

SEMANTIC RECOMMENDER SYSTEM WITH NATURAL LANGUAGE INTERFACE: MALAYSIAN TOURISM INDUSTRY

Tirad M. Almalahmeh, Sameem Abdul Kareem, Mansoor A. Abdulgaber

Abstract- Tourism economy is one of the fastest growing activities in developed countries. Information and communications technology (ICT) continues to be one of the greatest influences fuelling dramatic changes in reducing uncertainties and perceived risks to enhance the quality of trips by providing renewed web services. Many research findings caution that tourists are often overloaded by relevant and irrelevant information and resources. However, information searching and extraction as well as the interpretation of the information are left to the human user. In this paper we show how the promising integration of Semantic technology, ontology, recommender system and natural language interface can be used for next-generation tourism information systems.

Index Terms - Semantic Technology, Ontology, Recommender System, Natural Language Interface, Content Base Filtering, Natural Language Processing, Tourism

1. INTRODUCTION¹

The fast development of Information and communications technology (ICT) and the increase of Internet users have reshaped the tourism Industry around the world. ICT has been the backbone of many process innovations. As a result, incorporating ICTs in tourism businesses enrich the field with greater productivity, decreased costs, increased revenues and improved customer service[1-3]. However, Information is generated and published by multiple official or unofficial tourism sources. Searching for information on a desired spot for vacation is usually difficult and very time-consuming. For example tourists usually have problems when trying to find the available choices of accommodation and what is more suitable to their needs. Moreover, the open and distributed nature of the Web makes it difficult for Web search engines to find information related to user information. Users feel overwhelmed before finding the intended information, in

which an individual cannot process all the communication input. Tourists are often overloaded by relevant and irrelevant information and resources. Providing personalized information with particular interest to the user is still a challenging task for tourism information systems.

This paper is structured as follows. Section 2 presents the background and motivation. Section 3 discusses the novelty of the Semantic Malaysia Tourism Recommender System. Section 4 shows the implementation of the system. Section 5 discusses the evaluation of the system. Finally, Section 6 and 7 concludes and summarizes the contribution.

2. BACKGORUD AND MOTIVATION

Tourism is an information based domain[4]. Researchers in the field regard tourism as a hybrid industry since: it is information-based services while the core product is mainly physical services. Hence, this combination requires integration between information and physical services [5]. This information-based domain adopted the World Wide Web in order to improve the process of retrieving tourism related information. Therefore, the WWW is the main source of information. Information is considered as one of the biggest needs for tourists. If they have the appropriate information, it will help them in making their choices about (what to do, where to stay, and how to get there) the trip [6].

• Author **Tirad Almalahmeh** is currently pursuing PhD degree program in Artificial Intelligent field in University of Malaya, Malaysia, PH-+962797823300. E-mail: trad@siswa.um.edu.my

• Co-Author Assoc. Prof. **Sameem Abdul Kareem** a staff member in Artificial Intelligent Department, University of Malaya, Malaysia, E-mail: Sameem@um.edu.my

• Co-Author **Mansoor A. Abdulgaber** Obtained his PhD in Artificial Intelligent from University of Malaya. His current research activity in Semantic Technology and Knowledge Management Systems. Email: mansoorhak@me.com

Faculty of Computer Science & Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia.

However, this type of information are often isolated or intended with predefined programmes which simply broadcasts the tourism information, and provide the same kind of information to users with different interests. On the other hand, information-overload is another issue, since providers seemed to focus on delivering as much information as possible and fail to take into account specific users' needs [7]. Furthermore, information searching mechanism depends on users' effort while machine agents are still unable to perform, as the information is not machine readable. Meanwhile, moving from the current broadcasting information to intelligent machine readable information is not an easy task as satisfying users' requirements should be considered during the designing and implementation of these intelligent services[8].

In this context, Semantic Technology is perceived as a very promising technology. It enhances the web by enabling machines to understand the information available. Semantic technology enables data representation in a machine-readable form. Such representation facilitates the integration of tourist resources and data exchange among systems, which may include semantic descriptions of users and products provided[9, 10]. Meanwhile, Ontology plays an essential role in realizing the Semantic Web, which is a set of shared, explicit and formal concepts used to organize and classify contents of the domain (which in this case is the tourism domain). The task of inferring new knowledge from facts and rules is expressed in an ontology language so it can be used to reason about most important concepts of that domain, their attributes and relations between concepts [11]. Several ontology query languages have been developed for extracting this knowledge from ontologies such as SPARQL, but for casual end-users it is highly impossible to learn and use one of these query languages [12]. Also, users need to understand the contents of the ontology in order to build a query, where users actually prefer to query the ontology using their natural language such as (English). Supporting natural language is more user-friendly and will bring the advantages of this knowledge closer to the casual users [13].

According to the interface evaluation conducted in Kaufmann and Bernstein [13], systems developed to support Natural Language Interfaces (NLI) are perceived as the most acceptable by end-users. Familiarity with the natural language used in these systems is a key to simplify the information retrieval processes. Natural language interfaces have the possibility to answer tourist questions about tourism related items. Tourists may find many answers provided by the Natural languages interface for each question. These answers are not

personalized information with a particular interest to each individual user. A Recommender System (RS) is one of the applications that provide personalized information; the goal of RS is providing customized information in the World Wide Web environment. It actively constructs the personalized information environment, provides service according to the users' interest and filters the information resources according to the users profile [14]. Tourists have individual preferences so the user profile plays a vital role in the personalization process. The tourism User Profile is a structured representation of the tourists' preferences. Accurate User Profile enhances the information customization efficiently. In general, user profiles distinguish the needs of different users. It assists in providing customized answers to a particular tourist's query which matches these requirements.

Therefore, we reiterate the scope of this paper in adopting the natural language interface to provide personalized information in the Malaysian tourism domain. The aim is to enrich the semantic answers with recommendations based on the tourists' preferences.

3. THE SEMANTIC MALAYSIA TOURISM RECOMMENDER SYSTEM

The main purpose of this work was to get to the bottom of the tourism information finding process, to assist tourists with relevant information. The Semantic Malaysia Tourism Recommender System (SMTRS) was designed and developed with this purpose in mind. SMTRS adapts earlier techniques and include new techniques such as a Relation Search Algorithm (RSA)

In setting out how we achieved this goal, this research involved three major parts:-

- I. The first part is concerned with the development of an ontological model for representing the tourism resources.
- II. The second part is concerned with the development of a specific architecture for answering tourists' queries based on semantic natural language.
- III. The third part is concerned with enriching the queries' answers from the second part with Content-based recommendation based on the classification of the user preferences.

4. IMPLEMENTATION OF SMTRS

SMTRS integrates personalization and semantic, implemented in Java as a web application, using client server architecture. SMTRS will not only answer tourists' question semantically, but it enriches the semantic answers with a recommendation based on the tourists' preferences using interfaces that are easy to use, as well as the provision of ontology to harness the real power of the semantic representation.

The process undertaken to develop SMTRS culminated in the following procedures:-

4.1. System Process

In this process, we developed the modeling diagram, which describes the main components of SMTRS as shown in Figure 1.

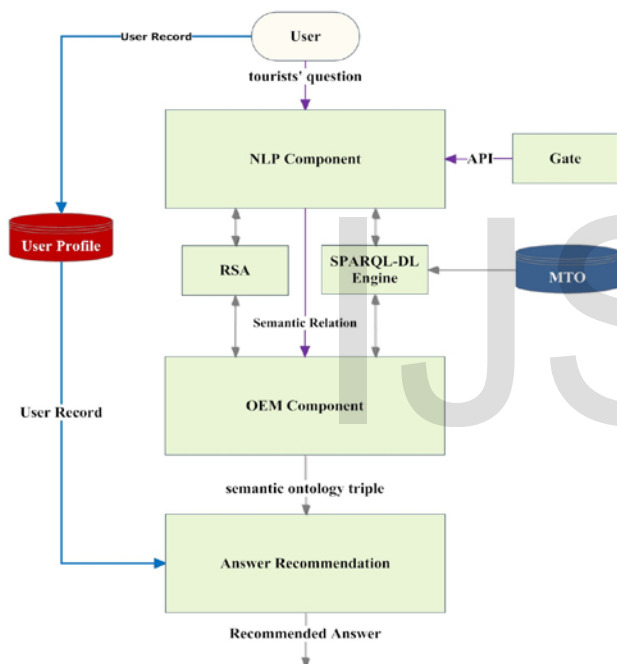


Figure 1 SMTRS' Component Architecture

a. Natural Language Processing (NLP) Component

The task of the NLP component in our prototype is to recognize the semantic relations in the Natural Language query, and identify a query triple, considered as the simplest form of a sentence structure (Subject, Predicate, and Object). In SMTRS we generate the triples by using the General Architecture for Text Engineering (GATE) as discussed in [15]. Furthermore, we built a RSA as a method to overcome the limitation of the GATE, for instance, in the case when the GATE is not able to capture the potential relation or terms to form the triples. RSA

tokenizes the tourists' question and search each word in the Malaysia Tourism Ontology (MTO) relations (either object or data property) to identify the semantic relation.

b. Ontological Entity Mapping OEM

The task of this component is to identify the entities (terms) in the ontology to the possible semantic relation found in the tourists' question and extracted by the NLP component. There are two main procedures in the OEM:

-Generate Ontology Triple List: In order to retrieve the ontology triple list for the Semantic Relation which is extracted by the NLP, SPARQL-DL engine was used. In this step our semantic concern by using the SPARQL-DL is to retrieve the ontology triples from MTO according to the given Semantic Relation.

-Comparison: Extracted terms from the NLP are compared with the Classes and Instances of the ontology triples retrieved from the Semantic Relation. In the comparison process we used the Levenshtein distance [16] between terms and the Ontology Triples to find the minimum number of operations needed to transform one into the other. The range of score is from 0 to 1 where 0 is distantly similar and 1 is closely similar. The output from these processes is to generate the Semantic Ontology Triple (SOT) list.

c. Answer Recommendation

If there is more than one item in the SOT list the Answer Recommendation component will match up the attributes of the user profile with the SOT items to generate a recommended list. The list will contain the nearest matching items to the user's preference based on the content base filtering and subsequently form the Recommended Answer.

4.2. Design of the User Profile

The tourism User Profile is a structured representation of the tourists' preferences. We utilize a knowledge-based approach to implement the User Profile. In the knowledge-based approach we create a model of the user to closely match the users' needs in order for the SMTRS to build a recommendation. The users are asked to fill an online registration form when visiting the web site for the first time. Implementing explicit User Profile has the advantage of letting users specify directly their interests [17]. The information in the user profile is used to generate three types of recommendation: dining - recommends user where to eat, accommodation - recommends users where to stay and activity - recommends users what to do.

4.3. Design Malaysia Tourism Ontology (MTO)

Representing the Malaysian tourism information is a core component of our prototype. MTO is a formal and an explicit description of Malaysian tourism terms, written in OWL-DL to gain the required reasoning support. OWL-DL is based on formalizing the semantics in Description Logic.

The MTO was built based upon the Thesaurus on Tourism and Leisure Activities, published by the World Tourism Organization [18] as a specialized agency of the United Nations and an international standard for the sector of e-tourism. Furthermore, extra terms representing the speciality of Malaysian tourism were added consistently.

The general classes of MTO is shown in Figure 3

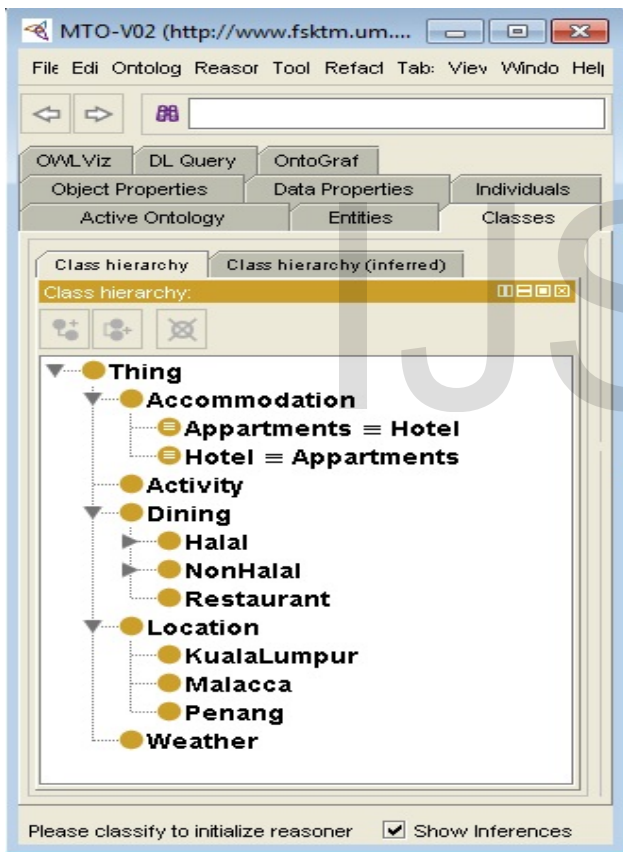


Figure 2 STO general classes

5. EVALUATION OF SMTRS

By evaluating the novelty presented in this work, we prove the significance and benefits of answering the

tourists' question semantically, and enriching the semantic answers with recommendations based on tourists' preferences. The design of the evaluation is as follows :

5.1. Correctness of MTO

MTO is an OWL-DL ontology built to represent the Malaysian tourism information. Pellet and Fact++ are the state of the art reasoners in ontology inference engines specializing in DL technology [19, 20]. Both reasoners were plugged-in with Protégé. Both support the inference services during the ontology development stages. The results used with the two reasoners showed that the concepts are consistent and the taxonomies are classified in MTO.

5.2. System-based Retrieval Performance Evaluation

In general, the aim for information retrieval systems is to optimise both precision and recall, this optimisation ensure the quality of the retrieval performance. Furthermore users are concerned about relevant retrieved answers measured by precision and recall as it determines the output of any information retrieval system [21-23]. We evaluated our retrieval performance approach using 16 sample questions obtained from Answers Yahoo [24]. These questions are reformulated to address the Malaysian cities and places without changing the structure of the original questions. Our questions are formulated taking into account that the classes and properties of the MTO are thoroughly covered.

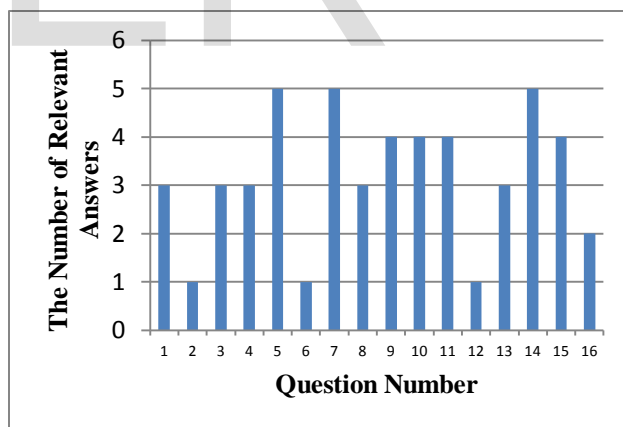


Figure 3 Histogram of number of relevant answers for each question in the sample set

SMTRS was able to find the relevant answers as shown in Figure 4.1 after processing the sample questions. Obviously, the number of relevant answers varies among the questions since the number of items (i.e. Hotel, Restaurant and Activity) varies in our MTO.

5-1 Precision and Recall for the sample Questions

| Question | Retrieved | Relevant | Relevant & Retrieved | Recall | Precision |
|----------|-----------|----------|----------------------|----------|---------------|
| 1 | 3 | 3 | 3 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 |
| 3 | 3 | 3 | 3 | 1 | 1 |
| 4 | 3 | 3 | 3 | 1 | 1 |
| 5 | 5 | 5 | 5 | 1 | 1 |
| 6 | 1 | 1 | 1 | 1 | 1 |
| 7 | 5 | 5 | 5 | 1 | 1 |
| 8 | 3 | 3 | 3 | 1 | 1 |
| 9 | 4 | 4 | 4 | 1 | 1 |
| 10 | 4 | 4 | 4 | 1 | 1 |
| 11 | 4 | 4 | 4 | 1 | 1 |
| 12 | 1 | 1 | 1 | 1 | 1 |
| 13 | 3 | 3 | 3 | 1 | 1 |
| 14 | 5 | 5 | 5 | 1 | 1 |
| 15 | 4 | 4 | 4 | 1 | 1 |
| 16 | 3 | 2 | 2 | 1 | 0.6667 |
| | | | | 1 | 0.9791 |

From the table the average precision is 97.91% and the average recall is 100%. The results show that SMTRS is capable of retrieving high relevant answers with high performance. Hence, this proves that SMTRS is competent to provide tourists, visiting Malaysia, with the relevant information to help them in planning their vacations efficiently.

5.3. User-based usability evaluation

SMTRS was built in order to allow tourists to get recommended information about tourism in Malaysia. In order to know if the system recommendation is used easily and effectively, we evaluated the system usability. In general the aim of measuring the usability of SMTRS is to evaluate the systems' core features specifically the recommended items from the user's point of view.

The methodology we used to perform the experiment was to observe users in a session of the system. Users were given a period of time with SMTRS and then asked to fill a questionnaire to express their views on the different features of SMTRS. The questionnaire used for evaluating SMTRS was driven from the System Usability Scale (SUS) [25] as SUS is one of the most popular questionnaires containing a standardized collection of

questions. Moreover specific questions were added to measure the users' satisfactions of the recommendation feature.

The result of the questionnaire is a value between 1 and 100, where 1 signifies that a user found a system absolutely useless and 100 that a user found a system optimally useful.

We chose a total of 15 participants to perform the test. All participants were international visitors who were familiar with tourism sites in Malaysia. Therefore, they were able to give us good feedback regarding the recommended answers without doubts as they have been in Malaysia for quite some time and knew the places the SMTRS recommended. Each participant was asked to fill the registration page first, to create a personal profile and navigate through the system features. Then, they were asked to formulate three questions in simple English and contain only one fact to find Hotel, Restaurant and Activity within Kuala Lumpur area. Finally, after reviewing the system recommendations and output the participants were given a questionnaire. The participants were asked to rate the system with a scale of 1 as strongly disagrees to 5 as strongly agree.

The results from the questionnaire about the usefulness of the SMTRS are shown in figure 4.5. The results indicate that the participants found that: the use of SMTRS is attractive, the system is easy to use and it provides the participants with information specially recommended for them.

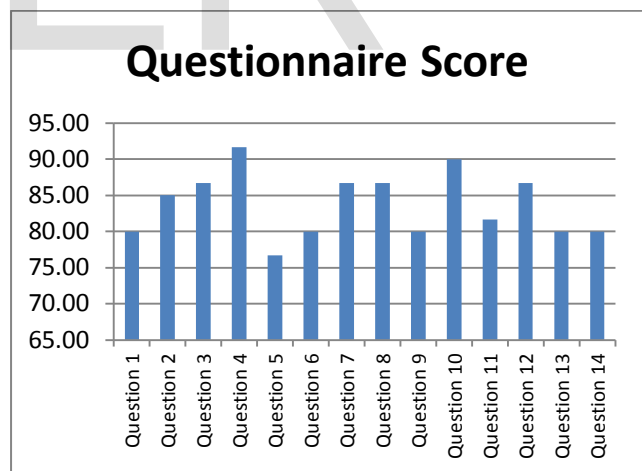


Figure 4 Questionnaire Results

As the score shows, users found SMTRS significantly suited to the required task. From a range of 0 to 100, users gives the SMTRS an average score of 83.69. The interpretation of the scores describing the acceptability of the system shows that the SMTRS is EXCELLENT [26].

SMTRS has overcome the problem of irrelevant information retrieval in the tourism systems

mentioned in section 1. The SMTRS retrieval information consists of recommended items based upon a description of the items and a profile of the user's interests. Our results show that these compositions are satisfactory to users. This is obvious as Q11 and Q12 gain the score between 80% and 90%. Hence we conclude that utilizing the user profile allows us to provide personalized information with particular interests to the user.

6. DISCUSSION

Literature reveals that the full-sentence query option was significantly preferred to any other method as discussed in section 2. The main purpose of this research was to get to the bottom of the tourism information finding process, to assist tourist with relevant information. To achieve this goal, the Semantic Malaysia Tourism Recommender System (SMTRS) was designed and developed. SMTRS adapts earlier techniques and include new techniques such as a Relation Search Algorithm (RSA). It analyzes the tourist full-sentence query by capturing the semantics between the terms in the query and converting it to semantic triples. Then it matches these triples with the MTO in order to retrieve answers. These answers are analyzed and prepared so they could be matched with the user profile. Finally, the system creates customized recommendation. This involves recommending unseen items that are similar to the preferences of the user. The similarity of items is calculated based on the features associated with the compared items. With this, we give the tourist the qualified information and the necessary recommendation to gain their satisfaction. Consequently, tourist satisfaction will lead to more tourism development.

This paper presented the evaluation of the SMTRS model as follows:-

- ✓ Results were reported in terms of recall and precision that the SMTRS is capable of retrieving high relevant answers with high performance.
- ✓ Results were reported in terms of building and using MTO as a fundamental component in SMTRS, MTO succeeded in representing the Malaysian tourism knowledge domain and retrieving the knowledge.

- ✓ Results were reported in terms of system usability, SMTRS achieved users satisfactions as it utilize the user profile in the information retrieval, and SMTRS provide personalized information with particular interest to the user.

These results demonstrate that SMTRS is applicable in the real-world scenario.

7. CONCLUSION

Searching for information on a desired spot for vacation is usually difficult and very time-consuming. Tourists are often overloaded by relevant and irrelevant information and resources. Providing personalized information with particular interest to the user is still a challenging task for the tourism information systems. SMTRS is an ontology-based application that provides semantic answers to the tourist full-sentence query, and support those answers with recommendation based on the tourists' preferences.

To prove the strength of our SMTRS, we evaluated the system by using a series of standardized models in the field. The results presented here, if adopted, will assist in providing customized answers to a particular tourist's query which matches their requirements. This makes the searching process more efficient.

REFERENCES

- [1] A. Armijos, et al., "IT trends in the lodging industry," *International Journal of Hospitality Information Technology*, vol. 2, pp. 1-17, 2002.
- [2] H. Werthner and F. Ricci, "E-commerce and tourism," *Commun. ACM*, vol. 47, pp. 101-105, 2004.
- [3] M. Sigala, et al., "ICT paradox lost? A stepwise DEA methodology to evaluate technology investments in tourism settings," *Journal of Travel Research*, vol. Journal of Travel Research, Nov 2004, Volume: 43 Issue: 2 pp.180-192, 2004.
- [4] F. Garzotto, et al., "Ubiquitous access to cultural tourism portals," in *Database and Expert Systems Applications, 2004. Proceedings. 15th International Workshop on*, 2004, pp. 67-72.
- [5] H. Werthner and S. Klein, *Information technology and tourism: a challenging relationship*. New York: Springer Verlag Wien, 1999.
- [6] W. V. Siricharoen, "Learning semantic web from e-tourism," presented at the *Proceedings of the 2nd KES International conference on Agent and multi-agent*

- systems: technologies and applications, Incheon, Korea, 2008.
- [7] A. Hinze and G. Buchanan, "Context-awareness in Mobile Tourist Information Systems: Challenges for User Interaction," in 'Proc. Workshop on Context in Mobile HCI, in conjunction with Mobile HCI', Salzburg, Austria., 2005.
- [8] P. Pu and C. Li, "User-Involved Preference Elicitation for Product Search and Recommender Systems," *AI Magazine*, vol. 29, pp. 93-103, 2009.
- [9] D. Damljanović and V. Devedžić, "Applying semantic web to e-tourism," in *The Semantic Web for Knowledge and Data Management: Technologies and Practices*, Z. Ma and H. Wang, Eds., ed IGI Global, New York, 2008, pp. 243-265.
- [10] C. Choi, et al., "Travel Ontology for Intelligent Recommendation System," presented at the Proceedings of the 2009 Third Asia International Conference on Modelling \& Simulation, 2009.
- [11] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," Stanford University, USA2001.
- [12] V. Tablan, et al., "A natural language query interface to structured information," presented at the Proceedings of the 5th European semantic web conference on The semantic web: research and applications, Tenerife, Canary Islands, Spain, 2008.
- [13] E. Kaufmann and A. Bernstein, "How useful are natural language interfaces to the semantic web for casual end-users?," presented at the Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference, Busan, Korea, 2007.
- [14] F. Ricci, et al., "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, F. Ricci, et al., Eds., ed: Springer US, 2011, pp. 1-35-35.
- [15] H. Cunningham, et al., "GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications," in *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, 2002.
- [16] *Algorithms and Theory of Computation Handbook*: CRC Press, Inc., 1998.
- [17] S. E. Middleton, et al., "Ontological User Profiling in Recommender Systems," *ACM Transactions on Information Systems*, vol. 22, pp. 54-88, 2004.
- [18] WTO, *Thesaurus on Tourism & Leisure Activities*: World Trade Organization 2001.
- [19] B. Parsia and E. Sirin, "Pellet: An owl dl reasoner," in *Proceedings of the International Workshop on Description Logics(DL2004)*, Whistler, BC, Canada, 2005.
- [20] D. Tsarkov and I. Horrocks, "FaCT++ Description Logic Reasoner: System Description." vol. 4130, U. Furbach and N. Shankar, Eds., ed: Springer Berlin / Heidelberg, 2006, pp. 292-297.
- [21] T. Mandl, "Recent Developments in the Evaluation of Information Retrieval Systems: Moving Towards Diversity and Practical Relevance " *Informatica*, vol. 32, pp. 27-38, 2008.
- [22] C. D. Manning, et al., *Introduction to Information Retrieval*: Cambridge University Press, 2008.
- [23] K. Zuva and T. Zuva, "Evaluation of Information Retrieval Systems," *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 4, 2012.
- [24] Yahoo. (2012, 20/September/2012). Answer Yahoo. Available: <http://answers.yahoo.com/>
- [25] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability evaluation in industry*, P. W. Jordan, et al., Eds., ed: Taylor and Francis, 1996.
- [26] A. Bangor, et al., "An Empirical Evaluation of the System Usability Scale," *International Journal of Human-Computer Interaction*, vol. 24, pp. 574-594, 2008/07/29 2008.