

Web Content Analysis of Online Grocery Shopping Web Sites in India

Arindam Banerjee
Tanushri Banerjee

W.P. No. 2016-03-01
March 2016

The main objective of the working paper series of the IIMA is to help faculty members, research staff and doctoral students to speedily share their research findings with professional colleagues and test their research findings at the pre-publication stage. IIMA is committed to maintain academic freedom. The opinion(s), view(s) and conclusion(s) expressed in the working paper are those of the authors and not that of IIMA.



INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD-380 015
INDIA

(This paper was accepted for publication in International Journal of Business Analytics
5(4). Copyright 2018, IGI Global,
<https://www.igi-global.com/journal/international-journal-business-analytics-ijban/67141> Posted
by permission of the publisher)

Web Content Analysis of Online Grocery Shopping Web Sites in India

Arindam Banerjee¹

Tanushri Banerjee²

Abstract

In this paper the authors evaluate online grocery shopping web sites catering to customers primarily in India. The process of evaluation has been carried out in 3 parts; by comparing the web content on their homepages, analysing customer reviews and also analysing their business performance as summarized on public web sites that use search optimization tools and analytical processes. This paper aims to study attributes from structured and unstructured data that lead to success of online grocery business in India. Results of the study will help identify the keywords that Indian consumers prefer to use while searching for information on online grocery websites. It will also identify consumer preferences from the customer review analysis. Additionally, it will identify the parameters from web site traffic metrics that drive per day revenue for the online retailer.

¹ Faculty, Indian Institute of Management Ahmedabad. Email: arindam@iima.ac.in

² Faculty, School of Petroleum Management, PDU. Gandhinagar. Email: Tanushri.Banerjee@spm.pdu.ac.in

Introduction

Shopping for grocery and everyday household products is an integral part of life. In India, buying from a nearby kirana store (small retail store in the neighbourhood in India), ordering products via phone call from the kirana store, picking up groceries from the supermarket are some of the methods adopted by consumers. With changing time, consumers have started becoming technology savvy, resulting in widespread use of Internet and mobile phones. E-tailers have identified an opportunity to reach consumers via the online platform, thus beginning the era of online shopping. The online grocery portal provides features such as comparative pricing, deals and coupons, home delivery options and such others. According to a Nielsen survey¹⁰, the urban and affluent consumers in India have opted for the e-commerce as an alternate means of shopping. This platform has provided the Indian customer the product, service, conversations, query resolutions, content all under the same umbrella of digital shopping.

Grocery market growth in the online space presents a challenge for supermarket chains competing for share, where their investments in online and offline need to be balanced⁹. Earlier research related to internet grocery shopping has concentrated on comparing online and offline purchase behaviour in various ways like the importance of brand names⁶, brand loyalty⁵, shopping pattern¹, and consumers' perceptions of the advantages and disadvantages of online grocery shopping⁷. Study has also been conducted on consumer traits of internet shoppers, either in terms of their general shopping orientation⁴, their web-usage-related lifestyle³, or psychographic characteristics². Additionally⁸, based upon motivations for shopping online attempts at developing a typology of internet grocery shoppers.

Another Nielsen survey¹¹ evaluated the parameters that drive increased consumer interest in online grocery shopping. *Convenience*, *selection*, and *savings* have been stated as primary reasons for choosing online grocery over store shopping. *Shipping costs*, *inability to inspect perishables*, and having to *wait for deliveries to arrive*, have been mentioned as disadvantages of online grocery shopping. There is a significant number of people who continue to shop from stores in-person and have not yet

experienced online shopping. Reasons stated have been the habit and preference of the individual or insufficient means to use it due to lack of awareness/availability of online grocery options.

There is opportunity for growth of online grocery due to the following reasons:

1. Convenience – Online grocery shopping from home avoids the trips that have to be made to the grocery stores. It frees up the time for other activities and helps maintain better work life balance.
2. Generation Y – Generation Y has a large number of online users. Hence habit and preference of in store shopping that exists with earlier generations is less with Generation Y. Usage of the internet and mobile applications are leading to increase in online shopping.
3. Broadband and Media – Growth and penetration of Internet has increased at a rapid pace. Technology savvy customers have ventured into the online shopping space looking for information. Advertisements on media have promoted the advantages of online shopping.
4. Customization – Online grocery shopping tracks customer purchase behaviour and hence is able to provide customized recommendations to the customer during following visits to the web site based on previous shopping habits (e.g., product recommendations or promotions for combinations of offerings/kits).

With growing number of e-tailers providing online platform for grocery shopping, competition has entered the online grocery shopping market in India. Research and academic discussion forums debate competitive strategies for their growth and survival. Some of the common attributes of measuring online presence in a competitive market include the bouquet of product categories offered, the ability to service PAN India, competitive pricing, on time delivery of products and services, discounts offered through deals and coupons, maintaining quality, managing suppliers, handling consumer complaints effectively. While the content displayed on the web site is important along with the look and feel, easy loading of pages, quick navigation between web page links, appearing in searches, referrals from other web sites and a desirable revenue model; equally important is to maintain the optimum products in the inventory, on time delivery,

cashback options, deals and discounts, payment options and handling customer queries and complaints. All of these activities generate a huge amount of data that are structured as well as unstructured in nature. Some of the data is consciously captured by the organizations which are used by their Analytics teams to do descriptive and predictive analysis and devise competitive strategies and metrics for measurement of their success online.

A huge amount of data exists in unstructured form like the textual data of the website content, customer buying habits, queries and complaints and so on. Analysts have developed ways of mining this rich source of data for developing insights that can be useful for business growth and profitability. Statistical methods like web and text mining for evaluating web and textual content on the web sites have become important in analysing unstructured data.

Text mining, also known as text data mining or knowledge discovery in textual databases, is the semi-automated process of extracting patterns (useful information and knowledge) from large amount of unstructured data sources. Text mining is the same as data mining in that it has the same purpose and uses the same processes, but with text mining, the input to the process is a collection of unstructured (or less structured) data files such as Word documents, PDF files, text excerpts, XML files and so on. In essence, text mining can be thought of as a process (with two main steps) that starts with imposing structure to the text based data sources followed by extracting relevant information and knowledge from this structured text-based data using data mining techniques and tools.

The World Wide Web is perhaps the world's largest data and text repository, and the amount of information on the Web is growing rapidly every day. A lot of interesting information can be found online: whose homepage is linked to which other pages, how many people have links to a specific Web page, and how a particular site is organized. In addition, each visitor to a website, each search on a search engine, each click on a link, and each transaction on an e-commerce site creates additional data. Although unstructured textual data in the form of Web pages coded in HTML or XML are the dominant content of the Web, the Web infrastructure also contains hyperlink information (connections to other Web pages) and usage information (logs of visitors' interactions

with Web sites), all of which provide rich data for knowledge discovery. Analysis of this information can help us make better use of Web sites and also aid us in enhancing relationships and value to the visitors of our own Web sites. Web mining (or Web data mining) is the process of discovering intrinsic relationships (i.e. interesting and useful information) from Web data, which are expressed in the form of textual, linkage, or usage information¹².

This paper evaluates 10 online grocery shopping sites in India comparing the content on their homepages, customer reviews and also analysing their business performance as summarized on public websites that use search optimization tools and analytical processes. This paper aims to study attributes from structured and unstructured data that lead to success of online grocery business in India.

The public web sites that have been considered for this study are as follows:

1. Bigbasket.com (Bangalore, Mumbai, Delhi, Mysore, Chennai, Hyderabad, Pune, Madurai, Coimbatore, Vijayawada, Ahmedabad, Nashik, Lucknow & Vadodara.): bigbasket.com has over 18,000 products and over a 1000 brands in their catalogue that include fresh Fruits and Vegetables, Rice and Dals, Spices and Seasonings to Packaged products, Beverages, Personal care products, Meats⁵¹.
2. Aaramshop.com (97 cities): AaramShop is a hybrid retail platform that is completely focused on ensuring that customers can shop for daily essentials & grocery brands from the comfort of their homes or offices⁵².
3. Reliancefreshdirect.com (Mumbai): Reliance Fresh Direct provides the convenience of ordering groceries online or on phone. Customers can buy various brands and pack sizes of fruits and vegetables, dairy and bakery, staples, packaged food, household and personal care products⁵³.
4. Askmegrocery.com (Pan India): Askmegrocery.com aims at fulfilling customers' daily grocery requirements and brings a gamut of products ranging from staples, dairy products, personal care products to other household utilities and thousands of brands⁵⁴.

5. Zopnow.com (Bangalore, Hyderabad, Mumbai, Navi Mumbai, Thane, Pune and Gurgaon.): As a starting point for their vision they plan to achieve first mission of delivering “daily needs” items at customers’ doorstep in less than 3 hours. They currently carry items ranging from home needs, personal needs, staples, beverages, spices, edible oil, cleaning utilities, baby products and many more⁵⁵.
6. Fabgrocery.com (Kolkata): Fabgrocery.com a virtual departmental store, serving across Kolkata. Fabgrocery.com is created to fit into the lifestyle of every Kolkatan. They want to ease the grocery shopping experience for every household with wide range of products, on time delivery, competitive prices and much more⁵⁸.
7. Bazaarcart.com (Delhi, Noida, Gaziabad): At BazaarCart.com one can expect the best prices in Delhi/NCR on favourite grocery items because when they sell online, they can sell stuff cheaper⁵⁹.
8. Grocermax.com (Gurgaon): They promptly deliver what you really want, from an assortment of 10,000+ quality merchandise, at customers’ doorsteps, always at a price less than the MRP⁵⁶.
9. Naturesbasket.com (Mumbai, Delhi, Bangalore, Hyderabad, Pune): Godrej Nature’s Basket is the retail venture of Godrej Group started in 2005 as a single fresh food store. They have today morphed into a 36-store chain of premium gourmet stores strategically located at high street locations in Mumbai, Delhi/NCR, Pune, Hyderabad and Bangalore and have created online presence via this portal⁶⁰.
10. SaltnSoap.com (Kolkata): SaltnSoap, a BlueBeaks initiative, intends to provide its consumers a convenient, social, enjoyable and rewarding experience of shopping their daily grocery needs online from the comfort of their home and offices⁵⁷.

Methodology

The authors have conducted this study in 3 parts, all based on secondary data collected from public web sites. See Appendix 1 for information on all metrics used for evaluation.

Part A

In part A, the authors have captured unstructured data (textual) from homepage of 10 grocery shopping web sites using the Rapidminer tool's web mining operators. Hence, they established a cosine similarity index (Appendix 1) amongst them. The obtained result allows the authors to understand the similarity in terms of the homepage content and keywords amongst the web sites.

The authors have observed that web sites that are designed as retail interface for online grocery shopping have certain design aspects in common. A quick look at the websites indicates that certain content like product categories, product names, login corner, payment link, mobile app link and so on appear in all the pages. However, in the competitive space the content and design remains different in their own unique ways. In this research work, in part A, the authors have attempted to identify the similarity amongst the homepage textual content of the web sites, thus indicating the difference that exists between them in terms of the online interface content. Using Rapiminer operators, the authors have used the web crawler to connect to the specified web sites and read the content from the homepage. A comparative statistical analysis of the textual content obtained using web mining establishes the similarity index as shown in Table 1.

Output of the analysis for Part A

From the result obtained from Rapidminer (Table 1), it can be concluded that the web sites are significantly different in terms of content on the homepage. The highest similarity is seen between the 2nd (askmegrocery.com) and 5th (fabgrocery.com) web sites of 8.1%. The least similarity is seen between the 1st (bigbasket.com) and 2nd (aaramshop.com) and the 3rd (naturesbasket.com) and 5th (fabgrocery.com) websites of 0.28%.

Table 1: Similarity index in %

First website	Second website	Similarity in %
1. Bigbasket.com	2. Askmegrocery.com	0.28
1. Bigbasket.com	3. Naturesbasket.co.in	1.03
1. Bigbasket.com	4. Bazaarcart.com	0.5
1. Bigbasket.com	5. Fabgrocery.com	0.72
1. Bigbasket.com	6. Saltsoap.com	1.91
2. Askmegrocery.com	3. Naturesbasket.co.in	1.63
2. Askmegrocery.com	4. Bazaarcart.com	3.27
2. Askmegrocery.com	5. Fabgrocery.com	8.13
2. Askmegrocery.com	6. Saltsoap.com	1.89
3. Naturesbasket.co.in	4. Bazaarcart.com	3.93
3. Naturesbasket.co.in	5. Fabgrocery.com	0.28
3. Naturesbasket.co.in	6. Saltsoap.com	1.25
4. Bazaarcart.com	5. Fabgrocery.com	0.77
4. Bazaarcart.com	6. Saltsoap.com	0.94
5. Fabgrocery.com	6. Saltsoap.com	1.68

In addition to homepage content, the keywords specified for organic search for the respective web sites have been documented in Table 2. Organic search is used by majority of online users for seeking information by typing search words in search engines. In India, due to the recent interest and growth of online shopping, search keywords becomes an important dimension of web site design. Keyword optimization (the act of researching, analyzing and selecting the best keywords to target to drive qualified traffic from search engines to your website)⁶⁸ is an important consideration in the overall web site design for reaching maximum number of prospective customers. It is observed that

the most common keywords appearing across all the sites are “online grocery shopping”, “home delivery” and “supermarket”.

Table 2: List of keywords

NO	WEB SITE	KEYWORDS
1	Bigbasket.com	online supermarket , online grocery store, buy groceries online, grocery online , Buy food online, grocery, online grocery shopping , online grocery, food shopping online, online grocery shopping india ²⁹
2	Aaramshop.com	CPG, FMCG, FREE Home Delivery , online grocery shopping , aaramshop, india, lizol, neighborhood retailers, hybrid commerce, allout mosquito repellent ³⁰
3	Reliancefreshdirect.com	reliance fresh, buy grocery online, online grocery shopping , online grocery, reliance one card ³¹
4	Askmegrocery.com	grocery shopping online , online grocery store, online grocery shopping , online grocery store in india, online vegetable shopping, Online Fruits shopping, askme grocery, askmegrocery, ask me grocery, askme grocery offers ³²
5	Zopnow.com	buy online, supermarket , Grocery delivery, ZOP, hypermarket, grocery, online grocery shopping , online grocery, tata tea gold, zopnow ²⁸
7	Fabgrocery.com	online grocery kolkata, best online grocery store in kolkata, online grocery shopping in kolkata, online grocery store Kolkata, grocery store in Kolkata, grocery shopping in Kolkata, online grocery products in Kolkata, fabgrocery, grocery product kolkata, grocery home delivery in kolkata ³³
8	Grocermax.com	vegetables, frozen, grocery online, non veg, grocery, online grocery shopping , online grocery, online grocery gurgaon ³⁵
9	Saltsoap.com	online supermarket, online grocery store, buy grocery online, grocery, online grocery shopping , grocery store india, home delivery, best online grocery store, online grocery kolkata, grocery store Kolkata ³⁴
10	Naturebasket.co.in	online grocery shopping , grocery shopping online, online grocery, online supermarket , food online, gourmet food india ²⁷

Part B

In part B, the authors have analysed customer reviews from the online grocery web sites. Fourteen customer reviews have been considered for the study. The authors have used Rapidminer’s text mining operators and association rules operators to analyse them and establish association rule from the textual content. Association rule mining finds interesting associations and correlation relationships among large sets of data items. Association rules show attribute value conditions that occur frequently together in a given data set. They provide information in the form of if-then statements and are probabilistic in nature. In addition to the

antecedent (if) and the consequent (then), an association rule has two numbers that express the degree of uncertainty about the rule. In association analysis, the antecedent and consequent are sets of items (called item sets) that are disjoint (do not have any items in common). The first number is called the “Support” for the rule. “Support” is the number of transactions that include all items in the antecedent and consequent parts of the rule. The other number is known as the “Confidence” of the rule. Confidence is the ratio of the number of transactions that include all items in the consequent, as well as the antecedent (the support) to the number of transactions that include all items in the antecedent. “Lift” is a value that gives us information about the increase in probability of the then (consequent) given the if (antecedent) part⁶⁹. The authors have evaluated the customer reviews using these parameters and text mining operators of Rapidminer.

The shopping experience of a customer who shops online is considered satisfactory only when the customer is happy with the overall shopping experience. The authors have addressed this hypothesis from the textual content documenting the customer reviews. Using “happy” as the indicator, the authors have identified the words that are associated with happiness. The strength of the “Support” of those words existing together across the reviews indicates the areas of focus for the e-tailer to provide an overall satisfactory experience to their customers. The authors have further evaluated the “Confidence” and “Lift” for the associated words that have high “Support” value.

Output of the analysis from Part B

Analysing the “Support”, “Confidence” and “Lift” values from association rule for conclusion word “happy” identifies the words to which happiness of a customer is associated. As seen from Image 1, the “Support” value is 35.7%, meaning 35.7% times of the customer reviews, the word “happy” exists together with the words “delivery” and “products”. This indicates that the mention of happiness due to delivery of products has been made only in 35.7% of the total number of reviews. The authors have further noted that there is an 83.3% chance of “happy” existing in customer reviews that mention delivery of products indicating a “Confidence” value of 83.3%. This indicates that the if the

customer mentions about delivery of products, there is 83.3% chance of them mentioning happiness. Additionally, the value of “Lift” is 1.667, indicating that occurrence of words “delivery” and “products” has a positive effect on the occurrence of “happy” or that they are positively correlated. We also note that words “delivery” and “groceries” along with “time” and “happy” has a “Lift” value of 3.5, indicating that customers who receive delivery of groceries view it as time saving and are happy with the experience.

Image 1: Association rule output from rapidminer

No	Premises	Conclusion	Support	Confidence	LiftPlace	Gain	p-s	Lift	Conviction
803	delivery, products	happy	0.357	0.833	0.950	-0.500	0.143	1.667	3
804	delivery, customer	happy	0.357	0.833	0.950	-0.500	0.143	1.667	3
819	I, customer	happy	0.357	0.833	0.950	-0.500	0.143	1.667	3
842	delivery, customer	I, happy	0.357	0.833	0.950	-0.500	0.204	2.333	3.667
843	I, customer	delivery, happy	0.357	0.833	0.950	-0.500	0.143	1.667	3
844	delivery, I, customer	happy	0.357	0.833	0.950	-0.500	0.143	1.667	3
913	delivery, groceries	happy	0.286	1	1	-0.286	0.143	2	∞
964	time, customer	happy	0.286	1	1	-0.286	0.143	2	∞
983	I, groceries	happy	0.286	1	1	-0.286	0.143	2	∞
1007	products, customer	happy	0.286	1	1	-0.286	0.143	2	∞
1009	customer, groceries	happy	0.286	1	1	-0.286	0.143	2	∞
1060	delivery, groceries	shopping, happy	0.286	1	1	-0.286	0.184	2.800	∞
1061	delivery, shopping, groceries	happy	0.286	1	1	-0.286	0.143	2	∞
1133	time, customer	delivery, happy	0.286	1	1	-0.286	0.143	2	∞
1134	delivery, time, customer	happy	0.286	1	1	-0.286	0.143	2	∞
1138	delivery, groceries	time, happy	0.286	1	1	-0.286	0.204	3.500	∞
1139	delivery, time, groceries	happy	0.286	1	1	-0.286	0.143	2	∞
1186	delivery, groceries	I, happy	0.286	1	1	-0.286	0.184	2.800	∞
1187	I, groceries	delivery, happy	0.286	1	1	-0.286	0.143	2	∞
1188	delivery, I, groceries	happy	0.286	1	1	-0.286	0.143	2	∞
1218	products, customer	delivery, happy	0.286	1	1	-0.286	0.143	2	∞
1219	delivery, products, customer	happy	0.286	1	1	-0.286	0.143	2	∞
1221	delivery, groceries	happy, customer	0.286	1	1	-0.286	0.184	2.800	∞
1224	customer, groceries	delivery, happy	0.286	1	1	-0.286	0.143	2	∞
1225	delivery, customer, groceries	happy	0.286	1	1	-0.286	0.143	2	∞
1259	time, customer	shopping, happy	0.286	1	1	-0.286	0.184	2.800	∞
1260	shopping, time, customer	happy	0.286	1	1	-0.286	0.143	2	∞
1282	I, groceries	shopping, happy	0.286	1	1	-0.286	0.184	2.800	∞
1283	shopping, I, groceries	happy	0.286	1	1	-0.286	0.143	2	∞
1307	customer, groceries	shopping, happy	0.286	1	1	-0.286	0.184	2.800	∞
1308	shopping, customer, groceries	happy	0.286	1	1	-0.286	0.143	2	∞
1317	time, customer	I, happy	0.286	1	1	-0.286	0.184	2.800	∞

Part C

In part C, the authors have captured the parameters (attributes) used by search engine optimization) SEO sites for analysing the online grocery web sites. Then we have tried to identify the correlation amongst these attributes. Several online sites like stattools.com, statvoo.com, informer.com, similarityweb.com and so on have used website usage metrics information and have analysed it using SEO tools and statistical methods. Information used for our study include the following parameters for the 10 websites: daily visitors, daily page views, Alexa rank, page speed score, images on the site in % compared to other content on the site, traffic source on the website of which organic search traffic in %, web site worth in dollars, income per day in dollars. The parameters

have been defined in Appendix 1. Using a correlation matrix, the authors have tried to understand the relationship between income per day in dollars made by the web sites to the other parameters, thus identifying the ones that are strongly (> 0.9) correlated. The parameters that have strong correlation to income per day in dollars are the ones that have higher impact on the revenue.

Output of the analysis from Part C

The correlation matrix in Table 3 indicates a strong correlation between “Daily page views” – 0.97, “Daily visitors” - 0.98, “web site’s estimated worth in dollars” – 0.99 to “income per day in dollars”. The Alexa ranking is negatively correlated as less the Alexa ranking more visited is the web site. However, from the findings, the Alexa ranking does not seem to be affecting the income per day. Similarly, the amount of images on the website, the time taken by the pages to load and the traffic source via search engines do not have significant impact.

Table 3: Correlation matrix

	<i>INCOME PERDAY (\$)</i>	<i>DAILY VISITORS</i>	<i>DAILY PAGEVIEWS</i>	<i>ALEXA RANK</i>	<i>PAGE SPEED SCORE (%)</i>	<i>ESTIMATED WORTH (\$)</i>	<i>IMAGES (%)</i>	<i>SEARCH TRAFFIC (%)</i>
INCOME PERDAY (\$)	1							
DAILY VISITORS	0.978	1						
DAILY PAGEVIEWS	0.970	0.998	1					
ALEXA RANK	-0.392	-0.366	-0.338	1				
PAGE SPEED SCORE (%)	0.225	0.161	0.167	-0.103	1			
ESTIMATED WORTH (\$)	0.988	0.969	0.968	-0.313	0.277	1		
IMAGES (%)	-0.459	-0.436	-0.452	0.278	-0.318	-0.523	1	
SEARCH TRAFFIC (%)	-0.077	-0.152	-0.155	0.022	0.768	-0.051	-0.270	1

Overall Findings

In part A, we find the following:

1. Maximum similarity between the homepage content of online grocery shopping websites (10 considered for this study) is 8.1% and minimum similarity is 0.28%.
2. The most common keywords appearing across all the 10 web sites are “online grocery shopping”, “home delivery” and “supermarket”.

In part B, we find the following:

1. Happiness with the experience of using the online grocery web site is positively correlated to delivery of groceries.
2. The correlation is stronger when the delivery of the grocery is done timely.

In part C, we find the following:

1. The income per day in dollars made by the e-tailers of grocery online is strongly correlated to number of visitors on the website, number of page views and web sites' worth in dollars.
2. It is not affected by Alexa ranking, images on the website, the time taken by the pages to load and the traffic source via organic search engines.

Practical implications

Results of the study will help identify the keywords that e-tailers of grocery online anticipate the Indian consumers to use while using a search engine to get to online grocery web sites. Common keywords used in search engines would fetch all the web sites containing the keywords. Thus the results will encourage e-tailers to uniquely differentiate from competitor websites by keyword optimization. It will also identify consumer preferences from the customer review analysis. With the statistically obtained result, the e-tailer is reassured that timely delivery of groceries makes an Indian customer happy. It may be important to identify what else is associated with happiness which did not show significantly due to the limited number of samples used for the study. The result obtained from the correlation matrix will help identify the parameters that drive per day revenue for the online retailer. Due to the limited sample size further statistical analysis could not be carried out to establish the strength of the impact of the independent variable on per day revenue, however the strength of the correlation identifies the impacting

variables. Further studies can help devise creative strategies for optimizing web site metrics for increasing revenue.

Originality/value

While study on individual online grocery shopping sites in India have been conducted, the authors have not come across a comparative study of the sites both from the point of content analysis as well as usage information. Results obtained from the study will be useful for online retailers in designing their web sites for generating more revenue and increase number of happy customers. The similarity index will identify the similarity in terms of textual content that exists between their and competitor web sites. The keywords similarity and difference highlight what competing sites use to show up in organic searches. The academic researchers who work in online retailing space can use the results of the study to carry out similar research with larger data sets.

Future directions and limitation of study

The online grocery shopping web sites catering to customers are at its nascent stage and small in number. Hence, the limitation of this study has been the small sample size both in terms of the number of grocery online web sites and the number of customer reviews available from the web sites. Similar studies can be carried out with other e-tailer web sites and web portal sites (assortment of products from different vendors) that are comparatively larger in number. Larger data sets and sample size thus obtained can be used to test the association rule and the correlation matrix for further statistical outcomes.

References:

1. Andrews, R.L. and Currim, I.S. (2004), "Behavioural differences between consumers attracted to shopping online versus traditional supermarkets: implications for enterprise design and marketing strategy", *International Journal of Internet Marketing and Advertising*, Vol. 1 No. 1, pp. 38-61.
2. Barnes, S.J., Bauer, H.H., Neumann, M.M. and Huber, F. (2007), "Segmenting cyberspace: a customer typology of the internet", *European Journal of Marketing*, Vol. 41 No. 1/2, pp. 71-93
3. Brengman, M., Heuens, M., Weijters, B., Smith, S.M. and Swinyard, W.R. (2005), "Segmenting internet shoppers based on their web-usage related lifestyle: a cross cultural validation", *Journal of Business Research*, Vol. 58 No. 1, pp. 79-88.
4. Brown, M., Pope, N. and Voges, K. (2003), "Buying or browsing? An exploration of shopping orientations and online purchase intention", *European Journal of Marketing*, Vol. 37 No. 11/12, pp. 1666-84.
5. Danaher, P.J., Wilson, I.W. and Davis, R.A. (2004), "A comparison of online and offline consumer brand loyalty", *Marketing Science*, Vol. 22 No. 4, pp. 461-76.
6. Degeratu, A., Rangaswamy, A. and Wu, J. (2000), "Consumer choice behavior in online and traditional supermarkets: the effects of brand name, price and other search attributes", *International Journal of Research in Marketing*, Vol. 17 No. 1, pp. 55-78.
7. Ramus, K. and Nielsen, N.A. (2005), "Online grocery retailing: what do consumers think?", *Internet Research*, Vol. 15 No. 3, pp. 335-52.
8. Rohm, A.J. and Swaminathan, V. (2004), "A typology of online shoppers based on shopping motivations", *Journal of Business Research*, Vol. 57 No. 7, pp. 748-57.
9. Francesca, R. and Rettie, R. (2009), "Online Grocery Shopping: The Influence of situational factors", *European Journal of Marketing*, Vol 43, No. 9/10, pp. 1205-1219.
10. Terron Adrian, The Nielsen Company, 2012, Nielsen Featured Insights: The Advent of 'Me'-tailing and how India's shoppers are now buying online, retrieved from <http://www.nielsen.com/content/dam/corporate/india/reports/2012/Me-tailing-and-how-In-dia-shoppers-are-now-buying-online.pdf>
11. Swedowsky Maya, The Nielsen Company, September 2009 "Analyst Report: Online Grocery Shopping: Ripe time for Resurgence" , retrieved from http://www.nielsen.com/content/dam/corporate/us/en/newswire/uploads/2009/10/Nielsen-OnlineGroceryReport_909.pdf
12. Turban, E., Sharda, R., Delen, D. and King, D. (2011), *Business Intelligence: A Managerial Approach*, Second Edition, PEARSON, ISBN: 978-93-325-1814-8
13. <http://bigbasket.com.statstool.com/> accessed on 12 February, 2016
14. <https://aaramshop.com.statstool.com/> accessed on 12 February, 2016
15. <https://sixstat.com/ratings-reviews/askmegrocery.com> accessed on 12 February, 2016

16. <http://typo.tools/sites/askmegrocery.com> accessed on 12 February, 2016
17. <https://www.website-worths.com/askmegrocery.com.html> accessed on 12 February, 2016
18. <https://www.website-worths.com/reliancefreshdirect.com.html> accessed on 12 February, 2016
19. <http://www.similarweb.com/website/reliancefreshdirect.com> accessed on 12 February, 2016
20. <http://webstats.report/www/reliancefreshdirect.com> accessed on 12 February, 2016
21. <https://statvoo.com/www/reliancefreshdirect.com> accessed on 12 February, 2016
22. <https://statvoo.com/www/aaramshop.com> accessed on 12 February, 2016
23. <https://statvoo.com/www/grocermax.com> accessed on 12 February, 2016
24. <https://statvoo.com/www/zopnow.com> accessed on 12 February, 2016
25. <https://statvoo.com/www/fabgrocery.com> accessed on 12 February, 2016
26. <https://statvoo.com/www/saltnsoap.com> accessed on 12 February, 2016
27. <http://naturesbasket.co.in.websiteoutlook.com/> accessed on 12 February, 2016
28. <http://website.informer.com/zopnow.com> accessed on 12 February, 2016
29. <http://website.informer.com/bigbasket.com> accessed on 12 February, 2016
30. <http://website.informer.com/aaramshop.com> accessed on 12 February, 2016
31. <http://website.informer.com/reliancefreshdirect.com> accessed on 12 February, 2016
32. <http://website.informer.com/askmegrocery.com> accessed on 12 February, 2016
33. <http://website.informer.com/fabgrocery.com> accessed on 12 February, 2016
34. <http://website.informer.com/saltnsoap.com> accessed on 12 February, 2016
35. <http://website.informer.com/grocermax.com> accessed on 12 February, 2016
36. <http://www.naturesbasket.co.in/> accessed on 12 February, 2016
37. <http://www.freewebsitereport.org/www.askmegrocery.com> accessed on 12 February, 2016
38. <http://www.similarweb.com/website/bigbasket.com> accessed on 12 February, 2016
39. <http://www.similarweb.com/website/aaramshop.com> accessed on 12 February, 2016
40. <https://www.woorank.com/en/www/askmegrocery.com> accessed on 12 February, 2016
41. <http://www.similarweb.com/website/zopnow.com> accessed on 12 February, 2016
42. <http://www.similarweb.com/website/fabgrocery.com> accessed on 12 February, 2016
43. <http://www.similarweb.com/website/grocermax.com> accessed on 12 February, 2016

44. <http://sitegur.com/bazaarcart.com> accessed on 12 February, 2016
45. <http://www.similarweb.com/app/google-play/com.godrej.naturesbasketltd/statistics> accessed on 12 February, 2016
46. <http://selectok.com/pagerank/bazaarcart.com> accessed on 12 February, 2016
47. <http://selectok.com/pagerank/naturesbasket.com> accessed on 12 February, 2016
48. <http://www.freewebsitereport.org/www.zopnow.com> accessed on 12 February, 2016
49. <https://sixstat.com/ratings-reviews/askmegrocery.com> accessed on 12 February, 2016
50. <http://www.rankstand.com/www/naturesbasket.co.in> accessed on 12 February, 2016
51. <http://www.bigbasket.com/about-us/> accessed on 12 February, 2016
52. <https://www.aaramshop.com/pages/about-us> accessed on 12 February, 2016
53. <https://www.reliancefreshdirect.com/> accessed on 12 February, 2016
54. <http://www.askmegrocery.com/home/aboutus> accessed on 12 February, 2016
55. <http://www.zopnow.com/pages/about#whatis> accessed on 12 February, 2016
56. <https://grocermax.com/about-us/> accessed on 12 February, 2016
57. <http://saltnsoap.com/about-us.php> accessed on 12 February, 2016
58. <http://www.fabgrocery.com/main/aboutus> accessed on 12 February, 2016
59. <http://www.bazaarcart.com/pages/About-Us/pgid-1026436.aspx> accessed on 12 February, 2016
60. www.naturesbasket.co.in/HTML/AboutUs.aspx accessed on 12 February, 2016
61. https://en.wikipedia.org/wiki/Cosine_similarity accessed on 24 February, 2016
62. <https://moz.com/learn/seo/page-speed> accessed on 24 February, 2016
63. <http://seositecheckup.com/articles/alt-tags-and-other-image-optimization-tools> accessed on 24 February, 2016
64. <http://help.weebly.com/stats--seo.html> accessed on 24 February, 2016
65. <https://www.tendenci.com/help-files/meaning-of-hits-visits-page-views-and-traffic-sources-web-analytics-definitions/> accessed on 24 February, 2016
66. <http://www.seo-detective.com/scan/> accessed on 24 February, 2016
67. <http://www.shoutmeloud.com/how-does-alexa-rankings-work.html> accessed on 24 February, 2016
68. <http://www.wordstream.com/blogs/ws/2010/04/14/keyword-optimization> accessed on 24 February, 2016
69. <http://www.solver.com/xlminer/help/association-rules> accessed on 24 February, 2016

Appendix 1

Definition of terms used in this research study:

1. Cosine Similarity index: Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and it is less than 1 for any other angle. It is thus a judgment of orientation and not magnitude: two vectors with the same orientation have a cosine similarity of 1, two vectors at 90° have a similarity of 0, and two vectors diametrically opposed have a similarity of -1, independent of their magnitude. Cosine similarity is particularly used in positive space, where the outcome is neatly bounded in $[0,1]$. In information retrieval and text mining, each term is notionally assigned a different dimension and a document is characterized by a vector where the value of each dimension corresponds to the number of times that term appears in the document. Cosine similarity then gives a useful measure of how similar two documents are likely to be in terms of their subject matter⁶¹.
2. Images on site: Search engines send out automated programs called “spiders” that crawl web sites and search for what type of content it contains. Although spiders aren’t able to “see” images, they are able to read the text which is associated with those images⁶³.
3. Alexa traffic ranking: Alexa Traffic Rank is traffic rank of a site. The traffic rank is based on 3 months of historical traffic data collected from the users of Alexa Toolbar. It is combined with different traffic data source and page views. Alexa counts the number of visitors and number of page views of all the sites for Alexa traffic rank⁶⁷.
4. Page load time: Page speed can be described in either "page load time" (the time it takes to fully display the content on a specific page) or "time to first byte" (how long it takes for your browser to receive the first byte of information from the web server). No matter how you measure it, a faster page speed is better⁶².
5. Revenue per day: Revenue of the web site on a daily basis.

6. Web site's worth: There are several factors that determine the website valuation price. These include age of the website, domain authority, SEO analysis, Page Rank, traffic Rank, social media interaction, website content and many other factors. These factors are all considered to calculate the estimate of how much the website is worth⁶⁶.
7. Number of unique visitors: A visit is one individual visitor who arrives at your web site and proceeds to browse. A visit counts all visitors, no matter how many times the same visitor may have been to your site. A unique visit will tell you which visitors are visiting your site for the first time. The website can track this as unique by the IP address of the computer. The number of unique visits will be far less than visits because a unique visit is only tracked if cookies are enabled on the visitor's computer⁶⁵.
8. Number of page views: Once a visitor arrives at a web site, they will search around on a few more pages. On an average, a visitor will look at about 2.5 pages. Each individual page a visitor views is tracked as a page view⁶⁵.
9. Traffic source via organic search: This indicates the percentage of visitors who visit the site via a search engine. There are primarily four different types of traffic sources (the way visitors find the web site)⁶⁵:
 - a. Direct Navigation: (type URL in traffic, bookmarks, email links w/o tracking codes, etc.)
 - b. Referral Traffic: (from links across the web, social media, in trackable email, promotion & branding campaign links)
 - c. Organic Search: (queries that sent traffic from any major or minor web search engines)
 - d. PPC: (click through from Pay Per click sponsored ads, triggered by targeted key phrases)