3.1 REVIEWS ON EXISTING TAG RECOMMENDATION APPROACHES

In order to analyze the Tag Recommendation approaches, reviews are carried out on various recommendation approaches such as content based (Lu et al 2009), Item based (Wetzker et al 2009) Folksonomy oriented approaches (Lipczak 2008). Consequently, merits and demerits of these approaches are clearly pointed out and efficacious and expressiveness of tag recommendation approaches are also explained.

3.1.1 Content Based Recommendation Approach

Content based approach is used to suggest the tag with or without preceding tag information. Recommendation is applied to a webpage not only for suggestions but also to comment the page. Entropy metrics are used to define the tags that indicate an annotated document. Similarity measure is computed between two documents for tags and terms. Vector space model represents the terms and tags as vectors. Cosine similarity is calculated for these tag vector and term vector. Tags are useful in improving search results and tag suggestion reduces difficulties of users in tagging and supports them to make use of more tags to decrease the complexity. In content based tag recommendation approach, the problem is divided into a problem of
supervised learning in which classifier is trained with page text, anchor text, available tag information and surrounding hosts for each tag.

This approach obtains higher precision and yields good results with small tag information using association rules. Similarity between tag vector and term vector is computed and combination of these similar vectors became the similarity of URLs. Propagated tag weight is computed for a URL. Term propagation is also used since URL tags the other one if they are same. The main drawback of this approach is that it has opportunity to improve the precision for suggesting useful tags. Content based approach suggests the popular tags, which is relatively unclear and not as informative as particular tags. Popular tags are related to only lesser quantity of web pages therefore, it is difficult to suggest tags for all resources (Lu et al 2009).

3.1.2 Item Based Recommendation

An user can tag his interested items for item recommendation in folksonomies. Probabilistic latent semantic analysis (PLSA) approach is extended and recommendation model is presented to develop from item user and item tag occurrences. In this approach, topic model is computed from item user and item tag annotations. It facilitates users to assign tags in order to add and handle the bookmarks. PLSA relates the co-occurrence of annotations with topic variable. Co occurrence matrix in collaborative filtering provides relevant annotation to item bookmarking. Probability of bookmarked item was calculated on adding all the underlying variables. Weighted items can be suggested to the users based on probability and items bookmarked in training set is assigned to 0 weight and it is added to the end of the suggested item list. Fusion independent model improves the Quality of recommendation for underlying topics, even though it reduces with greater k values. Using PLSA algorithm, tagging information and its usage is combined to unified model.
Time taken for suggesting tags is also reduced with this approach. There is also room for improving recommendation quality (Wetzker et al 2009).

3.1.3 Folksonomy Oriented Recommendation

Suggested tags are mainly focused on folksonomies of individual user (Lipczak 2008). The following three steps are involved in tag suggestion system. Initially, tags are taken from resource title words and score is assigned to tags, denote efficiency for already tagged resources. Second thing is, lexicon oriented resource tag is utilized to offer tags associated with title tags. Later on, the user’s personomy checks and filters the tag. Generally, suggested tags are amalgamation of large amount of tags generated using above mentioned procedure. Folksonomies permit users to save and share different kinds of internet resources. Folksonomy is categorized into two types. First is Bibsonomy and delicious based on the intention of tagging process. The major role of tags is to make a storage area personalized to individual user interests called Folksonomy based individual users. Another one is Flickr and Technorati public resources called Folksonomy based wide audience. Folksonomy based individual users are focused in this approach on noticing the straight post adjacent to suggest tags achieved from resource. Based on the Folksonomy character, tags are utilized from the resource. In Bibsonomy, resource is an entry of bibtex or webpage bookmark. Resource title is a strong foundation for tag suggestions. Bibsonomy snapshot generates any tags and title is related to the content of the resource. This draws low recall and gives unsuitable title. In resource tags, other users of Folksonomy assigned tags are not a fine source for tag suggestion since data sparsity exists. It confines the recall of basis tags and harms the retrieved tag precision.

The lexicons do not provide nature of relation information, but lexicon can list the relevant tags for possible suggestions. Lexicon has
relations among tags and can be utilized for resources. This lessens negative contact of sparse data. Furthermore, lexicon is applicable only for specific Folksonomy and can obtain particular relations among the tags. A Personomy tag is based on the representation of user interests within limited number of tags. User is probable to attach with lexical form of appearance or words using singular or plural type of noun continually. This will provide positive user interests. User will go for extra tags to define more precise resource. This causes huge amount of less frequent tags and make complex of the recommendation retrieval process from personomy. Personomy of user is more accurate basis for accurate tag recommendations. But, the one disadvantage is that the evaluation of dataset characteristics reduces its consistency (Lipczak 2008).

3.1.4 Flickr and Zoomr

Garg and Weber (2008) proposed a personalized method to tag recommendation for Flickr. Flickr and Zoomr are the online photo services to share users’ photos with friends, family and online group. This existing work analyzed how user tags photos and what information exists in tagging. Online photo services discussed a tag characterization and performance. This is a basis for tag recommendation system. By observing photo tag distribution, only few tags are used to comment on photos. Tag co occurrence data have been extracted based on delegated models of Flickr. Tag aggregation methods are efficient but, it is necessary to get co occurrence values of the contender tags while collecting the outcomes in an ordered list of suggested tags. This approach is mainly good at suggesting environments, work of art and objects. Flickr is an open tagging method which always develops vocabularies. Tag co occurrence was increasingly updated when new comments are available. Related tags are recommended efficiently for diversity of photos with various levels of far-reaching of original tagging.
3.1.5  **Document and Graph Centered Approach**

Document centered methods were presented for effective tag recommendations. They are graph centered and prototype approaches. Graph centered approach denotes the words relationship, documents and tags into two bipartite graphs, followed by separating the graph into secondary graphs as topic group. It denotes the frequency based tag ranking in each topic, executes a new document categorization and most likelihood tags are recommended. Prototype centered approach is utilized to decrease the learning difficulty. This approach is intended to find the most descriptive subset inside the training set instead of utilizing the whole training set. In this approach, documents are categorized into predefined classes. Based on joint likelihoods tags are ranked and suggested to a new document as graph centered approach. These processes are same for both approaches. Most related tags are selected from the classes as machine suggested tags (Song et al 2011).

### 3.2  **REVIEWS ON RECOMMENDATION ALGORITHMS**

An adaptation of user based collaborative filtering and a graph based recommender systems are used for tag recommendation. These are constructed on Folk Rank and adaptation of the notorious PageRank algorithm can deal with undirected triadic hyper edges (Jaschke et al 2007). Both algorithms are examined and compared on significant real life datasets and experimental results are demonstrated better outcomes than non tailored baseline schemes. Particularly, the graph based recommender performs better. But, this method yields low precision as it constantly returns predetermined number of tags. This section will analyze and discuss the existing recommendation algorithms such as Hybrid ANN (Lee et el 2007) and Poisson Mixture Model (Song et al 2008) algorithms, Latent Dirichlet Allocation (Blei et al 2003), Multinomial Naive bayes categorization
(Illig et al 2011) and Spreading activation algorithm. It adds the merits and demerits of these approaches.

### 3.2.1 Hybrid ANN and Poisson Mixture Model

Hybrid artificial neural network (HANN) is employed to forecast the best collection of tags to use (Lee et al 2007). This research has used a collective intelligence which is extracted from Web 2.0 collaborative tagging to create tag recommendation to blog based on the blog entries content. Collaborative tagging permits user to widely tag and share with other users. A two way Poisson Mixture Model (PMM) was proposed to model the document allocation into mixture mechanism (Song et al 2008). Two real world tagging datasets are examined for scientific and web pages. The two systems used are CiteULike and Delicious. Words, documents and tags are the tagged training sets and embodied in two bipartite graphs. Using ranking algorithm, topical clusters are ranked.

Collaborative tagging or Folksonomies allow internet users to share, comment and look for online resources with user chosen tags. K-Nearest Neighbor (K-NN) is a suggestion algorithm in Information Retrieval. Group of suggested items is built from the neighbors and it gets a collection of users based on user query. Untagged resources are ignored to modify the K-NN algorithm. Weight of the tags and average relationship of adjacent nodes are computed once similar set of adjacent users are discovered. Boosting feature b is used to support tags in a user profile and b can be added with tag weight if a user has already applied tag to another resource. It improved the coverage and accurateness of K-NN. Modified K-NN with Folksonomies makes good suggestions with huge datasets and it can be applied to both narrow and wide Folksonomy. In contrast, K-NN performs a bulk calculation during query processing and therefore, it is called a lazy algorithm (Gemmell et al 2009).
3.2.2 Latent Dirichlet Allocation

Blei et al (2003) proposed and applied LDA to explain several tasks with topic recognition, entity statement and spam categorization. A scalable and real time approach has been proposed for tag recommendation. Tag LDA method models the document, terms and tags. Latent Dirichlet Allocation method is extended using tag variable. Real time suggestion is made about the probability of tuning a tag to a new document using tag LDA method. Tags probability is used to create tag recommendation. Distributed training structure is executed to manage high level data set effectively from the web. Using a tag variable, a new method tag LDA is derived and it is used to link the tags to the document topics and catch the tag semantic in the structure of topic allocation. Training process is divided into a sequence of steps and the document group is divided into various parts in each step. The entire creation process is document independent since Dirichlet prior is added to the document topic and topic word distributions. The tag LDA method was trained and the tag recommendation result outperformed well (Si and Sun 2009). LDA is important in improving the search and suggesting resource tags (Blei et al 2003).

On clustering the features of data into unnoticed sets, data similarity is elucidated with LDA. Tags belong to a particular topic is recommended for the new resource. Association rules are applied to improve the precision, recall and it suggests more exact tags. In addition, expanding resources with suggested tags considerably develops search for new resources. Latent Dirichlet Allocation (LDA) is also one of the main approaches in suggesting tags for resources (Krestel and Fankhauser 2009). User commented resources are prepared with a quite stable. The entire tag collections are used to derive underlying topics denoted as a combination of description symbols and tags. In order to suggest the probable tags from the
underlying topics, new resources are also drawn to underlying topics based on the content. This method balances simple tag suggestion based on the frequently occurring tags but for new resources, it can recommend tags only to a few posts.

3.2.3 Multinomial Naive Bayes Categorization

Recommendation algorithms and multiclass classifiers are very useful in allocating tags to a bookmark in social bookmarking systems (Illig et al 2011). Multinomial Naive Bayes categorization mechanism is employed for categorizing the tag. New feature from the training set which is unknown by the set has been treated evenly to the wildcard feature during categorization. A novel co occurrence algorithm has been presented for recommending tags on demand driven to improve the quality of recommendation based on primary group of tags used in objects. These tags are mainly used for larger society of users and individual users. Tag recommendation improves the explanation of an object on recommending the tags that explain its content properly and entirely. Tags are connected with object which is given to the tag recommendation system results in relevant tags. There may be numerous options to define an object but this existing strategy dealt with tag exists as a class for an object. This explains that the tag recommendation is a multi label categorization but it does not treat well with problems of thousands of classes (Menezes et al 2010).

3.2.4 Spreading Activation for Recommender System

Most recommender systems produce recommendations based on the environment. This work was addressed how recommendations are made using associative spreading activation. Classical Associative Spreading Activation Network formalism (CASAN) obtains only concept of context. CASAN deals with the following three aspects. First is regarding additions to
CASAN, second is how classical spreading algorithm should change to hold these additions. Finally, third is respect to translation of theoretical formalism into CASAN engine. Spreading activation works as graph in which nodes are denoted as content and links are connected from source to destination. Link is constantly changing from -1 and 1. Links contain relevant environment node and corresponding link type node as well. Link type and environment nodes are class nodes and they may be source or destination of links. All the nodes in a graph or network might participate in spreading algorithm. Relatedness of environment node and its corresponding links are denoted using triangle adjacent. It connects the environment node with triangle and link type node using link. These are called components of formalism and based on the environment any node can act as link type node or content, or context. These three kinds of nodes should take activation values and it should be added with previously given associative networks. It plays a major role in CASAN formalism (Kovács and Ueno 2006).

Four-dimensional power tensor was introduced to share activation among link types. If current environment is changed and three kinds of nodes are divided, matrix computation is required to be less concentrated on the context and link types using CASAN engine. It activates only trusted users extended in the environment layer. Getting an activation model denotes the environment of the current request. This activation pattern is exploited to compute efficient links strength among items. Illegitimate users are turned off because it has zero activation. These inactive links would not participate in the resulting computation and these links are disregarded when spreading activation is in content layer. On expressing the likeness to the system user creates the query. The system then spreads the activation from user to the nodes and highest resulting activation is retrieved to the user as suggestions. Collection of trusted users is suggested in current environment and changes made by the user can result in another different environment. The main
drawback is that, it brings different suggestions in topology of content layer even if user submits the same query. According to the highest activation of nodes and links strength in environment layer, trusted user is retrieved as a result (Kovács and Ueno 2006).

3.2.5 Dynamics of Collaborative Tagging

Generative model was developed to know the dynamics of tagging and power law distribution of tag occurrence. Delicious data is being exploited to evaluate the dynamics of collaborative tagging. Power law distribution defines the frequently occurred tags for sites with its history. Tagging history is evaluated to find out how distribution occurs over time. Tagging is one of the processes of categorization. A similar concept within a concept is divided into groups and this process is called categorization. Tagging differs from classification in which user share and retrieves data more effectively. Tags and their frequencies are arranged using rank frequency to the specific resource is called tag distribution. User, tags and resource are the entities useful in tagging system. These three entities contain a collection of nodes with spaces and these nodes are interlinked with edges. In user space, each node is assumed as a user, tag space is a collection of tags that refers to terms and resource space is represented using URI. It is viewed as two edges link user to a tag and tag to the resource. Tagging is related to date with its tuple of user, tags and resource. This analysis shows several proportions of tagging that are frequently highlighted (Golder and Huberman 2007).

Tagging is mainly useful in information retrieval with major differences. User obtains the resource and adds tags to it in collaborative tagging in which a relationship exists between stability of tag distribution and common nature of the distribution. When the qualitative perception about tagging is correct, distribution must pursue a power law. Most of the sites
attain peak popularity in a given period. The disadvantage of tagging application is that, it must be able to discover power law distribution and relevant entropy measure in which point tagged resource is alleviated to power law. Tags are more general in tagging distribution in which one must know the real meaning of collective categorization system. In generative model, on searching the tags, quantity of pages is retrieved with information value. This model is addressed the information on how it is employed with several tags to envisage association graphs that provide insight into the categorization and existing perception concept relation (Golder and Huberman 2007).

3.3 REVIEWS ON TAG SUGGESTION TECHNIQUES

In order to reduce the creation of irrelevant tags, large number of techniques such as Automatic (Lipczak 2008), Personalized (Byde et al 2007) Music (Zhao et al 2010) and Tagommenders (Sen et al 2009) can perform automatic tag recommendation that automatically suggest the tags which are more relevant to the content of the weblogs.

3.3.1 Automatic Tag Suggestion

Automatic Tag Suggestion is the process of recommending tags to the users when they require annotating a resource. Word trigger method (WTM) is created to recommend tags based on the words in resource description. WTM is efficient and strong when compared to other tag recommendation approaches. WTM is used widely in Web applications and social tags represent the resource content and reveal user interests. There are three processes for recommending tags using word trigger method. The first one is describing annotation pair to learn conversion probabilities with word alignment representation for set of annotated resources. Secondly, IBM model-1 is adopted to learn conversion probabilities between description
words and annotation of tags. The third process is that suggesting tags for resource description. Words in description are restricted to hundreds of words. In the meantime, numerous users can annotate the popular resources with thousands of tags. WTM can only recommend tags that are shown in the translation models (Lipczak 2008).

### 3.3.2 Personalized Tag Suggestion

This approach addresses the tag suggestion through tagging and content oriented similarity metrics. In content based approach tags are suggested which is not already used by anyone. Delicious data set is exploited to investigate the suggestion process in which tagged URL is taken for each user. Common tags are obtained for URL and produce the list to the top. Similarity was computed for each URL that user has tagged already. Similarities are added to provide particular weight for each and every tag and weight is depended on the user. Tag weights present tag ranking from which top ranked tags are chosen as suggestion. This approach only dealt with number of tagged URLs and not the sum of all URLs tagged by each user (Byde et al 2007).

Personalized tag suggestion approach is geared only for realistic interest. Similarity metrics are cosine and variants from information retrieval and text mining. For URL with huge amount of tags, users favored tags thinned using other tags related to various interested areas, different vocabularies and different languages. Tags are suggested from user’s own interest field and therefore, it is more relevant. URL is denoted as a vector and hence, vector frequencies are considered as its metric and vector of word frequencies are noticed in URL content in content based similarity. This approach presents personalized suggestions that render existing tag suggestion based on frequently used tags. Tag based approach is less efficient at
providing good suggestions than content based approach and much poorer coverage (Byde et al 2007).

3.3.3 Music Tag Suggestion

Music tags are suggested based on explicit multiple attributes on semantic relationship of tags and music content. Music attributes may be any number of superficial dimensions. An attribute space is controlled at the starting to a set that reduces semantic loss and tag noise while guaranteeing attribute multiplicity. System suggests related tag list after uploading or browsing the song by user. System is largely deployed to process set of large-scale music data. Recent image domain tag suggestion approach is constructed by assigning explicit multiple attributes to music domain which imposes attribute multiplicity for music detection with greater semantic clarity. Music domain explicit multiple attributes are developed first, scalable content analysis and similar tags analysis algorithms are implemented to tackle the millions of songs. Finally, rapid tag suggestion engine was developed to give efficient online service. The system which is developed contains content based explicit multiple attributes (CEMA) and Social tags based explicit multiple attributes (SEMA). These modules maintain list of indexed Multiple attribute fuzzy music semantic vectors (MA-FMSVs) and multiple attribute tag distance vectors (MA-TDVs). Songs database is explored throughout offline processing (Zhao et al 2010).

Parallel multiple attributes concept detector (PMCD) and parallel occurrence co-occurrence (POCO) algorithms are used to produce MA-FMSVs and MA-TDVs for every song. The system suggests different attribute tags for a user offered song throughout online processing. MA-FMSVs songs are envisaged with the help of Concept detector and it is used to index into CEMA and then adjacent neighbors is discovered. These neighbors are indexed into SEMA, as a result list of tags are ranked and
sorted for each attribute. Finally, system ensures the suggested tags that can be different attributes. Parallel music content is analyzed and Map Reduce based social tags mining algorithm is used to concept detection and parallel social tags mining algorithms based on the Map Reduce framework to maintain large scale offline and online tag suggestion for each predefined attribute. Suggested tags are more diverse attributes, scalable to huge data sets and ranking is more efficient. This approach is applicable to the task of other multimedia content analysis as well as tag suggestion. But, the performance of framework is not evaluated using unequal and bigger sized CEMA and SEMA attribute spaces (Zhao et al 2010).

3.3.4 Tagommenders

Tagommenders is introduced to envisage user favorites for items using contingent preferences for tags (Park et al 2011). Tagommenders propose the automation of usual suggested systems but maintain the reliability of tagging system. Tagommenders allow recommenders to employ the proportions of items that users reflect on most significant. Implicit data and explicit data tag recommendations are also presented. Explicit data is exploited in tag based suggestion algorithms to help sites devoid of item ratings. Tag preferences are taken as input and gives appropriate score for item ranking in a list of recommendation as an output. These systems normally would not collect and envisage ratings. Hence, the output of implicit tag suggestion algorithms is more applicable to the task of suggestion and not to the task of envisage.

**Implicit tag:** Implicit tag computes user profile vector and a document term vector similarity in which columns communicate with tags and dot product is computed among user preferences for tags to produce prediction. Envisaged score is calculated using implicit tag preference.
**Implicit tag pop:** Implicit tag pop deals with the calculation of popularity based on the number of tags, clicks, users and taggers. Popularity signal is examined to see the performance of implicit tag pop using signal. Tags outperform users, clicks and log transforming signal.

**Explicit tag based algorithms**

The following explicit tag based algorithms are proposed to rate item for sites. Tag preferences expressed by all taken as input and results between values 1.0 and 5.0 using implicit and explicit signals. These algorithms improve the tasks of both envisage and suggestion.

**Cosine tag:** cosine tag envisages user ratings that are average weight of user preference for particular tags. It weights a specific tag based on attuned cosine similarity among ratings of tag and preferences for a tag. The attuned cosine similarity and its average for diverse tags are computed.

**Linear tag:** Linear tag represents more composite relations among preference tags and envisaged ratings. Weighting is calculated since the performance is improved and it provides greater significance to more exact fits.

**Regress tag:** Tag relations performed better than other methods. It generates linear equation for each tag and input values are user tag preferences consequences in user ratings.

Tagommenders and tag preference suggestion algorithms are examined that utilize the tags and different signals performed well. There is no surety of more precise suggestion while large amount of tags exist for delicious system. Primary problems occur in relationship between preference and quality (Sen et al 2009).
3.4 REVIEWS ON CLUSTERING APPROACHES IN TAG RECOMMENDATION

In this section, clustering approaches such as Trust based clustering (DuBois et al 2009), Web Recommender systems (Mabroukeh and Ezeife 2011) and social interest based clustering (Li et al 2008) will be discussed and the clustering support necessary to find a similarity between the tags in each of these approaches are also explained.

3.4.1 Trust Clusters for Recommendation

Trust cluster is also one of the recommendation approaches incorporated into memory based collaborative filtering algorithms. Trust network is being exploited to produce conditional trust values among network pairs. Trust values are then used as an origin in a network. Trust metrics are capable of clustering people with high trust within metric space. Similarity is computed using trust value from one node to another. Output of trust cluster algorithm can differ from one implementation to another since it is randomized. One node is randomly selected to be center and new cluster can be created out of all nodes within preset radius. Various iterations can be performed to generate a collection of clusters in inferred trust network. Trust inference algorithm was successfully performed on Film trust data to get clusters for recommendation systems and also some extensions are needed for improving its accuracy. Retrieving poor suggestions constantly over a time may lessen their trust and make them to look for delivering high-quality service. If it is a continuous means user interest can be vanished. Trust inference algorithm is only static and it does not consider the characteristics of community as it changes because it is time based problem (DuBois et al 2009).
3.4.2 Web Recommender System

Web recommendation is based on user surfing activities which is related to browsing history of existing users. Recommendation can be of items such as products, articles, titles, etc. It exploits the knowledge of blogs to envisage user preferences and suggests certain items to the user. Content based Web Recommender Systems would not pursue domain ontology to take out information about suggested items or user history which falls in some restrictions like the problem of new user without previous browsing history and sparsity problem. On clustering the users, recommendation of items can be provided to the users. Clustering is expensive to use for all recommendation process. Similarity between item and ontology concept is computed on click stream mapping in which web log is indicated as click stream includes items and corresponding tags. When a new user enters into the web page, corresponding item is retrieved and vectors are produced. Vector is matched with stored concepts in matrix and top most matched item is suggested to the users. This suggestion set is extended on increasing n times and also active user vector is extended with semantic relations in ontology in order to add several concepts and to draw the suggestion. User tags are mapped to the concepts of ontology. Spreading activation is applied over ontology to demonstrate on how suggestion set is extended. Semantic relations in ontology are taken to present relatedness measures for suggesting closely related concepts only (Mabroukeh and Ezeife 2011).

3.4.3 Social Interest Based Clustering

Internet Social interest discovery mechanism (ISID) has been proposed based on user created tags to find out the user interests generally, cluster users, saved URLs and different topics of interest (Li et al 2008). User entered tags are reliable with web content and user tags that occurred frequently are utilized to classify and confine the users’ topic interests. User
tags are employed in order to find out the user interests shared via social networks. Using the user tags, an ISID groups the users and URLs since tags explain the users’ interests completely (Li et al 2008). User profile recommender is a basis for possible tag recommendations. User profile consists of tags with dynamic user interests and their behaviors. Frequently used tags in the past are not essential for the present tag recommendation. To overcome this issue, the user recommender uses regency based scoring mechanism to balance the frequency based mechanism. These two mechanisms generate the same tag sets with different scores. These two are combined and the linear amalgamation of both mechanisms is called final score. Result of this user profile recommender is a collection of user relevant tags. User and resource relevant tags are combined to generate the final tag recommendation (Lipczak and Milios 2010 a).

3.5 REVIEWS ON TOPIC ONTOLOGY CONSTRUCTION TECHNIQUES

To support efficient tag recommendation automatically, it is necessary to construct semantic web (topic ontology) in a machine-processable way. To enable this, web resources such as wikipedia (multilingual online encyclopedia) and WordNet (Controlled vocabulary of lexical database terminology) has been enriched. Various topic ontology construction approaches such as semi automatic ontology construction, automatic ontology enrichment based on ODP will be analyzed and also provide the merits and demerits of each approach.

3.5.1 Enrich Wordnet Concepts to Construct Topic Ontology

In this approach, documents related to specific domain is collected from the World Wide Web and required information about that domain is extracted from those collected documents (Agirre et al 2001). Consecutively,
synset for that domain is extracted from the English language lexical dictionary called WordNet which is used to derive the semantic relation between the extracted keywords. This type of ontology enrichment is an automatic process that enriches the concept in existing ontology using extracted semantic relation between the topics in the documents. The core concept of this approach is that, weights are assigned to the topics according to the number of similar words available for the specified concept available in wordnet. Due to the enrichment of conceptual relation between the set of topics, efficient topic ontology has been constructed.

### 3.5.2 Wikipedia Categories Extraction

In order to construct the topic ontology, concepts related to the specific topic are needed. It is achieved through the extraction of information from wikipedia articles where group of related concepts are organized in the form of hierarchical categories. In this approach, Persian Wikipedia is used to gather the set of pages related to the specific domain and Wikipedia parser is deployed to extract the title, keywords and existing links (Farhoodi et al 2009). Metadata repository maintains the extracted information which is used to construct the topic ontology. Among the extracted metadata, titles in each page are considered as the topic of the ontology and associated links between the pages are used to derive the semantic relation between the topics in ontology. Concept hierarchy between set of topics is formed according to the weights of the links. Depending on the information in the page, weights are assigned to the associated links.

### 3.5.3 Wikipedia Semantic Relation Extraction

A new approach called WikiSemNet is introduced to enrich the extraction of semantic relation between the wikipedia articles (Stoutenberg et al 2009). At first step, pre processing is performed on the collected web
documents to extract the relevant categories related to the required domain. Tokenization is performed that can identify the semantic relation between the extracted keywords using the links between the wikipedia articles. Regular expression is applied to the set of concepts according to the identified semantic relation. Most relevant category plays an important role to form the strong semantic relation between the set of topics. The strong semantic relations are computed through two measures called number of links between the categories and connectivity ratio (Chernov et al 2006). The extracted relationships are maintained in the database which is used to emphasize the meaningful semantic relation between the relevant categories and irrelevant categories are not considered. From the extracted semantic relation, knowledge is acquired about the content of Wikipedia page that results in the construction of topic ontology. Finally, OWL on the web is used to visualize the constructed topic ontology.

3.5.4 **Semi Automatic Topic Ontology Construction**

In semi automatic approach, data mining algorithms such as i) Centroid Vectors ii) Support Vector Machine is used. Among the two, Centroid vector technique extracts keywords associated inside the topic from the large collection of documents. Support Vector Machine (SVM) binary classifier can classify the documents based on the computed cosine similarity measure between the extracted keywords (Fortuna et al 2006). After detecting the semantic similarity between the topics, k-means semantic clustering approach can group the documents with high semantic similarity. The major drawbacks in this work are i) data mining algorithms used in this approach is not that much sufficient ii) It only helps to discover the topics in the large corpus of documents but does not provide any support to identify the semantic relation between set of concepts in the document iii) It is semi-automatic approach, that requires user interaction to construct topic ontology whereas
system gives only suggestion through providing documents related to the specified topic.

3.5.5 Automatic Ontology Enrichment Based on ODP

OntoCase is a pattern based semi automatic ontology construction approach deploys ontology learning techniques to extract semantic concepts from the large collection of documents. The core concepts of this approach are i) Background knowledge of the documents is represented to provide the top structure to the ontology ii) Taxonomical depth and relation to concept ratio is improved to provide the overall enrich ontology structure (Blomqvist 2009). In order to improve the quality of existing ontology, Ontology Design Patterns (ODP) provides the precise design principles to enrich the overall structure of ontology using extracted semantic background knowledge. It is a language independent approach and also implemented through OWL ontologies. The retrieve and reuse phases in enrichment process is improved through various approaches such as pattern ranking, concept coverage, relation coverage and utility measure.

3.6 REVIEWS ON ONTOLOGY BASED RECOMMENDATION APPROACH

Various ontology based recommendation approaches such as domain ontology based recommendation (Pudota et al 2010), ontological relation based recommendation (Alexopoulos et al 2011) and ConTag based recommendation (Adrian et al 2007) are available in favor of recommending more relevant tags to the blog users. Review is carried out on various approaches that support efficient tag recommendation.
3.6.1 Automatic Domain Ontology Based Recommendation

This approach is a controlled and ontology based automatic tag recommendation that deploys a new component called Key phase Extraction Module (KpEM) that can automatically extract keywords from the textual document of web resources. Ontology Reasoner performs the keyword matching to find the exact match for the extracted keywords. Super class of the exact match is retrieved using parent-child relationship to determine concept mapping node (Pudota et al 2010). Spreading Activation Algorithm is used to carry out the special navigation strategy that can find the node which originates the corresponding match. It provides high-quality and disambiguated tags which summarize the semantic content of a Web resource. It is an unsupervised approach. Therefore, no prior knowledge about the document is needed. It can determine candidate key phrases using linguistic knowledge of the tags with arithmetical characteristics such as phrase depth, term frequency etc., Also it completely diminishes the manual effort in order to tag a Web resource. The relevant tags for the various number of web documents are computed using the different levels of documents’ subjectivity. The disadvantage of this approach is that key phase extraction process is not optimal and also domain ontology is deployed to form concept mapping. But, it is difficult to design and obtain the domain ontology in unsupervised approach. It must account the human subjectivity in key phrase extraction and also improves the adaptive personalization technique to extract the keywords according to the interest of the users (Baruzzo et al 2009).

3.6.2 Exploiting Ontological Relation for Tag Recommendation

This process is a novel approach that takes effort to reflect the semantic intended meaning of the web page. It can automatically generate and recommend the appropriate tags for the web documents based on utilizing the semantic relation between the concepts in domain specific ontology. In order
to select effective tags, this approach includes two components called Tagging context and Tagging Recommendation process (Alexopoulos et al 2011). According to the domain and specific application requirements, candidate tags will be selected from the textual documents based on the relative importance of ontological relation between semantic concepts in domain ontology. Keywords are extracted using the ontological relations based on the matched relation to the tag concepts. Extracted sets are grouped and termed as candidate tags. After identifying the candidate tags, tag recommendation process determine the tags which have high potential with high confidence score. In Tag Recommendation Process, the confidence score of the tags is computed from the degree at which tags can characterize the terms in the document using ontological support. For each of the selected tags, candidate score is computed based on the ontological ambiguity. Finally, tags which have high confidence score are labeled as relevant tags and recommend the tags to the user. This approach is an automated and optimal approach to recommend semantic tags for the web documents in a more accurate and complete way. The main disadvantage is that quantity and quality of the content of the documents mainly rely on the ontological relation in the domain ontology. Therefore, this approach is suitable only for less complex, unambiguous domain with small datasets. Furthermore, it does not have capability to adapt and manage the knowledge incompleteness and inconsistency.

3.6.3 Contag Recommendation

ConTag is also one of the recommender systems to tag or annotate the semantic web ontology concepts. In order to extract the document content, a web 2.0 services in ConTag presents text and phrase analysis functions such as term extraction, thesauri, dictionaries, categorizations and term relations. Web 2.0 information is merged with semantic web ontology facilitates
suggested semantic tags for documents. ConTag approach creates semantic tag suggestions for documents according to Web 2.0 services and semantic web ontology. Formalized concepts in Personal Information Model Ontology (PIMO) vocabulary is utilized to create tag recommendations. Articles occurred in a document is called topics. In categorization process, if class of a topic communicated directly to ontology class, topic is suggested as new object of ontology class. Alignment design is used as a tag recommendation list and each entity drawn in an alignment is considered as a tag for the document. ConTag approach with combination of semantic web ontology and web 2.0 services facilitate the relevant semantic tags suggestion for documents. In this recommendation, ConTag focused on retrieval of equivalent things in documents and ontology but not the similarity of things (Adrian et al 2007).

3.6.4 User Interest Ontology on Blog Community

In blog community, the entries posted by the user are constructed into service domain ontologies. User interest is extracted from the information of the blog and might be constructed into user interest ontology. In user interest ontology, interests of the users are expressed using specific weights semantically in the form of hierarchy of classes in top down approach. Consequently, interest of the user is updated periodically using top-down approach. In this approach, similarities between the users’ interests are detected automatically using the constructed ontology with their weight of each instance and classes. This approach is very useful to recommend the new innovative tags to the blog users. Inspite of providing similar tags to the users, this approach is very helpful to provide the most innovative tags which are different from the tags created by other user and also most relevant to the content of the blog entries (Nakatsuji et al 2006).
The above mentioned approaches explain the significant aspects of tag recommendation problem, but still, there is lack of precise relations among tags. There is no meaningful structure to the relevant tags or keywords for easy recommendation. To provide the fine grained structure, the topic ontology has been constructed from Wikipedia and WordNet information. Topic ontology renders the topics in hierarchy order with semantic relationship and spreading activation algorithm is used to calculate the weight of the tags and based on it, tags are recommended to the resources.