

Recommending Access Policies in Cross-domain Internet

Nuno Bettencourt¹, Nuno Silva¹ and João Barroso²

¹*GECAD and Instituto Superior de Engenharia, Instituto Politécnico do Porto, Porto, Portugal*

²*INESC TEC and Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal*

Keywords: Recommendation, Access Policy, Unknown-unknown.

Abstract: As the amount of content and the number of users in social relationships is continually growing in the Internet, resource sharing and access policy management is difficult, time-consuming and error-prone. In order to aid users in the resource-sharing process, the adoption of an entity that recommends users with access policies for their resources is proposed, by the analysis of (i) resource content, (ii) user preferences, (iii) users' social networks, (iv) semantic information, (v) user feedback about recommendation actions and (vi) provenance/traceability information gathered from action sensors. A hybrid recommendation engine capable of performing collaborative-filtering was adopted and enhanced to use semantic information. Such recommendation engine translates user and resources' semantic information and aggregates those with other content, using a collaborative filtering technique. Recommendation of access policies over resources promotes the discovery of known-unknown and unknown-unknown resources to other users that could not even know about the existence of such resources. Evaluation to such recommender system is performed.

1 INTRODUCTION

The Internet has recently grown to over three billion users. On certain social networks, more than two hundred thousand photographs are uploaded every minute. Such rate of content generation and social network building make the task of sharing resources more difficult for users.

Standard resource sharing in the Internet is achieved by granting users with access to resources, but they are commonly restricted to resources hosted on a single domain. Access policies are consequently issued to users registered on the same domain. Sharing resources with users that are not registered on the same domain has proven insecure or difficult to achieve. Referencing and accessing resources protected by access policies in other web domains (apart from where they are hosted) is practically unsupported by existing web applications.

In cross-domain sharing, such difficulties encourage:

- the cloning of the resource to different domains;
- the multiplication of users' internal and social identity.

The goal of this work is to provide a seamless cross-web-domain infrastructure that provides secure,

rich and supportive resource managing and sharing processes. It proposes a distributed and decentralised architectural model by fostering cross-web-domain resource sharing, resource dereferencing and access policy management. It adopts the principles of the Web and of World Wide Web Consortium (W3C) standards or recommendations.

In order to support user management of access policies, a recommendation provider capable of recommending access policies to users is included in the architecture (section 2). The proposed recommendation engine features a hybrid engine consisting on the combination of different filtering techniques that exploit user profiles, their social networks, resources content, (distributed) provenance and traceability information (section 3).

A prototype to demonstrate the infrastructure's feasibility was designed and implemented to prove that the architecture model can be deployed in a real world scenario. The hybrid recommendation process was tested using an available data set where information was interpreted to simulate human behaviour in the system (section 4 and 5).

Finally, the last section gives an overview of the proposed solution and suggests further research.

2 BACKGROUND KNOWLEDGE

This work allows the recommendation of access policies to resource authors. This section provides an insight about recommendation processes, resources that are currently not easily shared because of access policy restrictions and Authentication, Authorisation and Accountability (AAA) architectures.

2.1 Recommendation

Recommendation is something that has become part of everyone's daily lives. To reduce uncertainty and help coping with information overload when trying to choose among various alternatives, people usually rely on suggestions given by others, which can be given directly by recommendation texts, opinions of reviewers, books, newspapers, etc. (Shardanand and Maes, 1995).

Users are willing to follow others' recommendations and to give back recommendations to the community.

When deciding between which product to buy, users want to be able to read opinions from other buyers (MacKinnon, 2012) and tend to follow them as they are considered experienced users (Wasserman, 2012).

Currently, recommendation is widely used in electronic commerce (Adomavicius and Alexander, 2011; Linden et al., 2003; Schafer et al., 2001). In e-commerce web applications, trust is based on the feedback of previous online interactions between members as shown by the authors in (Resnick et al., 2000; Ruohomaa et al., 2007).

In the Internet perspective, there are other areas in which recommendation is also relevant, such as resource recommendations on websites (*e.g.* Pinterest), documents (*e.g.* Slideshare, Pocket) and users (*e.g.* LinkedIn, Facebook, Google+).

With the Internet's continual evolution, recommender systems have also evolved. While initially recommendation was only used in e-commerce websites for recommending similar or most bought items to users, nowadays the process of recommendation has improved such that the recommendation of friendship and/or relationship between users of a social network has become a quite common task on typical social web applications.

Every recommender system is typically based on two elements:

User/Item Actions. Represents user actions upon items and may include a possible rating.

Item Similarities. Represents the associations between users or between items. Some recom-

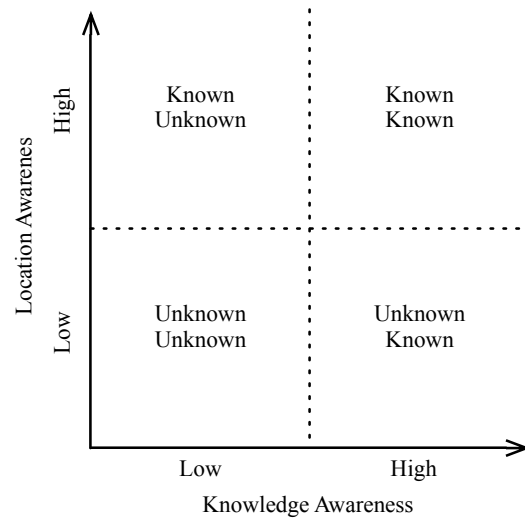


Figure 1: Location Awareness vs. Knowledge Awareness.

mender systems provide algorithms to calculate item similarity during the recommendation process, while others even allow the usage of external pre-computed item similarities during the process.

The output of a recommender system is a scored list of recommended items that are recommended to a list of users. The maximum number of retrieved recommendations is specified by the value of AT.

A systematisation of the user's consciousness about resources is presented next, which will be helpful to perceive the importance of the recommendation process in the scope of this work.

2.2 Known-known, Unknown-unknown

A user's consciousness about something can be characterised according to two dimensions: perception of reality and reality of perception.

Applying such rationale to resources' location and users' knowledge awareness of those, a particular resource can be classified as (*cf.* Figure 1):

Known-knowns. These are resources whose existence and location are known by the user *e.g.* a photograph is taken of a person, and the person knows about its existence and its location.

Known-unknowns. These are resources a user recognises he/she knows nothing about until he/she finds them *e.g.* a person finds a photo by chance on which he/she appears, knows its location but was not aware of its existence.

Unknown-knowns. These are resources that the user does not know how to find, but knows about their existence, *e.g.* a photo is taken of a person, the person knows about its existence but does not

know about its location. With time and searching investment the person might get to its location.

Unknown-unknowns. These are resources whose existence the user is not even aware of *e.g.* a photo is taken of a person but the person does not know its existence or where to locate it. These type of resources would only come up on searches related to the user if contextual information is used.

This classification emphasises the fact that the same existing information is perceived differently by users. There are different reasons for these different perceptions, including (i) access policy restrictions and (ii) information overload. Recommender systems are conceptually fit to help users perceive resources as (useful) known-knowns.

Access policy restrictions prevent users to access resources that would be of their interest. The recommender system mediates between the owner (that has the resource and can grant access to it) and the beneficiary (that is interested in the resource). Recommender systems will:

- recommend the owner with access policies to grant access permissions to another user upon the resource;
- recommend the beneficiary to request access permissions for a certain resource that is not accessible and that is a known-unknown, unknown-known or unknown-unknown to the reader.

Information Overload “occurs when the amount of input to a system exceeds its processing capacity” (Speier et al., 1999). In this context, information overload occurs because the owner is not able to match the large number of his/her protected resources with the potentially large number of interested readers. In that sense, recommender systems will:

- recommends the owner with suggestions of potentially interested users that are not able to access the resources;
- recommends the beneficiary, which is overloaded by the quantity of users that he/she would have to contact to request access to known-unknown, unknown-known or unknown-unknown resources.

2.3 Conceptual Architecture

Based on the nomenclature and responsibilities proposed by the Internet Engineering Task Force’s reference architecture for AAA in the Internet (Vollbrecht et al., 2000), this section describes the architecture for a system capable of accomplishing the envisaged goal.

When included as part of a multi-domain decentralised AAA system, the conceptual architecture sets the stage for defining protocol requirements between engaged systems.

Commonly accepted names for the various entities involved in the architecture are:

- Policy Enforcement Point (PEP) (Parducci and Lockhart, 2013; Vollbrecht et al., 2000; Westerinen and Schnizlein, 2001; Yavatkar et al., 2000);
- Policy Decision Point (PDP) (Parducci and Lockhart, 2013; Vollbrecht et al., 2000; Westerinen and Schnizlein, 2001; Yavatkar et al., 2000);
- Policy Information Point (PIP) (Parducci and Lockhart, 2013; Vollbrecht et al., 2000);
- Policy Retrieval Point (PRP) (Nair, 2013; Vollbrecht et al., 2000);
- Policy Administration Point (PAP) (Convery, 2007; Parducci and Lockhart, 2013; Stephen et al., 2008).

Other existing architectures use the concept of an IDentity Provider (IdP) that provide features for creating and maintaining users identity.

The decentralised structure is capable of providing authentication, authorisation, access control management and recommendation based on resources, users, provenance and traceability information in a distributed and decentralised system, by promoting the usage of action sensors, metadata generators and semantic rules (*cf.* Figure 2). This architecture is novel in respect to the following aspects:

- despite most of the components maintaining the same names as in typical architectures, their responsibilities and features are enhanced to address the defined requirements;
- adds a recommendation component that is responsible for the recommendation of access policies;
- boosts these components by replacing legacy and traditional non-standard formats and procedures with new data representation by using semantic web standards, capable of a better and explicit knowledge and information description.

In particular, managed and exploited information is a cornerstone of this work:

- resource, *i.e.* anything in the world that can be referred, either physical or virtual, that is identified;
- user, a special kind of resource representing a human or artificial agent in the system.

The conceptual architecture uses components that have been previously addressed by the authors in

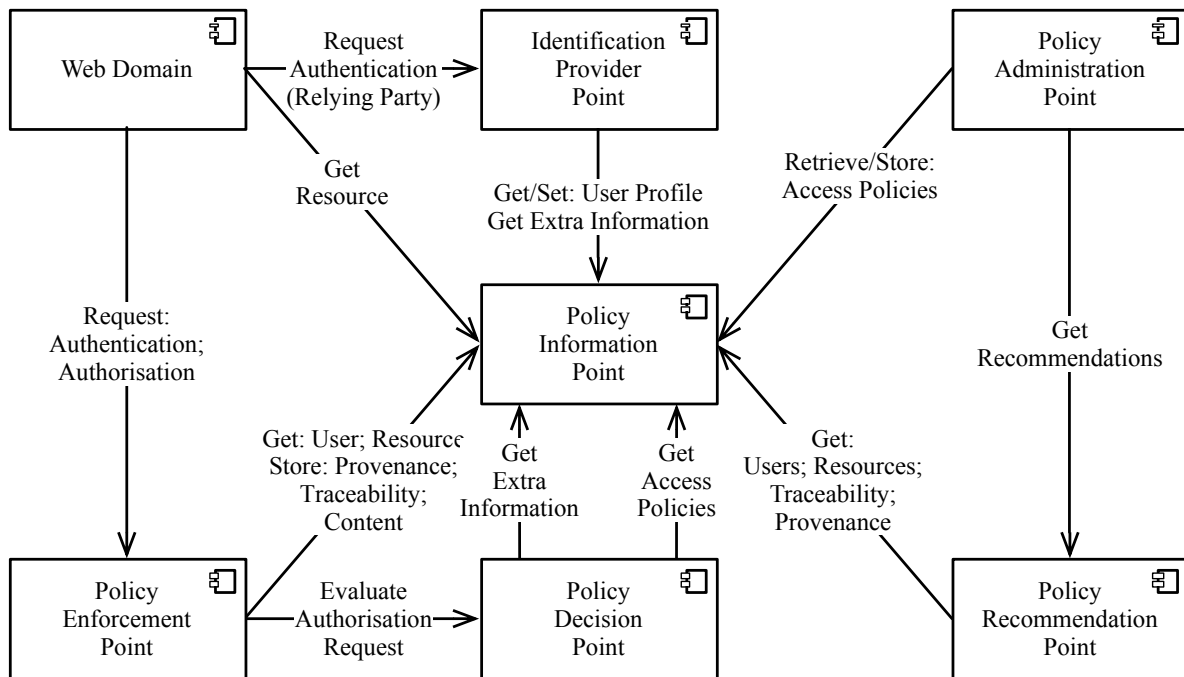


Figure 2: Conceptual Architecture.

(Bettencourt and Silva, 2010) and proposes the adoption of a new component which is the focus of the work being presented. Each component has specific responsibilities and features (*cf.* Figure 2).

In (Nimmons, 2012) the author suggests a typical operation pattern for providing resource authorisation. In this operation pattern, the PEP is responsible for intercepting access requests sent from the user to perform some type of action upon a resource. The PEP, on behalf of the user, requests authorisation for accessing the resource. This request is forward to the PDP, which is the entity that has the engine for evaluating access policies. It uses the information provided by the PEP and the specified access policies to determine if the user should be allowed or denied access to the resource.

The PDP uses the PRP and PIP to retrieve policies and attributes referenced in the policies. When the PDP finishes the evaluation of access policies, it returns an answer to the PEP stating whether access has been granted or denied to the user. If access is granted, the resource is retrieved from the hosting server. The PAP is the system entity used for managing the access policies. For that it uses the features of PRP to retrieve existing policies and store changes to those. Some of the component features are described next.

2.3.1 Identification Provider Point

Provides users with a new identity and appropriate credentials. The following features are enhanced or added:

- allows identity generation and credentials creation to new users;
- allows managing each user's internal and social identity in the virtual world;
- provides an authentication relying party service that allows legacy domains that do not provide Friend-Of-A-Friend + Secure Sockets Layer (FOAF+SSL) authentication to validate users credentials.

2.3.2 Policy Enforcement Point

Enforces user's authentication and guarantees controlled and authorised access to resources. The following features are enhanced or added:

- typical basic authentication methods are replaced by FOAF+SSL cross-domain authentication;
- enforcement is no longer achieved by using local access policies, but instead it is replaced by a distributed and decentralised method;
- action sensors capture User-Generated Content (UGC) and actions.

2.3.3 Policy Decision Point

Evaluates access policies in order to decide if a user should or not be granted access to a resource. The following features are added or changed:

- replaces traditional role or attribute based authorisation mechanisms by an authorisation mechanisms capable of handling semantic, declarative and expressive access policy languages;
- provides decentralised access policy evaluation that is used in a cross-domain perspective;
- obtains, if necessary, semantic information from the PIP for evaluating a particular policy;
- offers reasoning capabilities over more expressive access policy rules that exploit the system's semantics.

2.3.4 Policy Information Point

Manages the information needed for the authentication, authorisation and recommendation processes. The following features are added or enhanced:

- information management of:
 - resources' content, including their type, attributes/properties and preferred hosting domain;
 - provenance and traceability information over UGC and content;
- generating and publishing information according to an explicit and public semantic specification (*i.e.* ontology).

2.3.5 Policy Administration Point

Enables users to manage access policies over existing resources. The following features are added or enhanced:

- access policies are specified by rules instead of directly assigning users to resources or placing users in particular roles;
- proprietary access policies over resources imposed by closed domains are replaced by far more flexible and expressive rules that capture the rationale behind a particular access policy beyond current approaches;
- provides and promotes the means to create access policies based not only on user attributes and relationships, but also on resource attributes;
- provides and promotes the means to define more complex access policies through semantic reasoning over contextual information and meta-information.

2.3.6 Policy Recommendation Point

This component is a novelty in AAA systems. It recommends access policies that are applied to users and resources. These are some of the envisaged responsibilities and features:

- recommend access policies by combining collaborative, social content and semantic filtering methods, allowing the recommendation of known-unknown and unknown-unknown resources to users;
- allow customising the recommendation process, namely the weights for each filtering method.

This component proposal is addressed in the next section.

3 PROPOSAL

The Policy Recommendation Point recommends known-unknown and unknown-unknown resources to users in a cross-domain perspective, by exploring the information gathered by the system, namely user profiles, social network relationships, provenance and traceability information.

Having an access policy definition based on similarities between resources, users or domain knowledge, is being half-way to enabling an automatic recommendation system based on information such as FOAF profiles, interest topics and contexts to provide the sharing of resources.

Traditionally, the responsibility of sharing resources always comes down to the resource's author, based on his/hers restricted perception/knowledge of the whole network of users and resources. Resource access policy recommendation is a process that is introduced to widen that vision by which a system notifies the resource author when other users would probably benefit or rejoice from having access to a particular resource.

The access policy recommendation process aids resource authors in granting or denying access to existent resources by making use of similarity factors between resources and social relationships, suggesting which users should be given access to each resource.

It also eases resource authors' task of sharing resources by finding similar access policies that could be reapplied to similar resources. It is envisaged that recommendation can aid users in the access policy management process regarding their resources, and give other users access to resources that would not have previously been accessible to them.

This is achieved by enriching and enhancing the access policy recommendation process with existing users' and resources' meta-information, and creating a hybrid recommendation method capable of understanding not only the concepts of users and resources but also provenance and traceability annotations gathered from user actions.

A resource context is produced by the analysis of each resource's content and meta-information, while a relationship context is created based on the existing relationship depth between users (Wasserman and Faust, 1994), each user's profile, linked resources and consequent relationships.

One of the outcomes of this proposal is the creation of semantic rules that match similarities between contexts (Ghita et al., 2005). Therefore, for every resource or relationship, a context is generated and multiple contexts may exist for the same resource.

This PRP is responsible for:

- the implementation of a hybrid recommendation engine;
- guiding users through the resource-sharing process by suggesting access policies for their resources:
 - by evaluating feedback actions regarding the acceptance or rejection of recommended resource sharing;
 - avoiding rejected recommendations from being recommended again;
- recommending known-unknown and unknown-unknown resources.

3.1 Hybrid Recommendation Engine

When an application responsible for ensuring access control is aware of all users' resources and social relationships, such application is capable of recommending resources to new users that have recently become part of the resource author's social network.

This already happens on typically closed applications (*e.g.* Slideshare, Research Gate, etc.) but is still not being used in a cross-domain perspective for all user resources. Contrary to such closed environments, this proposal consists on performing such task in a cross-domain perspective.

The recommendation is enhanced with semantic information for cross-domain web applications relying on an open and distributed social network based on FOAF profiles, provenance and traceability information.

Users' public resources are used in the recommendation process to enable associations between users, between resources or between users and resources.

Despite already being publicly accessible, recommendation of publicly accessible resources is performed because other users that do not know of their existence can eventually have interest in them.

The proposed recommendation process consists of a hybrid approach accomplished by the combination of users' profiles, resources' meta-information, traceability and provenance annotations, social network analysis and domain knowledge.

The semantic filtering relates to problems as recommending known-unknown and unknown-unknown resources that users had little or even no knowledge about. The recommendation service is built on top of these three filtering methods that are capable of dealing with different sets of information.

The following methods are therefore suggested for the PRP:

Content-based Filtering Method. Recommends existing resources by comparing resource attributes, content and meta-information to the user's profile attributes and topic preferences in order to verify the resource's relevancy to the user. This relevancy is given by the similarity between resource attributes and the user's topic preferences. The content-based filtering method is enriched mainly by exploiting resources' content, resources' generated meta-information and users' interest topic preferences.

Collaborative Filtering Method. It recommends resources based on the following pairs of connections: (users, users), (resources, resources) and (users, resources). This process is content-agnostic, meaning that it only recommends resources based such these collaboration patterns, where similarities between users linked to resources are used to infer other new possible connections between users and resources. The collaborative filtering method uses information that associates users' actions to resources.

Context Filtering Method. Recommends resources that match the proposed user's topic preferences or semantically related topics. This filtering method expands the capabilities of the content-based filtering method by introducing reasoning over knowledge concepts. When the user context and resource context match, the recommender system recommends that resource to the user. Interest topics are semantically described, providing not only hierarchical relations between topics but also a graph of other connections between semantic information. Contexts are obtained through the usage of ontologies and semantic rules that provide grounding to this filtering method. The filtering method is enriched by semantic informa-

tion derived from multiple domains, that include users' FOAF profiles, topic interests, social network graphs, resources' meta-information, provenance and traceability annotations.

While the recommendation process runs continually, it is triggered by several changes in the system, namely:

User-generated Content. When users create, edit or change an existing resource, resource content is analysed by specific meta-information generators that generate semantic information. The recommendation process is triggered because changes in content might affect the result of the content-filtering (*i.e.* new content can be added or removed), collaborative-filtering (*i.e.* changing or adding a resource increases the number of times the resource has been accessed) and context-filtering (*i.e.* changing content may derive new context information) methods.

User-generated Actions. When users perform actions over resources, they are implicitly building their profile. When their profile changes, it is necessary to trigger a recommendation process because a change in a user profile might suggest access to other resources as it influences the collaborative and context-filtering method. Notice that revoking access permission might also be suggested if the resource is evolved through time and its applicability is over. In case the resource does not change and if it has been shared before, it makes no sense in revoking access rights because the resource might have been duplicated by others elsewhere.

Access Policy Modification. When users create, change or remove access policies, the recommendation process is triggered because other users may now have access to resources that they did not have before, which also influences the collaborative-filtering method.

Social Network Changes. Whenever a user becomes part of or is removed from another user's social network. In fact, this process is quite similar to the addition of new resources because a new user is actually a special case of a new resource that is identified by a corresponding Uniform Resource Identifier (URI). As a result, the user's context might change, which would trigger the recommendation process. The inclusion of a new relationship might change a user's context, which has an impact on the resources the user may have access to.

3.2 Notifications & Feedback

When the recommendation process succeeds in recommending access to resources, the resource author is notified with a message containing:

- the resource to be shared;
- the user to whom the resource is being shared;
- an explanation of why the resource is being recommended.

When a resource sharing is recommended, the system checks with the resource's author if he wishes to assign the access privilege to the proposed user. If the author wants to assign the privilege to the proposed user, the PRP takes the necessary actions to notify the proposed user. When authors accept resource sharing recommendations, these are translated into access policies over resources.

The author may receive recommendation notifications of access policies granting access to users that may not be part of his/hers social network. When sharing is recommended to users outside the author's social network, the inclusion of that user in the author's social network must be achieved prior to the sharing act, otherwise sharing is not permitted. To this end, the inclusion of a new relationship is proposed. If accepted, the author's FOAF profile is changed accordingly.

The proposed user who should be given access to the resource also receives a notification message stating:

- that a resource exists that might be suitable for the user;
- an explanation of why the resource is being recommended.

Each user receives a list of resources that were shared with him/her, and a request to express whether or not that resource is relevant to him/her, thus providing feedback to the recommender system. This feedback is captured in the form of traceability information and will be used as supporting information.

3.3 Known-unknown and Unknown-unknown Resources

In order for a system to be able to recommend the sharing of known-unknown and unknown-unknown resources, it must be possible to establish associations between resources, between users and between users and resources that are not possible to establish by means of content or collaborative analysis.

The semantic-filtering method uses ontologies to map existing information and allow the inference of

new knowledge by providing associations between resources that would not have been associated before.

Consider that resources (*e.g.* photos) have been annotated for having recognised but not identified a person that may or not be part of the resource author's social network. This unidentified person represents any possible user that may be interested in that resource, not because of any relationships with the user but because that person was at the same time and place where some photos were taken and could eventually appear in one or more. It is possible to narrow down the possibilities of people that could be passers-by at that location and time if the recognised but unidentified person is in the same context on which the photos were taken, and as a result recommend the resource sharing to that unidentified user, by using the following information:

- user profile;
- user contextual information:
 - users' geo-referenced position;
 - users' geo-referenced position's time;
- resource creation time and location;
- provenance and traceability information from user actions:
 - event records of their physical performance while practicing sports.

When the system discovers which unidentified users were at the same time and place, by comparing their location at a given time with the resource time and location, the resource's author is notified in order to share those resources with those particular users.

This type of recommendation can only be derived if different resources' contexts are matched. In this situation, time and location create the context for the presented resources. Nevertheless, this is just an example of a possible context. The conditions for specifying contexts can be fully captured by ontologies and semantic rules, thus being easily extended and reused by multiple recommendation system.

4 EXPERIMENTS

The aim of the experiments was to prove that even with a large dataset of information, semantic information would improve existing algorithms. For that, a larger set of information and a recommender system are required.

The recommender engine should feature a hybrid mechanism that makes use of collaborative, content

and semantic filtering techniques. Yet, these features are not natively supported by mainstream recommender systems.

Mahout recommender engine is a framework that provides advanced expansion features and makes use of collaborative filtering but it does not provide content or semantic filtering techniques, as these must use domain-specific approaches (Owen et al., 2011).

In order to provide this support with content and semantic filtering techniques, Mahout's recommendation process was modified to enable the aggregation of similarities between items and between users, together with Mahout's similarities generation.

Conducting the evaluation in a real world would be time-consuming and would hence face cold-start problems typically associated with collaborative filtering techniques. For these reasons, it was decided that the system should be evaluated according to an existing dataset.

Several datasets used on the Second International Workshop on Information Heterogeneity and Fusion in Recommender Systems (Hetrec'2011), were analysed in order to prove their appropriateness to the desired evaluation.

After a careful inspection of the content of the LastFM dataset it was clear that it would provide more useful information than the one in the Delicious Dataset or MovieLens, thus promoting the content and semantic filtering. For this reason, LastFM was the chosen dataset for the experiments as it suits the evaluation needs, considering a carefully planned interpretation and mapping to the ontology used in the system. The LastFM dataset is further enhanced with data from the Freebase and Music Brainz datasets.

Due to the lack of integration and explicit semantics of the source datasets, it is necessary to derive and integrate the implicit semantics from the existing datasets into a domain ontology. The mapping stage is responsible for converting the source datasets into a domain ontology. This mapping process is depicted in Figure 3. The dotted lines represent mappings from the source datasets to the domain ontology.

Each `lastfm:User` individual/instance gives origin to a `domain:User` individual. Listen and Tag actions are combined into the general domain's Action because Mahout recommender system does not distinguish between different types of user actions. Each LastFM Musical Artist individually originates a domain's Musical Artist.

The original `lastfm:Tag` individuals are interpreted as domain Musical Genres' individuals. This is the result of users' manual tagging of each Musical Artist. Yet, while these users' actions complement the Musical Artists with associations to Musical Genres, the

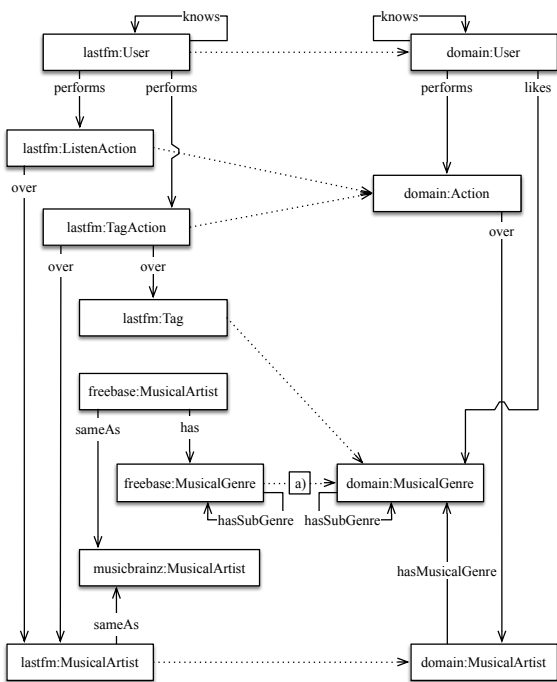


Figure 3: Source Datasets to Domain Ontology Mapping.

original LastFM dataset does not provide information about each Musical Artist and their related Musical Genres. In order to simulate the generation of semantic information, when UGC is captured, an enrichment process is performed for providing an association between domain:MusicalArtist and domain:MusicalGenre.

Domain’s Musical Genre individuals are obtained by the union of any Freebase Musical Genre:

- whose description matches LastFM’s Tag’s value by using a reconcile process. In the end of this process, 4698 of the initial 11946 tags were correctly reconciled to their semantic equivalent domain Musical Genre;
- that are tagged against the Musical Artist. Freebase’s and LastFM’s Musical Artist are not directly associated. Nevertheless, when a Music Brainz Musical Artist is the same for both Freebase and LastFM, one may conclude they are the same.

A transitive property “hasSubGenre” is added to the domain ontology to relate sub-genres. This “hasSubGenre” relation provides the necessary information for semantic filtering recommendation.

The process of generating the recommendation’s dataset consists in obtaining the following sets of information from the system’s ontology, to comply with the recommendation model presented in Figure 3.

The specific domain ontology is translated to a

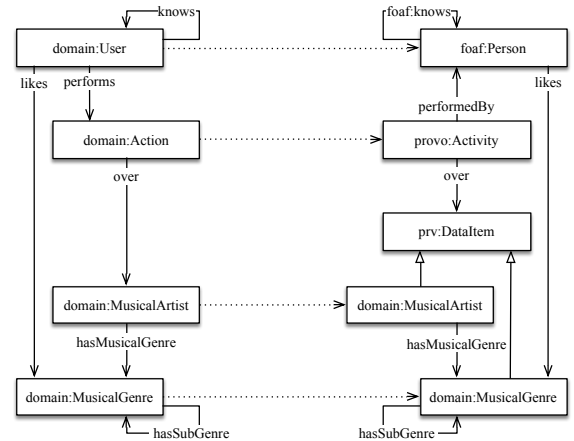


Figure 4: Domain Ontology to System Ontology Mapping.

generic ontology that is used by the system. The mapping between both ontologies is depicted in Figure 4.

Mahout’s recommendation process recognises users, items, and similarities between users or between items, user actions and their weights.

Because Mahout’s recommender system does not recognise or handle ontologies, a mapping between the system’s ontology and Mahout’s recommender model is necessary. It converts the system’s ontology data into a format that the recommendation engine can use (cf. Figure 5).

According to Figure 5, it is possible to derive the following concepts:

User. Derived from the foaf:Person concept. Each “foaf:Person” from the system’s ontology is mapped to “rec:User” in the recommendation dataset.

Item. Derived from the “prv:DataItem” concept. Each “prv:DataItem” is mapped from the system’s ontology to the “rec:Item” concept in the recommendation dataset.

Actions. Derived from the “provo:Activity” concept. Each “provo:Activity” from the system’s ontology is mapped into a rec:Action in Mahout’s. For each mapped activity, respective relationships with the user (“performedBy” property) and items (“over” property) are created.

User/User Similarities. Derived from the “foaf:knows” property. Each “foaf:knows” property originates a “rec:UserSimilarity” individual.

Item/Item Similarities. Derived from the “isSimilarTo” property. This similarity set is the outcome of the semantic filtering approach.

The weight of each user action, item similarity and user similarity is obtained by the number of rep-

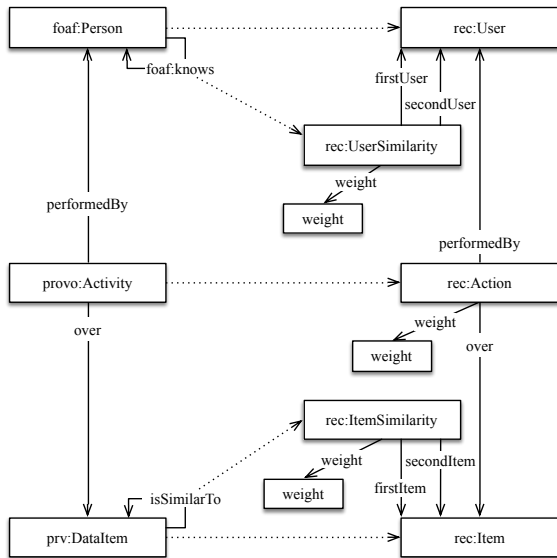


Figure 5: System Ontology to Recommendation Dataset Mapping.

itions that occur during the mapping process. The resulting dataset represents the input data for the recommender system.

5 EVALUATION

This evaluation suite gathers measurements of the recommendation evaluation execution under different runtime configurations. Some of the most relevant configurations are shown.

This section describes each configuration’s experiment and respective results. Experiments are characterised according to the following dimensions:

- the recommendation dataset;
- the process of generating the training model and relevant items;
- the process of generating and aggregating similarities;
- the recommendation engine configurations (*e.g.* AT).

Each experiment has its own configuration of these dimensions. The experiments were conducted for a top AT of 25, 50 and 150.

Each configuration evaluation consists in the calculation of average precision, recall and f1.

Experiment’s results are compared to those of an initial baseline experiment that is obtained by using the dataset with the simplest possible configuration.

Baseline configurations were created using Mahout’s algorithms without injecting any extra similar-

ities in the process, as depicted in Table 1 as configuration C1.

The configurations derived from the C1 baseline configuration are configured with an item-based boolean recommender that uses the Log-Likelihood Similarities algorithm as shown in Table 1.

By using the baseline recommender configuration solely with semantic similarities (*i.e.* C105), precision and recall values drop when compared to the baseline (*i.e.* C1).

Yet, when aggregating both the recommender system similarities and the semantic similarities, using an approach without averaging both similarity sets weights (*i.e.* C104), it produces much better results: precision is about six per cent higher, recall around twenty-nine per cent and f1 about ten per cent higher than the baseline and the normalised averaged approaches (*i.e.* C109 and C110) with union or average intersection.

Using a normalised approach with intersection provides worse results than a non-normalised union of all results.

Based on these results, it is possible to conclude that an item-based boolean recommendation is better when enriched with semantic similarities compared to the baseline configuration.

6 CONCLUSIONS

The access policy recommendation process aids resource authors in granting or denying access to existing resources by making use of similarity factors between resources and social relationships and suggesting which users should be given access to each resource.

As any other recommendation process, the one proposed is based upon three main parts: users, resources and associations between users and resources. Yet, provenance and traceability annotations, users’ social awareness, list of interest topics, resources’ and users’ context are used in the recommendation process to infer users’ interest in resources.

The addition of a semantic-filtering method that is capable of using contextual semantic information to the existing recommendation engine proved to enhance the results, even though it was achieved by only using a minimal subset of information that a semantic system can have.

The results demonstrate that introducing similarities calculated from content and semantic information into a collaborative filtering technique – either focusing on social networking, user profiles or resource

Table 1: Experiments Configuration and Results.

Configuration ID	Similarity		Prediction		Aggregation				Measures			
	Log-Likelihood	Semantic	User/Item-Based	Boolean/Weighted	Normalised/Std.	Union	Union Average	Intersection Average	AT	Precision	Recall	F1
C1	L	-	I	B	-	-	-	-	25	0,0814	0,4969	0,1399
C1	L	-	I	B	-	-	-	-	50	0,0345	0,6086	0,0654
C1	L	-	I	B	-	-	-	-	150	0,0193	0,6677	0,0376
C104	L	S	I	B	S	UN	-	-	25	+0,0605	+0,2874	+0,1005
C104	L	S	I	B	S	UN	-	-	50	+0,0143	+0,1988	+0,0266
C104	L	S	I	B	S	UN	-	-	150	+0,0054	+0,1496	+0,0103
C105	-	S	I	B	-	-	-	-	25	-0,0018	-0,0427	-0,0045
C105	-	S	I	B	-	-	-	-	50	-0,0078	-0,1516	-0,0149
C105	-	S	I	B	-	-	-	-	150	-0,0059	-0,2095	-0,0116
C109	L	S	I	B	N	-	UA	-	25	+0,0456	+0,2011	+0,0750
C109	L	S	I	B	N	-	UA	-	50	+0,0143	+0,1974	+0,0266
C109	L	S	I	B	N	-	UA	-	150	+0,0055	+0,1530	+0,0105
C110	L	S	I	B	N	-	-	IA	25	+0,0455	+0,2000	+0,0748
C110	L	S	I	B	N	-	-	IA	50	+0,0143	+0,1973	+0,0266
C110	L	S	I	B	N	-	-	IA	150	+0,0055	+0,1527	+0,0105

content – it is possible to improve recommendation results.

In (Said et al., 2011) the authors used the MovieLens dataset for measuring the system’s recommendation performance, using a mean average precision measure. Their precision values for an AT of 50 vary from 0.0272 to 0.0687, which are on par with the values obtained by the experiments conducted with the baseline configuration for the same AT (0.0255 to 0.0345). In the conducted experiments precision barely drops below 0.0500 hitting a maximum of around 0.0800 for a top AT value of 25, which is better than the best values (0.0699, AT=5) observed by the authors in (Said et al., 2011). This proves that precision measures produce quite small results but yet good enough for providing comparison between different systems in an evaluation phase.

The usage of similarities produced from semantic content injected in collaborative-filtering techniques, shows that precision values higher than ten per cent are easily achievable. Provenance and traceability information, together with enriched semantic information, can indeed make the resource recommendation better.

In summary, the proposed system architecture provides the following features and functionalities:

- the resource author is recommended with new

access policies that would facilitate sharing resources with other users;

- it allows discovering resources that users did not even know existed;
- users can be given a list of resources which match their interests or contexts, even though specific names and content are not shown unless the author gives them permission to access it, *i.e.* the resource author will know which user is requesting access but the requesting user does not know who is the author;
- semantically enhanced recommendations, allowing the creation of contexts (*e.g.* time and space) for resources and users.

The adoption of a hybrid access-policy-recommendation engine enables the enrichment of access policy recommendations by using additional information provided by the system.

Captured provenance and traceability information are used together with the user’s social networks and resources’ contents as to automatically propose which access policies should be added to a certain resource.

ACKNOWLEDGEMENTS

This work is supported by FEDER Funds through the “Programa Operacional Factores de Competitividade - COMPETE” program and by National Funds through “Fundação para a Ciência e a Tecnologia (FCT)” under the project Ambient Assisted Living for All (AAL4ALL - QREN 13852).

REFERENCES

- Adomavicius, G. and Alexander, T. (2011). Context-Aware Recommender Systems. In Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B., editors, *Media*, chapter 7, pages 217–253. Springer US.
- Bettencourt, N. and Silva, N. (2010). Recommending Access to Web Resources based on User’s Profile and Traceability. In *the Tenth IEEE International Conference on Computer and Information Technology, CIT’10*, pages 1108–1113, Bradford, UK. IEEE.
- Convery, S. (2007). Network Authentication, Authorization, and Accounting: Part One: Concepts, Elements and Approaches. *The Internet Protocol Journal*, 10(1):2–11.
- Ghita, S., Nejdil, W., and Paiu, R. (2005). Semantically Rich Recommendations in Social Networks for Sharing, Exchanging and Ranking Semantic Context. *Social Networks*, 3729:293–307.
- Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80.
- MacKinnon, K. A. (2012). User Generated Content vs. Advertising: Do Consumers Trust the Word of Others Over Advertisers? *The Elon Journal of Undergraduate Research in Communications*, 3(1):14–22.
- Nair, S. (2013). XACML Reference Architecture. <http://developers.xiomatics.com/blog/index/entry/xacml-reference-architecture.html>. Accessed on May 3, 2015.
- Nimmons, S. (2012). Policy Enforcement Point Pattern. <http://stevenimmons.org/2012/02/policy-enforcement-point-pattern/>. Accessed on May 4, 2015.
- Owen, S., Anil, R., Dunning, T., and Friedman, E. (2011). *Mahout in Action*. Manning.
- Parducci, B. and Lockhart, H. (2013). eXtensible Access Control Markup Language (XACML) Version 3.0. Technical Report January, OASIS.
- Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation Systems. *Communications of the ACM*, 43(12):45–48.
- Ruohomaa, S., Kutvonen, L., and Koutrouli, E. (2007). Reputation Management Survey. In *Second International Conference on Availability, Reliability and Security, ARES’07*, Vienna, Austria.
- Said, A., Kille, B., De Luca, E. W., and Albayrak, S. (2011). Personalizing Tags: A Folksonomy-like Approach for Recommending Movies. In *Proceedings of the Second International Workshop on Information Heterogeneity and Fusion in Recommender Systems, HetRec’11*, pages 53–56, Chicago, IL, USA. ACM.
- Schafer, J. B., Konstan, J. A., and Riedl, J. (2001). E-Commerce Recommendation Applications. *Data Mining and Knowledge Discovery*, 5(1):115–153.
- Shardanand, U. and Maes, P. (1995). Social Information Filtering: Algorithms for Automating “Word of Mouth”. In the Proceedings of the ACM Conference on Human Factors in Computing Systems, editor, *Katz, I R and Mack, R and Marks, L and Rosson, M B and Nielsen, J*, volume 1 of *CHI ’95*, pages 210–217. ACM Press/Addison-Wesley Publishing Co.
- Speier, C., Valacich, J. S., and Vessey, I. (1999). The Influence of Task Interruption on Individual Decision Making: An Information Overload Perspective. *Decision Sciences*, 30(2):337–360.
- Stephen, L., Dettelback, W., and Kaushik, N. (2008). Modernizing Access Control with Authorization Service. *Oracle - Developers and Identity Services*, (November).
- Vollbrecht, J. R., Calhoun, P. R., Farrell, S., Gommans, L., Gross, G. M., de Bruijn, B., de Laat, C. T., Holdrege, M., and Spence, D. W. (2000). AAA Authorization Framework [RFC 2904]. The Internet Society.
- Wasserman, S. (2012). The Amazon Effect. <http://www.thenation.com/print/article/168125/amazon-effect>. Accessed on April 27, 2015.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Structural analysis in the social sciences. Cambridge University Press, 1 edition.
- Westerinen, A. and Schnizlein, J. (2001). Terminology for Policy-Based Management [RFC 3198]. The Internet Society.
- Yavatkar, R., Pendarakis, D., and Guerin, R. (2000). A Framework for Policy-based Admission Control [RFC2753]. The Internet Society.