

Automated Ontology-based Customer Needs Translation and Representation

Chen Xingyu¹, Chen Chun-hsien¹ and Leong Kah Fai¹

¹Department of Systems and Engineering Management, Nanyang Technological University, Singapore
{chen0532, mchchen, mkfleong}@ntu.edu.sg

Abstract—Intense competition and high failure product rate calls for a deeper understanding of customer needs. However, the process of translation and interpretation of voice of customers (VoC) into customer needs statements involves much imprecise information with linguistic vague descriptions. In order to get accurate customer need statements and further to enhance customer satisfaction for product success, we have to take a review of how to well interpret, translate and represent customer needs in the front-end product design process. We endow the use of ontology is an efficient approach for accurate custom needs translation and representation. Although ontology is promising for our target, it is known that manually building ontology is a tedious work, which requires much human effort. To solve this problem, we present a framework that automatically translates and represents customers' needs in the form of ontology in this paper. We first employ natural language processing tools to pre-process the customer statements. Then, a set of algorithms are used to extract concepts and relations from the processed statements, building the final ontology. We have conducted a case study of the framework. In particular, the customer statements about digital camera products are collected from customer reviews. Then, we build ontology from the collected statement. The experimental results demonstrate the efficacy of our framework.

Keywords: *Customer needs representation; natural language processing; ontology construction; ontology learning*

I. INTRODUCTION

In this fast moving consumer market, it is known that over 80% of new products fail [1]. As customers are better educated and have greater access to information nowadays, they are increasingly more demanding than ever before. This trend challenges companies in terms of satisfying customers' growing sophisticated needs and improving customer satisfaction [2]. The journey of achieving customer satisfaction through the design process starts with customer needs identification [3]. Only companies that can effectively elicit customer's genuine requirements and embed these requirements into product design concepts may succeed in the marketplace [4]. It is, therefore, crucial to get accurate customer need statements and use them as inputs to the succeeding product design stages.

Customer needs are typically expressed using written statements and are the results of interpreting the needs underlying the raw data gathered from the customers [5]. The process of translation and interpretation ineluctably involves much imprecise information with linguistic vague descriptions.

It requires a very explicit and clear interpretation and translation to customer's linguistic needs.

Usually such imprecise customer needs information still cannot be effectively handled by methods such as qualitative methods and quantitative methods such as fuzzy logic and rough set based methods in the traditional way. To fill up this gap, ontology is used for its ability in establishing explicit formal vocabulary to solve the heterogeneity of customer statements.

Manually build ontology is a tedious and time-consuming process which requires much time and human resources. To reduce the heavy burden in ontology building, ontology learning has been developed to extract concepts, relations from a specific domain [6]. By using natural language (NLP) process tools including a morphologic analyzer, a part-of-speech (POS) tagger, and a chunk parser, unstructured customer statements on a specific product can be translated, analyzed and modeled automatically.

As we know, customer statements may account for the needs for both tangible functions (e.g. forms) and intangible product characteristics (e.g. social and emotional related product characteristics) [7]. It is difficult to embody intangible product characteristics so as to satisfy the needs derived from the customer statements. However, recent studies have suggested that customers' needs for product form and functions are by and large motivated by symbolic product meanings [8]. In other words, need statements toward intangible product characteristics, which is based on the subjective perception on high level product characteristics, can be reflected with tangible product characteristics. Therefore, apart from ontology building from concept with explicit relations, another problem is how to represent higher-level concepts with implicit relations. Ontology is able to comprehensively formulate upper-level ontology by providing general-purpose terms and acting as a foundation for more specific domain ontology.

By combining ontology with natural language processing, Concepts and taxonomic (IS-A) relations can be extracted from customer statement. This method can also facilitate in extracting abstract themes hiding behind the extracted concepts and relations. The main contribution of this paper is by combining NLP, ontology building and ontology learning for automated customer needs translation and representation. Latent relationships between different terms are also obtained through non-taxonomy relation extraction. Customer statements are further classified into several themes and sub-

themes with different levels of abstractions to determine customer needs for intangible product characteristics.

The rest of the paper is organized as follows. The next section presents the system's framework and proposed methodology. The third section illustrates the whole process using a case study. Conclusions and future work are given in the final section.

II. METHODOLOGY

A. Overview

The overall framework for modeling the customer needs translation and representation consists of three stages, namely knowledge acquisition, natural language pre-processing, and knowledge representation. The whole process is illustrated in Fig. 1. First, VoC is acquired from raw text in customer reviews. Second, linguistic processing is conducted to normalize unstructured customer statements into structured need statements. After this stage, set of terms are extracted from documents. Semantic analysis, which is also a technique in natural language processing, is used to analyze the relationships between a set of documents and the terms they contain by producing a set of concepts. Third, ontology which is based on tangible product characteristics is built with sets of attributes and relations. Concepts are classified and well represented using taxonomies tree. These taxonomies which are constituted by concepts, attributes and relations are extracted from customer need statements. Several semantic rules are developed to describe the low level product characteristics of customer needs. Lastly, an extended ontology based semantic representation method is used to further represent the implicit concepts for high level themes generation.

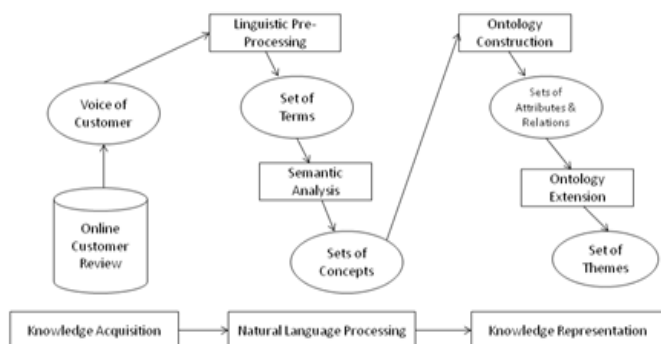


Figure 1. Overall framework of the proposed model

B. Natural language processing

Natural language processing runs into many stages [9]. It starts with tokenization which is the task to divide a text into single lexical tokens. Tokenization includes activities such as simple space identification, dehyphenation, sentence boundary detection, proper name recognition, and language detection. After tokenization, a lexical analysis is conducted on the tokenized text. A main task of the lexical analysis is Part of

Speech (POS) tagging, which is a prerequisite to derive grammatical words (e.g. determiners or prepositions) from lexical words (e.g. nouns, verbs or adjectives). Then, syntactic analysis is used to determine the underlying structure in a string of words. In syntactic analysis, a wide range of grammar formalisms should be applied in computational linguistics and techniques for chunk parsing and grammar formalisms on the whole sentence level are employed. Subsequently, semantic analysis is used to return the sentence's proposition using formal language [9]. Semantic analysis is able to locate the main components of a sentence (e.g. subject, verb, object, and so on), and identify name entities (e.g. John is a person; Singapore is a place). At the level of design application, product semantics can be considered as meanings associated with product characteristics such as form, dimensions, color, graphics, texture, transparency, fragility and so on.

Existing customer statements are typically unstructured text which needs designers' analysis to extract useful information. Meanwhile, the original information collected from customers mostly through survey, transaction data, and customer review should be filtered. With the explosion of the information, automatically extracting structured information from raw data becomes necessary, which requires analysis of syntactic structures and semantic meanings. Lexical terms are natural language words or phrases which are used to map the need and to explicitly represent the vocabularies of different ontology concepts. That is to say, word morphs, abbreviations, acronyms and synonyms of the words or phrases derived from the semantic analysis are lexical terms and share same concept with original lexical terms. Besides, the noun, verb, adverb and prepositional phrases can also be extracted from the original source. From lemmatization and POS tagging to different forms of both shallow and deep parsing together with statistical analysis, natural language processing is the preparation steps for both customer statement translation and ontology building.

C. Customer needs representation

As mentioned in earlier sections, ontology modeling, which defines a set of representational terms by using the hierarchical correlations or tree structures provides an effective method for knowledge representation of domain-specific concepts with different abstraction [10].

In this paper, we extracted semantic relations through both taxonomic relations and non-taxonomic relations. Taxonomic relations are the most important semantic relations in a domain ontology, the extraction of which has been well studied in the field of lexicon building [11]. Taxonomic relations can be extracted by using the well-known lexico-syntactic patterns giving a sentence containing a pattern. Taxonomic relations can also be extracted based on the term structure through string matching. WordNet [12] which contains terms and glosses like a traditional dictionary is used for taxonomic relation extraction.

Quite similar as the conventional approaches of non-taxonomic relations extraction [13], we hypothesize that verbs indicate non-taxonomic relations between concepts. A semantic relation of the (Concept, Relation, Concept) is extracted if its lexical relation can be founded from the raw text

in the form of (Noun, Verb, Noun). We adopted a rule-based method for such tuples extraction from the text. The parse tree is utilized for its effectiveness in non-taxonomic relations extraction [11].

We try to integrate ontology with lexicon resource for customer needs representation. For ontology modeling, the most fundamental relation of these concept objects is the taxonomic relation. Different from the existing approaches that only use concepts and taxonomic relations for ontology modeling, the proposed model also utilizes non-taxonomic relations to capture customer conceptual needs to product characteristics.

III. ILLUSTRATIVE CASE STUDY

In this section, we describe the entire translation process by using an industrial case study, in which certain customer needs from online customer reviews are identified and analyzed.

A. Obtaining the customer statements

In this study, we used data obtained from online customer reviews on Epinions.com. A set of 8000 online digital camera reviews were collected in 20 January, 2011.

B. Data processing for text analysis

After obtaining the raw data of customer reviews, semantic pre-processing was conducted for linguistic translation. Natural language processing mechanism was used for linguistic pre-processing and text analytics. The tasks in natural language processing included part-of-speech tagging, chunking, syntactic parsing, and semantic analysis. The implementations of these tasks were shown in Table I.

TABLE I. AN EXAMPLE OF NATURAL LANGUAGE PROCESSING TO CUSTOMER REVIEW

Tasks	Example
Raw text	... It gives me the ability to capture many special moments in an instant...
POS	... ('It', 'PRP'), ('gives', 'VBZ'), ('me', 'PRP'), ('ability', 'NN'), ('capture', 'VB'), ('many', 'JJ'), ('instant', 'NN')...
Chunk	[S:It] [V:gives] [O:me] [O: the ability] ...
Syntactic analysis	Parse tree

After implementing these natural language processing tasks, written statements from customer reviews were chunked into units of syntactic terms. Subsequently, words were further treated with syntactic analysis and semantic analysis. Further statistical analysis was then performed to correctly infer relations of human meaning-based judgments and performance. After this, we managed to extract correct reflections and word meanings, and a set of structured words of customer statements were generated accordingly.

C. Data Evaluation

The evaluation consisted of two sub-jobs that are automatic customer statement translation and ontology construction. The first job was the comparison between manually generated translation and automatically generated translation. Here, the

major task was to check whether the method can be done automatically. The second job involved checking if the established ontology can well represented knowledge acquired from customer statements. Specifically, we wanted to find out whether the non-taxonomic relations can well represent those customer needs on intangible product characteristics. In this study, we compared the ontology derived from our approaches with publicly available expert-generated ontology from online buying guides. The online buying guides we used are also from epinions.com. The 11 product characteristics from online buying guides are listed in Table II.

TABLE II. DIGITAL CAMERA CHARACTERISTICS FROM ONLINE BUYING GUIDES

Megapixels	Price Range	Optical Zoom & Digital Zoom
Size	Memory	Battery Life/Power
TV Connections	LCD Viewfinders	Realistic Expectations
Digital Video Format/Movie Mode		Computer Connectivity

D. Results and discussions

After we got the structured customer statements, ontology on product characteristics was built. Since customer statements typically include information on both tangible and intangible product characteristics, an ontology based on tangible product characteristics (e.g. color, function, cost et al.) was built first. As illustrated in Fig. 2, these needs were further broken into more concrete product characteristics (e.g. color-> bright color, dark color et al.)

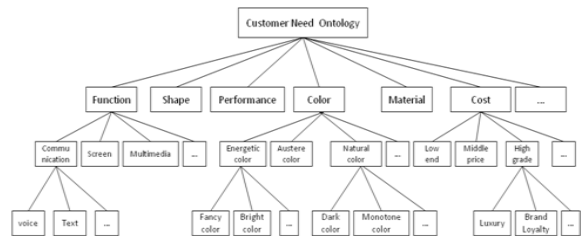


Figure 2. The taxonomy of explicit product characteristics ontology

Second, an extended ontology based on customer’s needs for intangible product characteristics was built. Different themes were extracted from the established ontology as shown in Fig. 3.

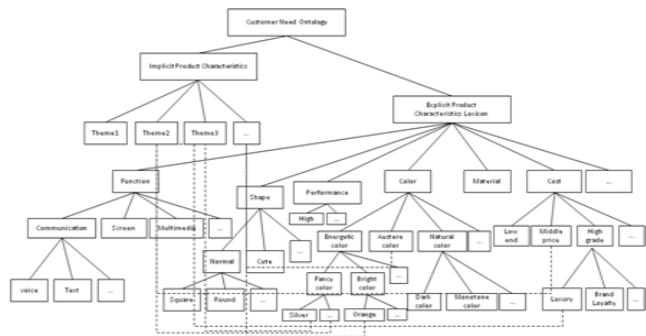


Figure 3. Theme-extraction based on the different levels of abstraction with ontology representation

Themes and sub-themes were generated based on the previous ontology. The built ontology was intended to capture knowledge from customers at different levels. Based on the established taxonomy for explicit product characteristics, high level semantic themes and sub themes were extracted from different levels of abstraction of customer statements on product characteristics.

The whole process consisted of two major steps which are linguistic pre-processing and ontology construction. At the first step, by breaking down and analyzing the splitted units of words using semantic analysis techniques, the translation from VoC to customer statements can be done automatically. At the second step, the ontology construction process was implemented to get different levels of abstraction from customer statements to product characteristics.

Based on the evaluation from experts, the proposed method can well solve the problems and meet the goals. According to the criteria, the ontology for digital cameras built using the proposed method outperformed the ontology generated by experts from online buying guide. The experts' evaluation results are summarized in Table III. These results are mainly about the feasibility of the methodology in achieving the research objective.

TABLE III. EVALUATION RESULTS

Main goals	Sub goals	Solution	Results (Solved: V)
Customer Statement Representation	Explicit customer statements	Ontology Construction	V
	Implicit customer statements		V

Although the evaluation is intended to illustrate the viability of the proposed approaches and framework, it leaves a huge space for extensibility. The originality of this research lies in the way of representing customer needs from a natural language processing perspective. Ontology combined with the lexicon opens a new way for customer requirement management. Furthermore, we distinguished concrete and implicit entities through different levels of abstraction when building the ontology. Based on this, we were able to represent customer needs for intangible product characteristics from a novel perspective.

IV. CONCLUSIONS

In this paper, an ontology-based automatic text extraction and customer statement representation model is proposed to get a set of well-organized and deeply-analyzed customer need statements from raw customer data for the succeeding stages in the product development process. By using online customer reviews, the uttered and unuttered customers' needs are first normalized into structured customer statements automatically by using natural language processing treatment. Customer

statements on both tangible and intangible product characteristics are well represented with different levels of abstraction. Customer statements on tangible product characteristics are mapped onto ontology with attributes and relations. Extended ontology is built to get high level abstraction from existing ontology. Different themes are extracted from the established ontology based on different levels of abstraction. In this sense, a set of useful customer need statements can be passed to the succeeding stages in product design and be used as inputs to product specification.

This study has demonstrated that our work has great potential in improving product design through well managing customer statements which served as the most important inputs for the following stages in product design. Future research includes developing ontology learning models for semantic rule generation and application. Through ontology learning, ontological knowledge from various forms of data can be found automatically or semi-automatically. It can overcome the bottleneck of ontology acquisition in ontology development. Natural language processing techniques can also be integrated into ontology learning to generate linguistic rules from the text automatically.

REFERENCES

- [1] J. Delin, et al., "Linguistic support for concept selection decisions," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, vol. 21, pp. 123-135, 2007.
- [2] I. Alam and C. Perry, "A customer-oriented new service development process," *Journal of Services Marketing*, vol. 16, pp. 515-534, 2002.
- [3] I. Alam, "New service development process: Emerging versus developed markets," *Journal of Global Marketing*, vol. 20, pp. 43-55, 2007.
- [4] A. McKay, et al., "Requirements management: A representation scheme for product specifications," *CAD Computer Aided Design*, vol. 33, pp. 511-520, 2001.
- [5] J. Kušar, et al., "Finding and evaluating customers' needs in the product-development process," *Ugotavljanje in vrednotenje potreb kupcev y postopku razvoja izdelka*, vol. 53, pp. 78-104, 2007.
- [6] A. Gómez-Pérez and D. Manzano-Macho, "A survey of ontology learning methods and techniques," *A Survey of Ontology Learning Methods and Techniques*, 2003.
- [7] V. Hubel and D. B. Lussow, *Focus on designing*. Toronto: McGraw-Hill Ryerson, 1984.
- [8] M. E. H. Creusen and J. P. L. Schoormans, "The different roles of product appearance in consumer choice," *Journal of Product Innovation Management*, vol. 22, pp. 63-81, 2005.
- [9] J. Davies, et al., Eds., *Towards the semantic web: ontology-driven knowledge management*, 2002, p. ^pp. Pages.
- [10] W. Van Der Vegte, et al., "Coping with unintended behavior of users and products: Ontological modelling of product functionality and use," 2004, pp. 683-692.
- [11] X. Jiang and A. H. Tan, "CRCTOL: A semantic-based domain ontology learning system," *Journal of the American Society for Information Science and Technology*, vol. 61, pp. 150-168, 2010.
- [12] C. Fellbaum, *WordNet: An Electronic Lexical Database*: Bradford Books, 1998.
- [13] M. Ciaramita, et al., "Unsupervised learning of semantic relations between concepts of a molecular biology ontology," *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pp. 659-664, 2005.