ISSN 2320 - 2602
Volume 3, No.1, January 2014
International Journal of Advances in Computer Science and Technology
Available Online at http://warse.org/pdfs/2014/ijacst07312014.pdf

Swarm Intelligence Based Fuzzy with Personalized Ontology Model for Web Information Gathering

S.Reha [1], M.Geetharani [2]
[1]Assistant Professor, Department of Information Technology , Sri Eshwar College Of Engineering , Coimbatore, Tamilnadu, India, rekhasece@gmail.com
[2]Assistant Professor, Department of Information Technology, Sri Eshwar College Of Engineering , Coimbatore, Tamilnadu, India, geethaitsece@gmail.com

ABSTRACT

Ontology is expansively used to represent the knowledge clarification and formalization of user profiles for gathering personalized web information. For representing user profiles from both global information and local information several models have been studied in literature. Personalized user profile based ontology model revise user profiles for mutually a world knowledge base and user local design repositories. But still it becomes less result in examination of local knowledge base to corresponding the demonstration of a global knowledge base in well-organized way. To overcome these issues proposed work analysis the improved method for local knowledge phase by using particle swarm optimization (PSO) algorithm. PSO based ontology method is used to find the best optimal information from local repository for each and every user profiles. Every particle in method is considered as local information for each user profile ,it move one position to another best web information user profiles by finding the best possible user profile. Changing the values of velocity of every particle toward its p_{id} and p_{gd} locations if founds the greatest optimized web information in simultaneously with local and global knowledge representation. It is clearly proved that swarm intelligence optimization of user profile provided that improved performance than a presented ontology for the equivalent system. The proposed swarm intelligence based ontology model is evaluated by comparing it alongside standard models in web information gathering.

Keywords: Ontology, Personalized Ontology, PSO (Particle Swarm Optimization) and web information gathering system.

1. INTRODUCTION

Due to the rapid development of World Wide Web (WWW), there has been an extraordinary gush in information sources available internationally in current years. Even though web users get benefit from free of charge access and dissemination of information. But while throughout this process they may also to face a number of challenges present are retrieving, filtering, and monitoring and ever-changing information, possibly of astronomical extent. A number of the issues that might impede quick and well-organized information recovery include size, heterogeneity, and lack of central control. The number of documents available on the web is disreputably huge, disseminated and rising. Continuous changes of the web pages, it might be complicated to keep up-to-date page indices [1]. Whilst the lack of central control permits free flow of information,

uncontrolled publishing on the web might result in “noisy” information sources. Moreover, information in the web frequently comes in dissimilar formats. For example, a webpage might include text documents or graphic images and text documents can make use of dissimilar terms for the similar conception.

Even though there are numerous of existing search engines, they are mostly being used to establish information sources and URL links containing associated information or documents that are related and of wellbeing to users’ queries [2]. Although traditional search method rest greatly on the users profiles, because they examine user information and task information from the listing of recommended URL links from information sources. Therefore, even if users might be familiar with somewhere to retrieve advantageous information they have to 1) Essentially visit the website(s) and 2) frequently and regularly visit the websites to retrieve up-to-date information. It becomes a time consuming task and might cause redundant Internet traffic [3].

Researchers and practitioners in the field of databases and information combination contain a large organization of study to make easy interoperability among diverse systems. This investigation ranges starting from methods for matching database schemas to answering queries using numerous sources of data. Ontology research is a new discipline that deals by

means of semantic heterogeneity in structured data. Study the detail procedure of ontology, their divergence starting database schemas, and difficulties in semantic integration that the ontology area faces [4].

Particular knowledge services consequently necessitate to search and extract particular information from unstructured text on the Web, to perform this guided by ontology that details of which type of information to gather. Classification of domain knowledge using ontology identifies the relationship between concepts and relations [5].Whereas there are numerous definitions specifies what is ontology [6], the regular explanation for these definitions is that an it is various prescribed explanation of a domain of discourse, wished-for sharing amongst diverse applications, and articulated in a language that can be used for analysis. These features of ontologies emphasize the major trends that differentiate semantic-integration investigate in the ontology: Primarily the fundamental objective of ontology growth is to generate artifacts that diverse applications that can then be extensive for other precise domains and applications. These developed ontologies are used for reasoning engines and semantics of ontology languages are specific with analysis of result, conclusion and analysis takes middle point in ontology-integration approaches.

In this article, PSO based ontology system is anticipated for knowledge illustration and examination in overkill of user profiles in local repositories as fine as global knowledge phase. PSO based ontology system analysis ontological user profiles from mutually a world knowledge base and user local instance repositories. The ontology model is evaluated by comparing it against existing models in web information gathering, it shows that PSO ontology better results.

2. RELATED WORK

Open Directory Project(ODP) to identify users' preferences and wellbeing search result are analyzed by creation of personalized ontology profiles by authors Gauch et al. [7] and Sieg et al. [8]. King et al. [9] introduced IntelliOnto based on Dewey decimal classification in disseminated web information recovery. Downey et al. [10] uses the Wikipedia concept to help identify with essential user interests in queries. These works successfully exposed user background information; though, their performance was imperfect by the excellence of the global information bases. Aiming at learning personalized ontologies, numerous works mined user background information from user restricted information.

Li and Zhong [11] second-hand pattern recognition and ARM methods to discover information from user local documents for ontology creation. Tran et al. [12] covert the basic keyword queries into Description Logics' conjunctive queries and used ontologies to characterize user background information. Zhong [13] anticipated a domain ontology learning approach that working a variety of data mining and natural-language sympathetic techniques.

Moreover, ontologies were used in numerous workings to get better the performance of knowledge invention. Lau et al. [14] build fuzzy domain ontology to map the concept based on the posts on online dialogue forums. Quest and Ali [15] used ontologies to analyze data mining in biological databases. Jin et al. [16] included data mining and information retrieval methods to additional improve information discovery.

Creation of user profiles from interviewing results is developed by using semi-automated techniques. These techniques regularly give users with a listing of categories either interesting or noninteresting category. One characteristic instance is the web training dataset illustration introduced by Tao et al. [17], which extracts training sets from the web based on user feedback categories. Noninterviewing methods doesn't entail users at every one, except ascertain user interests as an alternative. They obtain user profiles by observing user action and discovering user background information [18].

3. SWARM INTELLIGENCE BASED FUZZY WITH PERSONALIZED ONTOLOGY MODEL

Perform personalized web information gathering system first collect the set of documents and their corresponding associated information for each documents with user profiles .Extracting information from knowledge phase keywords are used. This keyword extraction is reducing the analyzing procedure. The users study every document and gave a positive or negative opinion to the document beside a specified topic. Since, user only has their own interest and preferences to accurately reproduce user background knowledge or information. User profiles are specified with help of semi automated methods with few involvement of user. These techniques regularly give users with a listing of categories either interesting or noninteresting category. Generally

three way of process can be performed to acquire user profiles that are interviewing, semi-interviewing, and non interviewing user correspondingly. Ontology is generated from user local information web instance collections. It is named as local instance repository (LIR). These documents are perfect to create the instances for ontology populace. Similarly dissimilar global knowledge bases are used for ontology and mapping local knowledge to global knowledge with ontology maps dissimilar representation of user profiles. These generated web document with ontology might be formulated or non-formulated in structured manner, it have straight indication to the concepts those specified in ontology either global or local information.

Globally information are locally stored in ontology are developing separately, then once in a while new web collection documents are added and their determination require to be a number of method for the additions and any changes made in local ontology added to global ontology automatically. Every local group would have a delegate; it would possibly too be a superior scheme to encourage professional classifiers to contribute. Throughout the meetings, the subsequent behavior would take place: investigation of innovative contents of local ontologies; recognize what requirements to be combined and promoted to the comprehensive stage; identify what requirements to wait local, but should be mapped to conditions in the global ontology.

PSO is a robust stochastic optimization system based on the progress and aptitude of particles. It uses a numeral of agents that represent a swarm moving approximately in the exploration space looking for the most excellent result. It is also a computational technique that optimizes a difficulty by iterative process to improve results. Search space of best solution is performed by formulation of each particle with known velocity and position. Every particle's group is predisposed by its restricted best well-known location and is moreover guided in the direction of the best well-known position in the search, simultaneously updated by each and every step. Additionally, every individual has a memory, remembering the best location of the search space it has increasingly visited. Thus, its group is an aggregated quickening towards the greatest character of a topological locality. PSO based ontology system is anticipated for knowledge representation and analysis in excess of user profiles in local repositories as well as global knowledge phase

Proposed PSO with local instance repository ontology are organized according to some information are collected from web related to personalized information they often to find the best personalized web information related to concepts. After selection of personalized user information constructed from ontology, and then studies their related web information with predefined selection of web information with concepts and generate suitable web information Every particles go towards in the direction of j^{th} user personalized local ontology in i's neighborhood of personalized local ontology of user both together similar to the same concept that are searched by user. Each particle communicates with some other particles with related concepts of both user and is affected by the best point local personalized ontology discovered by any member of its current local instance personalized ontology pi. The vector p_i for that best local instance personalized ontology , which we will indicate with p_g. Initialize the particle's location best identified personalized ontology related to the web concepts to its initial personalized ontology position: $p_i \leftarrow x_i$., then likewise update the location of the personalized user profiles concepts and their velocity to know the best global web information position .In PSO algorithm repeat the steps ,until efficient personalized ontology web information are gathered in local instance repository .Finally after finding the global best web information only considers as best information result .

Particle swarm optimization (PSO) with local knowledge phase

1. Initialize a population array of user profiles is considered as particles with random positions of profiles with collected information in ontology and velocities on D dimensions in the search space of best local information representation.
2. loop
3. For each user profile in ontology for both local and global knowledge particle, evaluate the desired optimization fitness function in D variables.
4. Compare particle's that is user personalized ontology local knowledge fitness evaluation with its $pbestloc_i$. If current value is better than $pbestloc_i$, then set $pbestloc_i$ equal to the current value of the best user profile for web information gathering , and Initialize the particle's that is user personalized ontology in local knowledge location best known position to its initial position: $p_i \leftarrow x_i$
5. Identify the particle in the neighborhood with the best success and assign its index to the variable g.
   5.1. If ($f(p_i) < f(g)$) update the swarm's best known position: $g \leftarrow p_i$
6. Initialize the particle's velocity: $v_i \sim U(-|b_{upv}-b_{lov}|, |b_{upv}-b_{lov}|)$

7. Change the velocity and position of the particle local instance ontology according to the following equation
8. Until a termination criterion is met ,repeat:
   8.1. For each particle that is personalized local ontology (i = 1, ..., S)
   8.2. do
   8.3. For each dimension d = 1, ..., n
   8.4. do
   8.5. Pick random numbers r_p, $r_g \sim U(0,1)$
   8.6. Update the particle's velocity $v_{i,d} \leftarrow \omega\ v_{i,d} + \varphi p_{rp}\ (p_{i,d}-x_{i,d}) + \varphi grg\ (g_d-x_{i,d})$
   8.7. Update the particle's position: $x_i \leftarrow x_i + v_i$
   8.8. If $(f(x_i) < f(p_i))$ do:
   8.9. Update the particle's best known position: $p_i \leftarrow x_i$
   8.10. If $(f(p_i) < f(g))$ update the swarm's best known position: $g \leftarrow p_i$
   8.11. Now g holds the best found solution.
9. end loop

.

Mapping among two ontologies

When dissimilar global knowledge bases are second-hand, ontology mapping techniques can be used to match the concepts in dissimilar representations. Mapping together two ontology by using the measurement of similarity between the terns in LIR and their corresponding concepts in both LIR. Based on the similarity calculation personalized ontology model based documents are classified. Mapping methods have been used to map the features that are found by using either clustering or classification methods in this work clustering methods with dissimilar representation of features are considered as input .Mapping results create a new ontology model for every local and global knowledge repository .The world information repository creates the taxonomic construction for the personalized ontology. It discovers knowledge from local instance repository for given topics.

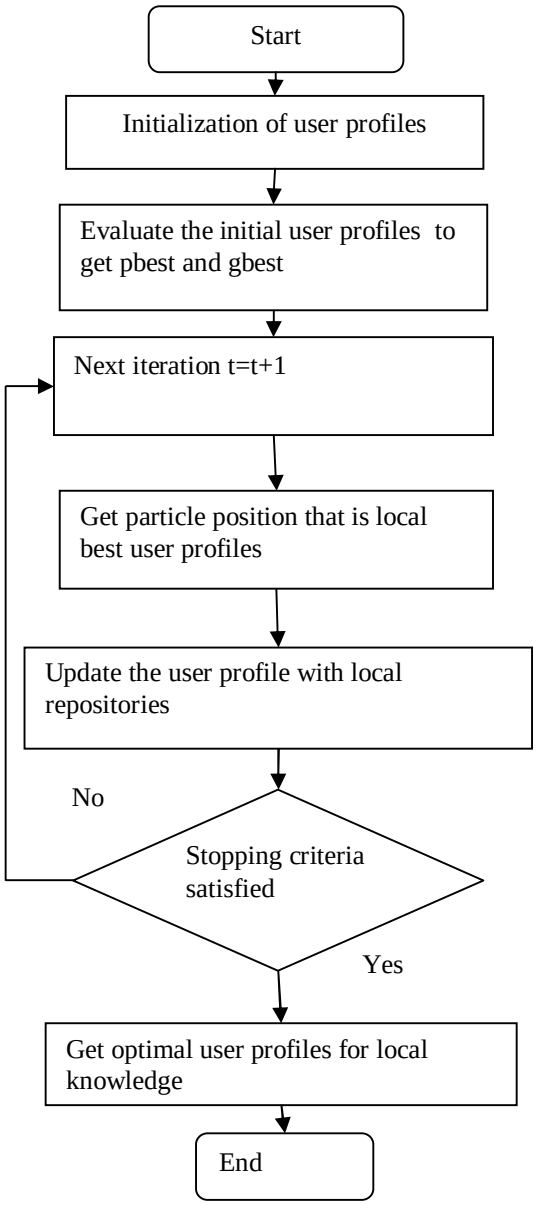

Figure 1: Flowchart of proposed personalized ontology model

4. EXPERIMENTAL RESULTS

The experiments were designed to compare the information gathering result from performance results shown by proposed PSO personalized ontology model with existing personalized ontology model in terms of the parameters are mean average precision (MAP), and the F1 Measure for information gathering evaluation [19]. Precision is the capability of a system to recover solitary relevant documents. Recall is the capability to recover all applicable documents.

4.1 F1 Measure

The F_1 Measure is calculated by

$$F_1 = \frac{2 \times precision \times recall}{precision + recall} \quad (1)$$

Where precision and recall are evenly weighted average for information gathering system it is called as F-measure. For every topic, the macro-F_1 Measure averages the precision and recall and then calculates F_1Measure, while the micro-F_1 Measure estimates the F1 Measure for every returned effect and then averages the F_1 Measure values. The superior F_1 values specify the better performance. Figure 4.1 shows the performance results of the methods with parameters specified in Table 1 and their results are also tabulated.

Table 1: The MAP and F_1 Measure Experimental Results

	Web	**Category**	**Ontology**	**PSO with ontology**
MAP	0.2875	0.2712	0.2916	0.38978
Micro-FM	0.3558	0.3345	0.3756	0.4125
Macro-FM	0.3789	0.3664	0.3989	0.4987

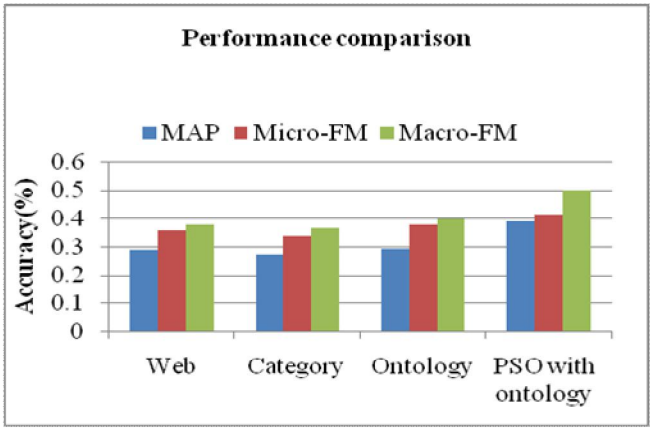

Figure 4.1: Performance comparison vs parameters

4.2 Statistical test

The statistical tests were also performed for the reliability of the evaluation. Regularly, a consistent significance test concern the dissimilarity in the mean of a measuring metric (e.g., MAP) and the significance level p-value calculated for the probability that a value could have occurred under a given zero value [20], [21].The percentage change in performance is used to calculate the differentiation in MAP and F_1 Measure results occurred among the Ontology model and a target model. It is calculated by,

$$Percentage = \frac{1}{N} \times \sum_{i=1}^{N} \frac{result_{Ontology} - result_{target}}{result_{target}} \times 100\% \quad (2)$$

A larger $percentage$ value shows additional considerable upgrading achieved by the Ontology model. Table 2 presents the average $percentage\ values$ results in this test. As shown, the Ontology results and their corresponding are shown in Figure 4.2, 4.3, 4.4.

Table 2: Significance Test Results

PSO with fuzzy ontology vs ontology	MAP		Macro-FM		Micro-FM	
	$\%percentage$	$p-value$	$\%percentage$	$p-value$	$\%percentage$	$p-value$
TREC	7.86%	0.891	7.15%	0.612	6.85%	0.529
Web	9.45%	0.031	8.86%	0.1	8.51%	0.0545
Category	20.12%	0.0002	18.89%	0.05	16.97%	0.001

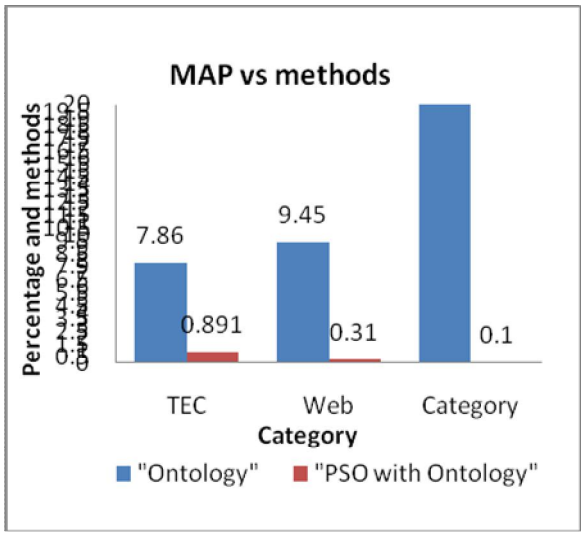

Figure 4.2: MAP vs methods with stastical test

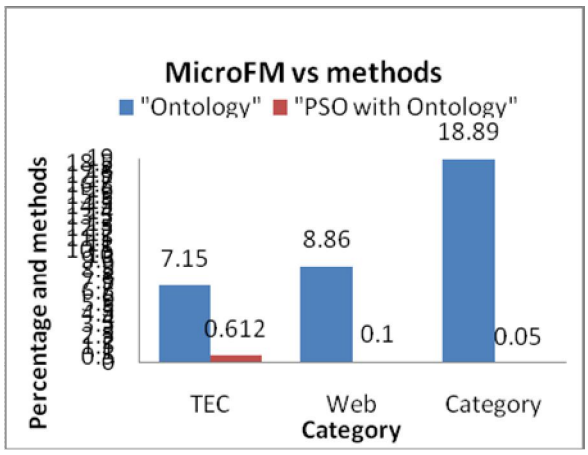

Figure 4.3: MicroFM vs methods with stastical test

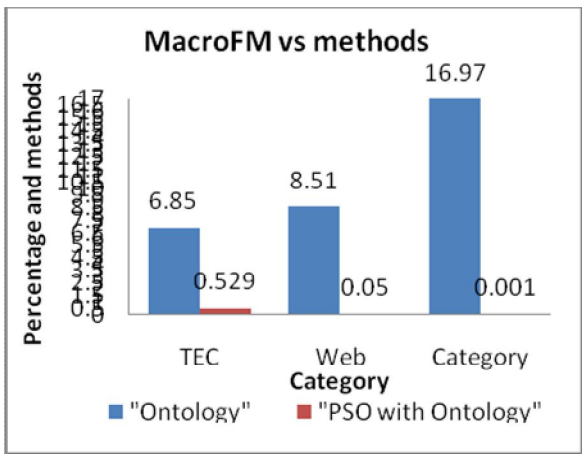

Figure 4.4: MacroFM vs methods with stastical test

5. CONCLUSION AND FUTURE WORK

In this paper, the construction of an improved PSO based personalized ontology model is obtainable as in this proposed ontology model a Similarity model is working in organize to present a semantic similarity inside the procedure also have been evaluated the approach. PSO based personalized ontology constructs personalized user profiles by management system related world information and finding corresponding user background information from user local illustration repositories. The experimental results shows that our proposed PSO based personalized ontology model performs improved than any other models compared. This experimental study says that this proposed model can be used for the growth of definite web information gathering systems and such models.

In future work focus apply Non-Latent Similarity model is employed in order to improve the present a semantic similarity within the process also we have evaluated the approach.

REFERENCES

1. O. Etzioni and D. Weld, "**A softbot-based interface**," *in Readings in Agents*. San Mateo, CA: Morgan Kauffman, 1997, pp. 71–81.
2. K. M. Sim, "**Toward holistic web-based information retrieval: An agent based approach**," *in Proc. 2003 IEEE/WIC Int. Conf. of Web Intelligence*, pp. 39–46.
3. M. Anandarajan, "**Profiling web usage in the workplace: A behavior based artificial intelligence approach**," *J. Manage. Inf. Syst.*, vol. 19, no. 1, pp. 243–266, 2002.
4. M. Uschold and M. Gr¨uninger," **Ontologies and semantics for seamless connectivity**" *SIGMOD Record*, 33(3), 2004.
5. M. Crub´ezy and M. A. Musen," **Ontologies in support of problem solving**", *In S. Staab and R. Studer, editors, Handbook on Ontologies*, pages 321–342. Sringer, 2003.
6. C. Welty," **Ontology research**", *AI Magazine*, 24(3), 2003.
7. 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.

8. A. Sieg, B. Mobasher, and R. Burke, "**Web Search Personalization with Ontological User Profiles**," *Proc. 16th ACM Conf. Information and Knowledge Management (CIKM '07)*, pp. 525-534, 2007.

9. 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.

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. Y. Li and N. Zhong, "**Mining Ontology for Automatically Acquiring Web User Information Needs**," *IEEE Trans. Knowledge and Data Eng.*, vol. 18, no. 4, pp. 554-568, Apr. 2006.

12. T. Tran, P. Cimiano, S. Rudolph, and R. Studer, "**Ontology-Based Interpretation of Keywords for Semantic Search**," *Proc. Sixth Int'l Semantic Web and Second Asian Semantic Web Conf. (ISWC '07/ ASWC '07)*, pp. 523-536, 2007.

13. N. Zhong, "**Representation and Construction of Ontologies for Web Intelligence**," *Int'l J. Foundation of Computer Science*, vol. 13, no. 4, pp. 555-570, 2002.

14. 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.

15. D. Quest and H. Ali, "**Ontology Specific Data Mining Based on Dynamic Grammars**," *Proc. IEEE Computational Systems Bioinformatics Conf.* (CSB '04), pp. 495-496, 2004.

16. W. Jin, R.K. Srihari, H.H. Ho, and X. Wu, "**Improving Knowledge Discovery in Document Collections through Combining Text Retrieval and Link Analysis Techniques**," *Proc. Seventh IEEE Int'l Conf. Data Mining (ICDM '07)*, pp. 193-202, 2007.

17. X. Tao, Y. Li, N. Zhong, and R. Nayak, "**Automatic Acquiring Training Sets for Web Information Gathering**," *Proc. IEEE/WIC/ ACM Int'l Conf. Web Intelligence*, pp. 532-535, 2006.

18. J. Trajkova and S. Gauch, "**Improving Ontology-Based User Profiles**," *Proc. Conf. Recherche d'Information Assistee par Ordinateur (RIAO '04)*, pp. 380-389, 2004.

19. C. Buckley and E.M. Voorhees, "**Evaluating Evaluation Measure Stability**," *Proc. ACM SIGIR '00*, pp. 33-40, 2000.

20. G.E.P. Box, J.S. Hunter, and W.G. Hunter, "**Statistics for Experimenters** ".*John Wiley & Sons*, 2005.

21. M.D. Smucker, J. Allan, and B. Carterette, "**A Comparison of Statistical Significance Tests for Information Retrieval Evaluation**," *Proc. 16th ACM Conf. Information and Knowledge Management (CIKM '07)*, pp. 623-632, 2007.