Bachelor Degree Project

Spring 2012

School of Health and Society
Department Design and Computer Science

Inspirational Filter Application for Fashion E-Commerce

Writer
Brian Dewangga

Instructor
Christian Andersson

Examiner
Eric Chen

School of Health and Society
Department Design and Computer Science
Kristianstad University College
SE-291 88 Kristianstad
Sweden

Author, Program and Year:
Brian Dewangga, Computer Software Development, 2009

Instructor:
Christian Andersson, HKr
Dr. Jesper Brøndum, Chief Technology Officer, Boozt Fashion AB

Examination:
This graduation work on 15 higher education credits is a part of the requirements for a *Degree of Bachelor in Computer Software Development* (as specified in the English translation).

Title:
Inspirational Filter Application for Fashion E-Commerce

Abstract:
Electronic commerce is becoming a growing industry as information technology has been gradually diversifying the business industry. As a result, there is greater competition between online businesses to attract and retain customers. To overcome this challenge, it is proposed that by implementing a new Inspirational Filter Application to a company's web store's filter will increase the growth of returning costumers. This new application is based on loose filters, which goes beyond traditional sorting mechanisms based on color and size. This technology expands searches into inspirational categories such as occasion, material and style, allowing the customer to online shop with greater variety and efficiency. By developing this application alongside with Scandinavian web shop, Boozt Fashion AB, the success rate is tracked after the establishment of this application.

Language:
English

Approved by:

_____

Eric Chen Date
Examiner

Table of Contents Page

1 Introduction

1.1 Background

1.1.1. E-Commerce

For decades the Internet has made a significant impact on everyday human life. With the advancements in technology, social interaction has been greatly transformed. Information Technology has significantly influenced the realm of business and as a result, e-commerce has enabled a more efficient way of life. E-commerce is the business process of selling and buying goods and services through on-line communications. Consumerism is vastly expanding via the Internet, and has become further diversified by e-commerce. Individuals now heavily rely on the Internet to research information on goods and services before making a final purchase. Whether the final purchase is made directly in person or online, the Internet has greatly influenced the decision making process.

1.1.2. Socially Fashionable

Fashion shopping environments fundamentally involve social and interpersonal activities [1], [2]. Social and interpersonal aspects are closely associated and may often influence an individual's style. However, fashion plays a primary role when one decides what to wear, how to wear it and why they have chosen to wear a specific item of clothing. Multiple variables may influence what type of clothing is appropriate for a specific occasion. Hence, the social shopping aspect plays a crucial role in the fashion e-commerce world. Fashion e-commerce has made a huge impact in today's consumerism world as it has made online shopping easy for everyone of all ages.

1.1.3. Problem of Retaining Customers in Online Shopping

There are a vast number of vendors available on the Internet and are giving marketers an increased challenge to obtain and keep customers. In today's fierce competition among businesses to acquire and retain customers, it has become one of the main issues in the fashion e-commerce industry. In agreement with Rosenberg and Crepiel, the cost to attract a new customer can be as much as six times the cost to keep a current one [3].

1.1.4. Hard Filters versus Loose Filters

Filtering is a methodology widely used within the entire web. Filters are used in online shopping to "filter out" unwanted search results. To overcome this, "filters" are needed in order to rule out the results that do not match. For example, by selecting: women clothing items, pants category, and size medium. The results given are only pants in size medium, which are intended for women. This is called using hard filters, which is commonly used on web shops.

To differentiate the norm of using hard filters, the introduction of loose filtering will be implemented which will be the focus of my project. Loose filtering is a way of filtering through a more fuzzy approach. Unlike hard filters which have strict controls, loose filters can be used to include products which are not clearly defined.

1.2 Context

Together with the company I have been working for, Boozt Fashion AB, we are developing an Inspirational Filter Application. This highly responsive filter application will allow the user to narrow down a product-listing page based on an inspirational input across various brands and categories. This application is centered on loose filtering which allows an individual to narrow down item, which may be vague, and not clearly defined. The filter selection will be based on a fuzzy clustering of the products such as: product description, material, colors, etc., instead of solely using hard filters. This new concept uses filters for the type of occasion and style the customer wishes to seek for. For instance, the different types of occasions include: Date Night, Everyday Casual, Dinner Party, and Outdoor Adventure. Different fashion styles may include: Digital Prints, Classic, Contemporary, and Flirty. This form of filtering will allow a shopper to be inspired by certain styles which fall under a specific category allowing shopping to become more efficient.

1.3 Aim and Purpose

By adopting an Inspirational Filter Application, this new cutting edge technology will allow the possibility to overcome the challenge of maintaining existing customers. This new concept allows an individual who is online shopping to filter various clothing by selecting different occasions and styles. This inspiration filter can start a revolution on how current and future web shops categorize products. Other competitor web shops have not sought this idea out, and the aim of this project is to be the first web shop to implement such a new way of categorizing products under loose filtering. The objective of my project is to streamline customers to have an unforgettable experience shopping on a web shop so the growth of returning customers can drastically increase.

1.4 Presentation of the Company

Boozt Fashion AB is a full-service e-commerce company that is dedicated to helping new developing brands build a profitable on-line business in Europe. The company's has proven expertise in selling to European customers making it a one-stop shop for selling products across all of Europe. Boozt's market-leading e-commerce platform is designed to maximize Web sales while giving customers a simple and satisfying shopping experience.

Boozt has a web shop found online at Boozt.com. It offers a wide variety of fashion products through an assortment of brands for women, men, and children. With a carefully selected selection of clothing & footwear including dresses, skirts, jeans, bags and shoes for women, men's shirts, jeans, tops and underwear we aim to have something for everyone. It is the leading Scandinavian and International clothing and jewelry brands including Vero Moda, Jack & Jones, Soaked in Luxury, and Vila Clothes.

1.5 Role at Boozt Fashion AB

I was previously working as an Operations Assistant in the Operations department at Boozt Fashion AB. I was recently transferred to the Product team, and I have been with this company for a year and a half. Growing with the company, I have decided to undergo my degree project at Boozt Fashion AB and intertwine my skills as an aspiring software developer with the fashion e-commerce industry. To incorporate what I have learned at the University of Kristianstad, I joined the Platform team at Boozt Fashion AB in which suited the type of work I have been studying.

The Inspirational Filter Application was created after sitting down with Chief Technology Officer, Jesper Brøndum. I have been given a great opportunity to carry out my degree work at Boozt Fashion AB. The decision for the company to hand over this project for me to manage had been a role which was new to me. My new role which I had now inherited as project manager for the Inspirational Filter Application project had expectations of me to attend status meetings and workshops that the platform team has organized. This was vital in my learning and knowledge towards my degree project as to working in a real world working environment.

1.6 Report Organization

The report is divided up into seven chapters. The whole project is presented in the analysis, design, and implementation chapters.

The first chapter introduces the project's background, aim and purpose, presentation of the company, and my role in the company.

The second chapter is the incorporation of Project Management. It shows how I have planned for the entire project and which methods I have undertaken.

The third chapter is where the Analysis of product data begins. It tackles the selection process of product data as well as includes the database management, PHP introduction, and stop word definition.

The fourth chapter illustrates the Design of how product data is placed through PHP scripts to achieve certain formats to comply with further implementations.

The fifth chapter demonstrates the Implementation of how everything is placed together. The technique of data mining is used with Weka, software for our data clustering needs.

The sixth chapter sums up our Conclusions to revisit our aim and purpose to recap our analysis, designs, and implementations.

The last chapter explains further work to be done with including customer purchase history with the concept of collaborative filters.

1.7 Acknowledgements

I would like to express my heartfelt gratitude to the staff at Boozt. I would like to first thank Chief Technology Officer, Jesper Brøndum, for his patience, dedication, and guidance over this project. I am grateful for the opportunity to have worked under his direction and am confident that the skills that I have gained at Boozt will benefit me throughout my career. I would also like to thank Frontend Developer, Dusan Novakovic, for his support in taking the time to teach me the programming language, PHP. Finally, I would like to thank Data Manager, Jonny Johansson, for his commitment and assistance with this project. He provided me access to the data warehouse and educating fundamental knowledge of product data. I could not have gotten through this process without him.

I would also like to thank my instructor, Christian Andersson, for motivating me to conduct this research. Being able to work around my schedule and have meetings at odd times helped me so much. His advice and generosity regarding this project is greatly appreciated.

I would like to express special thanks to my family for inspiring me to do the best that I can personally and professionally. They have been an incredible support system to me. I am evermore grateful for their nonstop encouragement, and I hope to honor them in all I do. I am very fortunate to have so many people who care about me.

1.8 List of Abbreviations

UI – User Interface

SQL – Structured Query Language

API – Application Programming Interface

PHP- PHP: Hypertext Preprocessor

WAMP – Windows,Apache, MySQL, PHP

e-commerce – electronic commerce

CSV – Comma-Separted Values

EM – Expectation Maximation

2 Project Management

2.1 Introduction

One of the most important tools in Software Engineering is to deploy project management concepts**.** Charting methods such as the use of a Gantt chart have been developed as tools to create a graphic representation of a project plan and its current status.

2.2 Data Sources

In the beginning of the degree project, data sources were identified. Shown in Figure 2.2.1, the three main headings are Boozt Data Sources, Inspirational Filter, and Taxonomy Module. The figure illustrates Boozt Data Sources and the Taxonomy Module pointing inwards to the Inspirational Filter. This is the actual application we will be working towards. Before this is possible, collection of data is needed from the Boozt Data Sources. Then, organization of the data is to be sought out by the Taxonomy Module. Afterwards, collection and organization of the data is to be put together for the Inspirational Filter Application to come in place.

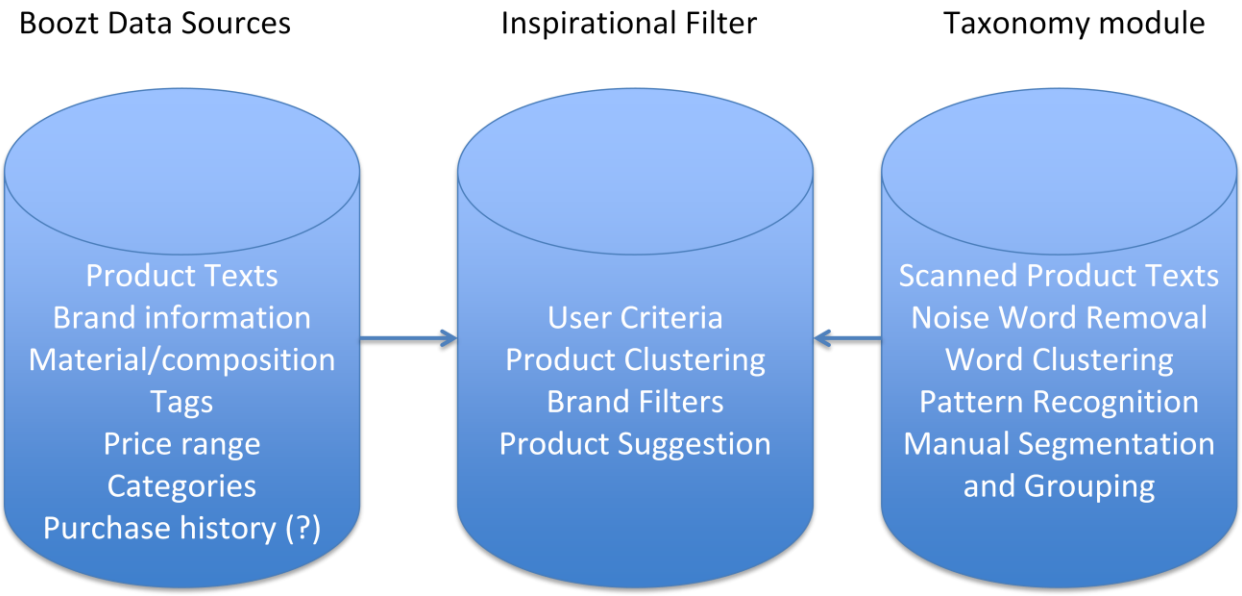

Figure 2.2.1 Data Sources Model

2.3 Modular Overview

Looking at the project into categories, Figure 2.2.1 is transformed to create Figure 2.3.1 as a modular overview. The same headings apply but within each heading, smaller parts are independently created for multiple functionalities.

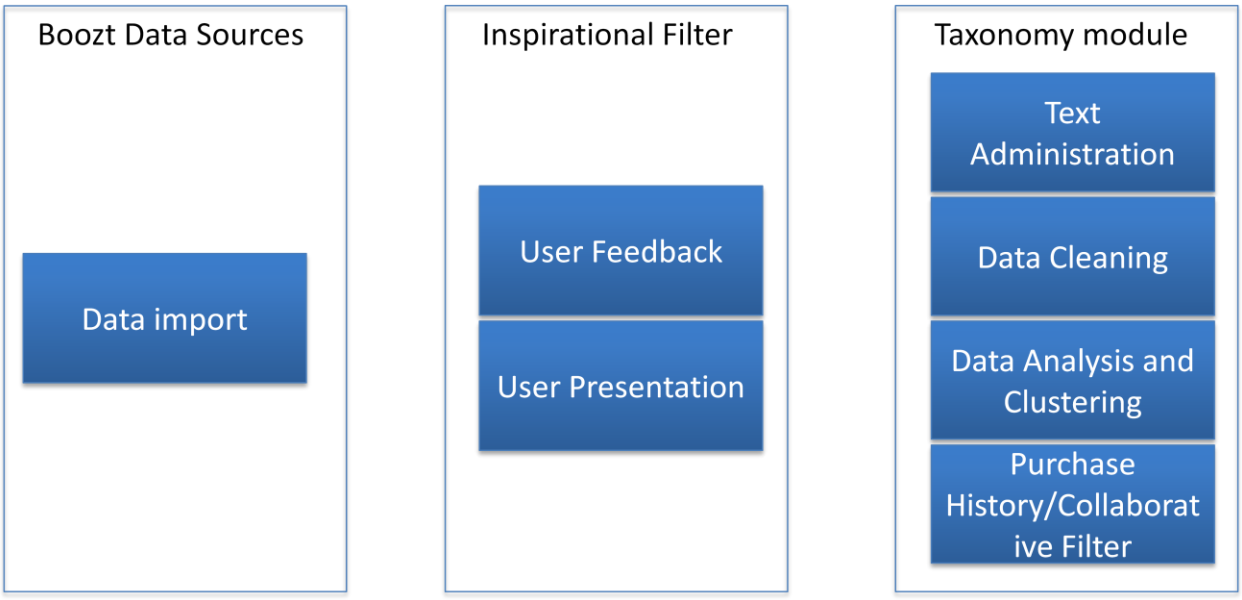

Figure 2.3.1 Modular Overview

2.4 Overall Project Plan

In relation with Figure 2.2.1 and Figure 2.3.1, an overall project plan is made in Figure 2.4.1. Topics taken from the modular overview, they are placed in a monthly timeline to show the overall project plan.

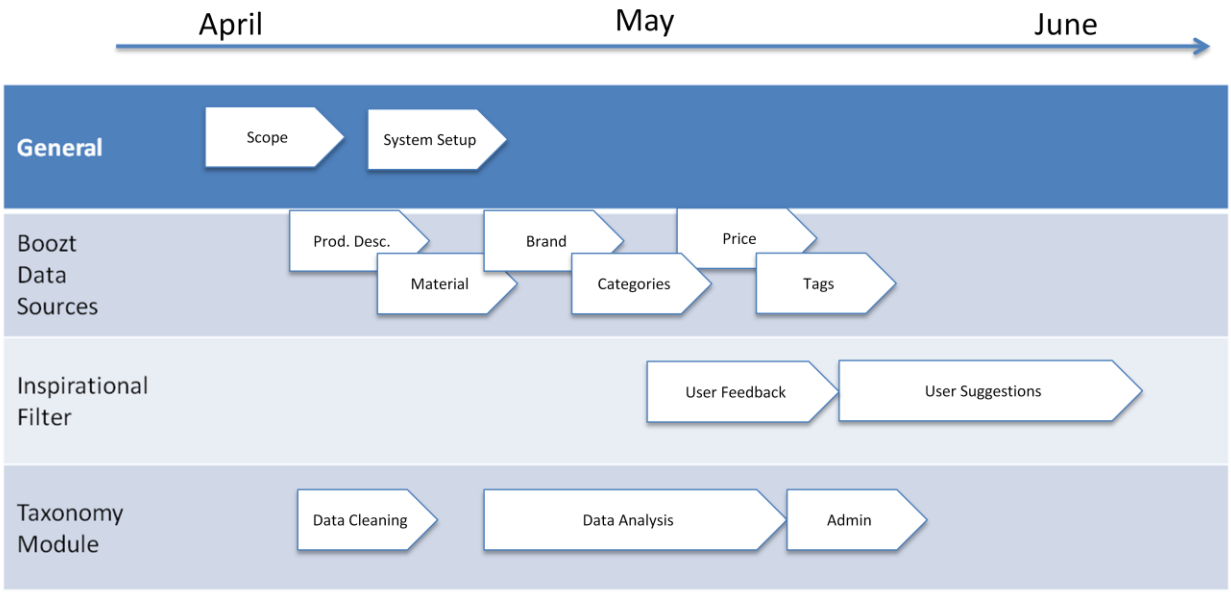

Figure 2.4.1 Overall Project Plan

The choice of charting method used is a Gantt chart. This will provide a graphical illustration of the schedule that helps plan, coordinate, and track specific tasks in a project. To create the Gantt chart shown in Figure 2.4.2, Microsoft Excel 2007 was used. The construction of this Gantt chart was made at the beginning of the project where deadlines were set and what specific tasks needed to be carried out. Of course all the tasks listed were to be completed by myself. The legend states that green meant tasks were completed on time, yellow meant that the tasks were completed earlier than scheduled, and red meant that the tasks were delayed. To keep the entire Platform team in the loop, a presentation of my current status was made every week on Mondays.

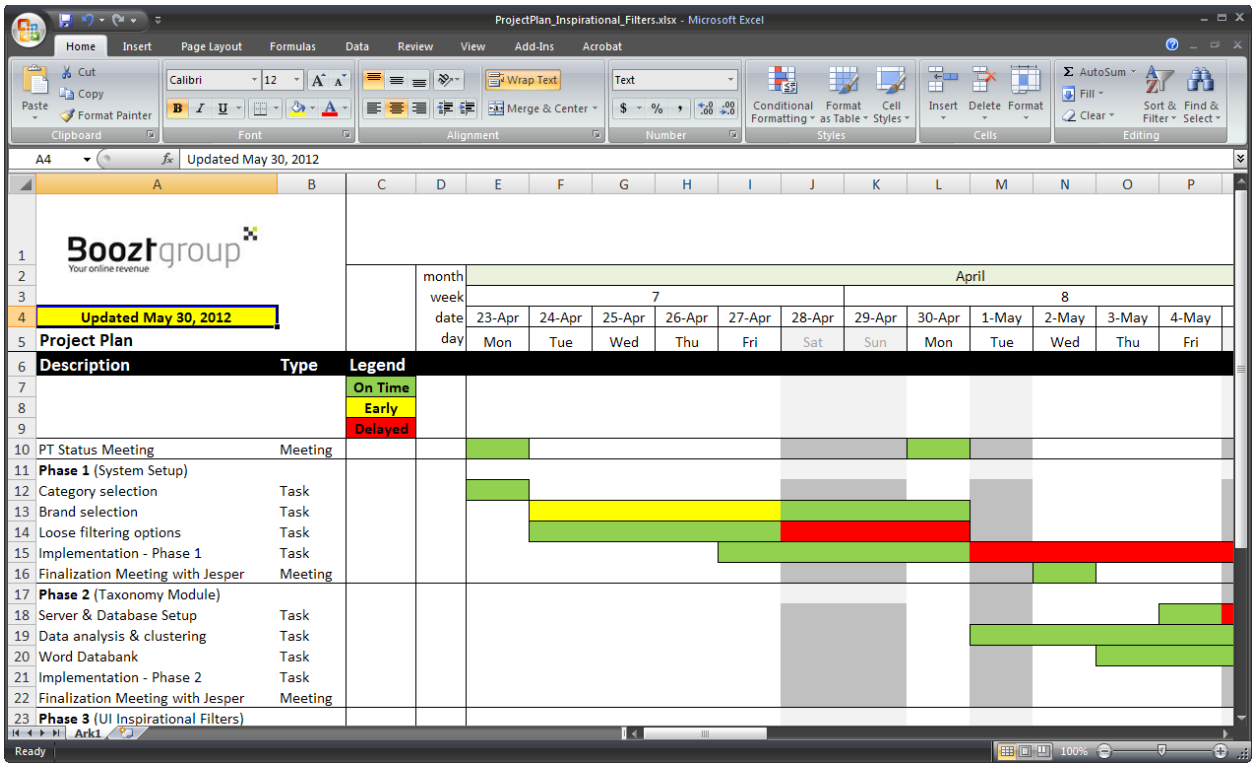

Figure 2.4.2 Project Gantt chart

For online access of my progress, an online project management and collaboration tool is used called activeCollab. This way anyone from Boozt can track my progress for the Inspirational Filter Application. In activeCollab, the tasks in the Gantt chart are called tickets. Creating a ticket includes stating the priority level, due date, and assignees. When all the tickets have been completed in a phase, a milestone is created. At times when working away from the office, the supervisor can track my development in the project.

2.5 Phase Breakdown

2.5.1 General Breakdown

Three phases are made to fulfill the completion of the project. Each phase has an implementation task which is a recap of all the tasks in the corresponding phase. To finalize each phase, a meeting is scheduled with my supervisor at the company, Jesper Brøndum.

2.5.2 Phase One – System Setup

Phase one is the system setup which includes the tasks, category selection, brand selection, and loose filtering options. Category selection is the process of choosing which clothing categories should be looked at for this inspirational filter application. Brand selection is the process of choosing which brands will be looked at. Lastly, to have a sense of what the filters may be like, a certain approach was made when selecting which occasions and styles to use as filters.

2.5.3 Phase Two – Taxonomy Module

Phase two is the taxonomy module which includes the tasks, server & database setup, data analysis & clustering, and word databank. Setting up the server & database had to be done by a colleague so I could have access. Data analysis & clustering

3 Analysis

3.1 Product Data

Product data is what fulfills the actual product by means of its product name, product description, brand, and category. The selection of product data is an essential step before tackling the inspirational filter application. The Boozt web shop offers thousands of products ranging from clothing to accessories and to shoes. To limit this range, a selection process needs to be done by category and brand.

3.1.1 Category Selection

Products are categorized amongst the web shop to have structure. Choosing all the categories will result in too many products to work with. Therefore, six categories are focused upon.

These categories are:

1. Dresses
2. Skirts
3. Tops
4. Knitwear
5. Shirts & Blouses
6. Outerwear

3.1.2 Brand Selection

Numerous brands are featured on the Boozt web shop which makes it very diverse when finding that right outfit. The same problem exists when choosing the categories. We cannot pick all brands or else again, too many products will be shown.

These brands are:

1. Esprit (Esprit - Bodywear Woman, Esprit - Casual, Esprit- Collection, Esprit - EDC Woman)
2. Vila
3. Day
4. Vero Moda
5. BZR
6. InWear
7. French Connection
8. Only
9. Signal
10. Soaked In Luxury

3.1.3 Filter Specification

Keeping the project name intact, inspirational filters come from the inspiration of choosing what clothes to wear for a certain occasion and style. This is crucial in choosing what occasions and styles will be used because based on customer preferences; this can be the deciding factor if the customer will be a returning one or not.

The occasion filters are:

1. Date Night
2. 9 to 5
3. Summer Weekend
4. Everyday Casual
5. Dinner Party
6. Outdoor Adventure

The style filters are:

1. Luxurious Lace
2. Horizontal Stripes
3. Digital Prints
4. Modern
5. Classic
6. Comfort
7. Contemporary
8. Elegant
9. Edgy
10. Flirty
11. Layer It
12. Sheer
13. Chic
14. Polka Dot

3.2 Product Data Specification

Once the product data that will be looked at is chosen regarding the categories and brands, the actual product data can be looked at individually. The most important data we need are the individual product descriptions. What makes products on Boozt unique is that each and every product is manually written by the product team staff. Since the product descriptions are what differentiate from product A with product B, this is what will be used as the core part of having inspirational filters.

3.3 Database

Acquiring database access from a company can be a problematic issue as it contains sensitive information. This contains customer information such order and purchase history. For the purpose of the inspirational filter application, no information is such needed. Therefore, the essential product data is cloned to a dedicated database for the use of this project.

MySQL is an open source relational database management system that uses SQL for processing data in a database. API's are supported for the language PHP in which will be further discussed in this chapter.

3.3.1 Online vs. Offline Database Access

The cloned database reserved for the inspirational filter application project is to be accessed only within the office network. This database runs dynamically correlating with the live database used for the web shop.

For offline database access, the database is cloned yet again and zipped into an SQL file. WampServer is a Windows web development environment which allows you to create applications with PHP and a MySQL database. This will act as the web server run by Apache. The SQL file is imported into phpMyAdmin where we can manage the database.

3.4 PHP

PHP has become one of the most widely used programming languages, on account of its popularity for web applications [4]. Through it's widely spread popularity, the purpose for using PHP is because it is a chance to learn and try out a new programming language. In this case, command line scripting is the main area looked at as simple text processing tasks can be carried out efficiently and effectively. Data in CSV format can be easily read and written through file arguments in PHP. Product data is closely looked in CSV format.

3.4.1 Command Line Usage

Simple text processing tasks will be run using PHP scripting. A method best used for running through product data quickly and retrieve wanted results. No heavy application is needed for such tasks needed for this project. Instead, performing operations through the command line through a PHP script had outputted the correct results.

3.5 Stop Words

A stop word is a commonly used word (such as "the") that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query [6]. Stop words are introduced due to the importance of product descriptions which defines each product. Certain keywords will be looked at in order for clustering to happen. An example product description is: Dress from Soaked in Luxury in an elegant matte finish. It has short sleeves and a defined waistline. The dress closes with a full front zipper. The stop words listed here would be *from, in, an, it, has, and, a, the, and with*. These stop words are not needed because we have no use for them but the words that are not stop words, those will be what determine what clusters we get.

4 Design

4.1 Filtering Mechanism

Figure 4.1.1 demonstrates how the filtering mechanism will work with the product data that contains product descriptions being extracted from the database to output the keywords needed to perform our clustering algorithm.

The product data extracted from the database with product descriptions is run through a list of stop words using a Stop Words PHP script. The Stop Words PHP script will rule out the stop words listed and output the keywords for each product.

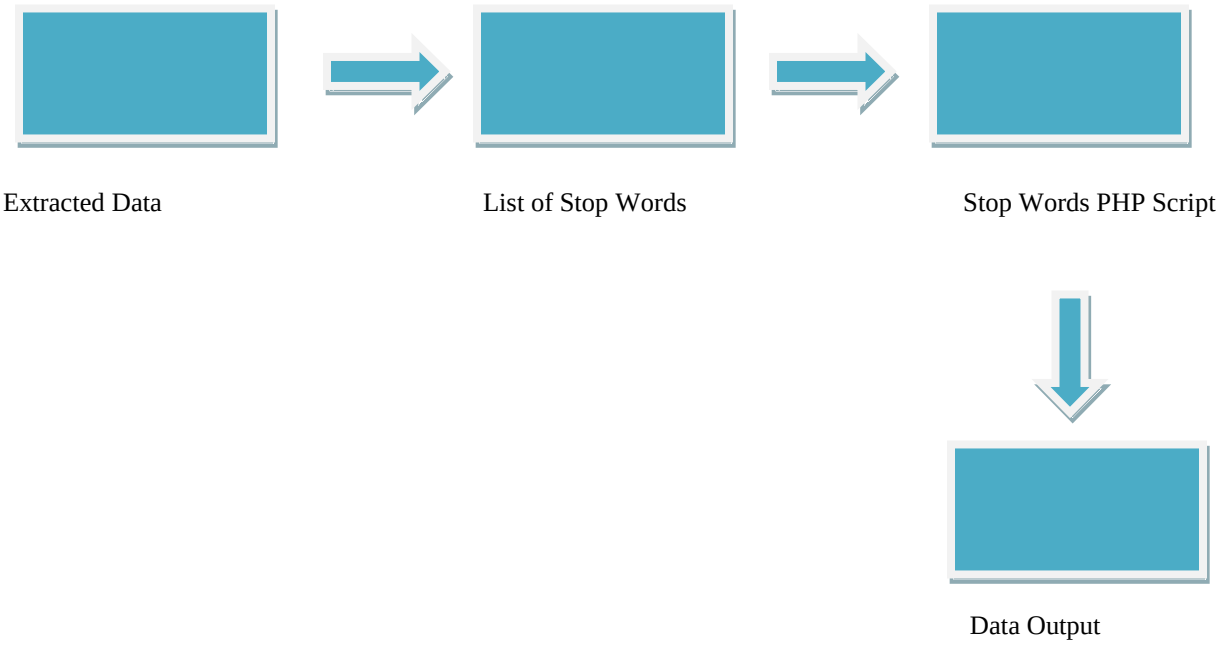

Figure 4.1.1 Flow Chart of Product Data being cross referenced with Stop Words

Once the stop words have been removed, the data needs to be prepared in a certain format. Figure 4.1.2 shows the flow chart of what the next step needs to be done. Keywords leftover are used as tags on the column level and product id number along with their keywords on the row level. This prepared data format is run through a Binary PHP Script and outputs a 1 if the keyword has been found under each tag column and outputs a 0 if it is not found.

Figure 4.1.2 Flow Chart of Product Data being cross referenced with Binary PHP Script

5 Implementation

5.1 Data Mining

The Webopedia defines data mining as a class of database applications that look for hidden patterns in a group of data that can be used to predict future behavior [7]. This is a concept in which will play an imperative role as data is dug deeper to fully satisfy results. As mentioned earlier, product descriptions are the core to being able to group similar products. With these groupings, form clusters in which are possible to analyze. Through this, Weka software is used to develop machine learning techniques and to apply them to real-world data mining problems.

5.2 Weka

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes [4]. Binary data is chosen to look at how the product data corresponds with the tags because the approach in this step is at a static state.

After the binary data is run through the Weka software, an algorithm is needed to be chosen. In this case, K-Means Algorithm is selected for a first try.

5.3 K-Means Clustering Algorithm

K-Means is a straightforward clustering algorithm for the purpose to split up a set of objects according to their attributes and features. Given *k* clusters, where *k* is a user-defined constant, the

data set is grouped into these *k* clusters. The procedure of this algorithm starts by assigning objects within our given data set into a random cluster. Following this, the algorithm iteratively goes through each cluster shifting objects from cluster to cluster depending on if they are similar from one another.

Weka introduces the K-Means Algorithm by naming it SimpleKMeans in the application under the Cluster tab. By selecting, SimpleKMeans, the default number of clusters is two. To relate clusters to the study, a cluster will represent a certain type of occasion and style.

To start off, twelve clusters are chosen and the data set is run iteratively through the K-Means Algorithm. Figure 5.2.1 displays a visualization of the cluster assignments. Shown here are the different color coded clusters and objects placed according to the clusters similarities. Objects are marked with a colored "x" and running twelve clusters does not work because the visual cluster assignments are scattered and no clear definition can be seen. Therefore, twelve clusters are too many to put into practice.

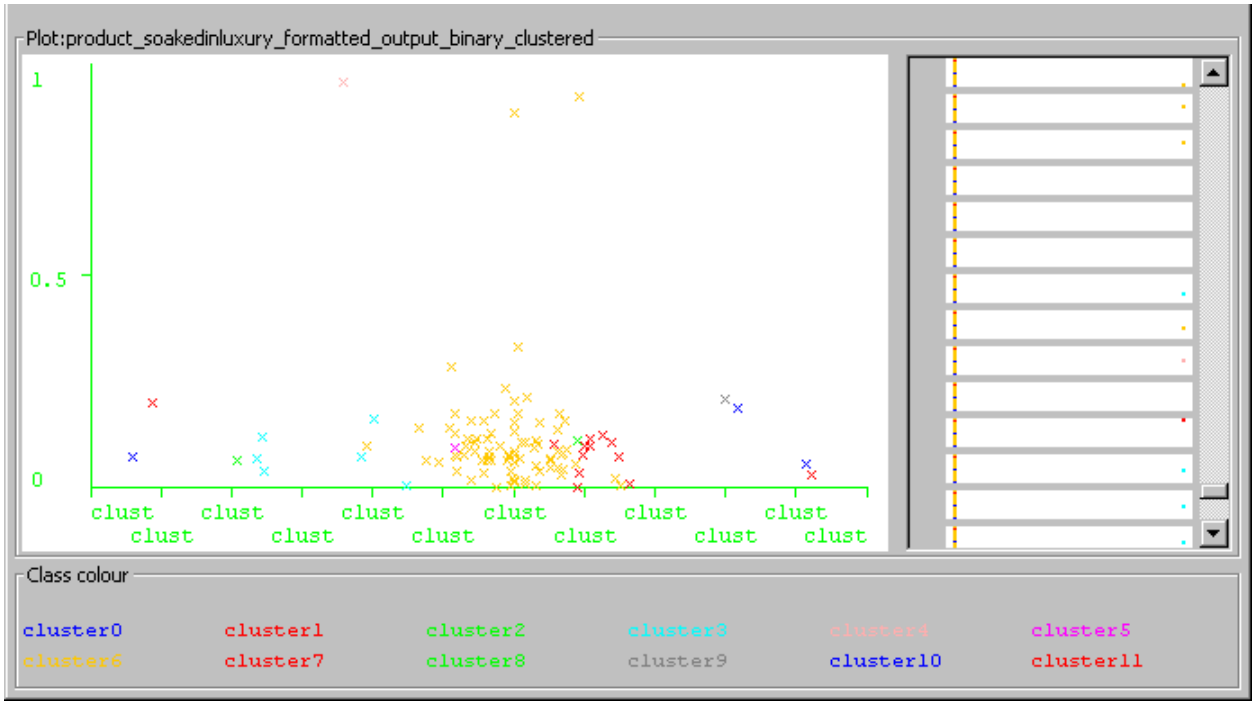

Figure 5.2.1 K-Means Clustering Algorithm visual cluster assignments running twelve clusters

Halving the initial clustering trial of twelve to six clusters, shown in Figure 5.2.2 displays defined clusters of the objects showing the boundaries within each cluster. In this scenario, a more defined visual cluster assignment is shown rather than choosing twelve clusters.

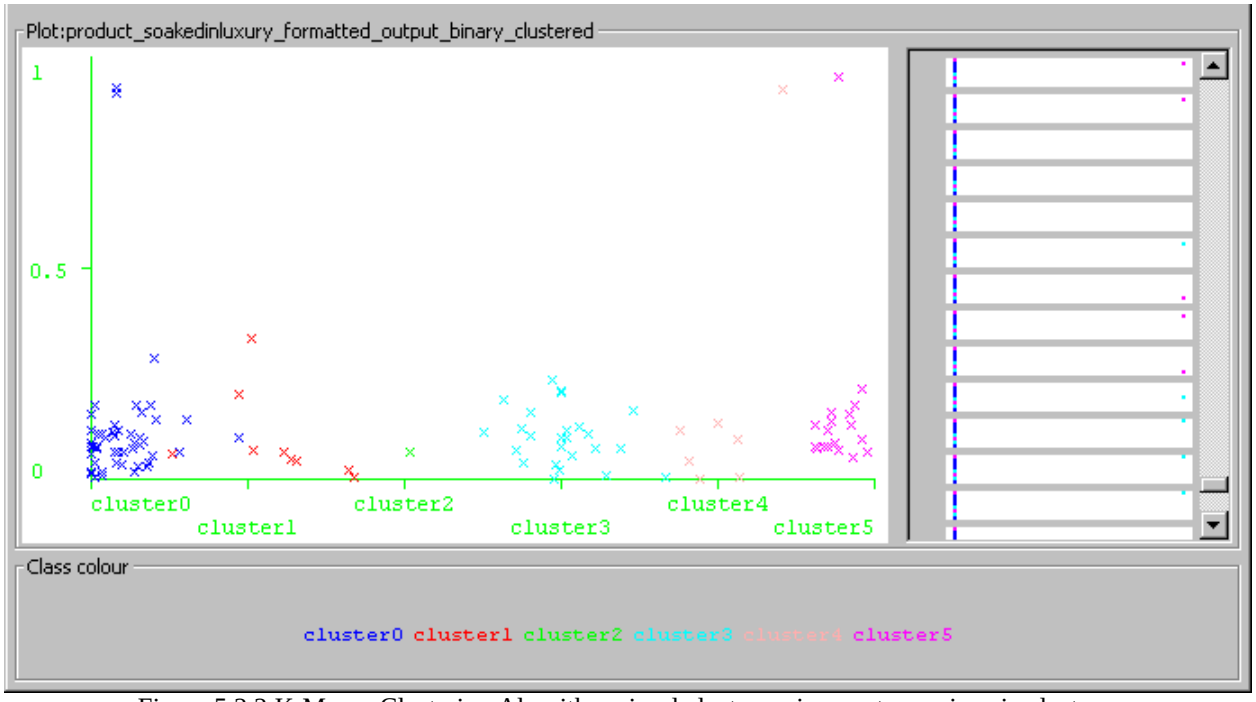

Figure 5.2.2 K-Means Clustering Algorithm visual cluster assignments running six clusters

Our results are shown in Table 5.2.3. Looking at the data more in depth, the cluster numbers in each column indicate the different groupings that can be made. Along the rows, keywords are displayed, also known as tags. To know what cluster group one is, the keywords distinguish that by looking to see how many of the full data is affected by each cluster.

By going down Table 5.2.3, clusters can be classified by what occasion and style according to the tag hits. For example, cluster 8 returns a binary 1 for the following attributes: luxury, detail, and cardigan. This cluster can be grouped as clothing which is of luxury with detail and specifically cardigans.

Table 5.2.3 K-Means Clusters

		Cluster#								
Attribute	Full Data	0	1	2	3	4	5	6	7	8
	(111)	(1)	(1)	(1)	(6)	(1)	(1)	(84)	(11)	(1)
luxury	0.8649	1	1	1	1	1	1	0.8214	1	1
elegant	0.036	0	0	0	0	0	0	0.0476	0	0
matte	0.009	0	0	0	0	0	0	0.0119	0	0
finish	0.009	0	0	0	0	0	0	0.0119	0	0
short	0.027	0	0	0	0	1	0	0.0238	0	0
sleeves	0.0991	0	0	1	0	1	0	0.0833	0.0909	0
defined	0.027	0	0	0	0	0	0	0.0357	0	0
waistline	0.027	0	0	0	0	0	0	0.0357	0	0
closes	0.027	0	0	0	0	0	0	0.0357	0	0
form-fitting	0.036	0	0	0	0	0	0	0.0476	0	0
skirt	0.0631	1	0	0	0	0	0	0.0714	0	0
casual	0.045	0	0	0	0	0	0	0.0595	0	0
occasionwear	0.009	0	0	0	0	0	0	0.0119	0	0
high-waist	0.009	0	0	0	0	0	0	0.0119	0	0
detail	0.1261	0	0	0	0.8333	0	0	0.0952	0	1
closure	0.1081	0	1	0	0	0	0	0.1071	0.1818	0
reverse	0.045	0	1	0	0	0	0	0.0476	0	0
drape	0.018	0	0	0	0.1667	0	0	0	0.0909	0
lovely	0.0631	0	0	0	0	0	0	0.0357	0.3636	0
cardigan	0.2072	0	0	0	0.8333	0	0	0.0476	0.9091	1
outfit	0.009	0	0	0	0	0	0	0	0.0909	0
complete	0.009	0	0	0	0	0	0	0	0.0909	0
long-sleeved	0.0901	0	0	0	0	0	0	0.0476	0.4545	0
linear	0.009	0	0	0	0	0	0	0	0.0909	0
pattern	0.027	1	0	0	0	0	0	0.0119	0.0909	0
cold	0.009	0	0	0	0	0	0	0.0119	0	0
weather	0.009	0	0	0	0	0	0	0.0119	0	0
wardrobe	0.018	0	0	0	0	0	0	0.0238	0	0
effortless	0.009	0	0	0	0	0	0	0.0119	0	0
style	0.1532	1	0	0	0.1667	0	0	0.1548	0	0

As a result, K-means clustering algorithm is not practical for the Inspirational Filter Application. The reason being the drawback of having to choose the number of clusters before the algorithm runs through.

5.4 Expectation-Maximization Algorithm

Expectation-Maximization clustering algorithm abbreviated as EM-clustering algorithm is a distribution-based clustering technique. Each object is specified some probability to belong to a cluster. The cluster centers are recomputed based on the average of all objects weighted by their probability of belonging to the cluster at hand.

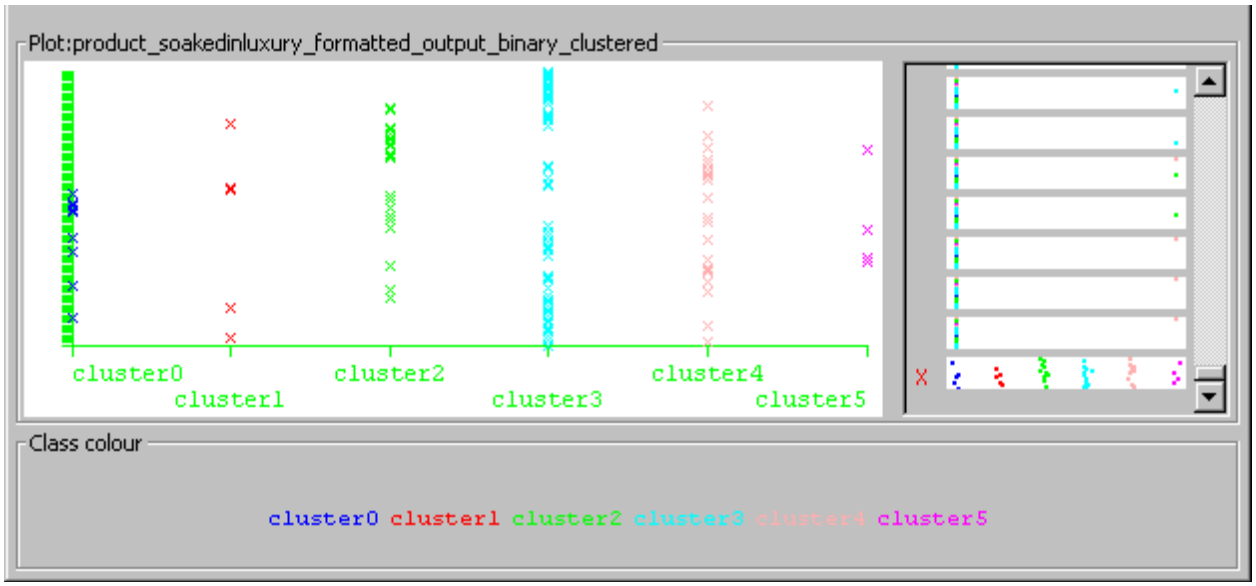

Figure 5.4.1 EM-Clustering Algorithm visual cluster assignments

In Figure 5.4.1 shows the visual cluster assignments using the EM-Clustering technique. The cluster centers are based on all the objects and each cluster is clearly defined. Instead of having to choose the number of clusters, EM-Clustering defines it for you. Figure 5.4.1 computes a total of six clusters based on the given data.

One problem that may occur when using EM-Clustering is known as over fitting. Over fitting is when the clusters start to overlap. This is when the cluster centers are too close from each other causing the clusters to overlap. Luckily, the given data does not give the problem of over fitting.

5.5 Why EM-Clustering Is Selected

After experimenting with K-Means clustering and EM-Clustering algorithms, the results shown have proved that choosing EM-Clustering algorithm is the best choice for the Inspirational Filter Application. With K-Means clustering algorithm, it is hard to choose how many clusters to work with because there is a lot of trial and error to be done with this algorithm. Hence, with EM-Clustering algorithm, the number of clusters is chosen for us depending on the given data set. EM-Clustering algorithm will be more effiecient yet effective in comparision with K-Means clustering algorithm.

5.6 Application Comparison

After research of competitor web shops, two companies have implemented a similar approach to the concept of the Inspirational Filter Application. The first company is Nordstrom which is based in the United States. The second company is Belle 96 which is also based in the United States but has made this concept as a mobile application.

5.6.1. Nordstrom

Nordstrom is one of the United States leading fashion specialty retailers both in-store and online. Their web shop introduces a theme of "Complete Looks" which is sorted "By Occasion" and "By Style". The same idea of the Inspirational Filter Application to sort clothes according to which occasion and style has been applied on Nordstrom's web shop. As for the concept of applying a dynamic filter application, they have not put into practice this type of design.

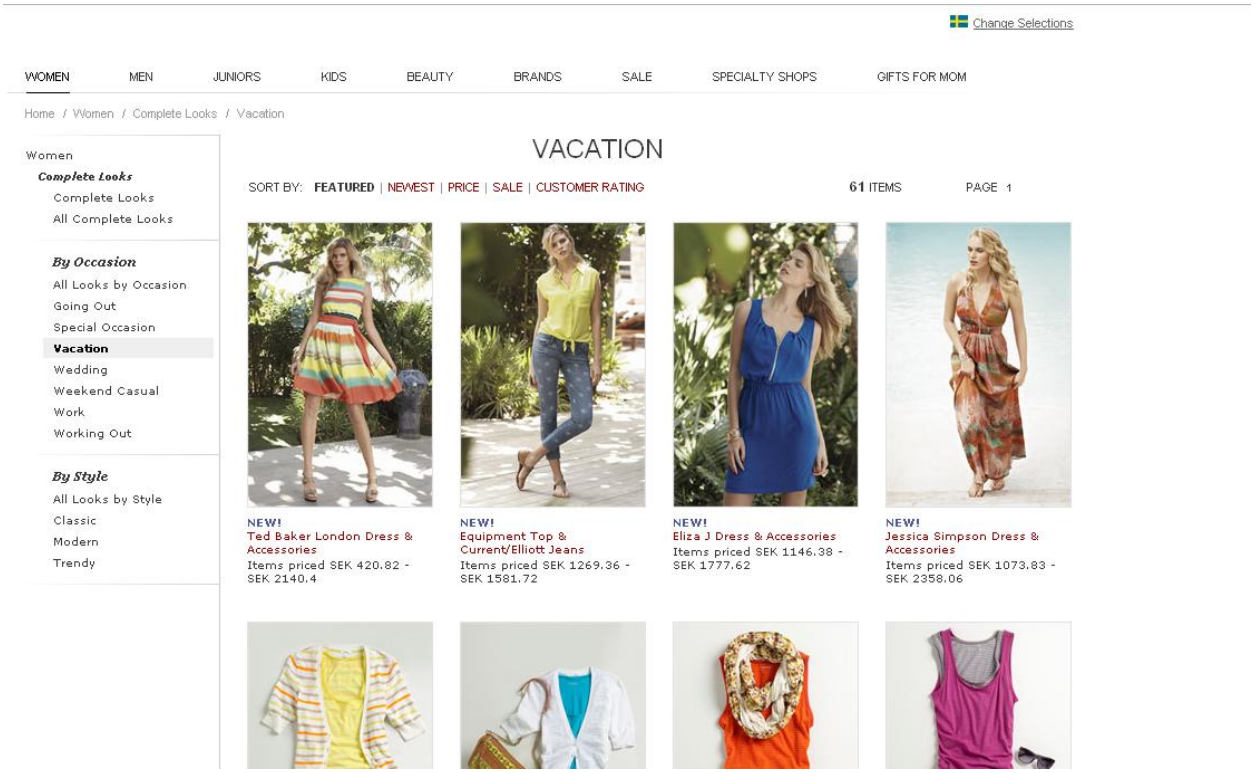

Figure 5.6.1.1 Nordstrom's Complete Looks Theme

Figure 5.6.1.1 exhibits Nordstrom's web shop through their "Complete Looks" theme. Their navigation on the left side demonstrates how they have clothes "By Occasion" and "By Style". Each shows the specific occasions and styles chosen to be used. In this case, clothes are handpicked to be placed in such categories "By Occasion" and "By Style" making it more static than dynamic. Yet, the concept is very similar but not quite what the Inspirational Filter Application is trying to achieve.

5.6.2. Belle 96

Belle 96 is a unique company in a sense of having this concept on a mobile application. Rather than having their own web shop, Belle 96 is going through a different channel of shoppers. Targeted for mobile users who love to shop online, this model has grown to be trending.

This approach is much more similar to the Inspirational Filter Application because it shows the usage of sorting clothing by occasion and style as well as implementing a filter application.

Figure 5.6.2.1 explicitly shows the Find screen on a mobile which has three separate filters: type of clothing, which occasion, and what style.

The only difference with this approach is that Belle 96 has chosen to create a mobile application instead of implementing such filters on a web shop. Again, what Belle 96 has accomplished is not quite what the Inspirational Filter Application is trying to achieve.

Figure 5.6.2.1 Belle 96 Mobile Application

6 Conclusion

The purpose of this study was to overcome the challenge to attract and retain customers in the fashion e-commerce industry through implementation of the Inspirational Filter Application. Conclusions are based on the results from statistical analysis of data clustering run by the Weka software. Clusters are formed through the binary data created by PHP scripts. These clusters are what can classify which products belong to the various occasions and styles. According to the frequency of occurrence a certain tag is matched, all the matching tags gathered can be classified

to the specific occasions and styles. This phase is beyond the project scope as it includes expertise knowledge from product coordinators in the product team from Boozt. Product coordinators are assigned brands in which clustered data can be passed onto them. With this data, they can conclude if the products in the clusters match or do not match. This end result will give us our outcome if such Inspirational Filter Application is wise to apply to the Boozt web shop.

Our comparisons with Nordstrom, one of the United States leading fashion specialty retailers both in-store and online, and with Belle 96, a company who has created a mobile application, has shown that such Inspirational Filter Application could work online in a web shop. Having variations of the concept has proven that to go above and beyond the norm of using "hard filters", The Inspirational Filter Application can be successful in the fashion e-commerce industry.

7 Recommendations for Further Work

7.1 Collaborative Filters

Further work for this application will focus on the process of including a customer purchase history through the use of collaborative filtering. Collaborative filtering is a common Web technique for generating personalized recommendations (Hoppe, 2009). Through the use of this web technique, customer purchase history can be viewed as valuable data. Memory-based technique will be used to remember what the customer has previously purchased. From this data, we can make an example assumption, "Customers who bought products A and B also bought product C." Therefore, other similar customers like A and B will be recommended C.

8 References

8.1 List of References

Journals

[1] P.H. Bloch, N.M. Ridgway & S.A. Dawson, "The shopping mall as consumer habitat," *J. Retailing*, Vol. 80, 1994.

[2]G.M.P. Swann, R. Cowan & W. Cowan, "Waves in Consumption with Interdependence Between Consumers", *Canadian Journal of Economics*, Vol.37, 2004.

[3]L.J. Rosenberg & J.A. Crepiel, "A marketing approach for customer retention", *The Journal of Consumer Marketing*, Vol.1, 1984.

[4] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann & I.H. Witten. "WEKA Data Mining Software: An Update*", SIGKDD Explorations*, Vol. 11, 2009.

World Wide Web

[5] M. Brampton. (2009, August 30). *PHP-Ace For and by top PHP developers* [Online]. Available at: http://php-ace.com/

[6]A. Brahaj. (2009, March 14). *List of English Stop Words*[Online]. Available at: http://norm.al/2009/04/14/list-of-english-stop-words/

[7] M. Rouse. (2005, September 1). *What is stop word?* [Online] Available at: http://searchsoa.techtarget.com/definition/stop-word

[8]Webopedia. (2012) *What is data mining?* [Online] Available at: http://www.webopedia.com/TERM/D/data_mining.html

[9] B. Hoppe. (2009). *Collaborative Filtering*. [Online] Available at: http://webwhompers.com/collaborative-filtering.html

[10] Boozt. (2012). *Designer Clothes For All The Family – Boozt*.com [Online] Available at: http://www.boozt.com/eu/en

[11] M. Abernethy. 2010. *Data mining with WEKA, Part 2: Classification and clustering*. [Online] Available at: http://www.ibm.com/developerworks/opensource/library/os-weka2/index.html

[12] M. Kaplan. 2011. *Behavioral Differences Between Men and Women Influence Shopping.* [Online] Available at: http://www.practicalecommerce.com/articles/3222-Behavioral-Differences-Between-Men-and-Women-Influence-Shopping

8.2 Personal Contacts

Jesper Brøndum, Chief Technology Officer, Boozt Fashion AB, jbr@boozt.com, (+46 (0) 40 12 80 05)

Appendix A – SQL Query to Extract Product Data

```
SELECT p.sku, p.name AS "product name", b.name AS "brand", c.id, REPLACE( REPLACE( p.description,
'\n', ) , '* ', ) AS "description", c.parent_id, cp.name AS "parent name", e.gender
FROM `product` AS p
LEFT JOIN ean AS e ON p.id = e.product_id
LEFT JOIN brand AS b ON p.brand_id = b.id
RIGHT JOIN product_category AS pc ON pc.product_id = p.id
LEFT JOIN category AS c ON c.id = pc.category_id
LEFT JOIN category AS cp ON c.parent_id = cp.id
WHERE b.name = "Soaked In Luxury"
AND (
c.name = "Skirts"
OR c.name = "Dresses"
OR c.name = "Tops"
OR c.name = "Knitwear"
OR c.name = "Shirts"
OR c.name = "Blouses"
OR c.name = "Outerwear"
)
AND e.gender = "F"
GROUP BY p.sku
```

Appendix B – Stop Words Source Code

The following PHP program removes stop words from an input file cross referenced with a stop words list file and writes to an output file.

```
<?php
/*This is a program to remove stop words read from an input file with a stop words list file and writes to an output file.
Brian Dewangga
10/5/2012
*/

if (count($argv) != 4) {
   echo("Usage: input file, stop words list file, output file.\n");
   exit;
}

if (!file_exists($argv[1])) {
   exit("Unable to open file $argv[1]!\n");
}

if (!file_exists($argv[2])) {
   exit("Unable to open file $argv[2]!\n");
}

$post = file_get_contents($argv[1]);
$stop_words = file($argv[2]);

foreach ($stop_words as $word) {
   $word = rtrim($word);
   $post = preg_replace("/\b$word\b/i", " ", $post);
}

$post = preg_replace("/[?;:!,.'\"*]/", " ", $post);

$output = fopen($argv[3], 'w') or
   exit("Unable to open file $argv[3]\n!");
fwrite($output, $post);
fclose($output);
?>
```

Appendix C – Binary Format Source Code

The following PHP program reads the input file and checks to see if contains the tags and writes to the output file.

```
<?php
/*This is a program to read the input file and checks to see if it contains the tags and writes to the output file.
Brian Dewangga
18/5/2012
*/

/* Configs */
$input = "product_soakedinluxury_formatted_output.csv";
$output = "product_soakedinluxury_formatted_output_binary.csv";

$handleInput = fopen( $input, "r" );
$handleOutput = fopen( $output, "w");
$tagHeader = array();

$firstLine = true;

while (!feof($handleInput))
{
   //Get line
   $line = fgets($handleInput);
   //Clean line that has extra white spaces
   $line = preg_replace("/[[:blank:]]+/",' ',$line);
   //Convert line to lowercase
   $line = strtolower($line);
   //Convert string in array using ' ' as delimiter
   $arrayLine = explode(' ', $line);

   //First line = TAG HEADER
   if ($firstLine)
   {
            $tagHeader = explode(' ', $line);
            foreach($tagHeader as $k => $v)
            {
                     //Trim the header
                     $tagHeader[$k] = trim($v);
            }
            $outputLine = str_replace(' ', ',', $line);
            $firstLine = false;
   } else {
            foreach($arrayLine as $key => $value) {
                     $arrayLine[$key] = trim($arrayLine[$key]);
            }
            $cnt = 1;
            //Copy SKU value followed by comma delimiter
            $outputLine = $arrayLine[0].',';
            //Start checking conditions for selected line
            for ($i=1; $i<count($tagHeader); $i++)
            {
                     // var_dump($tagHeader[$i]);
                     // var_dump($arrayLine);
```

```
                        //If header tag is set and are equal with current one
                        if (in_array(trim($tagHeader[$i]), $arrayLine))
                        {
                                $outputLine .= '1';
                        } else {
                                $outputLine .= '0';
                        }
                        if (++$cnt < count($tagHeader)) $outputLine .= ',';
                }
                $outputLine .= "\n";
        }
        fputs ($handleOutput, $outputLine);
}
fclose($handleInput);
fclose($handleOutput);
?>
```