MARKET INTELLIGENCE FROM THE INTERNET: AN ILLUSTRATION USING THE BIOMASS HEATING INDUSTRY

Prof Adee Athiyaman  
Western Illinois University  
an-athiyaman@wiu.edu  
Orcid no: 0000-0002-4684-4204

Mr. Thebe Magapa  
Vaal University of Technology  
thebem@vut.ac.za  
Orcid no: 0000-0002-6754-3435

—Abstract—

Extant research on marketing strategy suggests that most companies underuse web intelligence as publicly available data on the Internet are considered hard to access and analyse. This paper demonstrates how biomass home heating businesses can utilise the Internet for data collection and business insights. The market structure of the biomass heating industry was identified using the Google Correlate algorithm. The production rule ‘that newer the product the higher is consumer search for the product’ was operationalised using the correlations of the concept ‘home heating equipment’. Intra-industry competition was assessed using Google’s brand impression analysis and firm behaviour and performance were modelled using a differential equation relating product sales to marketing expenditures. Empirical analysis reveals that the product form “biomass home heating” is growing, pellet stoves and fireplace inserts top the lists of “stove” searches, there are two competitive clusters of biomass firms and the marketing spending for the industry is well below its optimum level needed to increase and maintain sales.

Keywords: Market intelligence, Biomass home heating, US biomass markets, Marketing optimisation, Google tools

JEL Classification: M-M3-M31.
1. INTRODUCTION

Rapid changes in political, economic, social and technological domains are impacting businesses around the globe. Global economic growth has slowed to an annual average of 2.1 percent, a 42 percent decline from the 2010 average (IMF Research Bulletin, 2016). In this competitive economy, winning companies utilise specific capabilities such as market research to understand and invest in pockets of growth. For example, research by Matsuyama (2017) suggests that companies that outperform peers in categories in which they compete invest considerable time and effort to acquire market structure information and simulate the impact of competitive strategies on firm performance.

Newer options in market research are making the task of gathering market intelligence less burdensome; machines can scan millions of documents on the Internet to generate relevant data and reports. For example, at Clearwell Systems, a Silicon Valley company that analyses legal documents for pre-trial discovery, machines routinely scan millions of documents on the Internet to retrieve data that are relevant for upcoming trials (Zimmerman & Blythe, 2017). Yet most businesses, specifically small businesses, tend to stick to off-the-shelf methods of data collection and thus fail to gain better competitor and consumer insights (Forsyth & Boucher, 2010; Llave, 2017).

This paper demonstrates how the Internet can be used to gather market intelligence for a business; the biomass heating industry in the US is the unit of analysis. The need to study biomass home heating stems from its potential for growth; unlike fossil fuels, biomass heating neither strip resources from the earth nor contributes to pollution and climate change (Athiyaman, 2018; Bullock & Yu, 2016; Gershoff & Irwin, 2012). Furthermore, scientific advancements in synthetic biology are expected to double biofuels usage globally, from its current 2 percent to 4 percent by 2020 (Manyika, Michael, Jacques, Richard, Peter & Alex, 2013).

1.1 What is biomass?

Biomass is any organic matter – wood, crops, seaweed, animal wastes – that can be used as an energy source. According to the Energy Information Administration (EIA), biomass constitutes slightly more than 5 percent of the total US energy production. https://www.eia.gov/totalenergy/data/monthly/pdf/sec10_3.pdf
. About 50 percent of the biomass used today for energy production comes from burning wood and wood scraps such as sawdust. According to the American Community Survey, one in ten households in the nation burn wood for heating, but few use wood as the only source of heat https://www.census.gov/programs-surveys/acs/. Fireplaces, wood/pellet stoves and outdoor wood boilers are the common heating appliances used at homes. Although most homes burn split logs or cordwood, wood pellet use is on the rise. Geographically, households in the north-east states rely on biomass heat the most. Socioeconomically, while low-income households burn more wood, it is the upper-income households that use the most fireplaces and woodstoves https://www.eia.gov/totalenergy/data/monthly/pdf/sec10_3.pdf.

1.2 Conceptual framework

In business, customers are won by fending off competitors. Hence, without proper understanding of both customers and competitors – market intelligence – a business cannot succeed or gain competitive advantage. This is why competitive advantage is defined as a business leveraging; a benefit valued by customers that is not easily copied by competitors (Qureshi, 2017; Rossiter, 1996).

How could we gain insights into competitive and consumer behaviour? The idea of Schumpeter (1939) and Fusari (2016) on innovation and business cycle that product innovation drives competition, can be used to garner insights into competitive behaviour. Marketing theory offers insights into consumer behaviour.

2. COMPETITIVE BEHAVIOUR

When a business creates a new product, often other companies in the industry each offer an improved product. This behaviour results in innovations clusters and consumers respond by buying the innovations (Schumpeter, 1939; Howard, 1989; Lambert, Mattson & Dorriere, 2017).

The concept of product life cycle (PLC) can be used to capture these ideas (Day, 1981; Steinhardt, 2017). PLC postulates four distinct stages in the life of a product: introduction, growth, maturity and decline. Schumpeter’s new product is introduced in the marketplace. Soon competitors follow, each trying to introduce a better product: the period of growth in PLC terminology. In time, market saturation or maturity is reached.

The classic PLC holds almost universally for product classes and very well for product forms (Kotler, 2017). As a product form, biomass was the primary fuel for heating and cooking in homes in the 1860s (Bracmort, 2015). In the early
1900s, coal was the preferred mode of heating for city dwellers; the country households stayed with biomass for heating and cooking. In the 1950s, electricity and natural gas replaced most biomass and coal usage at homes across the US. Starting in the 1970s, biomass heating became popular again due to high world energy prices. Since then, biomass home heating has been steadily on the rise; market potential calculations suggest vast potential for growth in certain segments of the population (Athiyaman, 2015a). Note that the form of biomass heating in the 1860s is not the same now; then it was trees and shrubs burnt in fire pits and now it is cordwood and pellets fired in central-air furnaces and stoves. To this supply-side explanation, we now add demand or consumer behaviour explanations.

2.1 Consumer behaviour: marketing theory

In marketing, consumer behaviour is often categorised into three decision stages: (i) extended problem solving (EPS), (ii) limited problem solving (LPS) and (iii) routinised response behaviour (RRB) (Howard 1989; Rossiter & Percy, 1997). Each of these stages of decision can be related to a stage of the PLC. For instance, in EPS, consumers have much to learn about a product. This learning is the result of consumers encountering a marketplace innovation that is in the introductory stage of its lifecycle. As more new products, or variations of the innovation, appear in the marketplace, consumers become familiar with the product and hence demand much less information – corresponds to LPS in consumer behaviour and the growth stage in the PLC. Finally, all consumers become familiar with the products and their behavioural response to the purchase of one or more products becomes routine: this depicts the RRB of the consumer and the mature stage of the product in the PLC.

2.2 Framework for market intelligence

Since PLC captures both the demand and the supply side of an industry, we employ it as a framework to garner market intelligence. Based on the industrial organisation’s market-structure approach we posit that a product’s position in the PLC defines the state of the industry or industry structure and this leads to certain competitive behaviour on the part of the company and its market performance (Aslan, 2017; Bain, 1968).

Marketing expenses in the introduction stage are typically high – a high level of advertising / sales promotion is needed to make people aware of the “innovation” or the newly introduced product and to communicate its benefits for potential
buyers’ considerations. This trend of high marketing spending continues during the product growth phase, hits a ceiling during maturity and declines to a minimum during the “end” stages of the product (Lukas, Spengler, Kupfer & Kieckhafe, 2017; Rossiter, & Percy, 1997).

2.3 Methodology

We employ Google’s search refinement operators and lesser-known features such as “stemming” to gain market intelligence about biomass home-heating products. In line with established search practices (Barker, Barker & Pinard, 2012) we begin with a basic search by arranging the search words in the order that they are likely to appear in the web pages. For example:

- **biomass home-heating product**

Note that a hyphenated word search will also yield pages with the unhyphenated version. Next, we add to the basic search Boolean operator ‘OR’ to recognise that pages may refer biomass product as biomass equipment:

- **biomass home-heating product OR equipment**

To prevent Google from matching variations of the search words (stemming) we enclose the first three words in double quotes:

- **“biomass home-heating” product OR equipment**

However, since home heating is synonymous with residential heating and house heating, we request search for such synonymous words using the tilde operator, “~”:

- **“biomass ~home-heating” product OR equipment**

To this basic search, we add several other refinements:

(i) **site**: Since only governments and companies publish information about biomass heating products we limit our search to “.gov, .com and .org” domains;

(ii) **fieltype**: Often company or industry data are published in pdf or xls/xlsx format. Hence we restrict our search to documents with these extensions;

(iii) **allintitle**: as a variation to the basic search we request Google to search for all these words in page titles:

> biomass home heating
(iv)  *daterange*: we also restrict the search results to pages added or updated since 2000;

(v)  *related*: to build a list of industry associations similar to pellet fuels institute, for example, we use the related operator:

> related:http://www.pelletheat.org

(vi) similarly, we explore webpages linked to biomass industry associations using the “link” operator, for example,

> link:http://www.pelletheat.org

While the above search strategy is expected to yield publicly available pages, we also probe confidential industry documents using both ‘filetype’ and ‘daterange’ operators:

> confidential “biomass heating” report filetype:pdf 2009..2014 u.s.a. | usa | u.s.a. | united states | america.

Another unavailable industry-source-of-information is Forrester Research Reports. These reports typically cost around $2,000. In spite of this prohibitive cost, some of these reports could have found their way online, albeit by accidental uploading of the report by a genuine purchaser or customer of Forrester. To find such reports, we make use of the fact that two words often appear in Forrester Reports: grapevine and endnotes (Spencer, 2011). Specifically, using the ‘filetype’ operator, we search for:

> forrester research grapevine endnotes filetype:pdf

### 2.4 Google tools

In addition to document searches, we also utilise:

(i)  Google Ngram Viewer: it lets one to visualise the frequency of word(s) usage in books over the last 200 years; this measure is a proxy for “popularity” of a concept (product form, product type, brand, etc.). We deploy this tool to assess the popularity of the product form ‘biomass home heating’ and product types such as wood stove and pellet stove.

(ii)  Google Trends: It analyses “search” proportion of search terms; for example, the proportion of “biomass” search as compared to all search terms.

(iii)  Google Correlate: We utilise this algorithm to visualise the correlates of biomass search terms across the US. The application’s URL is: [http://www.google.com/trends/correlate/](http://www.google.com/trends/correlate/)
(iv) Google Public Data Explorer: We utilise this application to view macro indicators of biomass heating: for example, combustible renewables as percentage of total energy in the US, and

(v) Google Think Insights: We use the “Brand Impression” tool to track pictures, words and actions related to biomass thermal energy.

2.5 Results

The Internet search for biomass heating should be optimal; the results should be the best that could be obtained by the search. To achieve the optimum, we utilise WordNet and semantic similarity measures. WordNet is a semantically oriented dictionary of English that is part of the Python NLTK module (Cocos, Apidianaki & Callison-Burch, 2017). WordNet defines biomass as ‘plant materials and animal waste used as fuel’. Based on the notion of lexical relations, that concepts are related in superordinate / subordinate structure (cf. the spreading activation model in cognitive psychology), various similarity measures have been calibrated (Saravanakumar, 2014). One such measure is path similarity; it assigns a score in the range of zero to one, based on the shortest path that connects the concepts in the hierarchy. Comparing a concept with itself will result in a perfect score of one. For example, WordNet’s definition of biomass includes the concept ‘fuel’ and this association produced a similarity score of 0.5. Biomass and product produced a 0.1 score; biomass and heater had a 0.09 score (Table 1). Using the logical process of induction by enumeration, we conclude that collocations ‘biomass home-heating product’ would produce the best search engine results.

<table>
<thead>
<tr>
<th>Semantic similarity</th>
<th>Path similarity score</th>
<th>Google hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomass and fuel</td>
<td>0.5</td>
<td>415,000</td>
</tr>
<tr>
<td>Biomass and product</td>
<td>0.1</td>
<td>87,900</td>
</tr>
<tr>
<td>Biomass and heater</td>
<td>0.09</td>
<td>17,100</td>
</tr>
</tbody>
</table>

The basic search with the tilde operator resulted in 79,600 hits. However, with the application of ‘sites’ and ‘filetype’ criteria, only seven pages had unique or unduplicated information and these were largely product ads and brochures. To gain better, more focused search results, we explored Google tools.
2.6 Industry structure

The conceptual model based on the industrial organisation’s market-structure approach suggests that industry subsumes products, for example, SIC code 3433 includes all heating equipment products except electrical ones. In general, products are functional-need satisfiers (Abell, 1980) albeit with different technologies (for example, outdoor hydronic heating systems, indoor pellet furnaces, etc.). To gain insights into the life stages of differing biomass heating equipment, we implemented the Google Correlate algorithm. Based on the notion that the newer the product the higher the consumer search for the product (cf. Howard’s (1989) EPS model), we looked for correlations of home heating. The term ‘stoves’ exhibited one of the largest correlations with ‘home heating’, the correlation coefficient was 0.93. Next, we examined the correlations for “stoves”. This run produced 90 relationships, each with greater than a 0.8 correlation coefficients.

Table 2 categorises these relationships either as a product form (for example, biomass, fuel oil, etc.), product category (for example, wood stove, pellet furnace, etc.) or manufacturer / brand (for example, Lopi Stoves). The correlations in Table 2 reveal that:

(i) Biomass form of home heating is growing;
(ii) Pellet stoves and fireplace inserts top the lists of “stove” searches and
(iii) Lopi Stoves is the top-ranked and sought-after manufacturer of home heating equipment.

The correlations, when interpreted within a consumer-information-processing framework such as the target-audiences learning framework discussed by Rossiter and Percy (1997), reveal that ‘pellet stove’ is in the growth stage and biomass furnace is in the introductory stage of its life. This categorisation is based on the reasoning that ‘more product awareness leads to more search’ (cf. the Attention ➔ Learning framework of Rossiter and Percy (1997)). Since more product awareness occurs during the growth stages of the product, the search correlations in Table 2 can be thought of as indicating product’s life-stages. The lowest ranked correlation in the product category, wood/ gas fireplaces is the mature product and the highest ranked correlation (wood/ pellet furnaces) is in the introductory stage. This pattern discovery, or association rule, as Agarwal, Aggarwal & Prasad (2001) put it, was validated by industry experts at the Pellet Fuels Institute’s Annual Conference in 2014 (Athiyaman, 2014).
Table 2: Correlates of ‘Stove’

<table>
<thead>
<tr>
<th>Category</th>
<th>Level</th>
<th>Strength of correlation</th>
<th>Lifecycle stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product form</td>
<td>Biomass</td>
<td>.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fuel oil</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Electric / Gas</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>Product category</td>
<td>Wood / pellet furnaces</td>
<td>.90</td>
<td>Introduction</td>
</tr>
<tr>
<td></td>
<td>Wood and pellet stoves</td>
<td>.89</td>
<td>Growth</td>
</tr>
<tr>
<td></td>
<td>Wood / gas fireplaces</td>
<td>.87</td>
<td>Mature</td>
</tr>
<tr>
<td>Brand (Top three stoves; see Appendix 1 for complete listing of brands)</td>
<td>Lopi Stoves</td>
<td>.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quadra-Fire</td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jotul</td>
<td>.93</td>
<td></td>
</tr>
</tbody>
</table>

2.7 Intra-industry competition

Competition among firms occurs when customers regard certain products and/or brands as close substitutes or purchase alternatives. To identify competitors in the wood/ pellet stove category and wood/ pellet furnace category, we employed a distance-and-similarity measure. Specifically, based on the principle that competition is intense among firms with strong brand presence; brand presence is analogous to consumers’ brand awareness concept (Biel, 1990). The brands listed in Appendix 1 were submitted to Google’s “brand impression” analysis. The resulting brand-visuals-metrics – the most viewed YouTube video of the brand – were scaled for brand-pair similarity using the formulae:

\[ d_{ij} = \sqrt{(x_i - x_j)^2} \]

where, \( x \) is the number of views each for brands \( i \) and \( j \).

A subsequent hierarchical clustering of \( d_{ij} \) resulted in two competitive clusters each for the wood /pellet stove category and the wood/ pellet furnace category.
(Maddala & Lahiri, 1992). On average, competitive cluster 2 firms were larger than cluster 1 firms (Appendix 1).

2.8 Firm behaviour and performance

The google search “confidential “SIC 3433” report filetype:pdf” took us to the Annual Manufacturing Survey (AMS) website. A search for SIC 3433 in the AMS website produced the information needed to construct firm behaviour and performance metrics for small and large firms (Table 3).

Table 3: Competitive Behaviour (SIC: 3433: Heating Equipment, Except Electric)

<table>
<thead>
<tr>
<th></th>
<th>Competitive Cluster</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 (Small Firms)</td>
<td>2 (Large Firms)</td>
<td></td>
</tr>
<tr>
<td>Net sales</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Materials cost (proportion of sales)</td>
<td>.4</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td>Direct labour cost</td>
<td>.047</td>
<td>.057</td>
<td></td>
</tr>
<tr>
<td>Gross profit (margin)</td>
<td>.59</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Advertising / marketing</td>
<td>.005</td>
<td>.005</td>
<td></td>
</tr>
<tr>
<td>EBT</td>
<td>.044</td>
<td>.039</td>
<td></td>
</tr>
</tbody>
</table>

Source: 2016 Annual Survey of Manufacturers

2.9 Discussion

Herewith the Internet-driven market intelligence for biomass home heating:

✔️ The ‘pellet stove’ product category is in the growth stage of its life and biomass-furnace category (central heating) is in the introductory stage of its life.
✔️ Customers regard Jotul and Whitfield wood/ pellet stoves as purchase alternatives.
✔️ Breckwell and Woodmaster biomass furnaces are perceived as substitutes.
✔️ Analysis of performance reports for SIC 3433 suggests a low 0.5 percent industry expenditure on advertising/ marketing.

For products in the introductory or the growth stage of life cycle, not only must the product be readily available, but the buyer must also be aware of the product and its benefits. Given a 0.5 percent sales-to-advertising ratio for SIC 3433, it is doubtful that customers are being educated about biomass home heating.

What should the advertising/ marketing expenditure be for the products?

Sethuraman and Tellis (1991) posit that the average advertising elasticity for
durable product is 0.23. Our interest is in assessing the sales-response of biomass-residential-heating industry’s marketing spending. Therefore, we employed a dynamic sales-response model of the Vidale-Wolfe (1957) form:

\[
\frac{dQ}{dt} = \frac{rx(V-Q)}{V} - \lambda Q
\]

where \( Q \) = sales volume expressed in monetary terms;
\[
\frac{dQ}{dt} = \text{change in sales at time } t; \ X = \text{marketing spending};
\]
\( V \) = market volume or potential in monetary terms;
\( r \) = sales response constant and
\( \lambda \) = sales-decay constant.

The model was calibrated using HPBAs industry data (hearth shipment statistics for the US), market potential data from Athiyaman (2015a) and industry marketing expenditure data from ASM. Table 4 shows data sources and sample values of variables.

### Table 4: Data for Sales Response Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sample value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>$30.3mil</td>
<td><a href="https://www.census.gov/programs-surveys/acs/">https://www.census.gov/programs-surveys/acs/</a></td>
</tr>
<tr>
<td>V</td>
<td>1,432,390 units</td>
<td>Athiyaman (2015a)</td>
</tr>
</tbody>
</table>

The results of the model run suggest a parameter estimate of 10.62 for \( r \), the monetary or $ value of sales response. The optimal marketing spending required for the sale of 1.43mil units was calculated by setting:

\[
\frac{dQ}{dt} = 0
\]

The results suggest that the industry has to increase their marketing spending to $16.83 mil from its current spending of $30.5mil.

Extant research on business use of market intelligence suggests that companies seldom act on market intelligence (Stephenson & Pandit, 2008). Given the positive impact of marketing spending on industry profits, it is hoped that biomass
home heating companies will act on the intelligence; the mismatch between industry’s current marketing expenditure and desired marketing expenditure is 55 percent.

3. SUMMARY AND CONCLUSION

A unique feature of this work is the exclusive use of the Internet to gather intelligence about the industry. Web business intelligence has two components: content acquisition and knowledge creation (Choo, Detlor & Turnbull, 2013). This paper shows how Google Tools can assist in information retrieval for industry analysis. Although this is a simple, manual gather, web pages and documents were downloaded by the author; this method gives the highest precision since only relevant information is downloaded (Srivastava & Cooley, 2003). Regarding knowledge creation, statistical tools such as clustering to gain insights into industry conduct were used.

Business decisions will not get any better without deep market insights. Biomass home heating industry must be open to experimentation to manage their product lifecycle better. Companies prepared to experiment with the strategy(s) recommended in this paper will have the opportunity to gain competitive advantage.
Appendix 1: Product Category and Brands

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Brand</th>
<th>Competitive Cluster</th>
<th>Brand Impression Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood / Pellet Stove</td>
<td>Breckwell</td>
<td>1</td>
<td>32982</td>
</tr>
<tr>
<td></td>
<td>Country Flame</td>
<td>1</td>
<td>25753</td>
</tr>
<tr>
<td></td>
<td>Dutchwest</td>
<td>1</td>
<td>11298</td>
</tr>
<tr>
<td></td>
<td>Earth Stove</td>
<td>1</td>
<td>1556</td>
</tr>
<tr>
<td></td>
<td>Englander</td>
<td>1</td>
<td>21914</td>
</tr>
<tr>
<td></td>
<td>Heatilator</td>
<td>1</td>
<td>34602</td>
</tr>
<tr>
<td></td>
<td>Jotul</td>
<td>2</td>
<td>62816</td>
</tr>
<tr>
<td></td>
<td>Pacific Energy</td>
<td>1</td>
<td>28901</td>
</tr>
<tr>
<td></td>
<td>Quadra-Fire</td>
<td>1</td>
<td>32992</td>
</tr>
<tr>
<td></td>
<td>Soapstone Stove</td>
<td>1</td>
<td>32173</td>
</tr>
<tr>
<td></td>
<td>Travis Industries</td>
<td>1</td>
<td>18027</td>
</tr>
<tr>
<td></td>
<td>Whitfield Pellet Stoves</td>
<td>2</td>
<td>70285</td>
</tr>
<tr>
<td>Wood/Pellet Furnace</td>
<td>America’s Heat</td>
<td>1</td>
<td>1021</td>
</tr>
<tr>
<td></td>
<td>Breckwell</td>
<td>2</td>
<td>3373</td>
</tr>
<tr>
<td></td>
<td>Fahrenheit Tech</td>
<td>1</td>
<td>1030</td>
</tr>
<tr>
<td></td>
<td>Harman</td>
<td>1</td>
<td>2044</td>
</tr>
<tr>
<td></td>
<td>Ozark Biomass</td>
<td>1</td>
<td>1319</td>
</tr>
<tr>
<td></td>
<td>WoodMaster</td>
<td>2</td>
<td>3420</td>
</tr>
</tbody>
</table>

REFERENCES


14


