Sex, Privacy and Ontologies

Adriel Dean-Hall
DRC School of Computer Science
University of Waterloo
Canada
adeanhal@uwaterloo.ca

Robert Warren
Math and Statistics
Carleton University
Ottawa, Canada
rwarren@math.carleton.ca

ABSTRACT

Personal profiling has long had negative connotations because of its historical association with societal discrimination. Here we re-visit the topic with an ontology driven approach to personal profiling that explicitly describes preferences and appearances. We argue that explicit methods are superior to vendor-side inferences and suggest that privacy can be maintained by both exchanging preferences independently from identity and only sharing preferences relevant to the transaction. Furthermore this method is an opportunity for additional sales through the support of anonymous 'drive by' shopping that preserve privacy. We close by reviewing the computational advantages of accurate profiling and how the ontology can be applied to complex real world situations.

1. INTRODUCTION

In this paper, we consider which additions to user profiles are helpful to support suggestion, matching and classical information retrieval needs in shopping, dating, and erotica contexts. The novel approach of communicating preferences without identity is explored, as well as the relationships between the notions of sex and gender within recommendation systems. This research is a continuation of previous research efforts on data extraction [15] and recommendation systems [4], using user gender and personal appearance preferences.

Keeping track of gender has historically only been important for a limited number of purposes: a) it is a high selectivity identifier for identification purposes, b) it permitted the automation of expected social conventions and salutations and c) allowed persons to be pre-qualified according to a marketeer's sales plan.

A strong desire for personal privacy is now preventing this information from being widely shared because preferences and identity have historically been unified. Concurrently, social mores have changed in that alternative interpersonal relationships are being recognized as mainstream and the classical definition of gender as male or female is being questioned. It follows that these changes will drive the creation of new tools which are used to seek entertainment, relationships and goods.

Far from classifying individuals in a box, we seek only to provide models accurate enough for a software agent to adjust its retrieval algorithms. We note that human preferences are notoriously fickle as well as the difficulty in getting a person to express their wants and desires when they themselves are unsure. The intent is not to turn the individual into a product, but to facilitate communications with the information retrieval agent and permit proper content negotiation with information providers. A number of research problems remain in the privacy-preserving processing of the information, but this research is a first step in documenting them.

This paper is organized as follows: we first review the previous work done in customizing recommender and IR systems using personal preferences and then motivate our research based on a number of use cases that occur in the IR field. An ontology of personal characteristics, gender and ethnicity is also presented as an experimental reference for assisting IR for positive and negative feedback. Finally, we report on some experimental results in the retrieval of erotica materials and how the use of a separate preference data structure improves retrieval performance. We conclude on the extension of the ontology to other IR and matching problems.

2. PREVIOUS WORK

The use of recommender systems and profiled IR systems is not new and several other research approaches have been attempted in the past. Approaches using mobile software agents were proposed early on to perform distributed information retrieval by Brewington et al. [2] while other researchers, such as Pipanmaekaporn[12] focused on learning a user's interest based on a single relevancy class using research paper collections.

Daoud [3] and Middleton [11] furthered this approach by utilizing web server logs to determine user interest which was then classified within topic ontologies based on the Open Mozilla Directory.

Schiaffino and Amandi [13] also supported the creation of user profiles through the use of demographic databases and

SEXI 2013 Rome, Italy

user questionnaires. Gasparini [6] used an ontology to report the needs, demographics and languages spoken by a user in order to customize the material for presentation. Sutterer et al. [14] made use of OWL [1] ontologies in an attempt to provide better context to retrieval situations.

Katifori [9] and Ghosh [7] also made attempts at building integratable user profiles using semantic web technologies. Ghosh [7] makes use of a simple "Shopping List" property and Sutterer [14] attempted to create context for preferences based on locations. One of the reoccurring issues is the lack of support to record preferences within personal profiles.

The relevance track that was part of the TREC 2012 conference explored the performance of recommender systems. For these systems, preference data was provided as a list of 50 venues, along with positive and negative feedback data supplied by the users. Out of 23 recommender systems submitted to TREC 2012, 14 systems performed better than simple baseline systems that did not incorporate user profile data and simply made general, non-user specific, suggestions [4].

In the following research, we present a modified approach to information retrieval using a novel ontology titled Appearances [2] that can support both preferences and generic identity information. Terms are provided for physical appearance traits, body measurements as well as sexual, romantic and entertainment preferences and aversions. This is also done in a manner that allows for the independent use of different aspects of the ontology without having to produce an actual identity, which is similar to the approach used by Gulyás and Imrel [8] for anonymizing social network applications.

3. SPECIFIC USES CASES

Figure 1 is a high-level representation of what we believe the appropriate use of the ontology should be. Here the profile information is completely stored on user's computer. Previously, most if not all user preferences would have been stored as part of a user account on the server side, which would require the user to register with the web-site before usage. Previous research [5] has shown that the effort required for registration is an effective deterrent to the buying behaviour.

Instead we propose for most of the profiling information to be stored on the user's client. The profile agent will communicate which aspects of the profile it could make use of to improve retrieval, these portions on the profile can then be dispensed to the server after the user has authorized it. The user can decide to withhold certain parts from the server or never include them in the profile. The profile can be sent to the server irrespective of whether the user has registered with the merchant. For privacy reasons, the profiling information can also be dispensed anonymously.

This is not unlike the mobile agent information retrieval paradigm of Brewington et al. [2], with the caveat that ontological structures are exchanged instead of computer code. To the best of our knowledge, providing partial elements of a personal profile to the seller without identity information in a novel contribution that has not been studied before. We now review several use cases where we think that this approach may be useful and a solution to specific corner cases within each one.

[1] www.w3.org/TR/owl-ref/

[2] http://rdf.muninn-project.org/ontologies/appearances.html

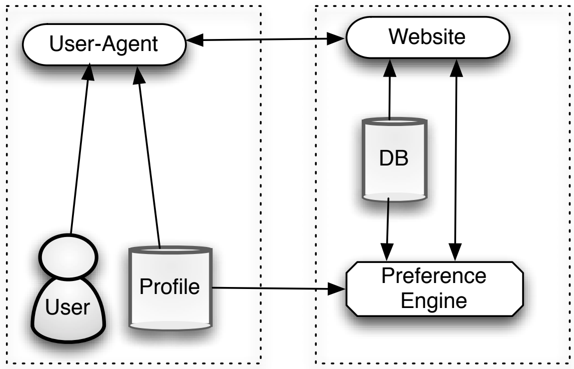

Figure 1: A user's profile will be stored locally and portions of the profile are used by the provided to improve information retrieval.

3.1 Shopping

Online shopping is an application that has traditionally made use of 1) short-term, cookie based preference modelling based on advertising click through, 2) the relationship between products "people who bought A also bought B", and 3) proprietary long-term preferences modelling for registered users. Method 1 has been reported as having conversion rates as high as 10% in trade magazines for large retailers, while method 2 is known to increase sales by providing the users with a set of products likely to be of interest. In both cases, these systems have been known to make recommendations that were not only ineffective, but actively detrimental to the relationship between vendor and customer, for example by recommending a book on anal sex as a Father's Day reading list and as complementary to an evangelist book[10, 16].

The third option of requiring customer registration is known to be a hindrance since it requires effort from the user to enter his personal information, additionally some users will not be comfortable sharing all the requested information. Vendors have a vested interest in getting the user to register since demographic and marketing information can be derived from logistic information such as billing address. This type of profiling or recommendation system works best for existing customers who have already expended effort on registering.

Lastly, an existing problem with recommendation and profiling systems is their inability to handle contexts, such as a difference between the person buying and the person receiving the gift. While florists and some online bookstores have had to tackle this problem, they do so primarily by separating shipping and billing addresses and focusing on calendar events, such as Mothers' day, instead of the customer's profile.

A re-occurring use case for third party purchases has to do with a husband shopping for lingerie for his wife. What makes this particular case so interesting is that the product is completely removed from the shoppers' experience and any profiling information likely to be available: the sex, gender, body measurement (which vary across country and sex) preferences are unusable and in some cases the husband is ignorant of their spouse's size. This particular use case has spawned several web-sites dedicated only to this problem which is dealt with as a presentation problem instead of a recommendation problem.

Besides inducing large errors in 3rd-party recommendation systems that monitor browsing behaviour (purchased goods drive the objective function), these specific cases represent lost sale opportunities in that the user is fighting the recommender system while attempting to search for a relevant gift.[3] The next-generation use case is one person shopping for another person whose lifestyle or culture are completely orthogonal to their own, and who are looking for gifts that are appropriate without having a complete understanding of the world of the gift receiver.

3.2 Dating

As one of the primary human drives, dating is an application where personalization and profiling are key. Furthermore, it is a matching problem in that the preferences of both potential matches must be taken into account concurrently. The removal of geographic restrictions and limited online anonymity have enabled the creation of new alternative communities for dating, such as academic or military singles dating. Currently the popular online classified ad web-site Craigslist has no less than 21 different types of relationships listed, ranging from traditional marriages to polyamourous relationships.

Interestingly, the mass customization of dating communities seems to rely primarily on the re-branding of the same back-end systems and / or the restriction of the sex field on the site registration form. We performed a short survey of alexa.com's directory of gay, lesbian and alternative dating sites, classifying them by their treatment of gender. We also performed the same classification for the dating websites found in the first 20 search results for the Google queries "gay dating" and "lesbian dating".

The results in Table 1 tabulate the number of dating sites according to their treatment of a person's gender. The first class, "Generic", identifies dating sites that are simply targeted advertisements for larger, brand name dating sites. The "Hard-coded" class contains the dating sites that use commercial off the shell dating website software, with only one gender as a choice. Lastly, the "Choice" class lists websites where any kind of gender differentiation opportunity is provided to the user.

[3]The authors note that the systems that generate and display online ads went to great lengths to try and find a relevant advertisement after several days of online searches on gender, appearances and interpersonal relationships. Results ranged from comical to insightful, but it is obvious that further work on profiling (including a "please-ignore-this-search" button) are needed.

Source	Generic	Hard-coded	Choice
Alexa	2	4	4
Google 'Gay Dating'	5	6	3
Google 'Lesbian Dating'	3	7	2

Table 1: Gender customization of different dating websites

This survey is by no means comprehensive, but is valuable in identifying the lack of support for the self identification of gender. In all three surveys, the majority of dating websites treat gender as a binary choice with no attempt at differentiation, even through this additional profiling would make matching easier.

What is interesting is that the marketing documentation of the websites makes it clear that the website operators are aware of the target community and its terminology. This understanding has not been ported to the search and matching functions of the website since its profiling system is incapable of recording the data.

In the last category of dating websites, some support was provided for the limited self identification of gender, primarily through the use of the 'Male Trans Female' and 'Female Trans Male' terms. The only other case was the crude use of sexual positions as a proxy for gender and this approach may not be appropriate for all demographics.

Personal preferences in romantic and / or sexual relationships are among the most complex, and can be at times contradictory. This limited support for gender terms within dating sites does not provide adequate support for the profiles specific to the target community. The ability to process a complex personal profile would allow matching engines to locate not only members of the users' preferred community but improve its matching algorithms based on its understanding of that community.

Furthermore, some of these preferences can be awkward to enumerate in public and can incur a certain public stigma. Examples can include a preference for persons with a specific hair colour, or an aversion to persons from certain cultural backgrounds. In these cases, profile preferences can be used as a social lubricant by avoiding unnecessary rejections and ensuring that incompatible matches are never introduced to one another.

Lastly, we note that terms, labels and nomenclature represent a tremendous opportunity for additional profiling by extrapolating additional information from the specific class of terms used within a personal profile. As an example, the terms "man, gentleman, dude, bro, boy, lad", all have been used by men of any age to describe themselves where each implies a set of demographic, social and temporal properties that can populate a personal profile. The key problem in their use, which we will not tackle here, lies in documenting those properties both in the user's semantics and its relationship to the merchant's semantics.

3.3 Erotica

Beitzel et al. [1] estimate that as much as 7% of queries in the 2006 AOL query log were pornography related. Sex being one of the primary human drives, it is no surprise that erotica searches are an important class of search problem.

Creating preferences sets for erotica is a complex endeavour in that there is a strong element of fantasy to the preferences and aversions. There does exist a probabilistic relationship between romantic, sexual and entertainment preferences, but its complexity is too high to easily infer one from the other.

We can, however, let the user specify their entertainment preferences using the ontological constructs for gender, ethnicity and physical appearances. Anecdotal review of erotica web-site folksonomies in Section 6 has identified Appearance, Gender and Race/Nationality as the most selective categories within collections, which makes vendor selection possible for the user and query pre-processing possible for the vendor.

Currently, erotic material is an adult-oriented product that requires special considerations (as with alcohol or medication) and which is a social taboo. These factors therefore dictate that these customers will likely be concerned about privacy and desire a discrete shopping experience. An overly aggressive attempt at getting the customer to register with a vendor will likely fail; the ontological preference terms therefore permits the customer to locate the entertainment that he desires without being driven away too early.

Without doubt, the use cases listed above and the suggested use of ontological terms for describing preferences or aversions will likely engender a new generation of spammers attempting to disguise their intentions. Vendor reputation and statistical normalization methods will therefore be required to perform expected quality control on vendor results and the suggested products.

4. APPEARANCES ONTOLOGY DESCRIPTION

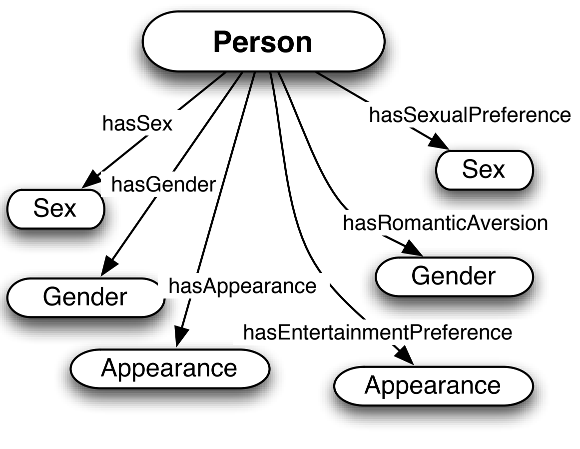

Figure 2: The profile contains attributes, preferences, and aversions for different situations.

User profiling and preferences remain an ongoing problem within e-commerce systems as there is a lack of standardization. What systems do exist are black box systems dedicated to specific tasks whose models are not portable to related modelling questions. Furthermore, what standards do exist vary according to the institution that published them and their intended audience, making data transformation problematic. Figure 2 shows an overview of the profile information which contains information such as the person's gender and entertainment preference.

We make use of OWL based ontological data structures because they have built in equivalence and inheritance for both properties and classes. This ensures that whenever preferences are communicated, a super-class of the desired property is available as a fall-back if the specific semantics desired is unavailable. Furthermore, the data structures are independent of any software package preventing lock-in and the use of terms and properties can describe profile data independently of identity which permits both anonymous "window shopping" and shopping with a third party's preferences.

In the following subsection, we review the aspects of the Appearances ontologies and how it represents a persons' attributes, preferences and aversions.

4.1 Gender, Sex and Orientation

The Appearances ontology was originally meant to deal with ambiguous gender references in text and was expanded to deal with soldiers' personal description. In the coming paragraphs we describe the working of the ontology and its application to information retrieval.

An ongoing problem for ontology design is that a number of social and linguistic conventions are in everyday use while being logically wrong or ambiguous. Examples include the use of sex and gender interchangeably, the use of contradictory genders ("Sarah was a airman") and the ambiguity of perception.

The ontology provides two sexes, which within the ontology is grounded to XX and XY phenotypes and three genders Male, Female, Transsexual[4]. The terms have no restriction on any combination of Gender, Sex or Relationship which gives the end user full descriptive power. Specifically, sexual, romantic and legal relationships can be separated and impose no combinatorial restrictions.

As the ontology has its roots in the processing of war records that straddle the Edwardian and Victorian Eras, a second set of terms are provided which are suffixed Simple. These terms are a replica of the generic terms previously enumerated, but include cardinality and disjoint restrictions that enforce gender assumptions held within official historical records. Hence, instance SimpleGenderM is the same as SinpleSexM and SexISO5218-1 while being disjoint with SimpleGenderF. This permits assertions that the wife of soldier on an 1915 form must be a woman and the mother of their child. Not all cases fit within this model, a number of women do

[4]We note that there exists a great deal more, but we provide only the most obvious ones here.

serve as soldiers as both men or women, but it provides a model that accounts for the majority of cases and that can locate exceptions worthy of study within a database.

4.2 Observed versus self-reported profile properties

One of the more useful aspects of an ontology as opposed to other schema based solutions is its ability to make use of sub-properties and sub-classes. In cases where a perfect match can not be obtained between the merchant and users' systems, it is always possible to obtain partial information. A typical example could be the `hasAppearance` property that can be branched into `hasAppearanceObserved` or `hasAppearanceSelfReported`: different vendors may choose the property as observed by an authority versus a self-reported property, but have a documented alternative to the `hasAppearance` property if the specific property they want is unavailable.

Similarly, an interesting element of the linked open data model is that the parts of ontologies can be separated and used independently without loss of semantic meaning. The utility of this is evident for "window shopping" e-commerce applications where anonymous preferences can be enumerated by the user agent without necessarily providing identity information.

4.3 Eyes, Skin and Hair

The ontology provides several different reference standards for identifying the colour of hair, eyes and skin. Because not all standards have the same degree of specificity or precision and equivalences are not always available from one standard's term to another. In some cases we are able to provide properties that indicate the inclusivity of a term within another through the use of skos:broader and skos:narrower properties. An example of this is the use of the Martin-Schultz eye colour scale which provides several different terms for shades of blue eyes that all map to the single BLUE eye colour for motor vehicle license terms.

Other standards such as the Von Luschan skin colour terms are known to be ambiguous, while some of the hair colour references in the US Federal Bureau of Investigation such as PINK or BALD have no equivalent within the Fischer-Saller hair colour scales.

There does exist a series of statistical relationships between macroetchnicity and skin, hair and eyes: it is not unreasonable to expect a person from Japan to have black hair. However because of the difficulty in reconciling statistical relationships into logical relationships, these are not currently recorded in the ontology.

Several standards are referenced within the ontology to improve interoperability with the caveat that not all known standards are directly interchangeable. Appearances makes use of three levels of properties to record the relationships between the standards: owl:sameAs for terms that are completely similar, skos:related for terms that can easily be confused for one another (eg: grey eyes versus blue eyes) and the skos:broader and skos:narrower properties for terms that can encompass a series of other terms (eg: light blue, dark blue versus blue). Whenever possible, common language labels have been provided and related to these terms.

4.4 Modelling preferences and aversions

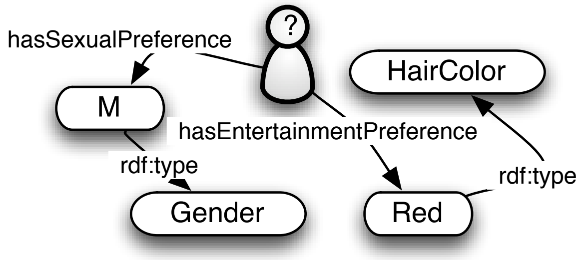

Figure 3: Enumeration of anonymous preferences

Figure 3 is a representation of the anonymous profile information for a person's sexual and entertainment preferences in the context of the information retrieval of erotica, in this case a person interested in males with red hair. An ongoing discussion is whether merchants will accept to process anonymous preferences, as some merchants wish to establish user accounts immediately in order to capture the customer.

Research suggests that online users shun complexity of user account registration and favour web-sites that provide effective search functions [5]. Notwithstanding, this is likely to lead to a second generation of spam and spam sites, with the caveat that the encoding will allow complex normalization.

We choose the term aversion to represent the antonym of a preference. As with preference, these properties do not represent values or moral judgments, but affinities towards certain concepts which recommender systems should use with a certain amount of flexibility. They are not meant to represent absolute requirements: as an example of food preferences one can record a preference for oranges and an aversion to mushrooms but not a deadly allergy to nuts.

Also note that between the set of preferences (P_s) and the set of aversions (A_s), there may exist an unknown region (U_s) within the universe S that is $U_s = S - P_s - A_s$. This has to do with both preferences and aversions not being a complete element set of possibilities.

Figure 4 is a limited representation of the taxonomy that links the different sub-classes of appearances, different measuring standards, and actual measurements.

Appearances makes use of a top level property that records all other aspects of the person appearances, including measurements, hair, eye and skin colour. All of these are accessible according to multiple formats. There are relationships that are encoded within the ontology that allow for the translation of equivalent or similar terms across different standards. This allows an agent and server to negotiate transactions even through both are not using standards that are not completely compatible. One item that created some difficulty is that information that is expected to be true statistically, such as age causing grey and white hair is not

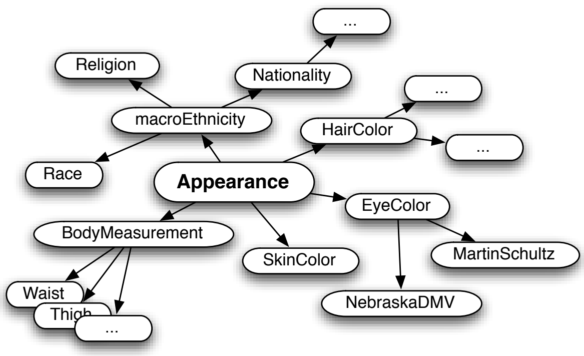

Figure 4: The ontology of appearances has multiple classes with sub-classes and classification standards.

easily encoded into an ontological framework.

4.5 Body Measurements

Body measurements are provided as part of the appearance class for both men and women. Their use-case for online shopping is clear and we use generic clothes fitting measurements, which are historically linked to sex. For this reason, all measurement properties are related to the sex ontological terms for men or/and women instead of their gender. As measurement units are currently problematic in ontologies, separate properties for both decimal inches and meters are provided.

Manufacturer's clothing and shoe sizes have been omitted as a number of different conflicting standards exists across the world for men and women and equivalences are not always available or consistent over time. Thus, we only provide a basic set of body measurement properties as a starting point. In future work, it would be interesting to add the ability to translate clothing size across standards. This would be useful in the use case of Section 3.1 where a person is purchasing a gift for a spouse from one vendor based on the label size of a current garment from a second vendor.

5. IR OF EROTICA MATERIAL

In order to estimate how including attributes, such as gender, into an IR system would improve results we calculated which subset of documents would have to be searched if the system knew a user's attribute preference and the expected speedup a system could have given this information.

Site	Num Documents
Site 1	472,283
Site 2	4,785,909
Site 3	76,531
Site 4	637,650
Site 5	5,441,078
Site 6	2,762

Table 2: Number of documents for each site.

We looked at the categorization of documents (videos) on six pornography sites. These sites were chosen from a list of the top sites in the "Adult" category of Alexa.com. They were all general interest sites, rather than sites that focused on only a certain category of documents, e.g., a site that only contains Japanese oriented documents. It is, however, interesting to see that genre specific categories exist, these sites are, effectively, limiting their results to a specific attribute that could have been communicated in a profile. All of the sites chosen also exposed how many documents were in each of the categories on the site. The number of documents in each site can be seen in table 2.

For each of the top 50 categories on each site, we labelled whether the category was relevant to each of four attributes: gender, ethnicity, age, and hair colour. For example, the category "college girl" is relevant to the age group and the female gender group. Most categories are only relevant to one of the four attributes. These four attributes were chosen as the attributes that had many relevant corresponding categories in our data-set. Each of the four attributes can have certain values. The possible values chosen are values associated with the top 50 categories rather than every possible value.

We then looked at how search performance would improve with knowledge about user preferences from profiles. For example, given the query "swimsuit brunette", a search system that does not consider profile data would have to search all documents for this query. However, with the user preferences from the profile, we could know that, for example, the user likes documents where the hair colour attribute is "brunette". With this knowledge we can not only find more relevant documents but also improve the time it takes to perform the query. This query would only have to be run on the subset of documents in the "brunette" category. The documents are categorized so we know, in advance, all possible values for all four attributes so we would be able to precompute which documents belong to which attribute values. This would enable us to speedup the search for documents where an attribute is specified. Note that not all categories have a value for each of the four attributes. For example, some categories do not specify anything about hair colour.

In order to estimate the speedup improvement we use calculations for the fraction of documents that fall within each attribute. The speedup is the inverse of the fraction of documents that would have to be searched. Each of the four attributes can have one of several values. In order to calculate the advantage of having preference information about an attribute we take the mean of the fraction of documents across all of the attribute values. This is done for each of the four attributes for each of the six sites. So, for hair colour, this would be the mean of the fraction of documents that would have to be searched over the blond, brunette, and redhead preferences.

6. EXPERIMENTAL RESULTS

Figure 5 shows the observed speed up of the system when certain parts of the profile are exposed to us. Again, these collections are general, rather than category specific, we expect that similar results will be found for other general collections. We calculated the average speed up given each of the four factors for each site. For example, if we knew the desired ethnicity from the profile, on site 1 there is a

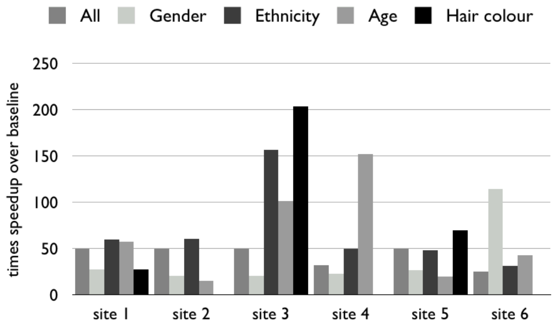

Figure 5: Observed speedup, over baseline of searching the entire collection, across the six sites.

60 times speed up to search time. This increase in speed comes from the fact that we only need to look at the subset of documents that are relevant to a given ethnicity and we have pre-computed which documents are related to which ethnicity.

As a comparison to the expected speed up for each factor we look at the expected speed up if the profile contained which specific category the user preferred. The "All" expected speed ups are the mean expected speed ups over all categories (the categories given on the site rather than the attributes). Note however that for the four attributes more than one category might be included in a particular attributes. For example, if there is a preference for the "male" gender there may be multiple categories that satisfy this preference, which will generally lead to less of an expected speed up than if the preference specific a single category. The "All" category is provided as a means to compare random category selection to given the selection of attributes. These speedups could be realized by picking categories at random and recording whether a user has a preference for that category in the profile.

Factor	Speed up
All	43
Gender	37
Ethnicity	68
Age	65
Hair Colour	50

Table 3: Observed speed up for attributes.

The differences in the expected speed ups between sites is seen because each site has a different number of categories relevant to each factor and a different number of documents within each category. The average expected speed ups for each factor can be seen in Table 3.

7. CONCLUSION

In this paper, we reported on current issues in the use of user preferences for e-commerce application including information retrieval engines. A novel ontology describing a person's preferences and appearances is described and its applications to multiple use cases presented. Finally its application to the problem of the retrieval of erotic material was reported with a speed improvement that can be achieved in a manner that permits customers to preserve privacy. Using an ontology, like the one describes, to record profile data gives the flexibility needed to describe a variety of a user preferences. At the same time it allows for anonymity and privacy to be preserved.

8. REFERENCES

[1] S. M. Beitzel, E. C. Jensen, O. Frieder, D. D. Lewis, A. Chowdhury, and A. Kolcz. Improving automatic query classification via semi-supervised learning. In *ICDM*, pages 42–49. IEEE Computer Society, 2005.

[2] B. Brewington, R. Gray, K. Moizumi, D. Kotz, G. Cybenko, and D. Rus. Mobile agents for distributed information retrieval. In *Intelligent Information Agents, chapter 15*, pages 355–395. Springer-Verlag, 1999.

[3] M. Daoud, L. Tamine-Lechani, M. Boughanem, and B. Chebaro. A session based personalized search using an ontological user profile. In S. Y. Shin and S. Ossowski, editors, *SAC*, pages 1732–1736. ACM, 2009.

[4] A. Dean-Hall, C. L. Clarke, J. Kamps, P. Thomas, and E. Voorhees. Overview of the trec 2012 contextual suggestion track. In *Online Proceedings of TREC 2012*, Gaithersburg, MD, USA, 2012. National Institute of Standards and Technology. To appear.

[5] N. Farag, M. Smith, and M. Krishnan. The consumer online purchase decision: A model of consideration set formation and buyer conversion rate across market leaders and market followers. In *AMCIS*, page 37. Association for Information Systems, 2003.

[6] I. Gasparini, M. S. Pimenta, J. P. M. de Oliveira, and A. Bouzeghoub. Combining ontologies and scenarios for context-aware e-learning environments. In J. C. Anacleto, R. P. de Mattos Fortes, and C. J. Costa, editors, *SIGDOC*, pages 229–236. ACM, 2010.

[7] R. Ghosh and M. Dekhil. Mashups for semantic user profiles. In J. Huai, R. Chen, H.-W. Hon, Y. Liu, W.-Y. Ma, A. Tomkins, and X. Zhang, editors, *WWW*, pages 1229–1230. ACM, 2008.

[8] G. G. Gulyás and S. Imre1. ï£ijanalysis of identity separation against a passive clique-based de-anonymization attack. *INFOCOMMUNICATIONS JOURNAL*, 3(4):11–20, December 2011.

[9] A. Katifori, M. Golemati, C. Vassilakis, G. Lepouras, and C. Halatsis. Creating an ontology for the user profile: Method and applications. In C. Rolland, O. Pastor, and J.-L. Cavarero, editors, *RCIS*, pages 407–412, 2007.

[10] S. K. Lam, D. Frankowski, and J. Riedl. Do you trust your recommendations? an exploration of security and privacy issues in recommender systems. In *Proceedings of the 2006 international conference on Emerging Trends in Information and Communication Security*, ETRICS'06, pages 14–29, Berlin, Heidelberg, 2006. Springer-Verlag.

[11] S. E. Middleton, D. D. Roure, and N. Shadbolt. Capturing knowledge of user preferences: ontologies in recommender systems. In *K-CAP*, pages 100–107. ACM, 2001.

[12] L. Pipanmaekaporn and Y. Li. Mining a data reasoning model for personalized text classification. *IEEE Intelligent Informatics Bulletin*, 12(1):17–24, 2011.

[13] S. N. Schiaffino and A. Amandi. Intelligent user profiling. In M. Bramer, editor, *Artificial Intelligence: An International Perspective*, volume 5640 of *Lecture Notes in Computer Science*, pages 193–216. Springer, 2009.

[14] M. Sutterer, O. Drögehorn, and K. David. User profile selection by means of ontology reasoning. In D. Collange, T. Atmaca, M. D. Logothetis, H. Mannaert, J. T. Yu, and C. Dini, editors, *AICT*, pages 299–304. IEEE Computer Society, 2008.

[15] R. H. Warren. Creating specialized ontologies using wikipedia: The muninn experience. In *Proceedings of Wikipedia Academy: Research and Free Knowledge. (WPAC2012)*, Berlin, Germany, June 2012. Wikimedia Deutschland.

[16] A. S. Weigend. Analyzing customer behavior at amazon.com. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 5–5, New York, NY, USA, 2003. ACM.