4.1 PREAMBLE

A specific semantic search is designed for the users who have domain knowledge with regard to their search queries, in order to help them to quickly retrieve the required data from the OKB. The meta data are marked with ontology markup languages. It is possible to use semantic web rule language and SPARQL based queries which retrieve details by querying their attributes. This helps the search requester to get the query meaning or disambiguation process and directly search metadata from the OKB. The search style is based on the premise that such user has some knowledge about a metadata. As a consequence the chapter proposes a specific semantic search technique and it is more suitable for tourism domain user.

Tourism has become a popular and global leisure activity. It has experienced consistent development over the recent years. The World Tourism Organization expects that by 2020 (Maswera et al 2009 & Kabassi 2010) tourism industry arrivals around the world would increase over 200%. Today, the WWW is the major source of tourist information for tourist requester. It has significantly influenced the way of information gathering, findings and exchanging in the area of tourism. However, manually planning for a trip, with such a huge amount of disparate information and services to suit one’s preferences it is still quite difficult and time-consuming. Neither is the automation of such tasks easier nor the semantic web and ontologies are good
candidates to access reliable and accurate information in a fraction of time from the web.

The internet information has increased dramatically such that it is very difficult for the user to find what they are looking for. The appropriate finding of tourism services is one of the main drawbacks when people seek information on the web, because different users have different preferences and needs. User perception plays an important role in semantic search and many proposals ignore this issue. In order to help this research work specified answer to this issue and utilized Bayesian network and Netica-J for calculating user preferred activities.

The rest of the chapter is structured as follows: section 4.2 defines the problem formulation and objectives. Section 4.3 explains the workflow of the proposed specific semantic search technique. Section 4.4 evaluates the performance analysis and discussion and section 4.5 presents summary of the chapter.

4.2 PROBLEM FORMULATION AND OBJECTIVES

The proposed work is concerned with the development of a technique towards the specific semantic search with the consideration of user’s preferences/profile. The technique has suggested the result based on user’s priority while searching the tourism domain of requester interest. Bayesian network is used to calculate the conditional probability of the given input based on user preferences/profile. The suggestion is that the system tells the probability of their interest on the type of places they prefer to tour. The user’s profile includes the age, the purpose for which the tour is being made, the user’s current status and the type of category he belongs to which indicate the behavior of the user. Apart from suggesting the tourism domain which suits him to search for, it also gives the complete details of tourist interest relating
to each of the tourism domain which includes nature, culture, entertainment and outdoor in which to search for, which consumes more time in the syntactic search system.

Hence, it is expected that the proposed specific semantic search technique yields better results than the existing approaches. The objectives are to quickly retrieve the required information and provide more accurate results to satisfy the needs of user in a fraction of time.

4.3 PROPOSED SPECIFIC SEMANTIC SEARCH TECHNIQUE

The specific semantic search technique of the semantic search system is highlighted in Figure. 4.1. The technique consists of three components, which are ontology knowledge base, Bayesian network approach and semantic query.

![Figure 4.1 Specific semantic search technique](image)
The ontology knowledge base is to store domain specific tourism ontologies and metadata base. The Bayesian network is used to estimate the users preferred activities and semantic query to retrieve the most accurate results in minimal time. The proposed work contributes the best results in time.

### 4.3.1 Process Flow Diagram

The process flow diagram is shown as in Figure 4.2. The user gives input as user profile/preferences, which is fed into Bayesian Network to find the probability of the user’s preference in each type of tourism (Nature, Culture, Entertainment, and Outdoor). Then it is retrieved by using semantic query from ontology knowledge base and provides relevant results to users.

![Figure 4.2 Process flow diagram](image)


4.3.2 Proposed Specific Semantic Search Method

The proposed method employs the following steps as shown in figure 4.3 and its elaborations in the following sections.

<table>
<thead>
<tr>
<th>Step 1: Ontology Knowledge Base</th>
</tr>
</thead>
<tbody>
<tr>
<td>( TourOKB = (C,P,I) )</td>
</tr>
<tr>
<td>( C = {c_{ti}</td>
</tr>
<tr>
<td>( P = {(p_{ti}, f(p_{ti}))</td>
</tr>
<tr>
<td>( I = {(i_{ti}, f(i))</td>
</tr>
</tbody>
</table>

| Step 2: (i) Find the highest probability value \( P(N|C) \), N is the net, C is the concept using Baye’s rule |
|\( P(N|C) = P(C|N)P(N)/P(C) \) |

(ii) Trying to and maximum of \( P(C|N)P(N) \) which is the same as maximizing the problem logarithm

\[ \log(P(C|N)) + \log(P(N)) \]

(iii) \( \log(P(N)) \) is the prior probability of net

(iv) \( \log(P(C|N)) \), C consists n independent cases \( c_1, c_2, c_3 ... c_n \)

the log possibility is

\[ \log(P(C|N)) = \log(P(c_1|N)P(c_2|N)) ... P(c_n|N)) \]

\[ = \log(P(c_1|N) + P(c_2|N)) + ... + P(c_n|N)) \]

\( \log(P(c_i|N)) \) term is easy to calculate, Netica inference is used to determine probability findings.

<table>
<thead>
<tr>
<th>Step 3: SPARQL and SWRL based querying.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 4: Metadata retrieval.</td>
</tr>
</tbody>
</table>

Figure 4.3 Proposed specific semantic search method
4.3.3 Ontology Knowledge Base for Tourism domain

Ontology is a main part in integrating heterogeneous information and is built by ontology web language (OWL). OWL has class, instances, and property as three main components in ontology. A class is used to represent a concepts defined in the ontology. Classes are typically arranged in taxonomy as classes and subclasses. A class may associate with a set of instances. A property is used to represent the concepts and attributes; including the relationships with other concepts. The ontology used in the tourism semantic search provides an understanding of the domain data set and all information is semantically linked and related, so that the user can search for relationship in information.

Definition of tourism ontology knowledge base: Tourism ontology knowledge base consists classes, properties and instances related to tourism data. It can be represented as follows:

\[ \text{TourOKB} = (C, P, I) \]

\[ C = \{ c_{ti} | c_{ti} \in \text{owl:Class} \}, \]

\[ P = \{ (p_{ti}, f(p_{ti})) | p_{ti} \in \text{rdf:Property}, f(p_{ti}) \in \mathcal{R}^+ \}, \]

\[ I = \{ (i_{ti}, f(i)) | i_{ti} \in c_{ti}, f(i_{ti}) \in \mathcal{R}^+ \}, \]

where \( C \) denotes the set of tourism related concepts, \( P \) denotes the set of tourism related properties and \( I \) denotes the instances set of all tourist concepts. \( c_{ti} \in \text{owl:Class} \) denotes a concept, \( p_{ti} \in \text{rdf:Property} \), denotes a property and \( i_{ti} \in c_{ti}, \) instance of tourist concepts. \( f(v) \) denotes a function for measuring the property and instances values. The function calculates the value of specificity. OWL has two types of properties, one is object property which shows the relationships between instances of two classes. The other is data type property which shows relationship between class instances with RDF literals and XML schema data types.
The tourism ontology provides a way of viewing the world tourism information. It organizes tourism related information and concepts. Tourism ontology is explicitly defined to share a contextual conceptualization of the tourism domain. This allows the users to make their resources more machines processable by collaboratively constructing an enriched layer of the semantic web that links tourism artifacts with formal semantics to support semantic query.

The ontologies are useful in improving the accuracy of web searches. The tourism ontology knowledge base has two main parts, which are domain ontology base and metadata base. It has two rules contained in the semantic relationship; the first one is a concept which may semantically relate to arbitrary metadata and the second one is a metadata which may also semantically relate to arbitrary concepts.

The aim of ontology knowledge base is to store the domain specific service ontology relevant to tourist attractions. The domain ontology is a multi-layer service concept hierarchy. The first layer is the root of the hierarchy, which represents the abstract concepts of all service domains. The next layer is the preliminary specialization of the abstract service concepts, which represents the service sub domain concepts. As in children concepts, all other concepts in this hierarchy automatically inherit its properties.

In this research, the online information of tourist domain is the source to derive concepts and relationships between these concepts in the tourist ontology. The background knowledge for the ontology is obtained from Linked Open Vocabulary (LOV) (http://lov.okfn.org/dataset/lov/), QALLME dataset (qallme.fbk.eu/index.php/location=ontology) and the following web sites (www.tamilnadutourism.org) (http://www.indianyellowpages.com/tour-services) which consists of many tourist concepts. Each
concept is defined by a certain number of concept descriptions, and the concept is related to the sub concept relationship. Concepts and relationships are basic components in ontology found that the main investigated concepts in this ontology are attractions, location, age, occupation, accommodation, shopping, places, transportation, seasons, foodstuff etc., in which each concept may have sub concepts.

For example, the age category has three sub concepts: youth, adult, and senior. The accommodation category has sub concepts as hotel and others. The sample tourism ontology screen shot is illustrated in Figure 4.4. Next Metadata Base is designed to store the estimated probability data and metadata information.

![Figure 4.4 Tourism ontology screen shot](image_url)
4.3.4 A Bayesian network to measure a user’s preferred activities

Bayesian network is employed to estimate the preferred activities in the semantic search of tourism domain (Baeza-Yates & Ribeiro-Neto 2011). The user wishes to get a tour plan cover in several aspects, such as preferred activities engaged in a tourist attraction or preferred transportation model etc.. Here this research work focuses on the measurement of the user preferred activities. The thought is that the preferences of activities perform an important grace in the choice of visiting tourist attractions.

The main aim is to estimate a user’s preferred activity of visiting tourist attractions in a Bayesian network based on that person and other users’ user behaviors. Three steps, namely qualitative, quantitative, and updating steps are carried out to complete this task. The qualitative step signifies the qualitative information of the Bayesian network to measure the user’s preferable activities, including a graphic depiction of variables and correlations between variables. The quantitative step defines the quantitative information of the Bayesian network, which includes denoting a probability distribution for root nodes without parent nodes and a conditional probability table for others. The last step updates the probability distribution of preferred activities served the proof of other variables within the Bayesian network. The consequence with the highest probability considers the best preferred activities for the user. Among these three steps, the first and second step depends on the OKB in the tourism domain, and the third step follows the Bays theorem.

4.3.4.1 Qualitative step

The attributes in a Bayesian network consists of a set of factors influencing the preferred activities of the user. In the tourist domain, a number of fictions can be investigated to identify the factors affecting preferred
activity choice of theoretical and experimental study. This research has considered a set of factors that are commonly mentioned by researchers in the tourist domain and are listed as below.

**Tour motivation**: This attributes clarifies the analysis that a rover visits a destination, and it is the most important factor influence the choice of preferred activities. The most common values for this variable, when visiting a city are, enjoying nature, learning something, seeking novelty, have a rest, visiting interesting places, historical places, famous places, business, study, visiting friends or families, attending workshops or conferences, and others.

**Journeymers’ type**: Journeymers’ type is also a powerful and common way to estimate preferred activities. Journeymers’ are often grouped based on their demographics and psychographic characteristics in different ways. In this research, four user types are used and they are: adventurer, multifarious, relaxation seeker, and urban. The adventurer prefers outdoor adventure and general sightseeing activities. The multifarious likes diverse activities. The relaxation seeker seeks amusement, relaxation, and general sightseeing activity. The urban prefers performing arts, local events etc..

**Age**: This attribute is a discriminating demographic factor that affects the preferred activity choice. Different age groups have different preferred activities. For example, older tourists have a high interest in culture, arts and temples. For this variable, three common age ranges are used: youth, adult, and senior.

**Occupation**: The preferred activities could be varied because of journeymers occupation. For this attribute, three widely used occupation labels were considered: employed, business and unemployed.
**Personality Type:** As a key psychographic factor, this variable is influential in shaping the choice of general preferred activities. Personality can be defined as ‘the reflection of a person enduring unique characteristics that urge him/her to respond in persistent ways in recurring environmental stimuli. For this attribute, triple common labels are considered and they are type A, type B and type C. For example, Type_A: can be Professional, Scientific, educational and Technical Services. Type_B can be Health, Social Assistance, Entertainment and recreation. Type_C can be Management of Companies, Enterprises, and Administrative Services.

**Preferred activities:** There are many types of preferred activities organized in different levels. For simplicity, this research only considers the activities at the top level, including five types of activities: natural-based sightseeing, cultural-based sightseeing, outdoor activities, relaxation activities, and local events. The assumption is that the activities of similar type have the same probability of user preferences.

After the definition of relevant variables, the relationships between the variables is established and presented in Figure 4.5. The attributes, age, occupation, personality type, and tour motivation are root nodes. The variables such as age, occupation, and personalities, influence the type of user, which subsequently combine with the motivation of tour to influence the preferred activities. In other words, if the probability distribution of any of the above three variables which are age, occupation, and tour motivation is changed, then it reflects on the probability of the type of user, and subsequently, the probability distribution of the preferred activities is changed. If the probability distribution of the tour motivation is changed, but others remain the same, the probability distribution of the preferred activities is also changed.
4.3.4.2 Quantitative step

In this step, the probability distributions of the four root variables such as age, occupation, personalities, and tour motivation and the conditional probability tables for other two variables namely user type and preferred activities are assigned. For example, the conditional probability table for the variable user type is assigned, based on the survey of user behavior. However, rather than doing an analysis by ourselves, summarized the empirical data by reviewing the existing research in the literature in user domain.

The conditional probability table for the preferred activities is assigned in the same way, through the data for a user is limited at the beginning. When more and more interests are made for this user and the feedback are obtained from that person, the conditional probability table can be reassigned, based on that user’s user behaviour. Second, the probability distributions of the two variables, namely user type and preferred activities are calculated based on the results in the first step using Bays theorem. Figure 4.5 shows probability distributions of these variables.

Figure 4.5 Bayesian network
4.3.4.3 Updating step

In this step, the probability distribution of the preferred activities is updated using Bayes theorem. The evidence of the four variables: age, occupation, personality, tour motivation are considered. The updating step involves two stages. Firstly, the user type is updated by analyzing the first three variables such as age, occupation and personality. Second, the preferred activity is updated by taking the last variable, namely tour motivation with the updated user type. Thus, a user’s preferred activity is estimated based on that person’s personal information, (age, occupation, personality) and tour characteristics such as tour motivation.

Figure 4.6 shows the estimation result of the preferred activity of a person for example (age = youth, occupation = employed, personality = type A, tour motivation = seeking) using Netica–Norsys Software Corporation, in this case, the probability distribution value of the journeyers motivation is nature = 0.118990384, culture = 0.7103365, outdoor = 0.056490384, entertainment = 0.11412696.
This estimation combines factors from both users’ socio-demographic characteristics and tour characteristics. The preferences are generated based on both the user behavior of the user and of other users. This provides a clear insight into the estimation of a user’s, user preferences. Each user destination is different. In order to endorse tourist attractions at a specific destination in the semantically enabled find for tour planning, a deal way is to assign the conditional probability table destination-specific, which is based on surveys in this destination. This is followed by the corresponding results retrieved from OKB using semantic query based on the high probability value.

### 4.3.5 Semantic Query

In this research, SPARQL query is being used to retrieve tourism relevant information from tourism OKB. SPARQL (Simple Protocol And RDF Query Language) is similar to SQL and used to access more reliable and accurate results. For example, in this query the region where the requested item to shopping is being present and is retrieved along with the place of shopping is filtered in the requested region as this sample code is shown in Figure 4.7.

```
PREFIX table: <http://www.owi-ontologies.com/Ontology1311200945.owl#> 
"" "SELECT ?place " +  "FROM <http://www.owi-ontologies.com/tour1.owl>" +  "WHERE ( " +  "?place table:belongsToRegion ?region Filter (?region = table:"+region+"))" +  "
  ) 

PREFIX table: <http://www.owi-ontologies.com/Ontology1311200945.owl#> 
```

Figure 4.7 Sample SPARQL code1
Another example for querying is to search for the hotel with the desired facilities such as laundry, allowing pets, etc. along with the type of food preferred such as Indian, Chinese etc., in the desired region of the visit as this sample code is shown in Figure 4.8.

```sparql
PREFIX table: <http://www.owl-ontologies.com/Ontology1325058945.owl#> " +
"SELECT ?shops ?location " +
"FROM <http://www.owl-ontologies.com/tour1.owl>" +
"WHERE {" +
"  ?shops table:isInRegion ?obj filter (?obj = table:country)"+" +
"  ?shops table:sells ?shopping filter (?shopping = table:shopping)" +
"  ?shops ?cond2 ?location filter(?cond2 = table:location)" +
" } .";
```

Figure 4.8 Sample SPARQL code2

4.4 PERFORMANCE ANALYSIS AND DISCUSSION

The proposed specific semantic search technique is built with JAVA, incorporated with Bayesian Network, Netica-J, JSP, Google Maps, Protege, Jena, SWRL, SPARQL with Intel Xeon 3.00-GHZ CPU, 3.59-GB RAM, and JAVA SDK 1.6. To test the performance using tourism data sets are collected from Linked Open Vocabulary (LOV), QALLME and various web sites.

Figure 4.9 shows the sample screen shot, the user selects the age category, occupation category, personality type and motivation as user preference, which is fed into Bayesian Network to find the probability of the user in each type of tourism (Nature, Culture, Entertainment, and Outdoor). The region of visit, products interested in buying and accommodation if needed, the user enters facility, he is in need of, food habits etc.. The most
preferable places belonging to that region and in that particular tourism category are suggested along with shopping and hotel details. Figure 4.10 shows, when the location map button is clicked the required location of the tourist places are displayed and outputs its route along with directions to reach it in a map and so on. This is retrieved by using a semantic query. The semantic query provides more accurate search results for the tourism user. In all the inputs results are averaged by 100%.

![Figure 4.9 Query Interface in specific search](image)

![Figure 4.10 Location map](image)
System Evaluation Results: To evaluate the performance of specific semantic search technique, four performance indicators are used, namely precision, mean average precision, recall and f-measure. The optimal threshold value set 0 to 1 with an increment of 0.1 and it is used to filter unrelated data. Table 4.1 presents test results of the specific semantic search technique. It is observed that along with the increase of the threshold value the precision results show a sharp rise in value from 22.82% to 92.72% of the threshold value 0 to 0.9. Mean average precision tests for the quickness and precision of a search result. The mean average precision values from 72.99% to 95.45% of the threshold value 0 to 0.9. In contrast, recall range from 77.49% to 42.95%. F-measure value ranges at a peak of 60.41% of the threshold value in 0.5. From the results, the proposed work is getting a high level of search performance.

Table 4.1 Testing results of specific semantic search

<table>
<thead>
<tr>
<th>Optimal Threshold Value</th>
<th>Precision %</th>
<th>Mean Average Precision %</th>
<th>Recall %</th>
<th>F-Measure %</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0</td>
<td>22.82</td>
<td>72.99</td>
<td>77.49</td>
<td>35.25</td>
</tr>
<tr>
<td>&gt;0.1</td>
<td>27.78</td>
<td>73.92</td>
<td>73.21</td>
<td>40.27</td>
</tr>
<tr>
<td>&gt;0.2</td>
<td>34.11</td>
<td>75.31</td>
<td>72.27</td>
<td>46.34</td>
</tr>
<tr>
<td>&gt;0.3</td>
<td>43.32</td>
<td>78.42</td>
<td>67.39</td>
<td>52.73</td>
</tr>
<tr>
<td>&gt;0.4</td>
<td>52.28</td>
<td>81.28</td>
<td>60.38</td>
<td>56.03</td>
</tr>
<tr>
<td>&gt;0.5</td>
<td>62.59</td>
<td>83.43</td>
<td>58.39</td>
<td>60.41</td>
</tr>
<tr>
<td>&gt;0.6</td>
<td>72.74</td>
<td>85.67</td>
<td>47.31</td>
<td>57.33</td>
</tr>
<tr>
<td>&gt;0.7</td>
<td>83.46</td>
<td>87.62</td>
<td>44.37</td>
<td>57.93</td>
</tr>
<tr>
<td>&gt;0.8</td>
<td>87.17</td>
<td>90.57</td>
<td>43.54</td>
<td>58.07</td>
</tr>
<tr>
<td>&gt;0.9</td>
<td>92.72</td>
<td>95.45</td>
<td>42.95</td>
<td>58.70</td>
</tr>
</tbody>
</table>
**Graph Results:** Figure 4.11 shows the performance results of a specific semantic search on precision with a variation of threshold values from 0 to 0.9 at an interval of 0.1. The precision value increases when the threshold value rises. The precision value increased from 22.82% to 92.72% at the threshold value 0 to 0.9. The highest score is obtained at the threshold value of 0.9. This experiment proves that the search technique works precisely.

![Threshold Values & Precision %](image)

**Figure 4.11 Performance result on precision**
Figure 4.12 shows the performance result of specific semantic search on mean average precision with a variation of threshold values from 0 to 0.9 at the interval of 0.1. The mean average ranges from 72.99% to 95.45%. From the result, it is proved that the search technique can be employed quickly and precisely.

Figure 4.12 Performance result on mean average precision
Figure 4.13 shows the performance result of specific semantic search on recall with a variation of threshold values from 0 to 0.9 at the interval of 0.1. The recall ranges from 77.49% to 42.95%. When the threshold value increases the recall value automatically decreases. This proves the effectiveness of the proposed system.
Figure 4.14 shows the performance results of specific semantic search on f-measure with a variation of threshold values from 0 to 0.9 at the interval of 0.1. F-measure combines the precision and recall values as an aggregated performance scale. The highest f-measure obtained at the threshold value 0.5, the value is 60.41%.

![Threshold Values & F-Measure %](image)

**Figure 4.14 Performance result on F-Measure**

Next, “Time” is an important factor in specific semantic search technique. The query processing retrieval time is compared with Yoo (2012). The response time for different queries to retrieve the relevant results by comparing the research work and Yoo (2012) work is shown in Figure 4.15. The Yoo approach has three methods, namely; Yoo Method1 (querying for inferring knowledge base), Yoo Method2 (querying for knowledge base) and Yoo Method 3 (hybrid inferred knowledge base). The three methods have similar components and different function. Yoo Method1 and Yoo method2 using SWRL domain rules for transferred user requirements and Yoo Method3 used individual rules templates to transferred user requirements and this
research proposed semantic query, Bayesian technique with inferred knowledge base as method 4.

Different input queries are made in the performance test using semantic query like SPARQL and SWRL. For example: query 1. User like to find pet allowed hotel, 2. User like to find hotel located near the university for attending seminar presentations, 3. User likes to search a reasonably priced hotel that offers various services like kids’ club, outdoor swimming pool, WiFi service in the room, and to drive extensively etc.. for these queries generate rule templates as

Example 1: Cheap hotel concepts: if it cost less than? Variable dollars per night \(\rightarrow\) then this hotel is a economic hotel.

Example 2: Nearby university or institute for these concepts the rule generates as: if it takes less than? Variable minutes walk or drive between sites \(\rightarrow\) then these sites are nearby.

All These methods and queries will bring in different results with varying time.

The response time for different queries to retrieve the relevant results by comparing the research work and Yoo (2012) work is shown in Figure 4.15.
The existing method 1, which is faster than method 2 and method 3, is said to retrieve irrelevant information to the user requirement quickly by using inferred knowledge base function. The existing method 2, is slower than other four methods, retrieves the latest information to the user requirements. This method is based on new inferred knowledge into the knowledge base to the user requirements. The existing hybrid method 3, is slower than method 1. This method is based on query rewriting and reasoning technology. The proposed method 4, (specific semantic search) is faster than the other three methods and retrieve results more relevant, accurate. This method is based on querying for inferring knowledge base with Bayesian technique.
**Discussion:** Table 4.1 and figure 4.11, 4.12, 4.13 and 4.14 presents performance testing results of specific semantic search. Figure 4.15 shows time comparison of our proposed method and existing work of Yoo (2012). The efficiency of the proposed methodology is confirmed from the testing results. Further, it is also analyzed and verified to show better performance, compared to the existing methods. Furthermore the recall and f-measure result is more enhanced in the next chapter.

### 4.5 SUMMARY OF THE CHAPTER

In this research work, an efficient and a reliable specific semantic search technique is presented. The proposed work, matches user request in the ontology and retrieves the appropriate record of query. The method also gives the sorted order of the tourist domain in which the user can search which makes their search and retrieval with a minimum time of the required record easy. The order of tourism domain is being based on the user preference/profile. There is also an effective plan of accommodation accompanied with food, shopping, special care for handicaps, parking facility, care for pet animals, etc.. This system suggests the spatial and temporal distance to the tourist. It also enables more convenient planning towards the goal of user destination by applying an ontology theory, Bayesian network and semantic query. It is confirmed from the experimental results, the efficiency of the proposed method is far outstanding than the existing method. The next chapter proposes QoS based ranking methodology to enhance semantic search performance.