ICED13
INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED13
19-22 AUGUST 2013, SUNGKYUNKWAN UNIVERSITY, SEOUL, KOREA

DISCOVERING CONTEXTUAL TAGS FROM PRODUCT REVIEW USING SEMANTIC RELATEDNESS

Soon Chong Johnson LIM (1), Ying LIU (2)
1: Universiti Tun Hussein Onn Malaysia, Malaysia; 2: National University of Singapore, Singapore

ABSTRACT
Nowadays, online product reviews has enabled product designers to better understand product related issues from the users' perspective. In the design community, there are a number of studies that have focused on studying product reviews in various analysis perspectives. While these are essential, we noticed that contextual annotation of tags has not been fully explored. We reckoned that such an annotation is equally important to better clarify the tags' context where tasks such as design experience analysis and faceted product comparison can be made possible. However, the challenge lies in automatic discovery of contextual tags from product reviews. Consequently, this paper proposed a learnable approach to address this issue. A ranking algorithm is proposed to rank important key terms along with an approach to discover contextual annotation of a given term. The performance evaluation of our proposal is done using two annotated corpus. A case study using a small laptop reviews corpus is also reported to showcase how our algorithm can be applied towards product understanding and product ontology development. Finally, we conclude this paper with some indications for future work.

Keywords: design informatics, information management, contextual tags, semantic relatedness

Contact:
Dr. Soon Chong Johnson Lim
Universiti Tun Hussein Onn Malaysia
Dept. of Engineering, Faculty of Technical & Vocational Education
Batu Pahat, Johor
86400
Malaysia
johnson@uthm.edu.my

1 INTRODUCTION

Nowadays, the rapid development of Internet technologies and the advent of Web 2.0 applications such as online forums, e-commerce portals and blogs have allowed Internet users to easily share their views online. From a product user perspective, it has become a common scene for customers to write reviews related to their feature preferences, views or actual user experience associated with a product. Whether the opinions come from an average customer or from a professional user point-of-view, the availability of these reviews has presented huge potential for mining useful product information. However, due to the sheer amount of product reviews available in various distributed sources, it is not possible for product designers to gather and analyze all of these reviews manually. Thus, automated processing of reviews is a more feasible and practical approach towards mining interesting patterns from these reviews.

Previously, there are a number of studies that focused on analyzing product reviews. Works in the area of opinion mining or sentiment analysis (Hu and Liu, 2004, Loh et al., 2009), product review summarization (Ling et al., 2008, Zhan et al., 2009) and determining design review helpfulness (Jin and Liu, 2010) are among the notable ones. While most of the current works emphasized on identifying semantic orientation of reviews towards certain product feature, or generating review summary according to elicited topical words, we noticed that contextual annotation has not been fully explored. Technically, these tags represent salient product features identified from review and corresponding annotation may describe various facets of a feature, such as preferences, emotions, or usage experience. Such an annotation serves as the important next step towards better understanding of the context of identified tags. For instance, in reviews about camera, contextual annotation enables the suggestion of related product features (e.g., "lens", "focus") with topical word of interest ("picture quality"), or discovery of other contextually similar product features ("video recording"). For product designers, this allows better understanding of component associations, usage experience, emotions, etc. from various customer or professional viewpoints.

Realizing the benefits of contextual product information, our interest is to discover contextual tags from product reviews. Nevertheless, the research issue is: how can we automatically learn relevant, contextual tags, which corresponds to a input query, from a collection of product reviews? To the best of our knowledge, there are relatively few studies that have explored on contextual tags discovery and particularly, on how its application can benefit the design community. In this paper, we propose an approach to automatically learn contextual tags from a collection of product review using semantic relatedness. Using this feature, an iterative ranking approach, FacetRank, is proposed for contextual annotation from review documents. The rest of this paper is discussed as follows: Section 2 presents the current state of research in product review mining, key term extraction and annotation of key terms followed by a summary of issues. Section 3 describes our proposal of discovering contextual tags from product documents, followed by evaluation of our approach using two annotated corpus in Section 4. In Section 5, we present a case study using a small corpus of laptop reviews and discussed some potential applications, and finally Section 6 concludes our work with further discussions.

2 RELATED WORK

2.1 Review Analysis for Product Design

Product review is an increasingly valuable user-generated content for various purpose of understanding customer concerns, rectifying product issues, etc. Unfortunately, the rapid growth of online customer reviews has hinders the possibility of manual processing. Automated processing tasks of these reviews, on the other hand, is non-trivial due to the inherent features of product reviews, such as heterogeneous descriptions, distributed locations and language ambiguity (Liu, 2010). In the past several years, automated parsing and analysis of online reviews has received notable attention in major research forums such as SIGKDD and SIGIR (Hu and Liu, 2004, Ding and Liu, 2007). These works focused mainly in sentiment analysis, either positive or negative oriented descriptions, that are related to various product features.

In relation, the application of such information discovered from review documents towards product design has just gained considerable attention in more recent years. From literature, among the notable ones are: Lee (2007) attempted a hierarchical, two-staged process that includes association rules for assessing changing user needs based on online reviews. Loh et al. (2009) proposed a hybrid opnion extraction framework that extracts features and predict semantic orientation of expressed opinions from free text. Ling et al. (2008), attempted the issue of generating multi-faceted semantic overviews of arbitrary topics in a text collection for a query term. Their work focused on generating faceted models using a few user-supplied keywords to indicate interested facets (e.g. cost). Zhan et al. (2009) proposed an approach to automatically summarize multiple customer reviews based on their internal topic structure. Jin and Liu (2010) studied the quality of product reviews and the correlation between the ratings by customers and those by designers in order to determine review helpfulness.

2.2 Key Term Extraction

A term refers to either a word (i.e. single word) or a phrase that capture the main topics discussed in a document (Nguyen and Kan, 2007). While the word “keyphrase” is widely used in literature to indicate a salient phrase or a word, this study use “key term” to avoid this confusion. In literature, the process of key term extraction is performed in two steps: candidate term extraction and key term selection. The first task aims to extract a list of potential terms. In the second task, significant candidates are selected based on certain document features. Generally, key term extraction approaches can be viewed as either supervised or unsupervised. Supervised approaches require labeled terms to train classifiers in order to correctly tag unseen or new key terms. Among the studies using supervised approaches are *Kea* (Witten et al., 1999), Nguyen and Kan (2007) and *Maui* (Medelyan et al., 2009). These studies have applied a number of features, such as term frequency (TF), inverse term frequency (IDF) and position of first occurrence from a small collection of labeled key terms for classifier training. Unsupervised approaches, on the other hand, do not require training corpora. In comparison, approaches such as POS tagging (Wu et al., 2006, Mihalcea and Tarau, 2004) and shallow parsing (Barker and Cornacchia, 2000) are adopted to improve the quality of terms extracted.

For key terms selection, majority of the supervised approaches use probabilistic classifier (e.g., Naïve-Bayes) to perform selection of key terms. While the number of features considered for training classifiers differs among the studies reviewed, features considered consist of document-related features (e.g. TFxIDF), corpora-related features (e.g. keyphraseness) and measures of semantic relatedness (e.g. co-occurrence). It is noted that a combination of these multiple features can improve the identification of key terms (Medelyan et al., 2009). In contrast, unsupervised approaches use similar features that are formulated using a scoring or ranking function, e.g., frequency-based weighted score (Barker and Cornacchia, 2000) and corpus-based scoring method by Wu et al. (2006). Mihalcea and Tarau (2004) proposed TextRank, an iterative ranking algorithm based on PageRank (Brin and Page, 1998) using co-occurrence statistics between words.

2.3 Annotation of Key Term

The aim of annotating key terms is to discover contextual and meaningful description of a key term and its relationship with other key terms. One of the approaches to tackle this issue is using statistical approaches. Previously, researchers have deployed techniques such as closed frequent patterns (Pasquier et al., 1999) or maximal frequent sequences (Ahonen-Myka, 1999) in order to highly summarize similar key term patterns into a general pattern that provides better information beyond word support. Another stream of researchers have tried document summarization techniques (Ledeneva et al., 2008, Ye et al., 2007) in order to discover meaningful topical phrases that describes a document. While these methods can successfully reduce the redundancy of key terms extracted and present to users only the meaningful ones, the further annotation of key terms is still very much limited to statistical information (e.g. support). In order to annotate meaningful key terms with semantically related terms, another approach is to use pre-defined controlled vocabulary list, such as WordNet (Miller, 1995) or some other domain-specific thesauri. Under this perspective, this issue can be viewed as a problem of term or category assignment. Example of related studies are medical text indexer (Aronson et al., 2000) and medical vocabulary-based topic generation (Markó et al., 2004) that emphasizes on terms matching. The drawbacks of this approach is that building and maintaining a controlled vocabulary requires considerable amount of efforts and are often limited to certain domain. In addition, a large training sets is often needed for machine learning based matching (e.g. classification), which limits its effectiveness over untrained key terms. Realizing this limitation, later studies (Medelyan et al., 2009) have applied open-domain corpus such as Wikipedia as a user generated and collaboratively maintained corpus. While open domain data such as Wikipedia may be a better option, it is only often restricted to known examples and may not be applicable for new entry of key terms.

2.4 Summary of Literatures

From our literature survey, current works in product design domain are mainly focused on identifying sentiment of reviews towards certain product feature and review summarization. Among these studies, identifying salient product features or topical words automatically from product review collections is a common processing task regardless of the analysis perspective. In the technical perspective of key term extraction, the main disadvantage of supervised approaches is the requirement of labeled training examples. While the inclusion of additional features can be helpful for training better classifiers, the issue lies in the best mix of these features where it can be corpus dependent. On the other hand, unsupervised approaches are independent of the features of a particular document set and can be applied even without training dataset. While such an approach offers greater flexibility, unsupervised approaches may also produce ill-formed key terms without significant meanings.

In general, we observed that the issue of contextual annotation of identified tags has not being fully investigated. To the best of the author's knowledge, Mei et al. (2007) is the first research group that has formally addressed the issue of semantic annotation of frequent patterns. They proposed a framework and dictionary analogy where semantic annotation of a frequent pattern consists of context models, a set of representative transactions and a set of semantically similar pattern. For studies related to product reviews, works by Ling et al. (2008) on generating faceted overview of topical words in review is another close example. Nevertheless, similar to the outcome by Zhan et al. (2009), their work is more focused on generating a summarized form of review and not contextually related term associations that is intended in this study. In this paper, we proposed an approach towards discovering contextual annotation that is relevant to a term. An unsupervised key term extraction approach that utilizes semantic relatedness information of domain specific corpus is preferred while avoiding the requirement of training examples. The idea is key term extraction of a document using its own semantic relatedness feature. Based on this feature, a suitable ranking approach is proposed to determine important key terms. For semantic tags discovery for key terms, the dictionary analogy as proposed by Mei et al. (2007) is adopted. We attempt to generate contextual annotation for a key term using similar analogy through building contextual/faceted model for the key term and to retrieve contextual tags through comparing faceted models of potential key terms.

3 DISCOVERING CONTEXTUAL TAGS FROM PRODUCT REVIEWS

This section details our proposal for discovering contextual tags from product reviews using semantic relatedness. Semantic relatedness generally refers to the degree of which a given pair of terms is related. Computationally, a semantic relatedness measure serves as a feature metric to indicate the strength of these relationships bonding. In this study, the strength of semantic relatedness is defined using pointwise mutual information (PMI) (Church and Hanks, 1990) as indicated in Equation (1). An assumption is made where candidate terms (t_1,t_2) that occur together are semantically associated. Using PMI measures, a suitable ranking algorithm is required to judge the importance of each terms based on these associations. In this study, the ranking problem is modeled using a graph-based ranking algorithm that is adapted from the PageRank algorithm (Brin and Page, 1998). The original PageRank is modified to form FacetRank (*FR*), a recursive ranking algorithm using semantic relatedness between terms(*V*) as in Equation(2). In this study, semantic relatedness measure using PMI is non-directed. Thus, *in(V)* and *out(V)* are similar representations of undirected term associations.

$$PMI(t_1, t_2) = \log \frac{\Pr(t_1, t_2)}{\Pr(t_1) \cdot \Pr(t_2)} \quad (1)$$

$$FR(V_i) = (1-d) + d \cdot \sum_{V_j \in in(V_i)} \frac{PMI_{i,j} \cdot FR(V_j)}{\sum_{V_k \in out(V_j)} PMI_{j,k}} \quad (2)$$

3.1 Key Terms Extraction

In this study, candidate term extraction process follows the pre-processing steps as proposed in *Kea* algorithm (Witten et al., 1999). The use of linguistic features, e.g. POS tagging and shallow parsing, is not considered in this study to avoid the use of extra tagging and selection process (which slows down the overall performance, especially on large, heterogeneous product review documents) and the use of specific linguistic corpus for shallow parsing. The stop-word list used in this study contains 425 words in nine syntactic classes (e.g. conjunctions, articles, etc.). Candidate terms are case-folded (i.e. to lower case) and stemmed using Porter stemmer (Porter, 1980) to discard any suffixes. The original form of candidate terms, however, is still retained for presentation purpose. Stemming is applied for comparison between candidate key terms and actual gold standard matching during evaluation. Candidate terms are then ranked using FacetRank for key terms selection. Statistical-based semantic relatedness metric in this study provides a flexible approach towards key terms extraction using different semantic relatedness information. The overall key terms extraction process is illustrated in Figure 1.

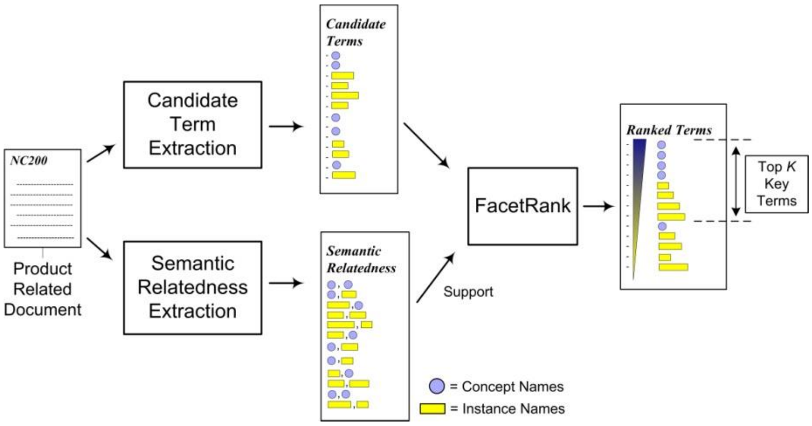

Figure 1. Key Term Extraction Process using FacetRank.

3.2 Contextual Annotation of Key Term

Upon the selection of candidate key terms, the next task is to generate faceted annotation of a key term. In a single document, different combination of terms can suggest different facets, i.e., specific perspectives of interest to user. For instance, for product review document that describes a digital camera model, possible product features extracted can be “flash”, “lens”, “image quality”, “image processor”, “auto focus” and “intelligent lighting”. Under the component context, the entities “flash”, “lens” and “image processor” represent the generic camera components; function wise, the phrase “auto focus” represents a camera function that is associated with “lens”; from a professional photographer’s perspective, “image quality” can be associated with both “image processor” and “lens”. Given a key term (say, "lens"), the aim of our annotation is to discover all these possible associations with other key terms and additional useful description to describe the term, that describe facets such as components or functions, subject to the context of a given corpus.

In this study, we suggest associations at smaller granularity of sentence level where we assumed that key terms contained in a sentence are semantically associated and describe a particular facet. Such a group of key terms corresponding to original document sentences is named entity set (ES). The algorithm for generating ES is as shown in Figure 2(a). The algorithm produces a collection of ES by comparing each extracted entity, $e \in E$ with every sentence in a document, $d \in D$. Using FacetRank, each key terms that are contained in an *es* is iteratively ranked. Ranked ES is named as faceted unit (FU), the basic building block of faceted modeling that describes an entry key term. For a FU, the highest ranked term is selected as faceted indicator, a representative key term that indicates the facet of an FU. In order to reduce redundant FUs, clustering is performed to aggregate similar FUs together. Hierarchical agglomerative clustering (HAC) (Jain et al., 1999) is used with algorithm as shown in Figure 2(b). For clustering, similarity between two FUs is determined using Euclidean distance measure. For similarity between clusters, single linkage scheme is used where distance between two cluster pairs is the smallest distance between two faceted units in both clusters. A maximum of two faceted indicators from each representative *fu* are aggregated as a cluster’s concepts.

Upon completion, we are able to apply the processed information for contextual annotation. Given a term *t*, the task of annotation is the process of selecting representative FUs where *t* occurs at least once in the sentences corresponding to the FUs. Once related FUs are selected, the corresponding faceted indicators of these FUs are identified to determine faceted weight, a measure of association strength between t and related faceted indicators using PMI. Consequently, a faceted model for a term is defined via faceted indicator with corresponding faceted weights. Contextual annotation for the term consists of faceted indicators, associated sentences and other related terms that have similar faceted models with that of the term's. In this case, related terms can be important terminology pre-determined by user or the important top few key terms from each document. For comparison, Let FM_{t1} and FM_{t2} denote two facet models for query terms t_1 and t_2 respectively. The two query terms are associated if the difference or distance between their faceted models, $diff(FM_{t1}, FM_{t2}) \leq k$, where k is a user defined threshold value. While there are a number of different similarities or distance measure that can be applied, the simplest Euclidean distance is applied in this study that only considers partial matches (i.e., only common key terms are considered).

Input: (i) Original dataset, D of m documents = $\{d_1, d_2, d_3, \dots, d_m\}$; (ii) A set of n extracted entities, E = $\{e_1, e_2, e_3, \dots, e_n\}$
Output: A set of j entity sets, ES = $\{es_1, es_2, es_3, \dots, es_j\}$

initialize empty set M, ES, SS
for each ($d_\alpha \in D$)
    initialize sentence sets $SS_\alpha = \{ss_1, ss_2, ss_3, \dots, ss_n\}$
    for each ($ss_u \in SS_\alpha$)
        for each ($e_v \in E$)
            match e_v with ss_u
            if (ss_u contains e_v)
                add e_v into matched set, M
        if (M is not empty)
            add M as new entity set, $es \in ES$
            update ES with es
output ES
end

(a) Entity Set (ES) Generation Algorithm.

Input: (i) A set of j Faceted Units, FU = $\{fu_1, fu_2, fu_3, \dots, fu_j\}$ (ii) Clustering Threshold, k where k~[0,1], $t \in R$
Output: A set of k clusters, C = $\{c_1, c_2, c_3, \dots, c_k\}$

initialize empty sets D
initialize m clusters $c \in C$, each contains a faceted unit, fu
compute distance set, D where $d_{ij} = d(c_i,c_j)$, $d_{ij} \in D$ among set C
find initial minimum distance, $d_{min} = argmax$ D
while ($d_{min} \leq k$) // clustering starts
    select $d_{i,j}$ where $(i,j) = argmax_{i,j}$ D
    merge clusters c_i and c_j into a new cluster c_u
    remove c_i and c_j from C
    remove $d_{i,*} = d(c_i,*)$ and $d_{j,*} = d(c_j,*)$ from D
    update C with c_u
    foreach $c_v \neq c_u$
        compute $d_{uv} = f_{dist}(c_u,c_v)$
        update D with d_{uv}
    find $d_{min} = \min(d_{ij})$
output C // clusters generated
end

(b) HAC Clustering Algorithm.

Figure 2. Algorithms Used for Annotation of Key Terms

4 PERFORMANCE EVALUATION

4.1 Key Terms Extraction

For key terms extraction, an annotated corpus, *CiteULike-180* (Medelyan et al., 2009) is selected for evaluation purpose. *CiteULike-180* is a collaboratively tagged corpus, where the corpus contains 946 tags that are agreed at least by two human taggers, resulted in accurate tag sets that contain an average of five tags per document. Collaborative tagged documents with mutually agreed key terms are preferred in this study to mitigate potential biases caused by a single human annotator. The ground truth for a document in *CiteULike-180* consists of at least three tags on which two users have agreed. Evaluation wise, we followed the standard performance metrics of precision (the ratio or percentage of "correct" tags out of all extracted tags), recall (the ratio or percentage of "correct" tags out of all manually assigned tags, i.e. by human taggers) and F1-measure (the harmonic mean of the two). The FacetRank proposed in this study follows an unsupervised approach in general. Therefore, only unsupervised key term extraction algorithms will be compared. Among the unsupervised key term extraction algorithms, a few has been identified for comparison: TextRank (Mihalcea and Tarau, 2004), Document Profile (DP) Model (Liu et al., 2007) and KIP (Wu et al., 2006). The details of experimental settings for each algorithm are as indicated in Table 1. In this study, evaluation is performed for top five key terms and top ten key terms for each algorithm. A summary of evaluation results is given in Table 2. From the table, it has been discovered that in overall, results using top five key terms are generally better than top ten ones in terms of F1 measure. The inclusion of extra key terms helped to boost recall at the expense of precision. Among the unsupervised approaches, KIP produces the poorest results of F-measure at 0.15. The performance results indicate that DP Model with averaged PMI selection comes second with F-measure at 0.28. FacetRank is better than DP Model at F-1 = 0.35. Comparatively, TextRank produces the best results with F-measure of 0.40. Compared with TFxIDF baseline method, it has been discovered that the performance of all unsupervised approach are better except for KIP.

Table 1. Summary of Experimental Settings for Algorithms in Comparison

Algorithm	**TextRank**	**DP Model**	**KIP**	**FacetRank**
Experimental Settings	• Undirected • Co-occurrence window = 3 • Damping factor, d = 0.85 • Iterative ranking threshold, δ = 0.001	• Support, s = [2,10] • Gap, g = [0,27] • 27 sets of DP • Averaged PMI for evaluation	• Default Settings without pre-weighted keywords • "1 word – 3 words" selected	• Damping factor, d = 0.85 • Iterative ranking threshold, δ = 0.001

Table 2. Summary of Evaluation Results for CiteULike-180 Dataset

Algorithms	Top 5 Key Terms			Top 10 Key Terms		
	Pr	Rec	F-1	Pr	Rec	F-1
TextRank	0.31	0.54	0.40	0.20	0.67	0.31
FacetRank	0.29	0.46	0.35	0.16	0.52	0.25
DP Model + averaged PMI	0.20	0.49	0.28	0.11	0.50	0.18
KIP	0.30	0.10	0.15	0.22	0.14	0.17
TFxIDF baseline (Medelyan et al., 2009)	0.14	0.16	0.15	N/A	N/A	N/A

Our experimental results show that FacetRank is good at generating a better variety of salient key terms that consists of keywords and keyphrases. The drawback, however, is slower iterative ranking computation compared to TextRank that uses only single words. TextRank is able to generate better candidate words that consist of nouns and adjectives annotated using POS tagging. This explains the relatively good performance of TextRank over FacetRank. In comparison, FacetRank is also able to generate promising candidate terms using much simpler pre-processing steps and better suited to heterogeneous descriptions of review documents. While the use of morphological analysis may improve performance by producing better candidate terms, the disadvantage of such an approach is that POS taggers are only limited to a few languages.

4.2 Contextual Annotation of Key Term

For evaluation of contextual annotation of key term, we are unable to find any annotated corpus that is specially designed for such an evaluation purpose. In order to judge the effectiveness and quality of annotations, an annotated and classified document corpus, Manufacturing Corpus Version 1 (MCV1) (Liu et al., 2009) is used. For evaluation, a few input terms are selected randomly. The only criterion for selection is that these terms must exist in document text. The goal of evaluation is to judge the quality of annotations for these input terms based on classified manufacturing concepts. Prior to the evaluation process, pre-processing tasks for MCV1 corpus are performed. ES are generated using top 15 key terms. For FacetRank, settings as in Table 1 are used. The distance threshold value used for clustering in this study is k = 1.0. Implementation wise, all the essential information at document level, such as file name, sentence id, FUs, cluster groups etc. are indexed using Lucene[1], a full-text search Java package. Faceted model for each main category and sub-category labels of MCV1 and top four terms (two keywords and two keyphrases) from each document are built for later comparison with input query terms.

Table 3. Evaluation Results for Contextual Annotation of Example Input Query

Input Query (hits)	automated guided vehicle (5)	acoustic emission (13)
Faceted Indicator (Weight)	concept (4.5221), high level(2.97), system (3.4124)	wavelet (7.381),wear (5.2434), sensor (5.8785), common (4.7029)
Representative Sentences	**concept (4.5221)** …a new ***automated guided vehicle*** (agv) dispatching algorithm based on a bidding *concept*… **system (3.4124)** … ***automated guided vehicle system*** (agvs) simulation *system* (agvsimnet)… … an ***automated guided vehicle*** (agv) is a mobile robot commonly used to carry loads in material handling *systems* (mhs)…	**wavelet (7.381)** … a flank wear estimation technique in turning through *wavelet* representation of ***acoustic emission*** (ae) signals… **sensor (5.8785)** … *sensor* fusion method using both an ***acoustic emission*** (ae) *sensor* and a built in force *sensor* is introduced… … two different types of *sensor*, the ***acoustic emission*** (ae) and the power *sensor* …
Contextual Category Labels	material handling, kanban, flexible manufacturing system, cad	electric discharge machine, process design, carbide
Contextually Similar Terms	control strategy, process planning, control macro, job shop, net model	detect cut, tool failure, tool condition, flank wear, tool wear

[1] http://lucene.apache.org

The results for four input terms for evaluation are selectively shown in Table 3. Feasible contextual annotations are in a dictionary-like format for various input terms. From experimental results, it is noted that feasible faceted models can also be generated for less occurring key terms, such as the first two examples of “automated guided vehicle” and “computer aided manufacturing”. Based on the available category labels, some interesting annotations (e.g. “materials handling” for “automated guided vehicle”) are discovered. Besides category label associations, other contextually similar terms are also extracted. For instance, the terms “control strategy” and “process planning” actually do not co-occur with “automated guided vehicle”, but are suggested because their facets are similar. Another example is “tool failure” and “tool condition”, which are feasible annotations to “acoustic emission”. In both cases, the annotations are quite meaningful to the example input query term.

5 CASE STUDY

In order to illustrate our approach, we have performed a case study using a small corpus of laptop computers(Lim, 2011). The corpus contains a collection of 47 web documents: with 26 documents related to the ThinkPad SL410 series and 21 documents related to the ThinkPad X200 series. There are about 8,000 words totally, with 1,700 unique words in about 500 sentences for the ThinkPad SL410 dataset. The ThinkPad X200 dataset is larger, consists of about 16,000 words, 2,800 unique words in about 970 sentences. Following the methodology for key term extraction using FacetRank as explained in Section 3, a list of key terms are initially extracted using FacetRank. Later, ES were generated using top 15 key terms using ES generation algorithm. FacetRank was applied to generate faceted units from ES that have at least three entities. As a result, there were about 150 FUs generated from the corpus. These FUs were clustered using cluster distance threshold of k = 1.8, half of the average distance value between all initial clusters. This setting produced a total of 40 clusters. Using the faceted units generated, faceted modeling for a collection of important key terms (top ten key terms from each document) were generated. There are about 240 highly important key terms with this in regard. Faceted models for these key terms were generated for later comparison with query term. Table 4 shows the contextual annotation generated for two example queries: “screen” and “business”. From the table, it has been shown that FacetRank is able to generate semantically related annotations. For faceted models, faceted indicators such as “widescreen” and “12.1 inch” for query term “screen” are informative to users. For query term “business”, the faceted indicators generated such as “owner”, “user” and “superb” are also descriptive. From the results, it is discovered that a few contextual terms are also annotated, such as “wide aspect” and “display” for query term “screen”; and “performance” and “travel” for query term “business”. These annotations provide useful indicators of a query term’s context according to the corpus in consideration.

Table 4. Contextual Annotation Results of Two Input Query

Input Query (hits)	screen (49)	business (59)
Faceted Indicator (Weight)	screen (6.1174), wide (4.3804), widescreen (4.1174), 12.1 inch (4.0019), inch (3.9352), program (2.5324), notebook (1.073)	owner (5.9896), appeal (4.4047), user (4.2527), superb (4.1152), haven (3.9896),like (2.2083)
Representative Sentences	**wide (4.3804)** … *wide* *screen* display features a sharp 1280x800 native resolution… **widescreen (4.1174)** … 12.1 inch *widescreen* not only lends extra on *screen* workspace, it also…	**owner (5.9896)** … lenovo for creating a laptop that the small *business* *owner* can afford… **appeal (5.6338)** … built to *appeal* to the small to medium *business* user…
Contextually Similar Terms	screen, adapt, wide aspect, display, size	business, performance, travel, notebook, quality, design, price

Contextual tags learned from product reviews can have a number of potential applications during product design. One of the useful ones is to contextual information search, retrieval and information presentation. Presenting information contextually (e.g., Table 4) allows designers to have a better overview of their product query term's context and how it is related to other contextually similar terms. For instance, in Table 4, the term "business" can be related to "owner" (user concept), "appeal" (affordance) and "travel" (usage). Such an annotation facilitates better understanding of products as perceived from the user's perspective. In relation, designers can better compare the context of a similar product feature under different user's angle (e.g., average user vs. professionals), or to compare two different products under the same user's perspective. On the other perspective, our proposal can also suite the design related ontology development process where annotation of tags and relationships between ontological concepts is concerned. In this context, contextual or faceted analysis of a key term allows an ontologist to learn a key term's context from multiple domain specific corpuses. For instance, for camera, it is also possible to deduce the associations between the input term (say, "lithium ion") and ontological concepts (e.g. "battery life"). Also, the product ontology can be annotated with customer experience, emotions, etc., allowing designers to better understand product from different angles. The realization of all these features will help to reduce the time and resources needed during ontology development process where new concept associations can be better discovered and erroneous annotation can be better avoided.

6 CONCLUSION & FUTURE WORK

In product design, the availability of vast online product review has presented a great resource for product designers to elicit useful design-related information. This paper has presented an approach towards contextual tags discovery from product review using semantic relatedness and FacetRank, a recursive ranking approach. The outcome of evaluation and case study shows that our approach is feasible in suggesting contextually similar tags towards a given term of interest. Nevertheless, there are a few limitations that are detected. Firstly, a query term is required to occur in the corpus for at least once. We noticed that there are situations where faceted model of an input term only consists of very few faceted indicators. There are also situations where the sentence that contain the input term may not have FUs, or is associated with very few FUs. This is a limitation caused by the initial number of selected key terms for FU generation. Strategies such as adaptive number of key terms according to document size may improve the situation. Secondly, while the quality of annotation can certainly be improved by including greater features such as information content, an annotated corpus that is specifically built for evaluating the quality of contextual annotation is, to the best of our knowledge, lacking. In constructing such a corpus, the overall corpus design, selection of annotated input terms, inter-consistency of human annotators etc. are all non-trivial issues. In spite of this, we believe that such a corpus is important for future studies, especially in evaluating the quality of tags generated. Scalability is another important issue to examine performance issues of our approach in processing large amount of review data. Application wise, we are also interested to see how our proposal can be helpful to product designers in designing better products, or novice engineers in better design understanding. User studies is recommended for validation purpose and on overall time and cost benefits in real world scenario.

REFERENCES

Ahonen-Myka, H. (1999). Finding all maximal frequent sequences in text. *In:* Mladenic, D. & Grobelnik, M., eds. Proceedings of the 16th International conference on Machine Learning ICML-99 Workshop on Machine Learning in Text Data Analysis, 1999 J. Stefan Institue, Ljubljana. pp. 1-17.

Aronson, A. R., Bodenreider, O., Chang, H. F., Humphrey, S. M., Mork, J. G., Nelson, S. J., Rindflesch, T. C. & Wilbur, W. J. (2000). The NLM Indexing Initiative. *In:* Annual Fall Symposium of the American Medical Informatics Associations, AMIA '00, 2000 Los Angeles, California. American Medical Informatics Association, pp. 17-21.

Barker, K. & Cornacchia, N. (2000). Using noun phrase heads to extract document keyphrases. *In:* Hamilton, H. J., ed. Advances in Artificial Intelligence. Proceedings 13th Canadian Conference on AI, 2000 Montreal, Canada. pp. 40-52.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *In:* 7th International Conference World WIde Web, WWW'08, 1998 Brisbane, Australia. pp. 107-117.

Church, K. W. & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics,* Vol. 16, No. 1, pp.22–29.

Ding, X. & Liu, B. (2007). The utility of linguistic rules in opinion mining. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval.* Amsterdam, The Netherlands ACM.

Hu, M. & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* Seattle, WA, USA: ACM.

Jain, A. K., Murty, M. N. & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys.,* Vol. 31, No. 3, pp.264-323.

Jin, J. & Liu, Y. (2010). How to interpret the helpfulness of online product reviews: bridging the needs between customers and designers. *SMUC '10 Proceedings of the 2nd international workshop on search and mining user-generated contents* Toronto, Canada: ACM.
Ledeneva, Y., Gelbukh, A. & García-Hernández, R. (2008). Terms Derived from Frequent Sequences for Extractive Text Summarization. *In:* Gelbukh, A. (ed.) *Computational Linguistics and Intelligent Text Processing.*
Lee, T. Y. (2007). Needs-based analysis of online customer reviews. *Proceedings of the ninth international conference on electronic commerce.* Minneapolis, MN, USA: ACM.
Lim, S. C. J. (2011). *A Semantically Annotated Multi-Faceted Ontology Modeling for Supporting Product Family Design.* Doctor of Philosophy, Hong Kong Polytechnic University.
Ling, X., Mei, Q., Zhai, C. & Schatz, B. (2008). Mining multi-faceted overviews of arbitrary topics in a text collection. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining.* Las Vegas, Nevada, USA: ACM.
Liu, M., Shen, W., Hao, Q. & Yan, J. (2009). An weighted ontology-based semantic similarity algorithm for web service. *Expert Systems with Applications,* Vol. 36, No. 10, pp.12480-12490.
Liu, Y. (2010). Beyond opinion mining: How can automatic online opinion analysis help in product design? *6th International Conference on Web Information Systems and Technologies (WEBIST)* Valencia, Spain.
Liu, Y., Loh, H. T. & Lu, W. F. (2007). Deriving taxonomy from documents at sentence level. *In:* Prado, H. A. d. & Ferneda, E. (eds.) *Emerging technologies of text mining: techniques and applications.* Idea Group Inc.
Loh, H. T., Sun, J., Wang, J. & Lu, W. F. (2009). Opinion Extraction from Customer Reviews. *In:* Proceedings of the ASME international design engineering technical conferences & computers and information in engineering conference (IDETC/CIE 2009), August 30-September 2 2009 San Diego, California, USA. ASME, pp.
Markó, K., Hahn, U., Schulz, S., Daumke, P. & Nohama, P. (2004). Interlingual indexing across different languages. *In:* International Conference Recherche d'Information Assistée par Ordinateur, RIAO'04, 2004 Avignon, France. pp. 100-115.
Medelyan, O., Frank, E. & Witten, I. H. (2009). Human-competitive tagging using automatic keyphrase extraction. *In:* 2009 Conference on Empirical Methods in Natural Language Processing, 6-7 August 2009 Singapore. pp. 1318-1327.
Mei, Q., Xin, D., Cheng, H., Han, J. & Zhai, C. (2007). Semantic annotation of frequent patterns. *ACM Transactions on Knowledge Discovery from Data,* Vol. 1, No. 3, pp.11:(1-30).
Mihalcea, R. & Tarau, P. (2004). TextRank–bringing order into texts. *Conference on Empirical Methods in Natural Language Processing.* Barcelona, Spain.
Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM.*
Nguyen, T. D. & Kan, M.-Y. (2007). Keyphrase extraction in scientific publications. *In:* International Conference on Asian Digital Libraries, ICADL'07, 2007 Hanoi, Vietnam. pp. 317-326.
Pasquier, N., Bastide, Y., Taouil, R. & Lakhal, L. (1999). Discovering frequent closed itemsets for association rules. *7th International Conference on Database Theory.*
Porter, M. F. (1980). An algorithm for suffix stripping. *Program: electronic library and information systems,* Vol. 14, No. 3, pp.130-137.
Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C. & Nevill-Manning, C. G. (1999). KEA: practical automatic keyphrase extraction. *In:* Proceedings of the Fourth ACM Conference on Digital Libraries. DL'99., 1999 Berkeley, California, United States. ACM, pp. 254-255.
Wu, Y.-f. B., Li, Q., Bot, R. S. & Chen, X. (2006). Finding nuggets in documents: A machine learning approach. *Journal of the American Society for Information Science and Technology,* Vol. 57, No. 6, pp.740-752.
Ye, S., Chua, T.-S., Kan, M.-Y. & Qiu, L. (2007). Document concept lattice for text understanding and summarization. *Information Processing & Management*, Vol. 43, No. 6, pp.1643-1662.
Zhan, J., Loh, H. T. & Liu, Y. (2009). Gather customer concerns from online product reviews - A text summarization approach. *Expert Systems with Applications,* Vol. 36, No. 2(1), pp.2107-2115.