

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**COLLEGE OF ENGINEERING GUINDY, ANNA UNIVERSITY**

|  |
| --- |
| Summarization of user reviews on e-commerce websites |

|  |
| --- |
| By  Pradeep T - 2012103570  Sai Shibi M R - 2012103056  Manikandan P - 2012103048  Project Guide : Dr. Arockia Xavier Annie, Assistant Professor |

**1.INTRODUCTION:**

**1.1. ABSTRACT:**

E-commerce is a rapidly growing industry at present, with the number of regular users of these sites increasing day by day. A recent survey suggests that more than 80% of users read the user-review forums of a product before they buy it. But the problem with these reviews is that they are not organised. Some users write about the product, while some people write about the service (like delivery, packaging etc.) and for popular products these reviews are big making it hard to read. These product reviews can further be categorized into positive, negative or both. Our aim is to organize the product reviews from all popular e-commerce websites and summarize those reviews, so that users can get a clear idea about the product.

**1.2. PROBLEM DOMAIN:**

Our project falls under the domains “Data Mining” and “Natural Language Processing” .Data Mining is an analytic process to explore data. Natural Language Processing is a field of computer science artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human–computer interaction.

**1.3. PROBLEM STATEMENT:**

To build a system to retrieve users’ feedback (reviews) from e-commerce websites like (Amazon, Flipkart, Snapdeal etc..,) and then to summarize the extracted reviews. The summarized content contains advantages and disadvantages of the product.

**2. LITERATURE REVIEW:**

There are a few research papers which serve as the base idea of the project.

## 2.1. “Recommender systems based on user reviews: the state of the art” Li Chen · Guanliang Chen · Feng Wang

## Advanced text analysis and opinion mining techniques enable the extraction of various types of review elements, such as the discussed topics, the multi-faceted nature of opinions, contextual information, comparative opinions, and reviewers’ emotions. In this article, they provide a comprehensive overview of how the review elements have been exploited to improve standard content-based recommending, collaborative filtering, and preference-based product ranking techniques. The review-based recommender system’s ability to alleviate the well-known rating sparsity and cold-start problems is emphasized. This survey classifies state-of-the-art studies into two principal branches: review-based user profile building and review-based product profile building. In the user profile sub-branch, the reviews are not only used to create term-based profiles, but also to infer or enhance ratings. The product profile can be enriched with feature opinions or comparative opinions to better reflect its assessment quality.

## 2.2. “[Information Extraction from Web Pages](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5284942&newsearch=true&queryText=information%20extraction%20from%20web%20pages)” [Novotny, Róbert](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Novotny%2C%20R%26%23x0F3%3Bbert.QT.&newsearch=true); [Vojtas, P.](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Vojtas%2C%20P..QT.&newsearch=true); Maruscak, Dušan

This paper presents a chain of techniques for extraction of object attribute data from web pages which contain either multiple object data or detailed data about a single object. We discover data regions containing multiple data records, which will be extracted with help of extraction ontology. Furthermore, we present an additional algorithm for detail-page extraction based on the comparison of two HTML subtrees.

## 2.3 “[Thai herb information extraction from multiple websites](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6287734&queryText=information%20extraction%20from%20webpage&newsearch=true)” [Chainapaporn, P.](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Chainapaporn%2C%20P..QT.&newsearch=true); [Netisopakul, P.](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Netisopakul%2C%20P..QT.&newsearch=true)

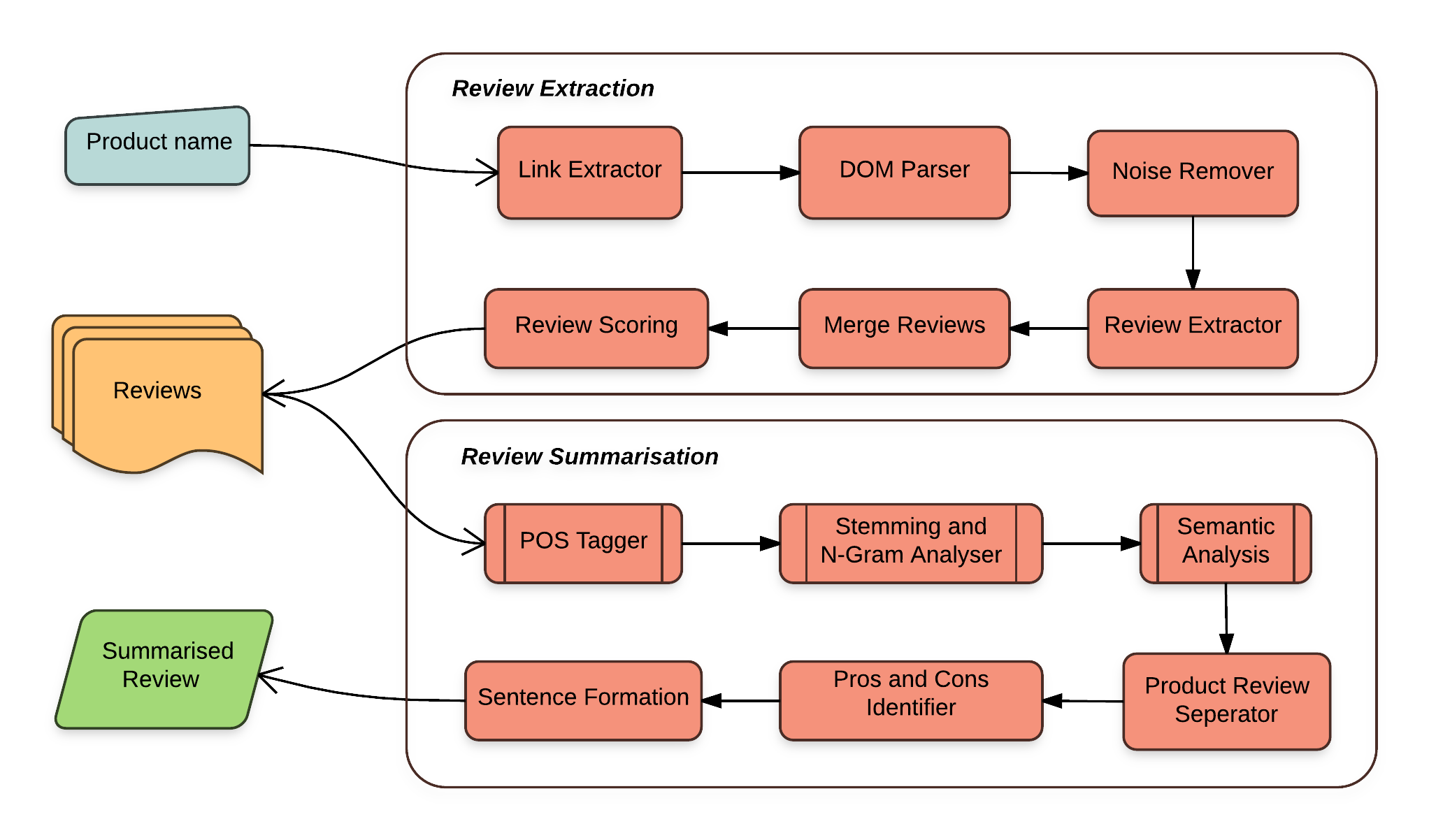
Thai herbs have increasingly gained public attention. Recently, there are a number of Thai herb websites. Each website has similar information but quite different details. For example, some webpages do not provide information indicating which part of Thai herb can treat the specified symptom. In order to collect more complete Thai herb information, we have developed information extraction process to extract Thai herb information from multiple websites. The process employed a HTML parser and file templates to recognize useful information in various webpage formats. Preliminary experiments gave satisfactory precision and recall over 85 percent.

## 2.4. “[Classification and Summarization of Pros and Cons for Customer Reviews](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5284936&newsearch=true&queryText=reviews%20summarization)” [Hu, Xinghua](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Hu%2C%20Xinghua.QT.&newsearch=true); [Wu, Bin](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Wu%2C%20Bin.QT.&newsearch=true)

As e-commerce is becoming more and more popular, the number of customer reviews for online products grows rapidly. For a popular product, there can be hundreds of reviews. This makes it difficult for a potential customer to read all of them in order to get as much information as possible and to make a decision on purchasing. Therefore, a summarization of product reviews would make purchase more convenient and reliable. The conventional way of summarizing a review is to select or rewrite a subset of the original sentences from the review, which is inefficient. In this paper, we propose to summarize all customer’s

reviews of a product as a list of phrases named pros and cons list, which can be perceived and understood at a glance. We employ a score algorithm which considers the strength of a word towards positive or negative orientation to calculate and weigh the sentiment of a sentence. To assess our algorithm, a number of existing classifiers are also presented. Our experimental results show that our Sentence Weight classifier is more accurate and effective than those compared.

**3.BLOCK DIAGRAM:**

****

**4.MODULE DESCRIPTION:**

**4.1. Fetching user reviews:**

Every product contains reviews written by the customers. This module brings all those reviews. Reviews can be categorized into product reviews which are actually the pros or cons about the product, delivery reviews which tells about the service. We do not need to worry about the delivery reviews which are out of our project’s aim. So the module also involves extraction of the product reviews separately which actually involves Natural Language Processing

**4.2. Summarizing reviews:**

This module involves elimination of unnecessary sentences in the extracted reviews. Natural Language is an ocean. So rather than eliminating unnecessary sentences, pick(include) the necessary sentences that actually describes the product. This is done with “phrase match” and “frequency”. Phrase match actually plays a great role in this module. Thus at the end of this module, the summary is available which then pass through another module where the pros and cons are separated.

**4.3. Separating Pros and Cons:**

The project basically aims at separating the pros and cons. In this module it is done. For this, decide whether a particular review is positive review or negative review. The decision is made by matching with set of positive and negative words.

**4.4. Sentence Formation:**

With the set of positive and negative words or reviews found so far, sentences are formed and appended to respective sections (Append positive sentences to pros and negative sentences to cons).

**5. REFERENCES:**

## [1] [Raut V.B](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Raut%2C%20V.B..QT.&newsearch=true) and [Londhe D.D](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Londhe%2C%20D.D..QT.&newsearch=true) :“[Opinion Mining and Summarization of Hotel Reviews](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7065546&newsearch=true&queryText=reviews%20summarization)” [*Computational Intelligence and Communication Networks (CICN), 2014 International Conference on*](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7058462) *Year: 2014*

[2] [Hu, Xinghua](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Hu%2C%20Xinghua.QT.&newsearch=true), [Wu and Bin](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Wu%2C%20Bin.QT.&newsearch=true) : “[Classification and Summarization of Pros and Cons for Customer Reviews](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5284936&newsearch=true&queryText=reviews%20summarization) “ *Web Intelligence and Intelligent Agent Technologies, 2013. WI-IAT '13. IEEE/WIC/ACM International Joint Conferences.*

## [3] [Novotny, Róbert](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Novotny%2C%20R%26%23x0F3%3Bbert.QT.&newsearch=true), [Vojtas,](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Vojtas%2C%20P..QT.&newsearch=true) [Maruscak and Dušan](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Maruscak%2C%20Du%26%23154%3Ban.QT.&newsearch=true) : “[Information Extraction from Web Pages](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5284942&queryText=information%20extraction%20from%20web%20pages&newsearch=true)” [*Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. IEEE/WIC/ACM International Joint Conferences*](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5284806)*.*

## [4] [Chainapaporn P](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Chainapaporn%2C%20P..QT.&newsearch=true) and [Netisopakul P](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Netisopakul%2C%20P..QT.&newsearch=true) : “[Thai herb information extraction from multiple websites](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6287734&newsearch=true&queryText=thai%20herb)” [*Knowledge and Smart Technology (KST), 2012 4th International Conference.*](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6275518)

.