

# Movie Recommendation on Web using Ontology and User Defined Tags

Ujwala Bharambe

Assistant Professor

Thadomal Shahani Engineering College, Bandra(W)  
Mumbai-50

Arti V Waghmare

M.E. Student

Thadomal Shahani Engineering College, Bandra(W)  
Mumbai-50

## ABSTRACT

Internet provides great amount of heterogeneous information. Thousands of news articles and blogs post each day. Millions of movies, books, music tracks are becoming available on internet. But we really need and consume only few of them. To recommend to us something we may like, we need an intelligent Web Recommendation system. Recommendation Systems are limited by several problems, of which are sparsity, and the new user problem[1]. They also fail to make full use and harness the power of domain knowledge and semantic web ontology[1]. Use of Ontology with relations provides better interpretability of recommendation results. Diversity of standards, languages, protocols, and hardware components leads to important incompatibility issues when designing and developing multiplatform multimedia systems. Furthermore, user and community requirements and preferences should be taken into account when instantiating and configuring these kind of systems [11]. The goal of the proposed system is to develop a web recommendation system by using the concept of domain ontology and user provided tags. Introduction of user provided tags which acts as the labels or meta data are becoming very familiar and popular. In the proposed system tags and the concepts of domain ontology are processed using hybrid similarity to provide the results of recommendation. The system would benefit those users who have to scroll through pages of results to find relevant contents. The proposed measure based on hybrid similarity can be adopted effectively in this application of movie recommendation.

## General Terms

Recommendation, weblog, sparsity, MS Word, web.

## Keywords

Ontology, Tag, SPARQL, wordNet, RDF, OWL.lcs.

## 1. INTRODUCTION

Recommendation system is an information filtering technique, which provides users with information which he/she may interested in. The filtering of information is often quantitative, based on collaborative or social filtering [5]. The alternate approach based on qualitative information retrieval deals with the content analysis. The semantic of the content leads to the more specific results of recommendation. The semantic mapping needs complete specification of the domain knowledge in machine learnable form. Ontology has been

widely used to represent many real world aspects and is prominently used as tool to facilitate shared understanding (and knowledge sharing) in a particular domain. Traditional recommendation systems mainly depends on web usage mining as an underlying architecture which is concerned with

finding user navigational pattern on the world wide web by extracting knowledge from web logs[1]. A web recommendation system uses this knowledge to predict user's preferences and recommends certain web sites or products. These content based web recommendation systems do not usually follow domain ontology to extract knowledge about the recommended items [1]. The proposed application deals with content based web recommendation based on domain ontology. It relies on user provided tags that are mapped to the concepts of ontology. User provided tags associate label to resources in order to find them again [2]. User provided tags may describe content, type or features of the resource [2]. These tags may lead to ambiguous and non useful information which cannot be mapped to a valid domain resource [2]. To overcome this drawback which is often found in social tagging network, the proposed system restricts the user provided tags to the specific types and allows nouns specifying actor/actress or director names in combination or alone which are very common in movie domain. Nouns provided as tags cannot be located in wordNet dictionary to process further. For this along with the wu-palmer, the string similarity is added to get the required results of recommendation. Replacing the traditional relational database with domain knowledge specification using ontology and the user provided tags makes an application more flexible and accurate in the area of recommendation on World Wide Web.

## 2. PROBLEM DEFINITION

Web discovery applications like Stumble Upon, Reddit, Digg, Dice (Google Toolbar) etc to name a few are becoming increasing popular on the World Wide Web. Information on the internet grows rapidly and users should be directed to high quality contents those are relevant to their personal interests. Displaying quality content to users based on ratings or past search results are not adequate. There is a lacking of powerful automated process combining human opinions with machine learning of personal preference. This leads to the recommendation system.

The reason manuals are not used to their full extent is that users often cannot find what they are looking for, regardless whether the manual is published online or printed on paper. A likely explanation for the user's problem with a manual is the incompatibility between a manual's index terms and the terms the users are likely to use for their searching. For instance, when you are looking for the programming language that is used in MS Word to create macros, searching the help system for the keywords 'programming language' will not result in the answer 'Visual Basic for Applications'. You would have to

know that it is referred to as script language or macro language. Often, more prior knowledge is needed than can be reasonably expected from users. This problem is not unique to the design of instructional documents, but is common to the design of all large searchable information systems. Some systems that contain large numbers of content items have introduced a more free-form way of indexing content. Instead of the content being indexed by professionals based on a fixed vocabulary, users themselves can classify and organize the content by means of tags. Tags are user-supplied access points for information resources [3], e.g. short labels that characterize pieces of information. Tags offer user-generated indexes to content. Apart from the collaborative construction of a shared vocabulary, an important advantage of tagging is that users are not restricted to a hierarchy when they attempt to find content, because tags are not explicitly related to each other [2]. Finally, tags are the product of a user's frame of reference. As such, they can inform the system about a user's preferences. These preferences can be used to adapt content to a user's needs in an adaptive information system such as a recommender system.

The above discussed points motivate us to develop a tag based web recommendation system mainly for the movie domain.

### 3. CONTRIBUTION AND OUTLINE

The proposed article contributes to the area of recommendation by showing (1) how domain knowledge can be incorporated into an application with the help of ontology. (2) The typical problems of recommendation system can be overcome with the use of tags rather than the using navigational pattern or the browsing history. (3) Hybrid similarity takes into account the wordNet based similarity and the string matching algorithm to work with words which are not defined in wordNet. Next in section 4, we introduce the proposed software architecture of recommendation system followed by four steps of recommendation, namely developing domain ontology, web log creation, similarity score matrix generation using wu-palmer and string matching for noun based recommendation and providing recommendation using cosine similarity. Then each step is detailed in section 4, section 5, section 6 and section 7,8 respectively. Section 9 will focus on corresponding results of recommendation followed by conclusion and future work in the section 10.

### 4. RECOMMENDATION OVERVIEW

In general the complete Recommendation system goes through following steps:

**Developing Domain Ontology:** Ontology construction is a highly relevant research issue depending on the extraction of information from web and emergence of ontologies. It is an iterative approach where every iteration can further refine the existing knowledge base.

Ontologies are constructed using some ontology supporting languages like RDF, OWL and the ontology editor such as Protégé. etc. and connected to each other in a decentralized

manner to clearly express semantic contents and arrange semantic boundaries to extract concrete information.

**Web log creation:** Every movie being tagged by user would be stored in the web log. Each entry of the log file corresponds to the movie title separated by the tag information. During recommendation this log file will be used to generate similarity between tag and the domain concept.

**Similarity score matrix generation using wordNet similarity and string matching:** This matrix will have a numeric value between 0 to 1 which corresponds to the similarity score between the tag word and the domain concept. The matrix will be generated on fly and will be use by the module of recommendation. In case of the movie recommendation generally nouns representing actor/actresses and the director are often used as a tag word. These nouns do not have dictionary meaning, as a result of which system could not map them to the domain concept. For this string similarity is added to the logic of recommendation.

**Recommendation using cosine similarity:** Cosine similarity is often used in vector based comparison. Cosine similarity value is further processed to generate set of recommended movies.

On the basis of steps stated above we have designed the following skeleton of ontology based movie recommendation system. The system consists of four basic modules: User Interface, Domain ontology construction, Tag processing and Tag based search as shown in the figure 1:

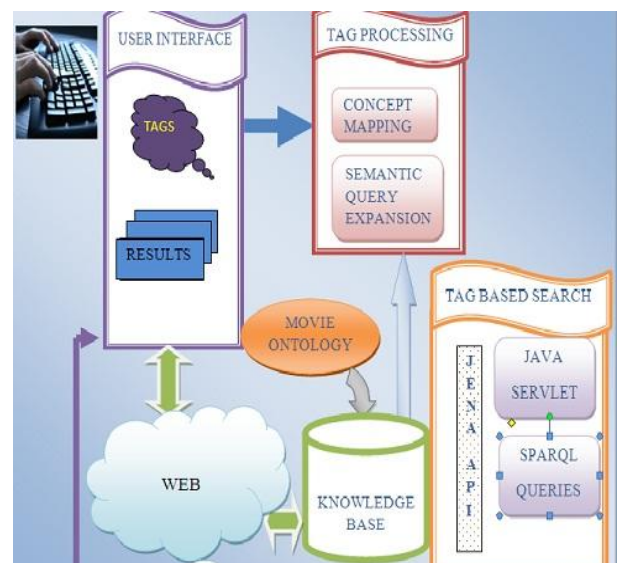


Figure 1

**User Interface:** A sophisticated user interface is provided wherein user can provide set of tag words. After processing of other subsequent modules user will get the results of recommendation.

**Domain Ontology construction:** Ontology is used to model the domain knowledge which will be an iterative process. With every iteration the ontology will be refined further with new classes and relationship get added/removed as a part of refinement.

**Tag processing:** As a result of input tag entered by user web log will be accessed to make the corresponding entry into it. This entry would later be used to generate similarity score value using wordNet based similarity measure. The item versus concept matrix will preserve this value for the corresponding tagged concept.

**Tag based Search:** For the inputted tag similarity score vector is generated. This vector is matched with every row of item versus concept matrix. The relatedness is found using cosine similarity. The item versus concept matrix is then sorted in descending order. The sorted entries will be finally used to generate the set of recommendation.

In case of casting or direction being specified as tag word, string similarity module will be invoked to match the tag word with domain instance. On successful match corresponding SPARQL query will be executed to generate the desired results of recommendation.

## 5. DEVELOPING DOMAIN ONTOLOGY

Since the beginning of 1990s ontology has become a popular research topic investigated by Artificial intelligence research communities including knowledge engineering, natural language processing and knowledge representation [09]. Domain ontology is the formal representation of domain knowledge by the set of concepts within the domain and relationship among them. Protégé is an effective ontology editor used to construct ontology.

The reason ontology is becoming so popular is because its promise a shared and common understanding of some domain that can be communicated between people and application systems. Because ontology aims at consensual domain knowledge, its development is often a cooperative process involving different people possibly at different locations. According to Wikipedia in computer science and information science ontology is formal representation of knowledge by a set of concepts/classes within domain and the relationship between those concepts. It is used to reason about the properties of the domain and may be used to describe the domain [10].

Ontology development involves following steps:

**Define scope of the domain:** Basic questions that can lead us to define scope of the domain are (1) What is the domain that ontology will cover. (2) For what we are going to use the ontology. (3) What type of question should ontology will provide answers. (4) who will use and maintain the ontology.

**Consider reuse:** We can use the ontology in the same domain knowledge if it exists.

**Enumerate terms:** Identify relevant terms Write down an unstructured list of all relevant terms. Nouns will be basis for class name and the verbs will form the property.

**Define taxonomy:** Organize relevant terms in taxonomy hierarchy.

The above steps of ontology development are taken into consideration while developing the movie ontology which is reflected in the following concept hierarchy in figure 2:



Figure 2

## 6. WEB LOG CREATION

The log file will store tag information associated with each movie. This file will be processed further to segregate the list of all tags associated with each movie.

Following is the sample log file entry in the domain of movie:

```
Log={<Movie1,["comedy","horror","romance"]>,<Movie2,["animation","fantasy"]>}
```

These shows the sample log file entries, where Movie1 and Movie2 are provided with set of specified tags.

## 7. SIMILARITY SCORE MATRIX GENERATION USING WU-PALMER AND STRING SIMILARITY

Similarity score matrix plays a valuable role in our proposed system. This matrix will have numeric values between 0 to 1. This matrix will be generated using wu-palmer similarity based on wordNet. Wu-palmer takes into account the depth of nodes together with the depth of their least common subsumer (lcs). The formula of which is as given below in equation 1:

$$wup(c1; c2) = \text{depth}(lcs1;c2) / \text{depth}(c1) + \text{depth}(c2) \text{ -----}(1)$$

Similarity score between two words representing tag word and the domain concept gives the numeric value that ranges between 0 to 1. As wordNet is the basis to compute similarity score between tag word and domain concept, nouns like name of actor/actresses or director cannot be located in wordNet and the similarity score cannot be generated. For this purpose the proposed application adds string matching logic and accordingly executes SPARQL (Simple protocol and RDF query language) to retrieve the corresponding movies having the specified casting or direction.

Samples entries for item versus concept matrix is as shown in figure 3.

Movies	Concepts				
	c1	c2	c3	c4	.....
m1	0.0	0.0	0.0	0.0	
m2	0.66	0.85	0.363	0.0	
m3	0.55	0.66	1.0	0.54	

Figure 3

## 8. PROVIDING RECOMMENDATION USING COSINE SIMILARITY

The item versus concept matrix is then processed further with the help of cosine similarity to generate the set of recommendation. The cosine similarity value is used as a basis for sorting the item versus concept matrix to get recommended results in descending order of cosine similarity.

## 9. EXPERIMENTS AND RESULTS OF RECOMMENDATION

To test the movie recommendation application the manual dataset is created with reference to the movie ontology developed using protégé editor. The dataset has around 200 instances of movie type with the slots values assigned to each of them. wordNet is used for similarity computation. Recommendation module refers to the item versus concept matrix to get the recommendation results. The SPARQL query execution results of string matching module are merged with the results of recommendation in case of name of the actors/actresses or director given as tag input.

The proposed application incorporates hybrid similarity by taking into account standard wordNet based similarity along with string similarity.

Graph in the Figure 5 shows the graphical results after testing the system performance with and without hybrid similarity. Horizontal axis represents the tag input given in combination of actor/actress/director and genre type tag words, vertical axis shows the number of relevant items recommended from the available dataset.

The System gives better and efficient performance when string similarity is introduced to produce the recommendation when input tag word represents actor name or actress name or director name alone or in combination. WordNet based similarity does not have provision to find the nouns defined into it, hence unable to compute similarity score between the casting of the movie and the domain concepts. String comparison along with SPARQL gives the desired results of movie recommendation.

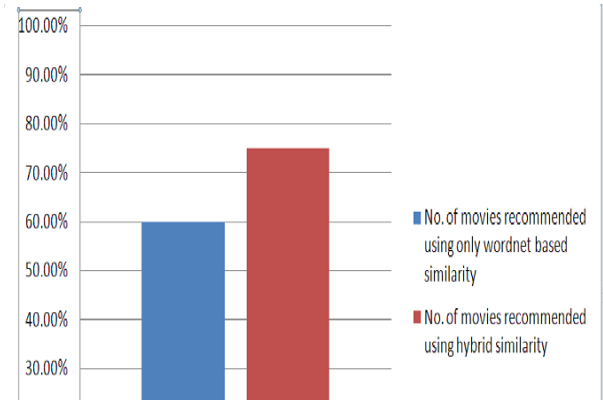


Figure 4

The movie ontology developed has various instances of film class with the necessary attributes representing the part of data set. A sample entry is shown in the figure 3.

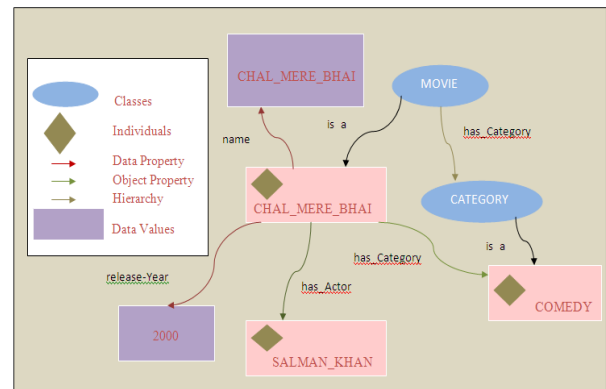


Figure 5

## 10. CONCLUSION AND FUTURE WORK

A content based web recommendation system is developed based on domain ontology. It relies on user provided tags that are mapped to the concept of domain ontology. Similarity measure is used during mapping and a matrix of item versus concept is built offline. Collecting the domain knowledge in the form of ontology has several benefits over traditional databases like relational database. Similarity based on wordNet can be further combined with other measures of similarity so as to incorporate hybrid similarity which can be helpful to work with any natural language terms to relate with each other.

In future we can connect our service with movie ticket reservation system where user can be recommended and can reserve the tickets of movies around at a time using our application.

## 11. ACKNOWLEDGMENTS

Our thanks to Mr. Jayant Gadge who had given us an inspiration to work in this domain. We would also like to thank all our family members who always encouraged & motivated us.

## 12. REFERENCES

- [1] G Nizar R. Mabroukeh, C. I. Ezeife , “Ontology based Web Recommendation from tags”, IEEE 2011 27<sup>th</sup> International conference on data engineering.
- [2] Mark Melenhorst Mark van Setten,” Usefulness of Tags in providing access to large information system”, IEEE 2007 International conference on hypermedia and multimedia.
- [3] Ana Belen Baragann Martinez, “What’s on TV tonight? Effective and efficient personalized recommendation system on TV programs”,IEEE 2009 transaction on consumer electronics.
- [4] A practical introduction to ontology and owl”, Online material published by University of Manchester.
- [5] Song Hui, Lu Penduyu,Zhau Kai,”Improving item based collaborative filtering recommendation system with tag”, IEEE 2011 International conference of world wide web.
- [6] Similarity measure based on edge counting”, world academy of science, engineering and technology in 2006.
- [7] Antonia Dattolo, Felice Ferrara, Carlo Tasso, “Role of Tags for recommendation a survey”, IEEE 2010 3rd International conference on human system interaction.
- [8] Freddy Lecue, “ Combining collaborative filtering and semantic content based approaches to common web services”, IEEE 2010 4<sup>th</sup> International conference on semantic web computing.
- [9] Kaushal Giri,”Role of Ontology in semantic web”,DESIDOC Journal of Library and Information Technology , Vol. 31, No. 2, March 2011, pp 116-120.
- [10] Gruber, T.A translation approach to portable ontology specifications. Knowledge Acquisition, 1998, 5, 199.
- [11] Ernesto,Expositon,Jorgge,Montalvoo,Myriamm,Laamoa ,”Mulimedia ontology driven architecture for Multimedia System”, 2010 second International Conference on Information, Process and Knowledge management.