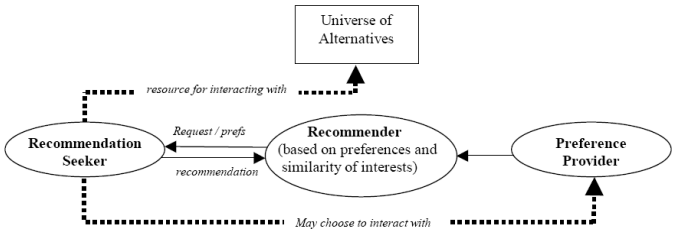

Figure 1

We can identify four main issues to characterize the design space for recommender systems. The issues concern preferences, roles and communication, algorithms, and human-computer interaction [2]

Preferences

Recommendation is based on preferences. Thus, an automated recommender system must obtain preferences from people concerning the relevant domain: whose preferences are used? And how are they obtained? Do recommendation users have to express their own preferences as part of the process of seeking a recommendation? What incentives are there for people to offer preferences? What is the form of a preference? How are preferences represented?

Roles & Communication

Do people play distinct roles, or do all users of a system play the same role? Are roles fixed, or do they evolve? How is the interaction between the recommendation user and the recommender initiated? Who initiates it? Is the recommendation directed to a specific person or is it broadcast to anyone who's interested? Is there opportunity for recommendation users to give feedback to the recommender? What information about the people whose preferences are used in computing a recommendation is revealed to the recommendation user? Is there an opportunity for communities of like-minded people to form? If information about preference providers is revealed, are any measures taken to safeguard privacy?

Algorithms for Computing Recommendations

How does an automated recommender system determine whose preferences to use in computing a recommendation? If we think of all the people who have expressed their preferences for a given domain as being placed in a large, multi-dimensional space, this is the problem of finding

neighbors in that space for the person seeking a recommendation. How are recommendations computed? For example, given that a set of neighbors for the recommendation seeker has been determined, how are the preferences of these neighbors weighted and combined?

Human-Computer Interaction

How are recommendations presented to the person who sought them? The most simple and common example is an ordered list

III. Major Types of Recommender Systems

The approaches can be distinguished by which of the four main issues they focus on, and how they address the issues.

Content-based systems use only the preferences of the seeker; they attempt to recommend items that are similar to items the user liked in the past. They are focused on algorithms for learning user preferences and filtering a stream of new items for those that most closely match user preferences.

Recommendation support systems do not automate the recommendation process; thus, they do not have to represent preferences or compute recommendations. Instead, they serve as tools to support people in sharing recommendations, helping both those who produce recommendations and those who look for recommendation.

Social data mining systems mine implicit preferences from computational records of social activity, such as Usenet messages, system usage history, citations or hyperlinks. These systems also have focused on the human-computer interface (HCI) issues involved in visualizing the results. These visualizations often have been presented to aid the navigation of information spaces like the World Wide Web; this helped motivate the term *social navigation*.

Collaborative filtering systems require recommendation seekers to express preferences by rating a dozen or two items, thus merging the roles of recommendation seeker and preference provider. These systems focus on algorithms for matching people based on their preferences and weighting the interests of people with similar taste to produce a recommendation for the information seeker.

The mechanisms to compromise the privacy depend on the kind of recommendation system.

IV. Privacy risks typology

The fact that the user is treated in a personalized way, which is the principle the recommenders rely on, poses a number of risks to the user privacy:

A. Unsolicited marketing [3]

Even if the consequences of unsolicited marketing are not that severe than other potential privacy risks, users are very concerned when targeted by a direct marketing campaign where they are offered products surprisingly aligned with their preferences.

The fact that the information they entered in e-commerce transactions may be used for targeted advertising, or even worst, sold to third parties that may want to market them, makes users very skeptical and mistrusting.

B. Recommendation is too accurate and potentially wrong

Moreover, user may get frustrated if she is given automatic predictions about her habits or interests that are off base and may lead somebody who finds them to draw erroneous conclusions about her (e.g.: if the user is recommended certain movie evidencing certain sexual orientation)

Jeffrey Zaslow starts his article *If TiVo Thinks You Are Gay,Here's How to Set It Straight [4]* as follows: *"Basil Iwanyk is not a neo-Nazi. Lukas Karlsson isn't a shadowy stalker. David S. Cohen is not Korean.*
But all of them live with a machine that seems intent on giving them such labels. It's their TiVo, the digital videorecorder that records some programs it just assumes its owner will like, based on shows the viewer has chosen to record."

Predictions that are too accurate and infers recommendations based on information user is not aware of having provided, may compromise the system acceptance.

C. Collected user data – user privacy perceptions mismatch

Additional concerns arise when there is a mismatch between users' perceptions about privacy and the types of data collection and use that actually occur [5]

Companies use this information to profile customers, which allows for pricing discrimination.

Even if theoretically seen, price discrimination would be a win-win situation for both consumer and sellers, consumer are often quite concerned to be profiled.

It is proven that users are concerned when they are treated differently from others (eventually being charged with higher prices) and with the amount of private information each transaction has associated.

D. Private information filtration to people in the environment

When private information is stored in a computer, the risk of being inadvertently revealed to other users of this computer is high (e.g.: cookies granting access to user's profile or authentication).

Frequently, having access to user's profile is the key for getting access to other accounts and sites

When private information is stored in a computer, the risk of being inadvertently revealed to other users of this computer is high (user account, personalized web site, etc).

Depending on the person gaining the access to the computer, the risk can be higher. Thus, if a member of your familiar environment sees what you purchased for its birthday, although the surprising factor is no longer there, the consequences are not that grave.

If rather stalker and identity thieves gain access to a user's profile, the implications might be immense.

Information derived from internet records –what someone has read, posted, eaten, purchased, downloaded, etc- is already being used in a lot of lawsuits situations (child custody, patent disputes, etc), and this information can certainly tip the scale in favor of the prosecution or of the defense. But personalized sites and among them, recommendation systems go far beyond: the information is not only the result of the navigation tracking, but preferences the user has proactively provided

(e.g.: users rating one item might reveal political interests or even more criminal proneness). Moreover, the information stored in personalized sites is not that volatile like the internet records use to be.

E. Geography location: I know where you were last summer

The ultimate goal of all the emerging mobile technologies is getting the user always connected, no matter where she is, or which device she is connecting from.

Applications that used to keep information about user activities and networking (calendar, address book) now take advantage of the user's precise physical location (by means of GPS data, etc). That substantially increases the privacy risks we have mentioned before.

F. Shilling

In June 2001 Sony Pictures had to admit, that quotes of non-existent movie critics had been used to promote newly released movies.

Ideally, items that are highly appreciated by the user community should be recommended more often than others, and their likelihood to be purchased is accordingly higher.

The economical benefit may lead to unscrupulous producers to seek for a mechanism to influence recommender systems in a way that their items are recommended to users more often. A recommendation system can be influenced by a group of users (*shills)* –automatic or human- that vouch for a given item in question. Shills false ratings are intended to mislead other users.

Shilling costs time and money by recommending bad items for the user, for the operator, as the level of trust decreases and for the retailer.

Recommenders based on collaborative filtering, where users are recommended items highly rated by users with similar tastes are especially sensible to shilling.

1) Recommendation algorithms

At a very high level, the recommendation algorithms identify these people that are similar under the assumption that the items these people enjoyed are the ones you will be also very likely to enjoy.

They can work in a prediction mode by predicting how much the user might like some items or a set/category of items, or in a recommendation mode, producing an shorted list of user the user might like.

Most of the algorithms behind almost all collaborative filtering recommenders are based on the statistical classification algorithm called k-Nearest-Neighbor.

This algorithm was first used by Resnik et all in 1994 [9] for a collaborative filtering recommender. Since them, the algorithm has been improved and 2 versions have appeared:

- *User-user*: with user-user based systems, a recommendation for an item depends on the recommendations of users with the same taste. A user-user recommender system has to select those rows in the user-item matrix that are most equal to the row representing the ratings of the user to make a recommendation for. Based on the recommendation of the selected users for the unrated item, and the similarity between the user and the selected users, a prediction can be made for the value the user under investigation would give to the item. The final recommendation is given to the user by ordering the list of empty items according to the expected value the user would give to it. To speed up the algorithm, a clustering technique can be used to first cluster users with similar taste instead of searching the whole matrix for similar users.
- *Item-item*: with item-item recommendation, a recommendation depends on how this item is related to other items in your taste. An item-item recommender system measures the equality between all ratings given to items, instead of the equality between users in user-user recommendation. For example, if a there is a perfect correlation between two items *a* and *b* (which means: everyone who likes *a* also likes *b* and vice versa) and user x likes a, it will be likely user x will also like item b. The correlations between all items are computed and stored in a model. If a user wants a recommendation, for all empty entries in the row representing the users ratings, the correlation with the other items are looked up in the model and based on the rating on the correlated item, a prediction of the rating of the user for the empty entry can be made. The final recommendation is again provided by sorting the list of empty entries according to the prediction of the value the user would give to it. Item-item based recommendation is often used in ecommerce and marketing, with the main reason people tend to buy something that is related to another product earlier bought that did meet the expectations.

.

Shyong and Riedl established a set of hypotheses about the shilling concerns for a recommendation system [8] which will allow qualifying better them.

They stated that the attack effects vary on the type of collaborative filtering algorithm (user to user, item to item, etc), and their impact depends directly upon the algorithm (if shilling affects only the last elements given in a recommendation the user might not ever read is not comparable with affecting the top 10)

Shilling attacks affect recommender algorithms differently from prediction algorithms

Shilling attacks can't be detected using traditional measures of algorithm performance (like Mean Absolute Error (MAE) to evaluate the overall predictive accuracy), as attacks can be subtle and focused enough that their overall effect on the system is minimal. Thus, one possible place to look for detecting shilling attacks is to understand which target items are most vulnerable to them, in other words, the number of ratings of an item, and the spread of those ratings over possible ratings values determine how effective an attack can be.

The rating distribution can be modeled by means of three variables: *likability* or average ranking, *popularity* or number of rankings and *entropy* or the variable that describe the ratings distribution.

The impact of those variables can be qualified as follows:

- The higher the likability of one item (the more well-liked an item is), the easier it is to cause that item to be recommended more often.
- The less the popularity of one item is, the easier it is to manipulate the predictions and recommendations for that item
- The higher the entropy of an item's ratings, the easier it is to manipulate the predictions and recommendations for that item

2) Shilling attacks

Shilling attempts consist of pushing (rising) or nuking (lowering the recommender's predictions for a given item). Usual attacks consist of introducing a set of new users into the system, trying to be similar to existing users (rating items others have already rated). Additionally, these new users rated the attack target item very highly to push it.

Both user-user and item-item implementations scale correlations according to the number of ratings in common. Thus, if the shill users rate too few items, the similarities will be scaled down and not considered by the algorithm. Therefore, to ensure the attack effectiveness, better rating the max number of items.

To read more about this topic, see [8]

V. Value of Information

The recommendation algorithms require as input certain information about the user the recommendations can be based upon. In general, the more information about the user, the better the recommendations are.

This presents a trade-off between getting the best recommendations and being more privacy invasive, that's why we should strive for a balance [16]

The accuracy of an algorithm with respect to the amount of information known about the user follows a diminishing returns curve. That is, once a certain amount is known about a user, obtaining further information is only marginally useful. This raises the possibility of finding a "sweet spot" that maximizes the recommendation accuracy per unit of information known about the user. [6]

Recommendation systems based on demographic information (ZIP code, age and gender) are not a big privacy threat, because the supplied information can't easily reveal user's identity.

On the other hand, recommendation systems based on collaborative filtering rely on much more personal information, what lead to pose privacy related question like the amount of information the user provides to recommend an item, if this information varies by item, or a quantification of the privacy lost vs the information gained by the recommendation system. Obviously, answers are highly dependant on the domain. Thus, a user may be more concerned sharing clinical information, than preferences about movies. The more harmful the information can be used against a user, the more concerned she is of sharing it.

Provided that you can calculate how useful a given piece of information is, then the system can be tuned to optimize its data collection process by choosing to solicit user preferences on items that carry the most value [7]

Ideally, quantifying the value of information would allow for deciding when the system is accurate enough and then stop collecting information from the user. Actually, going an step beyond would be thinkable, the system would be able to determine at which level the recommendations are accurate and at which level the user privacy is compromised.

VI. Privacy principles

The Organization for Economic Co-operation and Development (OECD) formulated the Guidelines on the Protection of Privacy and Transborder Data Flows of Personal Data [11] that are one of the best-known sets of fair information practice principles. Many other sets of guidelines and some privacy laws are based on these principles.

The eight OECD principles provide a useful framework for analyzing privacy issues related to ecommerce personalization:

1. Collection Limitation.
   Data collection and usage should be limited (personalization systems should collect only the data that they need, and not every possible piece of data that they might find a need for in the future)
2. Data Quality.
   Data should be used only for purposes for which it is relevant, and it should be accurate, complete, and kept up-to-date (data is meant to be used for relevant purposes, and not to make inferences that are irrelevant to the data). Additionally, the user should be provided with the ability for individuals to update and correct the information in their profiles.
3. Purpose Specification.
   Data controllers should specify up front how they are going to use data, and then they should use that data only for the specified purposes (users should be notified up front when a system is collecting data to be used for any other purpose). We will see in a next section how software tools make use of standard policies architectures (like P3P) to make users aware of privacy concerns.
4. Use Limitation.
   Data should not be used or disclosed for purposes other than those disclosed under the purpose specification principle, unless the user explicitly consents (data collected by personalization systems should not be used for other purposes without user consent.)
5. Security Safeguards.
   Data should be protected with reasonable security safeguards (encryption, secure transmission channels, etc)
6. Openness.
   Whenever the data collection and usage practices

start taking place, the user should be immediately and up front notified.

7. Individual Participation.
   Individuals should have the right to obtain their data (profiles, etc) from a data controller and to have incorrect data erased or amended.
8. Accountability.
   Data controllers are responsible for complying with the principles mentioned above.

VII. What can be done

A. *Usage of pseudonymous for the profiles*

User's identity is not always required to provide personalized services (which is typically the case of recommender systems). Users can be identified by pseudonyms or nicknames, so that the binding of nickname and the real life identity is not always manifested.

Actually, some combination of non-identifiable information contained in a pseudonymous profile with web usage logs might provide the real identity of one user.

For companies, it is much easier storing pseudonyms than real user names to comply with the privacy laws.

Going one step beyond towards improved privacy would lead applications to store internet records containing IPs, etc separately from pseudonyms. Moreover, internet logs should be scrubbed so that they don't contain information that would allow pseudonymous profiles to be linked with other data.

B. *User private data store at the client*

Shifting from the approach of storing the user profiles in the server to the one of storing the private information on the client helps reducing the privacy concerns.

For example, the usage of cookies, that are required by the server on the fly to perform personalization related tasks and are immediately discarded once the task has been completed. One key aspect is keeping the information encrypted to avoid requests being sniffed, or people having access to the client's machine or malware that looks for user cookies.

Another possibility is providing personalized services by means of client scripting, so that the privacy sensible information never travels to the server.

Canny proposed an architecture for a recommendation system in which participants compute a public "aggregate" of their data to share with members of their community. Individuals can then compute their own personal recommendations without revealing their individual data. He suggests that such an approach might be particularly useful in a ubiquitous computing setting where users may desire recommendations about everyday activities but are concerned that detailed information about their own activities not be revealed [13]

C. *Task-oriented personalization*

One approach based on a session cookie which is employed to store temporally information about the user, that can be deleted once the session is destroyed, ensures the compliance with the Collection Data principle, as only the required personalization data for the particular task the user is engaged in, is stored in a volatile way.

The key point relies on knowing which kind of data is required for a given task (we will discuss in more detail the Value of Information principle in a later section).

Actually, having at a time only the required information can contribute to improve the performance of the recommendation system, because the system doesn't provide recommendations based on one user's full profile information (e.g.: user's personal preferences might not be relevant for someone buying a gift for another person) Likewise, once a user completes a particular task, she may no longer be interested in receiving recommendations related to that task (e.g.: prompt interest in car dealers advertisement that goes away when the purchase has been done)

D. *Put the user in control*

Enabling that the user decides when this personalization takes place, which information should be collected, and for which purpose will ensure the privacy principles compliance.

Putting the user in control implies developing tools to allow user to control the information that is part of their profiles.

If the granularity of the tool allows for establishing privacy rules for each and every object in the system, using the tool would be for the user a very tedious task, which has to be performed again for each new object coming into the system. Thus, these tools should rather enable the users to specify policies that apply automatically to the objects that are encountered [14]

As we will see in the next section, where we will examine the current frameworks for specifying privacy policies, the task of formulating a privacy rule is quite tough, which assumes that the user has a deep understanding of privacy issues and the ability to foresee future activities that might concern her privacy.

The other side of the coin takes up the systems that give users access to their profiles, but users aren't either aware or interested in pro-actively customize their online experiences. [15] For example, many well-known recommendation systems, like Amazon.com, requires users to click on the Your Account Page link, and select from several items in a "Recommendations" section. Users can edit previous ranking they have given and review their transaction and rating history, and more important, specify which items should be excluded for further recommendations in the future. The problem is the one we have pointed above: granularity, as users have to make individual privacy decisions for every object in the system – time consuming activity-, and keeping the privacy settings up-to-date, as for each new purchase made by the user, a privacy settings update might be required.

One a priory suitable approach considering the trade-off of over-specifying and a well defined set of privacy rules would

be providing an interface where users define privacy policies that apply automatically (e.g.: certain categories of purchases should be excluded from recommendations, expiration of purchase history, use those items being purchased with the business credit card and not with the private one, etc). This approach could be combined with a mechanism to include the private policy specification as part of a given transaction (e.g.: when user indicated her shipping address and her credit card number, user might be asked if this transaction should be added to her profile. Moreover, the general privacy settings would apply automatically unless the user indicates otherwise at transaction level)

Typically, the relationship between physical person and recommendation profiles has a 1-to-n cardinality. For example, one user could have separate profiles for personal and business purchases, or one profile for each individual she buys gifts for. One approach enabling the definition of multiple personae would likely lead to better personalization because it could offer recommendations within the appropriate context.

VIII. Frameworks to specify privacy policies

The risk of opening a door to the private sphere is intrinsic to the personalization. As soon as the user is ready to provide private information and preferences to be better informed and take better purchase decisions, there is a certain chance this information to reach undesired destinations.

The World Wide Consortium launched an initiative to enable a secure exchanging of user profiles information: Open Profiling Standard (OPS) [10].

A big problem of the personalization is the data protection. The recommendation systems collect, store and evaluate important data about the users, but it is very rare that the recommendation system pages (and generally, all ecommerce applications) provide precise information about what they will do with the user information or explicitly make a statement about that in their general terms and conditions on the company site.

Ideally, the privacy safe recommender system would separate service and user profile. A user has a profile that can be used for all shops, recommenders and other systems. The problem is basically that the companies, that have put all efforts to collect the information about a particular user and therefore has a very complete user profile for their customer, are not interested in sharing this information with the competitors. Moreover, before the data gathering starts, it has to be proven that the customer agrees for a customization. Even if this doesn't take place, certain individual data can be collected about the customer.

Actually, the prerequisite for the user to provide information about her is that data protection and privacy expressed warranties are done. But it is a matter of fact, that the users are not very prone to provide private information, fearing misuse or simply because the information requestors don't provide sufficient context about what the user's data will be used for.

A. *Platform for Privacy Preferences*

Many efforts have been intended to introduce an electronic privacy policy to define an organization's information management practices. In particular, this information can be encoded in one of two XML-based policy definition languages: the Platform for Privacy Preferences (P3P), the W3C recommendation, or the Enterprise Privacy Authorization Languages (EPAL) [17], developed by IBM.

The policy is shredded, and the necessary information stored inside the relational database as tables called the "privacy meta-data." In practice, we found that it was possible to express privacy policies in the meta-data in a way that is largely language independent. Both P3P and EPAL encode rules for allowing or disallowing disclosure of data based on a combination of several factors, including some notion of purpose, data recipient, data category and condition. The primary type of condition is an *opt-in/opt-out* choice, specified by the user providing the private data, though EPAL also supports more complex conditions. The privacy meta-data stores a set of rules of the form <*purpose, recipient, data category, condition*>, indicating that the privacy policy allows for the disclosure of a particular category of data to a particular recipient for a particular purpose, provided that the indicated condition holds. For example, a rule might indicate that medical history is provided to external drug companies for research if an individual "opts in" to this choice. The meta-data also stores a mapping of data categories to relational attributes.

We will see in detail the basics of P3P in combination with APPEL language, which can be summarized in three bullet points as follows:

- P3P is Platform for Privacy Preferences
- P3P defines protocols and specifies languages
- P3P Schema for Websites, APPEL Schema for Clients. APPEL is the P3P Policies Exchange Language [19]

In the Figure 2 the P3P architecture is presented, which even still very coined with concepts like trust, reputation, etc, introduces encryption, certificates exchange, digital signature, etc.

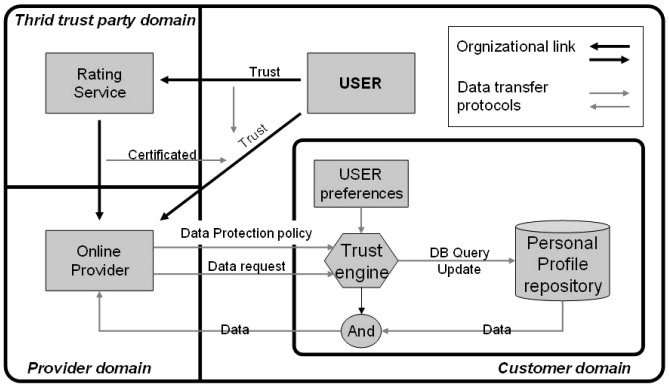

Figure 2

The Figure 3 shoes the roundtrips involved in the P3P

policy retrieval. A basic P3P interaction might proceed as follows:

1. The agent requests a Web page from a service.
2. The service responds by sending a reference to a P3P policy-reference in the header of its HTTP response. A policy-reference file lists parts of a Web site and the URIs of their corresponding privacy policies. A policy consists of one or more statements about a service's privacy practices.
3. The agent fetches the policy-reference file and determines the URI of the policy that applies to the requested page.
4. The agent fetches the policy, evaluates it according to the user's ruleset (which represents her preferences) and determines what action to take (e.g., simply informing the user about the privacy policy in place, or prompting her for a decision).

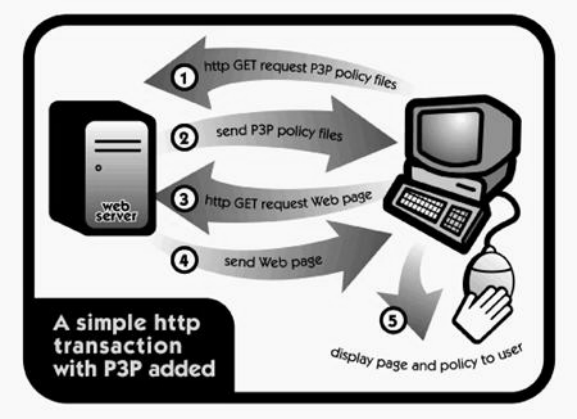

Figure 3

Following fragment shows how a P3P statement is written for "individual decision" purpose and "ours" as recipient, and how the negotiation language XPref would establish a rule to grant the access.

```
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1">
<POLICY discuri="http://p3pbook.com/privacy.html"
      name="policy">
 <ENTITY>
 <DATA-GROUP>
  <DATA
   ref="#business.contact-info.online.email">privacy@p3pbook.com
  </DATA>
  <DATA
   ref="#business.contact-info.online.uri">http://p3pbook.com/
  </DATA>
  <DATA ref="#business.name">Web Privacy With P3P</DATA>
 </DATA-GROUP>
 </ENTITY>
 <ACCESS><nonident/></ACCESS>
 <STATEMENT>
  <PURPOSE>< individual-decision /></PURPOSE>
  <RECIPIENT><ours/></RECIPIENT>
  <RETENTION><indefinitely/></RETENTION>
  <DATA-GROUP>
   <DATA ref="#dynamic.clickstream"/>
   <DATA ref="#dynamic.http"/>
  </DATA-GROUP>
 </STATEMENT>
</POLICY>
</POLICIES>

…..
<RULESET>
  <RULE behavior="request"
    condition="/POLICY[
        every $pname in STATEMENT/PURPOSE/*
        satisfies name($panme)="individual-decision"
              and
        every $rname in STATEMENT/RECIPIENT/*
        satisfies name($rname)= "ours"
      ]"/>
 <RULE behavior="block" condition="true"/>
</RULESET>
```

Pranam Kolari et al formulated three major drawbacks of P3P [20]:

1. P3P policies published by websites are not trusted by users
2. The languages available to describe user privacy preferences are not sufficiently expressive and
3. P3P framework does not provide a coherent view of available privacy protection mechanisms to the user.

Additionally, they introduced a model of user perspective of trust, as well as a populating ontology for the instance data (services for users to specify their preferences), what is known as web evaluation ontology (see Figure 4)

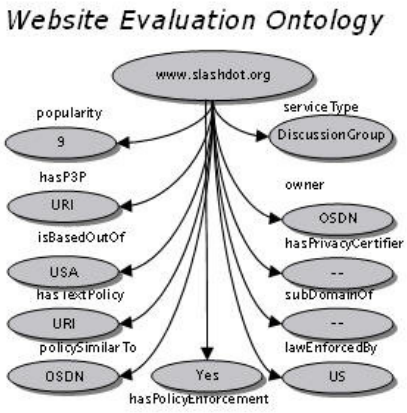

Figure 4

Bound to the previous idea, the Rei policy language is introduced [21], whose core features are given bellow:

- Encoded in (1) Prolog, (2) OWL
- Models deontic concepts of permissions, prohibitions, obligations and dispensations (actually more variety than P3P) (see Figure 6)
- Uses meta policies for conflict resolution
- Uses speech acts for dynamic policy modification
- We used it as a policy specification language
  - RDF specification capability (matches that of P3P)
  - Dynamic Policies as future extension to our work

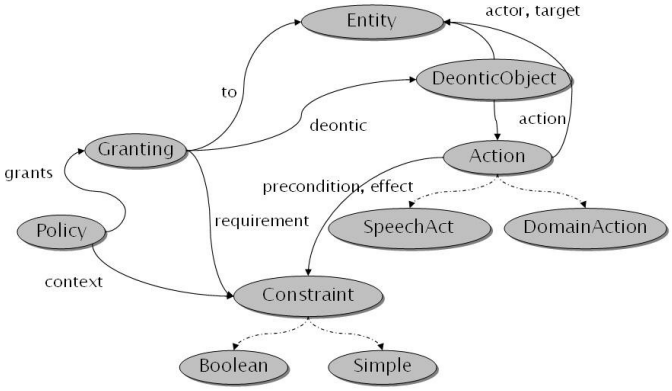

Figure 5

The possible usage of the web site evaluation ontology combined with the Rei is depicted in the Figure 6:

- Web Sites **optionally** publish P3P policies
- Clients specify privacy preferences using a policy language - **Rei**
- **Privacy Expert** is the privacy enhancement enabler by binding together entities of the system
- **Rei Engine** evaluates policies of users against website attributes
- **Website Recommender Network** propagates and builds a model of websites based on reputation
- **FOAF** – Enables the creation of the website recommender network

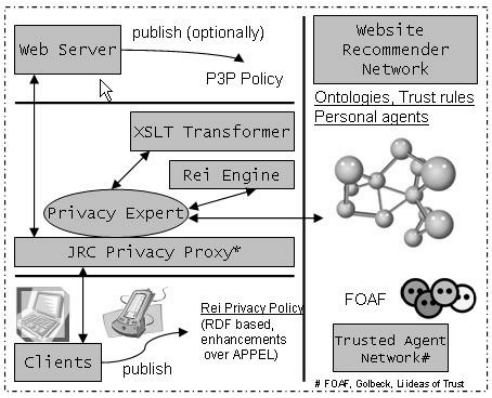

Figure 6

A comparison between P3P and Rei is out of the scope of this work.

IX. Conclusion and Outlook Statement

Recommendation systems usually relies on personal information given by their users or information derived from the interaction of the user with the system. Whenever personal information is required, certain privacy concerns are compromised.

A set of privacy preserving principle has been defined by international organizations to address the privacy risks. They can be summarized as follows: user should be aware of the fact that personal information is being collected, which information exactly and for which purposes, and be able of changing/updating the collected information that constitutes her user profile.

To achieve it, recommender systems should follow some design guidelines targeted to complying with the privacy preserving principles, or rather shift the control to the user, so that she takes decisions about her privacy information and its usage.

Frameworks are being introduced with the only purpose of enabling user to decide on the information to be collected and retained and on the purpose this information can be used. The main issue to adopt this approach is the usability penalty and the specification overhead to make it happen.

Intelligent information gathering frequently disregards the context of a transaction that the user performs on the system, attributing a user profile facts that are not necessary true in general, but only in a particular, controlled context.

The future will bring us easy-to-use privacy specification tools and systems able to discern which user profile of one user should be used to make a recommendation.

References

[1] M. D. Mulvenna, S. S. Anand, and A. G. Buchener. Personalization on the net using web mining. CACM, 43(8), 2000.

[2] Terveen, L. G. and Hill, W. C. 2001. Beyond recommender systems: Helping people help each other. In HCI In The New Millennium, J. Carroll, Eds. Addison-Wesley, Reading, Mass.

[3] L. Cranor, P. Guduru, and M. Arjula. User Interfaces for Privacy Agents. ACM Transactions on Computer-Human Interaction, June 2006, pp 135-178

[4] Zaslow, J. If TiVo Thinks You Are Gay, Here's How to Set It Straight: What You Buy Affects Recommendations On Amazon.com, Too; Why the Cartoons? The Wallstreet Journal, November 26, 2002. http://online.wsj.com/article_email/0,,SB1038261936872356908,00.html

[5] Adams, A. The Implications of Users' Multimedia Privacy Perceptions on Communication and Information Privacy Policies, in Proceedings of Telecommunications Policy Research Conference, (Washington DC, 1999). http://www.cs.mdx.ac.uk/RIDL/aadams/TPRC%20final.PDF

[6] Shyong K Lam and John Riedl, Privacy, Shilling, and The Value of Information in Recommender Systems. GroupLens Research Computer Science and Engineering, University of Minnesota, 2006

[7] Pennock, D.M., Horvitz, E., Lawrence, S., Giles, C.L.: Collaborative filtering by personality diagnosis: A hybrid memory and model-based approach. In: UAI '00: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, Stanford, CA, Morgan Kaufmann Publishers Inc. (2000) 473–480

[8] Shyong K. Lam and John Riedl. Shilling recommender systems for fun and profit. In Proceedings of the 13th international conference on World Wide Web, pages 393–402. ACM Press, 2004.

[9] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In Proceedings of CSCW 1994. ACM SIG Computer Supported Cooperative Work, 1994.

[10] Proposal for an Open Profiling Standard, available at http://www.w3.org/TR/NOTE-OPS-FrameWork

[11] Organization for Economic Co-operation and Development.Recommendation Of The Council Concerning Guide-Lines Governing The Protection Of Privacy And Transborder Flows Of Personal Data. Adopted by the Council 23 September 1980. http://www.datenschutz-berlin.de/gesetze/internat/ben.htm

[12] Arlein, R.M., Jai, B., Jakobsson, M., Monrose, F., and Reiter, M.K. Privacy-Preserving Global Customization. In Proceedings of EC'00, (Minneapolis, MN, October 17-20, 2000), ACM Press, 176-184. http://doi.acm.org/10.1145/352871.352891

[13] Canny, J. Collaborative filtering with privacy. In IEEE Symposium on Security and Privacy, (Oakland, CA, May 2002)

[14] Lau, T., Etzioni, O., and Weld, D. S. Privacy Interfaces for Informationanagement. Commun. ACM 42, 10 (October 1999), 89-94.

[15] Manber, U., Patel, A., and Robison, J. Experience with Personalization on Yahoo! Communi. ACM 43, 8 (August 2000), 35-39.
[16] S. K. Lam and J. Riedl (2005): Privacy, Shilling, and The Value of Information in Recommender Systems. Proc. of User Modeling Workshop on Privacy-Enhanced Personalization, 85-92
[17] Enterprise Privacy Authorization Language, available at http://www.zurich.ibm.com/security/enterprise-privacy/epal/Specification/index.html
[18] Platform for Privacy Preferences, available at http://www.w3.org/P3P/
[19] APPEL, available at http://www.w3.org/TR/P3P-preferences/
[20] Pranam Konami, Lalana Kagal, Anupam Joshi, and Tim Finin, "Enhacing P3P Framework with Policies and Trust", UMBC Technical Report and under review, 2004.
[21] Lalana Kagal, "Rei: A Policy Language for the Me-Centric Project", HP Labs Technical Report, 2002