Chen, Y.J. Chu, H.C. and Chen, Y.M. (2016), 'Acquiring consumer perspectives in Chinese eWOM', *Journal of Information Management*, Vol. 23, No. 2, pp. 187-216.

Acquiring Consumer Perspectives in Chinese eWOM

Yuh-Jen Chen
Department of Accounting and Information Systems, National Kaohsiung First University of Science and Technology

Hui-Chuan Chu
Department of Special Education, National University of Tainan

Yuh-Min Chen*
Institute of Manufacturing Information and Systems, National Cheng Kung University

Abstract

Purpose－This study develops a method for acquiring consumer perspectives in Chinese electronic word-of-mouth (eWOM) to assist enterprises to automatically generalize and acquire consumer comments from the Internet and to rapidly realize the hidden consumer perspectives.

Design/methodology/approach － The methodology can be developed by performing the following tasks: (1) designing a consumer perspective framework, (2) designing a process for acquiring consumer perspectives, (3) developing techniques related to consumer perspective acquisition, and (4) implementing a mechanism for acquiring consumer perspectives. Finally, a system evaluation with the Delphi method for firm satisfaction is conducted to demonstrate that the developed method and system worked efficiently.

Findings－After measuring firm satisfaction with the Delphi method, the findings indicated that over eighty percent of companies satisfied with the results from acquiring consumer perspectives in Chinese eWOM.

* Corresponding author. Email: ymchen@mail.ncku.edu.tw
2014/08/27 received; 2015/01/12 revised; 2015/04/02 accepted

Research limitations/implications－Future research can extend the development of eWOM polarity analysis to effectively help an enterprise rapidly and accurately realize the current situation of eWOM to improve customer relationships.

Practical implications－It is expected that the consumer perspectives offer enterprises an important reference of improving the internal environment and service innovation, and thus increase their global competitiveness.

Originality/value－This study firstly designs an acquisition process to help an enterprise automatically acquire consumer comments from the Internet and rapidly reveal hidden consumer perspectives in Chinese eWOM contents. Based on the designed consumer perspective acquisition process, the techniques for acquisition are developed to facilitate the implementation of the consumer perspective acquisition system. Finally, the system is implemented to instantaneously acquire the Chinese eWOM articles of consumers and effectively extract hidden consumer perspectives to assist enterprises in internal improvement and service innovation.

Keywords: customer relationship management (CRM), consumer perspective, electronic word-of-mouth (eWOM), data mining

陳育仁、朱慧娟、陳裕民（2016），『中文網路口碑之消費者觀點擷取』，
中華民國資訊管理學報，第二十三卷，第二期，頁 187-216。

中文網路口碑之消費者觀點擷取

陳育仁
國立高雄第一科技大學會計資訊系

朱慧娟
國立台南大學特殊教育學系

陳裕民*
國立成功大學製造資訊與系統研究所

摘要

消費者觀點係為企業產品創新、服務改善的重要依據；過去企業主要透過銷售人員與消費者之間的互動、專家訪談以及問卷調查等方式來瞭解消費者觀感，然而隨著網路技術的蓬勃發展，越來越多消費者會在網路上發表企業評論，這也意味著企業有著另一種不同的管道可以更客觀地瞭解消費者觀點。但是，大量且過載的網路資訊難以有效地被整理、歸納與分析，導致企業往往無法迅速且清楚地瞭解消費者觀點或需求，進而作出正確的決策。

本研究目的在於針對中文網路口碑發展一消費者觀點獲取方法，以協助企業能自動歸納與獲取消費者網路評論，快速瞭解網路評價內容中潛藏的消費者觀點，進而作為企業內部改善以及服務創新之依據，藉此提昇企業整體競爭優勢。針對上述目的，本研究主要研究項目包括：(1)消費者觀點架構設計、(2)消費者觀點獲取流程設計、(3)消費者觀點獲取方法發展以及(4)消費者觀點獲取機制實作。

關鍵詞：顧客關係管理、消費者觀點、網路口碑、資料探勘

* 本文通訊作者。電子郵件信箱：ymchen@mail.ncku.edu.tw
2014/08/27 投稿；2015/01/12 修訂；2015/04/02 接受

1. Introduction

With the rapid development of information and communication technology and the globalization of markets, the interaction between consumers and enterprises becomes increasingly frequent, and product varieties and competitors are increasing. Modern business environments therefore have become stricter, and the product-oriented business model has changed to a customer-oriented model to satisfy consumer demands (Lin et al. 2006; Ngai et al. 2009).

The consumer perspective is an important reference for enterprises in product innovation and service improvement. Enterprise owners have traditional understood consumer perceptions through the interaction between salespeople and consumers, expert interviews, and questionnaire surveys (Knight & Cavusgil 2004). The growth of Internet technology has caused consumers to increasingly issue corporate comments online, indicating that enterprises can more objectively understand consumer perspectives through different channels (Hennig-Thurau et al. 2004). However, the abundance of information available on the Internet, which creates a situation of information overload, creates difficulties in the effective organization, generalization, and analysis of information. This situation thus prevents enterprises from rapidly and clearly understanding consumer perspectives or demands in relation to making correct decisions. This circumstance has become a key issue in rapidly and correctly analyzing the abundant Internet comments made by consumers to enhance enterprise competitive advantage.

Recently, various methods have been developed to acquire consumer perspectives in eWOM. For example, Guo et al. (2009) extracted consumer perspectives on products from semi-structural product comments, and generalized these product perspectives using unsupervised learning. Furthermore, Xia, Zong and Li (2011) conducted a comparative study of the effectiveness of the ensemble technique for sentiment classification. The ensemble framework was applied to sentiment classification tasks, with the aim of efficiently integrating different feature sets (i.e., the part-of-speech based feature sets and the word-relation based feature sets) and classification algorithms (i.e., nai¨ve Bayes, maximum entropy and support vector machines) to synthesize a more accurate classification procedure. Archak, Ghose and Ipeirotis (2011) analyzed customer responses and sales quantity on the Amazon shopping platform to determine consumer preferences in relation to digital cameras. Meanwhile, Zhai et al. (2011)

applied semi-supervised learning to product perspective clustering. Moreover, Netzer et al. (2012) analyzed user comments on automobile and drug forums using semantic analysis and text mining techniques. However, these recent studies focused primarily on differentiating positive/negative opinions of consumers to determine consumer satisfaction, and ignored consumer perspectives in eWOM, thus preventing an enterprise from effectively understanding internal problems or consumer demands.

This study develops a method for acquiring consumer perspectives in Chinese electronic word-of-mouth (eWOM) to assist enterprises to automatically generalize and acquire consumer comments from the Internet and to rapidly realize the hidden consumer perspectives. These valuable consumer perspectives provide an important reference of the internal environment improvement and service innovation of enterprises, and thus increase their global competitiveness. The study objectives can be realized by performing the following tasks: (1) designing a consumer perspective framework, (2) designing a process for acquiring consumer perspectives, (3) developing techniques related to consumer perspective acquisition, and (4) implementing a mechanism for acquiring consumer perspectives. Meanwhile, developing techniques associated with consumer perspective acquisition involves the retrieval of eWOM content, the extraction of dimension instance name and consumer perspectives, the generalization of dimension instance name, and the integration of consumer perspectives. Finally, a system evaluation with the Delphi method for firm satisfaction is conducted to demonstrate the developed method and system worked efficiently.

2. Design of Consumer Perspective Framework

Based on the concept of service package in service science, this study designs the consumer perspective framework to represent the consumer perspective acquisition mode, as illustrated in Fig. 1. The consumer perspective framework comprises five levels, namely company, dimension, dimension instance, perspective, and perspective instance. Meanwhile, the company represents the target company in consumer perspectives. The dimension comprises the product, service, and facility of a target company, while the dimension instance represents the product, service, and facility items. The perspective is defined as consumer sense and psychological perceptions, while the perspective instance expresses the real experiences of consumers and their actual psychological perceptions.

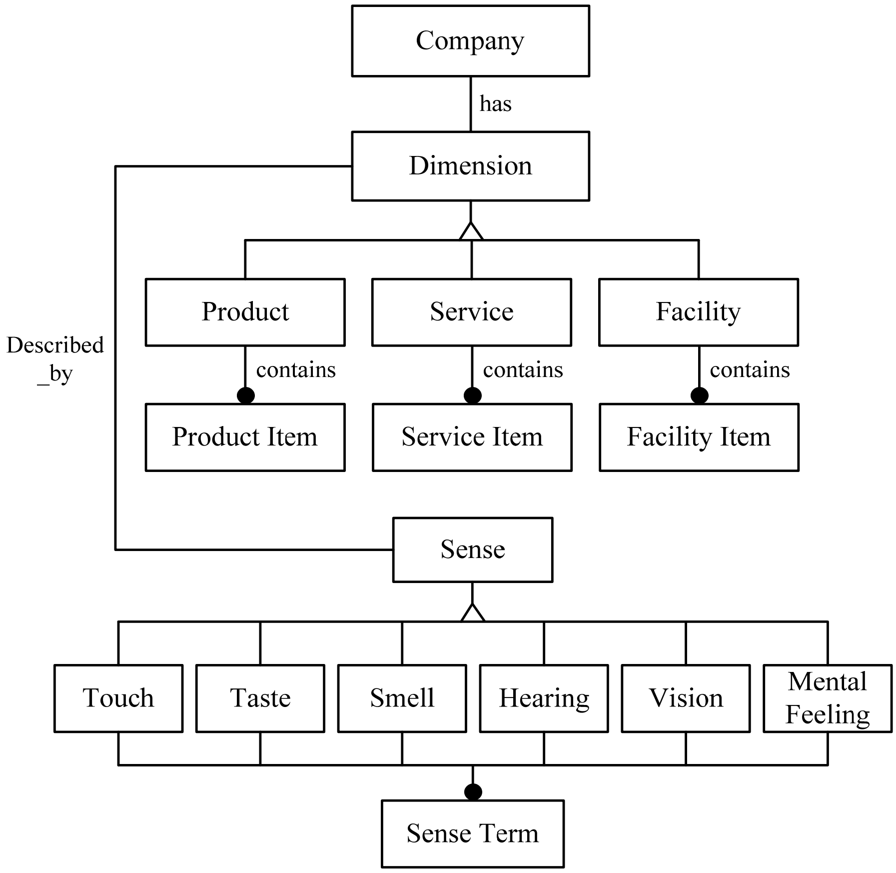

Figure 1: Consumer Perspective Framework

3. Design of Consumer Perspective Acquisition Process

Based on the consumer perspective framework designed in Section 2, this section establishes the consumer perspective acquisition process to effectively analyze consumer perspectives in Chinese eWOM, and provides enterprises with valuable references for making correct decisions. As shown in Fig. 2, the process includes five phases, namely eWOM content retrieval, dimension instance name extraction, consumer perspective extraction, dimension instance name generalization, and consumer perspective integration, which are summarized as follows.

3.1 eWOM Content Retrieval

eWOM content is complex, and mostly comprises semi-structured or unstructured data. In this case, two steps are conducted to retrieve Chinese eWOM contents to effectively acquire consumer perspectives.

1. eWOM article collection: The contents related to a target company inputted by the user are retrieved and stored in the Chinese eWOM article repository.
2. eWOM content preprocessing: According to the Chinese eWOM articles stored in the repository, these semi-structured or unstructured contents are transferred into structured contents through noise filtering, sentence splitting, tokenization, and part-of-speech tagging.

3.2 Dimension Instance Name Extraction

Dimension instance name extraction involves searching for dimension instance names from the preprocessed Chinese eWOM contents, through the following two steps.

1. Feature tagging for the dimension instance name: Based on the defined product named entities and POS features, the names of dimension instances are tagged as features.
2. Tagging of dimension instance name: The tagged features in step (1) are first trained to obtain optimal parameter values. The trained parameters are then used to tag the dimension instance names of the testing data, after which the tagged contents are analyzed.

3.3 Consumer Perspective Extraction

Consumer perspective extraction mainly acquires consumer perspective instances from the preprocessed Chinese eWOM contents using POS combination matching. For example, “taste” in cream with an excellent taste is a perspective instance.

3.4 Dimension Instance Name Generalization

Different consumers may describe the same dimension using distinct names. Thus, the dimension instance name is generalized via the following two steps.

1. Similarity calculation of the dimension instance name: NGD (Normalized Google Distance) is performed to determine the strength of the relationship among dimension instance names.
2. Clustering of dimension instance name: The dimension instance names can be

clustered with a threshold value according to the determined relationship strength among them.

3.5 Consumer Perspective Integration

In the consumer perspective integration, the product and consumer perspectives are mapped into the designed consumer perspective framework by the following steps.

1. Establishment of a consumer perspective document: Vocabulary relevant to the consumer perspective in Chinese eWOM documents is retrieved and combined to create a document for representing the consumer perspective.
2. Classification of the consumer perspective: The perspective features and the optimal combination of perspective features are first selected from the document generated in Step (1). Based on the combination of features, the clustering effect is then trained and tested.

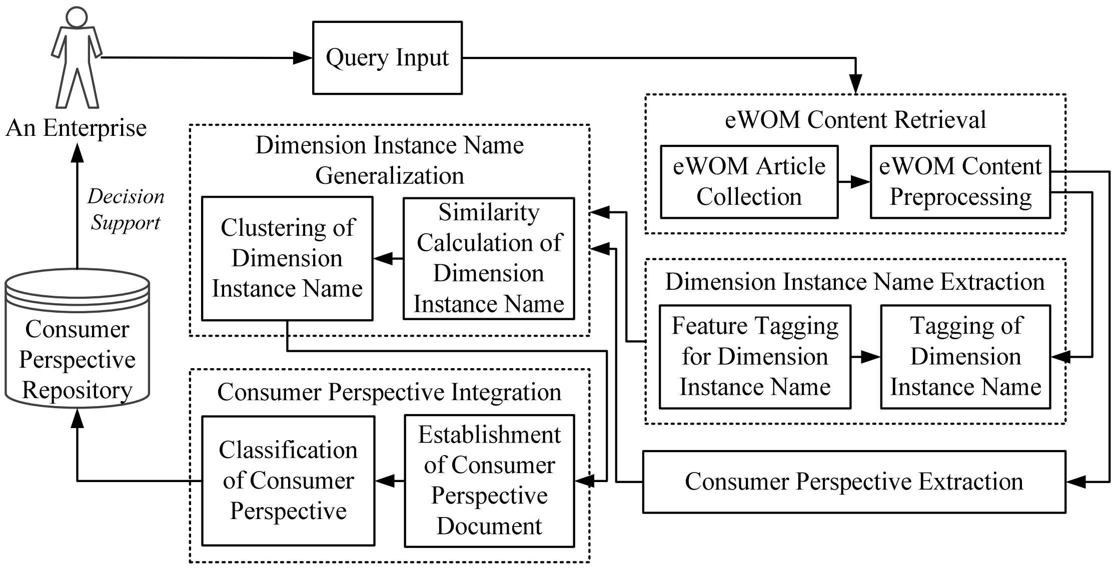

Figure 2: Consumer Perspective Acquisition Process

4. Development of Consumer Perspective Acquisition Techniques

Based on the consumer perspective acquisition process designed in Section 3, this section develops techniques for consumer perspective acquisition, including “eWOM content retrieval”, “dimension instance name extraction”, “consumer perspective extraction", "dimension instance name generalization", and "consumer perspective integration".

4.1 eWOM Content Retrieval

eWOM content retrieval includes the two steps of Chinese eWOM article collection and Chinese eWOM content preprocessing, as described below.

1. Chinese eWOM article collection

   Using a company name as the keyword, the Crawler is used to collect the relevant Chinese eWOM articles from the websites "bloger.com", "wordpress.com" and "google.com/bloggersearch". Figure 3 depicts the algorithm for Chinese eWOM article collection. In Fig. 3, URL list, seed array and crawled URL array denote three different URL storage structures for retrieving Chinese eWOM articles and contents.

2. Chinese eWOM content preprocessing

   Before preprocessing Chinese eWOM contents, all the Chinese eWOM articles retrieved from different websites are transferred with a consistent formatting, which can thus be preprocessed by the algorithm designed in Fig. 4. In this preprocessing, noise filtering, sentence splitting, tokenization and POS tagging are performed in order based on the collected Chinese eWOM articles. First, the noise filtering utilizes a filtering noise method proposed by Zheng, Cheng and Chen (2008), which can remove advertising noise from the web contents and external links to reserve the main Chinese eWOM content. Subsequently, sentence splitting segments the Chinese eWOM contents into sentences. Based on these sentences, tokenization is executed to segment named entities or general terms. Finally, the segmented named entities or general terms are tagged by POS. Figure 4 illustrates the algorithm for Chinese eWOM content preprocessing.

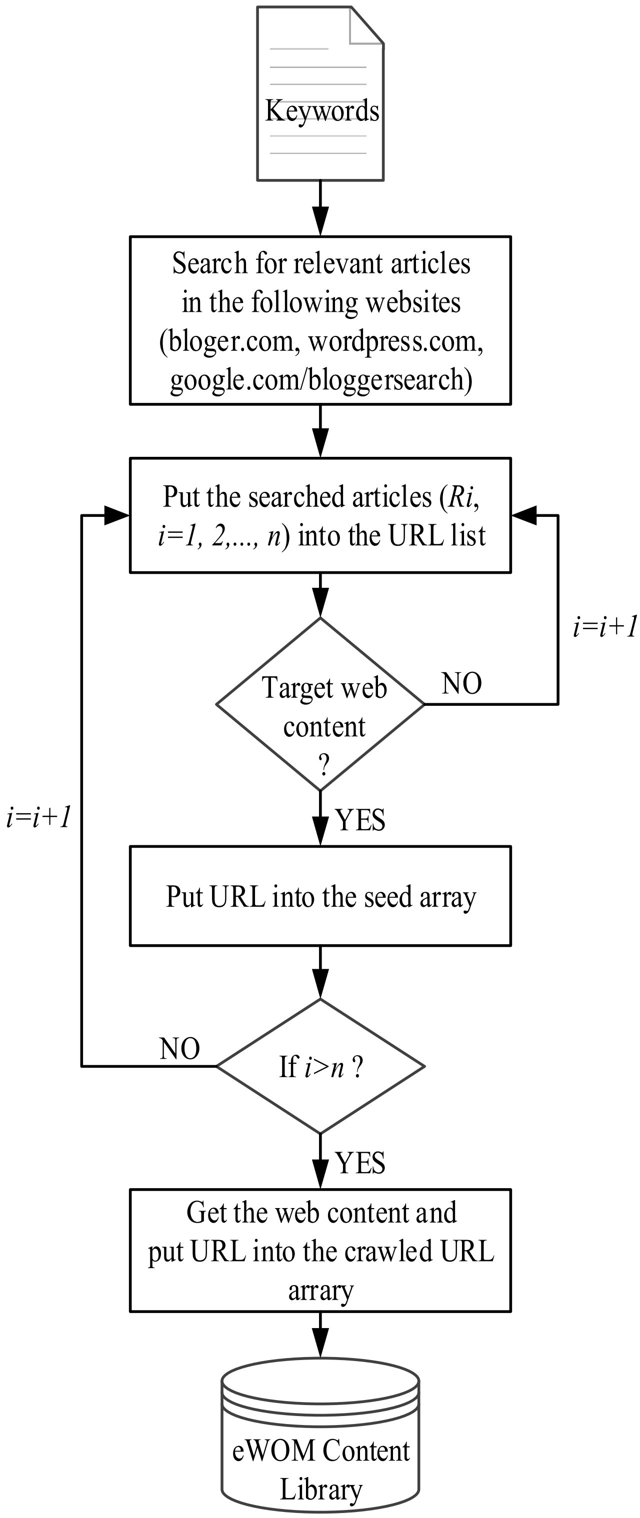

Figure 3: Algorithm for Chinese eWOM Article Collection

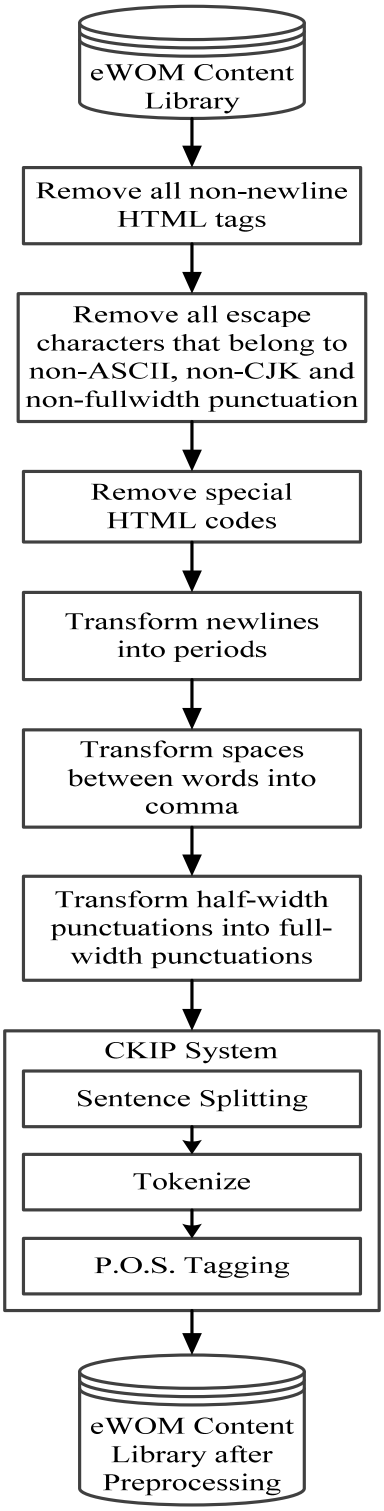

Figure 4: Algorithm for Chinese eWOM Content Preprocessing

4.2 Dimension Instance Name Extraction

According to the preprocessed Chinese eWOM contents from Section 4.1, the conditional random field (CRF), which presents favorable performance in natural language processing, is adopted to extract dimension instance names from the Chinese eWOM contents (Pang & Lee 2004; Pan & Wang 2011; Zhang et al. 2009). The steps are detailed below.

1. Feature tagging for the dimension instance name

   As indicated in Fig. 5, the dimension instance features are first defined. Based on the POS of the defined features, the dimension instance names of Chinese eWOM contents are tagged. Finally, the tagged dimension instance names are divided into training and testing datasets to pave the way for dimension instance name tagging, as follows.

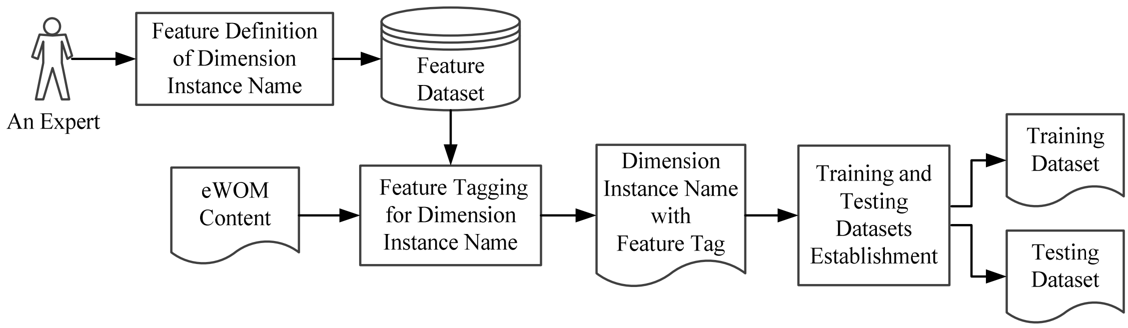

Figure 5: Feature Tagging Process for Dimension Instance Name

(1) Feature definition: The features of dimension instance name in Chinese eWOM are defined based on the named entity recognition field (Tsai & Chou 2011; Zhao & Liu 2008). These features include keywords of the named entity and POS feature, as well as the enlargement, bold, color, and quotation marks of words in the Chinese eWOM contents.

(2) Feature tagging: Based on the features defined in Step (a), the Chinese eWOM contents are tagged using "BIO" label. The "B" represents the beginning of a named entity, while "I" represents the content of a named entity, and "O" denotes a position that is not part of a named entity. Table 1 lists an example of feature tagging.

Table 1: An Example of Feature Tagging for Dimension Instance Name

Input Chinese Unit	POS Feature	Feature Tag
裕珍馨(Yu-Jan-Shin)	Nb	O
的(Of)	DE	O
奶油(Butter)	Na	B
酥餅(Crispy Cake)	Na	I
名聲(Reputation)	Na	O
很(Very)	Dfa	O
響亮(Prosperous)	VH	O

(3) Training and testing datasets establishment: This study conducts ten-fold cross validation to train and test the feature tagging for the dimension instance name. The data are divided into ten subsamples. Meanwhile, one subsample is considered as the testing dataset and nine subsamples are treated as the training dataset.

2. Tagging of dimension instance name

The established training datasets are first used to train the CRF parameters. Next, the testing datasets are tested by the trained parameters. Finally, CRF tagging and analysis are performed to determine the named entities in dimension instance names, as shown in Fig. 6.

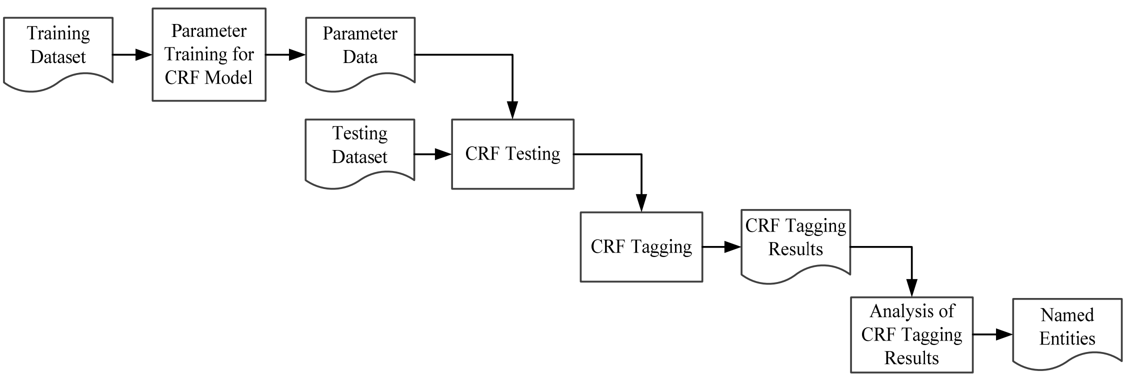

Figure 6: Dimension Instance Name Tagging Process for Predicting Named Entities

(1) Parameter training for the CRF model

The training dataset is used to train parameter λ_k in the CRF formula, as illustrated in Eq. (1).

$$P(Y \mid X) = \frac{1}{Z}\exp\left(\sum_t \sum_k \lambda_k f_k(y_{t-1}, y_t, x, t)\right) \tag{1}$$

Where Z denotes the normalized coefficient, which results in the probability sum of all state sequences being 1;
X denotes the input observation sequence;
Y represents the correspondent state sequence;
$f_k(y_{t-1}, y_t, x, t)$ denotes the feature function for a state sequence to be transferred from position t-1 to position t;
λ_k represents the weight of each feature function;

The feature function $f_k(y_{t-1}, y_t, x, t)$ is normally a binary-valued function, and parameter λ_k can be regarded as the weight of the function f_k. In this study, when a word is tagged as the symbol "B" (i.e., morpheme-initial of named entity), the next word appears a high probability of being tagged as the symbol "I". The feature function thus can be defined as Eq. (2). Through the feature function, the corresponding parameter λ_k can be trained to derive a higher weight.

$$f_k(y_{t-1}, y_t, X, t) = \begin{cases} 1, \text{if } y_{t-1} = B \text{ and } y_t = I \\ 0, \text{otherwise} \end{cases} \tag{2}$$

(2) CRF testing

Based on the weight of the parameter λ_k trained by CRF in step (a) and the testing dataset, the F-Measure is adopted to evaluate the testing results, and ensure the effectiveness of the CRF Tagging. Equation (3) illustrates the formula of the F-Measure.

$$F = \frac{2PR}{P+R} \tag{3}$$

where F represents the F-Measure results of CRF Tagging;
P denotes the precision of CRF tagging;
R is the recall of CRF tagging;

(3) CRF tagging

The trained and tested CRF model is used to tag Chinese eWOM contents.

Table 2 shows an example of CRF Tagging.

Table 2: An Example of CRF Tagging

Input Chinese Unit	POS Feature	CRF Tagging
相(Co)	ADV	O
呼應(Respond)	Vt	O
的(Of)	T	O
「	AA	O
大甲(Dajia)	N	B
積木(Block)	N	I
」	AA	O

(4) Analysis of CRF tagging results

The CRF tags contain three symbols, namely “B, “I”, and “O”. The dimension instance name can be acquired from the following analysis of CRF tagging results.

a. If the inputted Chinese unit is tagged as the symbol “I” or “B” and the previous inputted Chinese unit is tagged as the symbol “O”, or if the inputted Chinese unit is the first inputted Chinese unit of the dimension instance name, then the inputted Chinese unit is determined as the initial unit of the dimension instance name.

b. If the inputted Chinese unit is tagged as the symbol “I” and the previous inputted Chinese unit is tagged as the symbol “I” or “B”, then the inputted Chinese unit is determined to be part of dimension instance name.

c. If the inputted Chinese unit is tagged as the symbol “B” or “I” and the next inputted Chinese unit is tagged as the symbol “O”, then the inputted Chinese unit is determined as the last unit of the dimension instance name.

4.3 Consumer Perspective Extraction

Previous studies on customer perspective extraction have focused mainly on nouns or noun combinations. However, not all nouns and noun combinations revealed meaningful customer perspectives. For example, the sentence “I look for a pair of glasses on the Internet for three months”, in which there is no adjectives to modify the noun “glasses”. The noun “glasses” in the sentence thus is not a perspective instance. In this study, the POS combination proposed by Pan and Wang (2011) is utilized to extract consumer perspectives. As shown in Table 3, when the number of terms acquired from the Chinese eWOM content preprocessing is 2, 3 or 4, respectively, then the correspondent POS combinations 「na、an、vn」, 「nda、adv、nav」 or 「nvda、ndav」 is chosen to match the POS of the terms. Furthermore, while the number of terms acquired from the Chinese eWOM content preprocessing exceeds 4, the correspondent POS combination 「nav、na、nva」 is chosen as the POS of those terms. In the POS combination, “n” denotes a noun, “a” represents an adjective, “d” is an adverb, and “v” denotes a verb, as listed in Table 4.

Table 3: POS Combination Mode

Term Num.	POS Combination		
2	na	an	vn
3	nda	ndv	nav
4	nvda	ndav	---
exceeds 4	nav	na	nva

Table 4: POS Tag Comparison Table

Tag	POS
n	Noun
a	Adjective
d	Adverb
v	Verb

4.4 Dimension Instance Name Generalization

In Chinese eWOM, consumers generally use various names to describe the same dimension instance. For example, the name “4s” is used rather than the name “iphone4s”. Thus, this section first uses the normalized Google distance (NGD) to calculate the similarity of dimension instance name. Based on the similarity calculation, a Beta-similarity graph is then created for clustering dimension instance names.

1. Similarity calculation of dimension instance name

The NGD calculates the relationship strength among nodes based on information distance and Kolmogorov complexity (Cilibrasi & Vitanyi 2007). This study thus uses the NGD to calculate the similarity of the dimension instance name. The Normalized Google Distance for two terms x and y can be defined in Eq. (4).

$$NGD(x,y)=\frac{\max\{\log f(x),\log f(y)\}-\log f(x,y)}{\log N-\min\{\log f(x),\log f(y)\}} \tag{4}$$

where $f(x)$ and $f(y)$ are the number of search results returned by Google for the search query term x and y respectively;
$f(x,y)$ denotes the number of search results returned by Google for the search query containing both terms x and y;
N denotes the number of documents searched by Google, which is estimated to be around 10^{10};

Using the examples of the Chinese terms "奶油酥餅(butter crispy cake)" and "紫芋酥(taro pastry)" in a Taiwanese bakery, the similarity between the two terms is calculated using NGD, as follows. In this case study, the time interval is fixed to consecutive periods of 7 days each.

$$\begin{aligned}NGD(\text{奶油酥餅},\text{紫芋酥})&=\frac{max\{log\ f(\text{奶油酥餅}),log\ f(\text{紫芋酥})\}-log\ f(\text{奶油酥餅},\text{紫芋酥})}{log\ N-min\{log\ f(\text{奶油酥餅}),log\ f(\text{紫芋酥})\}}\\&=\frac{max\{log\ 2280000,log\ 229000\}-log\ 141000}{10-min\{log\ 2280000,log\ 229000\}}\\&=\frac{log\ 2280000-log\ 141000}{10-log\ 229000}\\&=0.26\end{aligned}$$

2. Clustering of the dimension instance name

The Beta-similarity graph is an undirected graph, in which each node represents a dimension instance name. While nodes i and j are connected, the similarity between them exceeds the threshold value Beta; moreover, when nodes i and j are not connected, the similarity between them is below the threshold value Beta. Additionally, when a new dimension instance name is included, the similarity between the new dimension instance name and other dimension instance names is calculated. Two dimension instance names with higher similarity exceeding the threshold value Beta are considered as a cluster.

The dimension instance names collected in this study form a set of vocabulary. The clustering of dimension instance name thus can be considered the clustering of vocabulary. This study adopts the NGD to calculate the similarity distance between terms. Lower value of NGD indicates higher similarity between terms. Subtracting the value of NGD from 1 expressing the similarity between terms, larger similarity reveals closer terms. Equation (5) illustrates the similarity calculation between terms with NGD. Figure 7 shows an example of the undirected graph for Chinese term similarity, where the value 0.99 is set (Hou et al. 2011) as the threshold value Beta for the Chinese term clustering, as shown in Fig. 8.

$$Similarity(x, y) = 1 - NGD(x, y) \tag{5}$$

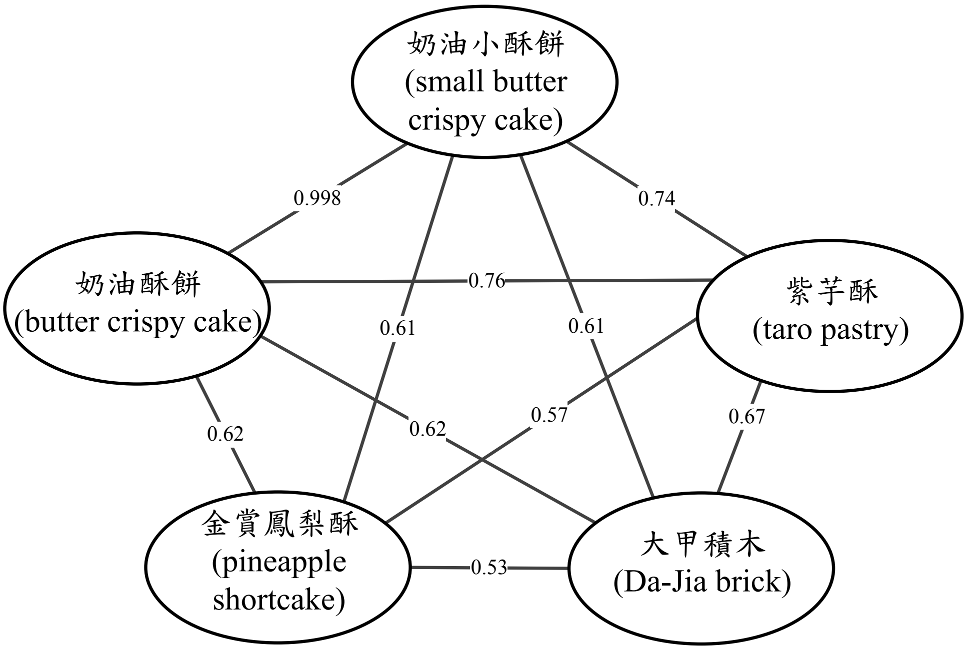

Figure 7: Undirected Graph of Chinese Term Similarity (Illustrative Example)

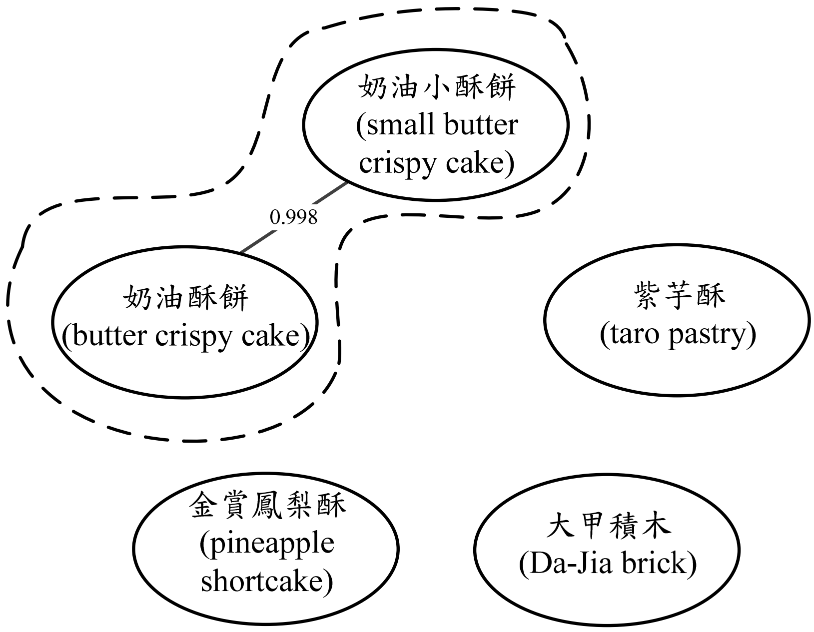

Figure 8: The Results of the Chinese Term Clustering (Beta is set as 0.99; Hou et al. 2011)

4.5 Consumer Perspective Integration

In Chinese eWOM, consumers are likely to express the same perspective using different terms. This study thus utilizes document classification to solve problems involving consumer perspective classification. Document classification mainly adopts the Naïve Bayesian classifier with high efficiency and precision (Chen et al. 2009) to classify Chinese eWOM documents. The detailed steps involve the establishment of a consumer perspective document and the classification of the consumer perspective, as illustrated below.

1. Establishment of the consumer perspective document

   Based on the perspective instance names extracted in Section 4.3, sentences containing the perspective instance names are first selected from the Chinese eWOM documents. The terms defined in Table 5 are then extracted from the selected sentences. Finally, these extracted terms are stored in a document, which can represent the consumer perspective. Figure 9 depicts the algorithm used to establish a consumer perspective document.

Table 5: Term Extraction Scope for Establishing Consumer Perspective Documents

Scope of Term Extraction	Description
[-t, t]	The range of the terms which appears near the perspective X.
$adj_{-1}(X)$, $adj_{+1}(X)$, $adj_{-2}(X)$, $adj_{+2}(X)$	The nearest four adjective terms to the perspective X.
$adj_{-1}(X-1, X+1)$, $adj_{+1}(X-1, X+1)$, $adj_{-2}(X-1, X+1)$, $adj_{+2}(X-1, X+1)$	The nearest four adjective terms to the previous and the next perspective of the perspective X.
X-3, X-2, X-1, X+3, X+2, X+1	The previous/next three perspectives of the perspectives X.

2. Classification of the consumer perspective

In the classification of the consumer perspective, the documents established in step (1) are used to train the Naïve Bayesian classifier model, as shown in Eq. (6).

$$\begin{aligned} P(C_i \mid D) &= \frac{P(D \mid C_i)P(C_i)}{P(D)} \\ &\propto P(D \mid C_i)P(C_i) \\ &= P(w_1 \mid C_i) \times P(w_2 \mid C_i) \times P(w_3 \mid C_i) \times \cdots \times P(w_n \mid C_i) \times P(C_i) \\ &= \prod_{j=1}^{n} P(w_j \mid C_i) \times P(C_i) \end{aligned} \tag{6}$$

where C_i is the *i-th* consumer perspective class;

$D = (w_1, w_2, w_3, ..., w_n)$ denotes a consumer perspective document;

$P(C_i \mid D)$ represents the probability that a consumer perspective document D being classified to the *i-th* consumer perspective class C_i;

$P(D \mid C_i)$ is the probability that a consumer perspective document D appears in the *i-th* consumer perspective class C_i;

$P(C_i)$ is the probability of class C_i occurred;

$P(D)$ is the probability of a consumer perspective document D appeared;

$P(w_n|C_i)$ denotes the probability that the *n-th* term w_n appears in consumer perspective documents from the *i-th* consumer perspective class C_i;

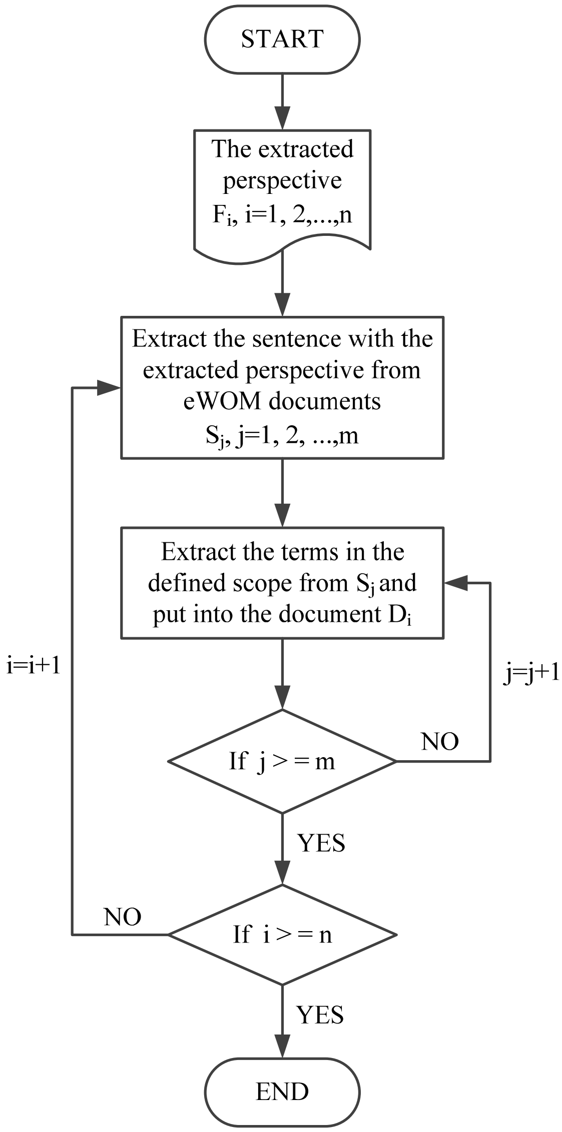

Figure 9: Algorithm for Establishing a Consumer Perspective Document

Additionally, Eq. (7) classifies the consumer perspectives given a specific consumer perspective document and maximum posterior probability. Restated, the classification value of consumer perspective document D is designated as c^* when the classification value C_i of the maximum $P(C_i \mid D)$ is designated as c^*.

$$c^* = \arg\max_i P(C_i \mid D) \tag{7}$$

5. System Implementation with a Case Study

Based on the proposed techniques for consumer perspective acquisition, this section presents a system for the acquisition of consumer perspectives in Chinese eWOM. The following subsections describe the implementation environment, results with a foodstuff case, and a system evaluation for company satisfaction.

5.1 Implementation Environment

This study implemented a prototype of the consumer perspective acquisition at the Knowledge Engineering and Management Laboratory, National Kaohsiung First University of Science and Technology. The implementation environment consisted of an Intel Core 2 Duo E7500, 2.93GHz PC running Microsoft Windows 7 Professional, Python 2.7, C sharp, and My SQL 5.0.51b. Figure 10 illustrates the framework of the consumer perspective acquisition system, which includes the three layers of user service, business service, and knowledge repository.

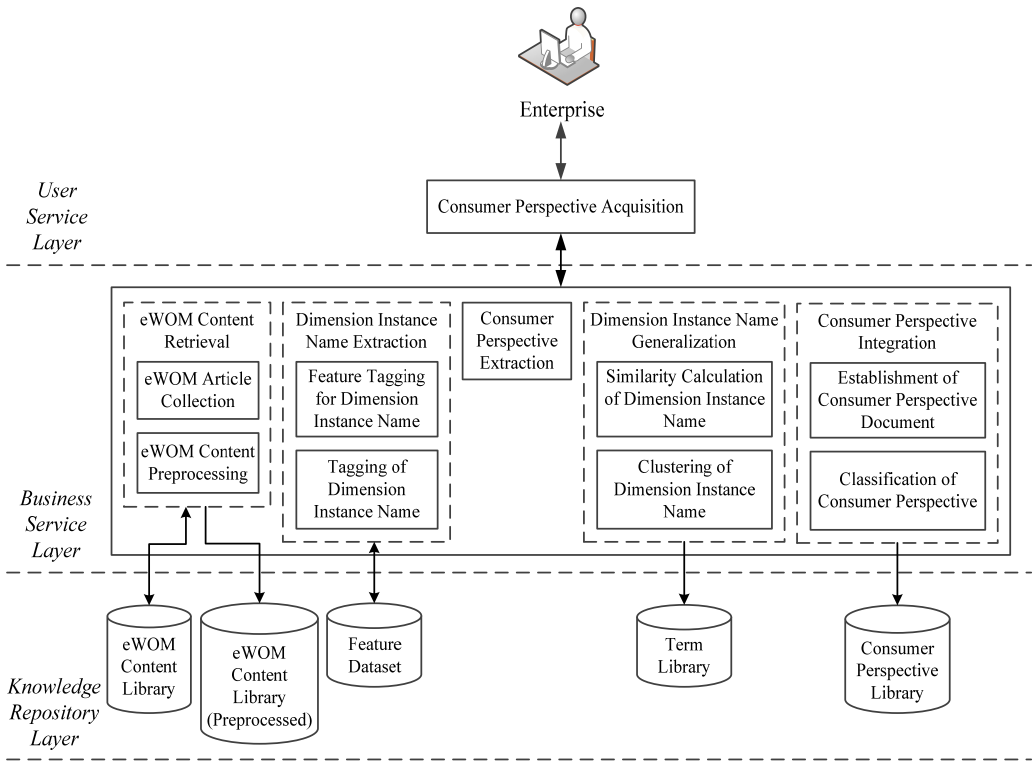

Figure 10: Framework of the Consumer Perspective Acquisition System

5.2 Implementation Results

This section uses the example of foodstuffs to explain the implementation results of the system of consumer perspective acquisition in Chinese eWOM. Figure 11 displays 481 Chinese eWOM articles retrieved from the website http://blog.yam.com/. Figure 12 shows 10 dimension instance names extracted from these articles by feature and CRF tagging. Based on the preprocessed Chinese eWOM content and the POS combination mode defined in Table 3, Fig. 13 illustrates the results of POS combination matching and consumer perspective extraction. Moreover, Fig. 14 shows the results of similarity calculation and clustering for the dimension instance name. Finally, Fig 15 depicts 29 consumer perspective classifications obtained from the Chinese eWOM document classification.

Figure 11: Results of Chinese eWOM Articles Collection and Content Preprocessing

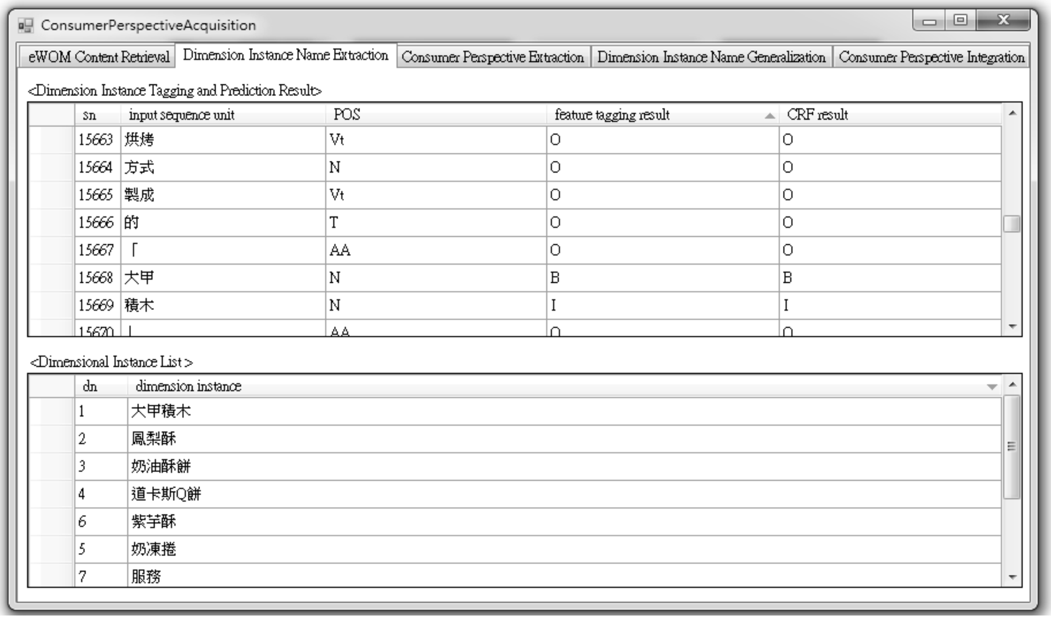

Figure 12: Results of Feature and CRF Tagging for Dimension Instance Name Extraction

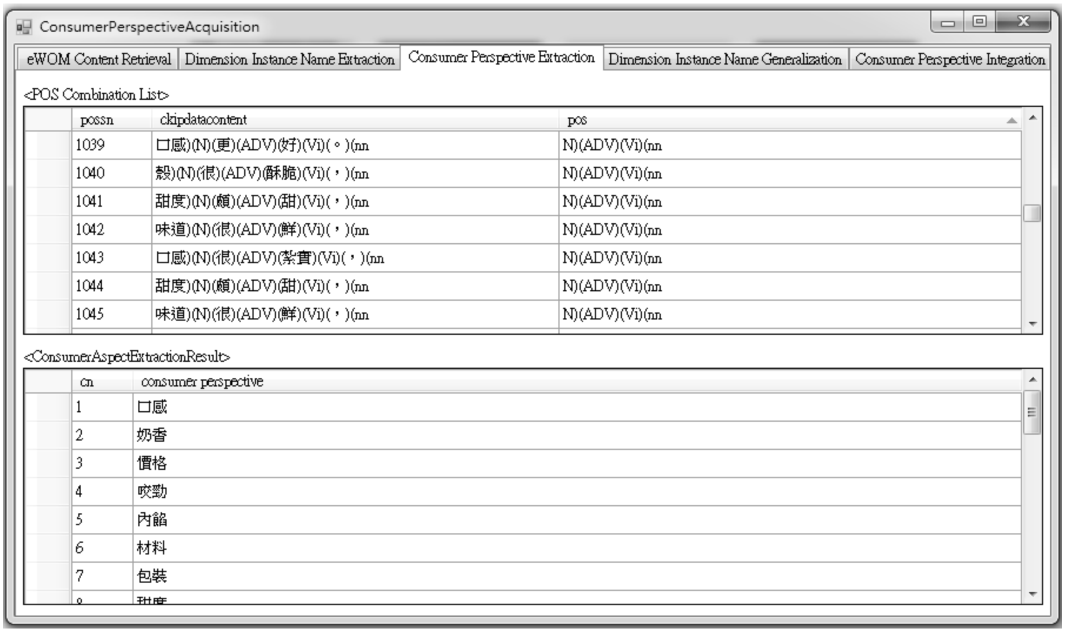

Figure 13: Results of POS Combination Matching and Consumer Perspective Extraction

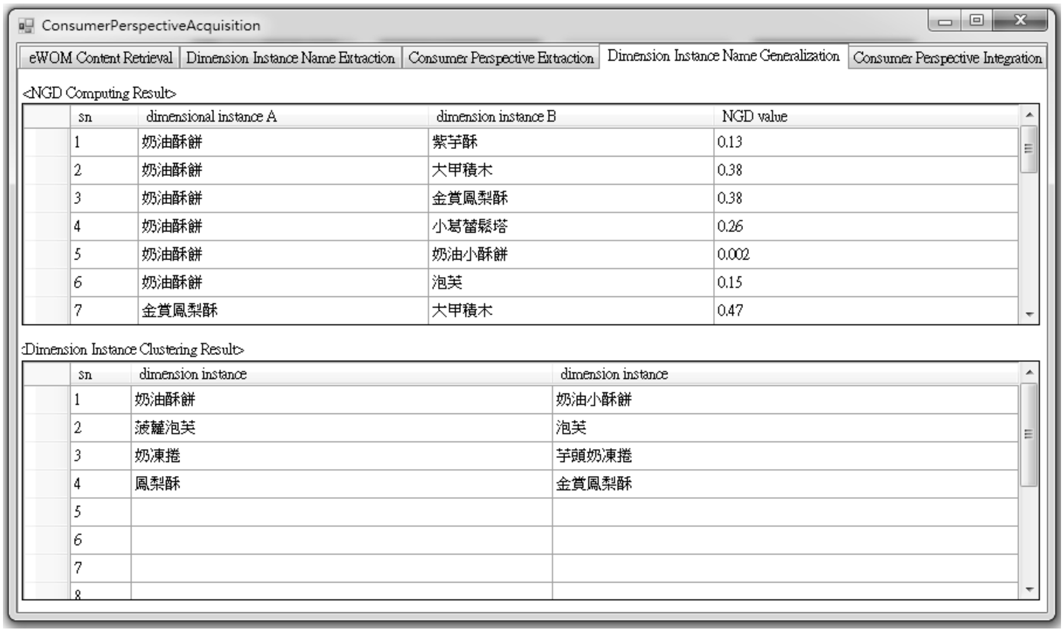

Figure 14: Results of Similarity Calculation and Clustering for Dimension Instance Name

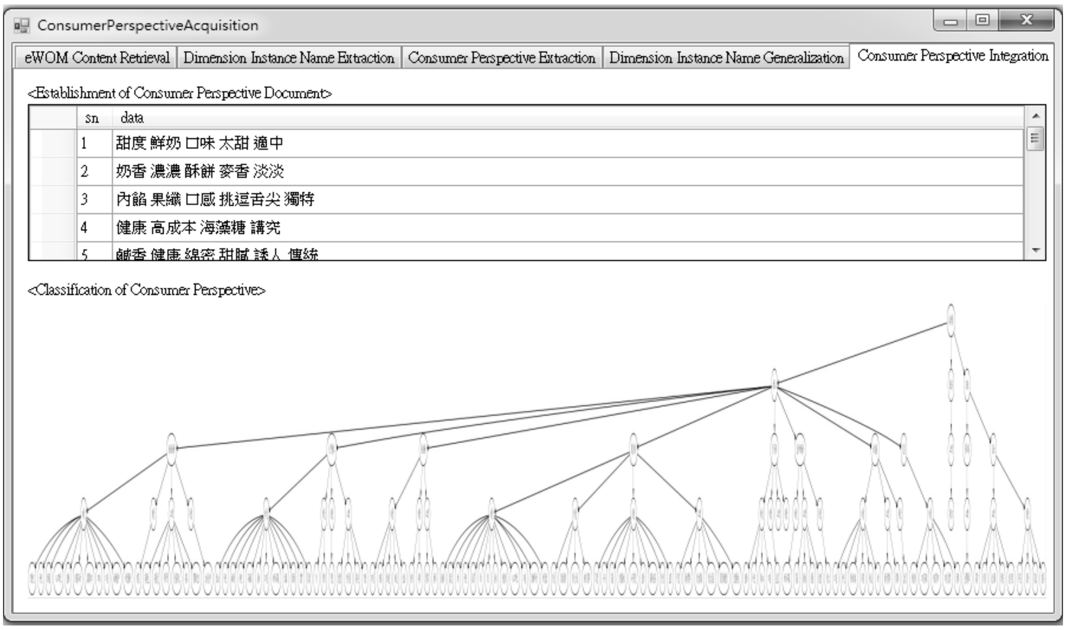

Figure 15: Results of Consumer Perspective Document Establishment and Consumer Perspective Classification

5.3 System Evaluation for Company Satisfaction

In measuring user satisfaction (Chou 2002; Herlocker et al. 2004; Manning et al. 2008; Pu et al. 2008), the Delphi method has become a widely used tool for acquiring a consensus-based opinion from a panel of experts. This study thus applied the Delphi method to examine firm satisfaction with the proposed system. In this case, 15 food companies were randomly chosen using simple random sampling.

Using the questionnaires shown in Fig. 16, the findings indicated that over eighty percent of companies satisfied the results for acquiring consumer perspectives in Chinese eWOM, as shown in Table 6.

1. The degree to which system function "consumer perspective acquisition" supports companies in acquiring consumer perspectives in eWOM.
   ☐Very Useful (5) ☐Useful (4) ☐No Comment (3) ☐Useless (2)
   ☐Very Useless (1)
2. The degree to which the system function "consumer perspective acquisition" assists companies in reducing search time of consumer perspective within eWOM.
   ☐Very Useful (5) ☐Useful (4) ☐No Comment (3) ☐Useless (2)
   ☐Very Useless (1)
3. User interfaces take less time than that by manual evaluation.
   ☐Very Useful (5) ☐Useful (4) ☐No Comment (3) ☐Useless (2)
   ☐Very Useless (1)
4. The user interface is easy to use.
   ☐Very Useful (5) ☐Useful (4) ☐No Comment (3) ☐Useless (2)
   ☐Very Useless (1)
5. The user interface useful for identification.
   ☐Very Useful (5) ☐Useful (4) ☐No Comment (3) ☐Useless (2)
   ☐Very Useless (1)

Figure 16: Questionnaire for Assessing Company Satisfaction

Table 6: Company Satisfactions Results

Question Item	Very Useful	Useful	No Comment	Useless	Very Useless	Total	Satisfaction
Q(1)	6	4	2	0	0	12	83.33%
Q(2)	9	3	0	0	0	12	100.00%
Q(3)	10	1	1	0	0	12	91.67%
Q(4)	6	3	3	0	0	12	75.00%
Q(5)	7	2	3	0	0	12	75.00%
Average	7.6	2.6	1.8	0.0	0.0	12.0	85.00%

6. Conclusions

This work develops a technology for acquiring consumer perspectives in Chinese eWOM through the design of a consumer perspective acquisition process and the development of acquisition techniques, and a consumer perspective acquisition system. The techniques for consumer perspective acquisition contain Chinese eWOM content retrieval, dimension instance name extraction, consumer perspective extraction, dimension instance name generalization, and consumer perspective integration. Finally, this study implements a system for consumer perspective acquisition based on the aforementioned techniques. The main results and contributions of this work are summarized as follows.

1. Consumer perspective acquisition process

   Based on the proposed consumer perspective framework, this study designs a consumer perspective acquisition process to help an enterprise automatically generalize and acquire consumer comments on the Internet and rapidly realize hidden consumer perspectives in Chinese eWOM contents for the important reference of internal environment improvement and service innovation of enterprises.
2. Consumer perspective acquisition techniques

   Based on the designed consumer perspective acquisition process, the techniques for consumer perspective acquisition are developed to facilitate the implementation of the consumer perspective acquisition system.
3. Consumer perspective acquisition system

   The system implemented in this study can instantaneously acquire the Chinese eWOM articles of consumers and effectively extract hidden consumer perspectives to assist enterprises in internal improvement and service innovation.

Acknowledgements

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract Nos. NSC100-2410-H-327-003-MY2 and NSC102-2410-H-327-028-MY3.

References

Archak, N., Ghose, A. and Ipeirotis, P.G. (2011), 'Deriving the pricing power of product features by mining consumer reviews', *Management Science*, Vol. 57, No. 8, pp. 1485-1509.

Chen, J., Huang, H., Tian, S. and Qu, Y. (2009), 'Feature selection for text classification with Naïve Bayes', *Expert Systems with Applications*, Vol. 36, No. 3, pp. 5432-5435.

Chou, C. (2002), 'Developing the e-Delphi system: a web-based forecasting tool for educational research', *British Journal of Educational Technology*, Vol. 33, No. 2, pp. 234-236.

Cilibrasi, R.L. and Vitanyi, P.M.B. (2007), 'The Google similarity distance', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19, No. 3, pp.370-383.

Guo, H., Zhu, H., Guo, Z., Zhang, X. and Su, Z. (2009), 'Product feature categorization with multilevel latent semantic association', *Proceedings of the Eighteenth Association for Computing Machinery Conference on Information and Knowledge Management (CIKM 2009)*, Hong Kong, China, November 02-06, pp. 1087-1096.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G. and Gremler, D.D. (2004), 'Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet?', J*ournal of Interactive Marketing*, Vol. 18, No. 1, pp. 38-52.

Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. (2004), 'Evaluating collaborative filtering recommender systems', *ACM Transactions on Information Systems*, Vol. 22, No. 1, pp. 5-53.

Hou, X., Ong, S.K., Nee, A.Y.C., Zhang, X.T. and Liu, W.J. (2011), 'GRAONTO: a graph-based approach for automatic construction of domain ontology', *Expert Systems with Applications*, Vol. 38, No. 9, pp. 11958-11975.

Knight, G.A. and Cavusgil, S.T. (2004), 'Innovation, organizational capabilities, and the born-global firm', *Journal of International Business Studies*, Vol. 35, No. 2, pp.124-141.

Lin, Y., Su, H.Y. and Chien, S. (2006), 'A knowledge-enabled procedure for customer relationship management', *Industrial Marketing Management*, Vol. 35, No. 4, pp. 446-456.

Manning, C.D., Raghavan, P. and Schutze, H. (2008), '*Introduction to information retrieval*', Cambridge University Press, Cambridge, UK.

Netzer, O., Feldman, R., Goldenberg, J. and Fresko, M. (2012), 'Mine your own business: market-structure surveillance through text mining', *Marketing Science*, Vol. 31, No. 3, pp. 521-543.

Ngai, E.W.T., Xiu, L. and Chau, D.C.K. (2009), 'Application of data mining techniques in customer relationship management: a literature review and classification', *Expert Systems with Applications*, Vol. 36, No. 2, pp.2592-2602.

Pan, Y. and Wang, Y. (2011), 'Mining product features and opinions based on pattern matching', *International Conference on Computer Science and Network Technology (ICCSNT)*, Vol. 3, pp. 1901-1905.

Pang, B. and Lee, L. (2004), 'A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts', *Proceedings of the Forty second Annual Meeting on Association for Computational Linguistics(ACL 2004)*, Barcelona, Spain, July 21–26, pp. 271.

Pu, P., Chen, L. and Kumar, P. (2008), 'Evaluating product search and recommender systems for E-commerce environments', *Electronic Commerce Research*, Vol. 8, No. 1, pp. 1-27.

Tsai, R.T. and Chou, C.H. (2011), 'Extracting dish names from Chinese blog reviews using suffix arrays and a multi-modal CRF model', *Proceedings of the First International Workshop on Entity-Oriented Search (EOS 2011)*, Beijing, China, July 28, pp.7-13.

Xia, R., Zong, C. and Li, S. (2011), 'Ensemble of feature sets and classification algorithms for sentiment classification', *Information Sciences*, Vol. 181, No. 6, pp. 1138-1152.

Zhai, Z., Liu, B., Xu, H. and Jia, P. (2011), 'Clustering product features for opinion mining', *Proceedings of the Forth ACM International Conference on Web Search and Data Mining (WSDM 2011)*, Kowloon, Hong Kong, February 09 - 12 pp. 347-354.

Zhang, C., Zeng, D., Li, J., Wang, F.Y. and Zuo, W. (2009), 'Sentiment analysis of Chinese documents: from sentence to document level', *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 12, pp. 2474-2487.

Zhao, J. and Liu, F. (2008), 'Product named entity recognition in Chinese text', *Language Resources and Evaluation*, Vol. 42, No. 2, pp.197-217.

Zheng, Y., Cheng, X.C. and Chen, K. (2008), 'Filtering noise in web pages based on parsing tree', *The Journal of China Universities of Posts and Telecommunications*, Vol. 15, pp. 46-50.