THE DESIGN OF AN ANALYSIS MODULE FOR ONLINE TRAVEL ACCOMMODATION REVIEWS

Phichayasini Kitwatthanathawon

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Information Science in Information Technology

Suranaree University of Technology

Academic Year 2013

การออกแบบมอดูลสำหรับวิเคราะห์บทวิจารณ์ออนไลน์
เกี่ยวกับสถานที่พักเพื่อการท่องเที่ยว

นางสาวพิชญ์สินี กิจวัฒนาถาวร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาการสารสนเทศดุษฎีบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

มหาวิทยาลัยเทคโนโลยีสุรนารี

ปีการศึกษา 2556

THE DESIGN OF AN ANALYSIS MODULE FOR ONLINE TRAVEL ACCOMMODATION REVIEWS

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Information Science in Information Technology.

Thesis Examining Committee

____

(Assoc. Prof. Dr. Weerapong Polnigongit)

Chairperson

____

(Dr. Jitimon Angskun)

Member (Thesis Advisor)

____

(Assoc. Prof. Dr. Punpiti Piamsa-nga)

Member

____

(Assoc. Prof. Dr. Nittaya Kerdprasop)

Member

____

(Dr. Suphakit Niwattanakul)

Member

____

(Dr. Thara Angskun)

Member

____

(Prof. Dr. Sukit Limpijumnong)

Vice Rector for Academic Affairs and Innovation

____

(Dr. Peerasak Siriyothin)

Dean of Institute of Social Technology

พิชญสินี กิจวัฒนาถาวร : การออกแบบมอดูลสำหรับวิเคราะห์บทวิจารณ์ออนไลน์เกี่ยวกับสถานที่พักเพื่อการท่องเที่ยว (THE DESIGN OF AN ANALYSIS MODULE FOR ONLINE TRAVEL ACCOMMODATION REVIEWS) อาจารย์ที่ปรึกษา : อาจารย์ ดร. จิติมนต์ อั่งสกุล, 263 หน้า.

ปัจจุบัน นักท่องเที่ยวส่วนใหญ่ใช้อินเทอร์เน็ตในการสืบค้นข้อมูลเพื่อนำไปใช้ประกอบการตัดสินใจเลือกแหล่งท่องเที่ยวให้ได้ตรงกับความต้องการของตนเอง ซึ่งวิธีที่นิยมใช้กันโดยทั่วไปคือ การตัดสินใจโดยใช้ข้อมูลจากบทวิจารณ์ของนักท่องเที่ยวที่เคยไปมาแล้ว แต่อย่างไรก็ตาม การเลือกแหล่งท่องเที่ยวให้ได้ตรงกับความต้องการนั้น นักท่องเที่ยวต้องอ่านบทวิจารณ์เป็นจำนวนมาก และจากการสำรวจมอดูลสำหรับการเขียนบทวิจารณ์ที่มีอยู่ในปัจจุบันพบว่า มอดูลดังกล่าวยังไม่สามารถสกัดความสัมพันธ์ของคุณลักษณะของสถานที่พักออกมาได้ และมีการประเมินผลความพึงพอใจเพียง 2 ระดับเท่านั้น คือ ดี หรือควรปรับปรุง ซึ่งต่างจากการประเมินผลของนักท่องที่ยวซึ่งมี 5 ระดับ ได้แก่ ดีมาก ดี ปานกลาง ควรปรับปรุง และต้องปรับปรุง

ในงานวิจัยนี้ได้ออกแบบมอดูลวิเคราะห์บทวิจารณ์ออนไลน์เกี่ยวกับสถานที่พักในการท่องเที่ยว โดยมอดูลที่ออกแบบนี้ได้นำเทคนิคต่าง ๆ ได้แก่ ออนโทโลยี มาใช้เป็นฐานความรู้ในการสกัดและจัดเก็บความรู้ และเทคนิคการวิเคราะห์เชิงความหมาย มาใช้ในการสกัดคุณลักษณะที่สำคัญของสถานที่พักออกจากบทวิจารณ์ออนไลน์ รวมทั้งได้พัฒนาวิธีการคำนวณระดับคะแนนความพึงพอใจของนักท่องเที่ยวที่มีต่อบริการและ/หรือสิ่งอำนวยความสะดวกต่าง ๆ ของสถานที่พักนั้นด้วยเทคนิคตรรกศาสตร์คลุมเครือ และในการแสดงผลความรู้ที่สกัดได้นั้น มอดูลแสดงให้เห็นถึงความสัมพันธ์ของคุณลักษณะของสถานที่พักในรูปแบบของแผนภาพโครงสร้างต้นไม้ พร้อมทั้งแสดงจำนวนการวิจารณ์และระดับคะแนนความพึงพอใจที่แตกต่างกัน 5 ระดับของการวิจารณ์แต่ละครั้งอย่างละเอียด

ในการประเมินมอดูลที่ได้ออกแบบนั้น ข้อมูลที่นำมาทดสอบได้จากการสุ่มเก็บบทวิจารณ์จำนวน 200 บทวิจารณ์จากโรงแรมต่าง ๆ ซึ่งอยู่ในระดับความพึงพอใจตามที่ผู้ใช้ให้คะแนนแตกต่างกัน 5 ระดับ จำนวนระดับละ 40 บทวิจารณ์ โดยการประเมินมอดูลและการวิเคราะห์ผลการทดลองจะแบ่งออกเป็น 2 ส่วนตามกระบวนการของส่วนอนุมานความรู้ ได้แก่ กระบวนการสกัดคุณลักษณะที่สำคัญของสถานที่พัก และกระบวนการคำนวณระดับคะแนนของสถานที่พัก ผลการทดลองพบว่า กระบวนการสกัดคุณลักษณะที่สำคัญของสถานที่พักได้รับค่าความถูกต้องโดยรวมร้อยละ 79.22 ค่าความแม่นยำโดยรวมร้อยละ 100 และค่าความระลึกโดยรวมร้อยละ 76.05 ส่วนในการประเมินกระบวนการคำนวณระดับคะแนนของสถานที่พัก พบว่า ค่าการคำนวณระดับคะแนน

ของการวิจารณ์โดยรวมมีความแตกต่างจากระดับคะแนนที่นักท่องเที่ยวกำหนดเท่ากับ 0.378 คะแนน โดยได้รับค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเท่ากับ 0.489 ทั้งนี้ผู้ประกอบธุรกิจการท่องเที่ยวอิเล็กทรอนิกส์สามารถนำมอดูลที่ถูกออกแบบขึ้นมานี้ไปประยุกต์ใช้ในการสกัดและค้นคืนความรู้ที่ได้จากบทวิจารณ์ของนักท่องเที่ยว แล้วนำความรู้เหล่านั้นไปพัฒนาสินค้าหรือบริการของตนเอง เพื่อสามารถตอบสนองความต้องการของนักท่องเที่ยวให้ได้มากที่สุด

สาขาวิชาเทคโนโลยีสารสนเทศ	ลายมือชื่อนักศึกษา__________________________
ปีการศึกษา 2556	ลายมือชื่ออาจารย์ที่ปรึกษา____________________
	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม_________________

PHICHAYASINI KITWATTHANATHAWON : THE DESIGN OF AN ANALYSIS MODULE FOR ONLINE TRAVEL ACCOMMODATION REVIEWS. THESIS ADVISOR : JITIMON ANGSKUN, D.ENG., 263 PP.

REVIEW ANALYSIS MODULE/FEATURE EXTRACTION/RATING/FUZZY

Currently, most tourists use the Internet to retrieve information for supporting their decision in selecting the tourist places that conform to their preferences. The most common method is the decision based on reviews of experienced tourists. However, tourists must read enormous reviews in order to select their preferred tourist places. According to a survey of existing review analysis modules, the results show that those modules could not extract relationships among the accommodation features. Additionally, the satisfaction evaluation is illustrated in only two levels that are good or bad whereas the experienced tourists evaluate the accommodation features in 5 levels that are excellent, good, average, bad, or poor.

This research designs an analysis module for online travel accommodation reviews. The analysis module combines several techniques, such as using ontology as a knowledge base for knowledge extraction and storage, and using the semantic analysis technique to solve the feature extraction problems. In addition, this module provides a fuzzy-based method for calculating a tourists' satisfaction level with accommodation services and facilities. In order to present extracted knowledge, this module illustrates the relationships of features in the form of hierarchy diagram, the number of criticisms, and the tourists' satisfaction level with 5-rating scale of each criticism thoroughly.

In order to evaluate this designed module, a new dataset of 200 reviews were randomly selected from several accommodations. It covers all 5 satisfaction levels (40 reviews in each level). The module evaluation and experimental result analysis are divided into two parts that are a feature extraction process and an accommodation rating process according to the processes of knowledge inference engine. The experimental results of the feature extraction process are achieved in 79.22% of overall accuracy, 100% of overall precision, and 76.05% of overall recall. Moreover, the evaluation of accommodation rating process reveals that overall review points calculated by the module are 0.378 different from those specified by tourists while the root mean square error (*RMSE*) is 0.489. Thus, e-tourism operators can apply this designed module to extract and retrieve knowledge from online tourist reviews, and then the knowledge is applied to develop their products or services for satisfying the needs of tourists as much as possible.

School of Information Technology Student's Signature ____________________

Academic Year 2013 Advisor's Signature ____________________

Co-advisor's Signature ____________________

ACKNOWLEDGEMENTS

There are many people to whom I would like to express my sincere gratitude for their assistance, advice and encouragement which I have received in completing this dissertation.

First, my deepest respect and sincere gratitude belong to my research advisor, Dr. Jitimon Angskun and Dr. Thara Angskun. I am very grateful for their supervision, useful comments, kind assistance, continuous encouragement, and understanding given to me. Without them, my research work would not have been completed.

I would like to extend my gratitude to Assoc. Prof. Dr. Weerapong Polnigongit, the chairman of my dissertation committee, for his valuable advice. I express my sincere gratitude to the committees for my dissertation who are Assoc. Prof. Dr. Punpiti Piamsa-nga for his insightful comments, Assoc. Prof. Dr. Nittaya Kerdprasop for sharing her expertise and giving me guidance, and Dr. Suphakit Niwattankul for giving me encouragement and inspiration in all aspects of this endeavor.

I wish to thank Asst. Prof. Dr. Issra Pramoolsook who willingly spared his precious time for providing close supervision about English writing on my dissertation.

My grateful thanks go to Suranaree University of Technology that provides me with a scholarship for studying a doctoral degree. I would also like to express my gratitude to all of my lecturers at the School of Information Technology for their patience in teaching me throughout my study, kindness, and warm encouragement.

Finally, I wish to convey thanks to my beloved parents, sister and brother for their endless love and care, encouragement and moral support.

Phichayasini Kitwatthanathawon

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

LIST OF TABLES

Table **Page**

LIST OF TABLES (Continued)

Table **Page**

LIST OF TABLES (Continued)

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES (Continued)

Figure **Page**

LIST OF FIGURES (Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Introduction

In traditional businesses, the business operators focus on face-to-face or direct physical contact and have storefronts to facilitate their customers. They utilize some office automation functions in limited areas. Recently, the Internet technology has been used to support the entire operations of modern business, e.g. buying, selling and service process. Customers can access the business via the Internet anytime and anywhere, which is known as electronic commerce (or e-commerce) (Sukhothai Thammathirat Open University, 2007). E-commerce refers to a wide range of online business activities for products and services (Rosen, 2000: 5). It is usually associated with conducting any transaction through the Internet. The use of information and communication technology in e-commerce has enhanced productivity, encouraged customer participation, and enabled customization, besides reducing the costs (Andam, 2003: 5). In addition, the business operators can conduct the e-commerce along with their storefronts.

The tourism business is one of businesses that could benefit from the Internet technology, such as searching tourism information or purchasing online services. Recently, travel-related organizations have transformed their own organizations into electronic business (called e-tourism) for encouraging tourists to pay for their products or services. The e-tourism also facilitates tourists to reserve a hotel, book a flight or rent a car more rapidly.

Most tourists use the e-tourism websites to retrieve information for supporting their decision in selecting the tourist places, accommodation, and related information. In addition, the most common method that tourists use in selecting the tourist places is comments from online reviews of experienced tourists. Recently, there are several e-tourism websites that provide online reviews to support tourists' decision, such as Booking.com, Agoda.com, and TripAdvisor.com. The world's largest e-tourism website enabling tourists to plan and have the perfect trip is TripAdvisor.com (TripAdvisor, www, 2013: 1). It also offers trusted advice from experienced tourists who write reviews on tourist destinations, services or products. It collects tourism-related information from tourists in 30 countries worldwide and there are more than 260 million different monthly visitors, 47-million registered members, over 100 million reviews and opinions, and 60 new reviews and opinions posted per minute. These records conform to the findings of consumer reliability surveys. For example, 70% of global consumers trust consumers' opinions posted online (Nielsen, www, 2009: 3), and 79% of younger consumers between 16 and 34 years old trust online customers' reviews as much as personal recommendations, whereas older consumers are less likely to trust online opinions of strangers (Anderson, www, 2010: 1). Therefore, from the above findings, it can be observed that the online reviews contribute to prospective buyers' decision in selecting goods or services (Zhang, Narayanan, and Choudhary, 2010), and it can be used as a reflection to see the weaknesses of business. If the business operators ignore improving their own business or satisfying customers with their products or services, the customer reviews may directly and negatively affect the viability of the business.

Although there are many websites in various businesses that comprise a write a review module, such as online bookstore websites, e-tourism websites and products manufacturers' websites, those reviews have rarely been utilized. It is due to the fact that those reviews have never been processed or extracted the valued information. For instance, the tourism websites store reviews of experienced tourists about tourist destinations, such as attractions, accommodations or restaurants. These reviews are useful for other tourists to conduct their travel plans. If tourists want to know about these destinations in details, they must read enormous reviews by themselves and then analyze or decide whether they should go or not (Zhang, Narayanan, and Choudhary, 2010). It is due to the fact that the existing websites provide only an overall rating of each destination from these reviews. Moreover, the existing websites could not rank the attractiveness of tourist destinations that have an equal rating; therefore tourists cannot utilize those data for supporting their decisions immediately. Finally, tourists could not search for information according to their individual needs because most tourism websites are still static (Nysveen and Pedersen, 2004).

Currently, there are several review analysis modules (Turney, 2002; Dave, Lawrence, and Pennock, 2003; Hu and Liu, 2004; Taboada and Grieve, 2004; Whitman and Ellis, 2004; Zheng and Ye, 2009; Jakob and Gurevych, 2010; Ramkumar, Rajasekar, and Swamynathan, 2010; Zhang, Narayanan, and Choudhary, 2010). Those modules consist of four common processes: accommodation feature extraction, accommodation rating, knowledge base and knowledge explanation. However, those processes in each module have different approaches in details. All existing review analysis modules extract only accommodation features (such as price, value, cleanliness, service, etc.), but they do not extract relationships between the features. For example, the features “cleanness”, “room”, and “hotel” are extracted but their relationships are not revealed. In addition, the feature scoring methods of the existing modules are illustrated in the form of binary-rating scale (e.g. “positive/negative” or “recommend/don’t recommend”) that differs from user opinions using 5-rating scale (i.e. “very poor”, “poor”, “average”, “good”, and “very good”) (Burke, 2002; Cesarano et al., 2006).

Typically, the existing modules are evaluated by comparing module-computed scores with scores converted to binary, not raw scores. The converted scores are defined by experts, while the raw scores are specified by experienced tourists. Thus, the evaluation results from the comparison may not align with opinions of the experienced tourists.

From the problems of the existing modules stated above, the problem statement of this study is provided as follows.

1. There are no relationships between accommodation features.
2. The existing feature rating methods compute in the form of binary-rating scale that contrast with tourist-defined rating in the real world (using 5-rating scale).
3. The evaluation is performed by comparing module-computed scores with scores converted to binary, not raw scores defined by experienced tourists.

Hence, this study aims to design an analysis module for online travel accommodation reviews. This designed module can extract accommodation features and their relationships. In addition, it can calculate the tourists' satisfaction level on each extracted feature and on the entire review. Finally, the designed module can explain more clearly reviews with 5-rating scale and visualize the accommodation feature relationships in hierarchy.

1.2 Research Objectives

1.2.1 To design a review analysis module that performs the following tasks.

a) Extract feature relationships and present them in hierarchy.

b) Design a feature scoring method in the form of 5-rating scale.

1.2.2 To evaluate the review analysis module by comparing module-computed scores with raw scores defined by experienced tourists.

1.3 Research Hypothesis

1.3.1 The analysis module for online travel accommodation reviews can extract feature relationships correctly with more than 80% of Recall.

1.3.2 The analysis module for online travel accommodation reviews can compute the entire review scores correctly, whereas the difference between the module-computed scores and the tourist-specified scores is less than 0.1 review points from 5 full scores.

1.4 Basic Assumption

1.4.1 The online travel accommodation reviews, i.e. hotel, resort, home stay, and national park, are used as a case study in the design of an online review analysis module.

1.4.2 The tourists have to write a travel accommodation review in the English language.

1.4.3 The travel accommodation review, entered into the system, must not have any errors such as typographical, grammatical, spelling, and vocabulary errors.

1.5 Scope of the Study

This study is the design of an analysis module for online reviews, which uses the tourism information on travel accommodations as a case study. This research focuses on travel accommodation reviews because tourism is one of important industries that create revenues in Thailand. This research proposes a module design for analyzing unstructured texts stored in the travel accommodation reviews. The proposed module comprises the methods of information extraction and accommodation rating. The information extraction method is used to solve a sentence analysis problem by applying the tourism ontology to extract and store accommodation information, called features (such as factors related to services and facilities). The accommodation rating method is used to compute feature scores and aggregate them to overall scores of the accommodation. In conclusion, the extracted accommodation information with rating scores is illustrated in a hierarchical structure as knowledge representation of the accommodation.

1.6 Expected Results

1.6.1 Direct expected results

1) An analysis module for online travel accommodation reviews that performs the following tasks.

- Extract accommodation features and their relationships including presenting them in hierarchy.
- Calculate the tourists' satisfaction level on each extracted feature and the entire review in the form of 5-rating scale.

2) The evaluation method by comparing module-computed scores with raw scores defined by experienced tourists.

1.6.2 Indirect expected results

1) For tourism business entrepreneurs:
   - The designed module in analyzing customers' opinions that affect the tourism business.
   - The extracted knowledge for developing their accommodation services and facilities in order to meet more customers' need and gain more advantages over the competitors.
   - The extracted knowledge for supporting strategic management and tourism business policy.
2) For tourists:
   - The extracted knowledge for supporting their decisions on selecting travel accommodations easily and quickly.
3) For general use:
   - The designed module for online travel accommodation reviews to analyze other reviews, such as book reviews, product reviews, restaurant reviews, shop reviews, and tourist attraction reviews.

1.7 Definitions of Terms

1.7.1 Module

A module means a portion of a program or functional unit that carries out a specific and important function. It may be used alone or connected with other components of the same program according to the purposes of the module. Moreover, it is also designed for reusability, that is, it can be deployed with other systems.

1.7.2 Accommodation

An accommodation means travel accommodations, i.e. hotel, resort, home stay, and national park.

1.7.3 Review

A review means an opinion of experienced tourists who criticize travel accommodation services and facilities on websites.

1.7.4 Criticism

A criticism means an opinion given in the review, which judges the qualities of the accommodations.

1.7.5 An Analysis Module for Online Travel Accommodation Reviews

An analysis module for online travel accommodation reviews means a module for extracting the tourists' satisfaction level on services and facilities of travel accommodations from online reviews given by experienced tourists.

1.7.6 Tourists' Satisfaction Level

Tourists' satisfaction level means the numerical scores representing tourists' satisfaction with accommodation services and facilities.

1.7.7 Expert

An expert means the person who has knowledge or expertise in the English language.

CHAPTER 2

REVIEW OF THE LITERATURE

This chapter contains the related literature on the design of an analysis module for online travel accommodation reviews. Firstly, the background and method of information extraction are described. Next, primary problems of language parsing, the context free grammar approach, and anaphora resolution are reviewed. Then, definitions and structures of ontology are presented. After that, fuzzy logic and tourists' satisfaction are described. Finally, the chapter concludes with related work on online review analysis.

2.1 Information Extraction

This research aims to design an analysis module for online reviews, which focuses on extracting tourist opinions, including a calculation of the tourists' satisfaction level on each extracted feature in details. Hence, literature on information extraction is reviewed for designing an approach that can extract useful information to support tourists' decisions. There are two matters to be discussed below.

2.1.1 Rationale

The explosive growth and popularity of the World Wide Web have resulted in a huge amount of information on the Internet. Because that amount keeps increasing, accessing this huge collection of information is almost impossible for a single user to read all the information under the time constraints (Zhang and Narayanan, 2010). Besides spending too much time on reading, users may lose opportunity to get crucial information. Therefore, mining these reviews to extract useful information efficiently is an important and challenging problem.

However, as the number of customers' reviews continues to increase, it is almost impossible for a single user to read and comprehend all the reviews to make informed decisions. Therefore, the process of information extraction plays an important role in transforming unstructured and/or semi-structured machine-readable documents into structured information. The extracted information can be used in other tasks effectively. However, there are important problems that affect the performance of extracting process, such as language processing, extraction modeling, and knowledge base maintenance (Kawtrakul, www, 2007: 1).

The information extraction not only allows users to access and use the extracted information easily, but also helps them to make decisions. Therefore, it is widely used for analyzing and creating useful results based on users' requirements, such as statistical analysis and possible solution recommendation (Kawtrakul, www, 2007: 1-21).

In this research, information extraction refers to the task of discovering valued information or specific knowledge in natural language, unstructured or semi-structured text. The discovery process is a data analysis in order to categorize and identify the relationships of the data. The final output of the extraction process varies; in every case, however, that output can be transformed for populating some type of database (Cowie and Wilks, 1996). In this context, the information extraction process implies an automatic review analysis for discovering the tourists' satisfaction level.

2.1.2 Information Extraction Techniques

Information extraction systems have been developed for a variety of domains. Hence, there are many research studies which provide different algorithms and methodologies, such as the information extraction based on corpus including various techniques for enhancing the accuracy of results.

For instance, an algorithm for word extraction from Thai texts employs the decision tree technique. Several attributes such as string length, frequency, mutual information, and entropy were chosen for word/non-word determination (Sornlertlamvanich, Potipiti, and Charoenporn, 2000).

Haruechaiyasak, Srichaivattana, Kongyoung, and Damrongrat (2004) proposed an alternative method to word segmentation for extracting important keywords from categorized text corpus, called Automatic Categorized Keyword Extraction (ACKE). The experiments were performed on Thai newspaper articles.

Romero, Olivas, Genero, and Piattini (2008) proposed a text mining technique for the automatic extraction of the most relevant terms used in Empirical Software Engineering (ESE) documents. The goal was to define a glossary of terms related to ESE based on an initial glossary published in http://lens-ese.cos.ufrj.br/wikiese. This glossary must be dynamically updated with information extracted from the relevant documents in the research domain. Finally, the relationships between terms are built by the ESE ontology.

Furthermore, there is a good amount of research on information extraction from web documents. Most of these research studies apply a wrapper to transform the HTML document into user-defined format, such as Peerawit, Yingsaeree, and Kawtrakul (2004) which included machine learning from the sample labeled document. However, the information extraction from web documents still has a limitation of domain difference because the dissimilar domain of data means the dissimilar web structure.

In addition, there are many research studies applying information extraction in other areas, i.e. studying and trying out structural transformation from the HTML document to tree structure (Chanlekha and Kawtrakul, 2001); finding and extracting tables from documents (Smitinand, 2001; Sirigayon, Chanlekha, and, Kawtrakul, 2004; Imsombut, Suktarachan, Yingsaree, and Kawtrakul, 2005; Chareonsuk, Sukvakree, and Kawtrakul, 2005); developing an information extraction system of agricultural expert by extracting the experts' information from multi-webpage. This system applies Simple Rule Language (SRL) for constructing extraction rules. However, the limitation of this study was to extract semi-structure websites only. The evaluation results were achieved in 91.5% of precision assessed from 5 websites (Wuttilerdcharoenwong, 2009).

2.2 Parser

This current research aims to design a semantic analysis algorithm for digesting significant issues from travel accommodation reviews and calculating the tourists' satisfaction level on each extracted issue which is stored in the knowledge base as useful resources for other tourists. Consequently, the parser theory is reviewed in order to clearly develop semantic analysis process and achieve the precious results for other tasks.

2.2.1 Background Problem and the Analyzing Process

Liddy (2001: 1) stated that Natural Language Processing (NLP) is considered a discipline within Artificial Intelligence (AI) which refers to a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications.

The notion of "*levels of linguistic analysis*" of this definition implies the fact that there are multiple types of language processing known to be at work when humans produce or comprehend language. It is thought that humans normally utilize all of these levels since each level conveys different types of meaning. But various NLP systems utilize different levels, or combinations of levels of linguistic analysis, and this is seen in the differences among various NLP applications.

In language processing, parser is used for recognizing an unstructured sentence in order to solve the syntactic and semantic analysis problems (Sirinaovakul, 2008: 386-387). Parsing is the task of recognizing a sentence and assigning a syntactic structure to it (Jurafsky and Martin, 2000: 357). There are many processes of computerized approach for processing texts stored in natural language. The following description of levels will be presented sequentially. The key point here is that meaning is conveyed by each and every level of language and that since humans use all levels of language to gain understanding, the more capable an NLP system is, the more levels of language it will utilize (Liddy, 2001: 6).

- **Morphological processing**

This level deals with the componential nature of words which are composed of morphemes – the smallest units of meaning. For example, the word "*preregistration*" can be morphologically analyzed into three separate morphemes: the prefix "*pre*", the root "*registra*", and the suffix "*tion*". Since the meaning of each morpheme remains the same across words, humans can break down an unknown word into its constituent morphemes in order to understand its meaning. Similarly, an NLP system can recognize the meaning conveyed by each morpheme in order to gain and represent meaning. For example, adding the suffix *–ed* to a verb conveys that the action of the verb took place in the past. This is a key piece of meaning, and in fact, that is frequently only evidenced in a text by the use of the *-ed* morpheme.

- **Lexical processing**

At this level, humans, as well as NLP systems, interpret the meaning of individual words. Several types of processing contribute to word-level understanding – the first of these being assignment of a single part-of-speech tag to each word. In this processing, words that can function as more than one part-of-speech are assigned the most probable part-of-speech tag based on the context in which they occur. Additionally at the lexical level, those words that have only one possible sense or meaning can be replaced by a semantic representation of that meaning. The nature of the representation varies according to the semantic theory utilized in the NLP system.

- **Syntactic processing**

This level focuses on analyzing the words in a sentence in order to uncover the grammatical structure of the sentence. This requires both a grammar and a parser. The output of this level is a representation of the sentence that reveals the structural dependency relationships between those words. There are various grammars that can be utilized, which will impact the choice of a parser. Not all NLP applications require a full parse of sentences, therefore the remaining challenges in parsing of prepositional phrase attachment and conjunction scoping no longer stymie those applications for which phrasal and clausal dependencies are sufficient. Syntax conveys meaning in most languages because order and dependency contribute to meaning.

For example, the two sentences: "*The dog chased the cat.*" and "*The cat chased the dog.*" differ only in terms of syntax, yet convey quite different meanings.

- **Semantic processing**

This is the level at which most people think meaning is determined, however, as described in the above defining of the levels, it is all the levels that contribute to meaning. The semantic processing determines the possible meanings of a sentence by focusing on the interactions among word-level meanings in the sentence. This level of processing can include the semantic disambiguation of words with multiple senses; in an analogous way to how syntactic disambiguation of words that can function as multiple parts-of-speech is accomplished at the syntactic level. Semantic disambiguation permits one and only one sense of polysemous words to be selected and included in the semantic representation of the sentence. For example, amongst other meanings, "*file*" as a noun can mean either a folder for storing papers, or a tool to shape one's fingernails, or a line of individuals in a queue. If information from the rest of the sentence were required for the disambiguation, the semantic, not the lexical level, would do the disambiguation. A wide range of methods can be implemented to accomplish the disambiguation, some of which require information as to the frequency with which each sense occurs in a particular corpus of interest, or in general usage, some which require consideration of the local context, and others which utilize pragmatic knowledge of the domain of the document.

- **Discourse processing**

While the syntax and semantics work with sentence-length units, the discourse level of NLP works with units of text longer than a sentence. That is, it does not interpret multi-sentence texts as just concatenated sentences, each of which can be interpreted singly. Rather, discourse focuses on the properties of the text as a whole that convey meaning by making connections between component sentences. Several types of discourse processing can occur at this level, two of the most common being anaphora resolution and discourse/text structure recognition.

- **Pragmatic processing**

This level is concerned with the purposeful use of language in situations and utilizes context over and above the contents of the text for understanding. The goal is to explain how extra meaning is *read into* texts without actually being encoded in them. This requires much world knowledge, including the understanding of intentions, plans, and goals. Some NLP applications may utilize knowledge bases and inference modules. For example, the following two sentences require resolution of the anaphoric term "they", but this resolution requires pragmatic or world knowledge.

- *The city councilors refused the demonstrators a permit because **they** feared violence.*
- *The city councilors refused the demonstrators a permit because **they** advocated revolution.*

2.2.2 Context Free Grammar

Jurafsky and Martin (2000: 326-328) stated that the most commonly used mathematical system for modeling constituent structure in English and other natural languages is the Context Free Grammar, or CFG. Context free grammars are also called Phrase-Structure Grammars, and the formalism is equivalent to what is also called Backus-Naur Form or BNF.

A context free grammar consists of a set of rules or productions, each of which expresses the ways that symbols of the language can be grouped and ordered together, and a lexicon of words and symbols. For example, as shown in Equation 2.1-2.3, the following productions express that a *NP* (or noun phrase) can be composed of either a *ProperNoun* or a determiner (*Det*) followed by a *Nominal*; a *Nominal* can be one or more *Nouns*.

$$NP \rightarrow Det\ Nominal \tag{2.1}$$

$$NP \rightarrow ProperNoun \tag{2.2}$$

$$Nominal \rightarrow Noun \mid Noun\ Nominal \tag{2.3}$$

Context free rules can be hierarchically embedded, so the previous rule could be combined with others which express facts about the lexicon as Equation 2.4-2.6.

$$Det \rightarrow a \tag{2.4}$$

$$Det \rightarrow the \tag{2.5}$$

$$Noun \rightarrow flight \tag{2.6}$$

The symbols that are used in a CFG are divided into two classes. The symbols that correspond to words in the language ("the", "flight") are called terminal symbols. The lexicon is the set of rules that introduce these terminal symbols. The symbols that express clusters or generalizations of there are called non-terminals. In each context free rule, the item to the right of the arrow ($\rightarrow$) is an ordered list of one or more terminals and non-terminals while to the left of the arrow is a single non-terminal symbol expressing some clusters or generalizations. Notice that in the lexicon, the non-terminal associated with each word is its lexical category, or part-of-speech.

A context free grammar is usually thought of in two ways: as a device for generating sentences, or as a device for assigning a structure to a given sentence. As a generator, the $\rightarrow$ arrow could be read as "rewrite the symbol on the left with the string of symbols on the right". So starting from the symbol *NP*, rule 2.1 can be used to rewrite *NP* as *Det Nominal*, and then apply rule 2.3 to derive *Det Noun*, and finally parse via rules 2.4 and 2.6 as *a flight*. In summary, the string *a flight* can be derived from the non-terminal *NP*. Thus a CFG can be used to randomly generate a series of strings. This sequence of rule expansions is called a derivation of the string of words. It is common to represent a derivation by a parse tree (commonly shown inverted with the root at the top). Here is the tree representation of this derivation. (Fry, 2004: 1-4; Jurafsky and Martin, 2000: 328)

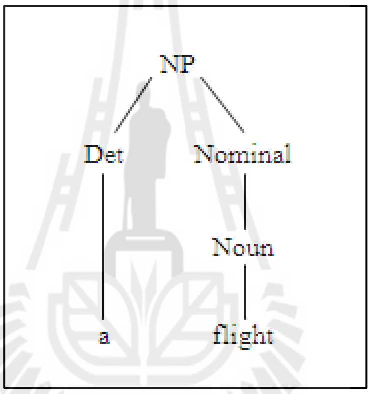


Figure 2.1 A parse tree for "*a flight*"

The formal language defined by a CFG is the set of strings that are derivable from the designated start symbol. Each grammar must have one designated start symbol, which is often called *S*. Since context free grammars are often used to define sentences, *S* is usually interpreted as the "sentence" node, and the set of strings that are derivable from *S* is the set of sentences in some simplified version of English.

Let's add to sample grammar a couple of higher-level rules that expand *S*, and a couple others. One will express the fact that a sentence can consist of a noun phrase and a verb phrase (Jurafsky and Martin, 2000: 329).

$S \rightarrow NP\ VP$

$NP \rightarrow Det\ Noun$

$VP \rightarrow V\ Adj$

$Det \rightarrow the$

$Noun \rightarrow room$

$V \rightarrow is$

$Adj \rightarrow clean$

We can use this grammar to generate sentences by starting with *S*, expanding it to *NP VP*, choosing a random expansion of *NP* and *VP*, and so on until the string *"the room is clean"* is generated. Figure 2.2 shows a parse tree that represents a complete derivation of *"the room is clean."*

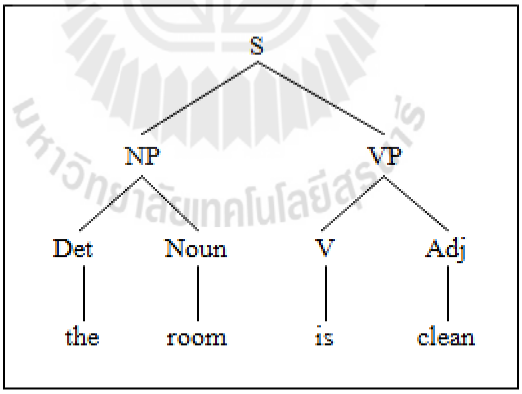


Figure 2.2 A parse tree for "*the room is clean*" according to grammar

In summary, parser is an important component of information extraction process. Parsing is a combination of recognizing an input string and assigning some structure to it. Then, syntactic parsing is the task of recognizing a sentence and assigning a syntactic structure to it. Since context free grammars are a declarative formalism and allow us to model the constituency facts (or groups of words may behave as a single unit or phrase), this research focuses on the kind of structure assigned by the context free grammars.

2.3 Anaphora Resolution

This research aims to extract and digest the meaning from online travel accommodation reviews contents. Thus, the discourse analysis theory in a matter of anaphora resolution is reviewed for identifying relationships between the words within review sentences.

2.3.1 Definition

Anaphora resolution is basically the problem of resolving what a pronoun or a noun phrase refers to. Various definitions of anaphora have been put forward, but the classical definition given by Halliday and Hasan (1976) which is based on the notion of cohesion: anaphora is cohesion (presupposition) which points back to some previous item. The "pointing back" (reference) is called an *anaphor* and the entity to which it refers is its *antecedent*. The process of determining the antecedent of an anaphor is called *anaphora resolution*. Usually, both the antecedent and the anaphor are used as referring expressions and having the same referent in the real world, they are termed *coreferential*. (Mitkov, 1999: 1-2)

Moreover, anaphora is defined by Hirst (1981) as "the device of making in discourse (group of sentences) an *abbreviated* reference to some entity (or entities) in the expectation that the perceiver of the discourse will be able to disabbreviate the reference and thereby determine the identity of the entity. The reference is called an *anaphor*, and the entity to which it refers is its *referent* or *antecedent*. A reference and its referent are said to be *coreferential*. The process of determining the referent of an anaphor is called *resolution*".

In Natural Language Processing (NLP), the discourse level works with units of text longer than a sentence. Rather, discourse focuses on the properties of the text as a whole that convey meaning by making connections between component sentences. Several types of discourse processing can occur at this level, two of the most common being anaphora resolution and discourse/text structure recognition. Anaphora resolution is the replacing of words such as pronouns, which are semantically vacant, with the appropriate entity to which they refer. Discourse/text structure recognition determines the functions of sentences in the text, which, in turn, adds to the meaningful representation of the text. For example, newspaper articles can be deconstructed into discourse components such as Lead, Main Story, Previous Events, Evaluation, Attributed Quotes, and Expectation. (Liddy, 2001: 8)

For example (Wilson, www, 2012): "Mary died. She was very old." The word "she" refers to "Mary", and is described as an anaphoric reference to "Mary". "Mary" is described as the antecedent of "she". Anaphoric references are frequently pronouns, such as "he" "her" "their" "one" "ones" "this" "that" "these" "those" showed in the example, but may also be definite noun phrases, as in: "Barack Obama frowned. The President was clearly worried by this issue." Here, "The President" is an anaphoric reference to "Barack Obama". Anaphors may in some cases not be explicitly mentioned in a previous sentence - as in "John got out his pencil. He found that the lead was broken." "The lead" here refers to a subpart of "his pencil". Anaphors need not be in the immediately preceding sentence, they could be further back, or in the same sentence, as in "John got out his pencil, but found that the lead was broken." In all our examples so far the anaphor and the antecedent are noun phrases, but verb phrases and sentence-anaphora is also possible, as in "I have today dismissed the prime minister. It was my duty in the circumstances." Here "It" is an anaphoric reference to the verb phrases "dismissed the prime minister".

2.3.2 Types of Anaphora Resolution

There are various types of anaphora. This section provided a brief description of basic five types of referring expression. (Hirst, 1981: 2-10; Jurafsky and Martin, 2000: 673-677; Mitkov, 1999: 2-3)

2.3.2.1 Indefinite Pronouns

An indefinite pronoun (Willis, 1996: 14) is a pronoun that refers to one or more unspecified beings, objects, or places which are not definite or specific or exact, e.g. "Everyone", "Somebody", "Anything", "Each", "Nothing", "One", and "No one". The most common form of indefinite reference is marked with the determiner "a" (or "an") but it can also be marked by quantifier such as "some", "this" or even the determiner "one" (called One-anaphora). One-anaphora is the case when the anaphoric expression is realized by a "one" noun phrase (Mitkov, 1999: 2-3), for instance, "John found the love of his life. The relationship did not last long. He started a new one." Here, the word "one" refers to "a relationship".

2.3.2.2 Definite Noun Phrase Anaphora

Definite noun phrase anaphora occurs in the situation that the antecedent is referred by a definite noun phrase representing either same concept (repetition) or semantically close concepts (e.g. synonyms and superordinate) (Mitkov, 1999: 3). Definite reference is used to refer to an entity that is identifiable to the hearer, either because it has already been mentioned in the discourse context, it is contained in the hearer's set of beliefs about the world, or the uniqueness of the object is implied by the description itself (Jurafsky and Martin, 2000: 674). For example, "John found the love of his life. The relationship did not last long." Here, a definite noun phrase "The relationship" refers to "the love" in previous sentence.

2.3.2.3 Pronominal Anaphora

The pronominal anaphora is the most widespread type of anaphora which is realized by anaphoric pronouns: "He", "She", "It", "They", and "That" (Mitkov, 1999: 2). For example, "John found the love of his life." Where "his" refers to the man named "John".

Moreover, the epithets can also be used pronominally, as "the poor guy" in sentence "Ross used his Bankcard so much, the poor guy had to declare bankruptcy.", or "the bastard" in sentence "When John found out about Mary's marital infidelity, the bastard punched her." (Hirst, 1981: 10)

2.3.2.4 Locative or Demonstrative References

Demonstrative pronouns, like "This" and "That", behave somewhat differently than simple definite pronouns like "it". They can appear either alone or as determiners, for instance, "this car" and "that car". The choice between two demonstratives is generally associated with some notion of spatial proximity:

"this" indicating closeness and "that" signaling distance. Spatial distance might be measured with respect to the discourse participants' situational context.

Moreover, the word "there" is often an anaphoric reference to a place, as in "The Church of Scientology met in a secret room behind the local Colonel Sanders' chicken stand. Sue had her first diabetic experience there."

Locative relations, like temporal relations, may also reference anaphorically, as "Across the street" in the sentence "The Church of Scientology met in a secret room behind the local Colonel Sanders' chicken stand. Across the street was a McDonald's where the Bokononists and The Church of God The Utterly Indifferent had their meetings."

2.3.2.5 Ellipsis (∅)

Some anaphoras are completely null, such as in "Nadia brought the food for the picnic, and Daryel ∅ the wine." Here the elided verb phrase is "brought to the picnic". Verb phrase ellipsis cannot in general be exospheric. This examples illustrated verb phrase ellipsis. However, almost any part of a sentence can be elided, as in "Ross carefully folded his trousers and ∅ climbed into bed." Here, the subject noun phrase "Ross" is elided.

2.3.3 Anaphora Resolution Method

Mitkov (1999: 3) stated that most of the anaphora resolution systems deal with resolution of anaphors which have noun phrases as their antecedents because identifying anaphors which have verb phrases, clauses, sentences or even paragraphs/discourse segments as antecedents, is a more complicated task. Typically, all noun phrases (NPs) preceding an anaphor are initially regarded as potential candidates for antecedents. Usually, a search scope has to be identified: most approaches look for NPs in the current and preceding sentence. However, an "ideal" anaphora resolution system should extend its scope of search: antecedents which are 17 sentences away from the anaphor have already been reported (Mitkov, 1995). Assuming that the scope of search for a specific approach has already been specified, the NPs preceding the anaphor within that scope are identified as candidates for antecedents and a number of anaphora resolution factors are employed to track down the correct antecedent.

Having described a variety of reference phenomena that are found in natural language, now can consider how one might develop algorithms for identifying the references of referential expressions. One step that needs to be taken in any successful reference resolution algorithm is to filter the set of possible referents on the basis of certain relatively hard-and-fast constraints. Some of these constraints are described below.

2.3.3.1 Constraints

Several constraints will be outlined and illustrated by examples. Coreferential items are given the same index (Mitkov, 1999:4; Jurafsky and Martin, 2000: 678-679).

- Gender, number, person and case agreement

This constraint requires that anaphors and their antecedents must agree in number, gender, person and case as shown in Table 2.1-2.3, respectively.

Table 2.1 Gender agreement in the English pronominal system

Singular	Plural	Unspecified
"She", "Her", "He", "Him", "His", "It"	"We", "Us", "They", "Them"	"You"

Table 2.2 Number agreement in the English pronominal system

Masculine	Feminine	Nonpersonal
"He", "Him", "His"	"She", "Her"	"It"

Table 2.3 Person and case agreement in the English pronominal system

	First	Second	Third
Nominative	"I", "We"	"You"	"He", "She", "They"
Accusative	"Me", "Us"	"You"	"Him", "Her", "Them"
Genitive	"My", "Our"	"Your"	"His", "Her", "Their"

- Syntactic constraints

Reference relation may also be constrained by the syntactic relationship between a referential expression and a possible antecedent noun phrase when both occur in the same sentence. Results in Government and Binding Theory (Haegeman, 1994) and Lexical Functional Grammar have provided useful constraints on the anaphors and their antecedents which have been successfully used in anaphor resolution for eliminating unacceptable candidates when searching for the antecedent.

(a) A non-pronominal noun phrase (NP) cannot overlap in reference with any noun phrase that c-commands it. For example, He_i told them about $John_j$.

(b) The antecedent of a bound anaphor must c-command it. For example, $John_i$ likes pictures of $himself_i$.

(c) A personal pronoun cannot overlap in reference with an NP that c-commands it. For example, $John_i$ told $Bill_j$ about him_k.

- Semantic consistency

This constraint stipulates that if satisfied by the anaphor, semantic consistency constraints must be satisfied also by its antecedent.

Vincent removed the DVD from the $\underline{computer_i}$ and then disconnected $\underline{it_i}$.

Vincent removed the $\underline{DVD_i}$ from the computer and then copied $\underline{it_i}$.

2.3.3.2 Preferences

Preferences, as opposed to constraints, are not obligatory conditions and therefore do not always hold. There are three preferences: syntactic parallelism, semantic parallelism and center of attention. (Mitkov, 1999:4)

- Syntactic parallelism

Syntactic parallelism could be quite helpful when other constraints or preferences are not in a position to propose an unambiguous antecedent. This preference is given to noun phrases with the same syntactic function as the anaphor.

The programmer$_i$ successfully combined Prolog$_j$ with C, but he$_i$ had combined it$_j$ with Pascal last time.

The programmer$_i$ successfully combined Prolog with C$_j$, but he$_i$ had combined Pascal with it$_j$ last time.

- Semantic parallelism

This is a useful (and stronger than syntactic parallelism) preference but only systems which can automatically identify semantic roles, can employ it. It says that noun phrases, which have the same semantic role as the anaphor, are favored.

Vincent gave the DVD to Sody$_i$. Kim also gave him$_i$ a letter.

Vincent$_i$ gave the DVD to Sody. He$_i$ also gave Kim a letter.

- Centering

Although the syntactic and semantic criteria for the selection of an antecedent are very strong, they are not always sufficient to distinguish between a set of possible candidates. Moreover, they serve more as filters to eliminate unsuitable candidates than as proposers of the most likely candidate. In the case of antecedent ambiguity, it is the most salient element among the candidates for antecedent which is usually the frontrunner. This most salient element is referred to in computational linguistics as focus (Sidner, 1979) or center (Grosz, Aravind, and Scott, 1983) though the terminology can be much more diverse (Hirst 1981; Mitkov, 1995).

For instance, neither machines, nor humans, would be able to resolve the anaphoric pronoun “it” in the sentence “Jenny put the cup on the plate and broke it." However, if this sentence is part of a discourse segment which makes it possible to determine the most salient element, the situation is different:

> Jenny went window shopping yesterday and spotted a nice cup. She wanted to buy it, but she had no money with her. Nevertheless, she knew she would be shopping the following day, so she would be able to buy the cup then. The following day, she went to the shop and bought the coveted cup. However, once back home and in her kitchen, she put the cup on a plate and broke it...

In this discourse segment, "the cup" is the most salient entity and is the center of attention throughout the discourse segment.

It is now clear that very often when two or more candidates "compete" for the antecedent, the task of resolving the anaphor is shifted to the task of tracking down the center/focus of the sentence (clause). Although, useful the term center (or focus) can be for anaphora resolution, it has suffered from two inconveniences: its intuitive nature and the use of different terms to describe concepts which either seem to be very close to "center" or even could be considered practically identical (e.g. focus, topic, and theme).

2.3.3.3 Computational Strategies

While a number of approaches use a similar set of factors, the "computational strategies" for the application of these factors may differ (here, the term "computational strategy" refers to the way antecedents are computed, tracked down, i.e. the algorithm, formula for assigning antecedents rather than computational issues related to programming languages, and complexity). Some approaches incorporate a traditional model which discounts unlikely candidates until a minimal set of plausible candidates is obtained (and then make use of center or focus, for instance), whereas others compute the most likely candidate on the basis of statistical or AI techniques/models.

2.4 Ontology

This research focuses on a semantic analysis of travel accommodation reviews using an ontology as a set of formal vocabulary definitions for accommodation extraction and storing. Thus, the literature of ontology is reviewed for developing a knowledge base completely.

2.4.1 Definition

The term ontology has its origin in philosophy, in which it refers to the subject of existence, and has been applied in many different ways. The core meaning within computer science is a model for describing the world that consists of a set of types, properties, and relationship types (Garshol, 2004).

In the context of knowledge sharing, several definitions of ontology are described. For example, Gruber (1993: 199) defined that ontology as an explicit specification of a conceptualization. That is, ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. This definition is consistent with the usage of ontology as set-of-concept-definitions, but more general. And it is certainly a different sense of the word than its use in philosophy.

Swartout, Patil, Knight, and Russ (1997: 138) stated that ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base.

Guarino (1998: 4-5) defined that an ontology is a logical theory accounting for the intended meaning of a formal vocabulary.

In this study, ontology refers to a hierarchically structured set of terms for describing a set of concepts within a designated domain and relationships among those concepts. It can be used to reason about the objects within that domain including the relations between them and may be used to describe the domain. Normally, ontologies consist of: *classes* - collection of objects, *attributes* - properties an object can have and share, *relations* - represent the way the objects are related, and *individuals* - which are instances of the class (Chandrasekaran, Josephson, and Benjamins, 1999).

Recently, ontologies are applied in many fields, such as artificial intelligence, semantic web, knowledge management, e-business, and information extraction as a form of knowledge representation about the world or some part of it. The resources in knowledge base or ontology regard as the essential background knowledge which can be improved the system performance (Hepp, Siorpaes, and Bachlechner, 2006: 1).

There are several researches in the area of information extraction using ontology For instance, Alani et al. (2003) developed the Artequakt project which seeks to automatically extract knowledge about artists from the Web. This system integrates a variety of tools in order to automate an ontology-based knowledge acquisition process and maintain a knowledge base which is used to generate customized biographies. The ontology supplies the needed knowledge about the domain and type of information to extract, and how to represent this information in a proper metadata format to insert automatically into a specific knowledge base.

Popov et al. (2003) extracted information by Knowledge and Information Management platform (KIM) which is a platform for semantic indexing, annotation, and retrieval. It combines information extraction based on the mature text engineering platform (GATE) with Semantic Web-compliant knowledge representation and management. The cornerstone is automatic generation of named-entity (NE) annotations with class and instance references to a semantic repository.

Bartolini et al. (2006) proposed a methodology for extracting multimedia information from product catalogues empowered by the synergetic use and extension of domain ontology. The result of the extraction process is a semantically rich representation of the content of catalogs, where knowledge extracted from texts (e.g. product descriptions) is integrated with knowledge extracted from pictures, and made available for any service one may want to build on top of it.

2.4.2 Tourism Ontology

Although ontologies are applied in many fields, the development of ontology still have crucial problem, there is no clear standard definition of product or service (Siricharoen, 2007: 1275), especially in the tourism industry. In practical application, these ontologies cannot be used to connect and share information effectively. Thus, defining standard term should be provided for ontology in order to develop the semantic web with data sharing capability. (Roopa, Mladen, and Nalin, 2007: 1)

There are a number of researches in tourism ontology application such as Ding, Herzog, Luger, Prantner, and Yan (2008) developed the OnTourism project (funded by Austria government). This research explores the possibility to apply semantic web technologies to eTourism area. It aims to create a semantic content management solution based on the existing Microsoft SharePoint employed by the Austrian Tourism call center in order to make full usage of both semantic and social metadata. The OnTourism architecture consists of 6 modules, i.e. sharepoint server module, ontology module, tagging module, mapping module, search module, and bookmarking module. The ontology is used to store semantic annotation or metadata. In the long-term, business implications on such system can be foreseen when their data need to be integrated with other tourism data and end users increases their expectation to find proper or precise information they request.

Chunhua, Pengfei, and Cong (2006) presented the architecture and the knowledge base of a destination decision system named TBJ (Traveling in Beijing), which carries out the design principle of the CTIS (Comprehensive Tourism Information System) which requires a tourism information system to cognize tourist's truly requirement and provides proper information based on this requirement. This cognition can be achieved by considering tourist's information, scenery information, general knowledge of tourism and situational variables in one context. This research applied ontology to provide a common vocabulary for supporting the sharing and reuse of knowledge.

Choi et al. (2009) proposed intelligent recommendation system based on Jeju travel ontology. The proposed system can recommend the tourist more intelligent information using properties, relationships of travel ontology. Next, the system is responsible for finding personalized attractions and plotting location of traveler on the AlMap.

Ananthapadmanaba and Srivats (2011) created sophisticated user profiles ontology which can improve the process of searching for the perfect tourism package by analyzing the user interest with help of user ontology for Tamilnadu tourism.

In this research, the tourism ontology was used to store a set of vocabulary definitions of travel accommodation, which was revised from class hierarchy for the e-tourism ontology version 8 (Siorpaes, Prantner, and Bachlechner, 2004) by analyzing keywords (or accommodation features) in the domain of travel accommodation from 400 accommodation reviews using the *Rocchio's TF-IDF* weighting approach (Salton and Buckley, 1988: 517). All the selected features are added to the standard ontology mentioned above. The revised ontology consists of 10 classes and 95 key properties, as partially shown in Figure 2.3. Nevertheless, the application of the existing ontology and information extraction method in various aspects needs a process modification. This modification aims to adapt suitable process for each aspect in order to extract information accurately and efficiently.

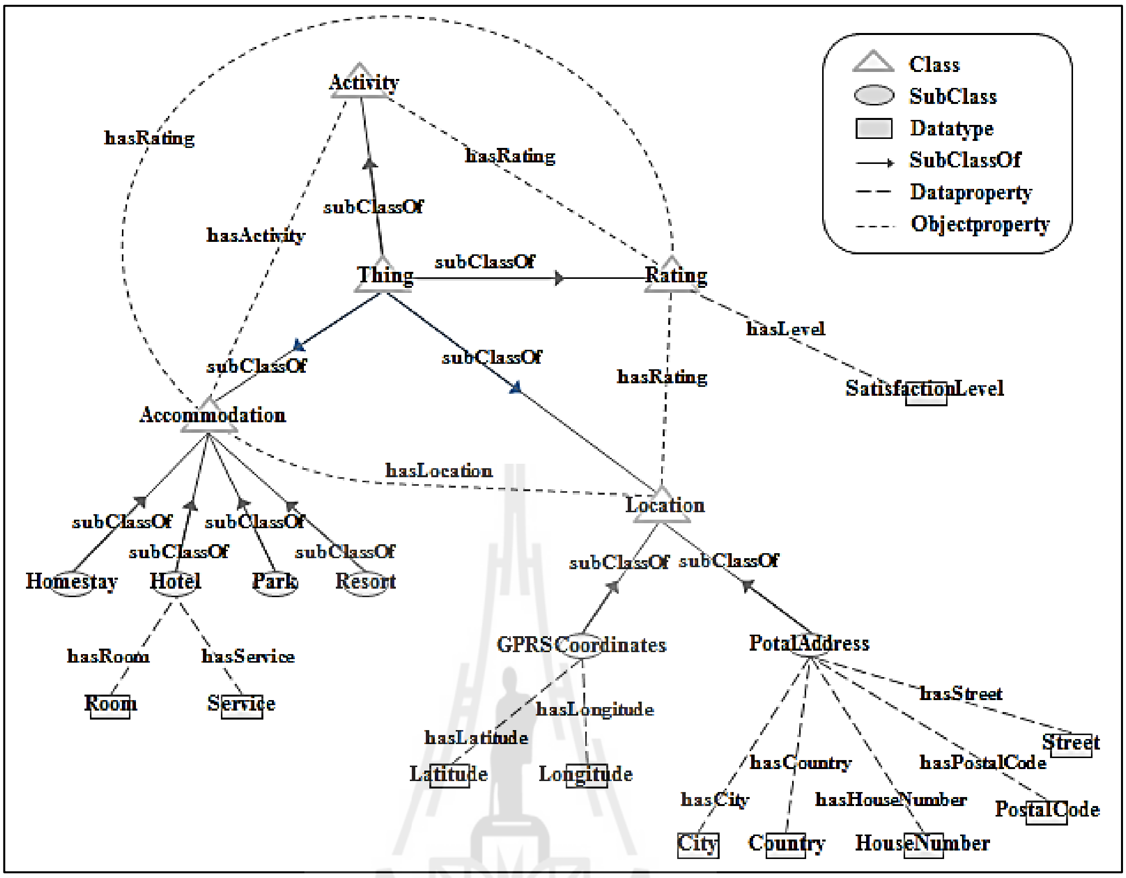


Figure 2.3 A partial structure of modified tourism ontology

2.5 Fuzzy Logic

Fuzzy logic is studied for developing the fuzzy-based method for calculating a tourists' satisfaction level with accommodation services and facilities. This calculation method aims to compute the feature rating of accommodation that closest to the tourist subject as much as possible.

2.5.1 Introduction

Fuzzy logic (Sumathi and Surekha, 2010: 9, 203) was initiated in 1965 by Lotfali Askar Zadeh, professor for computer science at the University of California in Berkeley. Since then it has grown and is found in several application areas. Fuzzy logics are multi-valued logics that form a suitable basis for logical systems reasoning under uncertainly or vagueness that allows intermediate values, as shown in Figure 2.4, to be defined between conventional evaluations like Boolean such as true/false, yes/no, and high/low. These evaluations can be formulated mathematically and processed by computers, in order to apply a more human-like way of thinking in the programming of computers. Fuzzy logic provides an inference morphology that enables approximate human reasoning capabilities to be applied to knowledge-based systems. The theory of fuzzy logic provides mathematical strength to capture the uncertainties associated with human cognitive processes, such as thinking and reasoning. Fuzzy systems are suitable for uncertain or approximate reasoning, especially for the system with a mathematical model that is difficult to derive. Fuzzy logic allows decision making with estimated values under incomplete or uncertain information.

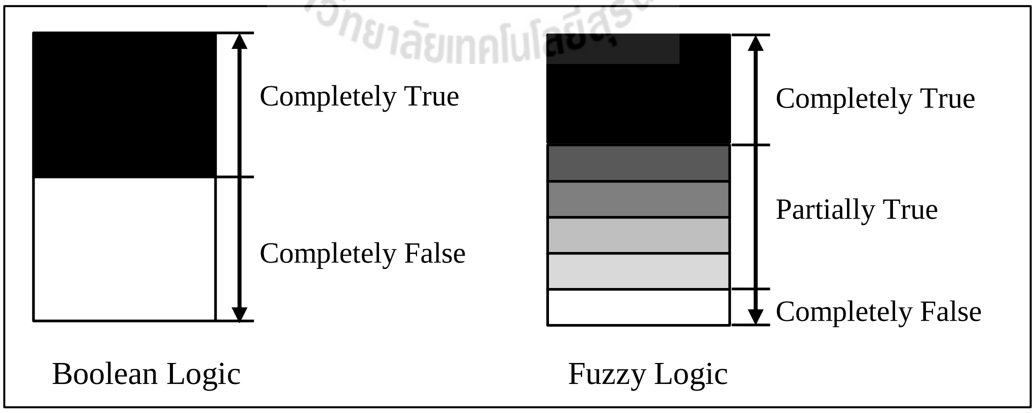


Figure 2.4 Boolean logic and Fuzzy logic

2.5.2 Imprecision and Uncertainty

Fuzziness (Sumathi and Surekha, 2010: 205) should not be confused with other forms of imprecision and uncertainty. There are several types of imprecision and uncertainty and fuzziness is just one aspect of it. Imprecision and uncertainty may be in the aspects of measurement, probability, or descriptions. Imprecision in measurement is associated with a lack of precise knowledge. Sometimes there are measurements that are inaccurate, inexact, or of low confidence.

Imprecision as a form of probability is associated with an uncertainty about the future occurrence of events or phenomena. It concerns the likelihood of non-deterministic events (stochastic uncertainty). An example is the statement “It might rain tomorrow” which exhibits a degree of randomness.

Imprecision in description is the type of imprecision addressed by fuzzy logic. It is the ambiguity, vagueness, qualitativeness, or subjectivity in natural language (linguistic, lexical, or semantic uncertainty). It is the ambiguity found in the definition of a concept or the meaning of terms such as “tall building” or “low scores”. It is also the ambiguity in human thinking, that is, perceptions and interpretations. Examples of statements that are fuzzy in nature are “Homoglobin count is very low.” and “Teddy is rather heavy compared to Ike.”

Fuzzy models and statistical models also possess philosophically different kinds of information: fuzzy memberships comprise similarities of objects to imprecisely defined properties, while probabilities convey information about relative frequencies. Thus, fuzziness deals with deterministic plausibility and not non-determinstic probability.

2.5.3 Crisp and Fuzzy Logic

Sumathi and Surekha (2010: 205-207) stated that fuzzy logic forms a bridge between the two areas of qualitative and quantitative modeling. Although the input-output mapping of such a model is integrated into a system as a quantitative map, internally it can be considered as a set of qualitative linguistic rules.

The term Fuzzy Logic implies that in some manner the methodological analysis is vague or ill-defined. The basic idea behind the development of fuzzy logic rose from the requirement to design the type of vague or ill-defined systems. These ill-defined systems cannot operate on traditional binary valued logic therefore the fuzzy methodological analysis is used based on mathematical theory.

The commonest binary valued logic and set theory is defined as "an element belongs to a set of all possible elements and given any specific subset, whether that element is or is not a member of it." Eventually, not all parameters can be described using binary valued sets. For instance, classification of a person into males and females are easy, but it is problematic to classify them as being tall or not tall. The set of tall people is far more difficult to define, since there is no exact precise value to define tall. Such a kind of problem is frequently twisted so that it can be delineated using the well-known existing technique. The height, e.g. 1.80 m, can be defined as tall, as shown in Figure 2.5 (a). This kind of crisp reasoning would not produce smooth results, since a person of height 1.79 m or a person of height 1.8 m would produce different results.

Fuzzy logic was suggested by Zadeh as a method for mimicking the ability of human reasoning using small number of rules and still producing a smooth output via a process of interpolation. It forms rules that are based upon multi-valued logic and so introduced the concept of set membership. Using fuzzy logic a component can be as decided as belonging to a set, this kind of allotment is carried out by membership functions. For example, a person of height 1.79 m would belong to both tall and not tall sets with a particular degree of membership. Equally the membership grade increases or decreases proportionate with the height of a person, as shown in Figure 2.5 (b). The output of a fuzzy logic thinking system would produce like results for similar inputs. The fuzzy logic theory is just a prolongation of traditional logic where partial set membership could exist, rule conditions could be satisfied partially, and system outputs are calculated by interpolation. Hence, the output is smooth over the equivalent binary-valued rule base. This property is especially crucial to control system applications.

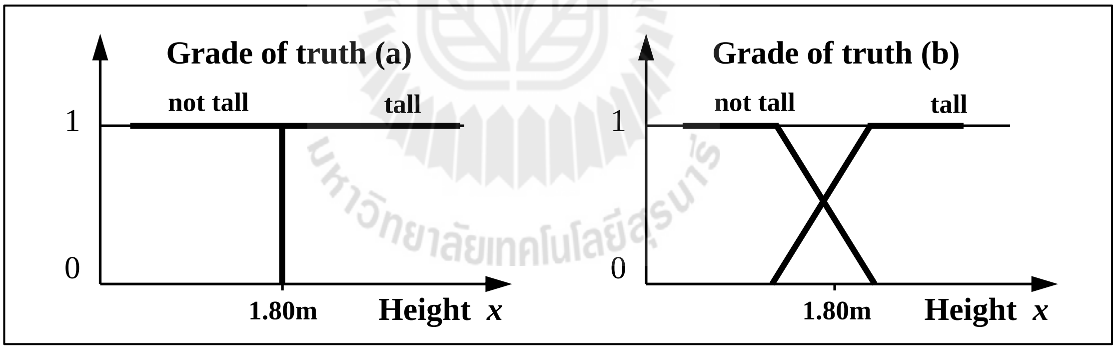


Figure 2.5 The difference between the grade of truth in

(a) Binary valued logic 0,1 and (b) Fuzzy logic [0,1]

2.5.4 Fuzzy Set

Sumathi and Surekha (2010: 207-209) stated that mathematical theory of sets have been extended to create fuzzy sets. There are two ways to describe a set:

explicitly in a list or implicitly with a predicate. In the classical set theory a set A can be represented by enumerating all its elements $f\, a_1, a_2, \ldots, a_n g$ using

$$A = f\, a_1, a_2, \ldots, a_n g.$$

The grade of membership for all its members thus describes a fuzzy set. An item's grade of membership is normally a real number between 0 and 1, often denoted by the Greek letter μ. The higher the number specify, the higher the membership get. If the elements of the above equation a_i (i=1, …, n) of A are together a subset of the universal base set X, the set A can be represented for all elements $x \in X$ by its characteristic function as Equation 2.7.

$$\mu_A(x) = \begin{cases} 1 & if\ x \in A \\ 0 & otherwise \end{cases} \tag{2.7}$$

In classical set theory $\mu_A(x)$ has only values 0 ("false") and 1 ("true"). Such sets are also called crisp sets. Non-crisp sets are called fuzzy sets, for which a characteristic function can also be defined. This function is a generalization of Equation 2.7 and called a *membership function*. The membership of fuzzy set is described by this membership function $\mu_A(x)$ of A, which associates to each element $x_0 \in X$ a grade of membership $\mu_A(x_0)$. In contrast to classical set theory a membership function $\mu_A(x)$ of a fuzzy set can have in the normalized closed interval [0,1] an arbitrary grade of truth. Therefore, each membership function maps elements of a given universal base set X, which is itself a crisp set, into real number in [0,1]. Each fuzzy set is completely and uniquely defined by one particular membership function. Consequently, symbols of membership functions are also used as labels of the associated fuzzy sets. That is, every fuzzy set and its membership function are referred by the same capital letter. Since crisp sets and the associated characteristic functions may be viewed, respectively, as special cases of fuzzy sets and membership functions, the same notation is used for crisp sets as well, as shown in Figure 2.6.

The base set *X* is introduced as a universal set. In practical applications, physical or similar quantities are considered that are defined in some interval. When such quantities are described by sets, a base set can be generalized seamless to a crisp base set *X* that exists in a defined interval. This is a generalization of fuzzy sets. Base sets are not always crisp sets. Another generalization is that the base set is itself a fuzzy set.

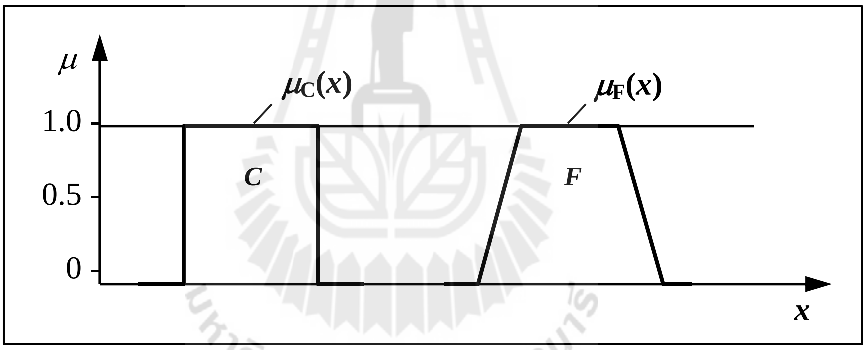


Figure 2.6 Membership functions of a crisp set C and a fuzzy set F

2.5.5 Universe

Sumathi and Surekha (2010: 209-210) stated that the constituents of a fuzzy set are acquired from a universe of discourse also referred to as universe. The universe comprises the complete elements that can inherit consideration. In case of dealing with a non-numerical quantity, for instance taste, which cannot be measured against a numerical scale, a numerical universe cannot be used. The elements are then said to be taken from a psychological continuum An example of such universe applied with Gaussian membership functions could be {extremely low, very low, low, medium, high, very high, and extremely high}.

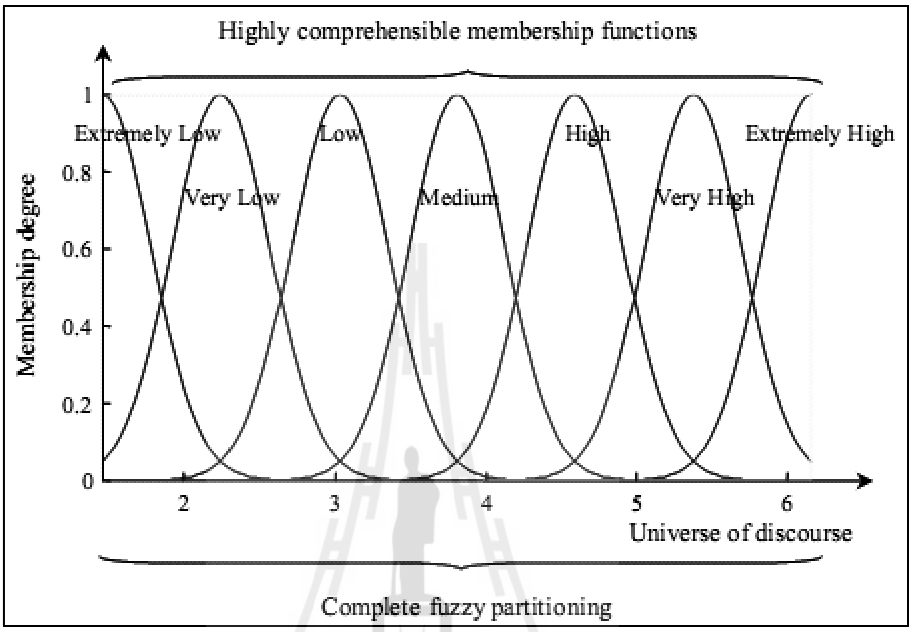


Figure 2.7 An example of universe

2.5.6 Membership Function

Sumathi and Surekha (2010: 210-211) stated that the membership function $\mu_A(x)$ describes the membership of the elements *x* of the base set *X* in the fuzzy set *A*, whereby for $\mu_A(x)$ a large class of functions can be taken. Reasonable functions are often piecewise linear functions, such as triangular or trapezoidal functions.

The grade of membership $\mu_A(x_0)$ of a membership function $\mu_A(x)$ describes for the special element $x=x_0$, to which grade it belongs to the fuzzy set *A*. This value is in the unit interval [0,1]. Of course, x_0 can simultaneously belong to another fuzzy set *B*, such that $\mu_B(x_0)$ characterizes the grade of membership of x_0 to *B*. This case is shown in Figure 2.8.

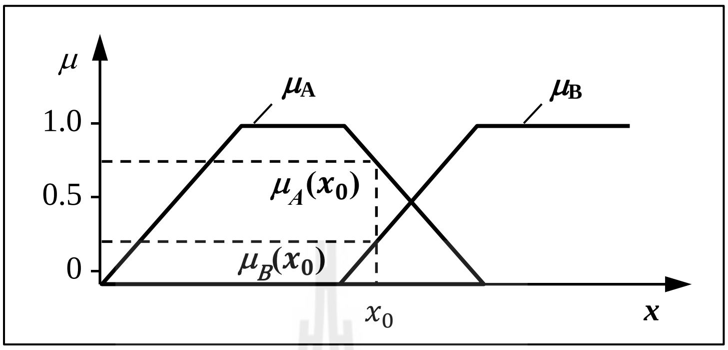


Figure 2.8 Membership grades of x_0 in the sets *A* and *B*:

$\mu_A(x_0) = 0.75$ and $\mu_B(x_0) = 0.25$

The membership for a 50-year old in the set "young" depends on one's own view. The grade of membership is a precise, but subjective measure that depends on the context. A fuzzy membership function is different from a statistical probability distribution.

In principle any function describes a membership function associated with a fuzzy set that depends not only on the concept to be represented, but also on the context in which it is used. The graphs of the functions may have different shapes and may have specific properties. Whether a particular shape is suitable can be determined only in the application context. In certain cases, however, the meaning semantics captured by fuzzy sets is not too sensitive to variations in the shape, and simple functions are convenient. In many practical instances fuzzy sets can be represented explicitly by families of parameterized functions, the most common being the following (Sumathi and Surekha 2010: 212):

1) Triangular Membership Function

The membership definition for a triangular function is given as Equation 2.8.

$$triangular(x:a,b,c)=\begin{cases}0 & x<a\\ (x-a)/(b-a) & a\le x<b\\ (c-x)/(c-b) & b\le x\le c\\ 0 & x>c\end{cases} \tag{2.8}$$

Where *b* is a modal value, *a* and *c* denote the lower and upper bounds, respectively, for nonzero values of triangular(*x*). For example, if a = 0, b = 5, and c = 10, the graphs of the functions is shown as Figure 2.9.

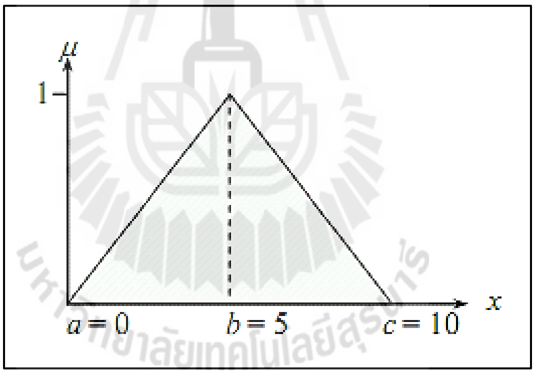


Figure 2.9 Triangular membership functions

2) Trapezoidal Membership Function

The membership definition for a trapezoidal function is given as Equation 2.9.

$$trapezoidal(x:a,b,c,d)=\begin{cases}0 & x<a\\ (x-a)/(b-a) & a\le x<b\\ 1 & b\le x<c\\ (d-x)/(d-c) & c\le x<d\\ 0 & x\ge d\end{cases} \tag{2.9}$$

For example, if a = 0, b = 2, c = 8, and d = 10, the graphs of the functions is shown as Figure 2.10.

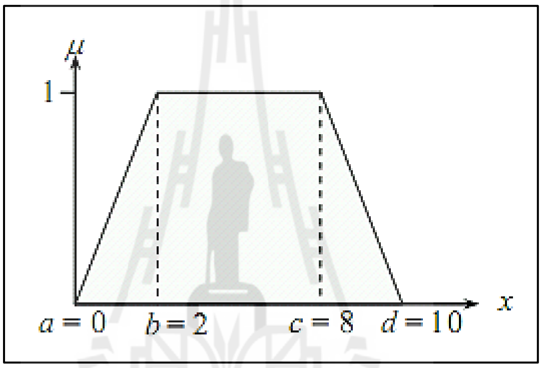


Figure 2.10 Trapezoidal membership functions

3) Gaussian Membership Function

The membership definition for Gaussian function is given as Equation 2.10.

$$Guassian(x:m,\sigma)=\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \tag{2.10}$$

For example, if m = 5 and σ = 1, the graphs of the functions is shown as Figure 2.11.

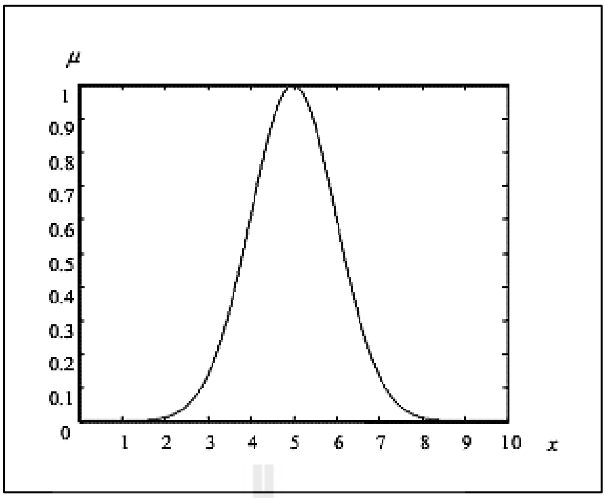


Figure 2.11 Gaussian membership functions

4) Bell-shaped Membership Function

The membership definition for a bell-shaped function is given as Equation 2.11.

$$bell - shaped(x : a, b, c) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}} \tag{2.11}$$

For example, if a = 2, b = 4, and c = 5, the graphs of the functions is shown as Figure 2.12.

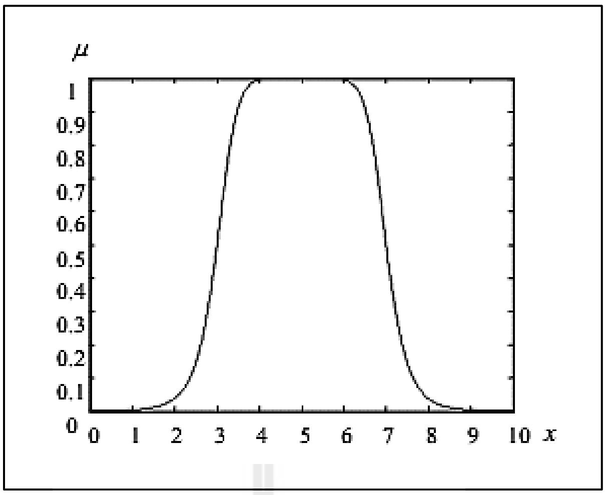


Figure 2.12 Bell-shaped membership functions

5) Smooth Membership Function

The membership definition for a smooth or S-function is given as Equation 2.12.

$$S(x:a,b)=\begin{cases} 0 & x<a \\ 2\left(\dfrac{x-b}{b-a}\right)^2 & a\le x<\dfrac{a+b}{2} \\ 1-2\left(\dfrac{x-b}{b-a}\right)^2 & \dfrac{a+b}{2}\le x<b \\ 1 & x\ge b \end{cases} \tag{2.12}$$

For example, if a = 2 and b = 8, the graphs of the functions is shown as Figure 2.13.

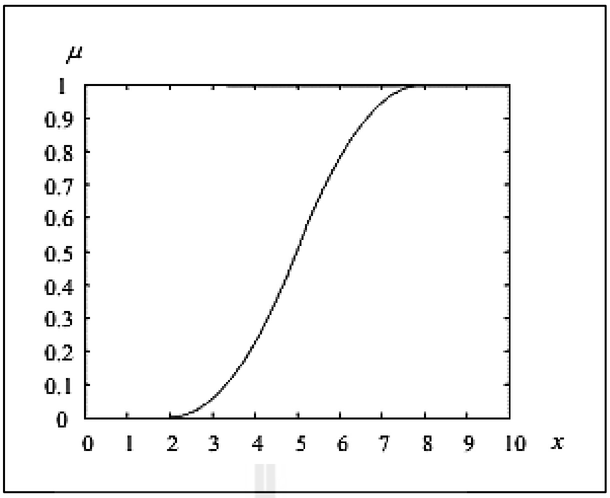


Figure 2.13 Smooth membership functions

6) Z-membership Function

The membership definition for a Z-function is given as Equation 2.13.

$$Z(x:a,b)=\begin{cases} 1 & x \le a \\ 1-2\left(\dfrac{x-a}{b-a}\right)^2 & a \le x \le \dfrac{a+b}{2} \\ 2\left(b-\dfrac{x}{b-a}\right)^2 & \dfrac{a+b}{2} \le x \le b \\ 0 & x \ge b \end{cases} \tag{2.13}$$

For example, if a = 2 and b = 8, the graphs of the functions is shown as Figure 2.14.

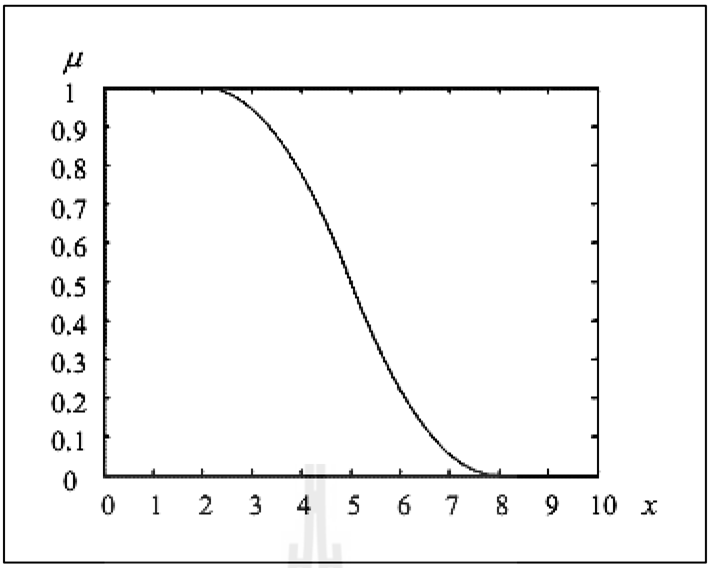


Figure 2.14 Z-membership functions

In this research, triangular membership functions are used for determining the grade of membership of the tourists' satisfaction level.

2.5.7 Fuzzy Rules

Sumathi and Surekha (2010: 261-262) stated that for any fuzzy logic operation, the output is obtained from the crisp input by the process of fuzzification and defuzzification. These processes involve the usage of rules, which form the basis to obtain the fuzzy output. A fuzzy if-then rule is also known as fuzzy rule, or fuzzy conditional statement or fuzzy implication. It is generally of the form

IF (*x is A*) AND (*y is B*), THEN (*z is Z*).

Where *x*, *y*, *z* represent the variables and *A*, *B*, *Z* are the linguistic values in the universe of discourse. Here the IF part is referred to as the antecedent or premise and the THEN part is referred to as consequent or conclusion. AND is the

Boolean operator which connects two or more antecedents. These fuzzy rules are multi-valued.

An individual fuzzy rule base can process more than one rule. Based on several set of rules an overall decision can be made from the individual consequents. This process of obtaining the overall decision is known as aggregation of rules.

Fuzzy rules are most commonly applied to control systems. The common types of fuzzy rules applied to control systems are the Mamdani fuzzy rules and Takagi-Sugeno (TS) fuzzy rules.

2.5.8 Fuzzy Inference

Fuzzy Inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned. The process of fuzzy inference involves all the topics such as fuzzification, defuzzification, implication, and aggregation (Sumathi and Surekha, 2010: 261, 278-281).

1) Fuzzification

The process by which the input values from sensors are scaled and mapped to the domain of fuzzy variables is known as fuzzification. The fuzzy variables also known as linguistic variables are determined based on intuition (from knowledge) or inference (known facts). These linguistic variables can be either continuous or discrete theoretically, but in practice it should be discrete. Fuzzification comprises two processes: 1) assigning fuzzy labels to the crisp input and 2) assigning numerical meaning (or range of the input value) to each label.

2) Combination

The Boolean operators such as “AND” (Min) and “OR” (Max) are used as connectives in the fuzzy rules. These operators are known as fuzzy combination operators since they are used to combine more than one antecedent part. Though the fuzzy AND/OR is similar to the Boolean AND/OR the difference is that, the Boolean AND/OR can perform with only 0 and 1, while the fuzzy AND/OR performs for the number between 0 and 1.

3) Implication or Consequence

Using the combination operators AND/OR the fuzzified inputs are combined and the rule strengths are determined. Then the output membership functions are clipped at the rule strength to obtain the consequence. The rule consequence is correlated with the truth value of the antecedent by cutting the consequent membership function at the level of the antecedent truth. This process is referred to as clipping or alpha cut. The top most membership functions is cut, therefore some information loss occurs. In order to preserve the unique shape of the membership function scaling is preferred over clipping. In scaling the degree of the membership function of the rule consequent is multiplied by the truth value of the antecedent, thus reducing the loss of information.

4) Aggregation

The process of unification of the outputs of all the rules is known as aggregation. The clipped or scaled membership functions are combined into a single fuzzy set. Each individual fuzzy rule yields a consequence, from which the overall output is to be computed. All the consequences are aggregated by using the “AND” or "OR" connectives. The process of aggregating the rules using "AND" connective is known as conjunctive aggregation and the process of aggregating the rules using "OR" connective is known as disjunctive aggregation.

5) Defuzzification

This is the final step of the inference method. In order to obtain a crisp output number several defuzzification methods can be used. The aggregated output from the previous step acts as the input to the defuzzification module and outputs a single crisp number. Though there are several defuzzification methods such as max membership principle, weighted average, centroid, center of sums, and max mean, and the most commonly used method is the centroid method. This method computes a defuzzification value which slices the aggregate set into two equal parts. Mathematically the defuzzified value is computed as

$$\frac{\int \mu_A(z) z dz}{\int \mu_A(z) z dz}.$$

Where z^* is the defuzzified output, $\mu_A(z)$ is the aggregated membership function, and z is the output variable.

2.6 Tourists' Satisfaction

This study used accommodation reviews of tourists as a case study because of the tourists' satisfaction concept is an important issue and it has considerably influenced tourism industry. Therefore, the concepts of tourist's interest and preference including related researches are reviewed for examining the factors influencing on the choices of tourist to make decision on their destination and utilizing these data for developing a module interface to meet the tourists' needs.

2.6.1 Definition of Satisfaction

In marketing, Sereerat et al. (1995: 11) mentioned that a customers' satisfaction is "*the level of customer feelings of expectation compared to quality of the products and services.*"

Lorpraditpong (2006: 26) concluded the meaning of customers' satisfaction is "*the feeling of customer (both satisfy and dissatisfy) arises from the comparison between the perceived performance of products or services and customer expectation*", as shown in Figure 2.15.

Customers' Satisfaction = Expectation - Perceived Performance

Figure 2.15 The customers' satisfaction equation

Furthermore, Lorpraditpong also defined the customers' loyalty is a customers' attitude towards products and services in the long term relationship. Loyalty comes primarily from a customers' emotional connection and experiences with an organization. The behavior of loyal or repeat customers is a result of satisfaction. Therefore, the organization should satisfy customers' needs continuously in order to make customers' royalty.

Moreover, customers' loyalty is driven by customers' experience with products and processes. There is differentiating between customers' loyalty and satisfaction. A customers' satisfaction is a kind of dynamic which always changes all the time whereas customers' loyalty is more static since they have been impressed by the products and services. This cause change less than customers' satisfaction.

In the area of tourism, tourists' satisfaction is a result of the interaction between a tourists' experience at the destination area (including quality of services) and the expectations they had about that destination. Notice that customers' satisfaction depend on their expectation and this expectation is individual. Thus, the expectation very affects the tourists' subjective perception (Hui, Wan, and Ho, 2007: 966). In addition, tourists' satisfaction can be assessed through these factors: perceived quality, safety and risk, reduction, novelty, destination competitiveness and past experience, and destination image (Yoon and Uysal, 2005: 47-48). Nevertheless, tourists are going to travel or not depend on motivations which are divided into four categories described below (Macintosh and Goeldner, 1986: 31-34).

1) Physical motivation: refreshment of body and mind (rest and relaxation) – beach holidays, lakes, and mountains, etc.; for health purposes (e.g. either medically prescribed or undertaken voluntarily) – spas, etc.; for participation in sports – skiing, canoeing, safari parks, pony-trekking, etc.; pleasure – fun, excitement, romance and entertainment, to shop.

2) Culture motivation: curiosity about foreign countries, people and places; interests in art, music, architecture, folklore – music festivals, theatre visits, etc.; interest in historical places (remains, monuments, churches); experiencing specific international and national events – Olympic Games, Oktoberfest, etc.

3) Interpersonal motivation: visiting relatives and friends; meeting new people and seeking new friendships; seeking new and different experiences in different environments – sailing, etc.; escaping from one's own permanent social environment (e.g. desire for a change); personal excitement of traveling; visiting places and people for spiritual reasons (e.g. pilgrimages); traveling for travel's sake.

4) Status and prestige motivation: pursuit of hobbies – craft or painting holidays, etc.; continuation of education or learning – study tours, etc.; seeking of business contacts and professional goals – fairs, etc.; conference and meetings; ego enhancement and sensual indulgence; fashion.

In conclusion, customers' satisfaction, in the area of marketing, is sensibility toward products and services compare to expectation. The customers' satisfaction and loyalty always go to the same direction. In other words, if organizations satisfy the need of customer continually with their products or services, customers' loyalty will be increased gradually. In the area of tourism industry, tourists' satisfaction is an emotional and sensational reward towards quality of services. It also depends on level of tourists' expectation like marketing area. Finally, the factor influence consumer's satisfaction is the past travel experiences which make tourists have higher level of expectation and may affect their attitude of the tourist destination.

2.6.2 Significance of Tourists' Satisfaction

According to the meaning of tourists' satisfaction, tourists make a decision on selecting products or services upon the quality of service staff and the satisfaction of products. They satisfy with products and services more than the reason of business competitors, price and quality, or other factors. On the contrary, they decide to quit buying products or using services easily, if they are dissatisfied. Can observe that, the organization do not compete with each other but oneself and customer, in terms of pursuing customers' need. They have to respond to this challenge. Hence, customers' satisfaction management is the important issue that organization must pay attention, especially in service business, for example, tourism industry. To have a competitive advantage, any organization have better customers' satisfaction measuring and benchmarking systems, they can manage customers' requirement immediately, respond to endless requests and better manage the different customers' needs and gain more market share. In addition, the importance of tourists' satisfaction is an asset to the company which has to collect and build all the time to gain competitive advantage in business. (Lorpraditpong, 2006: 17)

2.6.3 The Literature Reviews of Tourists' Satisfaction

This study focuses on the design of an analysis module for online travel accommodation reviews. In order to differentiate context and calculate the tourists' satisfaction level toward accommodation, the theories of tourists' satisfaction including related research are reviewed.

Pizam, Neumann, and Reichel (1978) empirically identified eight factors of tourists' satisfaction with Cape Cod, Massachusetts (USA) as a tourist destination area and suggested the means to measure them. By using a factor-analytic approach based on data obtained from a survey of 685 vacationing tourists, the following factors of tourists' satisfaction were derived: beach opportunities, cost, hospitality, eating and drinking facilities, accommodation facilities, environment, and extent of commercialization.

Chi and Qu (2008) offered an integrated approach to understanding destination loyalty by examining the theoretical and empirical evidence on the causal relationships among destination image, tourist attribute and overall satisfaction, and destination loyalty. The results supported the proposed destination loyalty model: (1) destination image directly influenced attribute satisfaction, (2) destination image and attribute satisfaction were both direct antecedents of overall satisfaction, and (3) overall satisfaction and attribute satisfaction in turn had direct and positive impact on destination loyalty.

Cracolici and Nijkamp (2008) assessed competitive attractiveness on the basis of individual visitors' perceptions regarding a holiday destination in the southern regions of Italy. They use individual survey questionnaires on the tourists' evaluation of the quality of tourist facilities and attributes in a given area (the regional tourist profile) as the basis for constructing an aggregate expression for the relative attractiveness of that area. The data were used to analyze tourist attractiveness and to identify policies for improving regional tourist competitiveness. As a result, the data show strengths, weaknesses, opportunity, and threats of the destination image. Also, in terms of tourists attracted (i.e. a quantitative performance) and tourists' satisfaction (i.e. a qualitative performance), simultaneously. The results show that a good quantitative performance is not always linked to an excellent qualitative performance.

Bosque and Martin (2008) explored the cognitive-affective nature of destination image and the role of psychological factors in its formation. As a result, tourists make their choices of where to travel based on the destination images being portrayed by the destination marketers but also based on their own images of a destination that come from many different sources including sometimes and past experiences with a destination.

Hsu, Tsai, and Wu (2009) identified the factors that influence the tourists' choice of destination and evaluated the preferences of tourists for destinations. A 4-level AHP model was proposed and tested using data collected from tourists visiting Taiwan to establish the relative importance of pre-selected factors (criteria). The results indicate that visiting friends/relatives and personal safety appear to be the 2 most important factors for inbound tourists to Taiwan, price is the least important and Taipei 101 is the first priority for travelers.

The design of an analysis module for online travel accommodation reviews aims to bring useful information about tourists' satisfaction to organization. The business entrepreneur can apply the extracted knowledge for developing their accommodation services and facilities in order to meet more customers' need and gain more advantages over the competitors. In addition, other tourists can apply the extracted knowledge from experienced tourists to support their decisions on selecting travel accommodations.

2.6.4 The Basis for Accommodation Selection

In order to develop an online review analysis module, the tourist criterions for choosing travel accommodation are reviewed. These criterions are useful for creating user friendly interface which satisfy the needs of tourists and available for making decision (Davenport and Prusak, 1998).

The related literature on factors influenced the tourists' choice of destination indicated that there are various factors in which tourists' interested and can be used as indicators of tourists' satisfaction. Therefore, the related and proper factors are selected for categorizing contents of tourists' opinions, e.g. services, environment, room, food, activity, price, security, cleanness, and transportation convenience (Hsu, Tsai, and Wu, 2009: 295; Cracolici and Nijkamp, 2008: 342; Chi and Qu, 2008: 629-631; Bosque and Martin, 2008: 562). Finally, the knowledge explanation of the designed module reveals the calculated scores of tourists' satisfaction on each factor in details.

2.7 Online Review

The objective of this research is to design an analysis module for online reviews, which uses the tourism information on travel accommodations as a case study. Thus, the significance of review and the literature of review analysis are reviewed and used as a guideline for designing an algorithm of the module.

A travel accommodation review in this study means an opinion of experienced tourists about travel accommodation visiting, which can be called tourists' opinion, tourist review, tourists' suggestion, and so on. According to the given meaning of review stated above, a travel accommodation review is information that expresses the tourists' satisfaction on product or service in tourism business.

Besides the tourists' satisfaction plays an important role in the organization as previously mentioned, the satisfaction of the experienced tourists also affect the decision of prospective buyers or other tourists in selecting goods or services. Due to the advancement of the information and communication technology, travel-related organizations have transformed their own organizations into electronic business for encouraging tourists to pay for their products or services. Moreover, this advancement also changes the most common method that tourists used in selecting the tourist places to the data analysis from online reviews of experienced tourists. The finding of global online consumer reliability survey (Nielsen, www, 2009: 3) and online reviews trust (Anderson, www, 2010: 1) indicate that there are more reviews and opinions given on the website and consumers tend to trust online opinions as much as personal recommendations. Therefore, the online reviews are not only contribute to prospective buyers' decision but also used as reflection to see the weaknesses of business.

The example of tourism websites that allow tourists to write a review, e.g. TripAdvisor.com, Agoda.com, and Booking.com are shown in Figure 2.16-2.18. Unfortunately, tourists must read enormous reviews in order to select the preferred tourist destinations. Thus, this study aims to design an analysis module for online reviews for extracting useful information that tourists can apply to support their decisions on selecting travel accommodations easily and quickly.

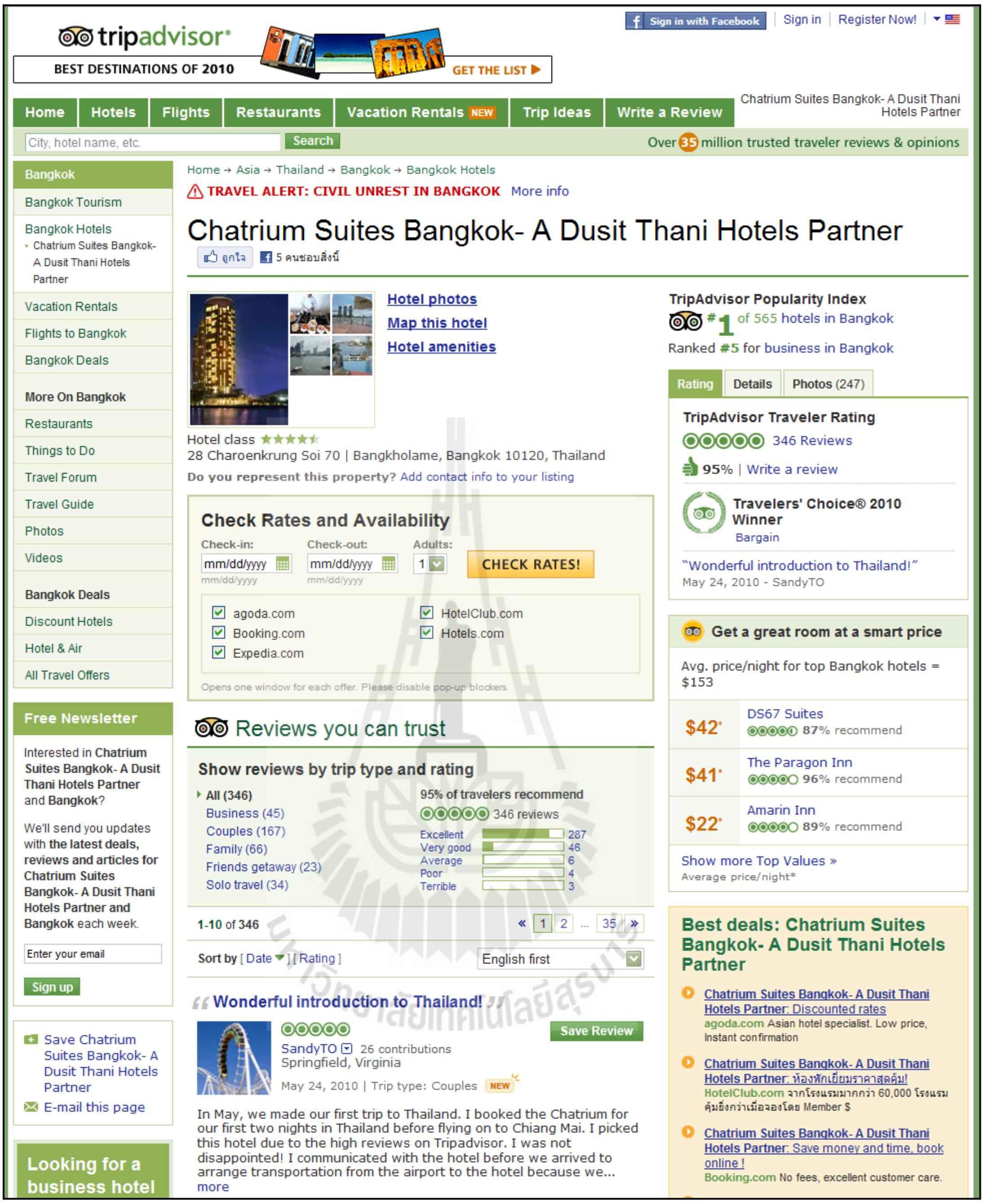


Figure 2.16 TripAdvisor.com

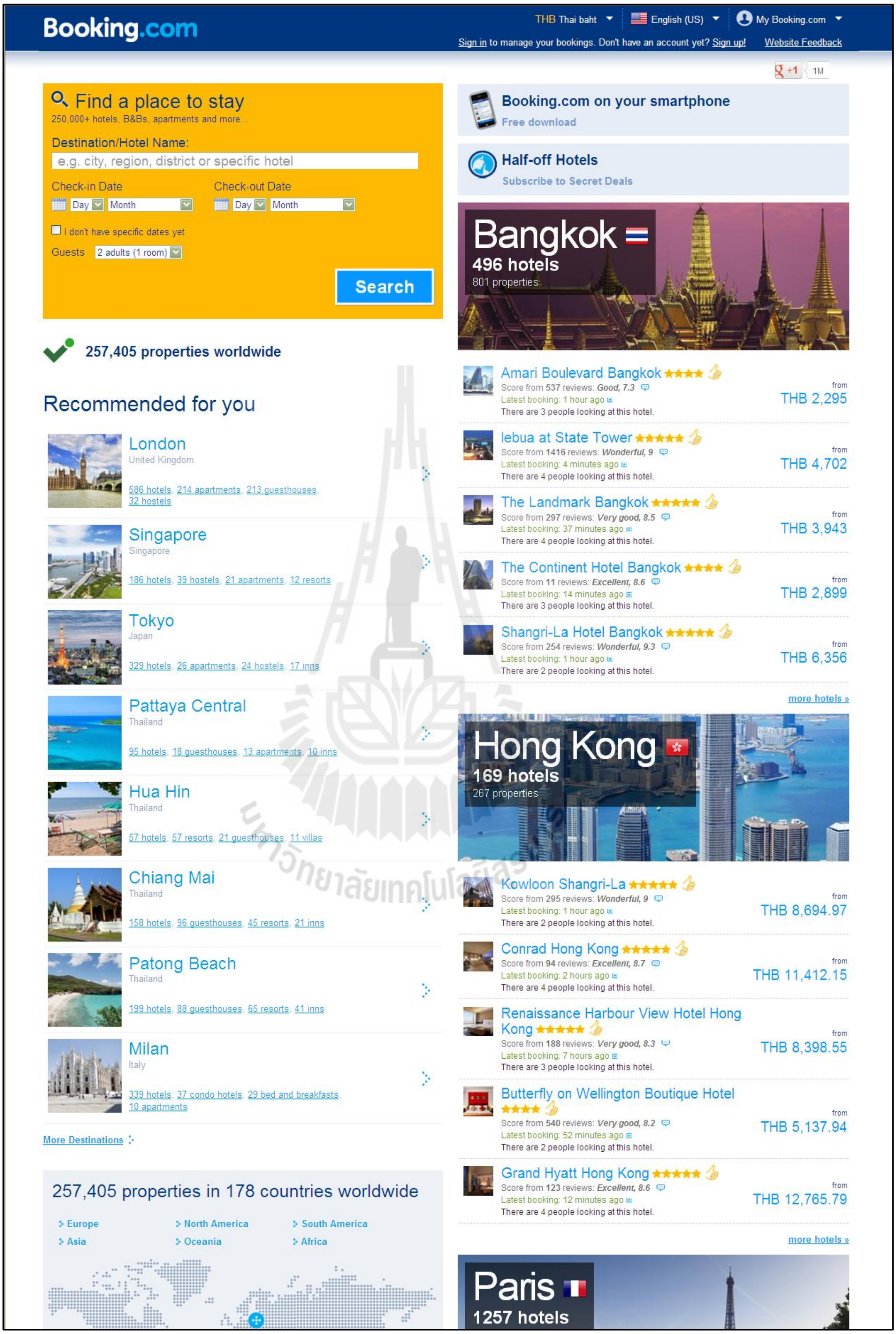


Figure 2.17 Booking.com

Figure 2.18 Agoda.com

2.8 Related Work

Turney (2002) created a simple unsupervised learning algorithm for classifying reviews as recommended (thumbs up) or not recommended (thumbs down). The algorithm classifies a review using the average semantic orientation of phrases in the review that contain adjectives or adverbs. A review is classified as recommended if the average semantic orientation of its phrases is positive. The algorithm achieves an average accuracy of 74% when evaluated on 410 reviews from Epinions.com.

Dave, Lawrence, and Pennock (2003) identified the unique properties of products and developed a method or a classifier for automatically distinguishing between positive and negative reviews. The classifier draws on information retrieval techniques for applying in feature extraction and scoring, and the results for various metrics and heuristics depending on the testing situation. The best methods work better than or equal to traditional machine learning. When operating on individual sentences collected from web searches, performance is limited due to noise and ambiguity. But in the context of a complete web-based tool and aided by a simple method for grouping sentences into attributes, the results are qualitatively quite useful.

Taboada and Grieve (2004) classified texts automatically based on a basic method for extracting a semantic orientation by taking into account the position for each of the adjectives in the text. The performance of the proposed system varies depending on the type of review under consideration. The Appraisal values are extracted for each of the reviews, also revealing different characteristics according to review type.

Whitman and Ellis (2004) analyzed a large testbed of music and a corpus of reviews for each work to uncover patterns and develop mechanisms for removing reviewer bias and extraneous non-musical discussion. This research showed analysis frameworks and results on learning the crucial relation between review texts and the music they describe using *TF-IDF* computation and regularized least-squares classification.

Zheng and Ye (2009) applied the supervised machine learning algorithm of SVM in sentiment classification for the online Chinese reviews on about forty hotels from Tianjin City and Chongqing City. The experimental results indicated that the SVM classifier, which has been proved well applied to the sentiment classification for traveler reviews written in English, also performs very well at classifying traveler reviews written in Chinese. Additionally, due to the recent study suggests that online traveler reviews have an important impact on online hotel bookings in China. Thereby this research can undoubtedly contribute to business performance of hotels in China if it could be applied to practice in the future.

Zhang, Narayanan, and Choudhary (2010) mined online customer reviews for product feature-based ranking by identifying subjective and comparative sentences in reviews and using a directed graph to determine the relative quality of products.

Ramkumar, Rajasekar, and Swamynathan (2010) scored products from online reviews using fuzzy logic to calculate the spam level scores of each review and the scores for each feature of a product.

Jakob and Gurevych (2010) extended an opinion mining algorithm with rule-based anaphora resolution algorithm, called CogNIAC, to improve opinion target identification in movie reviews. However, this algorithm does not yield high precision when resolving impersonal and demonstrative pronouns.

Hu and Liu (2004) mined and summarized online product reviews based on data mining and natural language processing methods including various techniques such as Part-of-Speech tagging, frequent feature identification, opinion word extraction and predicting the orientations of opinion sentences. It revealed a total number of positive and negative reviews for each product feature to users. The objective of this research is to produce a feature-based summary of a large number of customer reviews of a product sold online.

As discussed above, Table 2.4 illustrates a comparison of researches related to online review analysis. Most of them are applied with product reviews. They present a product rating based on their features in the form of binary scores, such as "positive/negative", "recommend/don't recommend", or "yes/no". Typically, tourists are interested in knowing the strength of opinion about an accommodation, therefore only a "positive/negative" binary score seems insufficient. It would be vastly preferable if tourists could give the accommodation a numeric score or at least grade it from a list of qualitative ratings (i.e. 5 means excellent, 4 means good, 3 means average, 2 means poor, and 1 means terrible) (Cesarano et al., 2006: 1). According to the findings of online consumers' or shoppers' requirements, the details and relevant product information and explanations are needed for decision making by consumers in order to select products or services (Burke, 2002: 411). Tourists also want to know about the accommodation in details (e.g. How about bed or air condition in the room? or How about services or cleanliness of the accommodation?) and use both rating scores and accommodation information for selecting their preferred accommodations promptly and efficiently.

Table 2.4 A comparison of researches related to online review analysis

Online Review Analysis	Related Work									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	*****
Feature extraction										
Feature scoring analysis										
- Association rule mining									√	
- Term frequency				√						
- Term frequency-inverse document frequency				√						√
- Information gain		√			√					
- Mutual information	√		√				√			
- Poisson distribution							√			
- Edit distance method							√			
- Frequency counts								√		
Morphological and Syntactic analysis										
- Part-of-speech tagging	√	√	√	√		√		√	√	√
- Context free grammar										√
- N-grams		√		√			√			
- Substrings		√					√			
- Lexical chunker				√						
- Named entity recognizer								√		
Semantic analysis										
- Semantic orientation scoring	√									
- Semantic similarity matching						√	√		√	
- Semantic rating										√
Discourse analysis										
- Anaphora resolution								√		√

*Related work: 1 = Turney (2002); 2 = Dave, Lawrence, and Pennock (2003); 3 = Taboada and Grieve (2004); 4 = Whitman and Ellis (2004); 5 = Zheng and Ye (2009); 6 = Zhang, Narayanan, and Choudhary (2010); 7 = Ramkumar, Rajasekar, and Swamynathan (2010); 8 = Jakob and Gurevych (2010); 9 = Hu and Liu (2004); * = This research*

Table 2.4 A comparison of researches related to online review analysis (Continued)

Online Review Analysis	Related Work									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	*****
Rating technique										
Hierarchical calculation										√
Fuzzy logic							√			√
Graph						√		√		
Machine learning										
- Support vector machine				√	√					
- Regularized least-squares				√						
- Naïve bayes		√								
Knowledge base										
Ontology										√
Thesaurus or Synonym set						√				
Terminology										
- Feature set or Opinion target list						√		√	√	
- Opinion word set								√	√	
- Positive-Negative word set	√					√	√			
- Adjective word set			√	√						√
- Adverb word set										√
- Negation word set						√				√
- Verb word set										√
- Noun phrases set				√						
- Comparative keyword						√				
Dictionary										√
WordNet		√							√	
Knowledge explanation										
Present reviews as positive/negative	√		√		√	√				
Present features as positive/negative		√							√	
Present products ranking by features						√			√	
Present reviews as 5-rating scale										√
Present features as 5-rating scale										√
Present features in hierarchy										√

*Related work: 1 = Turney (2002); 2 = Dave, Lawrence, and Pennock (2003); 3 = Taboada and Grieve (2004); 4 = Whitman and Ellis (2004); 5 = Zheng and Ye (2009); 6 = Zhang, Narayanan, and Choudhary (2010); 7 = Ramkumar, Rajasekar, and Swamynathan (2010); 8 = Jakob and Gurevych (2010); 9 = Hu and Liu (2004); * = This research*

Hence, this research presents the design of an analysis module for online travel accommodation reviews. The designed module focuses on a design of a semantic analysis approach (named "feature extraction process") for digesting accommodation features, feature modifiers, and relationship among these features from review contents. This feature extraction process applies the term frequency-inverse document frequency (or *TF-IDF*) weighting approach for feature scoring analysis. There are two reasons for selecting the *TF-IDF* approach as follows: 1) there are many evidences that this approach yields an efficient result for selecting a word with common characteristic of interest (or feature) in a document which is stored in a collection or corpus and 2) the *TF-IDF* approach can identify feature words in the same topic of interest (or class). This research aims to design a module that extracts features and calculates ratings from online reviews in the topic of travel accommodation. In addition, the final result of this module, i.e. the accommodation rating, is differentiated from 5-rating scale (or classes). Thus, the *TF-IDF* approach can be practically applied to the online review dataset of this research because these reviews are in the context of travel accommodation only.

Moreover, the part-of-speech tagging and context free grammar approach are applied to the morphological and syntactic analysis in the feature extraction process. As a matter of course, the part-of-speech tagging is a fundamental process to understand natural language. It is certainly a part of this research because of the suitability of input that is unstructured texts stored in the tourist reviews. Also, the context free grammar is an appropriate technique for analyzing the structure of sentences owing to the flexibility to rewrite an appropriate rule for analyzing the language usage based on individual expression which is an arbitrary distinction and not always in grammatical correction.

As previously described, this research aims to design a review analysis module that can interpret the meaning from online travel accommodation reviews and calculate the rating of tourists' satisfaction based on these meanings; in other words, the semantic analysis process of this research applies a semantic rating approach in the travel accommodation contexts. Furthermore, an anaphora resolution is used for language understanding in the semantic analysis process because it is one of techniques in the natural language processing, which can be applied to the review analysis module and can enhance the accuracy of results.

After digesting accommodation features, feature modifiers, and relationships among these features from review contents, all of these are used as inputs to the accommodation rating process. This research proposes methods for calculating a tourists' satisfaction level on each extracted feature and on the entire review by using the fuzzy logic technique. The fuzzy logic technique is multi-valued logics that form a suitable basis for logical systems reasoning under uncertainly or vagueness of tourist criticisms. Moreover, it allows calculating a tourists' satisfaction level on each feature and on the entire review whereas graph and machine learning techniques can only calculate on the entire review. Thus, the fuzzy logic is the most appropriate technique for this designed module according to the research objective. In addition, the designed module also explains more clear and easy understanding reviews with 5-rating scale and visualizes the accommodation feature relationships in hierarchy. These knowledge explanations conform to the research objective which is derived from the findings of online consumers' or shoppers' requirements. Hence, applying the hierarchical calculation for tourist rating and presenting them in 5-rating scale on both features and the entire review are proposed more details for decision making by tourists.

Finally, there are various resources for designing a review analysis module. This research uses a dictionary to collect frequently-used vocabularies that will be applied for syntactic analysis, and uses a tourism ontology to store the specific knowledge, such as accommodation features including their feature relationships. This tourism ontology is used to support the feature extraction, accommodation rating, and knowledge representation processes of the designed review analysis module. Furthermore, it is easy to attach some additional resources, such as a thesaurus or synonym set (named SKOS ontology) and other tourism websites can use or join the designed tourism ontology together which can assist tourists to search for information with various queries in the future. By the way, the commentary words used in the designed module are separately collected by types of criticism in order to calculate more precise rating. Notice that these commentary words are not included in the tourism ontology because they are used to explain the accommodation features in general, not specify any particular feature.

CHAPTER 3

RESEARCH PROCEDURE

The purpose of this chapter is to describe how the study will be carried out. This chapter explains the research methodology including problem analysis, knowledge base design, module design, and module evaluation. Then, research instruments are presented. Finally, this chapter describes how to analyze and interpret the obtained data.

3.1 Research Methodology

This research is an applied research which applies the System Development Life Cycle (SDLC) for developing an analysis module for online travel accommodation reviews. There are four processes as shown in Figure 3.1.

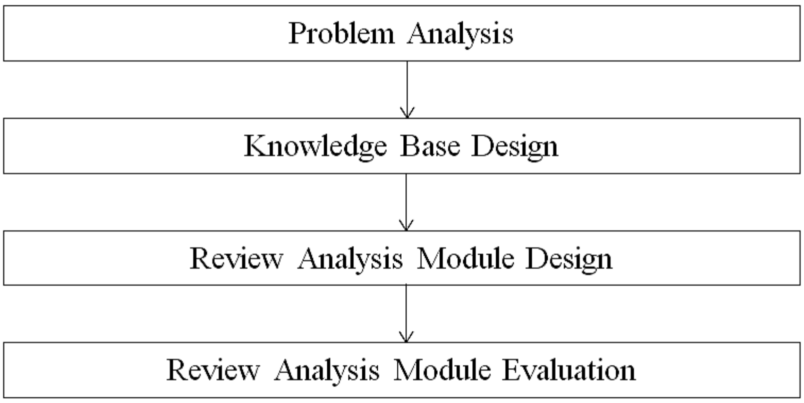


Figure 3.1 Research methodology

3.1.1 Problem Analysis

As stated in the Chapter 1, the most common method that tourists use in selecting the tourist places is the decision based on comments and reviews from online reviews of experienced tourists. Although there are many tourism websites that can write a review and store the reviews of experienced tourists about tourist places, tourists cannot utilize those data for supporting their decisions immediately. For example, if tourists want to know about these tourist places in details, they must read enormous reviews by themselves in order to select the preferred tourist places. Moreover, the existing review analysis modules extract only accommodation features (such as price, value, cleanliness, service, etc.), but they do not extract relationships between those features. In addition, the feature scoring methods of the existing modules are illustrated in the form of binary-rating scale (e.g. “positive/negative” or “recommend/don’t recommend”) that seems insufficient to know the strength of opinions and support tourist’s decisions as pointed out in the Chapter 2.

3.1.2 Knowledge Base Design

The knowledge base in this research is used to store extracted knowledge including necessary information for automatic information extraction. There are four components as follows:

- **Tourism Ontology**

The tourism ontology used in this research was revised from a class hierarchy of the E-tourism ontology version 8 (Siorpaes, Prantner, and Bachlechner, 2004). The purpose of modifying the ontology is to support the designed review analysis module. The new ontology revision is performed by analyzing accommodation features from 400 accommodation reviews using the *Rocchio's TF-IDF* weighting approach (Salton and Buckley, 1988: 517). Given a review collection D, a word w, and an individual review $d \in D$, the *TF-IDF* weighting approach are calculated as Equation 3.1.

$$W_d = f_{w,d} \times \log(|D| / f_{w,D}) \quad (3.1)$$

Where W_d is the weight of w appears in d, $f_{w,d}$ equals the number of times w appears in d, $|D|$ is the size of the review collection, and $f_{w,D}$ equals the number of reviews in which w appears in D.

Afterwards, the word w with the weight more than a threshold is selected as the accommodation feature. All the selected features are added to the tourism ontology. Figure 3.2 illustrates a structure of the tourism ontology used in this research by improving classes and subclasses on the previous E-tourism ontology. For instance, the new subclasses of Accommodation class are added to the tourism ontology consisting of Homestay, Hotel, Park, and Resort. Figure 3.3 depicts the properties of Accommodation class which are extracted from reviews by the *TF-IDF* weighting approach.

As a result, the revised ontology consists of 10 classes and 95 key properties including their relationships. The knowledge in the tourism ontology will be applied for the feature extraction process and the accommodation rating process in the knowledge inference engine of the review analysis module. According to the classes and properties relationships derived from the structure of the revised ontology, the designed module has an ability to extract feature relationships and also present them in hierarchy. This ability performs tourists to know that some vague features belong to which accommodation properties. For example, the feature “cleanness” is vague because it may belong to room or bathroom. When the vague features are criticized, the designed module would extract the features according to the structure of ontology and obviously reveal which accommodation properties that the features belong to.

1) **Accommodation**
2) **Activity**
3) **ContactData**
4) **DateTime**
    - OpeningHours
    - Period
        - DatePeriod
        - TimePeriod
    - Season
5) **Event**
6) **Infrastructure**
7) **Location**
    - GPSCoordinates
    - PostalAddress
8) **Room**
    - ConferenceRoom
    - Guestroom
9) **Ticket**

a) E-tourism Ontology

(*Siorpaes, Prantner, and Bachlechner, 2004*)

1) **Accommodation**
    - AccommodationType
        - HomeStay
        - Hotel
        - Park
        - Resort
    - Bathroom
    - Room
        - ConferenceRoom
        - Guestroom
2) **Activity**
    - Adventure
    - Relaxation
    - Sightseeing
    - Sport
3) **Amenity**
    - CarServices
    - EmergencyServices
    - Resturant
    - ServiceShops
4) **Attraction**
    - AgroTourism
    - CulturalTourism
    - HealthTourism
    - HistoricalTourism
    - NaturalTourism
5) **ContactData**
6) **DateTime**
7) **Event**
8) **Location**
    - GPSCoordinates
    - PostalAddress
9) **Rating**
10) **Ticket**

b) Tourism Ontology

(*Used in this research*)

Figure 3.2 A structure of the tourism ontology revised from the E-tourism ontology

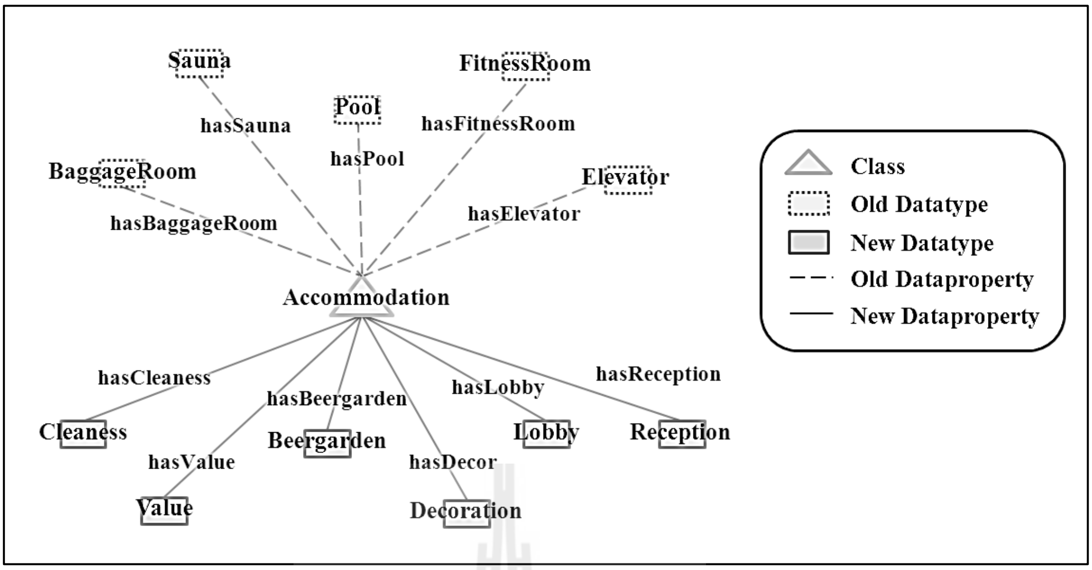


Figure 3.3 The properties of the tourism ontology revised from the E-tourism ontology

In addition, synonym words which have the same or very similar meanings, such as "fridge" and "cooler", "Air Conditioner" and "air condition" are also added to the SKOS ontology (World Wide Web Consortium, www, 2009) which is a part of the tourism ontology. These synonym words are collected from 400 accommodation reviews and required for a lexical analysis. For each synonym set, there is a designated word representing all synonym words.

Nevertheless, more details of the designed tourism ontology, i.e. classes, properties, and SKOS ontology are explained in Appendix A.

- **Dictionary**

The designed module uses a dictionary for the lexical and syntactic analysis. The module applies the LEXiTRON version 3.0 beta which is an online dictionary developed by the Human Language Technology Laboratory of Thailand's National Electronics and Computer Technology (NECTEC), Thailand since 2003

(Text Processing Technology, www, 2012). The LEXiTRON dictionary was originally constructed from a corpus which consists of frequently-used vocabularies in many topics from trusted publications. Currently, the database has more than 79,000 entries of English (Trakultaweekoon, Porkaew, and Supnithi, 2007: 43).

- **Terminology**

The designed module applies a terminology for the syntactic analysis and accommodation feature rating process. The terminology was created by analyzing keywords from 400 accommodation reviews as same as the tourism ontology. The terminology is a word collection assigned a satisfaction level in 5-rating scale for each word as shown in Figure 3.4. The assigned satisfaction level of words was confirmed by a language expert, where “rating = 1” implies terrible, “rating = 2” implies poor, “rating=3” implies average, “rating=4” implies good, and “rating=5” implies excellent. There are five types of words stored in the terminology as described below.

Figure 3.4 The 5-rating scale for each word

- ***Adjective***, each one is assigned a fixed satisfaction level such as “Excellent” = 5, “Effective” = 4, “Moderate” = 3, “Unfriendly” = 2, and “Awful” = 1.

- ***Special verb***, each one is assigned a fixed satisfaction level as same as an adjective, such as “Deteriorate” = 2 and “Work” = 4. Typically, verbs are not associated to any criticisms. However, some verbs can criticize an accommodation feature. For instance, the verb “deteriorate” in a sentence “The air quality has deteriorated.”. In this case, the verb “deteriorate” modifies the feature “air”.

- ***Special word***, each one is assigned a feature to which it implies and a fixed satisfaction level as same as an adjective. The special word can be noun, verb, adjective, or phrase and can be calculated without the feature word because it identifies the feature by itself. For instance, “Dirty” = Satisfaction level 2 and implies to the *Cleanliness* feature, “Walking distance” = Satisfaction level 4 and implies to the *Location* feature, etc.

- ***Adverb***, each one is assigned an adjustable rating such as “Very” = ±1, “So” = ±1, “Extremely” = ±1, “Most” = ±2, and so on. When this word type is calculated, a feature rating will obtain the same rating of adjective, that is, if tourists review in positive, the feature rating will increase.

- ***Negation adverb***, each one is assigned an adjustable rating as same as an adverb such as “Not” = ±2, “Almost” = ±1, “Never” = ±2, and so on. However, when this word type is calculated, a feature rating will contrast with the rating of adjective, that is, if tourists review is positive, the feature rating will decrease. Also, the satisfaction level of all word types are used in the knowledge extraction engine described later.

- **Context Free Grammar Rules**

The designed module applies a context free grammar (*CFG*) for the syntactic analysis and semantic analysis in the feature extraction process. The context free grammar is a set of rewrite rules that express the ways that symbols of the language can be grouped and ordered together (Fry, 2004: 1). An example of the context free grammar rules used in this research is shown in Figure 3.5.

AP -> ADJ ADJ	NP -> AP N	VP -> V AP
AP -> ADJ ADV	NP -> AP NP	VP -> V N
AP -> ADJ AP	NP -> comma N	VP -> V NP
AP -> ADJ CONJP	NP -> comma NP	VP -> V PP
AP -> ADJ PP	NP -> DET N	VP -> V PRON
AP -> ADV ADJ	NP -> DET NP	VP -> V VP
AP -> ADV ADV	NP -> N CONJP	VP -> VB ADJ
AP -> ADV AP	NP -> N CONJS	VP -> VB AP
AP -> ADV CONJP	NP -> N N	VP -> VB NP
AP -> AP CONJP	NP -> N PRON	VP -> VB V
AP -> comma ADJ	NP -> N NP	VP -> VB VP
AP -> comma ADV	NP -> N PP	VP -> VP ADV
AP -> comma AP	NP -> NP CONJP	VP -> VP AP
AP -> comma S	NP -> NP CONJS	VP -> VP N
AP -> NOT AP	NP -> NP PP	VP -> VP NP
AP -> NOT ADJ	NP -> NOT N	VP -> VP PP
PP -> P N	NP -> NOT NP	VP -> VP PRON
PP -> P NP	NP -> PRON CONJP	S -> comma S
PP -> P PRON	NP -> PRON CONJS	S -> CONJS S
PP -> P V	NP -> PRON N	S -> ADVT S
PP -> P VP	NP -> PRON NP	S -> IFS S
CONJP -> CONJ ADJ	NP -> PRON PP	S -> N V
CONJP -> CONJ ADV	VP -> ADV V	S -> N VP
CONJP -> CONJ AP	VP -> ADV VP	S -> NP V
CONJP -> CONJ AUXP	VP -> AUX N	S -> NP VP
CONJP -> CONJ NP	VP -> AUX NP	S -> PRON V
CONJP -> CONJ V	VP -> AUX PRON	S -> PRON VP
CONJP -> CONJ VP	VP -> AUX V	S -> S comma
CONJP -> CONJ N	VP -> AUX VP	S -> S CONJS
CONJP -> CONJ PRON	VP -> AUX AP	S -> S ADVT
CONJS -> CONJ S	VP -> comma VP	S -> S IFS
IFS -> IF S	VP -> NOT V	S -> S NP
NP -> ADJ N	VP -> NOT VP	S -> S S

Figure 3.5 An example of context free grammar rules

3.1.3 Review Analysis Module Design

The design of an analysis module for online travel accommodation reviews focuses on the semantic analysis of accommodation reviews in the English language by extracting accommodation information from reviews as accommodation features and calculating tourists' satisfaction levels on each accommodation feature and on the entire review. The extracted accommodation features and their satisfaction levels are stored in the knowledge base which will be retrieved later by other tourists. The module framework is depicted in Figure 3.6.

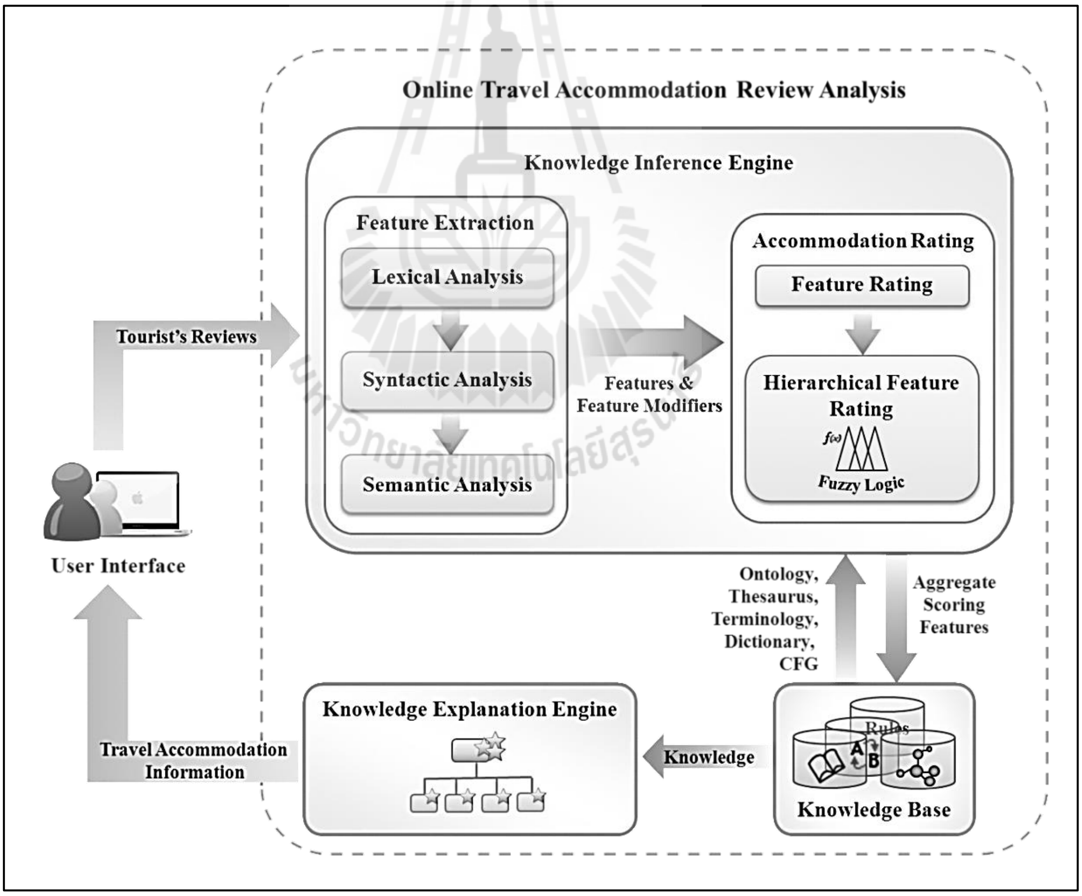
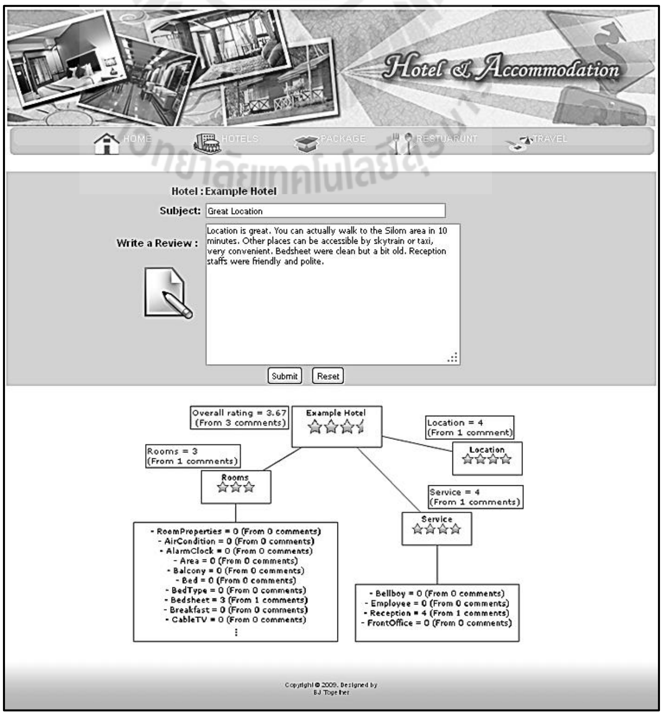


Figure 3.6 The framework of review analysis module

This designed module is an Internet-based application that is implemented with PHP, JavaScript, HTML, and other related web technologies. It consists of four components as follows: user interface, knowledge inference engine, knowledge base, and knowledge explanation engine.

1) User Interface

The user interface of an analysis module for online travel accommodation reviews is designed as a user-friendly graphic user interface (GUI). Tourists can access this analysis module via the user interface to write a review about a travel accommodation, which is used as input data for the module, and to examine an output of reviews summarization in a tree structure as shown in Figure 3.7. They can interact with the module through web browsers (e.g. searching for accommodation information).



1=*Terrible* 2=*Poor* 3=*Average* 4=*Good* 5=*Excellent*

Figure 3.7 Accommodation information of an online review analysis module

2) Knowledge Inference Engine

The knowledge inference engine performs two processes: the natural language parsing (named feature extraction process) and the tourists' satisfaction calculation (named accommodation rating process), as described below:

- **Feature Extraction**

The feature extraction is a process of digesting and selecting the significant keywords or features (noun), feature modifiers (adjective, special verb, special word, adverb, and negation adverb), and relationships among these features from review contents. These extracted features, feature modifiers, and feature relationships are used for calculating a tourists' satisfaction or accommodation rating described later. The feature extraction process is divided into three steps: lexical analysis, syntactic analysis, and semantic analysis.

– ***Lexical Analysis***

The lexical analysis performs word segmentation and eliminates symbols or special characters, such as acute accent (′), exclamation mark (!), at sign (@), and caret (^). Moreover, synonym words are transformed into the designated words of synonym sets in the SKOS ontology.

As depicted in Figure 3.8, a criticism "*The air conditioning was not very effective.*" is segmented by word space, and the synonym word "*air conditioning*" is transformed into the designated word "*aircon*". In addition, the lexical analysis involves basic correcting spelling and inflectional form analysis by searching for all the different tenses of a verb or both the singular and plural forms of a noun.

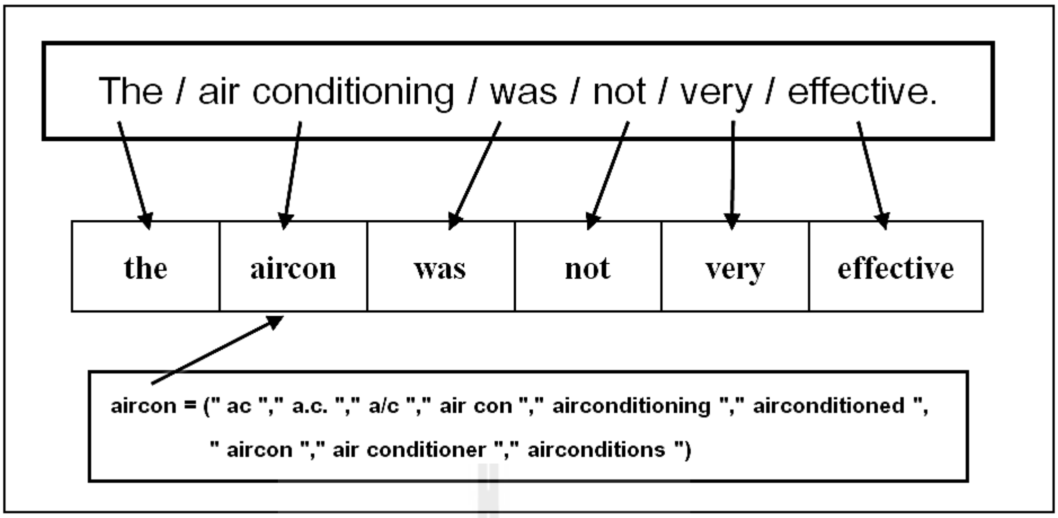


Figure 3.8 An example of the lexical analysis

– ***Syntactic Analysis***

The syntactic analysis performs relationship analysis between the words in a sentence (or part-of-speech) according to a context-free-grammar (CFG) parsing approach as previously described in the Chapter 2. Figure 3.9 shows an example of syntatic analysis of a sentence “the aircon was not very effective” using the CFG parsing approach. The language parsing method applies the tourism ontology to extract accommodation features and feature relationships, and the terminology to extract their feature modifiers as the following steps.

Step 1: All words of each criticism obtained from the lexical analysis are assigned a symbol in order to apply with the context-free-grammar. A noun are assigned as NOUN, verb to be as VB, verb as V, adjective as ADJ, as well as an adverb and nagation adverb as ADV, as illustrated in Figure 3.9.

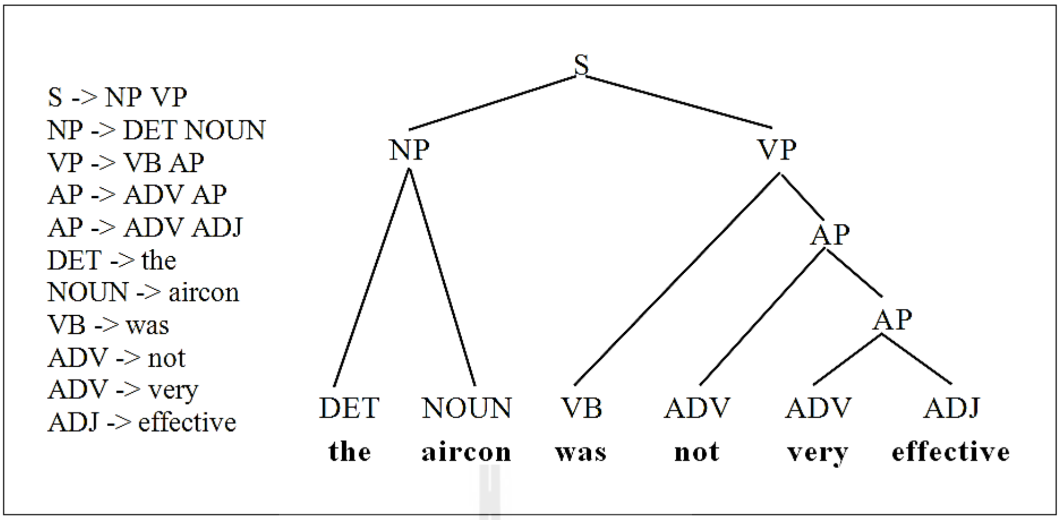


Figure 3.9 An example of the syntactic analysis

Step 2: All words assigned a symbol and all CFG rules defined as Figure 3.5 are passed through the CFG parsing approach. The approach analyzes accommodation feature, feature modifier, and feature relationship as follows:

- The accommodation feature is implied by a noun that matchs with the name of classes or properties in the tourism ontology. In the example sentence, the word "aircon" is extracted as an accommodation feature because it is a noun that matches with the property name in the tourism ontology.

- The feature relationship is identified by the relationships among classes and properties in the tourism ontology. In the example sentence, the feature "aircon" is extracted as a property of Room class, that is, the feature "aircon" is related with the feature "room".

- The accommodation feature modifier (adjective, special verb, special word, adverb, and negation adverb) is identified by matching with the words in the terminology. In the example sentence, the words “not”, “very”, and “effective” are extracted as a feature modifier of the accommodation feature “aircon”.

As a result, the relationship between accommodation features and its feature modifiers is used in the semantic analysis described later.

– ***Semantic Analysis***

The semantic analysis is a process for interpreting the meaning of reviews derived from the syntactic analysis. Its input is a parse tree of a criticism sentence with specified grammar (according to the context free grammar rules). The algorithm of the semantic analysis is as follows:

Step 1: Searching for an antecedent noun phrase in the parse tree according to the pronominal anaphora resolution adapted from Hobbs (1978)’s algorithm in cases of pronoun word “He”, “She”, “It”, and “They”. The intuition of this algorithm is to start with the target pronoun and walk up the parse tree to the root (*S*). For each noun phrase (*NP*) or sentence (*S*) node that it finds, it does a breadth-first left-to-right search of the node’s children to the left of the target. If no referent is found, the algorithm performs the same breadth-first search on preceding sentences (Jurafsky and Martin, 2000: 689)

Figure 3.10 illustrates an example of pronominal anaphora resolution. For the example, the pronoun word “it” refers to the noun phrase “aircon” according to the pronominal anaphora resolution adapted from Hobbs’s algorithm.

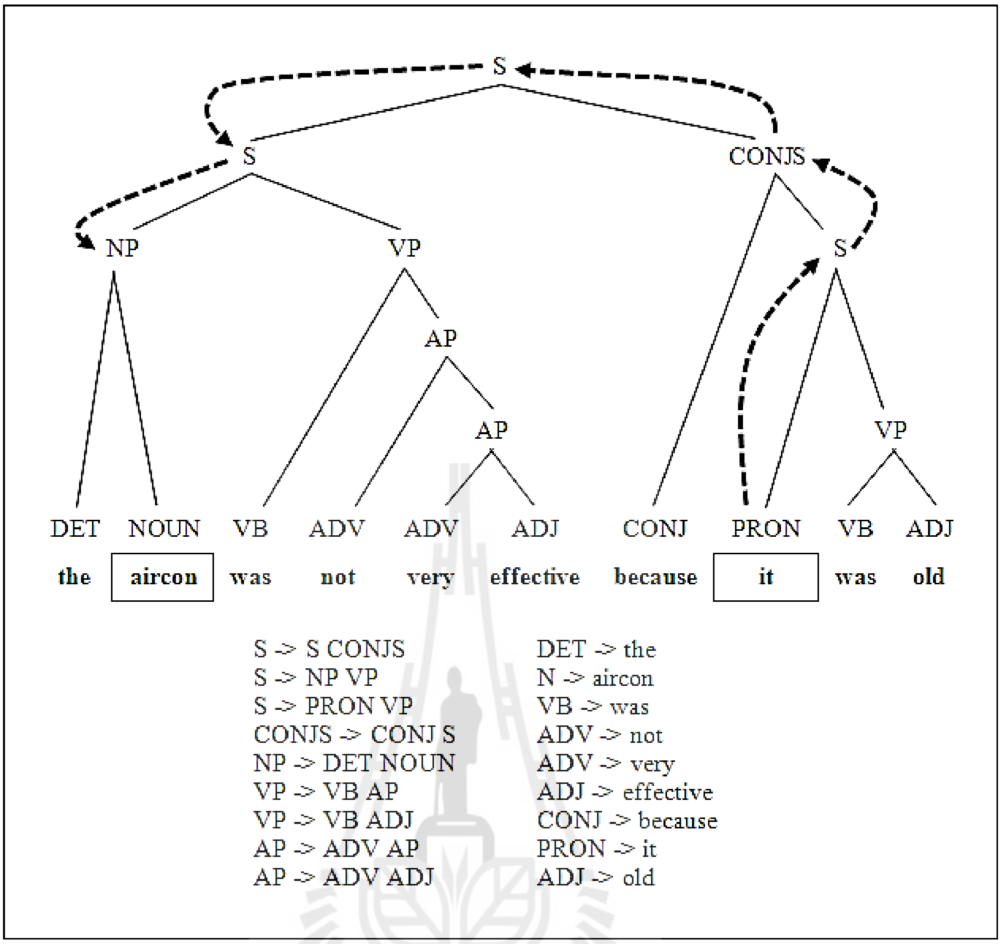


Figure 3.10 An example of the semantic analysis

Step 2: Searching for a pair of feature and feature modifier within the parse tree of the criticism sentence. The feature is implied by anoun that matchs with the name of classes or properties in the tourism ontology as previously discussed. The feature modifier consists of five types which are adjective, adverb, negation adverb, special verb, and special word. This step handle with the first three types of modifier only. First, an adjective will be identified as an adjective modifier of a feature, if it and the feature are in the same sentence with the nearest distance comparing with other nouns in the sentence, and there is no a preposition phrase node (*PP*) between them. Second, an adverb and a negation adverb will be identified as an adverb modifier of the feature, if it modifies an adjective that is the adjective modifier of the same feature.

In figure 3.11, the parse tree of the criticism sentence "the aircon was not very effective." was created and the word "aircon" was extracted as a feature by the syntactic analysis. Following the parse tree, the adjective "effective" will be identified as an adjective modifier of the feature "aircon", because this adjective and the feature are in the same sentence with the nearest distance and there is no a *PP* node between them. In addition, the adverb "very" is identified as an adverb modifier because the adverb modifies the adjective "effective". Finally, the adverb "not" is identified as an negation adverb modifier because this adverb modifies the adjective "effective" in opposite sense.

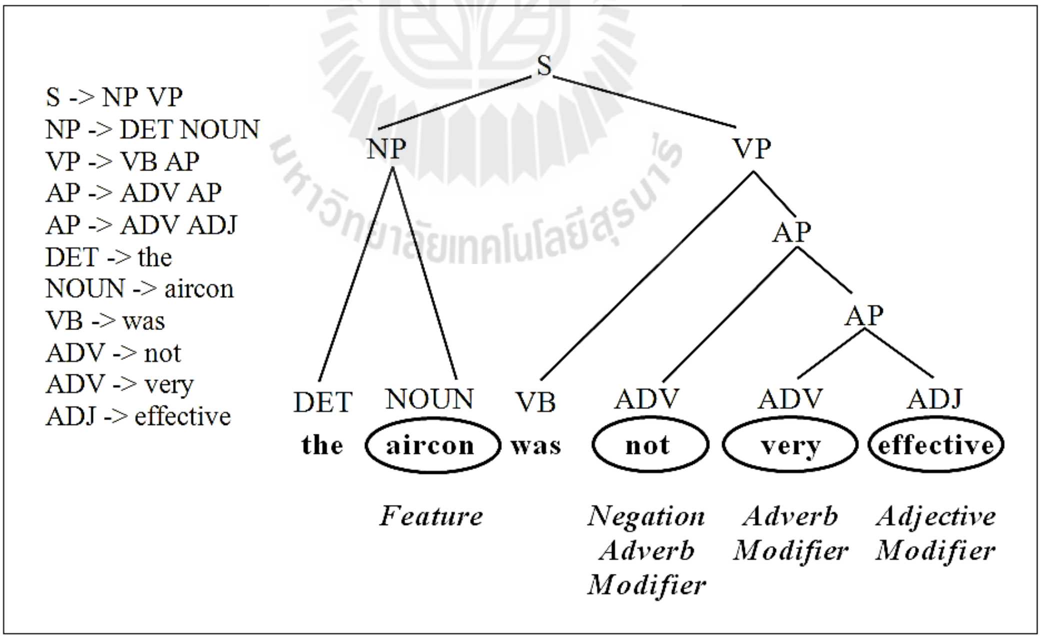


Figure 3.11 An example of searching for pairs of feature and adjective, adverb, and negation adverb modifier

Step 3: Searching for a special verb word in the criticism sentence. The special verb in this research means a verb word that is always criticized together with the feature word. This step will be performed by searching for a pair of feature and special verb within the parse tree of the criticism sentence as same as Step 2. The special verb word will be identified as a verb modifier of a feature, if it and the feature are in the same sentence with the nearest distance comparing with other nouns in the sentence, and there is no a preposition phrase node (*PP*) between them. All of the special verb words are presented in Appendix A.

In figure 3.12, the parse tree of the criticism sentence "the hotel has deteriorated since we last stayed there." was created and the word "hotel" was extracted as a feature by the syntactic analysis. Following the parse tree, the verb "deteriorated" will be identified as a verb modifier of the feature "hotel", because this verb and the feature are in the same sentence with the nearest distance and there is no a *PP* node between them.

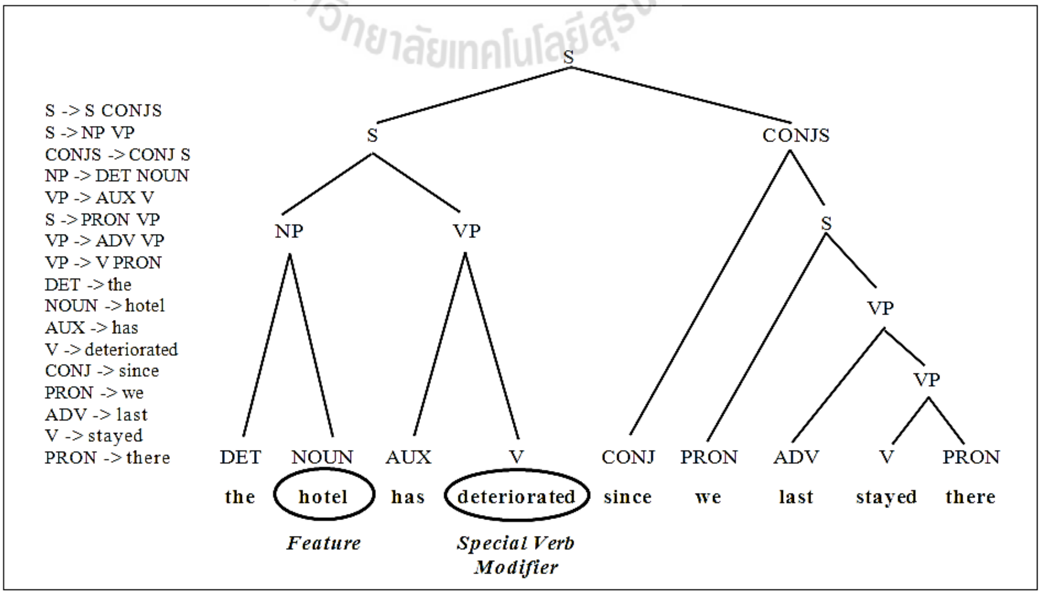


Figure 3.12 An example of searching for pairs of feature and special verb modifier

Step 4: Searching for a special word in a criticism sentence. The special word used in this research means a word that can be criticized an accommodation feature without the feature word in the criticism sentence, because it can identify the feature by itself. Hence, the special word is assigned a fixed satisfaction level including a feature to which it implies. For instance, the special word “expensive” which is assigned 2 points of satisfaction level and the “price” feature to which it implies. All of the special words are explained in Appendix A.

Step 5: Searching for vague domains of a feature implied in the special word. Note that some features implied in a special word can be criticized in various domains of accommodation property. For example, the feature “cleanness” implied in the special word “dirty” belongs to room, bathroom, and accommodation domains. Thus, a criticism about the special word “dirty” is vague about what domain of the “cleaness” feature is mentioned in the criticism.

This vague domain case is solved by finding surrounding words of feature which indicates the domain or accommodation properties that the feature belongs to. If no surrounding words are found, the vague feature is proposed that it is criticized in aspects of overall accommodation. This research assigns four accommodation features as the vague domain case, e.g. cleanness, curtain, decoration (which may be criticized in aspects of accommodation, bathroom, or room), and carpet (which may be criticized in aspects of accommodation or room).

Step 6: Searching for a real criticism of vague features in case of a special word that is matched with one feature in a criticism sentence, e.g. “The room is worth.”. There are two features extracted in the sentence, i.e. the feature “room” found in the sentence and the feature “value” implied in the special word “worth” as shown in Figure 3.13. The problem is what feature should be criticized in this sentence.

There are two cases of identifying an appropriate feature of a criticism as follows:

First, if the feature found in the sentence is not a domain of the feature implied in the special word, the feature found in the sentence is identified by the special word because it is directly criticized by a reviewer. Therefore, the feature found in the sentence is identified as a real criticism feature and the special word is considered as its feature modifier.

For instance of the criticism sentence “The room is worth.”. As illustrated in Figure 3.13, the feature “room” found in the criticism sentence is not the domain of the feature “value” implied in the special word “worth” as presented in the tourism ontology. The feature “value” is actually in the “accommodation” domain. Therefore, the feature “room” found in the sentence is the real critism feature explained by the special word “worth”.

Second, if the feature found in the sentence is the domain of the feature implied in the special word, the feature implied in the special word is a real criticism feature because the feature found in the sentence is less specific than the feature implied in the special word as illustrated in the tourism ontology.

For example, suppose that the criticism sentence is “The room is clean.”. There are two extracted features that are the feature “room” found in the criticism sentence and the feature “cleanness” implied in the special word “clean”. Unfortunately, the feature “cleanness” has vague domain. Thus, in the first step, a vague domain problem is solved by searching for an appropriate domain of the feature "cleanness". This example assumes that the appropriate domain is "room". In the second step, the vague feature problem is handled. The example indicates that the feature "room" found in the criticism sentence is the domain of the feature "cleanness". Therefore, the feature "cleanness" is the real criticism feature because the feature found in the sentence is less specific than the feature "cleanness".

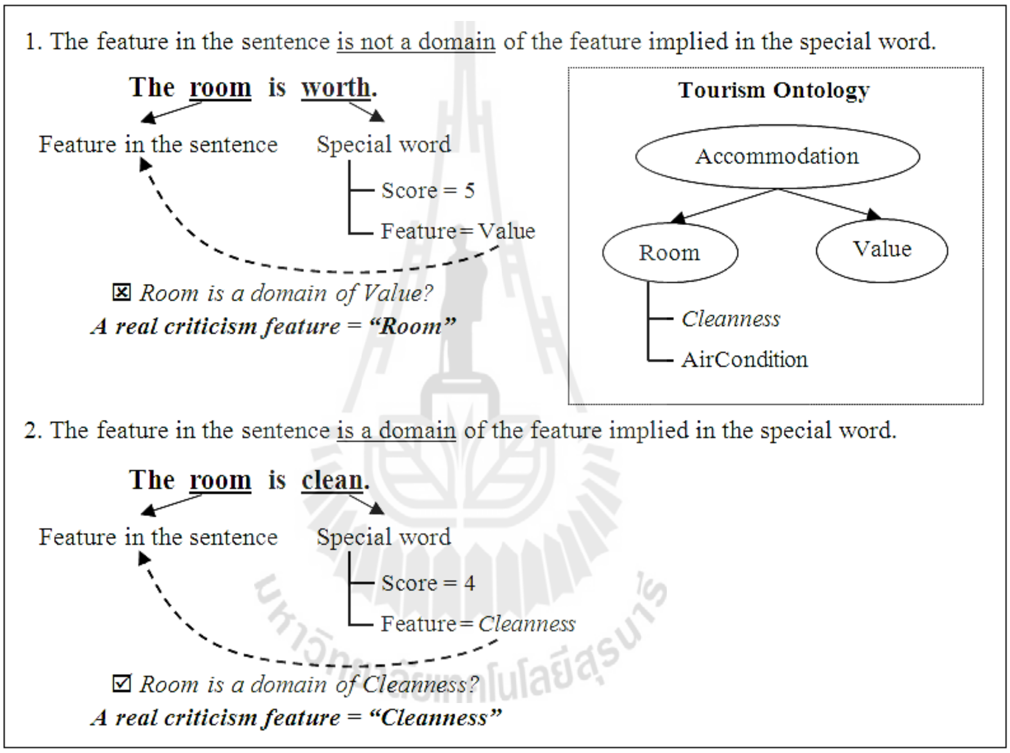


Figure 3.13 An example of searching for a real criticism of vague features in case of a special word that is matched with one feature in a criticism sentence

Step 7: Considering other commentary words or phrases, except feature modifiers. This step will be performed, if a feature is not matched with any feature modifiers and a commentary word or phrase such as "Need" and "In need of" are found after a feature word. As a result, the satisfaction level of a criticism will be decided as 2 points because the feature is regarded as a negative criticism.

For instance, suppose that the criticism sentence is "Facilities need upgrade.". The word "Facilities" is extracted as a feature but there is no feature modifiers matched with this feature. However, the commentary word "need" is related to the feature "Facilities". Thus, the satisfaction level of a criticism of this feature will be decided as 2 points.

Step 8: Considering other commentary words or phrases, except feature modifiers. This step will be performed, if a feature is not matched with any feature modifiers and a commentary word or phrase such as "Have", "Has", "There is", "There are", "There was", "There were", "With", and "No" are found and it precedes a feature word. As a result, the satisfaction level of criticism will be decided as 2 or 4 points in case of negative and positive criticism, respectively.

For instance, suppose that the criticism sentence is "There is a microwave oven.". The phrase "microwave oven" can be extracted as a feature but there is no feature modifiers matched with the feature. However, the commentary phrase "There is" is related to the feature "microwave oven" and the phrase precedes the feature. Thus, the satisfaction level of a criticism of this feature will be decided as 4 points.

After all the mentioned steps, the results of the knowledge inference engine, e.g. features with their relationships and feature modifiers, commentary words and special words, etc. are extracted and stored in the knowledge base in order to apply to accommodation rating in the next process.

- **Accommodation Rating**

This process performs the computation of tourists' satisfaction level on each accommodation feature and overall accommodation rating. The purpose of this process is to find the best computation method for an accommodation rating that is closest to the rating given by tourists (or human) as much as possible. The computation method is divided in two steps: feature rating and hierarchical feature rating as described below.

– ***Feature Rating***

The features are indicated by the nouns or significant keywords appeared in each sentence. The feature rating is a calculation of a tourists' satisfaction level on each extracted feature. Normally, the feature rating will be assigned by the score of adjectives which are stored in terminology. For instance, a sentence "the room is clean" has the feature "room", the adjective "clean", and "rating=4" as the score of adjective. However, there are many cases of the criticism sentence expressed in the form of Adverb Adjective Combination (AAC) (Benamara, Cesarano, Picariello, Subrahmanian, and Reforgiato, 2006: 1-7) such as the AAC "very good" consists of the adverb "very" and the adjective "good". Therefore the feature rating must derive from the aggregating score of both adverb and adjective.

In order to provide the aggregating score, this research proposes an AAC scoring method modified from Benamara et al. (2006). The AAC scoring method is based on measuring a relation score (r) between adjectives and adverbs in their semantic association in term of tourism subject. The AAC has three possible forms as follows: a unary, binary, and negation form.

a) Unary AAC Scoring

The unary AAC has a form: *<adverb><adjective>*.

For example: "*really nice*", "*extremely uncomfortable*", etc.

Suppose that *sc*(*adj.*), *sc*(*adv.*), and *sc*(*AAC*) are the scores of adjective, adverb, and the adverb-adjective combination, respectively. The variable *r* is the relation score. The unary *AAC* scoring algorithm works as follows:

- If $sc(adj.) \geq 3$, then $sc(AAC) = sc(adj.) + (r \times sc(adv.))$.
- If $sc(adj.) < 3$, then $sc(AAC) = sc(adj.) - (r \times sc(adv.))$.

Example 1: suppose the sentence is "*The reception staff is very friendly.*" Let $r = 0.9$, $sc(friendly) = 4$, and $sc(very) = \pm 1$. In this case, a feature is "*reception staff*", an adjective is "friendly", and an adverb is "very". The feature score would be calculated from the *AAC* "*very friendly*" as follows:

$$sc(very\ friendly) = sc(friendly) + (r \times sc(very))$$
$$= 4 + (0.9 \times 1) = 4.9.$$

b) Binary AAC Scoring

The binary AAC has a form: *<adverb*$_2$*><adverb*$_1$*><adjective>*.

For example: "*really so small*", "*very very clean*", etc.

There are two steps for calculating the feature score in this form. The first step is computing the score of AAC_1 as $sc(adv._1, adj.)$ according to the step of unary *AAC* scoring. The second step is setting this AAC_1 score as an input of computing the AAC_2 score. As a result, the score of AAC_2 is returned as the feature score as follows:

- If $sc(adj.) \geq 3$, then

$$sc(AAC_1) = sc(adj.) + (r_{adv.1} \times sc(adv._1))$$

$$sc(AAC_2) = sc(AAC_1) + (r_{adv.2} \times sc(adv._2))$$

- If $sc(adj.) < 3$, then

$$sc(AAC_1) = sc(adj.) - (r_{adv.1} \times sc(adv._1))$$

$$sc(AAC_2) = sc(AAC_1) - (r_{adv.2} \times sc(adv._2))$$

Example 2: Suppose that the sentence is "*The reception staff are really very friendly.*" and use the scores as shown in *Example 1* including $sc(really) = \pm 1$. In this case, a feature is "*reception staff*", an adjective is "friendly", the 1st adverb ($adv._1$) is "very", and the 2nd adverb ($adv._2$) is "really". The feature score would be calculated from the *AAC* "*really very friendly*". The binary AAC scoring algorithm would take the AAC_1 "*very friendly*" into account and assign it the score as:

$$sc(very\ friendly) = sc(friendly) + (r \times sc(very))$$
$$= 4 + (0.9 \times 1) = 4.9$$

Next, the AAC_1 would be used as an input of AAC_2 "*really very friendly*" scoring as follows:

$$sc(really\ very\ friendly) = sc(very\ friendly) + (r \times sc(really))$$
$$= 4.9 + (0.9 \times 1) = 5.8$$

Notice that, the final score of binary *AAC* exceeds 5-rating scale, therefore this score is recomputed and limited a range to 1-5 points.

c) Negation AAC Scoring

The negation AAC is the AAC including a negation adverb (*nadv.*) which has a form: <*negation adverb*><*adverb*><*adjective*>.

For example: “*not very helpful*”, “*never so clean*”, etc.

In order to calculate the score of a negation AAC, there are two steps like the binary AAC scoring. The first step is computing the score of AAC_1 as *sc*(*nadv*., *adj*.) by adjusting the score of adjective with the score of negation adverb. The second step is setting this AAC_1 score as an input of AAC_2. The AAC_2 score is computed by *sc*(*adv*., AAC_1) scoring. As a result, the score of AAC_2 is returned as the feature score as follows:

- If $sc(adj.) \geq 3$, then $sc(AAC_1) = sc(adj.) - sc(nadv.)$.
    - If $sc(AAC_1) \geq 3$, then

      $sc(AAC_2) = sc(AAC_1) + (r \times sc(adv.))$.
    - If $sc(AAC_1) < 3$, then

      $sc(AAC_2) = sc(AAC_1) - (r \times sc(adv.))$.
- If $sc(adj.) < 3$, then $sc(AAC_1) = sc(adj.) + sc(nadv.)$.
    - If $sc(AAC_1) \geq 3$, then

      $sc(AAC_2) = sc(AAC_1) + (r \times sc(adv.))$.
    - If $sc(AAC_1) < 3$, then

      $sc(AAC_2) = sc(AAC_1) - (r \times sc(adv.))$.

Example 3: Suppose that the sentence is “*The reception staff are not very friendly.*” and uses the scores as shown in *Example 1* including *sc*(*not*) = ±2. The feature score would be calculated from the *AAC* “*not very friendly*”. In this case, the negation *AAC* scoring algorithm would compute the feature rating from the AAC_1 “*not friendly*” by calculating as *sc*(*nadv*., *adj*.) as:

$$sc(not, friendly) = sc(friendly) - sc(not)$$

$= 4 - 2 = 2.$

Next, the AAC_1 would be used as an input of AAC_2 "*not very friendly*" scoring. The AAC_2 is $sc(adv., AAC_1)$ scoring as follows:

$$sc(very, not\ friendly) = sc(not\ friendly) - (r \times sc(very)).$$

$$= 2 - (0.9 \times 1) = 1.1$$

The three AAC forms as discussed above are applied with accommodation reviews in case of both simple and complex sentence. An example of room rating from the complex sentence "*The room is dirty but the air conditioning was almost very effective*", is shown in Figure 3.14. Each simple sentence is parsed and then a feature and feature score are extracted, i.e. the *RoomProperties* feature with rating score = 2 is extracted from the first simple sentence "*the room is dirty*" and the *AirCondition* feature with rating score = 3.9 is extracted from the second simple sentence "*the air conditioning was almost very effective*".

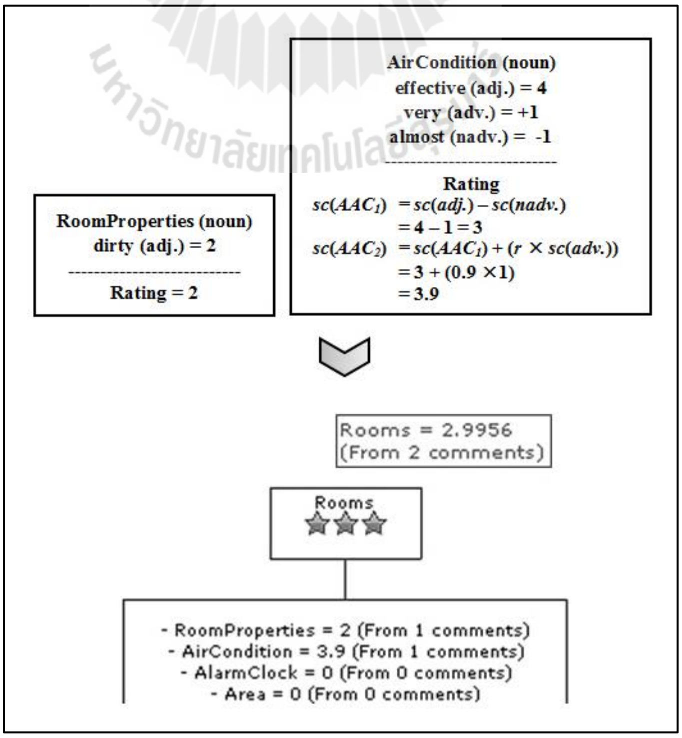


Figure 3.14 An example of feature rating

– ***Hierarchical Feature Rating***

In order to calculate the hierarchical feature rating or overall score of tourists' satisfaction, the scoring features (from the feature rating steps) will be used as an input data in bottom-up hierarchy of accommodation information, i.e. each feature score of the higher layer is calculated from the feature scores of the lower layer. For instance, the Room rating is calculated from the Air condition and Bed scores. Then, the Hotel rating is calculated from the Room and Location scores, as shown in Figure 3.15.

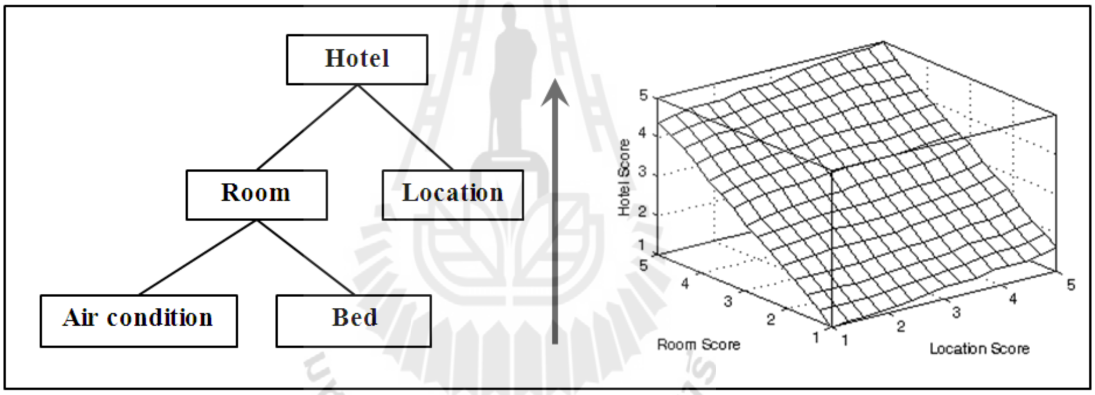


Figure 3.15 An example of hierarchical feature rating

There are several alternative methods for calculating the hierarchical feature rating. In this research, the hierarchical rating methods are divided into two groups which are non fuzzy-based method and fuzzy-based method.

a) Non fuzzy-Based Method

The non fuzzy-based method computes the overall score of features in the higher layer using an average of associated feature scores in the lower layer as Equation 3.2.

$$Overall\ Score = \frac{\sum_{i=1}^{n} Score_i}{n} \tag{3.2}$$

Where the *Overall Score* means the overall rating or hierarchical feature rating, $Score_i$ means the rating score of each feature i, and n means the number of all extracted features.

In Figure 3.16, an example of the hierarchical feature rating using the average scoring method is illustrated.

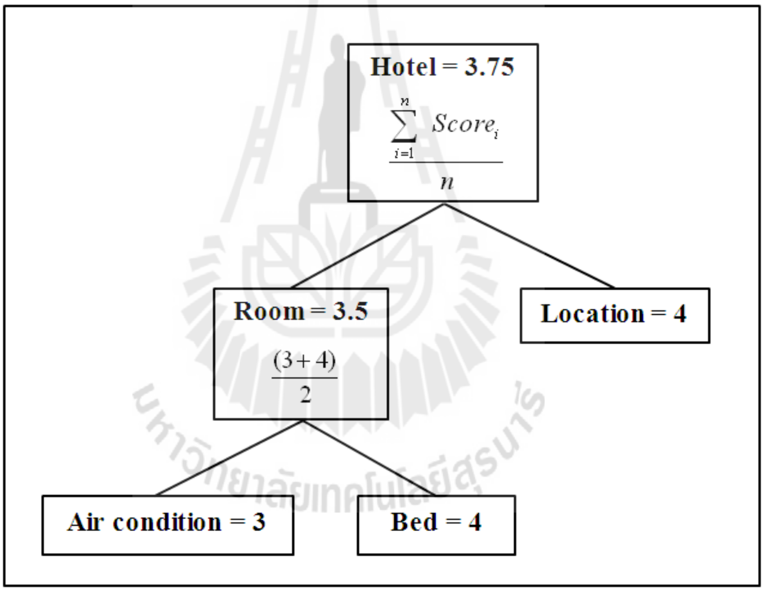


Figure 3.16 The hierarchical feature rating using the average scoring method

b) Fuzzy-Based Method

The fuzzy-based method computes the overall score of each feature by applying a fuzzy inference technique. The fuzzy inference technique is composed of four processes:

Process 1: Fuzzification

This process aims to transform feature scores into fuzzy values. It starts by assigning input and output linguistic variables. The inputs are 95 scoring features according to the properties of ontology. The outputs are the overall scores in the higher layer of each feature. Then, this process defines linguistic terms of the feature score and overall score in 1 to 5 degrees. In this research, a triangular membership functions are used for determining the degree of membership (DOM) which are defined by three parameters {a, b, c}. Where *b* is a modal value, *a* and *c* denote the lower and upper bounds, respectively, for non-zero values of the triangular(x). Equation 3.3 shows a triangular function (Nguyen and Walker, 2000: 56):

$$Triangular(x: a, b, c) = \begin{Bmatrix} 0 & x < a \\ (x-a)/(b-a) & a \leq x \leq b \\ (c-x)/(c-b) & b \leq x \leq c \\ 0 & x > c \end{Bmatrix} \quad (3.3)$$

The triangular membership functions used in the designed module describe the membership of the feature scores in the lower layer with 5 fuzzy sets (called terrible, poor, average, good, and excellent) having different parameters for each one. For example, defining the membership of the fuzzy set “poor” by setting parameters to a=1, b=2, and c=3, as shown in Figure 3.17.

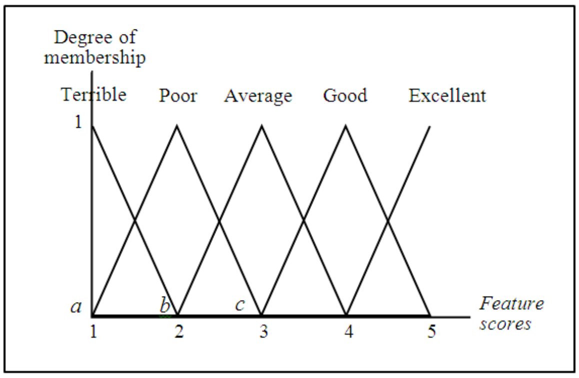


Figure 3.17 A triangular membership function

Process 2: Fuzzy Rule Base

In order to calculate the overall scores using the fuzzy inference, the fuzzy inference rules with weight between 0 and 1 for each rule are needed. A fuzzy rule base contains a number of fuzzy rules, such as "*If Location=Good, Then Hotel=Good (Weight = 0.5)*", where *Location* and *Hotel* are a linguistic variable, *Good* is a linguistic value or label that are characterized by the membership function, and *Value* (*0.5*) is this a rule weight. The rule weight tuning process used the 200 reviews from TripAdvisor.com as a dataset, covering all 5 satisfaction levels (40 reviews for each level). There are two weighting methods presented in this research as follows.

- **Equal Weighting**

This weighting method assigns the rule weight of all features to 1. It implies that all features affect the hierarchical feature rating with the same impact.

- **Probability Weighting**

For this method, the weight of each rule has been assigned based on the concept of "*How many feature rating affects the overall accommodation rating?*". For instance, if a review criticizes that an accommodation is good and an accommodation feature in the review is also good, then the feature rating affects the hierarchical feature rating. Thus, more weight implies that the feature rating has more contribution to the hierarchical feature rating. The steps of weighting each fuzzy rule or each feature are as follows.

Step 1: Compare the feature scores of each review with an adjacency range of each satisfaction level in Table 3.1 (derived from the triangular membership functions). If the feature score of any reviews is in the adjacency range of its level, an adjacency value is assigned as 1. On the other hand, if the feature rating of any reviews is not in the adjacency range of its level, the adjacency value is assigned as 0. Then, all adjacency values of every review are combined as a combined adjacency value (*V*).

Table 3.1 The criteria for proximity considering of the feature rating with the satisfaction levels

Satisfaction Levels	**Adjacency Range**
Level 1 (Terrible) ★	1.00-1.99
Level 2 (Poor) ★★	1.01-2.99
Level 3 (Average) ★★★	2.01-3.99
Level 4 (Good) ★★★★	3.01-4.99
Level 5 (Excellent) ★★★★★	4.01-5.00

Step 2: Count the number of reviews (*L*) which the designated feature is criticized at least once. An example of the reception feature, if there are 2, 4, 9, 4 reviews of Level 1, 3, 4, 5 criticizing this feature respectively while there have no reviews of Level 2, the number of reviews criticizing this feature is 19 reviews, derived from 2+4+9+4.

Step 3: Calculate a percentage of the adjacency ratio (*P*) by dividing the combined adjacency value (*V*) derived from step 1 with the number of reviews (*L*) derived from step 2, and then multiply by 100 as shown in Equation 3.4.

$$P = \frac{V}{L} \times 100\% \tag{3.4}$$

This percentage of adjacency ratio corresponds to the concept of "*How many feature rating affects the overall accommodation rating?*".

Step 4: Transform the percentage of the adjacency ratio into a 0-1 weight value by comparing the calculated adjacency ratio derived from step 3 with the adjacency ratio in Table 3.2 to obtain an actual weight of fuzzy rule. This research assigns the weight values as 1, 0.5, 0.1, and 0 to imply that criticizing the designated feature has high, moderate, little, and no effects on the overall or hierarchical feature rating, respectively.

Step 5: Repeat step 1 to 4 with all features.

Table 3.2 The criteria for transforming the percentage of adjacency ratio into an actual weight of fuzzy if-then rule

Percentage of Adjacency Ratio	**Actual Weight of Fuzzy Rule**
Greater than or equal to 80%	1
Between 50% and 79%	0.5
Between 1% and 49%	0.1
Equal to 0%	0

Suppose that there are 19 reviews criticizing the reception feature found in every satisfaction level: Level 1 has 2 reviews, Level 2 has no reviews, Level 3 has 4 reviews, Level 4 has 9 reviews, and Level 5 has 4 reviews, as depicted in Table 3.3. The steps of weighting the fuzzy rules are as follows:

Step 1: Compare the feature scores (or reception rating in this case) of each review in Table 3.3 with the adjacency range of each satisfaction level in Table 3.1. The results indicate that there are 2, 6, 3 reviews in satisfaction level 1, 4, and 5 are matched respectively while there have no reviews matching in satisfaction level 3. Thus, the combined adjacency value (*V*) is 11 reviews, derived from 2+6+3.

Table 3.3 An example of reception rating from reviews dataset

Satisfaction Levels	Reception Rating
Level 1 ★	(Found 2 reviews) Numbers of reviews matched with adjacency range = 2
Level 2 ★★	(Have no criticisms)
Level 3 ★★★	(Found 4 reviews) Numbers of reviews matched with adjacency range = 0
Level 4 ★★★★	(Found 9 reviews) Numbers of reviews matched with adjacency range = 6
Level 5 ★★★★★	(Found 4 reviews) Numbers of reviews matched with adjacency range = 3

Step 2: Count the number of reviews which the reception feature is criticized at least once (*L*). The result is 19.

Step 3: The percentage of adjacency ratio is 57.89% which is calculated from (11/19)*100, where the numerator 11 is the combined adjacency value (from step 1) and the denominator 19 is the number of reviews criticizing the reception feature.

Step 4: The actual weight of the fuzzy reception rule is 0.5 derived by comparing with the adjacency ratios in Table 3.2. This weight implies that criticizing the reception feature has a moderate effect on the hierarchical feature rating.

This research defines 70 rules derived from the classes and properties of ontology (or features of accommodation). These rules are divided into 4 groups: 20 bathroom rules, 20 room rules, 15 service rules, and 15 overall of accommodation rules. An example of the 6 fuzzy rules used in this research is shown in Table 3.4. One rule is composed of one output and one weight. Features that have the same output and weight are combined into one rule with the weight.

Table 3.4 An example of the fuzzy inference rules including weights for overall rating calculation

Rules	**If**	**Then**	**Weight**
1	(AlarmClock = Good) OR (InternetAccess = Good) OR (Microwave = Good) OR **(RoomCarpet = Good)** OR (TeaCoffeeEquipment = Good)	Room = Good	1
2	**(AirCondition = Poor**) OR (Hotwater = Poor) OR (RoomCleanness = Poor) OR (RoomDecoration = Poor) OR (OverallRoom = Poor)	Room = Poor	0.5
3	**(Location = Excellent**) OR (AccommodationCleanness = Excellent) OR (AccommodationDecoration = Excellent) OR (Dinner = Excellent) OR (Lobby = Excellent) OR (OffersChildCare = Excellent) OR (Restaurant = Excellent)	Accommodation = Excellent	0.1

Table 3.4 An example of the fuzzy inference rules including weights for overall rating calculation (Continued)

Rules	If	Then	Weight
4	(Bathroom = Poor) OR (Breakfast = Good) OR (Café = Good) OR (Elevator = Good) OR (FitnessRoom = Good) OR (Food = Good) OR (Furniture = Good) OR (OverallAccommodation = Good) OR (Parking = Good) OR (Pool = Good) OR **(Room = Good**) OR (Security = Good) OR (Service = Good) OR (Value = Good) OR (WellnessFacilities = Good)	Accommodation = Good	0.5
5	(Bathroom = Poor) OR (Breakfast = Average) OR (Café = Average) OR (Elevator = Average) OR (FitnessRoom = Average) OR (Food = Average) OR (Furniture = Average) OR (OverallAccommodation = Average) OR (Parking = Average) OR (Pool = Average) OR (**Room = Average**) OR (Security = Average) OR (Service = Average) OR (Value = Average) OR (WellnessFacilities = Average)	Accommodation = Average	0.5

Table 3.4 An example of the fuzzy inference rules including weights for overall rating calculation (Continued)

Rules	If	Then	Weight
6	(Bathroom = Poor) OR (Breakfast = Poor) OR (Café = Poor) OR (Elevator = Poor) OR (FitnessRoom = Poor) OR (Food = Poor) OR (Furniture = Poor) OR (OverallAccommodation = Good) OR (Parking = Poor) OR (Pool = Poor) OR (Room = Poor) OR (Security = Poor) OR **(Service = Poor**) OR (Value = Poor) OR (WellnessFacilities = Poor)	Accommodation = Poor	0.5

Process 3: Inference Engine

The inference engine is a process of fuzzy reasoning upon fuzzy rules. In other words, the overall rating is hierarchically calculated by the feature rating. However, the adjustment of the overall rating depends on the weight of associated rule. For example, the room carpet is reviewed as good (score = 4) and the air condition is reviewed as poor (score = 2). In inference engine, the 4 points of *RoomCarpet* and 2 points of *AirCondition* will be used as input data for comparing them with the membership functions to obtain the membership values according to the rule of *RoomCarpet* and *AirCondition* in Table 3.4 which have 1 and 0.5 of rule weight respectively as depicted in Figure 3.18. The room rating showed that the membership value of *RoomCarpet* is the maximum value because the weight of

RoomCarpet rule is assigned as 1 while the membership value of *AirCondition* is a half value as it should be because the rule weight is assigned as 0.5.

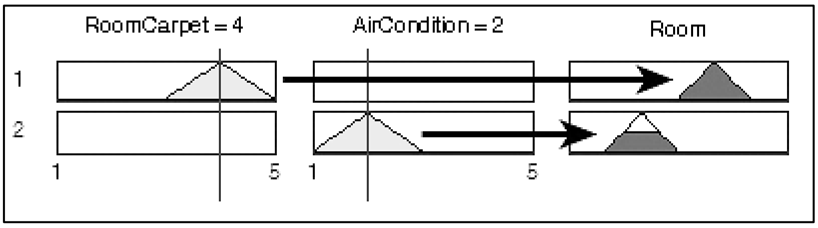


Figure 3.18 An example of the output inference, where the weight of rules are 1 and 0.5

Process 4: Defuzzification

After the previous step, the membership values of the feature rating will be combined for generating the qualified consequents depending on the weight of each rule and aggregating the qualified consequents to produce the room rating (as crisp output) using the Centroid of Area method (COA) (Jang, Sun, and Mizutani, 1997: 75), as illustrated in Figure 3.19.

In conclusion, the 4 points of *RoomCarpet* and 2 points of *AirCondition* are the feature rating in the lower layer used as inputs for calculating the *Room* rating or overall rating in the higher layer. As a result, 3.14 points of the *Room* rating is aggregated from the rule of *RoomCarpet* and *AirCondition* because the *RoomCarpet* and *AirCondition* features are used as inputs for inferring the output of the *Room* rating.

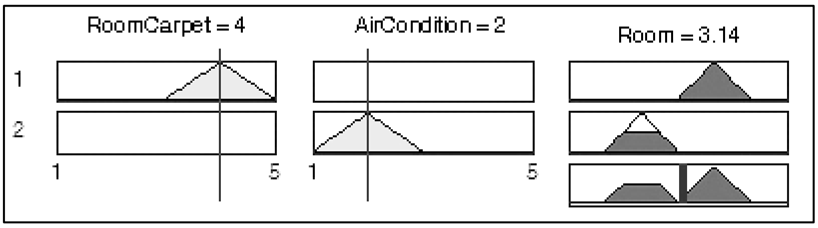


Figure 3.19 An example of the overall rating on the room feature, where the room carpet rating is 4 points and the air condition rating is 2 points

Furthermore, the overall accommodation rating is hierarchically calculated by the feature scores of the next layer, as shown in Figure 3.21, such as the 3.14 points of *Room*, 2 points of *Service*, and 5 points of *Location*. Consequently, 2.86 points of the *Accommodation* rating is aggregated from the rules in Table 3.4 which consecutively infer the *Accommodation* rating from *Room, Service*, and *Location* features as illustrated in Figure 3.20. Notice that the rating of *Location* has a little effect on the overall hotel rating because the weight of *Location* rule is assigned as 0.1 while the other is assigned as 0.5.

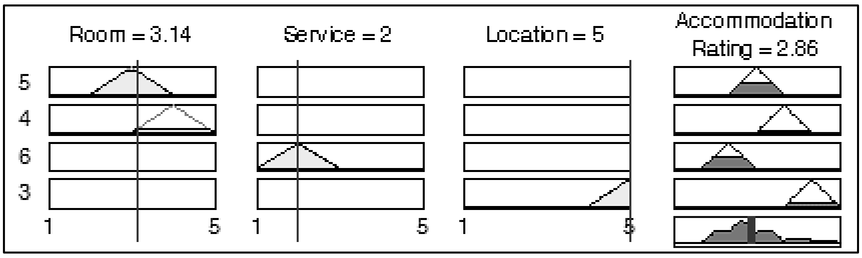


Figure 3.20 An example of the hierarchical feature rating calculation on *Accommodation*, where the room rating is 3.14 points, the service rating is 2 points, and the location rating is 5 points

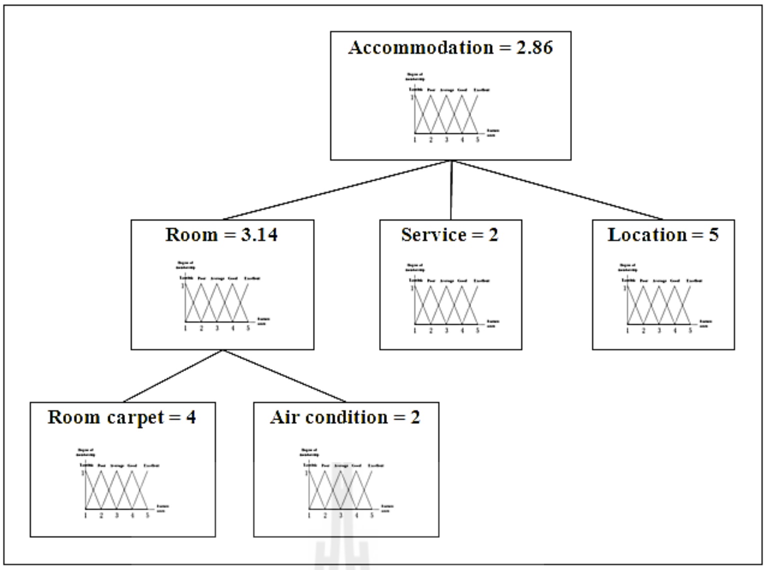


Figure 3.21 The fuzzy-based method

3) Knowledge Base

The knowledge base is an essential component of the designed module. It includes the tourism ontology, dictionary, terminology, and context free grammar rules which are used for a syntactic and semantic analysis in the feature extraction process. Moreover, the scores of words stored in terminology are also applied for a feature rating and hierarchical feature rating in the accommodation rating process. Finally, the extracted accommodation features and their satisfaction levels are gathered to the tourism ontology immediately.

4) Knowledge Explanation Engine

After all the mentioned steps, all extracted knowledge is stored in the knowledge base, and then retrieved by tourists via the user interface. The knowledge explanation engine aims to describe the knowledge as "information combined with experience, context, interpretation, and reflection", and "valuable information in action" (Davenport and Prusak, 1998). The knowledge explanation engine performs clear and easy understanding reviews with star rating and visualizing them into a tree structure according to the classes and properties of the tourism ontology. Thus, the knowledge explanation engine is discussed in two matters: the rating representation and the ontology implementation. The first matter describes transforming the tourists' satisfaction level into star rating, and the last matter describes the concept of applying to the tourism ontology which stores the knowledge of the designed review analysis model.

- **Rating Representation**

The tree structure is used for visualizing information retrieved from the knowledge base according to the search keywords. Figure 3.22 illustrated the knowledge representation of accommodation categories, which details are derived from the research study in tourist's satisfaction (Hsu, Tsai, and Wu, 2009: 295; Cracolici and Nijkamp, 2008: 342; Chi and Qu, 2008: 629-631; Bosque and Martin, 2008: 562) such as location, service, and room. In the tree structure, the overall rating is presented as stars. The number of stars is decided by following criteria:

– If the decimals part of the overall rating is between 0.00 and 0.24, the number of stars is displayed according to the whole number.

– If the decimals part of the overall rating is between 0.25 and 0.74, the number of stars is displayed according to the whole number and a half of star is added.

– If the decimals part of the overall rating is between 0.75 and 0.99, the number of stars is displayed according to the whole number and a full of star is added.

For example, in case the hierarchical feature rating is reviewed as 4 points, the star displaying is decided by criteria as shown in Table 3.5.

Table 3.5 The criteria for displaying the number of stars instead of 4 points of the overall rating

Decimals Part of Overall Rating Range	The Number of Stars
4.00-4.24	
4.25-4.74	
4.75-4.99	

Figure 3.22 illustrated an example of an accommodation that is reviewed as 3-star hotel (score = 2.86) by 255 comments got from various tourist aspects including: 1) Location, 5 points out of 100 comments, 2) Service, 2 points out of 100 comments, and 3) Rooms 3.14 points out of 55 comments. Thus, tourists can use this tree structure by exploring the accommodation in details to select an accommodation that conforms to their preferences.

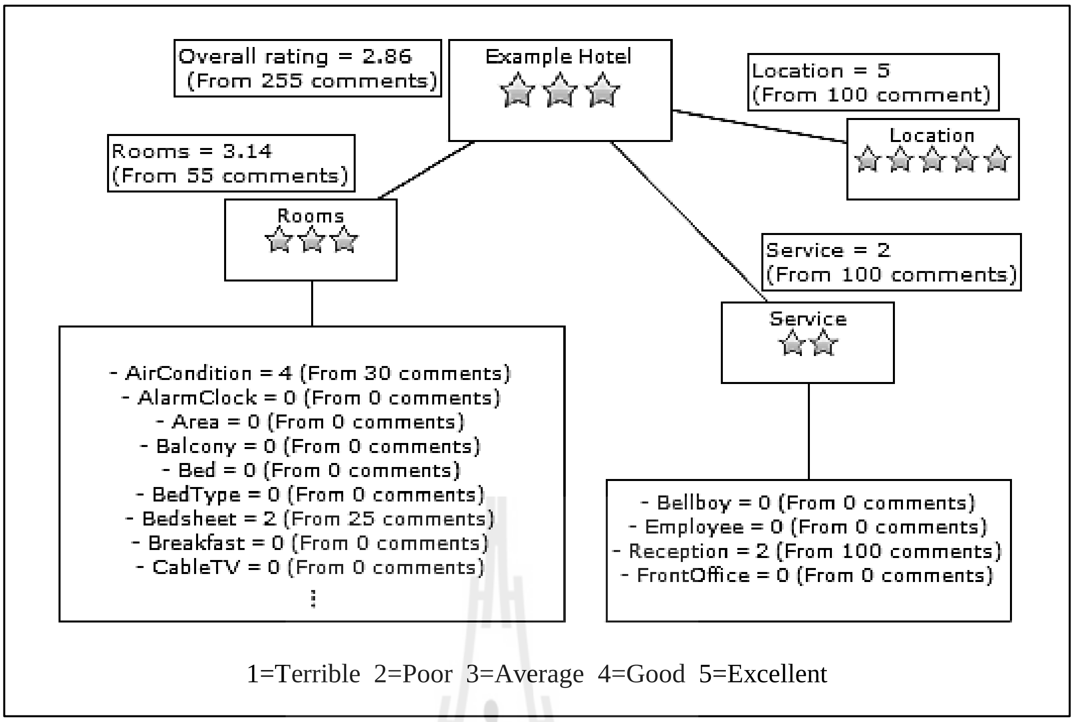


Figure 3.22 An example of the knowledge representation

- **Ontology Implementation**

This research modifies the tourism ontology by improving subclasses and properties of Accommodation class and adding new classes, such as Rating class as illustrated in Figure 3.23. The purpose of modifying the tourism ontology is to support the knowledge representation in the hierarchy as previously discussed. In addition, the other tourism websites can use or join the designed tourism ontology together which can assist tourists to search for information with various queries, e.g. a tourist can retrieve accommodation information by determining the activity, coordinate, rating, and so on.

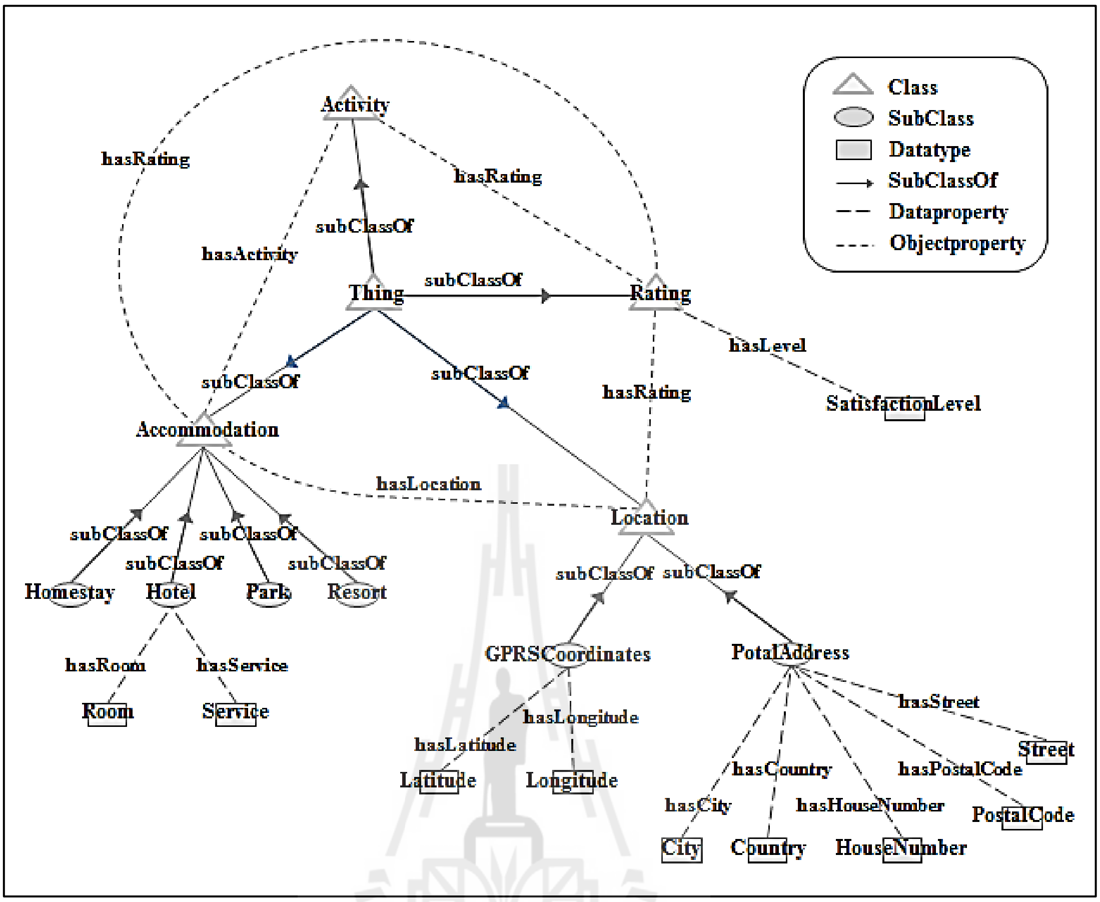


Figure 3.23 A partial structure of modified tourism ontology

In conclusion, the analysis module for online travel accommodation reviews performs three major tasks as follows.

1. It can extract accommodation features and their relationships.
2. It can calculate the tourists' satisfaction level on each extracted feature and on the entire review.
3. It can explain more clear and easy understanding reviews with 5-rating scale and visualize the accommodation feature relationships in hierarchy.

3.1.4 Review Analysis Module Evaluation

The review analysis module is evaluated by testing the accuracy of the feature extraction and the accommodation rating process. The Accuracy, Precision, and Recall measure (Miao et al., 2009: 9172) are used to evaluate the performance of the feature extraction process while the accommodation rating process is evaluated in the perspective of the difference of the tourist-defined rating and module-computed rating using the Mean Absolute Error (*MAE*) and Root Mean Squared Error (*RMSE*) measures. Moreover, this research presents the experimental results for finding the best relation score between adverbs and adjectives (*r*) that gives the overall rating closest to the tourists' subject. It is important to note that the test data of accuracy of both processes is separated from the data for module development. The details of module evaluation are thoroughly described at the data analysis section.

3.2 Research Instruments

The research instruments or development tools that are used for developing and evaluating the accuracy of review analysis module are described below.

1) Hardware: a laptop is used for developing an analysis module for online travel accommodation reviews, which have specifications as follows.

- Processor: Intel Core Duo T2050 1.60GHz
- Main memory: DDR2 2.5 GB DRAM
- Hard drive: 80 GB
- Internet connection: Integrated 802.11a/b/g
- Accessories: optical drive, keyboard, mouse, printer, etc.

2) Software: the operating system and various applications are used for developing the user interface of an online review analysis module which has the capability on developing web application and connecting with database. The software tool used in this research has specifications as described below.

- Operating system: Genuine Microsoft Windows XP professional version 2002 service pack 3
- Web server: Apache web server 2.2.8
- Web browser: Google chrome 21.0.1180.79
- Web development tool.: PHP script language 5.2.6
- Database: MySQL 5.0.51b
- Database management system: phpMyAdmin 2.10.3
- Ontology editor: Protégé 3.4.1
- Fuzzy logic tool: MATLAB 7.1
- Toolkit for connecting with ontology: RDF API for PHP (RAP)
- Hierarchical chart visualizer: PowerCharts version 3

3.3 Data Collection

This research collects the dataset of 400 reviews from TripAdvisor.com for developing the knowledge base of the designed module. It is randomly selected from several accommodations and divided into 5 satisfaction levels (80 reviews in each level).

In order to evaluate the designed module, this experiment uses the new dataset of 200 reviews that is not the same dataset used for the knowledge base construction. The dataset for the module evaluation also collects from TripAdvisor.com, which is randomly selected from several accommodations. It covers all 5 satisfaction levels (40 reviews in each level). Within those 200 reviews, they consist of 1,382 criticisms, 211 non-criticisms, and 97 criticisms with errors, such as typographical, grammatical, spelling, and vocabulary errors. These 97 criticisms with errors are pruned because they are out of scope in this research.

A criticism means an opinion that judges the qualities of the accommodations, which may be clearly expressed as the word of feature (including feature relationship) and feature modifier. On the other hand, a non-criticism means an opinion given in the review that does not judges the qualities of the accommodations. These criticisms and non-criticisms are classified using a language expert in answering a question "Is it a criticism?", as illustrated in Figure 3.24.

<u>Classifier question</u>: *Is it a criticism?*

Review 1

__×__ Where do I begin?
__✓__ The rooms are ok.
__✓__ The food is terrible (for the price).
__✓__ Would not recommend it at all.
__✓__ There are much better hotels in Sydney.

Review 2

__×__ We stayed at this hotel for two nights while we were vacationing in Bali January this year.
__×__ Checked in at midnight in the rain.
__✓__ Front desk guy was the worst.
__✓__ Staff was rude.
__✓__ Room smelled.

Figure 3.24 An example of review classification

3.4 Data Analysis

The knowledge inference engine, a core engine of the designed module, comprises the processes of feature extraction and accommodation rating. The feature extraction is a process for digesting accommodation features, feature modifiers, and relationships among these features from review contents while the accommodation rating performs the computation of tourists' satisfaction level on each accommodation feature and on the overall accommodation rating. Therefore the module evaluation and experimental result analysis is divided into two parts as follows.

3.4.1 Result Analysis of Feature Extraction Evaluation

As discussed above, each review comprises both criticisms and non-criticisms. To evaluate the accuracy of the feature extraction process, only criticisms given in the review dataset are concentrated. Thus, if feature, feature modifiers, and relationships among these features could be extracted, that means a criticism could be regarded as an extracted criticism.

For example, suppose that the criticism is *"The air condition was almost very effective."*. Certainly, it is regarded as a criticism because it is an opinion that judges the qualities of the accommodations. There are the feature *"air condition"* related to the "*room*" feature, and the feature modifier *"almost very effective"* appeared in the criticism. If the review analysis module cannot extract all of the features, the feature modifiers, and the feature relationships, the criticism is justified as a non-complete criticism or a non-extracted criticism because the accommodation rating process cannot handle with this criticism.

In the accommodation rating process, the tourists' satisfaction level cannot be calculated under the non-complete criticism because the calculation must use all of the features, feature modifiers, and feature relationships. Hence, the criticism would be regarded as a complete criticism or extracted criticism, if and only if feature, feature modifier, and feature relationship can be extracted.

The results of feature extraction evaluation are defined in four terms, as shown in Table 3.6 (Miao et al., 2009: 9172).

Table 3.6 The confusion matrix of the terms of criticism classification

	Data Extracted (+)	**Data Not Extracted (-)**
Criticism (+)	*TP*	*FN*
Non-Criticism (-)	*FP*	*TN*

- *TP (True Positive):* the number of criticisms that can be correctly extracted.
- *FP (False Positive):* the number of non-criticisms that is extracted.
- *FN (False Negative):* the number of criticisms that cannot be extracted.
- *TN (True Negative):* the number of non-criticisms that is not extracted.

Using these terms, the performance of the feature extraction process should be evaluated by Accuracy, Precision, and Recall measure as defined in Equation 3.5-3.7, respectively (Miao et al., 2009: 9172).

$$Recall = \frac{TP}{(TP + FN)} \times 100\% \tag{3.5}$$

$$Precision = \frac{TP}{(TP + FP)} \times 100\% \tag{3.6}$$

$$Accuracy = \frac{TP + TN}{(TP + FP + FN + TN)} \times 100\% \tag{3.7}$$

3.4.2 Result Analysis of Accommodation Rating Evaluation

The accommodation rating process is evaluated in the perspective of the difference of the tourist-defined rating and module-extracted rating. For this evaluation, the experiment uses all criticisms which mention to the travel accommodation feature (including criticisms that could not be extracted from the feature extraction process). All the criticisms in the review dataset have been manually extracted and then they are collected into the extracted criticism dataset. Finally, these extracted criticisms are calculated the accommodation rating and applied for measuring the Mean Absolute Error (*MAE*) and Root Mean Squared Error (*RMSE*) of the designed module as given in Equation 3.8 and 3.9 (Memmedli and Ozdemi, 2010: 15).

$$MAE = \frac{1}{n} \sum_{t=1}^{n} \left| A_t - F_t \right| \tag{3.8}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(A_t - F_t\right)^2} \tag{3.9}$$

Where A_t is the actual tourist-defined rating of the review *t*, F_t is the module-computed rating of the review *t*, and *n* is the number of all extracted reviews.

Evaluating in this respect aims to take into account the tourist-defined rating as a goal. Thus, this experiment compares the errors between the module-computed rating using the non fuzzy-based method and the module-computed rating using the fuzzy-based method.

CHAPTER 4

THE RESULTS OF THE STUDY AND DISCUSSIONS

This chapter presents the results of the design of an analysis module for online travel accommodation reviews. As previously discussed in the Chapter 3, the designed module is evaluated by testing an accuracy of the knowledge inference engine which is the feature extraction and accommodation rating processes. Furthermore, the experimental results for finding the best relation score between adverbs and adjectives that gives the overall rating closest to tourists' subject are included. Finally, all results of the study are discussed.

4.1 The Results of the Evaluation of Feature Extraction Process

The evaluation of the feature extraction process is performed by measuring the accuracy of criticism extraction in reviews. A criticism is composed of a feature, a feature modifier, and the feature relationship. To evaluate the performance of the feature extraction process, the new dataset of 200 reviews are used. Within those 200 reviews, they consist of 1,382 criticisms and 211 non-criticisms as previously described. The experimental results are separately considered in each level of tourists' satisfaction as follows.

- **Level 1 (Terrible)**: There are 244 criticisms and 38 non-criticisms in this satisfaction level. As presented in Table 4.1, there are 173 criticisms that can be correctly extracted (*TP* = 173), 71 criticisms that cannot be extracted (*FN* = 71), and 38 non-criticisms that are not extracted (*TN* = 38). Therefore, the experimental results are achieved in 74.82% of Accuracy, 100% of Precision, and 70.90% of Recall, as shown in Equation 4.1, 4.2, and 4.3, respectively.

Table 4.1 The confusion matrix of the terms of criticism classification in the satisfaction level 1

	Data Extracted (+)	**Data Not Extracted (-)**
Criticism (+)	*TP=173*	*FN=71*
Non-Criticism (-)	*FP=0*	*TN=38*

$$Accuracy = \frac{173+38}{(173+71+0+38)} \times 100\% = \mathbf{74.82\%} \tag{4.1}$$

$$Precision = \frac{173}{(173+0)} \times 100\% = \mathbf{100\%} \tag{4.2}$$

$$Recall = \frac{173}{(173+71)} \times 100\% = \mathbf{70.90\%} \tag{4.3}$$

- **Level 2 (Poor)**: There are 289 criticisms and 35 non-criticisms in this satisfaction level. As presented in Table 4.2, there are 209 criticisms that can be correctly extracted (*TP* = 209), 80 criticisms that cannot be extracted (*FN* = 80), and 35 non-criticisms that are not extracted (*TN* = 35). Therefore, the experimental results are achieved in 75.31% of Accuracy, 100% of Precision, and 72.32% of Recall, as shown in Equation 4.4, 4.5, and 4.6, respectively.

Table 4.2 The confusion matrix of the terms of criticism classification in the satisfaction level 2

	Data Extracted (+)	**Data Not Extracted (-)**
Criticism (+)	*TP=209*	*FN=80*
Non-Criticism (-)	*FP=0*	*TN=35*

$$Accuracy = \frac{209+35}{(209+80+0+35)} \times 100\% = \mathbf{75.31\%} \quad (4.4)$$

$$Precision = \frac{209}{(209+0)} \times 100\% = \mathbf{100\%} \quad (4.5)$$

$$Recall = \frac{209}{(209+80)} \times 100\% = \mathbf{72.32\%} \quad (4.6)$$

- Level 3 (Average): There are 325 criticisms and 44 non-criticisms in this satisfaction level. As presented in Table 4.3, there are 244 criticisms that can be correctly extracted (*TP* = 244), 81 criticisms that cannot be extracted (*FN* = 81), and 44 non-criticisms that are not extracted (*TN* = 44). Therefore, the experimental results are achieved in 78.05% of Accuracy, 100% of Precision, and 75.08% of Recall, as shown in Equation 4.7, 4.8, and 4.9, respectively.

Table 4.3 The confusion matrix of the terms of criticism classification in the satisfaction level 3

	Data Extracted (+)	**Data Not Extracted (-)**
Criticism (+)	*TP=244*	*FN=81*
Non-Criticism (-)	*FP=0*	*TN=44*

$$Accuracy = \frac{244+44}{(244+81+0+44)} \times 100\% = \mathbf{78.05\%} \quad (4.7)$$

$$Precision = \frac{244}{(244+0)} \times 100\% = \mathbf{100\%} \quad (4.8)$$

$$Recall = \frac{244}{(244+81)} \times 100\% = \mathbf{75.08\%} \quad (4.9)$$

- **Level 4 (Good)**: There are 271 criticisms and 42 non-criticisms in this satisfaction level. As presented in Table 4.4, there are 221 criticisms that can be correctly extracted (*TP* = 221), 50 criticisms that cannot be extracted (*FN* = 50), and 42 non-criticisms that are not extracted (*TN* = 42). Therefore, the experimental results are achieved in 84.03% of Accuracy, 100% of Precision, and 81.55% of Recall, as shown in Equation 4.10, 4.11, and 4.12, respectively.

Table 4.4 The confusion matrix of the terms of criticism classification in the satisfaction level 4

	Data Extracted (+)	**Data Not Extracted (-)**
Criticism (+)	*TP=221*	*FN=50*
Non-Criticism (-)	*FP=0*	*TN=42*

$$Accuracy = \frac{221+42}{(221+50+0+42)} \times 100\% = \mathbf{84.03\%} \quad (4.10)$$

$$Precision = \frac{221}{(221+0)} \times 100\% = \mathbf{100\%} \quad (4.11)$$

$$Recall = \frac{221}{(221+50)} \times 100\% = \mathbf{81.55\%} \quad (4.12)$$

- Level 5 (Excellent): There are 253 criticisms and 52 non-criticisms in this satisfaction level. As presented in Table 4.5, there are 204 criticisms that can be correctly extracted (*TP* = 204), 49 criticisms that cannot be extracted (*FN* = 49), and 52 non-criticisms that are not extracted (*TN* = 52). Therefore, the experimental results are achieved in 83.93% of Accuracy, 100% of Precision, and 80.63% of Recall, as shown in Equation 4.13, 4.14, and 4.15, respectively.

Table 4.5 The confusion matrix of the terms of criticism classification in the satisfaction level 5

	Data Extracted (+)	**Data Not Extracted (-)**
Criticism (+)	*TP=204*	*FN=49*
Non-Criticism (-)	*FP=0*	*TN=52*

$$Accuracy = \frac{204+52}{(204+49+0+52)} \times 100\% = \mathbf{83.93\%} \tag{4.13}$$

$$Precision = \frac{204}{(204+0)} \times 100\% = \mathbf{100\%} \tag{4.14}$$

$$Recall = \frac{204}{(204+49)} \times 100\% = \mathbf{80.63\%} \tag{4.15}$$

- **Overall**: There are 1,382 criticisms and 211 non-criticisms in this satisfaction level. As presented in Table 4.6, there are 1,051 criticisms that can be correctly extracted (*TP* = 1,051), 331 criticisms that cannot be extracted (*FN* = 331), and 211 non-criticisms that are not extracted (*TN* = 211). Therefore, the experimental results are achieved in 79.22% of Accuracy, 100% of Precision, and 76.05% of Recall, as shown in Equation 4.16, 4.17, and 4.18, respectively.

Table 4.6 The confusion matrix of the terms of criticism classification in overall satisfaction level

	Data Extracted (+)	**Data Not Extracted (-)**
Criticism (+)	*TP=1,051*	*FN=331*
Non-Criticism (-)	*FP=0*	*TN=211*

$$Accuracy = \frac{1,051+211}{(1,051+331+0+211)} \times 100\% = \mathbf{79.22\%} \tag{4.16}$$

$$Precision = \frac{1,051}{(1,051+0)} \times 100\% = \mathbf{100\%} \tag{4.17}$$

$$Recall = \frac{1,051}{(1,051+331)} \times 100\% = \mathbf{76.05\%} \tag{4.18}$$

According to the evaluation of the feature extraction process, the experimental results are illustrated in Figure 4.1 and expounded below.

1) There are no non-criticisms that are extracted (*FP=0*) in every level, because the accommodation features do not appear in these non-criticisms. As previously described, a criticism is regarded as an extracted criticism, if a feature, a feature modifier, and a feature relationship could be extracted. Therefore, all non-criticisms cannot be extracted owing to the feature absence. As a result, the Precision measure gets 100% because *FP* is zero according to the equation:

$$precision = \frac{TP}{(TP+FP)} \times 100\% = \frac{TP}{(TP+0)} \times 100\% = 100\% .$$

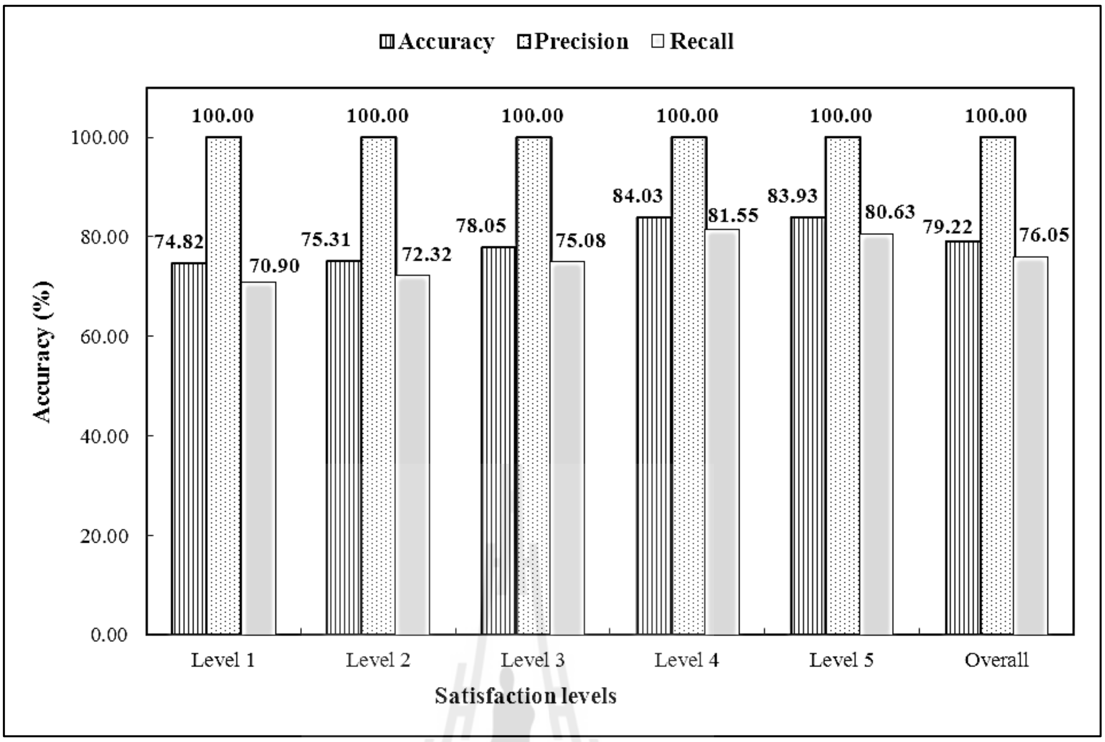


Figure 4.1 The Accuracy, Precision, and Recall of the feature extraction process

2) It is obviously seen that all of the satisfaction levels achieve more than 70% of Recall. This outcome implies that the extraction process correctly extracts all of accommodation features, their feature modifiers, and relationships among these features in every satisfaction level. All satisfaction levels arranged in descending order of Recall are as follows: level 4 (81.55%), level 5 (80.63%), level 3 (75.08%), level 2 (72.32%), and level 1 (70.90%). Nevertheless, the designed module could have the maximum benefit, if the reviews are written without any error.

By the way, the extraction process works incorrectly in some levels, particularly in level 1 and 2 which is obtained the minimum of Recall (only 70.90% and 72.32%). Because the tourist reviews in both levels always criticize with long explanations, for instance, *"I followed the front desk instruction for shuttle bus,* *because my flight was late. The front desk told me to pick me up. So I waited for the shuttle bus for a long time during midnight. But the shuttle bus never comes to the airport. 2:00 AM, I went to this hotel by taxi.... And there is a shuttle bus in front of the hotel.....without driver... I saw a creepy guy in the front desk....".*

Following the review example, there are many significant keywords that are useful for semantic analysis and appear in this long explanation criticism. These keywords are separated in many sentences and omitted in some cases. Thus, the designed semantic analysis approach does not support the long explanation criticism. Furthermore, many commentary words or phrases are in need of connotative meaning interpretation or pragmatic analysis that is out of scope of this research.

In addition, there are many pronouns or noun phrases that require determining the antecedent of an anaphor (or anaphora resolution), and some of the review contexts are inappropriate for applying the anaphora resolution algorithm. For instance, "*The room was clean and **one** was an ok size.*" and "*Chris and Maria were wonderful hosts. **Both** were incredibly friendly and **they** were especially helpful in directing **us** around London and booking tickets for **us** for different attractions right from their front desk.".* For these reasons, some criticisms are incorrectly extracted.

3) Due to that fact that *FP* is zero, *TN* is the total number of non-criticisms in the evaluation dataset. Notice that, the total number of non-criticisms (211 non-criticisms) is obviously less than that of criticisms (1,382 criticisms). Therefore, the Accuracy values are too close to the Recall values (about 3% difference in each level) because of the low *TN*. However, this research focuses on the correctness of criticisms, not non-criticisms. Therefore, the evaluation of the feature extraction process focuses on the Recall measure (derived from *TP* and *FN*).

4.2 The Results of the Relationship between Adverbs and Adjectives

As stated in the Chapter 3, the feature rating would be calculated from the scores of adjective, special verb, special word, adverb, and negation adverb words. In case of Adverb Adjective Combination (AAC), this research also presents the experimental results for finding the best relation score between adverbs and adjectives (*r*) that gives the overall rating closest to tourists' subject.

To achieve the objective of the study in computing the entire review scores correctly, this experiment aims to find out the relation score that completes the lowest errors of module-computed rating when compared with tourist-defined rating. The module-computed rating in this experiment is based on three feature rating methods that are two fuzzy-based methods and one non fuzzy-based method.

In Figure 4.2, various relation scores that are implemented in three feature rating methods are compared using a standard measure, named Mean Absolute Error (*MAE*). The experimental results indicate that the relation score 0.9 provides the lowest *MAE* for the fuzzy-based method with probability weighting and the non fuzzy-based method (or average rating method) whereas the fuzzy-based method with equal weighting gets the relation score 1.

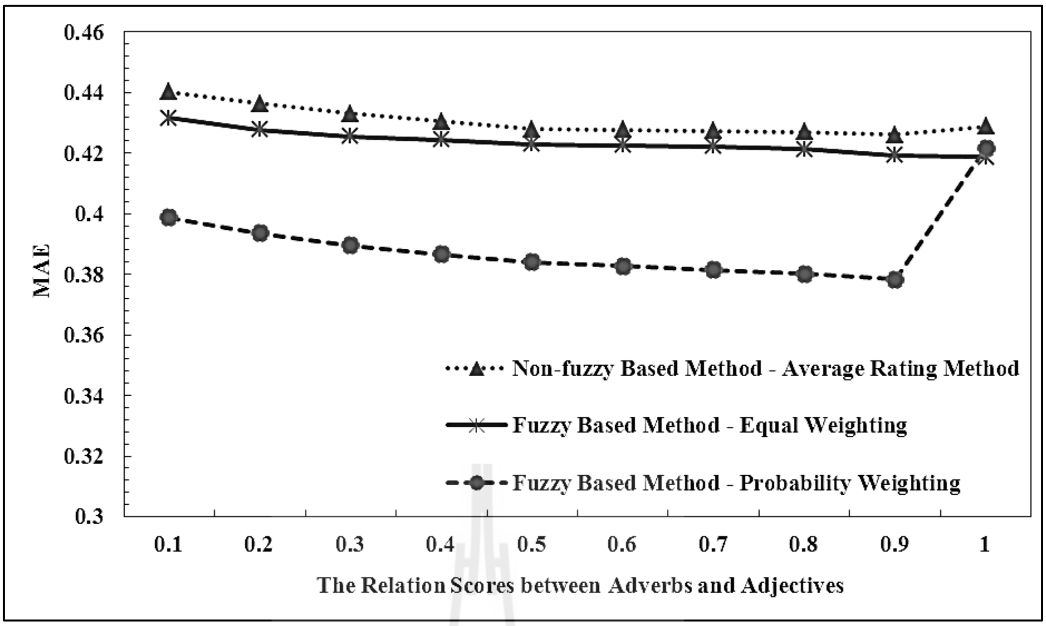


Figure 4.2 The *MAE* of the module-computed rating using the fuzzy-based methods and non fuzzy-based method in various relation scores

4.3 The Results of the Evaluation of Accommodation Rating Process

This section presents the evaluation of accommodation rating process which performs on the perspective of the difference of tourist-defined rating and module-computed rating. For this evaluation, the experiment uses the dataset of 200 reviews covered all 5 satisfaction levels (40 reviews for each level) as same as the evaluation of the feature extraction process presented in the Section 4.1.

As previously described, each satisfaction level of reviews has a different number of criticisms as follows: Level 1 has 244 criticisms, Level 2 has 289 criticisms, Level 3 has 325 criticisms, Level 4 has 271 criticisms, and Level 5 has 253 criticisms. In these criticisms, there are some criticisms that could not be extracted in the previous feature extraction process. However, the non-extracted criticisms are manually extracted to obtain the correct data, like the extracted criticisms. Thus, in the evaluation of the accommodation rating process, entire criticisms in the review dataset are used for measuring the Mean Absolute Error (*MAE*) and Root Mean Squared Error (*RMSE*) of the designed rating methods. Kindly note that, evaluating in this respect aims to take into account the tourist-defined rating as a goal. Thus, this experiment compares the errors of the module-computed rating between using the non fuzzy-based method versus fuzzy-based methods with equal weighting and probability weighting as illustrated in Table 4.7.

Table 4.7 The *MAE* and *RMSE* of the module-computed rating between the non fuzzy-based method and fuzzy-based method with equal weighting

Level	**Non Fuzzy-Based Method**		**Fuzzy-Based Method**	
	Average Rating		Equal Weighting	
	MAE	*RMSE*	*MAE*	*RMSE*
Level 1	0.410	0.536	0.406	0.523
Level 2	0.494	0.649	0.484	0.621
Level 3	0.507	0.604	0.479	0.581
Level 4	0.375	0.489	0.377	0.492
Level 5	0.344	0.447	0.347	0.453
Overall	***0.426***	***0.550***	***0.419***	***0.537***

In Table 4.7, the *MAE* and *RMSE* of the overall rating using the fuzzy-based method with equal weighting (at relation score (*r*)=1) and non fuzzy-based method (at *r*=0.9) are compared. The experimental results indicate that the *RMSE* of every level of both methods is approximately 0.5 which implies that every test case in each level obtains nearby errors while the *MAE* of the fuzzy-based method using equal weighting is marginally less than the others.

Due to the nearby errors between the fuzzy-based method with equal weighting and non fuzzy-based method stated above, the fuzzy-based method with probability weighting is designed for decreasing the errors.

The *MAE* and *RMSE* of the overall rating using the fuzzy-based method with probability weighting and non fuzzy-based method (at r=0.9) are compared in Table 4.8. The experimental results indicate that the *RMSE* of every level of both methods is approximately 0.5 which implies that every test case in each level obtains nearby errors. Furthermore, the *MAE* of the fuzzy-based method with probability weighting is more decreased but marginally less than other methods as same as the experiment results discussed in Table 4.7.

Table 4.8 The *MAE* and *RMSE* of the module-computed rating between the non fuzzy-based method and fuzzy-based method with probability weighting

Level	**Non Fuzzy-Based Method**		**Fuzzy-Based Method**	
	Average Rating		Probability Weighting	
	MAE	*RMSE*	*MAE*	*RMSE*
Level 1	0.410	0.536	0.352	0.454
Level 2	0.494	0.649	0.393	0.529
Level 3	0.507	0.604	0.448	0.544
Level 4	0.375	0.489	0.362	0.468
Level 5	0.344	0.447	0.338	0.439
Overall	***0.426***	***0.550***	***0.378***	***0.489***

Although, the overall error (*MAE*) is not significantly different, the percentage of test cases that errors are more than 0.5 points are clearly decreased in every satisfaction level by using the fuzzy-based method with probability weighting, as illustrated in Table 4.9. The results show that the number of test cases with errors less than 0.5 points reaches 67.5% by using the fuzzy logic-based method with probability weighting while using the non fuzzy-based method or average rating method gets only 58.50%. In other words, the fuzzy logic-based method with probability weighting performs closer rating to the tourist-defined rating (or tourists' subject) more than the non fuzzy-based method.

Table 4.9 A comparison of the percentage of errors between the module-computed rating using fuzzy-based method with probability weighting and non fuzzy-based method

Level	**Test Case (%)** (Non-Fuzzy)				**Test Case (%)** (Fuzzy)			
	Average Rating				Probability Weighting			
	MAE < 0.25	*MAE >= 0.25*	*MAE >= 0.5*	*MAE >= 0.75*	*MAE < 0.25*	*MAE >= 0.25*	*MAE >= 0.5*	*MAE >= 0.75*
Level 1	32.50	22.50	25.00	20.00	40.00	30.00	20.00	10.00
Level 2	37.50	27.50	2.50	32.50	45.00	22.50	12.50	20.00
Level 3	25.00	17.50	30.00	27.50	35.00	20.00	25.00	20.00
Level 4	42.50	22.50	17.50	17.50	45.00	25.00	17.50	12.50
Level 5	40.00	25.00	25.00	10.00	40.00	35.00	15.00	10.00
Overall	***35.50***	***23.00***	***20.00***	***21.50***	***41.00***	***26.50***	***18.00***	***14.50***

4.4 Discussions

According to the evaluation results of the knowledge inference engine, many reasons for the incorrect feature extraction and faulty accommodation rating are discussed as follows.

4.4.1 The lacking of words in the terminology (adjective, special verb, special word, adverb, and negation adverb), synonyms in the SKOS ontology, and accommodation features in the tourism ontology. These words, synonyms, and accommodation features are necessary resources for the feature extraction process. Therefore, the lacking of them causes the incorrect feature extraction.

4.4.2 The lacking of other related information units, like specific names (e.g. persons, organizations, and locations) and numeric expressions (e.g. time, date, quantities, monetary values, and percentages). These information units also affect the feature extraction process. Therefore, the Named Entity Recognition and Classification (NERC) task is needed.

4.4.3 The lacking of several of context free grammar rules. In the syntactic analysis of the feature extraction process, if the context free grammar rules are not cover the most of criticism sentence forms, the accuracy results of the feature extraction process is declined. Unfortunately, the research gathers inadequate number of context free grammar rules because there are various forms of criticism sentences based on individual language expression. Consequently, many criticisms could not be extracted as obviously seen in the satisfaction level 1 and 2 which obtained the lowest Recall. It is due to the fact that most tourists criticize with long explanation (or ambiguous semantic analysis), as well as many commentary words or phrases are in need of connotative meaning interpretation or pragmatic analysis.

4.4.4 The rating scores or assigned satisfaction level of each word in the terminology are assigned by a language expert that is the cause of the faulty accommodation rating. It is due to the fact that the rating scale does not conformed to tourists' opinion. Therefore, the module-computed rating differs from the tourist-defined rating.

4.5 The Results of the Hypothesis Testing

As mentioned in the Chapter 1, the main purposes of this research are to design a review analysis module that can extract the feature relationships and compute the accommodation rating that is closest to rating by tourists (or human) as much as possible. The two main hypothesis of this research are 1) The analysis module for online travel accommodation reviews can extract feature relationships correctly with more than 80% of Recall and 2) The analysis module for online travel accommodation reviews can compute the entire review scores correctly, whereas the difference between the module-computed scores and the tourist-specified scores is less than 0.1 review points from 5 full scores.

The First Research Hypothesis: the evaluation results of the feature extraction process showed that the extraction process correctly extracts accommodation features, their feature modifiers, and relationships among these features in every satisfaction level, especially level 4 and 5 which can be accomplished more than 80% of Recall. On the other hand, the evaluation results of level 1, 2, 3, and overall are marginally less than 80% of Recall.

The extraction process works incorrectly in the level 1-3 because the tourist reviews in these levels always criticize with long explanation. In this case, many significant keywords that are useful for semantic analysis are separated in many sentences and omitted in some cases. Thus, the designed semantic analysis approach does not support the long explanation criticism. Furthermore, many commentary words or phrases are in need of connotative meaning interpretation or pragmatic analysis that is out of scope of this research.

In addition, there are many pronouns or noun phrases that require determining the antecedent of an anaphor (or anaphora resolution), and some of the review contexts are inappropriate for applying the anaphora resolution algorithm. For these reasons, some criticisms are incorrectly extracted.

Following these results, the first research hypothesis can be concluded that the evaluation results conform to this hypothesis at the satisfaction level 4 and 5 while the satisfaction level 1, 2, 3, and the overall are inconsistency.

The Second Research Hypothesis: the evaluation results of the accommodation rating process showed that the designed module can compute the entire review scores (or overall rating) that marginally differs from the tourist-defined rating with 0.378 of the Mean Absolute Error (*MAE*) and 0.489 of the Root Mean Squared Error (*RMSE*). Following these results, the second research hypothesis can be concluded that the evaluation results do not conform to this hypothesis. However, the percentage of test cases that errors are more than 0.5 points is clearly decreased in every satisfaction level as illustrated in Table 4.9.

Although, some experimental results are not in line with both hypotheses, they obtain nearby results. Several improvements of the review analysis module are performed to enhance the correctness. Unfortunately, owing to many reasons that cause the incorrect feature extraction and faulty accommodation rating as previously discussed, those experimental results are marginally dissimilar to the hypotheses.

CHAPTER 5

CONCLUSIONS AND RESEARCH RECOMMENDATIONS

This chapter presents a summary of the research findings, the limitation of the study, the application of the study and recommendations for further research studies.

5.1 Summary of the Research Findings

This study is a design of an analysis module for online reviews, which uses the tourism information on travel accommodations as a case study. This research proposes a module design for analyzing unstructured text stored in the travel accommodation reviews. It starts with parsing sentences of accommodation reviews to identify keywords related to the accommodation topic, i.e. accommodation features (noun), feature modifiers (adjective, special verb, special word, adverb, and negation adverb), and relationships among these features. The process of parsing sentences is called feature extraction process. The feature extraction method is used to solve the sentence analysis problem by applying tourism ontology to extract and store accommodation feature. In addition, this research proposes a fuzzy-based method for calculating a tourists' satisfaction level on each extracted feature and on the entire review. Finally, it explains more clear and easy understanding reviews with 5-rating scale and visualizes the accommodation feature relationships in hierarchy as knowledge representation of the accommodation.

According to the related work on online review analysis, four components, i.e. user interface, knowledge inference engine, knowledge base, and knowledge explanation engine are designed. Moreover, the details of accommodation categories presented in the knowledge representation are derived from the research study in tourist's satisfaction in order to satisfy the needs of tourists as much as possible.

This designed module is an Internet-based application that is implemented with PHP language. The PowerCharts Version 3 is also used for visualizing accommodation features and ratings in hierarchy. Furthermore, MySQL and RDF API for PHP (RAP) are used as database and toolkit for connecting with ontology, respectively.

The review analysis module is evaluated by testing an accuracy of the feature extraction and the accommodation rating processes. The Accuracy, Precision, and Recall measure is used to evaluate the performance of the feature extraction process while the accommodation rating process is evaluated in the perspective of the difference of tourist-defined rating and module-computed rating using Mean Absolute Error (*MAE*) and Root Mean Squared Error (*RMSE*) measures. Moreover, this research provides the experimental results for finding the best relation score (*r*) that gives the overall rating closest to the tourist subject.

The research findings are summarized as follows:

5.1.1 In order to evaluate the feature extraction process, the new dataset of 200 reviews from TripAdvisor.com randomly selected from several accommodations were used. It covers all 5 satisfaction levels (40 reviews in each level) consisting of 1,382 criticisms and 211 non-criticisms. There are 1,051 criticisms that could be correctly extracted (*TP* = 1,051), 331 criticisms that cannot be extracted (*FN* = 331), and 211 non-criticisms that are not extracted (*TN* = 211). As a result, the feature extraction process is achieved in 79.22% of overall accuracy, 100% of overall precision, and 76.05% of overall recall.

5.1.2 In case of the feature rating assigned by the scores of Adverb Adjective Combination (AAC), the best score of relation between adverbs and adjectives (*r*) from the experiments is 0.9 in cases of the fuzzy-based method with probability weighting and the non fuzzy-based method (or average rating method) whereas the fuzzy-based method with equal weighting gets the relation score 1.

5.1.3 The designed score calculated from the fuzzy-based method with probability weighting for the feature rating is achieved in 0.378 of *MAE* and 0.489 of *RMSE*. These values indicate that the module-computed rating marginally differs from tourist-defined rating and the errors of every test case are nearby.

According to the processes of module design and evaluation, the design of a reviews analysis module conforms to the two main motivations of this research that are 1) Tourists want to know about the accommodation in details and 2) Tourists does not read enormous reviews. It is due to the fact that the designed module could automatically analysis the accommodation reviews with more details, such as accommodation features, feature relationships, and feature scores. In addition, the designed module can explain more clear and easy understanding reviews with 5-rating scale and visualize the accommodation feature relationships in hierarchy.

5.2 The Limitation of the Study

The limitation of the design of an analysis module for online travel accommodation reviews are described as follows.

5.2.1 This research aims to understand natural language by analyzing unstructured texts stored in accommodation reviews. Due to a natural language is typically used for human communication, such as speaking and writing, the language usage bases on individual expression which is an arbitrary distinction and not always in grammatical correction. Consequently, many criticisms could not be extracted by the designed module because the context free grammar rules are insufficient and not cover various forms of criticism sentences.

5.2.2 Since the natural language evolve and diversify all the time, many emerging words or phrases including idioms, proverbs, and slangs, are often written in online reviews. Certainly, these emerging words or phrases are used as criticisms. Unfortunately, the designed module does not automatically update those words or phrases in the knowledge base (i.e. terminology, synonyms in the SKOS ontology, and accommodation features in the tourism ontology). Therefore, the criticisms written by these emerging words or phrases could not be extracted.

5.3 The Application of the Study

The benefit of this research could be useful in tourism business. The tourists can apply the extracted knowledge to support their decisions on selecting accommodations easily and quickly. Moreover, tourism business entrepreneurs can use this module to analyze customers' opinions that affect their business and apply the extracted knowledge for developing their services and facilities in order to meet more customers' need and gain more advantages over the competitors. Furthermore, the extracted knowledge is also used for supporting strategic management and tourism business policy for tourism business. In addition, the designed module is applied to analyze other reviews, such as book reviews, product reviews, restaurant reviews, shop reviews, and tourist attraction reviews.

5.4 Recommendations for Further Study

There are some improvements that could be performed in the near future as described below:

5.4.1 Updating the knowledge base which contains the necessary resources of both feature extraction and accommodation rating processes in order to provide higher accuracy of results, such as collecting more related words in terminology and synonyms in SKOS ontology and increasing accommodation features in the tourism ontology.

5.4.2 Improving the semantic analysis process on aspects of word sense ambiguity. There are some adjective words with a neutral connotation that are used to criticize in either positive or negative senses, such as small, large, and silence. Therefore, what rating score should be assigned to these words? Considering each one of these words together with the feature word could be identified an appropriate rating score. Finally, a pair of the neutral connotation word with its rating score and the feature word should be assigned as a pattern for extracting in the future.

5.4.3 Reviewing pragmatic analysis in the Natural Language Processing and applying this knowledge to the review analysis module. The pragmatic analysis is concerned with the purposeful use of language in situations and utilizes context over and above the contents of the text for understanding. The goal of the pragmatic analysis is to explain how extra meaning is *read into* texts without actually being encoded in them. In this research, the pragmatic analysis concept is useful for interpreting the connotative meaning of the commentary words or phrases appeared in the accommodation reviews.

5.4.4 Improving the set of context free grammar rules in order to analyze complex sentences that partly omit constituents of a clause, such as subject or conjunction. For instance, the sentence “Room allotted to me was not good.” is a complex sentence.

5.4.5 Adjusting the designed module in order to extract reviews that are written in Thai language. To support this adjustment, all resources in the knowledge base need to adapt, i.e. related words in the terminology, synonyms in the SKOS ontology, accommodation features in the tourism ontology, frequently-used vocabularies in dictionary, and context-free grammar rules for the Thai language. Moreover, the feature extraction process should be specially adapted to apply for the Thai language.

REFERENCES

Alani, H. et al. (2003). Automatic ontology-based knowledge extraction from Web documents. **IEEE Intelligent Systems**. 18 (1): 14-21.

Ananthapadmanaba, K. R. and Srivats, S. K. (2011). Personalization of user profile: Creating user profile ontology for Tamilnadu tourism. **International Journal of Computer Applications**. 23 (8): 42-47.

Andam, R. Z. (2003). **e-Commerce and e-Business**. Manila, Philippines: The e-ASEAN Task Force and the UNDP Asia Pacific Development Information Programme (UNDP-APDIP).

Anderson, M. (2010). **79% of 'younger consumers' trust online reviews as much as personal recommendations** [On-line]. Available: http://www.brightlocal.com/press-releases/79-of-younger-consumers-trust-online-reviews-as-much-as-personal-recommendations/

Assis, A., Yerazunis, W., Siefkes, C., and Chhabra, S. (2006). Exponential differential document count a feature selection factor for improving Bayesian filters accuracy. In **Proceeding of the 2006 Spam Conference** (pp 1-6). Cambridge, MA: MIT.

Bartolini, R. et al. (2006). Multimedia information extraction in ontology-based semantic annotation of product catalogues. In **Proceeding of Semantic Web Applications and Perspectives 2006** (pp 1-8). Pisa, Italy: CEUR-WS.

Benamara, F., Cesarano, C., Picariello, A., Subrahmanian, V. S., and Reforgiato, D. (2006). Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In **Proceeding of the International Conference on Weblogs and Social Media** (pp 1-7). Boulder, Colorado, U.S.A.: The Association for the Advancement of Artificial Intelligence.

Bosque, I. R. D. and Martin, H. S. (2008). Tourism research. **Annals of Tourism Research**. 35 (2): 551-573.

Burke, R R. (2002). Technology and the customer interface: what consumers want in the physical and virtual store. **The Academy of Marketing Science**. 30 (4): 411-432.

Cesarano, C. et al. (2006). OASYS: An opinion analysis system. In **Proceeding of the AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs** (pp 21-26). Stanford, California: Stanford University.

Chandrasekaran, B., Josephson, J. R., and Benjamins, V. R. (1999). What are ontologies, and why do we need them? **IEEE Intelligent Systems**. 14 (1): 20-26.

Chanlekha, H., and Kawtrakul, A. (2001). Thai named entity extraction by incorporating maximum entropy model with simple heuristic information. In **Proceeding of International Joint Conference on Natural Language Processing (IJCNLP-04)**. Hainan, China: Springer.

Chareonsuk, J., Sukvakree, T., and Kawtrakul, A. (2005). Elementary discourse unit segmentation for Thai using discourse cue and syntactic information. In **Proceeding of the 9th National Computer Science and Engineering Conference (NCSEC 2005)** (pp 897-902). Bangkok, Thailand: University of the Thai Chamber of Commerce.

Chi, C. G-Q. and Qu, H. (2008). Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. **Tourism Management**. 29 (4): 624-636.

Choi, C. et al. (2009). Travel ontology for intelligent recommendation system. In **Proceeding of the 2009 Third Asia International Conference on Modelling & Simulation** (pp 637-642). Bandung, Bali, Indonesia: IEEE.

Chunhua, W., Pengfei, Z., and Cong, W. (2006). A comprehensive tourism information system. In **Proceeding of the 2006 IEEE International Conference on Industrial Informatics** (pp 720-724). Singapore: IEEE.

Cowie, J. and Wilks, Y. (1996). Information extraction. **Communications of the ACM**. 39 (1): 80-91

Cracolici, M. F. and Nijkamp, P. (2008). The attractiveness and competitiveness of tourist destinations: A study of Southern Italian regions. **Tourism Management**. 30 (3): 336-344.

Dave, K., Lawrence, S., and Pennock, D.M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In **Proceeding of the 12th International Conference on World Wide Web** (pp 519-528). Budapest, Hungary: ACM.

Davenport, T. H. and Prusak, L. (1998). **Working knowledge: How organizations manage what they know**. Cambridge, MA: Harvard Business School Press.

Ding, Y., Herzog, C., Luger, M., Prantner, K., and Yan, Z. (2008). OnTourism: Semantic eTourism portal. In **Proceeding of the 2nd International Scientific Conference of the e-Business Forum - E-Business in Travel, Tourism and Hospitality** (pp 1-7). Athens, Greece: The Greek National Research and Technology Network, The University of the Aegean, and the Technological Educational Institute of Thessaloniki.

Fry, J. (2004). **Context-Free Grammars for English** [Data file]. California, United States: San José State University.

Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? topic maps! making sense of it all. **Journal of Information Science**. 30 (4): 378-391.

Grosz, B. J., Aravind, J., and Scott, W. (1983). Providing a unified account of definite noun phrases in discourse. In **Proceeding of the 21st Annual meeting of the Association for Computational Linguistics** (pp 44-50). Cambridge, Massachusetts: Association for Computational Linguistics.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. **Knowledge Acquisition**. 5 (2): 199-220.

Guarino, N. (1998). Formal ontology and information systems. In **Proceeding of the International Conference on Formal Ontology in Information Systems** (pp 3-15). Trento, Italy: IOS Press.

Haegeman, L. (1994). **Introduction to government and binding theory**. Oxford: Blackwell.

Halliday, M. and Hasan, R. (1976). **Cohesion in English (Longman English language series 9)**. London: Longman.

Haruechaiyasak, C., Srichaivattana, P., Kongyoung, S., and Damrongrat, C. (2004). Automatic Thai keyword extraction from categorized text corpus. In **Proceeding of the 1st ECTI Annual Conference** (pp 1-4). Pataya, Thailand: Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology association of Thailand.

Hepp, M., Siorpaes, K., and Bachlechner, D. (2006). Towards the semantic web in e-tourism: lack of semantics or lack of content?. In **Proceeding of the 3rd Annual European Semantic Web Conference** (pp 1-2). Budva, Montenegro: ESWC.

Hirst, G. (1981). **Anaphora in natural language understanding: A survey**. NewYork: Springer-Verlag.

Hobbs, J. R. (1978). Resolving pronoun references. **Lingua**. 44 (4): 311-338.

Hsu, T-K., Tsai, Y-F., and Wu, H-H. (2009). The preference analysis for tourist choice of destination: A case study of Taiwan. **Tourism Management**. 30 (2): 288-297.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In **Proceeding of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining** (pp 168-177). Washington, USA: ACM.

Hui, T.K., Wan, D., and Ho, A. (2007). Tourists' satisfaction, recommendation and revisiting Singapore. **Tourism Management**. 28 (4): 965-975.

Imsombut, A., Suktarachan, M., Yingsaree, W., and Kawtrakul A. (2005). Country report and activity from Thailand: Ontology construction and maintenance system in agricultural domain. In **Proceeding of the International Joint Conference of the Asian Federation of Information Technology in Agriculture and the World Congress on Computer in Agriculture and Natural Resources** (pp 62-74). Bangkok, Thailand: Hydro and Agro Informatics Institute.

Jakob, N. and Gurevych, I. (2010). Using anaphora resolution to improve opinion target identification in movie reviews. In **Proceeding of the 48th Annual Meeting of the Association for Computational Linguistics** (pp 263-268). Uppsala, Sweden: Uppsala University.

Jurafsky, D. and Martin, J. H. (2000). **Speech and language processing : An introduction to natural language processing, computational linguistics, and speech recognition**. New Jersey: Prentice Hall.

Kawtrakul, A. (2007). **Research report topic: A development of information and knowledge extraction from unstructured Thai document (I-Know)** [On-line]. Available: http://naist.cpe.ku.ac.th/iknow/.

Liddy, E.D. (2001). Natural language processing (2nd Edition). **Encyclopedia of Library and Information Science**. January 2001: 1-14.

Lorpraditpong, N. (2006). **Customer satisfaction survey handbook**. 2nd Edition. Bangkok: Prachoomthong Printing Group.

Macintosh, R.W. and Goeldner, C.R. (1986). **Tourism principles, practices, philosophies**. 5th edition. New York: John Wiley & Sons.

Memmedli, M. and Ozdemi, O. (2010). An application of fuzzy time series to improve ISE forecasting. **WSEAS Transactions on Mathematics**. 9 (1): 12-21.

Miao, D., Duan, Q., Zhang, H., and Jiao, N. (2009). Rough set based hybrid algorithm for text classification. **Expert Systems with Applications**. 36 (5): 9168-9174.

Mitkov, R. (1995). **Anaphora resolution in natural language processing and machine translation**. Saarbrücken: Institut der Gesellschaft zur Förderung der Angewandten Informationsforschung (IAI).

Mitkov, R. (1999). Anaphora resolution: The state of the art. Working paper In **Proceeding of COLING'98/ACL'98 Tutorial on Anaphora Resolution** (pp 1-34). United Kingdom: University of Wolverhampton.

Nguyen, H.T. and Walker, E.A. (2000). **A first course in fuzzy logic**. Boca Raton, FL: Chapman & Hall.

Nielsen. (2009). **Nielsen global online consumer survey trust, value and engagement in advertising** [On-line]. Available: http://www.nielsen.com/content/dam/corporate/us/en/newswire/uploads/2009/07/trustinadvertising0709.pdf

Noroozi, A. and Fotouhi, Z. (2010). The influence of semantic web on decision making of customers in tourism industry. **International Journal of Information Science and Management**. 2010 (Special Issue 1): 77-98.

Nysveen, H. and Pedersen, P. E. (2004). An exploratory study of customers' perception of company web sites offering various interactive applications: moderating effects of customers' internet experience. **Decision Support Systems**. 37 (1): 137-150.

Peerawit, W., Yingsaeree, W., and Kawtrakul, A. (2004). The utilization of closing algorithm and heuristic information for broken character segmentation. In **Proceeding of Cybernatics and Intelligent Systems** (pp 775-779). Singapore: IEEE.

Pizam, A., Neumann, Y., and Reichel, A. (1978). Dimensions of tourist satisfaction with a destination area. **Annals of Tourism Research**. 5 (3): 314-322.

Popov, B. et al. (2003). Towards semantic web information extraction. In **Proceeding of the 2nd International Semantic Web Conference** (pp 1-21). Florida, USA: ISWC.

Ramkumar, V., Rajasekar, S., and Swamynathan, S. (2010). Scoring products from reviews through application of fuzzy techniques. **Expert Systems with Applications**. 37 (10): 6862-6867.

Romero, F. P., Olivas, J. A., Genero, M., and Piattini, M. (2008). Automatic extraction of the main terminology used in empirical software engineering through text mining techniques. In **Proceeding of the 2nd ACM-IEEE International Symposium on Empirical Software Engineering and Measurement** (pp 357-358). Kaiserslautern, Germany: ACM.

Roopa, J., Mladen, G., and Nalin, S. (2007). Connecting destinations with an ontology-based e-tourism planner. In **Proceeding of the 14th Annual Conference of IFITT, the International Federation for IT & Travel and Tourism** (pp 21-32). Ljubljana, Slovenia: Springer Vienna.

Rosen, A. (2000). **The e-commerce question and answer book**. USA: American Management Association.

Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. **Information Processing and Management**. 24 (5): 513-523.

Sereerat, S. et al. (1995). **The quality of services**. Bangkok, Thailand: Prachachon.

Sidner, C. L. (1979). **Towards a computational theory of definite anaphora comprehension in English discourse**. Cambridge, MA: MIT Artificial Intelligence Laboratory.

Siorpaes, K., Prantner, K., and Bachlechner, D. (2004). **Class hierarchy for the e-tourism ontology version 8** [On-line]. Available: http://e-tourism.deri.at/ont/etourism%20onto%2008.pdf

Siricharoen, W. V. (2007). E-commerce adaptation using ontologies for e-tourism. In **Proceeding of the 7th International Symposium on Communications and Information Technologies** (pp 1274-1279). Sydney, Australia: IEEE.

Sirigayon, C., Chanlekha, H., and Kawtrakul, A. (2004). Information extraction for agricultural information access. In **Proceeding of the International Joint Conference of the Asian Federation of Information Technology in Agriculture and the World Congress on Computer in Agriculture and Natural Resources** (pp 95-106). Bangkok, Thailand: Hydro and Agro Informatics Institute.

Sirinaovakul, B. (2008). **Artificial intelligence**. Bangkok, Thailand: Top.

Smitinand, T. (2001). **Thai plant names dictionary**. Thailand, The Forest Herbarium Royal Forest Department.

Sornlertlamvanich, V., Potipiti, T., and Charoenporn, T. (2000). Automatic corpus-based Thai word extraction with the c4.5 learning algorithm. In **Proceeding of the 18th Conference on Computational Linguistics 2000** (pp 802-807). Saarbrucken, Germany: Association for Computational Linguistics.

Sukhothai Thammathirat Open University. (2007). **Tourist destination management**. Nonthaburi, Thailand: Department of Management Sciences, Sukhothai Thammathirat Open University.

Sumathi, S. and Surekha, P. (2010). **Computational intelligence paradigms : theory and applications using MATLAB**. Boca Raton: CRC Press.

Swartout, B., Patil, R., Knight, K., and Russ, T. (1997). Toward distributed use of large-scale ontologies. In **Proceeding of the AAAI Spring Symposium on Ontological Engineering** (pp 138-148). Palo Alto, California: The Association for the Advancement of Artificial Intelligence.

Taboada, M. and Grieve, J. (2004). Analyzing appraisal automatically. In **Proceeding of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications** (pp 158-161). Palo Alto, California: The Association for the Advancement of Artificial Intelligence.

Text Processing Technology, National Electronics and Computer Technology Center. (2012). **About LEXiTRON** [On-line]. Available: http://lexitron.nectec.or.th/2009_1/index_en.php?q=common_manager/aboutlex

The Internet Movie Database. (2011). **IMDb charts: IMDb top 250** [On-line]. Available: http://www.imdb.com/chart/top

Trakultaweekoon, K., Porkaew, P., and Supnithi, T. (2007). LEXiTRON vocabulary suggestion system with recommendation and vote mechanism. In **Proceeding of the 7th International Symposium on Natural Language Processing** (pp 43-48). Chonburi, Thailand: Kasetsart University in Cooperation with the Thailand's National Electronics and Computer Technology Center (NECTEC).

TripAdvisor. (2013). **About TripAdvisor® media group** [On-line]. Available: http://www.tripadvisor.com/PressCenter-c6-About_Us.html

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In **Proceeding of the 40th Annual Meeting of the Association for Computational Linguistics** (pp 417-424). Philadelphia, America: Association for Computational Linguistics.

Whitby, A., Josang, A., and Indulska, J. (2004). Filtering out unfair ratings in bayesian reputation systems. In **Proceeding of 3rd International Joint Conference on Autonomous Agents and Multiagent Systems** (pp 1-12). New York, USA: IEEE Computer Society.

Whitman, B. and Ellis, D. (2004). Automatic record reviews. In **Proceeding of the 5th International Society for Music Information Retrieval** (pp 470-477). Barcelona, Spain: Universitat Pompeu Fabra.

Wigder, Z. D. et al. (2010). **Asia Pacific online retail forecast, 2010 to 2015** [On-line]. Available: http://www.forrester.com/rb/Research/asia_pacific_online_retail_forecast%2C_2010_to/q/id/57481/t/2

Willis, D. (1996). **Collins cobuild student's grammar self-study edition with answers**. Great Britain: HarperCollins.

Wilson, W. H. (16 June 2012). **The Natural language processing dictionary, aka NLP dictionary** [On-line]. Available: http://www.cse.unsw.edu.au/~billw/nlpdict.html

World Wide Web Consortium. (2009). **SKOS simple knowledge organization system reference** [On-line]. Available: http://www.w3.org/TR/2009/REC-skos-reference-20090818/

Wuttilerdcharoenwong, S. (2009). **Web based expert information extraction**. Department of Computer Engineering, Faculty of Engineering, Kasetsart University. (Unpublished manuscript).

Yoon, Y. and Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. **Tourism Management**. 26(1): 45-56.

Zhang, K., Narayanan, R., and Choudhary, A. (2010). Voice of the customers: mining online customer reviews for product feature-based ranking. In **Proceeding of the 3rd Conference on Online Social Networks** (pp 1-9). CA, USA: USENIX Association Berkeley.

Zheng, W. and Ye, Q. (2009). Sentiment classification of Chinese traveler reviews by support vector machine algorithm. In **Proceeding of 2009 the 3rd International Symposium on Intelligent Information Technology Application** (pp 335-338). NanChang, China: IEEE.

APPENDIX A

Knowledge Base

1. Tourism Ontology Tourism ontology is an essential component of a review analysis module because it is used for extracting and storing entire knowledge obtained from the module. The tourism ontology composes of classes, properties, and relationships as shown in Table A.1-A.15. Moreover, the properties of each class are divided into two types: object properties and datatype properties.

Table A.1 Class Accommodations

Slot name	Type	Allowed Values/Classes	Description
hasContactData	Object properties	ContactData	This property links an individual representing another individual's contact data.
hasPostalAddress	Object properties	PostalAddress	This property links an individual representing another individual's postal address.
hasRoom	Object properties	Room	This property links one or more individuals representing rooms.

Table A.1 Class Accommodations (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasGPSCoordinates	Object properties	GPSCoordinates	This property links an individual representing another individual's GPS coordinates.
hasOutdoorBathroom	Object properties	Bathroom	This property links an individual representing another individual's bathroom.
hasRating	Object properties	Rating	This property links an individual representing another individual's rating.
hasEvent	Object properties	Event	This property links an individual representing another individual's event.
hasDateTime	Object properties	DateTime	This property links an individual representing a date period.
hasActivity	Object properties	Activity	This property links an individual representing another individual's activity.
hasCleaness	Datatype properties	String	This property indicates whether an individual has cleanliness or not.

Table A.1 Class Accommodations (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasDécor	Datatype properties	String	This property indicates whether an individual has décoration or not.
hasSecurity	Datatype properties	String	This property indicates whether an individual has security or not.
hasValue	Datatype properties	String	This property indicates whether an individual has value or not.
hasWellnessFacilities	Datatype properties	String	This property indicates whether an individual has wellness facilities or not.
hasName	Datatype properties	String	This property indicates the name of an individual.
spokenLanguage	Datatype properties	String	This property indicates which languages are spoken.
hasHandicapAccessibility	Datatype properties	String	This property indicates whether an individual has handicap accessibility or not.
offersLaudryCleaning	Datatype properties	String	This property indicates whether an individual offers laundry service or not.
offersCot	Datatype properties	String	This property indicates whether an individual provides a cot on request or not.

Table A.1 Class Accommodations (Continued)

Slot name	Type	Allowed Values/Classes	Description
offersCrib	Datatype properties	String	This property indicates whether an individual provides a crib on request or not.
offersShuttleService	Datatype properties	String	This property indicates whether an individual offers a shuttle service or not.
hasMoneyExchangeOffice	Datatype properties	String	This property indicates whether an individual has money exchange office or not.
offersChildCare	Datatype properties	String	This property indicates whether an individual offers child care or not.
hasParking	Datatype properties	String	This property indicates whether an individual has parking or not.
hasBeergarden	Datatype properties	String	This property indicates whether an individual has a beergarden or not.
hasElevator	Datatype properties	String	This property indicates whether an individual has an elevator or not.

Table A.1 Class Accommodations (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasPool	Datatype properties	String	This property indicates whether an individual has a pool or not.
hasGarden	Datatype properties	String	This property indicates whether an individual has a garden or not.
hasEnglishNewspaper	Datatype properties	String	This property indicates whether an individual has english newspaper or not.
PetsAllowed	Datatype properties	String	This property indicates whether pets are allowed or not.
hasFrontOffice	Datatype properties	String	This property indicates whether an individual has a front office or not.
hasRestaurant	Datatype properties	String	This property indicates whether an individual has a restaurant or not.
hasCafe	Datatype properties	String	This property indicates whether an individual has a cafe or not.
hasLocalFoodStall	Datatype properties	String	This property indicates whether an individual has a local food stall or not.

Table A.1 Class Accommodations (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasSalon	Datatype properties	String	This property indicates whether an individual has a salon or not.
hasSolarium	Datatype properties	String	This property indicates whether an individual has a solarium or not.
hasBaggageRoom	Datatype properties	String	This property indicates whether an individual has a baggage room or not.
hasSauna	Datatype properties	String	This property indicates whether an individual has a sauna or not.
hasMassageStudio	Datatype properties	String	This property indicates whether an individual has a massage studio or not.
hasLobby	Datatype properties	String	This property indicates whether an individual has a lobby or not.
hasSpa	Datatype properties	String	This property indicates whether an individual has a spa or not.
hasSteamBath	Datatype properties	String	This property indicates whether an individual has a steam bath or not.

Table A.1 Class Accommodations (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasFitnessRoom	Datatype properties	String	This property indicates whether an individual has a fitness room or not.
hasCocktail	Datatype properties	String	This property indicates whether an individual has a cocktail lounge or not.
hasBoard	Datatype properties	FullBoard, HalfBoard, Breakfast, AllInclusive	This property indicates which board is offered by an individual. Allowed values are full board, half board, breakfast and all inclusive.
hasEmployees	Datatype properties	String	This property indicates whether an individual has employee or not.
hasReception	Datatype properties	String	This property indicates whether an individual has reception or not.
hasConcierge	Datatype properties	String	This property indicates whether an individual has concierge or not.
hasBellboy	Datatype properties	String	This property indicates whether an individual has bellboy or not.
hasPhoto	Datatype properties	String	This property indicates an individual's photo description.

Table A.1 Class Accommodations (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasCarpet	Datatype properties	String	This property indicates whether an individual has carpet or not.
hasCurtain	Datatype properties	String	This property indicates whether an individual has curtain or not.
hasFurniture	Datatype properties	String	This property indicates whether an individual has furniture or not.
hasHouseKeeper	Datatype properties	String	This property indicates whether an individual has housekeeper or not.
smokingAllowed	Datatype properties	String	This property indicates whether smoking is allowed or not.
hasDinner	Datatype properties	String	This property indicates whether an individual offers dinner or not.
hasFood	Datatype properties	String	This property indicates whether smoking is allowed or not.

Table A.2 Class Bathroom

Slot name	Type	Allowed Values/Classes	Description
hasRating	Object properties	Rating	This property links an individual representing another individual's rating.
hasName	Datatype properties	String	This property indicates the name of an individual.
hasHairDryer	Datatype properties	String	This property indicates whether an individual has a hair-dryer or not.
hasToilet	Datatype properties	String	This property indicates whether an individual has a toilet or not.
hasShower	Datatype properties	String	This property indicates whether an individual has a shower or not.
hasWaterPressure	Datatype properties	String	This property indicates whether an individual has a water pressure or not.
hasSoapBar	Datatype properties	String	This property indicates whether an individual has a soap bar or not.
hasBathtub	Datatype properties	String	This property indicates whether an individual has a bathtub or not.

Table A.2 Class Bathroom (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasCleaness	Datatype properties	String	This property indicates whether an individual has cleanliness or not.
hasRobes	Datatype properties	String	This property indicates whether an individual has robes or not.
hasTaps	Datatype properties	String	This property indicates whether an individual has taps or not.
hasTowel	Datatype properties	String	This property indicates whether an individual has a towel or not.
hasCurtain	Datatype properties	String	This property indicates whether an individual has curtain or not.
hasDécor	Datatype properties	String	This property indicates whether an individual has décoration or not.

Table A.3 Class Room

Slot name	Type	Allowed Values/Classes	Description
hasIndoorBathroom	Object properties	Bathroom	This property links an individual representing another individual's bathroom.

Table A.3 Class Room (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasRating	Object properties	Rating	This property links an individual representing another individual's rating.
hasRoomType	Datatype properties	String	This property indicates the room type of an individual.
hasName	Datatype properties	String	This property indicates the name of an individual.
hasInternetAccess	Datatype properties	String	This property indicates whether an individual has internet access or not.
hasVCR	Datatype properties	String	This property indicates whether an individual has a VCR (Video Cassette Recorder) or not.
hasHotwater	Datatype properties	String	This property indicates whether an individual has hot water or not.
hasTVSet	Datatype properties	String	This property indicates whether an individual has a TV set or not.
hasTelephone	Datatype properties	String	This property indicates whether an individual has a telephone or not.

Table A.3 Class Room (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasAirCondition	Datatype properties	String	This property indicates whether an individual has air condition or not.
hasFaxMachine	Datatype properties	String	This property indicates whether an individual has a fax machine or not.
hasArea	Datatype properties	String	This property indicates whether an individual has area or not.
hasCleaness	Datatype properties	String	This property indicates whether an individual has cleanliness or not.
hasChair	Datatype properties	String	This property indicates whether an individual has a chair or not.
hasCarpet	Datatype properties	String	This property indicates whether an individual has carpet or not.
hasCurtain	Datatype properties	String	This property indicates whether an individual has curtain or not.
hasDécor	Datatype properties	String	This property indicates whether an individual has décoration or not.

Table A.4 Class Guestroom

Slot name	Type	Allowed Values/Classes	Description
hasRating	Object properties	Rating	This property links an individual representing another individual's rating.
hasBedType	Datatype properties	String	This property indicates the bed type of an individual.
hasIce	Datatype properties	String	This property indicates whether an individual has ice or not.
hasSpringWater	Datatype properties	String	This property indicates whether an individual has a drinking water or not.
hasCableTV	Datatype properties	String	This property indicates whether an individual has cable TV or not.
hasMicrowave	Datatype properties	String	This property indicates whether an individual has a microwave or not.
hasBalcony	Datatype properties	String	This property indicates whether an individual has a balcony or not.
hasSafe	Datatype properties	String	This property indicates whether an individual has a safe or not.

Table A.4 Class Guestroom (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasRefrigerator	Datatype properties	String	This property indicates whether an individual has a refrigerator or not.
hasBed	Datatype properties	String	This property indicates whether an individual has a bed or not.
hasView	Datatype properties	String	This property indicates which view somebody has from an individual.
hasAlarmClock	Datatype properties	String	This property indicates whether an individual has an alarm clock or not.
hasBedsheet	Datatype properties	String	This property indicates whether an individual has a bed sheet or not.
hasMinibar	Datatype properties	String	This property indicates whether an individual has a minibar or not.
hasSlippers	Datatype properties	String	This property indicates whether an individual has slippers or not.
hasTerrace	Datatype properties	String	This property indicates whether an individual has a terrace or not.

Table A.4 Class Guestroom (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasKitchenetteDrawer	Datatype properties	String	This property indicates whether an individual has a kitchenette drawer or not.
hasTeaCoffeeEquipment	Datatype properties	String	This property indicates whether an individual has a tea and coffee equipment or not.
hasBreakfast	Datatype properties	String	This property indicates whether an individual offers breakfast or not.
hasDuvet	Datatype properties	String	This property indicates whether an individual has a duvet or not.
hasHeater	Datatype properties	String	This property indicates whether an individual has a heater or not.
hasPrice	Datatype properties	String	This property indicates whether an individual has price or not.

Table A.5 Class ConferenceRoom

Slot name	Type	Allowed Values/Classes	Description
hasRating	Object properties	Rating	This property links an individual representing another individual's rating.
hasVideoProjector	Datatype properties	String	This property indicates whether an individual has a video projector or not.
hasScreen	Datatype properties	String	This property indicates whether an individual has a screen or not.
hasFlipChart	Datatype properties	String	This property indicates whether an individual has a flip chart or not.
hasLectern	Datatype properties	String	This property indicates whether an individual has a lectern or not.
hasVideoConference System	Datatype properties	String	This property indicates whether an individual has a video conference system or not.
litByNaturalDaylight	Datatype properties	String	This property indicates whether an individual is lit by natural daylight or not.
hasStage	Datatype properties	String	This property indicates whether an individual has a stage or not.

Table A.5 Class ConferenceRoom (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasAudioEquipment	Datatype properties	String	This property indicates whether an individual has audio equipment or not.

Table A.6 Class Activity

Slot name	Type	Allowed Values/Classes	Description
canBeDoneAt	Object properties	Accommodation	This property links an individual representing the accommodation where an activity can be done.
hasRating	Object properties	Rating	This property links an individual representing another individual's rating.
hasPhoto	Datatype properties	String	This property indicates an individual's photo description.
hasName	Datatype properties	String	This property indicates the name of an individual.

Table A.7 Class Amenity

Slot name	Type	Allowed Values/Classes	Description
hasActivity	Object properties	Activity	This property links an individual representing another individual's activity.
hasRating	Object properties	Rating	This property links an individual representing another individual's rating.
hasGPSCoordinates	Object properties	GPSCoordinates	This property links an individual representing another individual's GPS coordinates.
hasPostalAddress	Object properties	PostalAddress	This property links an individual representing another individual's postal address.
hasContactData	Object properties	ContactData	This property links an individual representing another individual's contact data.
hasDateTime	Object properties	DateTime	This property links an individual representing a date period.
hasEvent	Object properties	Event	This property links an individual representing another individual's event.

Table A.7 Class Amenity (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasPhoto	Datatype properties	String	This property indicates an individual's photo description.
hasName	Datatype properties	String	This property indicates the name of an individual.
hasType	Datatype properties	String	This property indicates the type of an individual.

Table A.8 Class Attraction

Slot name	Type	Allowed Values/Classes	Description
hasActivity	Object properties	Activity	This property links an individual representing another individual's activity.
hasRating	Object properties	Rating	This property links an individual representing another individual's rating.
hasGPSCoordinates	Object properties	GPSCoordinates	This property links an individual representing another individual's GPS coordinates.
hasPostalAddress	Object properties	PostalAddress	This property links an individual representing another individual's postal address.

Table A.8 Class Attraction (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasContactData	Object properties	ContactData	This property links an individual representing another individual's contact data.
hasDateTime	Object properties	DateTime	This property links an individual representing a date period.
hasEvent	Object properties	Event	This property links an individual representing another individual's event.
hasTicket	Object properties	Ticket	This property links one or more individuals representing different kinds of tickets.
hasPhoto	Datatype properties	String	This property indicates an individual's photo description.
hasName	Datatype properties	String	This property indicates the name of an individual.
hasType	Datatype properties	String	This property indicates the type of an individual.

Table A.9 Class ContactData

Slot name	Type	Allowed Values/Classes	Description
hasEMail	Datatype properties	String	This property indicates an individual's e-mail address and is part of its contact data.
hasFaxNumber	Datatype properties	String	This property indicates an individual's fax number and is part of its contact data. It includes country and area codes separated by hyphens as well as possibly necessary extensions.
hasTelephoneNumber	Datatype properties	String	This property indicates an individual's telephone number and is part of its contact data. It includes country and area codes separated by hyphens as well as possible extensions.
hasWebsite	Datatype properties	String	This property indicates one or more of an individual's web site addresses and is part of its contact data. The protocol part of the URI must not be part of the value.

Table A.10 Class DateTime

Slot name	Type	Allowed Values/Classes	Description
hasStartDate	Datatype properties	String	This property indicates the date when a date period starts.
hasEndDate	Datatype properties	String	This property indicates the date when a date period ends.
hasStartTime	Datatype properties	String	This property indicates the time when a time period starts. Hours, minutes and seconds are separated by colons. The postfix stands for the hour difference compared to the Coordinated Universal Time.
hasEndTime	Datatype properties	String	This property indicates the time when a time period ends. Hours, minutes and seconds are separated by colons. The postfix stands for the hour difference compared to the Coordinated Universal Time.
hasWeekday	Datatype properties	Monday, Tuesday, Wednesday, Thursday, Friday	This property indicates a weekday.

Table A.11 Class Event

Slot name	Type	Allowed Values/Classes	Description
hasActivity	Object properties	Activity	This property links an individual representing another individual's activity.
hasGPSCoordinates	Object properties	GPSCoordinates	This property links an individual representing another individual's GPS coordinates.
hasPostalAddress	Object properties	PostalAddress	This property links an individual representing another individual's postal address.
hasContactData	Object properties	ContactData	This property links an individual representing another individual's contact data.
hasDateTime	Object properties	DateTime	This property links an individual representing a date period.
hasTicket	Object properties	Ticket	This property links one or more individuals representing different kinds of tickets.
hasPhoto	Datatype properties	String	This property indicates an individual's photo description.
hasName	Datatype properties	String	This property indicates the name of an individual.

Table A.11 Class Event (Continued)

Slot name	Type	Allowed Values/Classes	Description
hasType	Datatype properties	String	This property indicates the type of an individual.

Table A.12 Class GPSCoordinates

Slot name	Type	Allowed Values/Classes	Description
hasLatitude	Datatype properties	float	This property indicates the latitude of an individual in degrees and is part of its GPS coordinates. The WGS 84 (World Geodetic System 1984) is used as global reference frame.
hasLongitude	Datatype properties	float	This property indicates the longitude of an individual in degrees and is part of its GPS coordinates. The WGS 84 (World Geodetic System 1984) is used as global reference frame.

Table A.13 Class PostalAddress

Slot name	Type	Allowed Values/Classes	Description
hasHouseNumber	Datatype properties	String	This property indicates the house number of the building where an individual is located and is part of its postal address.
hasStreet	Datatype properties	String	This property indicates the street where an individual is located and is part of its postal address.
hasTambon	Datatype properties	String	This property indicates the tambon where an individual is located and is part of its postal address.
hasDistrict	Datatype properties	String	This property indicates the district where an individual is located and is part of its postal address.
hasProvince	Datatype properties	String	This property indicates the province where an individual is located and is part of its postal address.

Table A.14 Class Rating

Slot name	Type	Allowed Values/Classes	Description
Excellent	Datatype properties	Integer	This property indicates which rating an accomodation has at the excellent level. This value is represented by an integer.
Good	Datatype properties	Integer	This property indicates which rating an accomodation has at the good level. This value is represented by an integer.
Average	Datatype properties	Integer	This property indicates which rating an accomodation has at the average level. This value is represented by an integer.
Bad	Datatype properties	Integer	This property indicates which rating an accomodation has at the bad level. This value is represented by an integer.
Worse	Datatype properties	Integer	This property indicates which rating an accomodation has at the worse level. This value is represented by an integer.

Table A.15 Class Ticket

Slot name	Type	Allowed Values/Classes	Description
hasName	Datatype properties	String	This property indicates the name of an individual.
hasPrice	Datatype properties	Integer	This property indicates the price of an individual in Bahts.

2. SKOS Ontology SKOS ontology is a part of the tourism ontology. It is used for collecting the synonym words which have the same or very similar meanings as illustrated in Table A.16. Moreover, the SKOS ontology is also used for collecting the word with wrong written in order to correct the critirism words to the suitable word for feature extraction process. For each synonym set, there is a designated word which is the representation of all synonym words.

Table A.16 SKOS ontology

Designated word	Synonyms set
'	’
abit	a bit, a bit of, a modicum, a little
adream	a dream
advertisement	ad
aircondition	ac, a.c., a/c, air con, air conditioning, airconditioning, air conditioned, airconditioned, aircon, air con, air conditioner, air conditioners, airconditioner, air conditions, airconditions
ajoke	a joke
alarmclock	alarm clock, clock, clocks
allinall	all in all
alotof	alot of, a lot of, lots of
am	’m
and	&
are	’re
are not	aint, ain't, ainot
arealgem	a real gem
atfirst	at first

Table A.16 SKOS ontology (Continued)

Designated word	Synonyms set
atleast	at least, the least
audioequipment	audio equipment, stereo
balcony	terrace, balcany, balconies, patios, patio, patio area
bathroom	bath room, bath rooms, shower room, toiletries, bathroom wall, fan in the bathroom, soap or shampoo in room, wash room
bathroom curtain	shower curtain
bathtub	basin, bath tub, bath, baths, sink, private bath, tub, bathe, bath up, washbasin, bathroom tub, jacuzzi tub
bed	mattress, mattresses, matress, bedding, rollaway bed, king bed, double bed, sleeping bed
bedbug	bug, bed bug, insect, bug infested
bedsheet	bed sheet, bedsheets, bed cover, bedcover, bedspread, bed spread, tick, pillow, quilt, covering, bed linen, bedlinen, sheets, comforter
bellboy	bell boy, bellguy, bellhop, bellmen
best	the best
bigmistake	big mistake
breakfast	continental breakfast, free continental breakfast, english breakfast, brkfst, breakfast buffet, buffet breakfast, daily breakfast, buffet style breakfast, breakfast menu, breakfast room, breakfast area, breakfast spread
cabletv	satellite tv, cable tv
cafe	coffee shop
can not	cant, can't, cannot, cannt
carpet	rug, rugs, mat, bathmat, floor mat
chair	desk chair seat
charge	little extras

Table A.16 SKOS ontology (Continued)

Designated word	Synonyms set
cheerful	smile
childcare	child, children, amenities kids
childfriendly	child friendly
cleaness	cleanliness, cleanness
cockroach	roach
concierge	concierge service, doormen, concierge honor
could not	couldnt, couldn't, couldnot
curtain	drape
decor	decorated, decorating, decoration, decorations, decorative, furnish, furnished, physical appearance, outside decor, appointed, designed interiors
definitely	definately
delicious	deliciously, deliciousness
delight	a delight
did not	didnt, didn't, didnot
difficultcommunicate	difficult communicate
dinner	dinner buffet, buffet dinner, inroom dinning, dinning, dinning room, dinning area
do not	dont, don't, donot
does not	doesnt, doesn't, doesnot
duvet	doona, duvets, eiderdown, quilt
e	é
elevator	lift, elevators, lifts
englishnewspaper	english newspaper
especially	esp
fair	fairly

Table A.16 SKOS ontology (Continued)

Designated word	Synonyms set
employee	staff, wait staff, people, manager, assistant hotel manager, hotel manager, staff member, staff members, waiter, weekend manager, hotel staff, resident manager, lounge staff, bar and restaurant staff, hotel general manager, supervisor
faxmachine	fax machine
first	1st
firstofall	first of all
fitnessroom	fitness room, gym, exercise room
fivestar	5 star, 5* quality, 5star, first class
flipchart	flip chart, label
food	meal, food quality, meals pans, food spread, menu, fare
friendly	family friendly
frontoffice	front office, front staff, front, clerk, check in desk, front desk, front desk guy, front desk staff
furniture	fixture, woodwork, room furniture
hairdryer	hair dryer
handicapaccessibility	cripple, handicapped, defective
have	've, having
helpful	helpful answering questions
heater	heating, heatting, fireplace
highstandard	high standard
hotel	the place, this place, that place, hostel accommodation, motel, hotel management, overall experience, the accommodation, accommodations, resort, guesthouse, entire place, home stay, homestay, hotel itself, grounds, hotel experience, place, ambiance of the hotel, property, properties, ambience, ambiance

Table A.16 SKOS ontology (Continued)

Designated word	Synonyms set
hotwater	hot water, water heater
housekeeper	housekeeping service, homemaker, maid, housemaid, maidservant, servant, matron, housekeeping staff, house maid, house maids, house keeping, lady cleaners
inreality	in reality
internetaccess	wireless internet access, wireless internet, internet access, internet access computer, wifi, free wifi, wirelessconnection, wireless lan, access point, free internet, internet
inthemorning	in the morning
is	’s
kitchenettedrawer	kitchenette drawer, kitchen
large spacious	largespacious
laudrycleaning	laudry cleaning, laundry cleaning, laundry
lessthan	less than
litbynaturaldaylight	lit by natural daylight
lobby	lounge, foyer, executive lobby, executive lounge, hall, hallway, hallways, hall way, hall ways, hotel hall way, atrium, parlour, waiting room, waiting area
localfoodstall	food shop, local food stall, local shops
location	located, situated, neighborhood, locality
lowprice	low price
lowstandard	low standard, low class
makeof	make of, made of, make from, made from
massagestudio	massage
microwave	microwave oven, oven
midrange	mid range

Table A.16 SKOS ontology (Continued)

Designated word	Synonyms set
moneyexchangeoffice	money exchange
mosquito	anopheles, armigeres, culex, mansonia, donitz, aedes, aedes egypti, malaria mosquito, common house mosquito, malariacarrying mosquito
most	the most
motelfeel	motel feel
muststay	must stay
near	close to, closer to, next to, within minutes away from, convenient to, close by, close from, easy to get to, easy to get around
neveragain	never again, never come back again, never go back, never ever going to come back, never going back there again, we would not go again
nexttime	next time
noncooperative	non cooperative
not	n't
nothingelsebut	nothing else but
onarrival	on arrival
outdated	out of date, out dated
outoforder	out of order
overprice	overly priced, over priced
paperthin	paper thin
parking	car park, parking place, parking lot
petsallowed	pet, pets allowed, pets allow
price	rate, room rate
publicarea	public area, common area
rat	mouse, mice, rodent, vermin
pool	pond, swimming pool, indoor swimming pool, indoor pool, outdoor pool, pool area, poolside area, water slides, water rides

Table A.16 SKOS ontology (Continued)

Designated word	Synonyms set
photo	picture, image, website photo
readytoserve	ready to serve
reception	reception, reception staff, receptionist, owner, proprietor, host, hostess, customer service, reception desk, business centre staff, reception area
refrigerator	fridge, freezer, cooler, bar fridge, refridge
restaurant	hotel restaurant, eatery, ethnic restaurants, bar, indian restaurants, pubs
remodel	refurbish, redecorate, refurnish, renovate
ripoff	rip off
robes	bathrobe, bath robe, bathing gown, bathing wrap, robe
room	room suite, our room, suite, double room, single room, bedroom, bed room, hotel room, hotel rooms, bedroom suite, balcony room, sitting room, twin room, family room, standard room, sized room, room size, high ceiling, free standing fan
rundown	run down
salon	beauty shop, beauty parlor, hairdressing salon, beauty salon
spa	free spa
secondtonone	second to none
security	secure, secured, safety
service	room service, room services
shower	shower head
slippers	slipper
shuttleservice	shuttlebus, shuttle bus, shuttle service, shuttle service attendant, shuttle, car service, in house car, hotel car
smokingallowed	smoking allowed
soapbar	soap bar
spider	arachnid

Table A.16 SKOS ontology (Continued)

Designated word	Synonyms set
spokenlanguage	spoken language, speaking language, language speaking, spoke language, speaks language, speak english
springwater	spring water
stayagain	stay there again, stay again, stay here again
stayaway	stay away
steambath	steam bath
taps	tap, faucet, sink faucet, shower faucet
telephone	phone
teacoffeeequipment	tea coffee equipment, coffee equipment, teacoffee facilites, tea facilites, coffee facilites, tea making facilities
the walls	the room
theboutiquehotel	the boutique hotel
there is	theres, there are, there was, there were, there's, there're
threestar	3 star, 3star
topnotch	top notch
topoftheline	top of the line
tvset	t.v., television, tv, t v, tv lounge, tvs, color tv, wee old tv, vdo tv, remote for a vdo tv
u	you
ur	your
value	value of money, value for money
vcr	video cassette recorder
videoconferencesystem	video conference system, vdo conference, video conference
videoprojector	video projector, vdo projector, projector
walkingdistance	walking distance
was not	wasnt, wasn't, wasnot

Table A.16 SKOS ontology (Continued)

Designated word	Synonyms set
waterdown	water down, watered down
waterpressure	pressure, water pressure
website	web, the web, hotel website
welldone	well done
wellinformed	well informed
wellknown	well known
wellnessfacilities	recreation facilities, hotel facilities, amenities, hotel amenities, amenity, medical facilities, 4-5 star amenities, facility
wellpresent	well present
were not	werent, weren't, werenot
will not	wont, won't, willnot
wornout	worn out
worst	the worst
would	’d
would not	wouldnt, wouldn't, wouldnot

3. Terminology Terminology is a word collection assigned the satisfaction level in 5-rating scale for each word as shown in Figure A.1. This terminology is stored in the EXtensible Markup Language (XML) format with invented tags. The descriptions of each tag are described in Table A.17

Table A.17 Tag description in terminology

Tag name	Description
<words>	Define the entire words used in terminology.
<basics>	Define the entire adjective words.
<basic>	Define an adjective word with a fixed satisfaction level.
<vocab>	Define an adjective word.
<rate>	Define a fixed satisfaction level of an adjective word.
<advances>	Define the entire adverb words.
<advance>	Define an adverb word with an adjustable rating.
<vocab>	Define an adverb word.
<rateless>	Define an adjustable rating of an adverb word in case of negative criticism.
<ratemoreequal>	Define an adjustable rating of an adverb word in case of positive criticism.
<specials>	Define the entire negation adverb words.
<special>	Define a negation adverb with an adjustable rating.
<vocab>	Define a negation adverb.
<rateless>	Define an adjustable rating of a negation adverb word in case of negative criticism.
<ratemoreequal>	Define an adjustable rating of a negation adverb word in case of positive criticism.
<specialwords>	Define the entire special words.

Table A.17 Tag description in terminology (Continued)

Tag name	Description
<specialword>	Define a special word with a fixed satisfaction level and a feature to which it implies.
<word>	Define a special word.
<ecat>	Define part of speech of a special word.
<rate>	Define a fixed satisfaction level of a special word.
<feature>	Define a feature to which a special word implies.
<verbspecialwords>	Define the entire special verb words.
<verbspecialword>	Define a special verb word with a fixed satisfaction level.
<word>	Define a special verb word.
<rate>	Define a fixed satisfaction level of a special verb word.

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<words>
	<basics>
		<basic>
			<vocab>dark</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>disgusting</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>offputting</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>worse</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>acceptable</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>bearable</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>unacceptable</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>subpar</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>sparse</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>moderate</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>excellent</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>good</vocab>
			<rate>4</rate>
		</basic>
```


Figure A.1 Terminology

```
<basic>
        <vocab>minimal</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>minimum</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>bad</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>smooth</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>friendly</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>fun</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>unfriendly</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>obnoxious</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>helpful</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>weird</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>strange</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>passable</vocab>
        <rate>3</rate>
</basic>
<basic>
        <vocab>adequate</vocab>
        <rate>3</rate>
```


Figure A.1 Terminology (Continued)

```
        </basic>
        <basic>
                <vocab>inadequate</vocab>
                <rate>2</rate>
        </basic>
        <basic>
                <vocab>sympathetic</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>pathetic</vocab>
                <rate>2</rate>
        </basic>
        <basic>
                <vocab>fair</vocab>
                <rate>3</rate>
        </basic>
        <basic>
                <vocab>comfortable</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>fine</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>incredible</vocab>
                <rate>5</rate>
        </basic>
        <basic>
                <vocab>ideal</vocab>
                <rate>5</rate>
        </basic>
        <basic>
                <vocab>servicemind</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>serviceminded</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>have</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>has</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>thereis</vocab>
```


Figure A.1 Terminology (Continued)

```
            <rate>4</rate>
        </basic>
        <basic>
            <vocab>no</vocab>
            <rate>2</rate>
        </basic>
        <basic>
            <vocab>average</vocab>
            <rate>3</rate>
        </basic>
        <basic>
            <vocab>awful</vocab>
            <rate>1</rate>
        </basic>
        <basic>
            <vocab>basic</vocab>
            <rate>3</rate>
        </basic>
        <basic>
            <vocab>best</vocab>
            <rate>5</rate>
        </basic>
        <basic>
            <vocab>better</vocab>
            <rate>5</rate>
        </basic>
        <basic>
            <vocab>comfort</vocab>
            <rate>4</rate>
        </basic>
        <basic>
            <vocab>contented</vocab>
            <rate>3</rate>
        </basic>
        <basic>
            <vocab>cool</vocab>
            <rate>5</rate>
        </basic>
        <basic>
            <vocab>cold</vocab>
            <rate>4</rate>
        </basic>
        <basic>
            <vocab>hot</vocab>
            <rate>2</rate>
        </basic>
        <basic>
            <vocab>dated</vocab>
            <rate>2</rate>
        </basic>
        <basic>
```


Figure A.1 Terminology (Continued)

```
        <vocab>limited</vocab>
        <rate>2</rate>
    </basic>
    <basic>
        <vocab>decent</vocab>
        <rate>4</rate>
    </basic>
    <basic>
        <vocab>difficultcommunicate</vocab>
        <rate>2</rate>
    </basic>
    <basic>
        <vocab>disappoint</vocab>
        <rate>1</rate>
    </basic>
    <basic>
        <vocab>disappointed</vocab>
        <rate>1</rate>
    </basic>
    <basic>
        <vocab>disappointing</vocab>
        <rate>1</rate>
    </basic>
    <basic>
        <vocab>dreadful</vocab>
        <rate>1</rate>
    </basic>
    <basic>
        <vocab>efficient</vocab>
        <rate>5</rate>
    </basic>
    <basic>
        <vocab>fab</vocab>
        <rate>5</rate>
    </basic>
    <basic>
        <vocab>fantastic</vocab>
        <rate>5</rate>
    </basic>
    <basic>
        <vocab>fresh</vocab>
        <rate>4</rate>
    </basic>
    <basic>
        <vocab>great</vocab>
        <rate>5</rate>
    </basic>
    <basic>
        <vocab>grotty</vocab>
        <rate>2</rate>
    </basic>
```


Figure A.1 Terminology (Continued)

```
<basic>
	<vocab>happy</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>cheerful</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>happily</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>hate</vocab>
	<rate>1</rate>
</basic>
<basic>
	<vocab>unpolite</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>impolite</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>polite</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>inconsiderate</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>inconvenience</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>convenience</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>convenient</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>neat</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>suitable</vocab>
	<rate>4</rate>
```


Figure A.1 Terminology (Continued)

```
</basic>
<basic>
        <vocab>infestations</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>infest</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>infested</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>kind</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>love</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>lovely</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>modern</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>update</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>outdated</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>stylish</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>fancy</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>prime</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>narrow</vocab>
```


Figure A.1 Terminology (Continued)

```
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>claustrophobic</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>undersized</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>cramped</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>nice</vocab>
		<rate>4</rate>
	</basic>
	<basic>
		<vocab>ok</vocab>
		<rate>3</rate>
	</basic>
	<basic>
		<vocab>old</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>older</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>disappointment</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>snobby</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>crab</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>pretentious</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>insulting</vocab>
		<rate>2</rate>
	</basic>
	<basic>
```


Figure A.1 Terminology (Continued)

```
			<vocab>pokey</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>perfect</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>pleasant</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>unpleasant</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>fake</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>unpleasing</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>unpleasantly</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>pleased</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>creative</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>quiet</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>feeble</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>prompt</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>tasty</vocab>
			<rate>5</rate>
		</basic>
```


Figure A.1 Terminology (Continued)

```
<basic>
	<vocab>yummy</vocab>
	<rate>5</rate>
</basic>
<basic>
	<vocab>reasonable</vocab>
	<rate>3</rate>
</basic>
<basic>
	<vocab>civil</vocab>
	<rate>3</rate>
</basic>
<basic>
	<vocab>rubbish</vocab>
	<rate>1</rate>
</basic>
<basic>
	<vocab>rude</vocab>
	<rate>1</rate>
</basic>
<basic>
	<vocab>tatty</vocab>
	<rate>1</rate>
</basic>
<basic>
	<vocab>rusty</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>shame</vocab>
	<rate>1</rate>
</basic>
<basic>
	<vocab>shiny</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>smart</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>smelt</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>smooth</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>spectacular</vocab>
	<rate>5</rate>
```

Figure A.1 Terminology (Continued)

```
</basic>
<basic>
        <vocab>sullen</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>surprise</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>terrible</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>unhelpful</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>nonchalant</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>welcoming</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>welcome</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>unwelcoming</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>unwelcome</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>worst</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>frustrate</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>wornout</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>misleading</vocab>
```


Figure A.1 Terminology (Continued)

```
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>well</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>tear</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>wear</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>broken</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>cracked</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>unprofessional</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>repulsive</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>laughable</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>stupid</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>shabby</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>desperate</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>plain</vocab>
			<rate>3</rate>
		</basic>
		<basic>
```


Figure A.1 Terminology (Continued)

```
			<vocab>abusive</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>lousy</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>creepy</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>normal</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>okay</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>stuffy</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>hard</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>soft</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>squeak</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>tiny</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>presentable</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>plentiful</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>ample</vocab>
			<rate>4</rate>
		</basic>
```

Figure A.1 Terminology (Continued)

```
<basic>
        <vocab>fungus</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>varied</vocab>
        <rate>3</rate>
</basic>
<basic>
        <vocab>gorgeous</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>delightful</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>enamoring</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>ornate</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>beautiful</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>scenic</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>fluent</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>superb</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>functional</vocab>
        <rate>3</rate>
</basic>
<basic>
        <vocab>complimentary</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>discreet</vocab>
        <rate>4</rate>
```


Figure A.1 Terminology (Continued)

```
        </basic>
        <basic>
                <vocab>fabulous</vocab>
                <rate>5</rate>
        </basic>
        <basic>
                <vocab>regretful</vocab>
                <rate>2</rate>
        </basic>
        <basic>
                <vocab>regret</vocab>
                <rate>2</rate>
        </basic>
        <basic>
                <vocab>scented</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>calm</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>fashionable</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>chic</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>chicness</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>graceful</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>professional</vocab>
                <rate>5</rate>
        </basic>
        <basic>
                <vocab>fit</vocab>
                <rate>3</rate>
        </basic>
        <basic>
                <vocab>relaxing</vocab>
                <rate>4</rate>
        </basic>
        <basic>
                <vocab>relax</vocab>
```


Figure A.1 Terminology (Continued)

```
            <rate>4</rate>
    </basic>
    <basic>
            <vocab>phenomenal</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>impeccable</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>palatable</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>roomy</vocab>
            <rate>4</rate>
    </basic>
    <basic>
            <vocab>delicious</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>exceptional</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>wonderful</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>delectable</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>enjoyable</vocab>
            <rate>4</rate>
    </basic>
    <basic>
            <vocab>illy</vocab>
            <rate>1</rate>
    </basic>
    <basic>
            <vocab>ill</vocab>
            <rate>1</rate>
    </basic>
    <basic>
            <vocab>bustling</vocab>
            <rate>2</rate>
    </basic>
    <basic>
```


Figure A.1 Terminology (Continued)

```xml
			<vocab>crowded</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>clear</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>inspiring</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>uninspiring</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>proper</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>eager</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>lamentable</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>lamentably</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>personable</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>handy</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>nearest</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>large</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>spacious</vocab>
			<rate>4</rate>
		</basic>
```

Figure A.1 Terminology (Continued)

```
<basic>
        <vocab>big</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>huge</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>massive</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>enormous</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>small</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>wee</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>patronizing</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>narrow</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>lacking</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>daft</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>scary</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>scared</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>vast</vocab>
        <rate>5</rate>
```

Figure A.1 Terminology (Continued)

```
		</basic>
		<basic>
			<vocab>impressive</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>impressed</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>unimpressive</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>outstanding</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>empty</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>sad</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>tepid</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>aged</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>top</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>low</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>ordinary</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>accommodating</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>understanding</vocab>
```


Figure A.1 Terminology (Continued)

```
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>admirable</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>adorable</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>advisable</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>aesthetic</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>afraid</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>fearful</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>aggressive</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>aidful</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>aimful</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>airy</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>allergic</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>annoying</vocab>
			<rate>2</rate>
		</basic>
		<basic>
```


Figure A.1 Terminology (Continued)

```
		<vocab>annoyed</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>anxious</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>awesome</vocab>
		<rate>5</rate>
	</basic>
	<basic>
		<vocab>awkward</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>base</vocab>
		<rate>1</rate>
	</basic>
	<basic>
		<vocab>bland</vocab>
		<rate>2</rate>
	</basic>
	<basic>
		<vocab>brilliant</vocab>
		<rate>4</rate>
	</basic>
	<basic>
		<vocab>careful</vocab>
		<rate>4</rate>
	</basic>
	<basic>
		<vocab>meticulous</vocab>
		<rate>4</rate>
	</basic>
	<basic>
		<vocab>attentive</vocab>
		<rate>4</rate>
	</basic>
	<basic>
		<vocab>alright</vocab>
		<rate>4</rate>
	</basic>
	<basic>
		<vocab>ripoff</vocab>
		<rate>1</rate>
	</basic>
	<basic>
		<vocab>highest</vocab>
		<rate>5</rate>
	</basic>
```


Figure A.1 Terminology (Continued)

```
<basic>
        <vocab>bottom</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>easy</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>affordable</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>effective</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>dire</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>poor</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>creak</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>grumpy</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>slippery</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>simple</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>simply</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>horrible</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>trendy</vocab>
        <rate>4</rate>
```


Figure A.1 Terminology (Continued)

```
		</basic>
		<basic>
			<vocab>sloppy</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>ignorant</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>sick</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>elegant</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>elegance</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>gracious</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>hospitable</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>posh</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>cosy</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>amazing</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>slow</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>fast</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>private</vocab>
```

Figure A.1 Terminology (Continued)

```
            <rate>4</rate>
    </basic>
    <basic>
            <vocab>bountiful</vocab>
            <rate>4</rate>
    </basic>
    <basic>
            <vocab>special</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>much</vocab>
            <rate>4</rate>
    </basic>
    <basic>
            <vocab>extraordinary</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>courteous</vocab>
            <rate>4</rate>
    </basic>
    <basic>
            <vocab>discourteous</vocab>
            <rate>2</rate>
    </basic>
    <basic>
            <vocab>respectful</vocab>
            <rate>4</rate>
    </basic>
    <basic>
            <vocab>disrespectful</vocab>
            <rate>2</rate>
    </basic>
    <basic>
            <vocab>appreciated</vocab>
            <rate>4</rate>
    </basic>
    <basic>
            <vocab>harsh</vocab>
            <rate>2</rate>
    </basic>
    <basic>
            <vocab>ritzy</vocab>
            <rate>5</rate>
    </basic>
    <basic>
            <vocab>luxurious</vocab>
            <rate>5</rate>
    </basic>
    <basic>
```


Figure A.1 Terminology (Continued)

```
            <vocab>luxury</vocab>
            <rate>5</rate>
      </basic>
      <basic>
            <vocab>useful</vocab>
            <rate>4</rate>
      </basic>
      <basic>
            <vocab>comfy</vocab>
            <rate>4</rate>
      </basic>
      <basic>
            <vocab>steep</vocab>
            <rate>2</rate>
      </basic>
      <basic>
            <vocab>terrific</vocab>
            <rate>5</rate>
      </basic>
      <basic>
            <vocab>faultless</vocab>
            <rate>5</rate>
      </basic>
      <basic>
            <vocab>warm</vocab>
            <rate>3</rate>
      </basic>
      <basic>
            <vocab>warmed</vocab>
            <rate>3</rate>
      </basic>
      <basic>
            <vocab>uncomfortable</vocab>
            <rate>2</rate>
      </basic>
      <basic>
            <vocab>stunning</vocab>
            <rate>4</rate>
      </basic>
      <basic>
            <vocab>tired</vocab>
            <rate>2</rate>
      </basic>
      <basic>
            <vocab>arrogant</vocab>
            <rate>2</rate>
      </basic>
      <basic>
            <vocab>uncaredfor</vocab>
            <rate>2</rate>
      </basic>
```


Figure A.1 Terminology (Continued)

```
<basic>
	<vocab>shabby</vocab>
	<rate>1</rate>
</basic>
<basic>
	<vocab>upgrade</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>charming</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>charm</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>topnotch</vocab>
	<rate>5</rate>
</basic>
<basic>
	<vocab>powerful</vocab>
	<rate>5</rate>
</basic>
<basic>
	<vocab>new</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>highstandard</vocab>
	<rate>5</rate>
</basic>
<basic>
	<vocab>lowstandard</vocab>
	<rate>1</rate>
</basic>
<basic>
	<vocab>stale</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>weak</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>yuck</vocab>
	<rate>1</rate>
</basic>
<basic>
	<vocab>nonexistent</vocab>
	<rate>1</rate>
```


Figure A.1 Terminology (Continued)

```
		</basic>
		<basic>
			<vocab>delight</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>midrange</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>modernised</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>modernise</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>modernized</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>ajoke</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>adream</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>ridiculous</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>unethical</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>sufficient</vocab>
			<rate>3</rate>
		</basic>
		<basic>
			<vocab>useless</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>unorganized</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>exorbitant</vocab>
```


Figure A.1 Terminology (Continued)

```
        <rate>1</rate>
</basic>
<basic>
        <vocab>few</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>ghastly</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>confused</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>satisfied</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>repetitive</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>active</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>wornout</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>witty</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>classic</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>timeless</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>furnished</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>immaculate</vocab>
        <rate>5</rate>
</basic>
<basic>
```

Figure A.1 Terminology (Continued)

```xml
			<vocab>warmest</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>friendliest</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>tranquil</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>exotic</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>remarkable</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>soaked</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>boring</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>dreary</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>grand</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>unstable</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>wellpresent</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>dripped</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>rusted</vocab>
			<rate>2</rate>
		</basic>
```


Figure A.1 Terminology (Continued)

```
<basic>
	<vocab>chipped</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>lumpy</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>fuzzy</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>modest</vocab>
	<rate>3</rate>
</basic>
<basic>
	<vocab>unkempt</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>intermittent</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>torn</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>corroded</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>usual</vocab>
	<rate>3</rate>
</basic>
<basic>
	<vocab>neglect</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>impatient</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>mess</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>appalling</vocab>
	<rate>2</rate>
```


Figure A.1 Terminology (Continued)

```
		</basic>
		<basic>
			<vocab>reluctant</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>unwilling</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>characterless</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>knowledgeable</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>quick</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>bizarre</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>problematic</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>flawless</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>dismal</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>superior</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>mismatch</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>exposed</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>hideous</vocab>
```


Figure A.1 Terminology (Continued)

```
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>horrid</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>unspeakable</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>ancient</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>surly</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>musty</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>magnificent</vocab>
			<rate>5</rate>
		</basic>
		<basic>
			<vocab>deficient</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>peeling</vocab>
			<rate>2</rate>
		</basic>
		<basic>
			<vocab>reassuring</vocab>
			<rate>4</rate>
		</basic>
		<basic>
			<vocab>shoddy</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>shoddiest</vocab>
			<rate>1</rate>
		</basic>
		<basic>
			<vocab>suck</vocab>
			<rate>1</rate>
		</basic>
		<basic>
```


Figure A.1 Terminology (Continued)

```
        <vocab>depressing</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>largespacious</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>strong</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>flabbergasted</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>indifferent</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>wellknown</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>readytoserve</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>secondtonone</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>topnotch</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>topoftheline</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>muststay</vocab>
        <rate>5</rate>
</basic>
<basic>
        <vocab>theboutiquehotel</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>fivestar</vocab>
        <rate>5</rate>
</basic>
```


Figure A.1 Terminology (Continued)

```
<basic>
	<vocab>threestar</vocab>
	<rate>3</rate>
</basic>
<basic>
	<vocab>arealgem</vocab>
	<rate>3</rate>
</basic>
<basic>
	<vocab>panoramic</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>paperthin</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>ugly</vocab>
	<rate>1</rate>
</basic>
<basic>
	<vocab>uninterested</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>congested</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>incompetent</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>degraded</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>thorn</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>unsanitary</vocab>
	<rate>2</rate>
</basic>
<basic>
	<vocab>sanitary</vocab>
	<rate>4</rate>
</basic>
<basic>
	<vocab>unhygenic</vocab>
	<rate>2</rate>
```


Figure A.1 Terminology (Continued)

```
</basic>
<basic>
        <vocab>hygenic</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>unhygienic</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>hygienic</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>illmannered</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>immodest</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>inefficiency</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>noncooperative</vocab>
        <rate>1</rate>
</basic>
<basic>
        <vocab>substandard</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>fascinating</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>seedy</vocab>
        <rate>2</rate>
</basic>
<basic>
        <vocab>wellinformed</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>accessible</vocab>
        <rate>4</rate>
</basic>
<basic>
        <vocab>belowpar</vocab>
```


Figure A.1 Terminology (Continued)

```
            <rate>2</rate>
        </basic>
        <basic>
            <vocab>motelfeel</vocab>
            <rate>2</rate>
        </basic>
        <basic>
            <vocab>bigmistake</vocab>
            <rate>1</rate>
        </basic>
    </basics>
    <advances>
        <advance>
            <vocab>very</vocab>
            <rateless>-1</rateless>
            <ratemoreequal>+1</ratemoreequal>
        </advance>
        <advance>
            <vocab>ultra</vocab>
            <rateless>-1</rateless>
            <ratemoreequal>+1</ratemoreequal>
        </advance>
        <advance>
            <vocab>pretty</vocab>
            <rateless>-1</rateless>
            <ratemoreequal>+1</ratemoreequal>
        </advance>
        <advance>
            <vocab>incredibly</vocab>
            <rateless>-1</rateless>
            <ratemoreequal>+1</ratemoreequal>
        </advance>
        <advance>
            <vocab>super</vocab>
            <rateless>-1</rateless>
            <ratemoreequal>+1</ratemoreequal>
        </advance>
        <advance>
            <vocab>so</vocab>
            <rateless>-1</rateless>
            <ratemoreequal>+1</ratemoreequal>
        </advance>
        <advance>
            <vocab>abundantly</vocab>
            <rateless>-1</rateless>
            <ratemoreequal>+1</ratemoreequal>
        </advance>
        <advance>
            <vocab>really</vocab>
            <rateless>-1</rateless>
            <ratemoreequal>+1</ratemoreequal>
```


Figure A.1 Terminology (Continued)

```
</advance>
<advance>
	<vocab>highly</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>strongly</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>totally</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>extremely</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>genuinely</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>especially</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>particularly</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>more</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>most</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>definitely</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
```


Figure A.1 Terminology (Continued)

```
<advance>
	<vocab>obviously</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>absolutely</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>certainly</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>perfectly</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>immaculately</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>deep</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>terribly</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>beyond</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>superficially</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
	<vocab>beautifully</vocab>
	<rateless>-1</rateless>
	<ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
```


Figure A.1 Terminology (Continued)

```
        <vocab>wonderfully</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>seriously</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>well</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>amazingly</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>badly</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>many</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>alotof</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>spotlessly</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>truly</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>nothingelsebut</vocab>
        <rateless>-1</rateless>
        <ratemoreequal>+1</ratemoreequal>
</advance>
<advance>
        <vocab>nicely</vocab>
```

Figure A.1 Terminology (Continued)

```
                <rateless>-1</rateless>
                <ratemoreequal>+1</ratemoreequal>
            </advance>
            <advance>
                <vocab>properly</vocab>
                <rateless>-1</rateless>
                <ratemoreequal>+1</ratemoreequal>
            </advance>
        </advances>
        <specials>
            <special>
                <vocab>too</vocab>
                <rateless>-1</rateless>
                <ratemoreequal>-1</ratemoreequal>
            </special>
            <special>
                <vocab>below</vocab>
                <rateless>-1</rateless>
                <ratemoreequal>-1</ratemoreequal>
            </special>
            <special>
                <vocab>lessthan</vocab>
                <rateless>-1</rateless>
                <ratemoreequal>-1</ratemoreequal>
            </special>
            <special>
                <vocab>ridiculously</vocab>
                <rateless>-1</rateless>
                <ratemoreequal>-1</ratemoreequal>
            </special>
            <special>
                <vocab>disgustingly</vocab>
                <rateless>-1</rateless>
                <ratemoreequal>-1</ratemoreequal>
            </special>
            <special>
                <vocab>nothing</vocab>
                <rateless>+2</rateless>
                <ratemoreequal>-2</ratemoreequal>
            </special>
            <special>
                <vocab>never</vocab>
                <rateless>+2</rateless>
                <ratemoreequal>-2</ratemoreequal>
            </special>
            <special>
                <vocab>not</vocab>
                <rateless>+2</rateless>
                <ratemoreequal>-2</ratemoreequal>
            </special>
            <special>
```

Figure A.1 Terminology (Continued)

```xml
        <vocab>hardly</vocab>
        <rateless>+2</rateless>
        <ratemoreequal>-2</ratemoreequal>
</special>
<special>
        <vocab>without</vocab>
        <rateless>+2</rateless>
        <ratemoreequal>-2</ratemoreequal>
</special>
<special>
        <vocab>no</vocab>
        <rateless>+2</rateless>
        <ratemoreequal>-2</ratemoreequal>
</special>
<special>
        <vocab>barely</vocab>
        <rateless>+1</rateless>
        <ratemoreequal>-1</ratemoreequal>
</special>
<special>
        <vocab>relatively</vocab>
        <rateless>+0.5</rateless>
        <ratemoreequal>-0.5</ratemoreequal>
</special>
<special>
        <vocab>almost</vocab>
        <rateless>+0.5</rateless>
        <ratemoreequal>-0.5</ratemoreequal>
</special>
<special>
        <vocab>quite</vocab>
        <rateless>+0.5</rateless>
        <ratemoreequal>-0.5</ratemoreequal>
</special>
<special>
        <vocab>fairly</vocab>
        <rateless>+0.5</rateless>
        <ratemoreequal>-0.5</ratemoreequal>
</special>
<special>
        <vocab>somewhat</vocab>
        <rateless>+0.5</rateless>
        <ratemoreequal>-0.5</ratemoreequal>
</special>
<special>
        <vocab>rather</vocab>
        <rateless>+0.5</rateless>
        <ratemoreequal>-0.5</ratemoreequal>
</special>
<special>
        <vocab>abit</vocab>
```


Figure A.1 Terminology (Continued)

```xml
                <rateless>+1</rateless>
                <ratemoreequal>-1</ratemoreequal>
            </special>
            <special>
                <vocab>little</vocab>
                <rateless>+0.5</rateless>
                <ratemoreequal>-0.5</ratemoreequal>
            </special>
            <special>
                <vocab>slightly</vocab>
                <rateless>+1</rateless>
                <ratemoreequal>-1</ratemoreequal>
            </special>
            <special>
                <vocab>kinda</vocab>
                <rateless>+0.5</rateless>
                <ratemoreequal>-0.5</ratemoreequal>
            </special>
            <special>
                <vocab>slowly</vocab>
                <rateless>-2</rateless>
                <ratemoreequal>-2</ratemoreequal>
            </special>
        </specials>
        <specialwords>
            <specialword>
                <word>smell</word>
                <ecat>VB,V,ADJ,N</ecat>
                <rate>2</rate>
                <feature>cleaness</feature>
            </specialword>
            <specialword>
                <word>odor</word>
                <ecat>N</ecat>
                <rate>2</rate>
                <feature>cleaness</feature>
            </specialword>
            <specialword>
                <word>stink</word>
                <ecat>V</ecat>
                <rate>1</rate>
                <feature>cleaness</feature>
            </specialword>
            <specialword>
                <word>stunk</word>
                <ecat>V</ecat>
                <rate>1</rate>
                <feature>cleaness</feature>
            </specialword>
            <specialword>
                <word>stinky</word>
```

Figure A.1 Terminology (Continued)

```
            <ecat>ADJ</ecat>
            <rate>1</rate>
            <feature>cleaness</feature>
    </specialword>
    <specialword>
            <word>stinking</word>
            <ecat>ADJ</ecat>
            <rate>1</rate>
            <feature>cleaness</feature>
    </specialword>
    <specialword>
            <word>dirty</word>
            <ecat>ADJ</ecat>
            <rate>2</rate>
            <feature>cleaness</feature>
    </specialword>
    <specialword>
            <word>clean</word>
            <ecat>ADJ</ecat>
            <rate>4</rate>
            <feature>cleaness</feature>
    </specialword>
    <specialword>
            <word>unclean</word>
            <ecat>ADJ</ecat>
            <rate>2</rate>
            <feature>cleaness</feature>
    </specialword>
    <specialword>
            <word>filthy</word>
            <ecat>ADJ</ecat>
            <rate>1</rate>
            <feature>cleaness</feature>
    </specialword>
    <specialword>
            <word>foul</word>
            <ecat>ADJ</ecat>
            <rate>1</rate>
            <feature>cleaness</feature>
    </specialword>
    <specialword>
            <word>soiled</word>
            <ecat>ADJ</ecat>
            <rate>2</rate>
            <feature>cleaness</feature>
    </specialword>
    <specialword>
            <word>stain</word>
            <ecat>N,V,ADJ</ecat>
            <rate>2</rate>
            <feature>cleaness</feature>
```


Figure A.1 Terminology (Continued)

```
</specialword>
<specialword>
        <word>grubby</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>messy</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>untidy</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>tidy</word>
        <ecat>ADJ</ecat>
        <rate>4</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>unsanitary</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>dusty</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>nasty</word>
        <ecat>ADJ</ecat>
        <rate>1</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>dingy</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>mildew</word>
```


Figure A.1 Terminology (Continued)

```
                <ecat>N</ecat>
                <rate>2</rate>
                <feature>cleaness</feature>
        </specialword>
        <specialword>
                <word>mildewy</word>
                <ecat>ADJ</ecat>
                <rate>2</rate>
                <feature>cleaness</feature>
        </specialword>
        <specialword>
                <word>mold</word>
                <ecat>N</ecat>
                <rate>2</rate>
                <feature>cleaness</feature>
        </specialword>
        <specialword>
                <word>moldy</word>
                <ecat>ADJ</ecat>
                <rate>2</rate>
                <feature>cleaness</feature>
        </specialword>
        <specialword>
                <word>mouldy</word>
                <ecat>ADJ</ecat>
                <rate>2</rate>
                <feature>cleaness</feature>
        </specialword>
        <specialword>
                <word>mould</word>
                <ecat>N,V</ecat>
                <rate>2</rate>
                <feature>cleaness</feature>
        </specialword>
        <specialword>
                <word>ant</word>
                <ecat>N</ecat>
                <rate>2</rate>
                <feature>cleaness</feature>
        </specialword>
        <specialword>
                <word>rat</word>
                <ecat>N</ecat>
                <rate>1</rate>
                <feature>cleaness</feature>
        </specialword>
        <specialword>
                <word>cockroach</word>
                <ecat>N</ecat>
                <rate>1</rate>
                <feature>cleaness</feature>
```


Figure A.1 Terminology (Continued)

```
</specialword>
<specialword>
        <word>spider</word>
        <ecat>N</ecat>
        <rate>2</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>bedbug</word>
        <ecat>N</ecat>
        <rate>1</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>mosquito</word>
        <ecat>N</ecat>
        <rate>2</rate>
        <feature>cleaness</feature>
</specialword>
<specialword>
        <word>charge</word>
        <ecat>N,V</ecat>
        <rate>2</rate>
        <feature>price</feature>
</specialword>
<specialword>
        <word>overprice</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>price</feature>
</specialword>
<specialword>
        <word>expensive</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>price</feature>
</specialword>
<specialword>
        <word>pricey</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>price</feature>
</specialword>
<specialword>
        <word>cheap</word>
        <ecat>ADJ</ecat>
        <rate>4</rate>
        <feature>price</feature>
</specialword>
<specialword>
        <word>budget</word>
```


Figure A.1 Terminology (Continued)

```
		<ecat>N,ADJ</ecat>
		<rate>4</rate>
		<feature>price</feature>
	</specialword>
	<specialword>
		<word>inexpensive</word>
		<ecat>ADJ</ecat>
		<rate>4</rate>
		<feature>price</feature>
	</specialword>
	<specialword>
		<word>lowprice</word>
		<ecat>N</ecat>
		<rate>4</rate>
		<feature>price</feature>
	</specialword>
	<specialword>
		<word>costly</word>
		<ecat>ADJ</ecat>
		<rate>2</rate>
		<feature>price</feature>
	</specialword>
	<specialword>
		<word>far</word>
		<ecat>ADJ</ecat>
		<rate>2</rate>
		<feature>location</feature>
	</specialword>
	<specialword>
		<word>near</word>
		<ecat>ADJ</ecat>
		<rate>4</rate>
		<feature>location</feature>
	</specialword>
	<specialword>
		<word>walkingdistance</word>
		<ecat>N</ecat>
		<rate>4</rate>
		<feature>location</feature>
	</specialword>
	<specialword>
		<word>central</word>
		<ecat>ADJ</ecat>
		<rate>4</rate>
		<feature>location</feature>
	</specialword>
	<specialword>
		<word>worth</word>
		<ecat>N,ADJ</ecat>
		<rate>5</rate>
		<feature>value</feature>
```


Figure A.1 Terminology (Continued)

```
</specialword>
<specialword>
        <word>insecure</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>security</feature>
</specialword>
<specialword>
        <word>secure</word>
        <ecat>ADJ</ecat>
        <rate>4</rate>
        <feature>security</feature>
</specialword>
<specialword>
        <word>safe</word>
        <ecat>N,ADJ</ecat>
        <rate>4</rate>
        <feature>security</feature>
</specialword>
<specialword>
        <word>dump</word>
        <ecat>N</ecat>
        <rate>1</rate>
        <feature>hotel</feature>
</specialword>
<specialword>
        <word>loud</word>
        <ecat>ADJ</ecat>
        <rate>2</rate>
        <feature>hotel</feature>
</specialword>
<specialword>
        <word>noisy</word>
        <ecat>ADJ</ecat>
        <rate>1</rate>
        <feature>hotel</feature>
</specialword>
<specialword>
        <word>noise</word>
        <ecat>N</ecat>
        <rate>1</rate>
        <feature>hotel</feature>
</specialword>
<specialword>
        <word>rundown</word>
        <ecat>N,V,ADJ</ecat>
        <rate>1</rate>
        <feature>hotel</feature>
</specialword>
<specialword>
        <word>damp</word>
```


Figure A.1 Terminology (Continued)

```
            <ecat>N,V,ADJ</ecat>
            <rate>2</rate>
            <feature>hotel</feature>
        </specialword>
        <specialword>
            <word>recommend</word>
            <ecat>V</ecat>
            <rate>4</rate>
            <feature>hotel</feature>
        </specialword>
        <specialword>
            <word>neveragain</word>
            <ecat>idm,N</ecat>
            <rate>1</rate>
            <feature>hotel</feature>
        </specialword>
        <specialword>
            <word>stayaway</word>
            <ecat>idm,V</ecat>
            <rate>1</rate>
            <feature>hotel</feature>
        </specialword>
        <specialword>
            <word>stayagain</word>
            <ecat>idm,V</ecat>
            <rate>4</rate>
            <feature>hotel</feature>
        </specialword>
        <specialword>
            <word>childfriendly</word>
            <ecat>ADJ</ecat>
            <rate>4</rate>
            <feature>offersChildCare</feature>
        </specialword>
        <specialword>
            <word>remodel</word>
            <ecat>V</ecat>
            <rate>2</rate>
            <feature>decor</feature>
        </specialword>
        <specialword>
            <word>lostreservations</word>
            <ecat>NP</ecat>
            <rate>1</rate>
            <feature>service</feature>
        </specialword>
    </specialwords>
    <verbspecialwords>
        <verbspecialword>
            <word>collapsed</word>
            <rate>2</rate>
```

Figure A.1 Terminology (Continued)

```
            </verbspecialword>
            <verbspecialword>
                <word>work</word>
                <rate>4</rate>
            </verbspecialword>
            <verbspecialword>
                <word>worked</word>
                <rate>4</rate>
            </verbspecialword>
            <verbspecialword>
                <word>working</word>
                <rate>4</rate>
            </verbspecialword>
            <verbspecialword>
                <word>deteriorated</word>
                <rate>2</rate>
            </verbspecialword>
            <verbspecialword>
                <word>enjoy</word>
                <rate>4</rate>
            </verbspecialword>
            <verbspecialword>
                <word>leak</word>
                <rate>2</rate>
            </verbspecialword>
            <verbspecialword>
                <word>lack</word>
                <rate>2</rate>
            </verbspecialword>
            <verbspecialword>
                <word>decor</word>
                <rate>4</rate>
            </verbspecialword>
            <verbspecialword>
                <word>maintained</word>
                <rate>4</rate>
            </verbspecialword>
            <verbspecialword>
                <word>downgraded</word>
                <rate>2</rate>
            </verbspecialword>
        </verbspecialwords>
    </words>
```


Figure A.1 Terminology (Continued)

APPENDIX B

Testcase Examples

In order to evaluate the review analysis module, the new dataset of reviews which covered all 5 satisfaction levels as shown in Figure B.1-B.5 are required. Kindly note that each review contains whether the criticisms mention to the accommodations.

Worst Hotel Ever!

Traveler rating: 1 2 3 4 5 1.0

September 1, 2009

I would not even classify this hotel as suitable hostel accommodation. Room was filthy. Didn’t feel comfortable there at all. Stayed there due to the location and rate. Was booked last minute and didn’t get a chance to check out the reviews. Assumed it couldn’t be that bad considering it is in a nice area of London. The bed sheets were stained. the chair in the room was covered in mould and the doona (duvet) was soiled. I dont know the last time it was washed. Didn’t even bother with breakfast there. Was tempted to not stay there at all upon checking in, luckily it was only for 1 night. This place needs to clean up its act if they expect to get any more business.

Figure B.1 An example review in the satisfaction level 1

Not top of the list

Traveler rating: 1 2 3 4 5 2.0

September 11, 2007

We stayed at this hotel following the reviews on Tripadvisor. The location is really good, as its close to Hollywood and all trips, however the hotel itself is very basic. It stands beside the very impressive Magic Castle club, but you can’t go there unless you wish to reserve a table for dinner. The rooms are very small, and just about clean. The bathroom was not very pleasant and there was just about enough room to get in and out of the room, even though it was a larger than normal.

Overall, we would not recommend staying here as its quite below par and the reviews were not really accurate.

Figure B.2 An example review in the satisfaction level 2

All the hype is just WRONG

Traveler rating: 1 2 3 4 5 3.0

July 4, 2012

Ok, so the doormen are pleasant when you arrive. The Valet service IS prompt. But for the same price there are far more elegant and gracious hotels in Los Angeles. While the public rooms are pleasant, they are nothing special. The rooms (we stayed in a Deluxe Exec Suite at $785 per night) are dated and cramped. At these rates, housekeeping should be able to clean in a timely fashion. After a hectic day, we came home at 4:45 PM hoping to take a shower before dinner - only to discover the maid in our room, who had not even taken out the trash! Next time, we'll go back to the Peninsula. Or Beverly Wilshire. For sure.

Figure B.3 An example review in the satisfaction level 3

Good hotel experience

Traveler rating: 1 2 3 4 5 4.0

November 27, 2011

We stayed for 7 nights at Lavender Hotels, its the second hotels I ever tried in Bali. We don't think pure luxury Although there is lovely rooms, good free spa offer for our long stayed and great breakfast, its lovely place and reasonable price. Really exceptional and of real value is the service and kindness of the staff, it’s lovely place and reasonable price. Actually this is not a hotel, but your second home on Bali! We for sure would recommend Lavender Hotel to our friends.

Figure B.4 An example review in the satisfaction level 4

Beautiful quiet surrounds.

Traveler rating: 1 2 3 4 5 5.0

September 5, 2012

I had reservations after booking my stay here because of the mention of ants in previous reviews. However I must have been lucky because I only saw about 2 in the room the whole four days we stayed here. We had a family room and it was very spacious. The bed was a little hard but bearable. The breakfast was basic but good. The staff were very nice.

For the price this is a great place to stay.

Figure B.5 An example review in the satisfaction level 5

APPENDIX C

Fuzzy Rules

The fuzzy inference rules are necessary for the accommodation rating process because it is used for calculating the tourist satisfaction scores in fuzzy inference method. This research defined 70 rules divided into 4 groups according to the properties of ontology (or features of accommodation): 20 bathroom rules, 20 room rules, 15 service rules, and 15 overall of accommodation rules as shown in Table C.1.

Table C.1 The fuzzy inference rules including weights

Rules	If	Then	Weight
1	(AlarmClock = Excellent) OR (InternetAccess = Excellent) OR (Microwave = Excellent) OR (RoomCarpet = Excellent) OR (TeaCoffeeEquipment = Excellent)	Room = Excellent	1
2	(AlarmClock = Good) OR (InternetAccess = Good) OR (Microwave = Good) OR (RoomCarpet = Good) OR (TeaCoffeeEquipment = Good)	Room = Good	1
3	(AlarmClock = Average) OR (InternetAccess = Average) OR (Microwave = Average) OR (RoomCarpet = Average) OR (TeaCoffeeEquipment = Average)	Room = Average	1
4	(AlarmClock = Poor) OR (InternetAccess = Poor) OR (Microwave = Poor) OR (RoomCarpet = Poor) OR (TeaCoffeeEquipment = Poor)	Room = Poor	1
5	(AlarmClock = Terrible) OR (InternetAccess = Terrible) OR (Microwave = Terrible) OR (RoomCarpet = Terrible) OR (TeaCoffeeEquipment = Terrible)	Room = Terrible	1

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
6	(AirCondition = Excellent) OR (Area = Excellent) OR (BedType = Excellent) OR (CableTV = Excellent) OR (FaxMachine = Excellent) OR (Hotwater = Excellent) OR (Ice = Excellent) OR (KitchenetteDrawer = Excellent) OR (Minibar = Excellent) OR (Refrigerator = Excellent) OR (Safe = Excellent) OR (smokingAllowed = Excellent) OR (SpringWater = Excellent) OR (Telephone = Excellent) OR (Terrace = Excellent) OR (VCR = Excellent) OR (RoomCleanness = Excellent) OR (RoomCurtain = Excellent) OR (RoomDecoration = Excellent) OR (OverallRoom = Excellent)	Room = Excellent	0.5
7	(AirCondition = Good) OR (Area = Good) OR (BedType = Good) OR (CableTV = Good) OR (FaxMachine = Good) OR (Hotwater = Good) OR (Ice = Good) OR (KitchenetteDrawer = Good) OR (Minibar = Good) OR (Refrigerator = Good) OR (Safe = Good) OR (smokingAllowed = Good) OR (SpringWater = Good) OR (Telephone = Good) OR (Terrace = Good) OR (VCR = Good) OR (RoomCleanness = Good) OR (RoomCurtain = Good) OR (RoomDecoration = Good) OR (OverallRoom = Good)	Room = Good	0.5

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
8	(AirCondition = Average) OR (Area = Average) OR (BedType = Average) OR (CableTV = Average) OR (FaxMachine = Average) OR (Hotwater = Average) OR (Ice = Average) OR (KitchenetteDrawer = Average) OR (Minibar = Average) OR (Refrigerator = Average) OR (Safe = Average) OR (smokingAllowed = Average) OR (SpringWater = Average) OR (Telephone = Average) OR (Terrace = Average) OR (VCR = Average) OR (RoomCleanness = Average) OR (RoomCurtain = Average) OR (RoomDecoration = Average) OR (OverallRoom = Average)	Room = Average	0.5
9	(AirCondition = Poor) OR (Area = Poor) OR (BedType = Poor) OR (CableTV = Poor) OR (FaxMachine = Poor) OR (Hotwater = Poor) OR (Ice = Poor) OR (KitchenetteDrawer = Poor) OR (Minibar = Poor) OR (Refrigerator = Poor) OR (Safe = Poor) OR (smokingAllowed = Poor) OR (SpringWater = Poor) OR (Telephone = Poor) OR (Terrace = Poor) OR (VCR = Poor) OR (RoomCleanness = Poor) OR (RoomCurtain = Poor) OR (RoomDecoration = Poor) OR (OverallRoom = Poor)	Room = Poor	0.5

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
10	(AirCondition = Terrible) OR (Area = Terrible) OR (BedType = Terrible) OR (CableTV = Terrible) OR (FaxMachine = Terrible) OR (Hotwater = Terrible) OR (Ice = Terrible) OR (KitchenetteDrawer = Terrible) OR (Minibar = Terrible) OR (Refrigerator = Terrible) OR (Safe = Terrible) OR (smokingAllowed = Terrible) OR (SpringWater = Terrible) OR (Telephone = Terrible) OR (Terrace = Terrible) OR (VCR = Terrible) OR (RoomCleanness = Terrible) OR (RoomCurtain = Terrible) OR (RoomDecoration = Terrible) OR (OverallRoom = Terrible)	Room = Terrible	0.5
11	(Bed = Excellent) OR (Bedsheet = Excellent) OR (TVSet = Excellent) OR (View = Excellent) OR (Price = Excellent)	Room = Excellent	0.1
12	(Bed = Good) OR (Bedsheet = Good) OR (TVSet = Good) OR (View = Good) OR (Price = Good)	Room = Good	0.1
13	(Bed = Average) OR (Bedsheet = Average) OR (TVSet = Average) OR (View = Average) OR (Price = Average)	Room = Average	0.1
14	(Bed = Poor) OR (Bedsheet = Poor) OR (TVSet = Poor) OR (View = Poor) OR (Price = Poor)	Room = Poor	0.1
15	(Bed = Terrible) OR (Bedsheet = Terrible) OR (TVSet = Terrible) OR (View = Terrible) OR (Price = Terrible)	Room = Terrible	0.1
16	(Balcony = Excellent) OR (Chair = Excellent) OR (Duvet = Excellent) OR (Heater = Excellent) OR (Slippers = Excellent)	Room = Excellent	0

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
17	(Balcony = Good) OR (Chair = Good) OR (Duvet = Good) OR (Heater = Good) OR (Slippers = Good)	Room = Good	0
18	(Balcony = Average) OR (Chair = Average) OR (Duvet = Average) OR (Heater = Average) OR (Slippers = Average)	Room = Average	0
19	(Balcony = Poor) OR (Chair = Poor) OR (Duvet = Poor) OR (Heater = Poor) OR (Slippers = Poor)	Room = Poor	0
20	(Balcony = Terrible) OR (Chair = Terrible) OR (Duvet = Terrible) OR (Heater = Terrible) OR (Slippers = Terrible)	Room = Terrible	0
21	(Concierge = Excellent) OR (FrontOffice = Excellent) OR (SpokenLanguage = Excellent)	Service = Excellent	1
22	(Concierge = Good) OR (FrontOffice = Good) OR (SpokenLanguage = Good)	Service = Good	1
23	(Concierge = Average) OR (FrontOffice = Average) OR (SpokenLanguage = Average)	Service = Average	1
24	(Concierge = Poor) OR (FrontOffice = Poor) OR (SpokenLanguage = Poor)	Service = Poor	1
25	(Concierge = Terrible) OR (FrontOffice = Terrible) OR (SpokenLanguage = Terrible)	Service = Terrible	1
26	(Bellboy = Excellent) OR (Employee = Excellent) OR (Reception = Excellent) OR (OverallService = Excellent)	Service = Excellent	0.5
27	(Bellboy = Good) OR (Employee = Good) OR (Reception = Good) OR (OverallService = Good)	Service = Good	0.5
28	(Bellboy = Average) OR (Employee = Average) OR (Reception = Average) OR (OverallService = Average)	Service = Average	0.5
29	(Bellboy = Poor) OR (Employee = Poor) OR (Reception = Poor) OR (OverallService = Poor)	Service = Poor	0.5

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
30	(Bellboy = Terrible) OR (Employee = Terrible) OR (Reception = Terrible) OR (OverallService = Terrible)	Service = Terrible	0.5
31	(HouseKeeper = Excellent)	Service = Excellent	0.1
32	(HouseKeeper = Good)	Service = Good	0.1
33	(HouseKeeper = Average)	Service = Average	0.1
34	(HouseKeeper = Poor)	Service = Poor	0.1
35	(HouseKeeper = Terrible)	Service = Terrible	0.1
36	(BathroomCurtain = Excellent) OR (Towel = Excellent)	Bathroom = Excellent	1
37	(BathroomCurtain = Good) OR (Towel = Good)	Bathroom = Good	1
38	(BathroomCurtain = Average) OR (Towel = Average)	Bathroom = Average	1
39	(BathroomCurtain = Poor) OR (Towel = Poor)	Bathroom = Poor	1
40	(BathroomCurtain = Terrible) OR (Towel = Terrible)	Bathroom = Terrible	1
41	(BathroomDecoration = Excellent) OR (Bathtub = Excellent) OR (HairDryer = Excellent) OR (Shower = Excellent) OR (SoapBar = Excellent) OR (Taps = Excellent) OR (Toilet = Excellent) OR (WaterPressure = Excellent) OR (OverallBathroom = Excellent)	Bathroom = Excellent	0.5
42	(BathroomDecoration = Good) OR (Bathtub = Good) OR (HairDryer = Good) OR (Shower = Good) OR (SoapBar = Good) OR (Taps = Good) OR (Toilet = Good) OR (WaterPressure = Good) OR (OverallBathroom = Good)	Bathroom = Good	0.5
43	(BathroomDecoration = Average) OR (Bathtub = Average) OR (HairDryer = Average) OR (Shower = Average) OR (SoapBar = Average) OR (Taps = Average) OR (Toilet = Average) OR (WaterPressure = Average) OR (OverallBathroom = Average)	Bathroom = Average	0.5

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
44	(BathroomDecoration = Poor) OR (Bathtub = Poor) OR (HairDryer = Poor) OR (Shower = Poor) OR (SoapBar = Poor) OR (Taps = Poor) OR (Toilet = Poor) OR (WaterPressure = Poor) OR (OverallBathroom = Poor)	Bathroom = Poor	0.5
45	(BathroomDecoration = Terrible) OR (Bathtub = Terrible) OR (HairDryer = Terrible) OR (Shower = Terrible) OR (SoapBar = Terrible) OR (Taps = Terrible) OR (Toilet = Terrible) OR (WaterPressure = Terrible) OR (OverallBathroom = Terrible)	Bathroom = Terrible	0.5
46	(BathroomCleanness = Excellent)	Bathroom = Excellent	0.1
47	(BathroomCleanness = Good)	Bathroom = Good	0.1
48	(BathroomCleanness = Average)	Bathroom = Average	0.1
49	(BathroomCleanness = Poor)	Bathroom = Poor	0.1
50	(BathroomCleanness = Terrible)	Bathroom = Terrible	0.1
51	(Robes = Excellent)	Bathroom = Excellent	0
52	(Robes = Good)	Bathroom = Good	0
53	(Robes = Average)	Bathroom = Average	0
54	(Robes = Poor)	Bathroom = Poor	0
55	(Robes = Terrible)	Bathroom = Terrible	0
56	(AccommodationCarpet = Excellent) OR (Garden = Excellent) OR (HandicapAccessibility = Excellent) OR (OffersShuttleService = Excellent) OR (Photo = Excellent) OR (Sauna = Excellent) OR (Spa = Excellent) OR (Website = Excellent)	Accommodation = Excellent	1
57	(AccommodationCarpet = Good) OR (Garden = Good) OR (HandicapAccessibility = Good) OR (OffersShuttleService = Good) OR (Photo = Good) OR (Sauna = Good) OR (Spa = Good) OR (Website = Good)	Accommodation = Good	1

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
58	(AccommodationCarpet = Average) OR (Garden = Average) OR (HandicapAccessibility = Average) OR (OffersShuttleService = Average) OR (Photo = Average) OR (Sauna = Average) OR (Spa = Average) OR (Website = Average)	Accommodation = Average	1
59	(AccommodationCarpet = Poor) OR (Garden = Poor) OR (HandicapAccessibility = Poor) OR (OffersShuttleService = Poor) OR (Photo = Poor) OR (Sauna = Poor) OR (Spa = Poor) OR (Website = Poor)	Accommodation = Poor	1
60	(AccommodationCarpet = Terrible) OR (Garden = Terrible) OR (HandicapAccessibility = Terrible) OR (OffersShuttleService = Terrible) OR (Photo = Terrible) OR (Sauna = Terrible) OR (Spa = Terrible) OR (Website = Terrible)	Accommodation = Terrible	1

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
61	(AccommodationCurtain = Excellent) OR (BaggageRoom = Excellent) OR (Bathroom = Excellent) OR (Beergarden = Excellent) OR (Board = Excellent) OR (Breakfast = Excellent) OR (Café = Excellent) OR (Cocktail = Excellent) OR (Elevator = Excellent) OR (EnglishNewspaper = Excellent) OR (FitnessRoom = Excellent) OR (Food = Excellent) OR (Furniture = Excellent) OR (LocalFoodStall = Excellent) OR (MassageStudion = Excellent) OR (MoneyExchangeOffice = Excellent) OR (offersCot = Excellent) OR (offersCrib = Excellent) OR (offersLaudryCleaning = Excellent) OR (OverallAccommodation = Excellent) OR (Parking = Excellent) OR (PetsAllowed = Excellent) OR (Pool = Excellent) OR (Room = Excellent) OR (Salon = Excellent) OR (Security = Excellent) OR (Service = Excellent) OR (Solarium = Excellent) OR (SteamBath = Excellent) OR (Value = Excellent) OR (WellnessFacilities = Excellent)	Accommodation = Excellent	0.5

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
62	(AccommodationCurtain = Good) OR (BaggageRoom = Good) OR (Bathroom = Good) OR (Beergarden = Good) OR (Board = Good) OR (Breakfast = Good) OR (Café = Good) OR (Cocktail = Good) OR (Elevator = Good) OR (EnglishNewspaper = Good) OR (FitnessRoom = Good) OR (Food = Good) OR (Furniture = Good) OR (LocalFoodStall = Good) OR (MassageStudion = Good) OR (MoneyExchangeOffice = Good) OR (offersCot = Good) OR (offersCrib = Good) OR (offersLaudryCleaning = Good) OR (OverallAccommodation = Good) OR (Parking = Good) OR (PetsAllowed = Good) OR (Pool = Good) OR (Room = Good) OR (Salon = Good) OR (Security = Good) OR (Service = Good) OR (Solarium = Good) OR (SteamBath = Good) OR (Value = Good) OR (WellnessFacilities = Good)	Accommodation = Good	0.5

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
63	(AccommodationCurtain = Average) OR (BaggageRoom = Average) OR (Bathroom = Average) OR (Beergarden = Average) OR (Board = Average) OR (Breakfast = Average) OR (Café = Average) OR (Cocktail = Average) OR (Elevator = Average) OR (EnglishNewspaper = Average) OR (FitnessRoom = Average) OR (Food = Average) OR (Furniture = Average) OR (LocalFoodStall = Average) OR (MassageStudion = Average) OR (MoneyExchangeOffice = Average) OR (offersCot = Average) OR (offersCrib = Average) OR (offersLaudryCleaning = Average) OR (OverallAccommodation = Average) OR (Parking = Average) OR (PetsAllowed = Average) OR (Pool = Average) OR (Room = Average) OR (Salon = Average) OR (Security = Average) OR (Service = Average) OR (Solarium = Average) OR (SteamBath = Average) OR (Value = Average) OR (WellnessFacilities = Average)	Accommodation = Average	0.5

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	**If**	**Then**	**Weight**
64	(AccommodationCurtain = Poor) OR (BaggageRoom = Poor) OR (Bathroom = Poor) OR (Beergarden = Poor) OR (Board = Poor) OR (Breakfast = Poor) OR (Café = Poor) OR (Cocktail = Poor) OR (Elevator = Poor) OR (EnglishNewspaper = Poor) OR (FitnessRoom = Poor) OR (Food = Poor) OR (Furniture = Poor) OR (LocalFoodStall = Poor) OR (MassageStudion = Poor) OR (MoneyExchangeOffice = Poor) OR (offersCot = Poor) OR (offersCrib = Poor) OR (offersLaudryCleaning = Poor) OR (OverallAccommodation = Poor) OR (Parking = Poor) OR (PetsAllowed = Poor) OR (Pool = Poor) OR (Room = Poor) OR (Salon = Poor) OR (Security = Poor) OR (Service = Poor) OR (Solarium = Poor) OR (SteamBath = Poor) OR (Value = Poor) OR (WellnessFacilities = Poor)	Accommodation = Poor	0.5

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
65	(AccommodationCurtain = Terrible) OR (BaggageRoom = Terrible) OR (Bathroom = Terrible) OR (Beergarden = Terrible) OR (Board = Terrible) OR (Breakfast = Terrible) OR (Café = Terrible) OR (Cocktail = Terrible) OR (Elevator = Terrible) OR (EnglishNewspaper = Terrible) OR (FitnessRoom = Terrible) OR (Food = Terrible) OR (Furniture = Terrible) OR (LocalFoodStall = Terrible) OR (MassageStudion = Terrible) OR (MoneyExchangeOffice = Terrible) OR (offersCot = Terrible) OR (offersCrib = Terrible) OR (offersLaudryCleaning = Terrible) OR (OverallAccommodation = Terrible) OR (Parking = Terrible) OR (PetsAllowed = Terrible) OR (Pool = Terrible) OR (Room = Terrible) OR (Salon = Terrible) OR (Security = Terrible) OR (Service = Terrible) OR (Solarium = Terrible) OR (SteamBath = Terrible) OR (Value = Terrible) OR (WellnessFacilities = Terrible)	Accommodation = Terrible	0.5
66	(Location = Excellent) OR (AccommodationCleanness = Excellent) OR (AccommodationDecoration = Excellent) OR (Dinner = Excellent) OR (Lobby = Excellent) OR (OffersChildCare = Excellent) OR (Restaurant = Excellent)	Accommodation = Excellent	0.1
67	(Location = Good) OR (AccommodationCleanness = Good) OR (AccommodationDecoration = Good) OR (Dinner = Good) OR (Lobby = Good) OR (OffersChildCare = Good) OR (Restaurant = Good)	Accommodation = Good	0.1
68	(Location = Average) OR (AccommodationCleanness = Average) OR (AccommodationDecoration = Average) OR (Dinner = Average) OR (Lobby = Average) OR (OffersChildCare = Average) OR (Restaurant = Average)	Accommodation = Average	0.1

Table C.1 The fuzzy inference rules including weights (Continued)

Rules	If	Then	Weight
69	(Location = Poor) OR (AccommodationCleanness = Poor) OR (AccommodationDecoration = Poor) OR (Dinner = Poor) OR (Lobby = Poor) OR (OffersChildCare = Poor) OR (Restaurant = Poor)	Accommodation = Poor	0.1
70	(Location = Terrible) OR (AccommodationCleanness = Terrible) OR (AccommodationDecoration = Terrible) OR (Dinner = Terrible) OR (Lobby = Terrible) OR (OffersChildCare = Terrible) OR (Restaurant = Terrible)	Accommodation = Terrible	0.1

CURRICULUM VITAE

Miss Phichayasini Kitwatthanathawon was born on September 21, 1984 in Nakhon Ratchasima Province. She received Bachelor and Master of Information Science from Suranaree University of Technology, Thailand in 2006 and 2009, respectively. In 2010, she has got a scholarship from Suranaree University of Technology research to pursue her doctoral degree in Information Technology Program at Suranaree University of Technology. Her major research interests are in information extraction, natural language processing and fuzzy logic.