

# Design and Development of an Automated Product Aspect Ranking Framework Using Probabilistic Aspect Ranking Algorithm for Real World Applications

Ayela Shalini<sup>1</sup>, Ms.G.Sangeetha<sup>2</sup>, Prof.Dr.G.Manoj Someswar<sup>3</sup>

<sup>1</sup>M.Tech.(CSE) from Narasimha Reddy Engineering College, Affiliated to JNTUH, Hyderabad, Telangana, India

<sup>2</sup>M.Tech. (CSE), Assistant Professor, Department of CSE, Narasimha Reddy Engineering College, Affiliated to JNTUH, Hyderabad, Telangana, India

<sup>3</sup>B.Tech., M.S.(USA), M.C.A., Ph.D., Principal & Professor, Department Of CSE, Anwar-ul-uloom College of Engineering & Technology, Affiliated to JNTUH, Vikarabad, Telangana, India

## ABSTRACT:

Numerous consumer reviews of products are now available on the Internet. Consumer reviews contain rich and valuable knowledge for both firms and users. However, the reviews are often disorganized, leading to difficulties in information navigation and knowledge acquisition. This research paper proposes a product aspect ranking framework, which automatically identifies the important aspects of products from online consumer reviews, aiming at improving the usability of the numerous reviews. The important product aspects are identified based on two observations: 1) the important aspects are usually commented on by a large number of consumers and 2) consumer opinions on the important aspects greatly influence their overall opinions on the product. In particular, given the consumer reviews of a product, we first identify product aspects by a shallow dependency parser and determine consumer opinions on these aspects via a sentiment classifier. We then develop a probabilistic aspect ranking algorithm to infer the importance of aspects by simultaneously considering aspect frequency and the influence of consumer opinions given to each aspect over their overall opinions. The experimental results on a review corpus of 21 popular products in eight domains demonstrate the effectiveness of the proposed approach. Moreover, we apply product aspect ranking to two real-world applications, i.e., document-level sentiment classification and extractive review summarization, and achieve significant performance improvements, which demonstrate the capacity of product aspect ranking in facilitating real-world applications.

**Keywords:** Probabilistic Aspect Ranking Algorithm; Document-level Sentiment Classification; Automated Product Aspect Ranking Framework; Product Aspect Identification; Extractive Review Summarization

## INTRODUCTION



**Figure 1: Structure of Data Mining**

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data

from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

**Conference Chair:** Prof.Dr.G.ManojSomeswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)



While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. **Generally, any of four types of relationships are sought:**

- **Classes:** Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.
- **Clusters:** Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.
- **Associations:** Data can be mined to identify associations. The beer-diaper example is an example of associative mining.
- **Sequential patterns:** Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

**Data mining consists of five major elements:**

- 1) Extract, transform, and load transaction data onto the data warehouse system.
- 2) Store and manage the data in a multidimensional database system.
- 3) Provide data access to business analysts and information technology professionals.[1]
- 4) Analyze the data by application software.
- 5) Present the data in a useful format, such as a graph or table.

**Different levels of analysis are available:**

- **Artificial neural networks:** Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- **Genetic algorithms:** Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.[2]
- **Decision trees:** Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.[3]
- **Nearest neighbor method:** A technique that classifies each record in a dataset based on a combination of the classes of the  $k$  record(s) most similar to it in a historical dataset (where  $k=1$ ). Sometimes called the  $k$ -nearest neighbor technique.[4]
- **Rule induction:** The extraction of useful if-then rules from data based on statistical significance.[5]
- **Data visualization:** The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

**Characteristics of Data Mining:**

- **Large quantities of data:** The volume of data so great it has to be analyzed by automated techniques e.g. satellite information, credit card transactions etc.

**Conference Chair:** Prof.Dr.G.ManojSomeswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)



- **Noisy, incomplete data:** Imprecise data is the characteristic of all data collection.
- **Complex data structure:** conventional statistical analysis not possible
- **Heterogeneous data stored in legacy systems**

### Benefits of Data Mining:

- 1) It's one of the most effective services that are available today. With the help of data mining, one can discover precious information about the customers and their behavior for a specific set of products and evaluate and analyze, store, mine and load data related to them
- 2) An analytical CRM model and strategic business related decisions can be made with the help of data mining as it helps in providing a complete synopsis of customers
- 3) An endless number of organizations have installed data mining projects and it has helped them see their own companies make an unprecedented improvement in their marketing strategies (Campaigns). [6]
- 4) Data mining is generally used by organizations with a solid customer focus. For its flexible nature as far as applicability is concerned is being used vehemently in applications to foresee crucial data including industry analysis and consumer buying behaviors. [7]
- 5) Fast paced and prompt access to data along with economic processing techniques have made data mining one of the most suitable services that a company seek

### Advantages of Data Mining:

#### 1. Marketing / Retail:

Data mining helps marketing companies build models based on historical data to predict who will respond to the new marketing campaigns such as direct mail, online marketing campaign...etc. Through the results, marketers will have appropriate approach to sell profitable products to targeted customers.

Data mining brings a lot of benefits to retail companies in the same way as marketing. Through market basket analysis, a store can have an appropriate production arrangement in a way that customers can buy frequent buying products together with pleasant. In addition, it also helps the retail companies offer certain discounts for particular products that will attract more customers.[8]

#### 2. Finance / Banking

Data mining gives financial institutions information about loan information and credit reporting. By building a model from historical customer's data, the bank and financial institution can determine good and bad loans. In addition, data mining helps banks detect fraudulent credit card transactions to protect credit card's owner.

#### 3. Manufacturing

By applying data mining in operational engineering data, manufacturers can detect faulty equipments and determine optimal control parameters. For example semi-conductor manufacturers has a challenge that even the conditions of manufacturing environments at different wafer production plants are similar, the quality of wafer are lot the same and some for unknown reasons even has defects. Data mining has been applying to determine the ranges of control parameters that lead to the production of golden wafer. Then those optimal control parameters are used to manufacture wafers with desired quality.[9]

#### 4. Governments

Data mining helps government agency by digging and analyzing records of financial transaction to build patterns that can detect money laundering or criminal activities.

#### 5. Law enforcement:

Data mining can aid law enforcers in identifying criminal suspects as well as apprehending these criminals by examining trends in location, crime type, habit, and other patterns of behaviors.[10]

#### 6. Researchers:

**Conference Chair:** Prof.Dr.G.ManojSomeswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)



Data mining can assist researchers by speeding up their data analyzing process; thus, allowing those more time to work on other projects.

## LITERATURE SURVEY

### 1. LexRank: Graph-based lexical centrality as salience in text summarization

We introduce a stochastic graph-based method for computing relative importance of textual units for Natural Language Processing. We test the technique on the problem of Text Summarization (TS). Extractive TS relies on the concept of sentence salience to identify the most important sentences in a document or set of documents. Salience is typically defined in terms of the presence of particular important words or in terms of similarity to a centroid pseudo-sentence. We consider a new approach, LexRank, for computing sentence importance based on the concept of eigenvector centrality in a graph representation of sentences. In this model, a connectivity matrix based on intra-sentence cosine similarity is used as the adjacency matrix of the graph representation of sentences. Our system, based on LexRank ranked in first place in more than one task in the recent DUC 2004 evaluation. In this paper we present a detailed analysis of our approach and apply it to a larger data set including data from earlier DUC evaluations. We discuss several methods to compute centrality using the similarity graph. The results show that degree-based methods (including LexRank) outperform both centroid-based methods and other systems participating in DUC in most of the cases. Furthermore, the LexRank with threshold method outperforms the other degree-based techniques including continuous LexRank. We also show that our approach is quite insensitive to the noise in the data that may result from an imperfect topical clustering of documents.[11]

### 2. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics

With the rapid growth of the Internet, the ability of users to create and publish content has created active

electronic communities that provide a wealth of product information. However, the high volume of reviews that are typically published for a single product makes harder for individuals as well as manufacturers to locate the best reviews and understand the true underlying quality of a product. In this paper, we reexamine the impact of reviews on economic outcomes like product sales and see how different factors affect social outcomes such as their perceived usefulness. Our approach explores multiple aspects of review text, such as subjectivity levels, various measures of readability and extent of spelling errors to identify important text-based features. In addition, we also examine multiple reviewer-level features such as average usefulness of past reviews and the self-disclosed identity measures of reviewers that are displayed next to a review. Our econometric analysis reveals that the extent of subjectivity, in formativeness, readability, and linguistic correctness in reviews matters in influencing sales and perceived usefulness. Reviews that have a mixture of objective, and highly subjective sentences are negatively associated with product sales, compared to reviews that tend to include only subjective or only objective information. However, such reviews are rated more informative (or helpful) by other users. By using Random Forest-based classifiers, we show that we can accurately predict the impact of reviews on sales and their perceived usefulness. We examine the relative importance of the three broad feature categories: “reviewer-related” features, “review subjectivity” features, and “review readability” features, and find that using any of the three feature sets results in a statistically equivalent performance as in the case of using all available features. This paper is the first study that integrates econometric, text mining, and predictive modeling techniques toward a more complete analysis of the information captured by user-generated online reviews in order to estimate their helpfulness and economic impact.[12]

**Conference Chair:** Prof.Dr.G.ManojSomeswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)



### 3. A survey of text summarization extractive techniques

The existence of the World Wide Web has caused an information explosion. Readers are overloaded with lengthy text documents where a shorter version would suffice. All computer users, be it professionals or novice users, are particularly affected by this predicament. There exists an urgent need for the discovery of knowledge embedded in digital documents. This paper intends to investigate techniques and methods used by researchers for automatic text summarization. Special attention is paid to Bio-inspired methods for text summarization.

### 4. Mining and summarizing customer reviews

Merchants selling products on the Web often ask their customers to review the products that they have purchased and the associated services. As e-commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. For a popular product, the number of reviews can be in hundreds or even thousands. This makes it difficult for a potential customer to read them to make an informed decision on whether to purchase the product. It also makes it difficult for the manufacturer of the product to keep track and to manage customer opinions. For the manufacturer, there are additional difficulties because many merchant sites may sell the same product and the manufacturer normally produces many kinds of products. In this research, we aim to mine and to summarize all the customer reviews of a product. This summarization task is different from traditional text summarization because we only mine the features of the product on which the customers have expressed their opinions and whether the opinions are positive or negative. We do not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization. Our task is performed in three steps: (1) mining product features that have been commented on by customers; (2) identifying opinion sentences in

each review and deciding whether each opinion sentence is positive or negative; (3) summarizing the results. This paper proposes several novel techniques to perform these tasks. Our experimental results using reviews of a number of products sold online demonstrate the effectiveness of the techniques.[13]

### Multiple aspects ranking using the good grief algorithm

We address the problem of analyzing multiple related opinions in a text. For instance, in a restaurant review such opinions may include food, ambience and service. We formulate this task as a multiple aspect ranking problem, where the goal is to produce a set of numerical scores, one for each aspect. We present an algorithm that jointly learns ranking models for individual aspects by modeling the dependencies between assigned ranks. This algorithm guides the prediction of individual rankers by analyzing meta-relations between opinions, such as agreement and contrast. We prove that our agreement based joint model is more expressive than individual ranking models. Our empirical results further confirm the strength of the model: the algorithm provides significant improvement over both individual rankers and a state-of-the-art joint ranking model.

### SYSTEM ANALYSIS

#### EXISTING SYSTEM

Existing techniques for aspect identification include supervised and unsupervised methods. Supervised method learns an extraction model from a collection of labeled reviews. The extraction model, or called extractor, is used to identify aspects in new reviews. Most existing supervised methods are based on the sequential learning (or sequential labeling) technique. On the other hand, unsupervised methods have emerged recently. They assumed that product aspects are nouns and noun phrases. The approach first extracts nouns and noun phrases as candidate aspects. The occurrence frequencies of the nouns and noun phrases are counted, and only the frequent ones are kept as aspects.[14]

**Conference Chair:** Prof.Dr.G.ManojSomeswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)

### DISADVANTAGES OF EXISTING SYSTEM

- The reviews are disorganized, leading to difficulties in information navigation and knowledge acquisition.
- The frequency-based solution is not able to identify the truly important aspects of products which may lead to decrease in efficiency of the review.

### PROPOSED SYSTEM

- We propose a product aspect ranking framework to automatically identify the important aspects of products from numerous consumer reviews.
- We develop a probabilistic aspect ranking algorithm to infer the importance of various aspects by simultaneously exploiting aspect frequency and the influence of consumers' opinions given to each aspect over their overall opinions on the product.
- We demonstrate the potential of aspect ranking in real-world applications. Significant performance improvements are obtained on the applications of document-level sentiment classification and extractive review summarization by making use of aspect ranking.

### ADVANTAGES OF PROPOSED SYSTEM

- Identifies important aspects based on the product, which increases the efficiency of the reviews.
- The proposed framework and its components are domain-independent

### SYSTEM DESIGN

#### SYSTEM ARCHITECTURE:

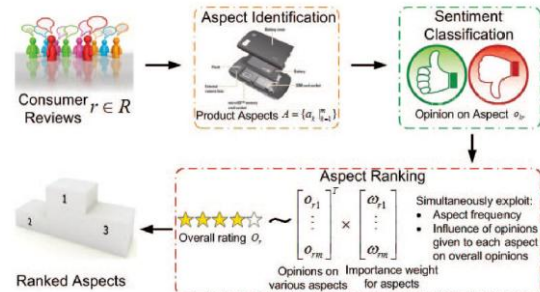
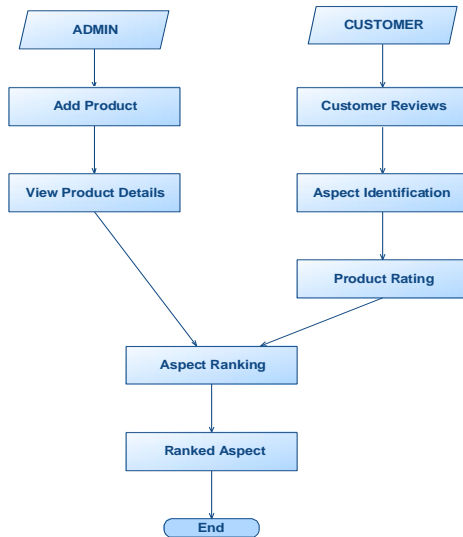


Figure 2: System Architecture

#### DATA FLOW DIAGRAM:

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.



**Figure 3: Data Flow Diagram**

## UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

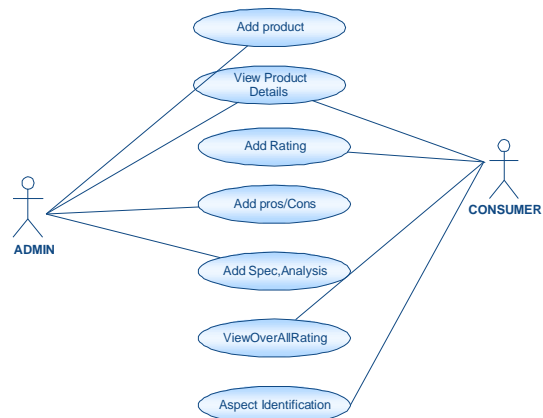
## GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

## USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



**Figure 4: Use Case Diagram**

**Conference Chair:** Prof.Dr.G.ManojSomeswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)

**CLASS DIAGRAM**

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

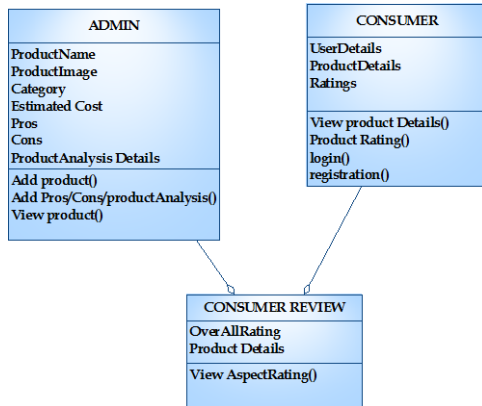


Figure 5: Class Diagram

**SEQUENCE DIAGRAM**

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

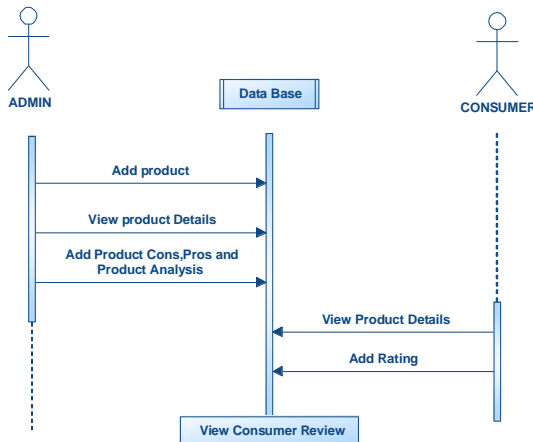


Figure 6: Sequence Diagram

**ACTIVITY DIAGRAM**

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

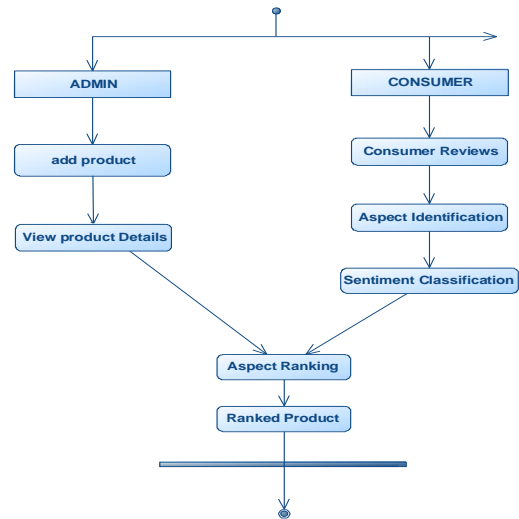


Figure 7: Activity Diagram

**SYSTEM STUDY**

**FEASIBILITY STUDY**

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ◆ ECONOMICAL FEASIBILITY
- ◆ TECHNICAL FEASIBILITY
- ◆ SOCIAL FEASIBILITY

**ECONOMICAL FEASIBILITY**

This study is carried out to check the economic impact that the system will have on the





organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased. [15]

### TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

### SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

### INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay,

avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

### OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

### OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

**Conference Chair:** Prof. Dr. G. Manoj Someswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)



1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- ❖ Convey information about past activities, current status or projections of the
- ❖ Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.

## **IMPLEMENTATION**

### **MODULES:**

- Product aspects
- Aspect Ranking
- Product Aspect Identification
- Consumer Review/Rating

### **MODULES DESCRIPTION**

#### **Product aspects**

- Generally, a product may have hundreds of aspects. For example, iPhone 3GS has more than three hundred aspects (see Fig. 1), such as "usability," "design," "application," "3G network."
- Identifying important product aspects will improve the usability of numerous reviews and is beneficial to both consumers and firms.
- Consumers can conveniently make wise purchasing decision by paying more attentions to the important aspects, while firms can focus on improving the quality of these aspects and thus enhance product reputation effectively.

#### **Aspect Ranking**

**Conference Chair:** Prof. Dr. G. Manoj Someswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)

- We propose a product aspect ranking framework to automatically identify the important aspects of products from numerous consumer reviews.
- We develop a probabilistic aspect ranking algorithm to infer the importance of various aspects by simultaneously exploiting aspect frequency and the influence of consumers' opinions given to each aspect over their overall opinions on the product.
- We demonstrate the potential of aspect ranking in real-world applications. Significant performance improvements are obtained on the applications of document-level sentiment classification and extractive review summarization by making use of aspect ranking.

#### **Product Aspect Identification**

- For the Pros and Cons reviews, we identify the aspects by extracting the frequent noun terms in the reviews.
- Previous studies have shown that aspects are usually nouns or noun phrases, and we can obtain highly accurate aspects by extracting frequent noun terms from the Pros and Cons reviews.
- For identifying aspects in the free text reviews, a straightforward solution is to employ an existing aspect identification approach.

#### **Consumer Review/Rating**

- Consumer reviews contain rich and valuable knowledge for both firms and users. However, the reviews are often disorganized, leading to difficulties in information navigation and knowledge acquisition.
- This article proposes a product aspect ranking framework, which automatically identifies the important aspects of products from online consumer reviews, aiming at improving the usability of the numerous reviews.
- The important product aspects are identified



based on two observations:

1) The important aspects are usually commented on by a large number of consumers and

2) Consumer opinions on the important aspects greatly influence their overall opinions on the products.

### SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

### TYPES OF TESTS

#### Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

#### Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfactory, as shown

by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

#### Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

#### System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

#### White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is used to test areas that cannot be reached from a black box level.

**Conference Chair:** Prof.Dr.G.ManojSomeswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)



### Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

### Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

### Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

### Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

### Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

### Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

**Conference Chair:** Prof.Dr.G.ManojSomeswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)

**Test Results:** All the test cases mentioned above passed successfully. No defects encountered.

### Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

**Test Results:** All the test cases mentioned above passed successfully. No defects encountered.

### CONCLUSION

In this research paper, we have proposed a product aspect ranking framework to identify the important aspects of products from numerous consumer reviews. The framework contains three main components, i.e., product aspect identification, aspect sentiment classification, and aspect ranking. First, we exploited the *Pros* and *Cons* reviews to improve aspect identification and sentiment classification on free-text reviews. We then developed a probabilistic aspect ranking algorithm to infer the importance of various aspects of a product from numerous reviews. The algorithm simultaneously explores aspect frequency and the influence of consumer opinions given to each aspect over the overall opinions. The product aspects are finally ranked according to their importance scores. We have conducted extensive experiments to systematically evaluate the proposed framework. The experimental corpus contains 94,560 consumer reviews of 21 popular products in eight domains. This corpus is publicly available by request. Experimental results have demonstrated the effectiveness of the proposed approaches. Moreover, we applied product aspect ranking to facilitate two real-world applications, i.e., document-level sentiment classification and extractive review summarization. Significant performance improvements have been obtained with the help of product aspect ranking.



## REFERENCES

- [1] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *J. Neural Parallel Scientific Comput.*, vol. 11, no. 4, pp. 351–368, 2003.
- [2] C. C. Chang and C. J. Lin. (2004). *Libsvm: A library for support vector machines* [Online]. Available:<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [3] G. Carenini, R. T. Ng, and E. Zwart, "Multi-document summarization of evaluative text," in *Proc. ACL*, Sydney, NSW, Australia, 2006, pp. 3–7.
- [4] China Unicom 100 Customers iPhone User Feedback Report, 2009.
- [5] ComScore Reports [Online]. Available: [http://www.comscore.com/Press\\_events/Press\\_releases](http://www.comscore.com/Press_events/Press_releases), 2011.
- [6] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *Proc. WSDM*, New York, NY, USA, 2008, pp. 231–240.
- [7] G. Erkan and D. R. Radev, "LexRank: Graph-based lexical centrality as salience in text summarization," *J. Artif. Intell. Res.*, vol. 22, no. 1, pp. 457–479, Jul. 2004.
- [8] O. Etzioni *et al.*, "Unsupervised named-entity extraction from the web: An experimental study," *J. Artif. Intell.*, vol. 165, no. 1, pp. 91–134. Jun. 2005.
- [9] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 10, pp. 1498–1512. Sept. 2010.
- [10] V. Gupta and G. S. Lehal, "A survey of text summarization extractive techniques," *J. Emerg. Technol. Web Intell.*, vol. 2, no. 3, pp. 258–268, 2010.
- [11] W. Jin and H. H. Ho, "A novel lexicalized HMM-based learning framework for web opinion mining," in *Proc. 26th Annu. ICML*, Montreal, QC, Canada, 2009, pp. 465–472.
- [12] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. SIGKDD*, Seattle, WA, USA, 2004, pp. 168–177.
- [13] K. Jarvelin and J. Kekalainen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inform. Syst.*, vol. 20, no. 4, pp. 422–446, Oct. 2002.
- [14] J. R. Jensen, "Thematic information extraction: Image classification," in *Introductory Digit. Image Process.*, pp. 236–238.
- [15] K. Lerman, S. Blair-Goldensohn, and R. McDonald, "Sentiment summarization: Evaluating and learning user preferences," in *Proc. 12th Conf. EACL*, Athens, Greece, 2009, pp. 514–522.

**Conference Chair:** Prof. Dr. G. Manoj Someswar, Director General, Global Research Academy, Hyderabad, Telangana, India.

Papers presented in Conference can be accessed from [www.edupediapublications.org/journals](http://www.edupediapublications.org/journals)