

# A Customized Ontology Shape for Web Data Congregation

V.Suguna<sup>1</sup>, M.Sai kishore<sup>2</sup>,

<sup>1</sup>M.Tech Student, <sup>2</sup>Asst. Prof

<sup>1,2</sup>Department of CSE,

Swetha Institute of Engineering & Technology,  
tirupathi, Andhrapradesh, India,

**Abstract—** As a sculpt for associate clarification and exemplification, ontology's are lengthily used to represent user profile in modified web in sequence parishioners. Equally, when in place of user profiles, several mold include used only associate from either a universal associate pedestal or user confined information.

In this paper, a modified ontology mold is designed for associate design and estimate over user profile. These molds learn ontological user profiles from equally a world associate stand and user curbed event repositories. The ontology shape is evaluated by compare it adjacent to measure models in web information parishioners. The penalty illustrate that this ontology mold is affluent.

**Keywords-** Ontology, world associate, local instance depository, user profiles, web information parishioners,

## I. INTRODUCTION

ON the previous decades, the extent of web-based in succession available has augmented fundamentally. How to assemble helpful in series from the web has rotate into a tough topic for user. Present web in succession parishioner's scheme effort to satisfy user provisions by incarcerate their data necessities. For this reason, client profiles are shaped for client location knowledge clarification.

Client profiles situate for the awareness replica gripped by client when parishioners web in series. A notion replica is completely obsessed by client and is produce from their situation knowledge. While this awareness replica cannot be established in laboratories, many system ontologists include untried it in client recital. When client read throughout documentation, they can simply terminate whether or not it is of their attention or consequence to them, a termination that arise from their implicit perception replica. If a user's beginning replica can be simulated, then a advanced revelation of client profiles can be construct.

To replicate user consideration representation, ontologies—a knowledge version and formalization model—are use in modified web in series congregation. Such ontologies are called ontological client profiles or custom-made ontologies [18]. To stand for client profile, a lot of researchers have attempted to decide user environment knowledge through entire or local search.

Global analysis utilizes offered global associate bases for user situation awareness design. frequently used command bases include general ontologies, thesauruses (e.g., digital libraries), and online acquaintance bases (e.g., online

categorizations and Wikipedia). The universal analysis system manufactures victorious exterior for consumer environment acquaintance removal. However, universal investigation is incomplete by the dominance of the used consciousness base. For example, WorldNet was report as supportive in capture client awareness in some locale but unpromising for others.

Local analysis investigates consumer local in progression or scrutinize client presentation in consumer profiles. For example, Li and Zhong discovered taxonomical pattern from the users' narrow text credentials to study ontologies for client profiles. A number of groups [12] well-read made to order ontologies adaptively from client's browsing the past. Otherwise, Sekine and Suzuki analyzed reservation kindling to realize client conditions knowledge. In some works, such as, regulars were provided with a set of recommendation and asked for implication response. Client conditions command was then revealed establishment this denigration for client profiles. However, because local investigation method rely on data elimination or organization technique for information detection, irregularly the bare results envelop noisy and uncertain in sequence. As a product, local scrutiny suffers from uselessness at capturing formal client acquaintance.

From this, we can assumption that user background attentiveness can be better naked and represent if we can put mutually worldwide and local inquiry within a mixture model. The acquaintance recognized in a universal awareness base will compress the background acquaintance recognition from the client constrained in sequence. Such a customized ontology model should assemble a bigger demonstration of client profiles for web in sequence gathering.

In this paper, an ontology illustration to assess this possibility is premeditated. This model simulate clients thought models by means of personalized ontologies, and endeavor to look up web in sequence parishioners performance by using ontological user profiles. The world acquaintance and a user's local happening reservoir (LIR) are used in the proposed depiction. World knowledge is commonsensical acquaintance acquire by people from understanding and edification [16]; an LIR is a user's not public collection of in sequence items. From a world acquaintance base, we assemble made to order ontologism by adopting user response on motivating knowledge. A multidimensional ontology removal technique, Specificity and Exhaustively, is also initiate in the projected replica for analyzing perception individual in ontologies. The projected

ontology model is evaluated by assessment touching numerous standard models through experiment use a large usual information set. The assessment results be confirmation for that the planned ontology replica is flourishing.

The research contributes to associate business, and has the in the near future to progress the design of made to order web in progression parishioners scheme. The contributions are novel and all the time more noteworthy, allowing for the immediate detonation of web in development and the escalating user-friendliness of online recommendation.

The paper is prepared as follows: Section 2 discusses the connected vocation; in Section 3, we inaugurate how modified ontologies are constructed for users; and in Section 4, we current the multidimensional ontology withdrawal method. After that, Section 5 gives the planning of the projected model; Section 6 discusses the estimation issues, and the penalties are analyzed in Section 7. Finally, Section 8 makes conclusions and addresses our vision work.

## II. RELATED WORK

### A. *Ontology Learning*

Global associate bases were used by lots of available replica to be taught ontologies for web in sequence congregation. For example, Gauch et al. [12] and Sieg et al. [15] learned customized ontologies from the unlock address list project to identify client's preferences and happiness in web explore. Wikipedia was used by Downey et al. [10] to help realize essential user happiness in queries. These works effectively revealed client background acquaintance; however, their concert was derisory by the authority of the global association pedestal.

Aiming at facts made to order ontologies, various works mined client background acquaintance from consumer confined in sequence. Li and Zhong [13] second-hand model acknowledgment and organization rule departure method to determine knowledge from client limited identification for ontology construct. Transept al. deciphers keyword queries to clarification Logic's conjunctive query and second-hand ontologies to symbolize client background associate. Zhong planned a sphere ontology learning approach that in employment various data extraction and natural-language selfless method. Navigli residential Onto study to notice semantic concept and associations from web credentials. Web contented captivating out method will be used by Jiang and Tan [16] to determine semantic associate from domain-specific satisfied credentials for ontology knowledge. Finally, Shehata et al. [14] captured client in order needs at the judgment level fairly than the deed level, and represented client profiles by the intangible Ontological Graph. The use of data withdrawal techniques in these models leads to more client background acquaintance being uncovered. However, the acquaintance exposed in these works prohibited noise and reservations.

In addition, ontologies were used in a lot of mechanism to look up the presentation of associate sighting. Utilize a fuzzy sphere ontology taking out algorithm, a equipment was developed by Lau et al. [19] in 2009 to make concept maps

based on the posts on online discussion forums. Quest and Ali [15] used ontologies to help data pulling out in ordinary databases. Jin et al. [17] integrated data removal and in sequence recovery technique to further enhance knowledge discovery. Doan et al. [8] proposed a model called stick and used machine education method to find alike concepts in different ontologies. Dou et al. [9] planned a construction for scholarship sphere ontologies using prototype collapse, clustering/classification, and association rules withdrawal techniques. These works attempt to walk around a direction to mold world contact extra competently.

### B. *Client Profiles*

Client profiles were second-hand in net in sequence congregation to construe the semantic meaning of query and imprison customer in sequence needs [12], [14]. User profiles were definite by Li and Zhong [14] as the motivating topics of a user's in sequence need. They also categorized client profiles into two diagrams: the data diagram user profiles acquired by analyzing a database or a set of communication [12], [11], [17], [13], [11]; the in sequence illustration user profiles acquired by using manual technique, such as questionnaires and interviews, or mechanical techniques, such as in sequence recovery and machine learning. Van der Sluijs and Huben projected a method called the frequent client Model constituent to look up the brilliance and consumption of user modeling. Wikipedia was in addition used by [10], to help establish client interests. In order to get hold of a consumer profile, Chirita et al. [6] and Teevan et al. worn a anthology of client desktop text recommendation and emails, and cached web pages to search client interests. Makris et al. acquire client profiles by a rank confined location of categories, and then utilized web pages to personalize look for outcome for a client. These instrument effort to get hold of client profiles in organize to discover client background acquaintance.

Client profiles can be categorized into three groups: interviewing, semi-interviewing, and non-interviewing. Interviewing client profiles can be deemed just the thing client profiles. They are acquired by means of physical techniques, such as review, interviewing client's, and analyze client confidential guidance sets. One distinctive example is the TREC Filtering Track teaching sets, which were generate by hand [13]. The users read each essay and gave a constructive or negative finale to the document alongside a given subject. Because, only client's finally know their security and preference, these instruction credentials exactly reflect user background associate. Semi-interviewing client profiles are acquired by partially mechanized techniques with restricted user association. These techniques typically provide client's with a inventory of categories and demand users for motivating or non-interesting categories. One distinctive example is the web training set attainment model introduce by Tao et al. [18], which extracts preparation sets from the net based on user feedback categories. Non-interviewing technique do not engage users at all, but resolve client wellbeing instead. They acquire client profiles by observe client movement and performance and discovering client background knowledge [11]. A typical model is OBIWAN,

planned by Gauch et al. [12], which acquires client profiles based on users' online browsing record. The interviewing, semi-interviewing, and non-interviewing client profiles can also be viewed as guide, semiautomatic, and habitual profiles, respectively.

### III. MODIFIED ONTOLOGY CONSTRUCTIONS

Modified ontologies are a conceptualization representation that properly describes and specifies client background acquaintance. From explanation in everyday life, we found so as to web users strength have dissimilar prospect for the same exploration query. For example, for the topic "New York," commerce travelers may command dissimilar in sequence from free time travelers. Occasionally even the same client may have dissimilar expectations for the similar search uncertainty if applied in dissimilar circumstances. A client may become a commerce traveler when development for a commerce trip, or a free time traveler when development for a family holiday. Based on this surveillance, a postulation is fashioned that web client's have a private concept model for their in sequence needs. A client's concept model may modify according to different in sequence requirements. In this section, a representation constructing made to order ontologies for web users' perception models is introduced.

#### A. World Knowledge Demonstration

World knowledge is significant for in sequence congregation. According to the description provided by [16], planet acquaintance is reasonable knowledge obsessed by people and acquired through their understanding and schooling. Also, as piercing out by Nirenburg and Raskin [19], "world information is compulsory for lexical and referential disambiguation, together with establishing co-reference relatives and resolving contraction as well as for establishing and maintaining connectivity of the conversation and observance of the text to the text producer's ambition and plans." In this planned model, client background acquaintance is extracted from a world acquaintance base programmed from the Library of legislature Subject Headings (LCSH).

We first require making the world acquaintance base. The world information base must cover up an comprehensive assortment of topics, since users may move toward from dissimilar backgrounds. For this reason, the LCSH arrangement is an ideal world acquaintance pedestal. The LCSH was developed for organizing and retrieving in sequence from a large quantity of documentation collections. For over a hundred years, the acquaintance contained in the LCSH has undergone unremitting reconsideration and enhancement. The LCSH represents the ordinary growth and allocation of human academic work, and covers all-inclusive and comprehensive topics of world acquaintance [5]. In addition, the LCSH is the most wide-ranging non-specialized proscribed terminology in English. In many respects, the arrangement has develop into a defects average for subject matter classification and indexing, and is used as a means for attractive topic admission to associate management scheme [5].

TABLE 1  
Comparison of Different World Taxonomies

	LCSH	LCC	DDC	RC
# of Topics	394,070	4,214	18,462	100,000
Structure	Directed Acyclic Graph	Tree	Tree	Directed Acyclic Graph
Depth	37	7	23	10
Semantic Relations	Broader, Used-for, Related-to	Super- and Sub-class	Super- and Sub-class	Super- and Sub-class

The LCSH agreement is greater evaluate with extra world associate taxonomy used in previous mechanism. Table 1 presents an estimation of the LCSH with the certification of lawmaking body classification (LCC) used by Frank in addition to Paynter [11], the Dewey Decimal Classification (DDC) used as a result of Wang and Lee [15] and King et al. [18], and the situation classification (RC) developed by Gauch et al. [12] using online categorizations. As shown in Table 1, the LCSH covers more topics, has a more precise structure, and specifies more semantic kindred. The LCSH descriptors are confidential by professionals, and the cataloging quality is definite by well-defined and continually refined category rules [5]. These skin tones make the LCSH an replica world associate base for associate manufacturing and administration.

The structure of the humanity knowledge pedestal used in this investigate is programmed from the LCSH reference. The LCSH classification contains three types of references: Broader term (BT), Used-for (UF), in accumulation to connected term (RT) [5]. The BT references are for two subject's recitation the same topic, but at disparate levels of construct (or specificity). In our replica, they are programmed as the is-a associations in the world associate base. The UF references in the LCSH are used for countless semantic situations, including improvement the semantic amount of an issue and recitation complex subjects and subjects subdivided by additional topics. The multiple practices of UF references make them tricky to educate. Throughout the assessment, we found that these orientations are time and over used to explain a realization or an entity. When entity A is used for an exploit, A becomes a portion of that use (e.g., "a divergence is used for dining"); when A is used for an additional object, B, A becomes a constituent of B (e.g., "a wheel is use for a car"). These cases can be fixed as the part-of kindred. Thus, we simplify the compound tradition of UF references in the LCSH and instruct them merely as the part-of relations in the world acquaintance stand. The RT orientation are for two subjects connected in some manner supplementary than by pecking order. They are automatic as the related-to relations in our world associate base.

The primitive contact element in our humankind acquaintance stand is subject. They are automatic from the topic heading in the LCSH. These subjects are dignified as follows:

Description 1. Let  $\mathcal{S}$  be a set of theme, an construction  $s \in \mathcal{S}$  pretentious as a 4-tuple  $s = (\text{label}; \text{neighbor}; \text{ancestor}; \text{successor})$ , wherever

- sticker is the description of s in the LCSH vocabulary;
- national is a assembly frequent the subject that have directly associations to s in the world knowledge support

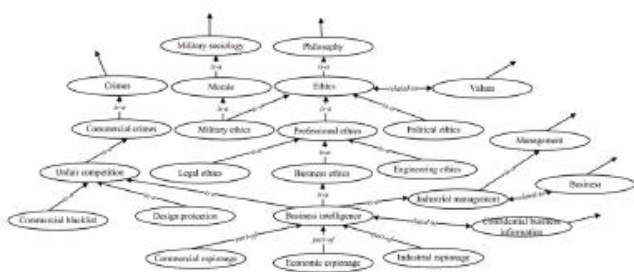


Fig 1: A sample part of the world knowledge base.

- Ancestor is a denotation regular the subjects that have a superior level of create than s and link to immediately or not directly in the humankind associate stand.
- Heir is a connotation frequents the subjects that are more detailed than s and link to s directly or diagonally in the world associate base.

The subject in the world associate base are connected to each other by the semantic kindred of is-a, part-of, and related-to. The families are formal as follows:

Definition 2 .Let to IR be a location of relatives, an component r belongs R is a 2-tuple  $r := \{edge; type\}$ , where

- an edge connect two subject that clutch a type of relative;
- a type of associations is an constituent of  $\{is-a; part-of; related-to\}$ .

With Definitions 1 and 2, the world associate stand can then be arrogant as follows:

Definition 3. Let WKB be a world associate base, which is taxonomy erect as a bound for acyclic chart. The WKB consists of a set of subjects associated by their semantic kindred, and can be suitably distinct as a 2-tuple  $WKB ::= \{S; IR\}$ , Where .

- S is a set of subjects  $S := \{s_1; s_2; \dots; s_m\}$ ;
- IR is a set of semantic relations  $IR := \{r_1; r_2; \dots; r_n\}$  linking the subjects in S.

**B. Ontology Production**

The subjects of consumer concentration are extracting from the WKB via user communiqué. A tool called Ontology knowledge environment (OLE) is developed to support client through such announcement. Concerning a argument, the gorgeous subject consist of two sets: optimistic subjects are the concepts applicable to the in progression need, and gloomy

subjects are the concepts resolving contradictory or confusing understanding of the in progression need. Thus, for a set topic, the OLE provides users through a set of candidate to distinguish assenting and pessimistic subjects. These applicant subjects are extract from the WKB.

Fig. 2 is a screen-shot of the OLE for the example topic “financial intelligence.” The subjects planned on the top-left plate of the OLE are the candidate subjects obtainable in hierarchical form. For each s 2 SS, the s and its acquaintances are

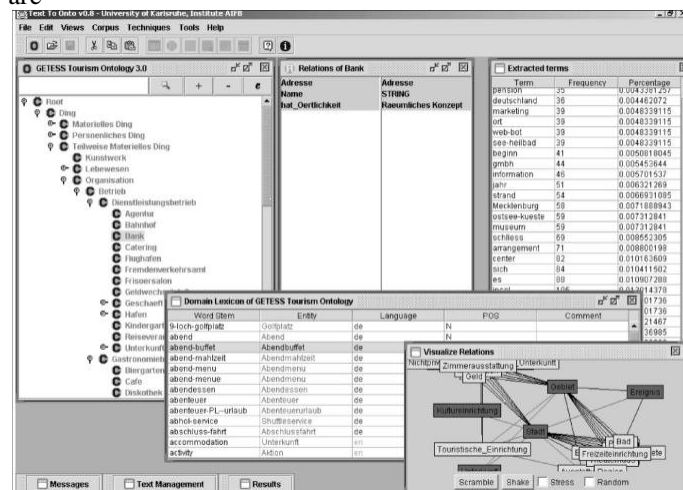


Fig 2. Ontology learning environment.

Recover if the tag of s contain any one of the query terms in the known topic (e.g., “financial” and “spying”). From these candidate, the user selects constructive subjects for the topic. The user-selected positive subjects are presented on the top-right section in hierarchical form.

The applicant negative subjects are the descendants of the user-selected positive subjects. They are shown on the bottom-left pane. beginning these negative candidates, the client selects the negative subjects. These user-selected distrustful subjects are listed on the bottom-right panel (e.g., “Political ethics” and “Student ethics”). Note that for the conclusion of the conformity, some optimistic subjects (e.g., “Ethics,” “Crime,” “Commercial crimes,” and “Competition Unfair”) are also incorporated on the bottom-right section with the negative subjects. These positive subjects will not be incorporated in the pessimistic set. The outstanding candidates, which are not fed back as either positive or halfhearted from the user, turn out to be the disinterested subjects to the specified topic.

Ontology is then built for the given topic by means of this client pointer subjects. The structure of the ontology is based on the semantic links concerning these data in the WKB. The ontology contains three types of contact: positive subjects, negative subjects, and impartial subjects. Fig. 3 illustrates the ontology (incompletely) build for the case in summit topic “Economic espionage,” where the white nodes are optimistic, the dark nodes are negative, and the gray



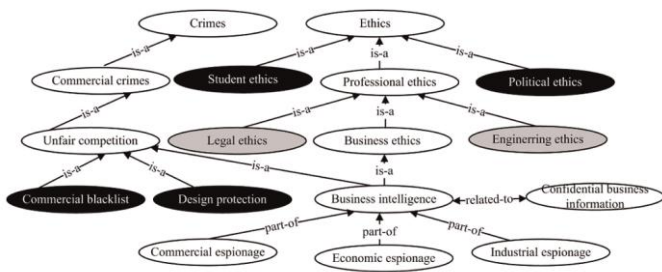


Fig.3 An ontology (partial) constructed for topic “Economic Espionage.”

#### IV. MULTIDIMENSIONAL ONTOLOGY MINING

Ontology elimination discover striking and on-topic associate from the notion, semantic kindred, and instances in an ontology. In this section, a 2D ontology mining method is introduced: Specificity and Exhaustivity. Specificity (denoted *spe*) describes a subject’s center on a given topic. Exhaustively (denoted *exh*) contain a subject’s semantic space dealing with the topic. This scheme aims to consider the subject and the potency of their relatives in ontology.

We quarrel that a subject’s specificity has two focuses: 1) on the referring-to concept (called semantic specificity), and 2) on the specified topic (called topic specificity). These need to be address independently.

##### A. Semantic Specificity

The semantic specificity is investigated based on the arrangement of  $O(T)$  innate from the world knowledge base. The strength of such a focus is influenced by the Subject’s locality in the taxonomic arrangement  $tax^S$  of  $O(T)$  (this is also argued by [42]). As stated in explanation 4, the  $tax^S$  of  $O(T)P$  is a graph linked by semantic relations. The subjects located at upper bound levels toward the root are further abstract than those at lower bound levels toward the “leaves.” The upper bound level subjects have more descendants, and thus refer to more conception evaluate with the lower bound level subjects. Thus, in terms of a concept being referred to by both an upper bound and lower bound subjects, the lower bound subject has a stronger focus because it has fewer concepts in its space. Hence, the semantic specificity of a lower bound subject is greater than that of an upper bound subject

#### V. CONSTRUCTION OF THE ONTOLOGY REPLICA

The projected ontology mold aims to determine user background associate and learns modified ontologies to symbolize consumer profiles. Fig. 6 illustrates the construction of the ontology mold. A tailored ontology is constructed, according to a given topic. Two acquaintance possessions, the global world knowledgebase and the user’s local occurrence depository, are utilized by the sculpt. The world associate base provides the taxonomic configuration for the tailored ontology. The client background associate is revealed from the user local instance depository. Against the given topic, the specificity and exhaustively of subjects are investigated for client background associate innovation.

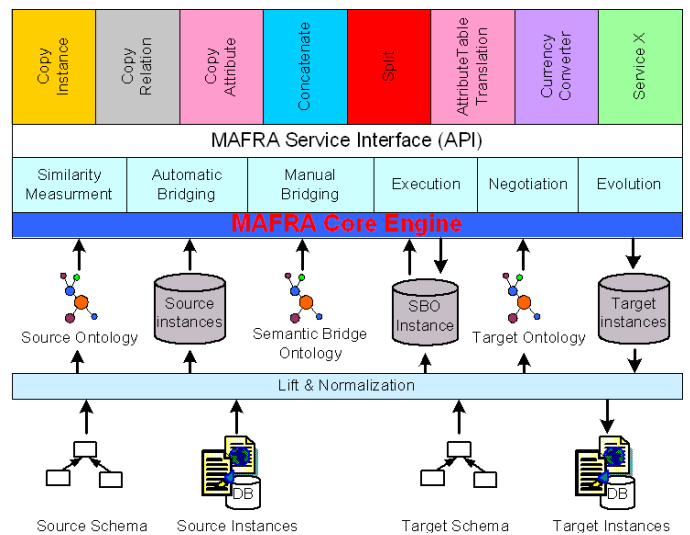


Fig: Architecture of the ontology model.

#### VI. EVALUATION

##### A. Experiment propose

The projected ontology sculpt was evaluated by intention experiments. Because it is tricky to evaluate two sets of acquaintance in diverse representations, the primary devise of the valuation was to evaluate the efficiency of an information congregation system (IGS) that used diverse sets of user background acquaintance for information congregation. The acquaintance discovered by the ontology sculpt was first used for a run of information congregation, and then the acquaintance manually particular by users was used for another run. The latter run set up a yardstick for the appraisal because the acquaintance was manually specified by users.

Under the similar untried circumstances, if the IGS might realize the same (or similar) presentation in two dissimilar runs, we could prove that the naked acquaintance has the same excellence as the user particular acquaintance. The projected ontology sculpt could then be proven capable to the domain of web information congregation. In information congregation evaluations, a common batchstyle research is developed for the evaluation of different models, using a test set and a set of topics connected with relevant judgments [36].

Our experiment followed this move en route for and was performed under the hesitant condition set up by the TREC-11 Filtering Track.3 This technique aimed to assess the methods of importunate client profile for separating apposite and non applicable papers in an conventional stream [12]. User atmosphere acquaintance in the experiments was represented by consumer profiles, such as those in the experiments of [13] and the TREC-11 Filtering Track. A consumer profile consisted of two copy sets: a positive document set containing the on-topic, appealing acquaintance, and a unnecessary manuscript set containing the contradictory, confusing concepts. Each manuscript d held a support value sustain the given topic. Based lying on this exhibition the baseline models in our experiment were suspiciously selected. User profiles can be categorized into three collections: session, semi-

interviewing, and non interviewing profiles, as previously discuss in Section 2. In an endeavor to estimation the planned ontology mold to the archetypal replica instead of these three group user profiles, four models were implemented in the experiment:

1. The Ontology mold that implement the designed ontology mold. user environment colleague was computationally naked in this mold.
2. The TREC sculpt that represented the perfect interviewing consumer profiles. User background familiarity was manually specified by users in this sculpt.
3. The grouping mold that represents the no consultation user profiles.
4. The Web mold that represent the semi-interviewing user profiles.

#### VII. CONCLUSIONS AND FUTURE WORK

In this paper, an ontology mold is probable for instead of user conditions knowledge for modified web information parishioners. The mold build user modified ontologies by extract world associate from the LCSH scheme and discover consumer milieu acquaintance from user local happening repositories. A multidimensional ontology mining technique, comprehensively and specificity, is also introduced for consumer background associate discovery. In appraisal, the normal topics and a enormous tested were used for experiments. The mold was compared against index models by applying it to a frequent system for information congregation.

The research consequences reveal that our planned sculpt is gifted. A sympathy investigation was also conducted for the ontology sculpt. In this analysis, we originate that the amalgamation of inclusive and local acquaintance works better than using any one of them.

In addition, the ontology sculpt using acquaintance with both is-a and part-of semantic kindred works better than using only one of them. When using only global acquaintance, these two kinds of kindred have the same assistance to the concert of the ontology sculpt. While using both global and local acquaintance, the acquaintance with part-of relations is more imperative than that with is-a.

The future ontology sculpt in this paper provides a explanation to emphasizing inclusive and local acquaintance in a single computational sculpt. The findings in this paper can be functional to the design of web information congregation systems. The mold also has general support to the fields of in sequence Rescue, web astuteness, recommendation scheme and Information scheme.

In our future work, we will observe the technique that engenders client local incident repositories to competition the depiction of a universal associate base. The nearby work assumes that all user restricted design repositories have content-based descriptors referring to the subjects, nonetheless a large quantity of papers presented on the web may not have such content-based descriptors. For this problem, in segment IV-B, strategies like ontology mapping and text

cataloging/clustering were suggested. These strategy will be investigated in prospect work to solve this problem. The assessment will enlarge the applicability of the ontology sculpt to the mainstream of the offered web papers and increase the donation and significance of the present occupation.

#### REFERENCES

- [1] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Addison Wesley, 1999.
- [2] G.E.P. Box, J.S. Hunter, and W.G. Hunter, *Statistics For Experimenters*. John Wiley & Sons, 2005.
- [3] C. Buckley and E.M. Voorhees, "Evaluating Evaluation Measure Stability," *Proc. ACM SIGIR '00*, pp. 33-40, 2000.
- [4] Z. Cai, D.S. McNamara, M. Louwerse, X. Hu, M. Rowe, and A.C. Graesser, NLS: A Non-Latent Similarity Algorithm," *Proc. 26th Ann. Meeting of the Cognitive Science Soc. (CogSci '04)*, pp. 180-185, 2004.
- [5] L.M. Chan, *Library of Congress Subject Headings: Principle and Application*. Libraries Unlimited, 2005.
- [6] P.A. Chirita, C.S. Firan, and W. Nejdl, "Personalized Query Expansion for the Web," *Proc. ACM SIGIR ('07)*, pp. 7-14, 2007.
- [7] R.M. Colomb, *Information Spaces: The Architecture of Cyberspace*. Springer, 2002.
- [8] A. Doan, J. Madhavan, P. Domingos, and A. Halevy, "Learning to Map between Ontologies on the Semantic Web," *Proc. 11th Int'l Conf. World Wide Web (WWW '02)*, pp. 662-673, 2002.
- [9] D. Dou, G. Frishkoff, J. Rong, R. Frank, A. Malony, and D. Tucker, "Development of Neuroelectromagnetic Ontologies(NEMO): A Framework for Mining Brainwave Ontologies," *Proc. ACM SIGKDD ('07)*, pp. 270-279, 2007.
- [10] D. Downey, S. Dumais, D. Liebling, and E. Horvitz, "Understanding the Relationship between Searchers' Queries and Information Goals," *Proc. 17th ACM Conf. Information and Knowledge Management (CIKM '08)*, pp. 449- 458, 2008.
- [11] E. Frank and G.W. Paynter, "Predicting Library of Congress Classifications from Library of Congress Subject Headings," *J. Am. Soc. Information Science and Technology*, vol. 55, no. 3, pp. 214-227, 2004.
- [12] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-Based Personalized Search and Browsing," *Web Intelligence and Agent Systems*, vol. 1, nos. 3/4, pp. 219-234, 2003.
- [13] R. Gligorov, W. ten Kate, Z. Aleksovski, and F. van Harmelen, "Using Google Distance to Weight Approximate Ontology Matches," *Proc. 16th Int'l Conf. World Wide Web (WWW '07)*, pp. 767-776, 2007.
- [14] J. Han and K.C.-C. Chang, "Data Mining for Web Intelligence," *Computer*, vol. 35, no. 11, pp. 64-70, Nov. 2002.
- [15] B.J. Jansen, A. Spink, J. Bateman, and T. Saracevic, "Real Life Information Retrieval: A Study of User Queries on the Web," *ACM SIGIR Forum*, vol. 32, no. 1, pp. 5-17, 1998.
- [16] X. Jiang and A.-H. Tan, "Mining Ontological Knowledge from Domain Specific Text Documents," *Proc. Fifth IEEE Int'l Conf. Data Mining (ICDM '05)*, pp. 665-668, 2005.
- [17] W. Jin, R.K. Srihari, H.H. Ho, and X. Wu, "Improving Knowledge Discovery in Document Collections through Combining Text Retrieval and LinkAnalysis Techniques," *Proc. Seventh IEEE Int'l Conf. Data Mining (ICDM '07)*, pp. 193-202, 2007.
- [18] J.D. King, Y. Li, X. Tao, and R. Nayak, "Mining World Knowledge for Analysis of Search Engine Content," *Web Intelligence and Agent Systems*, vol. 5, no. 3, pp. 233-253, 2007.
- [19] R.Y.K. Lau, D. Song, Y. Li, C.H. Cheung, and J.X. Hao, "Towards a Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning," *IEEE Trans. Knowledge and Data Eng.*, vol. 21, no. 6, pp. 800- 813, June 2009.