A USER CONCEPT MODEL FOR PERSONALIZED INFORMATION GATHERING AND RETRIEVAL

Nageswara Rao.Yarlagadda (M.Tech)
ComputerScience&Engineering(Dept)
BharathUniversity,Chennai,India
nageshchow@gmail.com

Mr.R.Chndrasekaran
Assoc.Prof &Project Coordinator
ComputerScience&Engineering
BharathUniversity,Chennai,India
chanindira@yahoo.com

Abstract: As a model f or knowledge description formalization, ontolog y are widely used to represent user profiles in personalized web informationgathering.Ho wever,when representing user profiles, many models have utilized only knowledge from either a global knowledge base or user local information. In this paper, a personalized ontolog y model is proposed for knowledge representation and reasoning over user profiles. This model learns ontological user profiles from both a world knowledge base and user local instance repositories. The ontology model is evaluated by comparin g it against benchmark models in web information gathering. The results show that this ontology model is successful.

Keywords-; Ontology, personalization, semantic relations, world knowledge, local instance repository, user profiles, web information gath ering.

1)INTRODUCTION: ON the last decades, the amount of web-based information available has increased dramatically. How to gather usef ul information from the web has become a challenging issue o1 for users. For this purpose, user profiles are created for user background knowledge description. . To represent user profiles, man y researchers have to user knowledge through global or local analysis. Global analysis uses existing global knowledge bases for user background knowled ge. Used bases include Generic ontolo gies(e.g.,Word Net),thesauruses and (e.g., digital), and online bases e.g., online categorizations and Wikipedia). Local analysis investigates user local information or observes user behaviour in user profiles. However, because local analysis relies on mining or classification techniques for knowledge discovery, occasionally the discovered contain noisy and uncertain information. As a result, local analysis suffers from ineffectiveness at capturing formal user knowledge.

2)World Knowledge Representation:

World knowledge is necessary for lexical and referential disambiguation, including establishing co reference relations and resolving ellipsis as well as for establishing and maintaining connectivity of the discourse and adherence of the text to the text producer’s goal and plans.

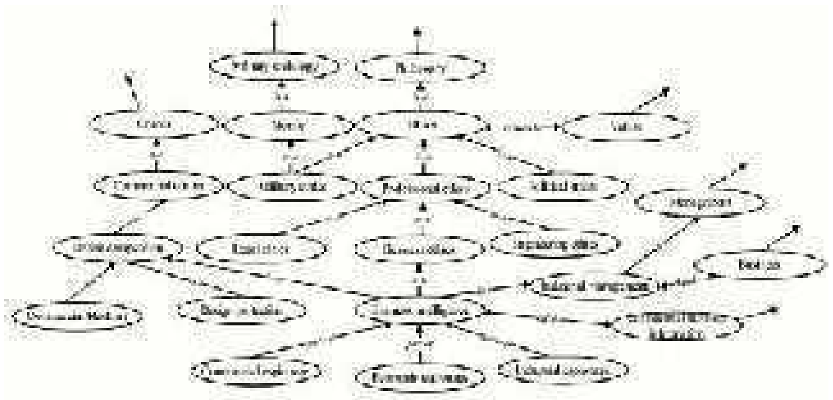


Fig1: A sample part of the world knowled ge base

3)OUR IDEA:

Ontological user profiles. The world knowledge and a user's local instance repository (LIR) are used in the proposed model. World knowledge is commonsense knowledge acquired by people from experience and education an LIR is a user's personal collection of information items. From a world knowledge base, we construct personalized ontologies by adopting user feedback on interesting knowl-edge.A multidimension al ontology minin g method, Speci-ficity and Ex haustivity, is also introduced in the proposed model for analyzin g concepts specified in ontologies. The users' LIRs are then used to discover background knowl-edge and to populate the personalized ontologies.

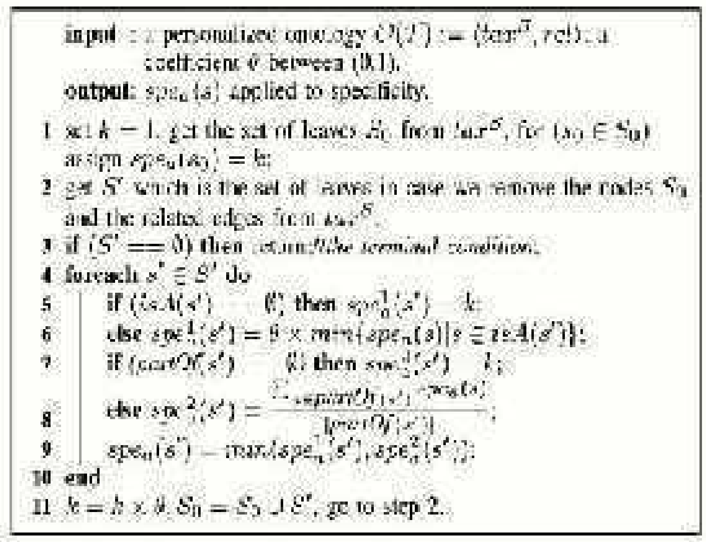
4)PERSONALIZED ONTOLOGY CONSTRUCTION:

Personalized ontologies are a con ceptualization model that formally describes and specifies user background knowl- edge. From observations in daily life, we found that web users might have different ex pectations for the same search quer y. For ex ample, for the topic "New York," busin ess travelers may demand different information from leisure travelers sometimes even the same user may have different expectations for the same search quer y if applied in a different situation. A user may become a busin ess traveler when planning for a business trip, or a leisure traveler when planning for a family holiday.

5)MULTIDIMENSIONALONTOLOGY MINING:

Ontolog y mining disco vers interesting and on -topic knowl-edge from the concepts, semantic relations, and instances in an ontolog y. In this section, a 2D ontology mining method is introduced: Specificityand Ex haustively. Specificicity describes a subject's focus on a given topic. Exhaustivity (denoted exh) restricts a subject's semantic space dealing with the topic.This method aims to investigate the subjects and the strength of their associations in an ontology.

Algorithm 1.Analyzing semantic relations for specificity


input : a personalized ontology [illegible]; a coefficient θ between (0,1).
output: $spe_a(s)$ applied to specificity.
1 set k = 1, get the set of leaves S_0 from [illegible], for ($s_0 \in S_0$) assign $spe_a(s_0) = k$;
2 get S' which is the set of leaves in case we remove the nodes S_0 and the related edges from [illegible];
3 if (S' == ∅) then return//the terminal condition;
4 foreach s' ∈ S' do
5 if [illegible] then [illegible];
6 else [illegible];
7 if [illegible] then [illegible];
8 else [illegible];
9 [illegible];
10 end
11 [illegible], go to step 2.

6)UserLocal Instance Repository:

i) User back ground knowledge can be discover ed from user local information collections, such as a user's stored documents, browsed web pages, and composed/received emails.

ii) The ontology OðT Þ constructed in Section 3 has only subject labels and semantic relations specified.In this section, we populate the ontology with the instances Generated from user local information collections.

iii)Generating user local LIRs is a challenging issue. The documents in LIRs may be semi structured.(e.g.,the browsed HTML and XMLwebdocuments) or unstructured (e.g., th e stored local DOC and TXT documents). In some semi structured web documents, content-related descriptors are specified in the metadata sections.

Fig2. An information item in QUT library catalogs

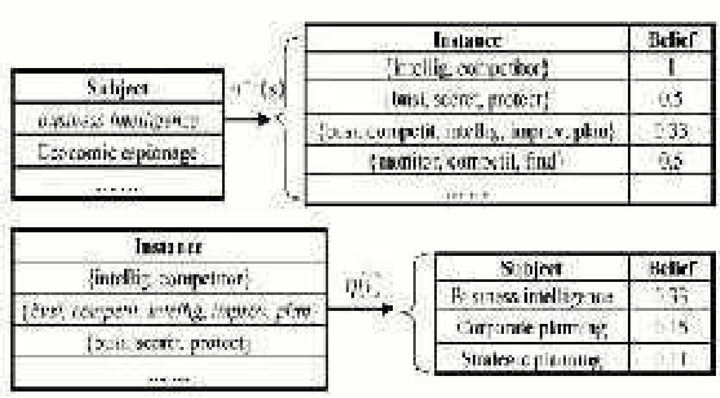


Fig. 3. Mappings of subjects and instances

7)Multidimensional Analysis of Subjects:

i)The exhaustively of a subject refers to the extent of its concept space dealing with a given topic. This space ex tends if a subject has more positive descendants regard-in g the topic. In contrast,ifasubjecthasmore negatived escendants,its exhaustively decreases.

ii) Based on this,let be a function that returns the descendants of s we evaluate a subject's exhaustively by aggregating the semantic specificity of its descendants :

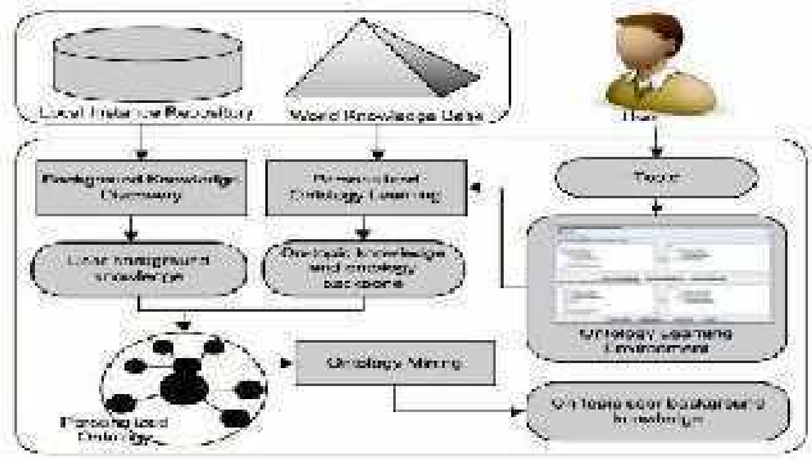


Fig. 4. Architecture of the ontology model

8)VALUATION:

8.1 Experiment Design:

i)The proposed ontology model was evaluated by objective experiments. Because it is difficult to compare two sets of knowledge in different representations, the principal design of the evaluation was to compare the ef fectiveness of an information gathering system (IGS) that used different sets of user background knowledge for information gathering.

ii) The latter run set up a benchmark for the evaluation because the knowledge was manually specified by user.

iii) The Ontology model that implemented the proposed ontology model. User background knowledge was computationally discovered in this model.

iV) The TREC model that represented the perfect interviewing user profiles. User background knowl- edge was manually specified by users in this model.

V)The Category model that represented the noninter viewing user profiles.

vi)The Web model that represented the semi-interviewing user pro files.

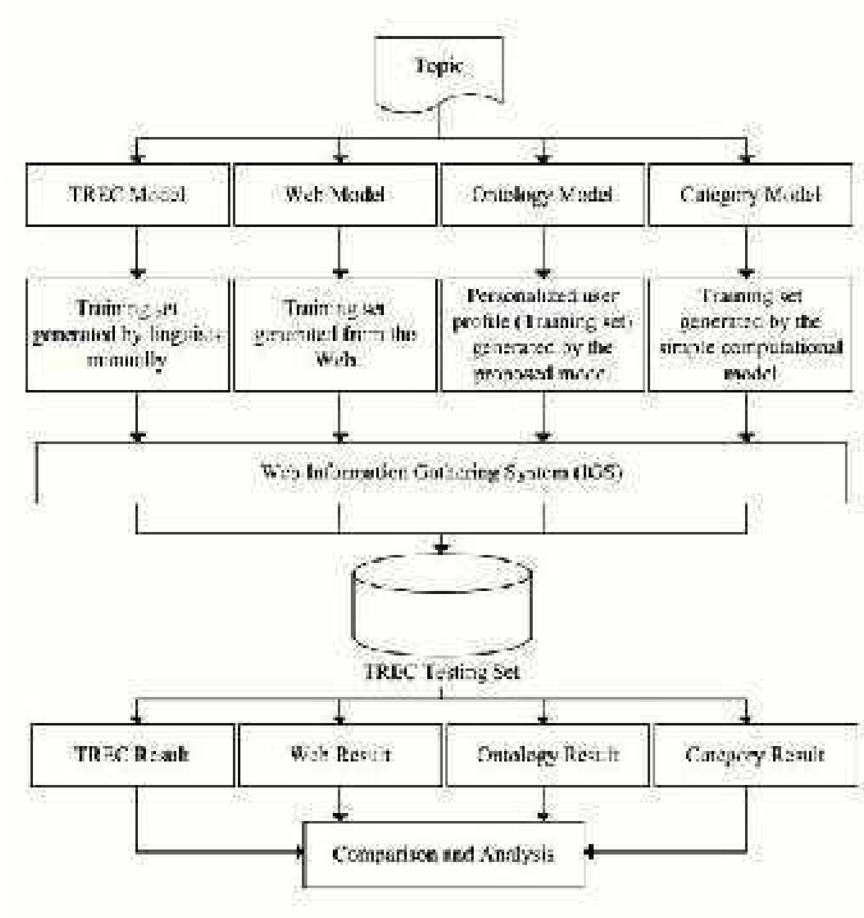


Fig5: Ex periment design.

9)Web Information Gathering System:

i) The information gathering system, IGS, was designed for common use by all experimental models.

ii) The IGS was an implementation of a model developed by Li and Zhong that uses user profiles for web information gathering.

iii) The input support values associated with the documents in user is also extensible in using support values of training information gathering .

10)Experimental Results:

i) The performance of the experimental models was measured by th ree methods: the precision averages at 11 standard recall levels (11SPR), the mean average precision (MAP).

ii) An 11SPR value is computed by summing the inter-polated precisions at the specified recall cutoff, and then dividing by the number of topics .

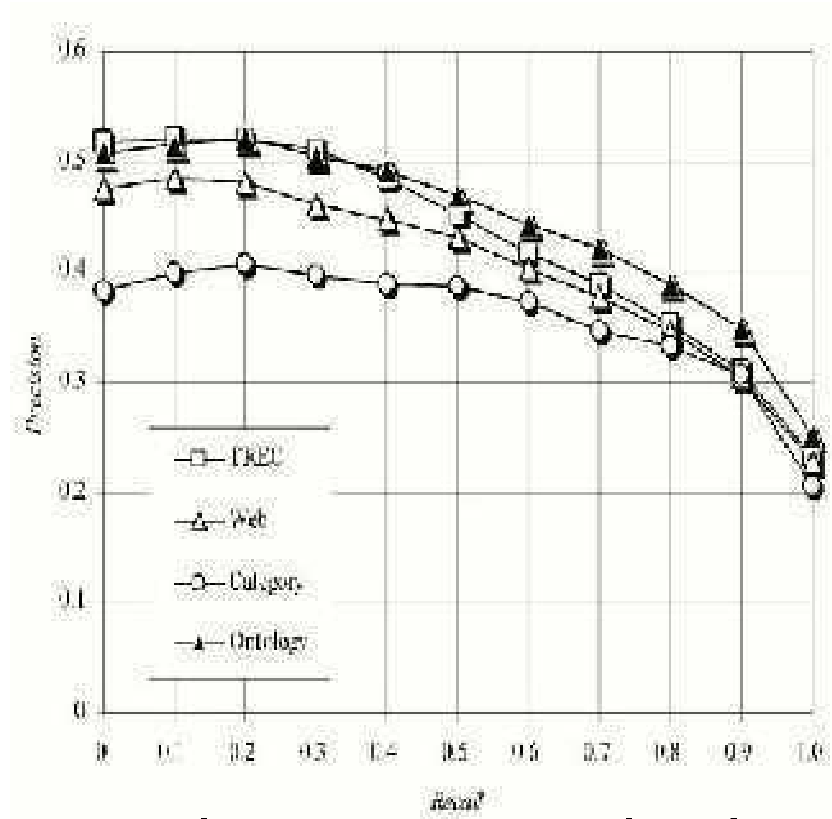


Fig6:The 11SPR ex perimental result. Table2 also presents the average macro-F1and micro-F1 Measure results. The F1 mesure is calculated by

TABLE 2

The MAP F1 Measure Experimental Result

	TREC	Web	Category	Ontology
MAP	0.2901	0.2775	0.2612	0.2856
Micro-FM	0.3559	0.3458	0.3288	0.3622
Macro-FM	0.3875	0.3759	0.3554	0.3941

TABLE3

Significan ce Test Results

	MAP		Macro-FM		Micro-FM	
Ontology vs.	*%Chg*	*p-value*	*%Chg*	*p-value*	*%Chg*	*p-value*
TREC	7.66%	0.882	7.00%	0.551	6.69%	0.519
Web	9.25%	0.026	8.57%	0.006	8.28%	0.005
Category	20.42%	0.0002	18.40%	0.0001	16.93%	0.0002

TABLE4

The Design of Experimental Models in the Sensitivity Test

	is-a only	*part-of* only	*is-a* and *part-of*	non-relationship specified
LTRs	-	-	-	Lac
WKB	GI	GP	GIP	-
LTRs + WKB	GLI	GLP	Ontology	-

11) Sensitivity Analysis:

i) The sensitivity analysis conducted in this paper aims to clarify the impacts made by different components in the Ontology model.

ii) As the architecture shows in Fig. 4, two knowledge resources, the global WKB and the LIRs, ar e used in the proposed mod el for user backgro und knowl-edge discovery.

iii) Does the model using all contributors have better performance than those using only one (or subcombina-tion of the four contributors .

iv)Which one is more important to the Ontology model, the is-a or part-of knowledge.

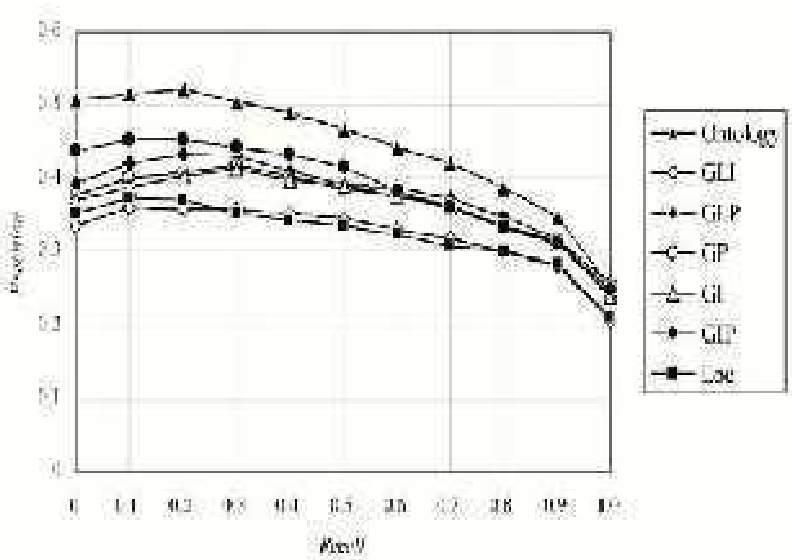


Fig. 7. The 11SPR results of sensitivity test.

TABLE5

The Average MAP and F-Measure Results of sensitivity Test

	Ontology	GIP	GLP	GP	GI	GLI	Loc
MAP	0.288	0.269	0.255	0.254	0.264	0.247	0.246
Micro-FM	0.362	0.337	0.335	0.332	0.332	0.313	0.309
Macro-FM	0.394	0.365	0.362	0.359	0.356	0.338	0.334

TABLE 6

T-Test Statistic Results for Sensitivity Test

		Ontology	GIP	GLP	GP	GI	GLI
GIP	MAP	0.002					
	Mic-FM	9.53E-06					
	Mac-FM	1.11E-06					
GLP	MAP	3.95E-06	0.435				
	Mic-FM	5.16E-06	0.755				
	Mac-FM	4.47E-06	0.674				
GP	MAP	1.59E-04	0.105	0.599			
	Mic-FM	2.46E-05	0.25	0.702			
	Mac-FM	1.86E-05	0.159	0.653			
GI	MAP	8.49E-05	0.127	0.541	0.846		
	Mic-FM	1.38E-05	0.263	0.633	0.998		
	Mac-FM	1.11E-05	0.177	0.625	0.927		
GLI	MAP	1.23E-08	0.006	9.89E-04	0.029	0.022	
	Mic-FM	1.33E-09	0.005	2.53E-04	0.028	0.020	
	Mac-FM	7.77E-10	0.004	2.52E-04	0.028	0.022	
Loc	MAP	1.80E-08	0.007	0.002	0.041	0.046	0.854
	Mic-FM	3.51E-08	0.008	0.001	0.036	0.035	0.555
	Mac-FM	3.46E-08	0.007	0.001	0.042	0.042	0.611

12)CONCLUSIONS AND FUTURE WORK:

We will investigate the methods that generate user local instance repositories to match the representation of a global knowledge base. The present work assumes that all user local instance repositories have content-based descriptors referring to the subjects. However, a large volume of documents existing on the web may not have such content-based descriptors. For this problem, strategies like ontology mappin g and text classification/clustering were su ggested. These strategies will be investigated in future work to solve this problem. The investigation will ex tend the applicability of the ontology model to the majority of the existing web documents and increase the contribution and significance of the present work.

13)ACKNOWLEDGMENTS:
i)This paper presents the extensive work of, but Significantly beyond, an earlier paper published in WI '07.
ii)The authors thank the Library of Congress and QUT Librar y for the use of the LCSH and librar y catalogs. The authors also thank also go to M. Carey-Smith, P. Delan ey, and J. Beale, for their assistance in proofreading and editing the paper.

14)REFERENCES:

[1] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison Wesley,

[2]W.J in, R.K.Srihari, H.H. Ho, and X. Wu, "Improving Knowledge Discovery in Document Collections through Combining Text Retrieval and Link Analysis Techniques.

[3]C.Makris,Y.Panagis,E.Sakkopoulos, "Categor y Ranking fo r Personalized Search,"Data and Knowledge

[4]A.-M.Popescu and O.Etzioni, "Extracting Product Features and Opinions from Reviews," Proc. Conf. Human Language Technologyand Empirical Methods in Natural Language Processing

[5]D.Quest and H.Ali, "Ontology Specific Data Mining Based on Dynamic Grammars," Proc. IEEE C omputational Systems Bonior-matics Conf.

[6]S.Sek ine and H. Suzuki, "Acquirin g Ontological Kn owledge from Query Logs," Proc. 16th Int'l Conf. World Wide Web

[7] R. M. C olomb,Information Spaces:The Ar chitecture of C yb erspace.

[8]G.A.Miller and F.Hristea,"WordNet Nouns:Classes and Instances, "Computational Lin guistics,

[9]G.E.P.Box,J.S.Hunter,and W.G.Hunter,Statistics For Experi-menters.John Wiley.

[10]M.D.Smucker,J.Allan,andB.Carterette,"A Comparison of Statistical Significance Tests for Information Retrieval Evalua-tion," Proc. 16th ACM Conf.