International Journal of Web Engineering 2012, 1(2): 11-14
DOI: 10.5923/j.web.20120102.01

# Building User Profiles Based on Preference Networks

**J. Jayanthi[1,*], Keerthana. B[2], K. S. Jayakumar[3]**

[1]Department of Computer Science Engineering, Sona College of Technology, Salem, 636005, India
[2]Department of Software Engineering, Sona College of Technology, Salem, 636005, India
[3]Department of Mechanical Engineering, SSN college of Engineering, Chennai, 603110, India
jayamithu2002@yahoo.co.in, keerthana.softy@gmail.com

**Abstract** As the number of internet users grows rapidly, it is important to provide more relevant results for the individual user. It can be achieved by the agent that tracks and monitors user's behavior. Users search behavior is stored in the form of profile.User's interest is analysed based on various factors like query,usage count,concept age etc.The proposed work in the profile construction includes following tasks,(i)Extraction of intents(set of keyword extracted from top k documents)(ii)Construction of preference network(iii)Creation and updating of the user profile.These profiles could be used to retrieve the prefered search results.Profile is organised in an hierarachial way so the particular concept along with the keywords can be viewed easily.User interest is tracked in a better way and it would improve the search results with more precision.

**Keywords** User Profile, Agent Monitor, Intent, Extent, Concept Age, Preference Network

## 1. Introduction

Search engine usage has been increased in recent years. Searching technique can be used in many ways as to provide the relevant results for the given query. The relevant results obtained can be either in World Wide Web or in the particular information retrieval database. The profile can be constructed where the unique data for each individual is stored. Information gathered can be general interest, demographic information like (name, age, country etc).These can be done using the usage logs where it contains both preferred and un preferred results.

In the proposed work profile is built for the authenticated users. When the query is submitted the search results are displayed ,from that documents the identical weighted keywords ,treated as intents are extracted using Term Frequency and Inverse Document Frequency(TF-IDF) schema.After extraction, a preference network is built for each user with respect to query along with intents.Agent monitor is arranged to track the user session and maintains the information in the user profile like deletion and updation of profiles.

## 2. Related Works

In paper[2] the user profile is constructed based on many data sources and framework uses three types of monitors. Various types of ontology and their relationship is discussed. In[1] explained using Spreading activation Algorithm the interest scores is assigned for each concepts.

Many hypothesis are framed in information processing regarding content relevance and self reference[5].In[7] the user profiling is done based on the personal data and search history.Different agents are used for improving relevance,response time,reducing the time and system extensibility. The users information can also be known with the help of preferences that is what type of document is viewed, based on the year of the document,type of document etc. The functionality of each agent discussed. Based on the click history the user model is developed[3] where the representation of user preference is given based on the topic and page.

For the given query the intent and extent is extracted based on FCA theory[4] and the keyword is extracted using tf- idf schema. The user interest session is supervised by the concept network. For personalization some client side algorithms are developed[6].With the help of domain list the relevant pages are identified and the user model is developed[8].

In[10] the teaching agent is designed for the learners to utilize the knowledge. The learning agents mainly include the teaching styles, psychological characterstics etc. The pedagogical agent is used for personalized learning units dynamically based on the information provided by user models and domain knowledge bases in order to improve the self-adaptability and pedagogical effects of the system. Internet shopping uses multi agent architecture to improve the commodity information and automate the process. The

* Corresponding author:
jayamithu2002@yahoo.co.in (J. Jayanthi)
Published online at http://journal.sapub.org/web

system consists of five types of agents are interface agent, buyer manager, buyer agent, evaluation agent and preference agent**,** which interacts with each other[9]. In[13] for using the e-learning model they use the agent which deleivers (SERP) gives the personalized suite materials based on users query. Various types of agent is used in the search engine model. Each agent performs the separate functions and have their own characterstics.

In[11] control of classroom hardware can be controlled using remote devices. A multi-tiered agent-based software architecture is proposed and a distributed deployment is presented in order to satisfy all the requirements. Resource management with higher level is used along with the components. In[12] Social Agents monitor the activities of Internet users to build and update profiles to create socio-culture community of the similar interests in the cyberspace. They have used the honeypot concept .They have generated the profiles and checks for profile compatibility.In[14] they are collecting the web queries from other repositories to increase the effectiveness in the information retrieval at the higher granularity levels.

# 3. Proposed Work

We have used the agent in the proposed work that involves in the profile building task . The architecture of the proposed system is illustrated in Figure 1

When the user submits the query through search engine it displays the top k search results.The intent is extracted from the extent top k search results.With this extracted results an preference network is constructed in every individual user profile for a submitted query.The long time unused concept is deleted from the user profile.Agent monitor is arranged to track the session for each individual user and update the information in the profile.

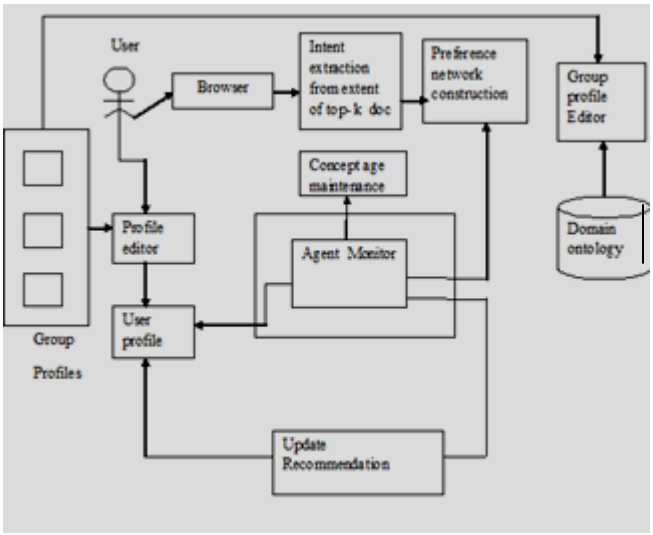



**Figure 1.** System architecture

### 3.1. Method

The proposed system proceeds through the below processes namely,

TF-IDF Measure Extraction

Construction of PN

Creation and updation of profile

The proposed "Agent Based system for Preference Network construction"where the authenticated users gives the query and it is monitored by the agent .For the given query the search results are displayed according to the profile maintained for the individual users.The preference network is constructed for each individual user profile for the given concept. The proposed framework is realized through three different processes and the data flow could be interpreted using Figure. 2.

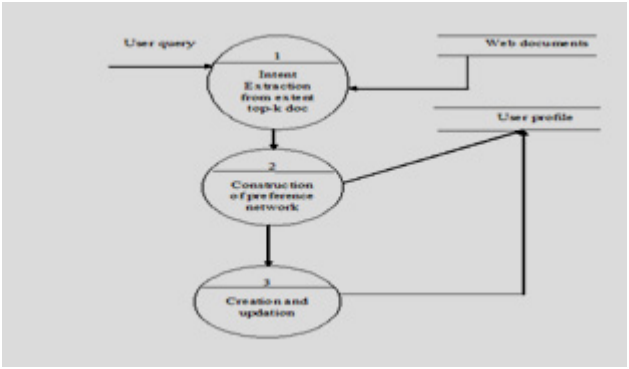



**Figure 2.** Process Flow Outline

#### 3.1.1. TF-IDF Measure Extraction

The top K documents from the web server are analysed for each term TF-IDF measure is computed and the same could be retained in the TF-IDF store. Terms are sorted based on the TF-IDF value measured and from this the top N terms with higher weights are used for further processing. From the above term set, the identical terms in all documents are collected and their weights are added up and from the outcome the higher weighted terms are again selected for building the personalized preference network. The above discussed process can be shown in Fig. 4.

Term frequency and Inverse document frequency can be obtained as below.

$$tf_i = \frac{n_i}{\sum_k n_k} \quad (1)$$

$n_i$ = No of occurrence of a term i

$n_k$ = Total no of terms in a document

$$idf_i = log \frac{N}{df_i} \quad (2)$$

N= Total number of documents that are relevant

$df_i$ =Number of documents that contain the term i at least once.

$$\text{TF-IDF weight} = TF_i * IDF_i \quad (3)$$

Thus the term frequency and inverse document frequency are computed.

#### 3.1.2. Construction of Preference Network

.The preference network is constructed based on TF-IDF weight of the intents for the particular concept.After finding the weight they are sorted from the higher value to the lower

value.The higher value intent is added to the network for the given user

3.1.3. Creation and Updation of the Profile

The profile should be created for the individual user by tracking the activity during their session.When the existing user gives the different query the intents is extracted and the weightage is calculated.Before updation the similarity measure is done between the existing query and new query.The query given by the user can also be the part of information from the existing preference network maintained

# 4. Algorithm

The algorithm for the preference network construction is explained as follows,

TF-IDF Calculation for Intents
Input query Q
Extract E Where E∈D
For(i=1 to N)
Compute Intent Weight (IW)
Add Weight of all intents
Sort(I,IW )
Append (I,IW) to Q in $P_i$ /*Profile of the user
Updation of the user profile

Input New query
Then IW>0
Add Subnode($SI_n$,$Iw_n$) to $C_1$
Else
In(α) < 0.45
Construct PN ⊅ $C_1$/*Independent from $C_1$
End
If $SI_n$ input query
Calculate $IW=\frac{n_i}{\sum_k n_k} * \log \frac{N}{df_i}$
Sort($SI_{n,m}$,IW)
Add ($SI_{n,m}$,IW) terms to $SI_n$ in $P_i(U_i)$

# 4. Results

For user given query "Artificial intelligence techniques" we are going to construct the preference for top k results.Let us consider that ,

i) The user selected documents can be{d1,d2,d4,d5,d6} from this set the tf-idf weight is calculated for all intents then the higher valued intents are considered to construct the preference network.

ii) For the "Artificial Intelligence " query the weight for the intents can be given as,

**Table 1.** Intent Extraction Measure for query 1

| Intents | weights |
|---|---|
| Neural networks | 0.88 |
| Natural language processing | 0.86 |
| Genetic algorithms | 0.80 |

**Case 1:** From the above values the preference network is constructed in the user profile Pi

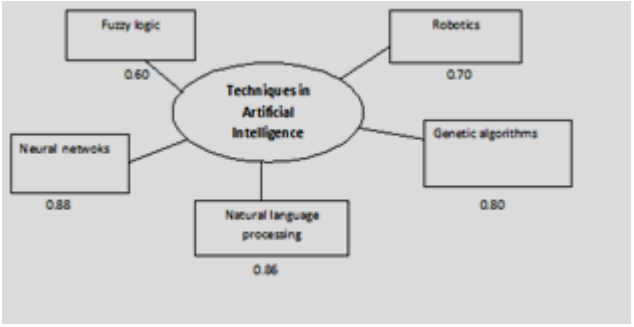



**Figure 3.** Preference network for a query

When the TF-IDF weight calculated is lower than the threshold value (α)<0.45 then they are independent from the existing preference network in the same user profile

**Table 2.** Intent Extraction Measure for query 1, α<0.45

| Intents | Weights |
|---|---|
| Back propagation | 0.37 |
| Time series prediction | 0.34 |
| Blind signal prediction | 0.30 |

**Case 2**:When the same user gives the different query "Genetic algorithm" ,TF-IDF method calculates the weight for all intents in the selected documents

**Table 3.** Intent Extraction Measure for query 2

| Intents | Weights |
|---|---|
| Fitness function | 0.85 |
| Mutation | 0.83 |
| Mutation cross over | 0.81 |

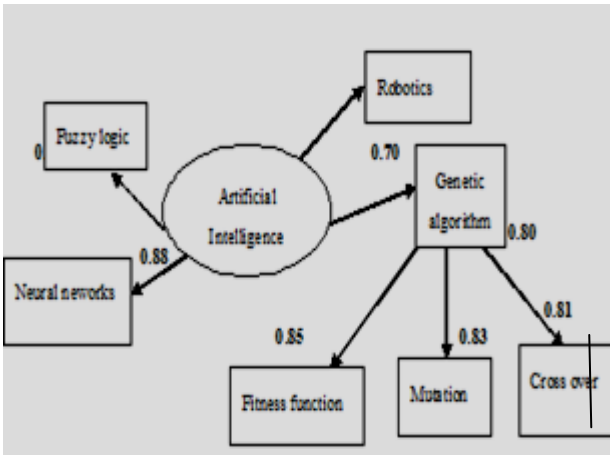



**Figure 4.** Preference network for query

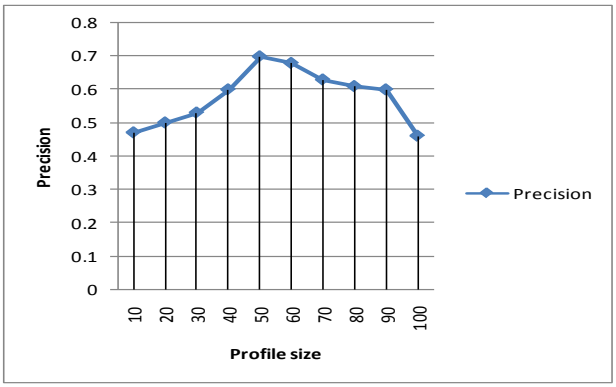
# 5. Experimental Results

Experiment was done with the profiles that ranges from {10,20……..100} terms.Precision may fall ,if the number of terms in the profile is large.In order to minimize the

precision fall, profile convergence factor needs to be analysed and evaluated properly.This would be done in the next level of implementation

**Table 4.** Precision calculation

| Profile size | Precision |
|---|---|
| 10 | 0.47 |
| 20 | 0.50 |
| 30 | 0.53 |
| 40 | 0.60 |
| 50 | 0.65 |
| 60 | 0.70 |
| 70 | 0.68 |
| 80 | 0.62 |
| 90 | 0.61 |
| 100 | 0.46 |




## 6. Conclusions

In our work, we have introduced the agent which tracks the activities of each individual user. A preference network is constructed using Personalized preference network construction(PPNC) algorithm. This approach would help us to retrieve personally preferred results from the search engines. Profile convergence factors need to be analysed further in order to improve retrieval efficiency. In future, this would be extended for providing service to the e-commerce applications.

## REFERENCES

[1] Ahu Sieg, Bamshad Mobasher, Robin Burke:"Ontological User Profiles for Representing Context in Web Search",ACM International Conferences on Web Intelligence and Intelligent Agent Technology,(2004)

[2] Plaban Kumar Bhowmick,Sudheshna sarkar and anupam basu:"Ontology based user modelling for personalized information access",Int'l Journal of Computer Science and Applications,pp.1-22(2010)

[3] Qui, F., Cho, J.:"Automatic Identification of User Interest for Personalized Web search",15th international conference on World Wide Web,2005.

[4] Han-joon Kim,Byungjeong Lee,Sungjick Lee*,Sooyong Kang.:"Building concept Network-based User profile for Personalized Web search".International Conference on computer and Information Science,2010

[5] Kar Yan Tam, Shuk Ying Ho:"Understanding the impact of web personalization on user information processing and decision outcomes,MIS Quarterly,2006

[6] Teevan, J. , Dumais, S.,T.,Horvitz, E.:"Personalizing search via analysis of Interest and activities" International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.449-456(2005)

[7] Kechid Samir , Drias Habiba.:"Mutli-agent system for personalizing Information source selection".ACM International Conference on Web Intelligence and intelligent Agent Technology,2009.

[8] Kavita ,Satokar ,D.,Gawali, S.,Z.:"Web search Result Personalization using Web Mining".International Journal of Computer Applications,2010

[9] Ziming Zeng , Bo Meng.: "A Multi-Agent Based Intelligent System for Internet shopping", International Conference on Services Systems and Services Management,2005.

[10] SUN Yu,LI Zhiping.:"A Multi-Agent Intelligent tutoring system".4th International Conference on Computer Science & Education,2009.

[11] Anchez, E.,S., Lama, M., Amorim, R.,Riera, A., Vila, J., Barro, S.: "A multi- tiered agent-based architecture for a cooperative learning environment". In.11th Euromicro Conference on Parallel,Distributed and Network-Based Processing,2003.

[12] Ghulam Ali, Noor,A., Shaikh, Dr. Zubair Shaikh, A.:"Agent-based User-Profiling Model For Behavior Monitoring".International Conference on Future Networks, 2009.

[13] Axita Shah, Sonal Jain,: "An Agent based Personalized Intelligent e-learning". International Journal of Computer Applications,2010.

[14] Chirita, P-A., Firan, C.,S., Nejdl,W.:"Personalized Query Expansion for the Web".30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval , pp.7-14(2007)