Proceedings of the TMCE 2010, April 12–16, 2010, Ancona, Italy, Edited by I. Horváth, F. Mandorli and Z. Rusák

GUIDING CONCEPT GENERATION BASED ON ONTOLOGY FOR CUSTOMER PREFERENCE MODELING

Dongxing Cao
School of Mechanical Engineering
Hebei University of Technology, Tianjin, PR China
e-mail: dongxingcao@gmail.com

Karthik Ramani
School of Mechanical Engineering
Purdue University, West Lafayette, Indiana, USA
e-mail: ramani@purdue.edu

Zhanjun Li
EaglePicher Medical Power, Plano, Texas, USA
e-mail: jameslee815@yahoo.com

ABSTRACT

Customers often show different preferences relative to the same products, such as function, shape, color, costs, etc. They will surely affect product market activities further. In this paper, a model of preference elicitation from customers is proposed to reduce the gap between the systematic representation of preferences and customers' actual preferences. Firstly, the attribute of customer preferences are classified by developing a kind of preference taxonomy which is analyzed to build customer preference ontological concepts and relationships. Secondly, the documents or catalogs of design requirements, perhaps containing some textual description and geometry data, are normalized by using ontology-based semantic expressions. Some semantic rules are developed to describe low-level features of customer preferences to build an ontological knowledge base towards customer preferences. Thirdly, customer needs are mapped to customer preference ontologies for driving high-level concept generations. A customer preference modeling framework is developed to construct a vector space model to measure the distance between sets of preferences. Finally, an empirical study is surveyed, and five different customer groups are queried towards the cell phone preferences. The query results are analyzed to represent validity of concept generation from customer preferences.

KEYWORDS

Ontology, Concepts, Customer preference, Semantic expression, Information extraction

1. INTRODUCTION

In today's rapidly changing market, an enterprise strategy is influenced by customer preferences through their consumption of products [1]. The idea in the mind of the customers is always flexible, and they do not know what they want exactly until they see it [2]. Their needs do not have a fixed benchmark. In general, customer preferences can be described by using words/phrases. In these cases, one can imagine a virtual product world or preference model that can be created by the designers' to elicit feedback from the customers [3]. Although the representation of customer preference is subjective, it can affect the customer preference itself. Furthermore they can produce a certain impact for product market activity. On the other hand, product structure, shape, color and so on, are directly perceived through the senses, and they belong to high-level features. However, the existing approach, such as in [4-6], represent preferences depended on some statistical measures which often are based on low-level features.

Generally speaking, a customer's preferences for a product can be viewed as a reflection of his or her inner world. They are related to customers' emotion and have a considerable influence on our purchase decisions [7]. Different culture differences maybe produce an important impact on customers' preference decisions [8]. The challenge makes a compelling case for us to focus on "how customers do it" rather than "what customers do."[9]. Also the psychographic activities still affect customers' interest in specific product preferences [10]. Owing to the varieties of preferences from different customers, for example, some customers may only like MP3 function of the cell phone, may be ambivalent about its communication function, while others may like it to have a camera function with high quantity. It is therefore not easy to exactly define a common customer preference.

Ontology can be viewed as an effective means of representing concepts. Ontology-based information extraction has been successfully applied to semantic indexing [11-12]. Traditional methods, such as quality function deployment (QFD) and the house of quality, rely upon the careful solicitation of representative customers from market, and guide designers for concept generation [13]. They mainly focus on special groups, surveys, and friendly-environmental studies to assess customer needs. However, few studies can be found that concentrate on customer preference modeling through using ontology knowledge indexing. Therefore, a practical approach that focuses on customer preference modeling is a challenge at the preliminary stages of product development. We will attempt to build a new model that addresses this issue by including ontology-based information extraction. Our approach also shares some similarity to the literature in adaptive text extraction [14]. The rest of this paper is organized as follows. In Section 2, we review related work. In Section 3, we introduce the research approach and ontology modeling for customer preferences in this context. The process of preference semantic extraction is described in Section 4. In Section 5, we give an evaluation analysis of customer preferences and show how our ontology-based model compares with traditional keyword-based search technique. An empirical study is presented in Section 6. Finally, in Section 7, we present the conclusions and discussions.

2. RELATED WORK

2.1. Low level features and high concepts

Customer satisfaction has been widely studied [1,15] and some methods, such as SCSB, ACSI and NCSB, have been successfully appiled to enterprises. In addition, laddering method can be used to reveal the deeper meaning of products or services for a customer in order to gain more insight about the perception of the market [16-17]. In general, the customer often shows different preferences relative to the same type products, such as function, shape, color, and costs [18]. The ideas in the mind of the customer are represented by higher level concepts. Also, the customers often do not know what they want and the concepts may be influenced by their browsing and experimenting with their options from new product concepts. Customer preference representations that can be described by using textual descriptions and shapes representations can be said to be lower level features. The activity of concept generation by the designers strives to reduce this gap between unconstrained/higher level customer beliefs and constrained/lower level representations as shown in Fig.1 [2]. A flexible representation which the customers may interact with, to indicate their preferences, can be modeled using new representations that combine shape and ontologies. The results of the customer interaction by search, relevance feedback and modifications can provide valuable input to the design activity and even group the customers into different categories or classes for further concept generation. Constrained/ lower level features are highly structured, such as final reports, transaction data, geometric shapes, CAD drawing, which we depict at bottom of design spectrum. At the same time, unconstrained/ high level concepts are unstructured, fragmentary documents, such as interviews, design logbooks, which they situate on the top of design spectrum [19]. The middle of design spectrum remedies defects at both ends, and essentially shortens the critical gap between low level features and high-level concepts shown in Fig.1.

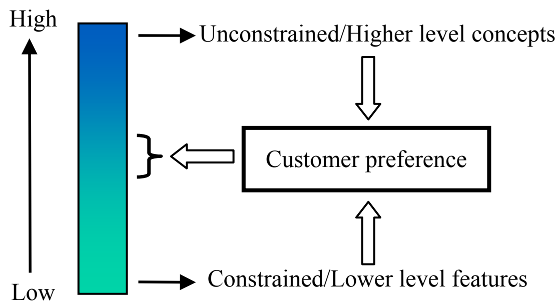

Figure 1 Different stage spectrum of concept generation

2.2. Customer preference for shape ontologies

Low-level features, such as shape, size, number of entities, are easily identified by humans. The shape can be specific from geometric point of view or it can be a perceived feature. They have considerable influence on customer purchase decisions [20]. In general, the shape can be easily represented in the hierarchy. This hierarchical representation describes the main shape categories that can be identified [21]. Also it can be indexed in the shape repository. The varied shape information is described by the knowledge represented in the shape repository. The domain knowledge is needed to describe shape information. A specific shape type is closely related with shape information. Some important properties of shape description are also described by using an associated textual description. The Shape type hierarchy captures information regarding the type of shapes that can be processed by a shape semantic description [22].

General shape ontology captures shape related metadata and constitutes an ontology-driven evolution of the metadata in the shape repository. A high level hierarchical relationship of the ontology describes the main concepts of shape ontology, in which includes shape program, shape repository, shape concepts and so on [20]. Shape program contains program rules and semantic structure. They can be extracted from the text information. Shape repository stores shape semantic information and structural information. File information can be used to describe shape concepts which capture some information regarding a product or shape associated with the various shape models stored in the repository. The concepts of shape classification can help us creating groups of shapes that share some common features [23]. Usually, several elements of a shape that differ in some key feature are placed in a group so that they can be examined together. File information can store information related to a shape, such as the size, material, color, etc. The shape representation is the central concept in the ontology and encapsulates information that is inherent to the shape model itself. It also constitutes a fundamental concept which is extended by concepts defined in the domain ontologies.

2.3. Customer preference for semantic ontologies

An important task is how to extract preference terms from all related design information and how to manage them efficiently. It is a challenging issue to discover, extract, and manage preference effectively in preliminary design stage [24]. A fundamental deficiency of current information retrieval methods is precision problems in which the meaning of the indexed words is not exactly the same as what the customer seeks. Sometimes, according to different contexts, different needs or linguistic habits, customers can express and describe the same preferences by using different terms and phrases. In fact, individual words often provide unreliable evidence about the conceptual topic or meaning of a document. There are usually many ways to express a given preference concept. For example, a simple term can consist of a concept and several terms can be clustered into a concept. Thus the literal terms in customer queries may not directly match corresponding terms of a relevant document. In addition, most words have multiple meanings. Therefore, terms in a user's query will literally match terms in documents that are not of interest to customer. To flesh out preferences, we may consider any document to consist of the scattered information that might come from customer market activities. Many semantic concept similarities and statistical word measures have been researched [25], and one well-known application tool is WordNet [26] which is an online lexical reference system used in semantic analysis and text information extraction across interdisciplinary domains.

3. THE PROPOSED APPROACH

3.1. Overview

Designers always like to extract some useful information from documents in order to carry out design tasks. As the input contents from customers are disorderly and unstructured documents, sometimes most of them are qualitatively described, such as user requirements, survey reports, transaction data and customer dialogs. We need analysis and transform them into formal documents. The transformation operation is to combine qualitative with quantitative aspects. Qualitative transformation characterizes some design information with disordered arrangements in an abstract manner. They allow designers to make a revision to improve concept description. On the contrary, quantitative transformation provides a canonical document description [27]. They allow designers to easily understand, evaluate and reuse previous design information.

In general, most of the existing information is disorderly and unsystematic. They need designers' analysis and abstraction in order to gain utility. First of all, the original materials of preference from customers, such as survey reports, transaction data, and customer dialogs should be filtered. A normal design information text or document is used to extract preference information after transformation [24]. Automatically extracting semantics from the normalized document requires recognizing the domain knowledge as well as the semantic structures of the text. Linguistic knowledge and semantic rules are needed to fulfill preference semantic extraction.

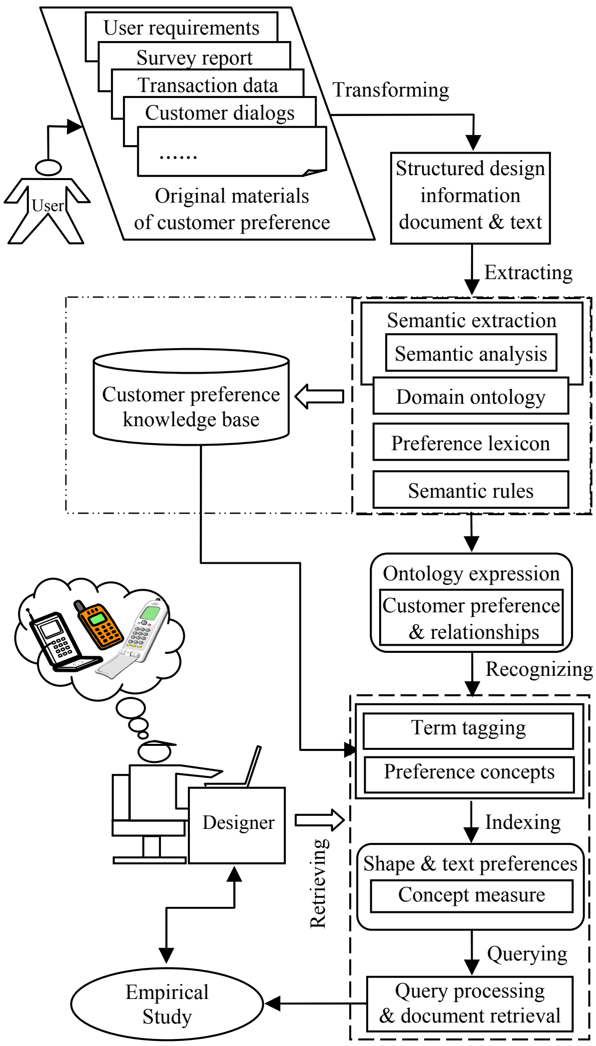

Figure 2 Infrastructure of customer preference modeling

To accurately represent the preference semantics in design information texts and documents, we need to extract as much relevant information from the document as possible. A preference knowledge base can be constructed by analyzing and collecting the varied product preference terms. It includes preference lexicon, domain ontology rules, semantic rules, and so on. It can be used to evaluate customer preferences of different products. Fig.2 gives the framework of customer preference modeling. It represents an ontology-based design document analysis and information extraction and indexing.

Preference terms and phrases can be extracted from the structured information texts, such as the noun phrases, verb phrases, adverb phrases, prepositional phases. The domain thesauri or taxonomies are built to conveniently capture preference concepts from the index documents.

The proposed approach is organized as follows. The taxonomy of preference is firstly discussed. Secondly, ontology expression and preference semantics extraction process are described, and the process of semantic extraction is based on a shallow NLP algorithm and the domain ontology. The preference ontology is automatically acquired after the extraction process. The extraction algorithm and the preference metrics are discussed. They are used for preference ontology modeling. Finally, the empirical studies for design preference extraction are introduced in this paper.

3.2. Ontology modelling for customer preference

Ontology is a formal, explicit specification of a shared conceptualization [28]. The formal means that it can be communicated across people and computer, and explicit means the concepts and relations are fully defined. Conceptualization refers to an intended model of the phenomenon identified by its concepts and relation. Therefore, ontology defines a set of formal terms we call concepts. The hierarchical correlations are described among concepts [29]. On the other hand, the taxonomy is only reviewed as concept classification in the hierarchy. It simply links concepts by domain-independent relationships.

Ontology concepts also have multiple parents and form the complex relations of inheritances. They share the genetic attributes. At present, considering ontology modeling for customer preferences, there are two main problems: one is the extraction of the semantic concepts from preference words and the other is the document indexing from users' requirements. As for the first problem, the key issue is to identify appropriate preference concepts that describe and identify documents. On the other hand, the preference terms can be indexed from customer documents. The precision problem of extraction is about semantic expression employed in customer requests. A hierarchical analysis process has been used to aggregate preferences in a group using a pair-wise approach [30]. However, a significant assumption for each of the methods is that the decision maker in the group is assumed to be equally important. That is, their information is handled equally without any preference given to one group member over another.

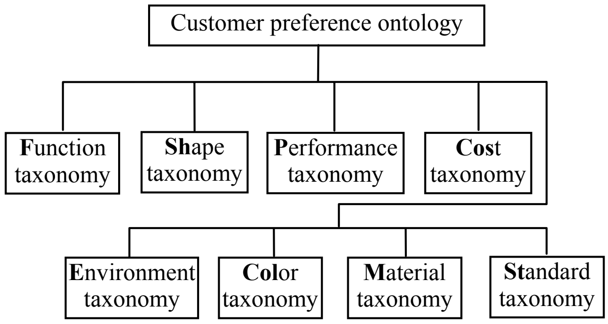

Figure 3 The taxonomy of preference ontology

Ontology modeling provides an effective approach to indexing terms/concepts which can be used to match with customer requests. However, the taxonomy acquisition of customer preferences of different products is of a certain subjective behavior. Their generation is either by brainstorming or by interviewing or dialogs with customers. Preference ontological acquisition and generation use the same methodology. Fig.3 presents the taxonomy of customer preference ontology, which comes from cell phone handbooks or knowledge resources. For example, cell phone handbooks often classify engineering components which can be clustered into an ontology model as concepts and taxonomy in the hierarchy. Each component is described in detail, including its attributes such as material, physical, geometric and functional properties which can easily be identified and mapped to ontologies as well as corresponding relationships.

Customer preference ontology includes concepts, taxonomies and relationships. Each taxonomical concept is acquired from various engineering knowledge resources. We can adopt terms or phrases to describe the concepts of the taxonomy as well as their relationships with other concepts. For example, multimedia belongs to Function taxonomy of cell phone. We can represent as F-MULTIMEDIA, where the prefix of each concept represents its taxonomy which this concept belongs to. Therefore, the relationships are structured between concepts across taxonomies. For example, has_feature (COL-SIVER, SH-KITTY-PHONE), in which COL-SIVER stands for a color concept in the color taxonomy, SH-KITTY-PHONE represents a shape concept in the shape taxonomy [12, 24]. Table 1 lists customer preference ontological concepts and acquisition

Table1 Customer preference ontological concepts and acquisition resources

Taxonomy	No. of concepts	Example of concepts	Acquisition resources or reasons
Function	52	Voice, text, multimedia, memory, picture, MP3, internet, digit camera, bluetooth, 3G, GPS, TV, etc.	Statistic current cell phone functions based on Website.
Environment	18	Radiation protection, man-machine friendly, recycle, health risk, etc.	Cell phone and human act environment each other.
Shape	20	Flip phone, kitty phone, hand-writing & PDA, lighterphone, moustache phone, Miniphone, wide screen phone, etc.	http://www.halfbakery.com
Performance	31	Good performance, signal strength, coverage area, large speaker, long talk time, clear call, etc.	Investigating different cell phone performances in light of customers.
Cost	12	Top grade phone, middle price phone, low end phone, etc.	Separating them according to their price difference in 50$
Color	12	Black, white, green, red, yellow, silver, oyster color, etc.	According to existing cell phone on the market.
Material	28	Metal, polycarbonate, plastic, steelless, synthese materials, etc.	Manufacturing materials used as main parts of cell phones.
Standard	12	Communication protocols, power, votige, Wi-Fi, port, AM/FM, etc.	Cell phone use standards in different areas and countries.
Brand	31	Blackberry, Motorola, Samsung, Nokia, Sanyo, HTC, LG, etc.	Different brands for customer uses on the market.
Accessory	24	Headset, lanyard, leather portfolio, clip, charger, battery, software, etc.	http://www.amazon.com

resources of a cell phone, including the number of concepts. The classification of their relationships is represented in Table 2.

At present, we have collected 10 taxonomies, 240 concepts and 7 types of relationships in customer preference ontology. The standard worksheets are developed to easily acquire preference ontology and lexicon. At the same time, they can automatically upload the required data into the Protégé editor [31]. Therefore, the proposed customer preference ontological concepts can be also presented by using Protégé 3.1, which is one of the most widely used ontology editor. Protégé provides a visual tool for preference ontology editing, including concept, taxonomy, and relationship building as well as preference ontology visualization.

Table 2 Classification of the relationship

Relationship	Concept	Definition
is_a	F-VOICE/ F-MP3	Parent-son relation
has_part	E-HEALTH RISK/ E-CYCLE	Part to whole relation
has_function	F-VOICE/ P-LONG TALK	Refer to the connection between two concepts
use_material	COS-LOW END PHONE/ M-METAL	The type of materials
has_property	SH-FLIP PHONE/ M-STEELLESS	Physical attribute/ geometric attribute
has_feature	COL-SIVER/ SH-KITTY PHONE	Geometric shape
has_standard	F-3G/ ST-NETWORK	Domain specific standards

4. PREFERENCE SEMANTIC EXTRACTION

4.1. Preference semantic representation

Semantic ambiguity often occurs in the design queries when customers cannot know the exact expressions or the related issues they want to pursue though they may have some contextual clues, such as the functional preference of the design and other interacting parts of the querying product. Preference lexicon is a better way to evaluate customer preferences. Lexical terms are the natural language phrases of the corresponding concept. They are used to map the concepts with words of texts and queries and explicitly represent the vocabularies of different ontology concepts. Therefore, words morph, abbreviation, acronyms, and synonyms of the word/phrase are also lexical terms and share the same concept with the original lexical term. Also some noun phrases, verb phrases, adverb phrases, prepositional phrases can be extracted as preference terms. The morphs of original lexical term can be easily and automatically obtained by WordNet (http://wordnet.princeton.edu/)[26], whereas other terms are acquired manually because WordNet is a general lexical resource but not a specific preference lexicon. We aim to extract implicit customer preferences from disorderly and unsystematic original materials, such as user requirements, survey report, transaction data, etc. As the existing case studies mostly are special products, the extracted texts have certain limitations. If a preference lexicon is built, it can be extended to improve preference evaluation.

However, it is difficult to model and extract the information of implicit customer preferences from design texts which embed into natural language documents. In order to identify linguistic forms of customer preference, we should build a preference lexicon to support the computer indexing. Logically speaking, such preference information is implicit within engineering design texts, but can be difficult to extract from unstructured information. The challenge is in linguistically modeling these preferences and mapping them into a ontology concept suitable for supporting design decision-making. We identify linguistic forms of preference, produce a specific preference lexicon, develop

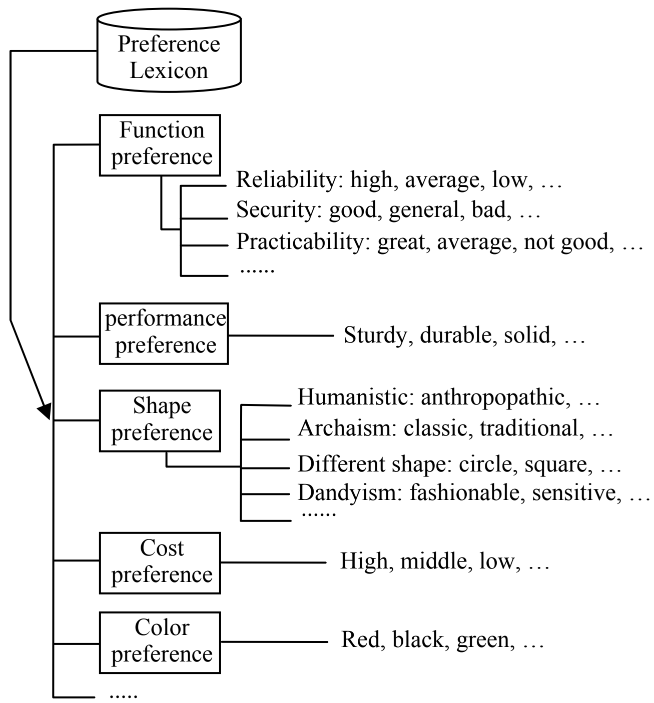

Figure 4 The lexicon of customer preferences

customer preference ontology concepts, and generate design alternatives. According to customer requirements, some documents are needed to analyze and extract. Preference lexicon can describe what the customers want. Perhaps the customers show different preferences for a special product. Fig. 4 represents a common preference lexicon for cell phone, where describes cell phone function, performance, shape, cost, color, and so on, in which each can be decomposed in the hierarchy.

Our preference lexicon includes the specific thesauri that can be used for design information retrieval to aid in the search for design information. Customer original materials are first input and the validity check is carried out. If there is any conflict among these materials, they will be rejected. Then according with the demands, the redundant check will be implemented. Finally all satisfied synonyms will be put into preference thesauri [19]. In addition, some original materials can be transformed or edited and directly put into preference thesauri. When indexing across contexts, the thesaurus can improve performance representation.

4.2. Text information extraction

We assume that input design information is expressed in plain English. If input is transaction data, it needs to be quantitatively changed into identified texts. They all need to transform into structured text information. Tokenization is carried out from the text of the customer request after stemming and removing stop words. According to preference lexicon, customer preference words are tagged to mark their position. Preference terms and phrases are recognized on the basis of indexing preference domain knowledge base. Using a list of synonyms, these tokens are associated with concepts in the ontology through depth first search (DPF) or breadth first search (BFS) [24, 32]. Therefore, after preference semantic extraction is embedded in the customer requests, the concepts are generated by matching terms and phrases in the ontology. The algorithm operations (as shown in Fig. 5) are described below.

• Stemming stop words and tokenizing

Some auxiliary verbs and articles are removed from the phrases. The tokens/words and punctuation symbols are marked by analyzing input texts.

• POS tagging

Each word is first inquired from preference lexicon and marked with its most likely POS tags as defined in the preference lexicon. The combination operation of automatic POS assignment and manual correction is carried out to improve the speed and accuracy of the mapping process. If the word does not have a match in the lexicon, then the word is assigned an unknown tag. After manual correction any incorrect tags will be removed [32].

• Recognizing terms and phrases

The purpose of recognizing a concept is to select the most appropriate terms or phrases in the domain ontology. This stage can be divided into two steps.

(1) Concept matching: Assigning the tagged terms/ phrases to the concepts it refers to. Words that match with a preference lexicon term will be assigned the pertinent ontology concept. Note that multiple concepts may be assigned to a single word or a series of words/phrases because different concepts may have the same lexicon term.

(2) Concept disambiguation: A word or term which matches with multiple concepts causes ambiguities. This ambiguity exists in polysemy and ellipsis semantic structures [14]. It can be disambiguated by referring to the contexts of the term/phrase meaning. The context of a term refers to the concepts to which its adjacent words/phrases are tagged.

• Joining relationships

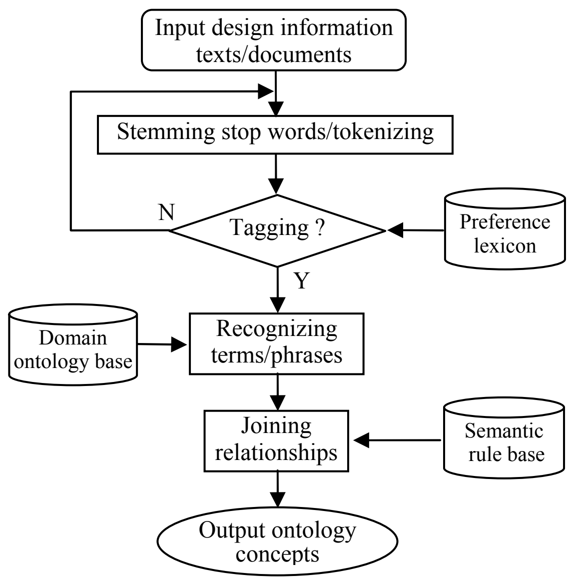

Figure 5 The process of customer preference extraction

The relationship between two concepts is joined together by a certain design hypernymy. The joining phase scans the sentences iteratively to generate relationships of two concept instances according to the semantic rules. Both concepts maybe exist in meronymy, hyponymy, causality, etc. These lexicon relationships are being used to include: has_part, is_a, has_property, etc. The similarity degree of different concepts should be also considered. The lexical relationships among the key words will be built and the semantic analysis will be employed to extract the information as described in the next section.

Preference ontology concepts, semantic structural expression and analysis are acquired from customer text information. A more accurate language model for the elicitation of preference expressions will be further developed in the future.

4.3. Preference concept disambiguation

In the process of customer querying and ontological indexing, semantic ambiguities often result in a lower retrieval precision, and often an even erroroneous retrieval. Theoretically speaking, three ambiguities may appear in text indexing as follows.

1) Polysemy: a term or phrase perhaps matches several concepts in order to result in semantic ambiguities.
2) Accuracy of term description: some concepts can be expressed by using different terms or phrases or synonyms, but they are of a little difference in semantics.
3) Ellipsis and acronym: part structures/letters of a sentence or a word are omitted to lead to semantic error or misapprehension.

These ambiguities are direct reasons that result in lower concept retrieval precision. For example, If customers like the price is about \$80 for a cell phone, in addition volume $80\times40\times10\text{cm}^3$, the two numbers “80” often appear ambiguous. In preference ontology concepts, we have divided customer preferences into different classifications (see 4.2 section). By marking different taxonomical signs during tagging terms, such as, COS-MIDDLE PRICE EIGHTY and SH-SIZE EIGHTY, we can distinguish from them. A detailed algorithm of concept disambiguation is described in 5.2 Section.

5. CUSTOMER PREFERENCE EVALUATION

5.1. Vector space model

In the traditional vector space model [33], a vector is used to represent each item or document. Each element of the vector includes a certain keywords associated with the given document. The value assigned to that element reflects the importance of the term in representing the semantics of the document. A database containing a total of documents described by terms is represented as a term-by-document matrix [34]. The rows of matrix are called the document vectors, and the columns of matrix are the term vectors. Thus, the matrix element is the weighted frequency at which term occurs in document. Therefore, a corpus matrix of document-terms is built on the basis of customer preference taxonomy and attribute, in which the column stands for the terms that appear in the documents as shown in Fig.6. The rows mean document descriptions from different cell phone brands. The matrix values a_{ij} are weights that represent the importance of terms in documents. One general approach is term-frequency (*tf*) and inverse document frequency (*idf*) weighting in which the calculation of weights is rooted in empirical studies that show the relevance of a term of taxonomy is related to the frequency with which it appears in a document.

$$\begin{array}{rl}
 & \overbrace{\begin{array}{cccccc}\text{F-terms} & \text{SH-terms} & \cdot & \text{P-terms} & \cdot & \text{M-terms}\end{array}}^{\text{Terms/concepts}} \\
\text{Documents}\left\{\begin{array}{l}\text{“Blackberry Bold}^{\text{TM}}\text{”} \\ \text{“Motorola Tundra”} \\ \cdot \\ \text{“MOTO W233”} \\ \cdot \\ \text{“Nokia 6010”}\end{array}\right. & \begin{bmatrix} a_{11} & a_{12} & \cdot & a_{1j} & \cdot & a_{1m} \\ a_{21} & a_{22} & \cdot & a_{2j} & \cdot & a_{2m} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{i1} & a_{i2} & \cdot & a_{ij} & \cdot & a_{im} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & a_{nj} & \cdot & a_{nm} \end{bmatrix}
\end{array}$$

Figure 6 Vector space model of corpus matrix

Therefore, we calculate the weight value w_{ij} of characteristic item as follows [33, 35].

$$\begin{aligned} w_{ij} &= tf_{ij}\times idf_j \\ &= tf_{ij}\times(\log_2(N/n_j)+1) \end{aligned} \quad (1)$$

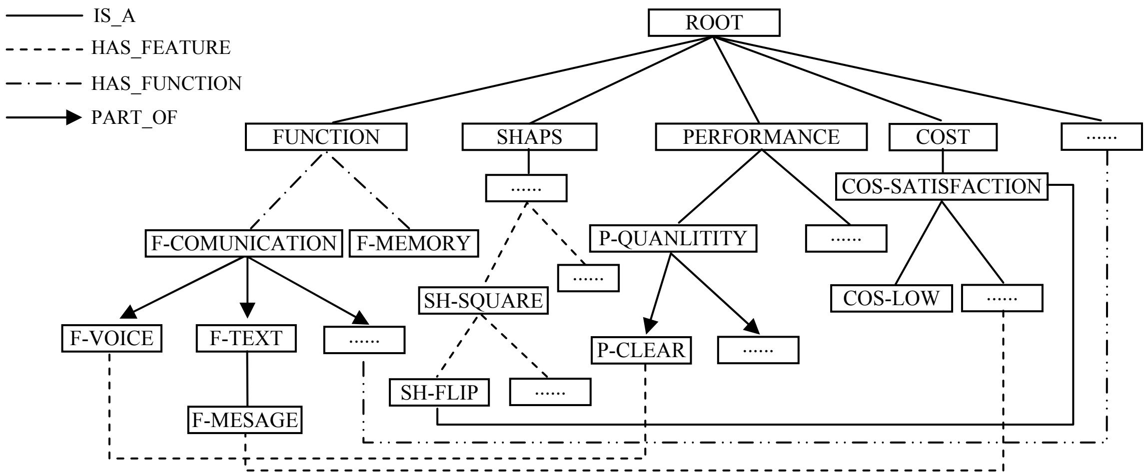

Figure 7 Customer preference ontological relationships of cell phone

Where tf_{ij} is the frequency of term t_j in document d_i, and N is the number of documents, n_j is the number of documents that involved the term t_j. From the formula (1), we can obtain that the value of w_{ij} increases with tf_{ij} and decreases with n_j. The distance between two document vectors is represented by similarity. The similarity between document d_i andd_j is defined as the cosine of the angle between two vectors below.

$$Sim(d_i, d_j) = \frac{\sum_{k=1}^{m} w_{ik} \times w_{jk}}{\sqrt{(\sum_{k=1}^{m} w_{ik}^2)(\sum_{k=1}^{m} w_{jk}^2)}} \tag{2}$$

When carrying out query operation, above model d_i could be viewed as query document from customers. By measuring customer' query key words and different brand cell phone document similarity, we can realize document retrieval.

5.2. Concept similarity measure

Lexical ambiguity can be distinguished from concept similarity measure. We can measure the distance between two concepts of the phrase/keyword clusters corresponding to special product attributes. In our model the customer preference ontology will conveniently execute indexing, that is, the preference ontology easily provides index terms/concepts which can be used to match with customer querying. For example, given a customer query: "want a camera and multimedia cell phone with a red flip phone", the keywords "camera and multimedia" is function taxonomy, stands for "F-CAMERA" and "F-MULTIMEDIA"; keywords "red" and "flip" belong to color and shape taxonomy, stand for "COL-RED" and "S-FLIP". By putting a taxonomical label in front of keywords, we can easily index preference terms from lexicon. When concepts are correlated, the associated concepts will be endowed greater weight based on their minimal distance from each other in the ontology and their own matching scores based on the number of words they match. In general, an ambiguous concept when related to other concepts will have a higher score, and retain a greater probability than non-correlated ambiguous concepts.

In order to calculate the similarity between two concepts, we need to build interrelationships of ontology concepts. Here, we represent our ontology as a directed acyclic graph (DAG)[12]. Each node in the DAG expresses a concept which it includes a label name and a synonym list. The synonym list of a concept contains a set of keywords through which the concept can be matched with customer queries. Fig. 7 represents a small portion of preference ontological relationships of cell phone. Each line type represents different ontology concept interrelationships [33]. Suppose matched concepts of query keyword K_i: C_1, C_2, …, C_i, …, C_n; and each selected concept (C_i) contains a score based on the number of lexical terms (T_{i1}, T_{i2}, …, T_{ij}, …, T_{im}) from the list of synonyms that have been matched with the customer queries. Keywords in customer queries are sought based on DFS or BFS which match each keyword on the lexical terms of a concept. The calculation of the score is based on the number of T_{ij} matched keywords of shown as follows.

$$T\text{score}_{ij} = \frac{\#\text{ of keywords in query } T_{ij} \text{ matches with}}{\#\text{ of keywords in } T_{ij}} \quad (3)$$

The shortest distance or least number of arcs between two matched concepts in preference ontology is

defined the concept distance (*CD*) as follows.

$$CD_{ij} = 1 + Min\ (\text{Number of arcs } (C_i, C_j)) \quad (4)$$

Note that if concepts are at same level and no path exists, the concept distance is infinite (see Fig.7). For example, the concept distance between "FUNCTION" and "SHAPE" is infinite, and "F-COMMUNICATION" and "F-VOICE" are linked by "Part_of" relation, their distance is 1. Similarly, the concept distance between "F-COMMUNICATION" and "F-MESAGE" is 2. And the concept distance between "F-VOICE" and "P-CLEAR" is 1.

The semantic ambiguity between two concepts can be also distinguished by calculating CD. Given customer queries "push button to sent text message with red kitty phone," the query keywords are first processed by Equation (3). All of the matched concepts are calculated and added to a candidate list. Query expansion is executed on the matched concepts. Here, assuming all these concept scores are F-PUSH (1.0), F-BUTTON (1.0), F-TEXT (1.0), F-MESSAGE (1.0), CLO-RED (0.5), P-KITTY PHONE (1.0), and breadth first search (BFS) is employed to obtain the scores of concepts. All of the selected concepts are added into a list in descending order according to their scores. They are further expanded as some relevant concepts into a list. Also we can use Equation (4) to calculate the semantic distances between two concepts. They are added into the selected concept list in descending order. Therefore, according to the above calculations, we can discriminate between ambiguous concepts and provide the fittest concepts for customer queries.

5.3. Evaluation analysis

Evaluation analysis uses the collected preference catalogs as the benchmark and compares the retrieval performance of the ontology-based search and keyword-based search. The test was executed by the five different groups. Each has different domain knowledge about cell phone preferences and part shopping experiences. The objective is to briefly describe what kind of cell phones they like. Each of them needs to provide at least 3-5 queries that they hope cell phones have the kind of function, performance, shape, and so on. These constitute design specifications of new cell phones. The effectiveness of retrieval is usually measured by two equations as follows.

$$Re = \frac{\#\text{ Of relevant concepts that are retrieved}}{\#\text{of relevant concepts}} \quad (5)$$

$$Pr = \frac{\#\text{ Of relevant concepts that are retrieved}}{\#\text{ Of retrieved concepts}} \quad (6)$$

Where *Re* stands for Recall, *Pr* means Precision. Two metrics are usually used to describe the quality of preference retrieval. Recall is the proportion of relevant concepts retrieved by the system and precision is the proportion of retrieved concepts that are relevant. Precision is an accuracy measure, while recall is a measure of how good is the information retrieved. Generally speaking, it is necessary for the customer preferences to evaluate recall versus precision to determine which overall strategies are most important [12].

6. EMPIRICAL STUDY

We design a virtual experiment platform to simulate customer preference ontology with respect to concept generation. Three brand cell phones, Blackberry, Motorola and Nokia, are selected to implement this empirical study. Each brand includes ten model series of cell phones as follows.

Blackberry: Bold™, Curve 8320, …, Pearl 8110
Motorola: Hint QA30, Tundra, …, MOTO W233
Nokia: Nokia N97, Nokia E75, ..., Nokia 6010

We assume that customer queries are focused on some terms or keywords of cell phone. Different customers may be concerned about different problems that depend on customer prefenerces and domain knowledge. In order to implement this empirical study, we have collected above 30 model description document data from some websites. In the processing of querying, these terms/phrases are recognized through DFS or BFS. On average, the length of each document is about 16.8 sentences and 312.4 words.

Five different group customers are investigated and experiment data and texts are gotten. Here five groups are represented as follows.

G_1 a group of retired people;
G_2 a group of university faculty;

G_3 a group of company sellers;
G_4 a group of graduate students;
G_5 a group of undergraduate students

The objective of concept generation involves identifying customer needs and then mapping those needs into a set of cell phone attributes or specifications. Considering this case study where the designer would like to generate a new cell phone concept, it is necessary to satisfy the following basic requirements.

- A hybrid cell phone with a touch screen and a hardware keyboard
- Push to Talk, Bluetooth, MP3, Video-fairly easy to master and operate
- Nice long battery life (above 3 weeks), lots of functions and extras-pretty easy to master
- Digital camera / digital player
- Added security features

Based on these requirements, we could first of all formulate the queries from users in order to extract customer preference terms. For example, retired people like old-fashioned and inflexible and easy to operate cell phone. College faculties and doctors like performance reliable cell phones with good voice effects. Company sellers like high-grade and luxurious cell phones with wide screen to show enterprise status. The graduate students like emotion-feeling features with multimedia function. And undergraduate students like small and exquisite cell phones with flip or kitty shapes in bright colors. They will inquire about function, externality, price, battery life, use environment, color, etc. The number of query keywords corresponding to different taxonomies is shown in Fig.8.

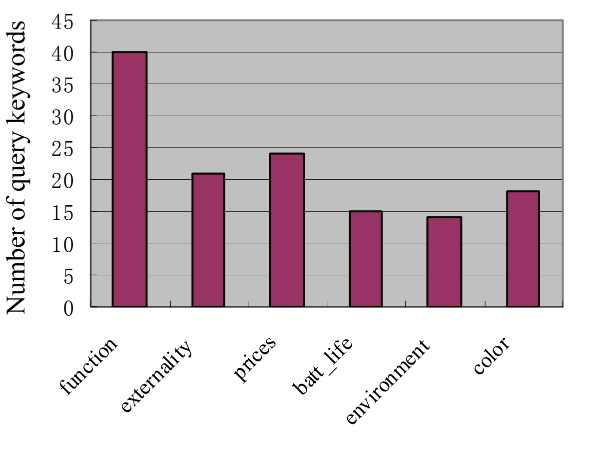

Figure 8 Distribution of keyword in different taxonomies

We collected a total of 20 queries from five different groups, in which 2 queries are eliminated, because they aren't related to customer preferences. The rest 18 queries are classified as general queries, specific queries, and context queries. The general queries are associated with the upper-level concepts of the ontology, such as customer preferences towards function and performance description, while the specific queries are associated with the lower-level concepts, e.g., shape features and material attributes. The third category is context queries that cannot be easily described except in context expression. In context expression a customer specifies a certain context in order to make the query unambiguous, such as cell phone performance parameters and quantitative indices. We can use Protégé 3.1 (http://protege.stanford.edu/) to generate domain ontology, in which preference taxonomy was generated as the basis of concept hierarchies. The lexical terms of the concept were modeled as the slot attribute of each concept class. It also supports the domain ontology model in several formats, such as XML, OWL and RDF. Domain ontology model was translated into XML script as input to the system [31].

Ontology concepts are built based on customer preference. Their interrelationships and number of types are statistically calculated and shown in Fig.9.

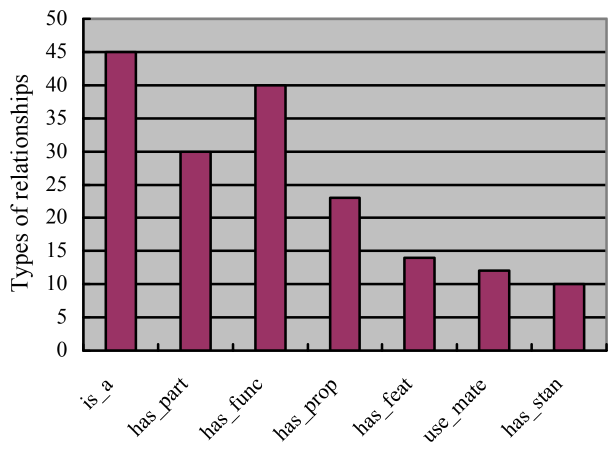

Figure 9 Distribution of different interrelationships

Tab.3 gives the comparison of query empirical results. The result shows ontology retrieval is superior to traditional keyword retrieval. As for G_1 group query, they are mainly concerned about easy operation, word clarity, monotonous color, hackneyed shape, simple function, and so on. After carrying out document retrieval, we obtain "Nokia 6010" model

Table 3 Recall/Precision based on ontology & keywords

types of queries		Recall		Precision	
		Ontology	Keyword	Ontology	Keyword
General	9	92%	42%	91%	46%
Specific	6	83%	78%	79%	81%
Context	3	86%	69%	80%	35%

close to G_1 group requirements. However, some concepts still need to be added in order to satisfy customer preferences, such as emergency call, voice help, accident discernment, etc. Also, on the basis of G_2 group query, we obtain "Tour 9630" model close to customer requirements, but there are still differences. It needs to add some new concepts to fit customer preferences, such as input and edit text, script, less radiation, etc. Although the results of retrieval provide the fittest brands to customers, they still can not satisfy customer preferences completely and need to attach some additional concepts as the supplement. Figure 10 presents different groups corresponding customer preference models and the number of concept supplements. In fact, all brand name cell phones always confine their functions within a certain range, and perfection cannot be obtained with the given limit cost. Sometimes performance is good, but their functions are not remarkable. Some designs are pretty good, but performance and customer service are unlikely to entirely satisfy customers' desires. The companies have been attempting to play a game of product function that satisfies customer preferences.

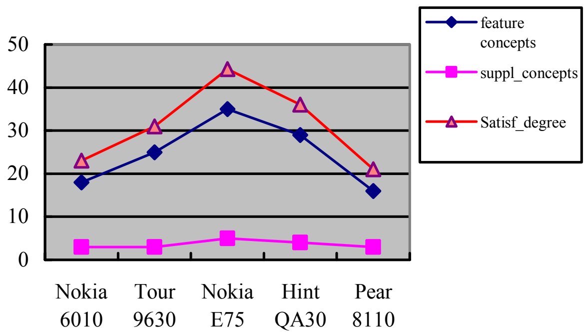

Figure 10 Different G preference brands and concepts

7. CONCLUSIONS AND DISCUSSIONS

In this paper, the customer preference ontology is developed and preference information is extracted to build preference lexicon. An ontology-based model is developed for the information extraction. The concept generation and selection of information are based on customer preference ontology. We have shown how the ontology can be used to generate and measure design concepts in customer queries. We have used the preference domain knowledge of cell phone for describing the proposed approach, while the results can be applied to other similar products. Our ontology-based retrieval demonstrates its superiority to keyword-based search techniques by evaluating recall and precision. However, the further research work is needed and discussed as follows.

- As many new technologies are being developed as well as converge with each other, the varieties of product will continue to increase. As a result, design requirements will become more and more complicated. Thus the large amount of unstructured and informal design information is steadily increasing, such as engineers' log, product image, nonstandard language description, etc. These texts are less likely to comply with the formal documentation format [34]. At the same time, they are still a part of customer preference document extraction. However, it is difficult to extract the ontology concept semantics from these documents. Further research work is worth in the future.

- Although we can calculate the result of cell phone models close to customer preference from VSM modeling, and propose the number of concept supplements, they are independent each other and depend on the designers' preferences. In addition, ontology preference retrieval can easily measure the distance between two concepts and find the fittest concepts of the customer queries. How to combine these concepts into a new configuration scheme still needs an effective algorithm. How to reach a preference compromise between customers and designers needs methods and algorithmic development.

- Customer preferences are not static and are indeed changeable. At a particular time, customers show a strong liking for certain cell phones. But later their preferences perhaps change and they show a liking for another cell phone. Therefore, we would like to build an ontology that is easy to update and can dynamically adapt to customer preference changes. In addition, as time goes on, customer perceptions and product concepts are constantly changing around customer preferences [36]. An automatic analysis approach to keeping abreast of the changes is needed to satisfy with a fast and simple response to customer preferences and changes in the market.

• Information extraction of customer preferences is currently based on indexing the sentence semantic rules, in which the preference lexicon and domain knowledge are crucial to achieve information retrieval. However, we don't consider document syntactic structures and syntactic rules. If we develop an automatic document indexing system for customer preferences in order to minimize human intervention, the sentence structures have to be analyzed based on syntactic rules [24]. Further work is needed for next step in our research.

• The preference lexicon, in this paper, only collects the most positive context terms and phrases. However, some negative context terms [19], such as the negative adverbs, "no, not, hardly, rarely" or the negative adjectives, "bad, ridiculous, impracticable, troublesome." Actually, double-negation equates to affirmation. We often use this in writing and speaking. Such word frequencies may be useful for some customers, but not for all. A more accurate language model for the elicitation of customer preferences will be developed to take this aspect of the matter into account.

• Iindividual preference often is a simple expression of group preferences. Sometime it has a big difference from group preferences. Perhaps everyone has a different insight about customer preferences [37]. Under these circumstances, "customized preferences" needs to be studied as well. Perhaps it will add the variety and complexity of customer preferences and move us towards mass customization concepts for customer preferences. In addition, product personality and brand personality cannot completely depend on their shapes [38]. The differences maybe exist in among customers. Undoubtedly, best-sellers surely have their product personality. For example, a cell phone looks a good in appearance. We cannot say everyone likes it. On the contrary, we should get to know its function, performance, etc. that is, its personality catches our eyes, but not simply saying that it is good in appearance alone.

ACKNOWLEDEMENT

We would like to acknowledge partial support of Product Lifecycle Management (PLM) center at Purdue University and the Center for Advanced Manufacturing (CAM).

REFERENCES

[1] Claes, F., (1992), "A national customer satisfaction barometer: The Swedish experience"; *Journal of Marketing*, Vol.56, pp. 6-21.

[2] Zhao, R., Grosky, W. I., (2002), "Narrowing the semantic gap---improved text-based web document retrieval using visual features"; *IEEE Transaction on Multimedia*, Vol. 4, No. 2, pp.189-200.

[3] Jiao, J., Chen, C. H., (2006), "Customer requirement management in product development: A review of research issues"; *Concurrent Engineering*, 14(3):173-185.

[4] West, P. M., Brockett, P. L., Golden, L. L., (1997), "A comparative analysis of neural networks and statistical methods for predicting consumer choice"; *Marketing Science*, Vol.16, pp.370-391.

[5] Ji, H., Yang, M. C., Honda, T., (2007), "A probabilistic approach for extraction design preferences from design team discussion"; Pro. of the ASME International Design Engineering Technical Conference & Computer and Information in Engineering Conference, Sept. 27, Las Vegas

[6] Lilien, G.L., Rangaswamy, A., and Bruyn, A.D.,2007, "Principles of marketing engineering"; Trafford Publishing

[7] Desmet, P., Erp, J.V., and Karlsson, M., 2008, "Design & emotion moves"; Cambridge Scholars Publishing.

[8] Chen, C.H., Khoo, L.P., and Yan, W., (2003), "Evaluation of multicultural factors from elicited customer requirements for new product development"; *Research in Engineering Design: Theory, Applications and Concurrent Engineering*, Vol.14, No. 3, pp.119-130.

[9] Gerald, Z., 2003, "How customers think: Essential insights into the mind of the market"; Harvard Business School Press

[10] Erin, F. M., Richard, G., and Panos, Y. P., (2009), "Preference inconsistency in multidisciplinary design decision making"; *Journal of Mechanical Design*, Vol.131/031009-1-13

[11] Thomas, Y. L., (2009), "Adaptive text extraction for new product development"; Pro. of the ASME International Design Engineering Technical Conference & Computer and Information in Engineering Conference, Aug. 30, San Diego

[12] Khan, L., McLeod, D., Hovy, E., (2004), "Retrieval effectiveness of an ontology-based model for information retrieval"; *Int. J. Very Large Data Base* (VLDB), Vol.13, pp.71-85.

[13] Griffin, A., Hauser, J. R., (1993), "The voice of the customer"; *Marketing Science*, Vol.12, No.1, pp.1-27.

[14] Li, Z., Raskin, V., and Karthik Ramani, (2008), "Developing engineering ontology for information retrieval"; *Journal of Computing and Information Science in Engineering*, Vol. 8/ 011003-1-13.

[15] Claes, F., Michael, D., Johnson, E. W. A., Jaesung, C., Barbara, E. B., (1996), "The American Customer Satisfaction Index: nature, purpose, and findings. *Journal of Marketing*"; Vol.60, Oct.

[16] Klaus, G. G., Tino, B. L., (2005), "Explaining choice option attractiveness by beliefs elicited by the laddering method"; *Journal of Economic Psychology*, Vol.26, pp.223-241.

[17] Thorsten, G., Alexander, R., Isabelle, S., Roediger, V., (2008), "Revealing the expectations and preferences of complaining customers by combining the laddering interviewing technique with the Kano model of customer satisaction"; *Qualitative Market Research: An International Journal*, Vol.11, No.4, pp. 400-413.

[18] Linoff, G. S., Berry, M. J. A., (2001), "Mining the web: Transforming customer data into customer value"; *John Wiley & Sons*, New York

[19] Yang, Maria C., Wood III, W. H., Cutkosky, M. R., (2005), "Design information retrieval: a thesauri-based approach for reuse of informal design information"; *Engineering with Computers*, Vol. 21, pp.177-192.

[20] Orsborn, S., Cagan, J., Boatwright, P., (2009), "Quantifying aesthetic form preference in utility function"; *Journal of Mechanical Design*, Vol.131/ 061001-1-10.

[21] Berkowitz, M., (1987), "Product shapes as a design innovation strategy"; *J. Prod. Innovation Manage*, Vol. 4, No. 4, pp.274-283.

[22] Crilly, N., Moultrie, J., and Clarkson, P. J., (2009), "Shaping things: Intended consumer response and the other determinants of product form"; *Design Studies*, Vol. 30, pp.224-254.

[23] McCormack, J., and Cagan, J., (2002), "Supporting designer's hierarchies through parametric shape recognition"; *Environ Plann B Plann Des*, Vol.29, pp.913-931.

[24] Li, Z., Yang, M. C, and Ramani, K., (2009), "A methodology for engineering ontology acquistion and validation"; Artifical Intelligence for Engineering Design, Analysis and Manufacturing, Vol. 23, pp. 37-51.

[25] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R., (1990), "Indexing by latent semantic analysis"; *J. Amer. Soc. Inf. Sci*. 41(6): 391-407.

[26] Miller, G. A., (1995), "Wordnet: a lexical database for English"; *Communication of the ACM*, 38 (11): 39-41.

[27] Zeng, Y., (2004), "Environment-based formulation of design problem"; Integrated Design and Process Technologies, Vol. 8, No.4, pp.45-63.

[28] Studer, R., Benjamins, V. R., and Fensel, D., (1998), "Knowledge engineering: principles and methods"; *Data and Knowledge Engineering* (DKE), 25 (1-2): 161-197.

[29] Kitamura, Y., and Mizoguchi, R., (2004), "Ontology-based systemization of functional knowledge"; *J. Eng. Design*, 15(4): 327-351.

[30] Chwolka, A., and Raith, M. G., (2001), "Group preference aggregation with the AHP-implications for multiple-issue agendas"; *European Journal of Operational Research*, 132: 176-186.

[31] Lim, S. C. J., Liu, Y., Lee, W. B., (2009), "Product analysis and variants derivation based on a semantically annotated product family ontology"; Pro. of the ASME International Design Engineering Technical Conference & Computer and Information in Engineering Conference, Aug.30, San Diego

[32] Marcus, M., Santorini, B., & Marcinkiewicz, M. A., (1994), "Building a large annotated corpus of English: The Penn Treebank"; *Computational Linguistics*, 19(2):313-330.

[33] Salton, G., 1998, "Automatic text process"; Wokingham, MA: Addison-Wesley

[34] Castells, P., Fernandez, M., and Vallet, D., (2007), "An adaptation of the vector space model for ontology based information retrieval"; *IEEE Transaction on Knowledge and Data Engineering*, 19 (2): 261-272.

[35] Shen, Li., (2009), "A semantic vector retrieval model for desktop documents"; *J. Software Engineering & Application*, 2: 55-59.

[36] Lin, Y.C., Lai, H.H., and Yeh, C.H., (2007), "Consumer-oriented product form design based on fuzzy logic: A case study of mobile phones"; *International Journal of Industrial Ergonomics*, 37: 531-543.

[37] Zaltman, G., 2003, "How customers think-Essential insights into the mind of the market"; Harvard Business School Press

[38] Mugge, R., Govers, P.C.M., Schoormans, J.P.L., (2009), "The development and testing of a product personality scale"; *Design Studies*, 30: 287-302.