Estimate the Market Share from the Search Engine Hit Counts

Robert Viseur

Faculté Polytechnique (Université de Mons), Rue de Houdain, 9, 7000 Mons, Belgium

Keywords: Marketing, Advertising, Market Share, Media Share, Share of Voice, Word-of-Mouth, Management of Innovation, Search Engine, API, Webometrics.

Abstract: The knowledge of the competitive environment (and, in particular, market share) is an important factor in the management of innovation. This type of information is not always accessible to small and medium enterprises. In addition, some sectors are changing rapidly under the pressure of technological change. We propose in this research a method for estimating the market share based on media share, based on the hit counts returned by search engines for each brand. We show the potential of this approach with a real example (the automotive industry) and discuss the limitations associated with the operating mode of search engines.

1 INTRODUCTION

In an article published in 2005 entitled "When search engines occupy the media space...", Olivier Andrieu, a French specialist in commercial search engines, noted that the Google search engine exceeded its competitors (Yahoo, MSN and, on the French market, Exalead) not only in terms of market share but also in terms of media coverage. Google produced from two to three times more adverts than its competitors. The measure of the number of adverts was based on the press review of a major news website devoted to commercial search engines.

In the management of innovation, market share is an important parameter for the understanding of the competitive environment (Lambin, 1998); (Porter, 1992). Information on market share is generally provided by panels of consumers and retailers. In the absence of official or professional statistics, the company must buy or edit this information. Access to this information presents several challenges. Firstly, the definition of market share can vary depending on the data source (Lambin, 1998). Does it speak of market shares in volume or revenue, absolute or relative market share, etc.? Secondly, the VSEs (Very Small Enterprises) and SMEs (Small and Medium Enterprises) do not always have the resources to collect, acquire and operate this type of data. Thirdly, some markets have experienced rapid and significant turbulence under the influence of new technologies (Millier, 1997). Historical data may therefore be quickly outdated.

The need for updated information on the competitive position of a company (and, in particular, the market shares of its competitors) and the existence of a possible link between market share and media coverage of the company concerned motivated us to study the relationship between media share and market share, as well as the feasibility of deriving a simple and affordable way of estimating market share. Market share is a well-known concept, which means "*the percentage of sales held by each competitor in the market*" (Kotler and Dubois, 2000; p255). The concept of "media share" that is exploited here will be defined by analogy with market share, as "*the percentage of published documents citing a company compared to the number of documents about companies in the relevant market*". It is here estimated from hit counts related to the results of a commercial search engine. Our research is applied in the automotive market.

Our research is divided into three parts. Firstly we will describe the state of the art. We shall examine the link between advertising and market share, and deepen the definition of media share. We will then document the potential biases in the method, including those induced by practical operating conditions of commercial search engines. Secondly we will present our methodology. Thirdly we will present our results and discuss them, particularly with regard to results from comparable approaches.

2 STATE OF THE ART

2.1 Advertising and Market Shares

Lambin (1998) defines marketing communications as "*the set of signals issued by the company towards its audiences*" (p. 615). The company has a set of tools to facilitate the matching of supply and demand. Advertising is part of the communications mix, alongside sales promotion, public relations, salesforce and direct marketing. It is used by organizations to develop their reputation or that of their products, services or ideas (Kotler, 1999).

The relationship between advertising and market share has long been a subject for study in marketing. The existence of a causal link between advertising on the one hand and sales on the other hand is thus assumed (Eagle et al., 2005). However its importance according to the market, the company size or the type of product is still discussed.

2.2 Concept of Share of Media

The concept of "media share" will be defined by analogy with market share, as "*the percentage of published documents citing companies compared to the number of documents about companies in the relevant market*".

This concept is differentiated from "*share of voice*", "*social share of voice*" and "*word-of-mouth*". Share of voice is the share of advertising expenditure of the company compared to total advertising spending in the relevant market. Social share of voice refers to the importance of a term compared to a set of terms on social medias (Emerson and al., 2012). This notion is closer to word-of-mouth, which means existing interpersonal communication, especially between the consumer and the environment, face to face or, increasingly, on the Web (Kotler and Dubois, 2000).

The concept of media share is clearly differentiated from share of voice (there is no question of budget). It is, however, closer to social share of voice. Its scope is broader than just social networking. Share of media can be related to word-of-mouth, as it includes its effects. Share of media also includes the communication about the brand in the press or communications operations relayed by the company itself (e.g. press releases).

2.3 Possible Biases within the proposed Method

2.3.1 Reliability of Hit Counts by Commercial Search Engines

Using results from a search engine involves a decision on how to interact with it. For example, Google (google.fr) offers two ways of accessing its search engine: manually, through its WUI (Web User Interface), and through its API (Application Programming Interface) (code.google.com). The latter has the apparent advantage of permitting the automation of queries, thus facilitating the repetition of various tests over time. However this opportunity comes up in practice against large differences in results observed between the results from the Web interface and those derived from the API, with the same search parameters (same keywords, same geographical area or language). This would not be crippling to the researcher if Google did not prohibit the execution of automated queries on the WUI, sparking a spontaneous use of the API in this particular context of use (McCown and Nelson, 2007).

The use of quantitative data from the search engines is part of a research field called webometrics (Thelwall et al., 2005). The behaviour of APIs has been studied in the literature. McCown and Nelson (2007) found significant differences in hit counts between the WUI and API of MSN, Yahoo! and Google. They confirmed the results obtained by Mayr and Tosques (2005) with the Google API. However, problems are encountered beyond the simple use of search engines API. Thus recovery is weak between the results of different search engines (Véronis, 2006). Despite the global reach of major commercial search engines, geographic bias may exist. In addition, the hit counts of search engines are not stable over time, while complex queries do not always give results consistent with the theory of sets (Boolean logic) (Viseur, 2012a).

The importance of the problems varies from one tool to another. Early in 2012, Bing presented a more predictable behaviour than its competitor Google, when using both API or complex queries (Viseur, 2012). Due to the speed of technological developments it is nevertheless necessary to periodically reassess these conclusions. Using the Web user interface (WUI) for the collection of search results using simple queries globally poses no problem whatsoever on Google (google.com) or Bing (bing.com).

2.3.2 Representativeness of Web Users

Several media can be selected for a campaign: press, television, billboard, radio and cinema (Kotler and Dubois, 2000). A sixth media was added in the nineties: the Internet. It has gradually grown in importance. In 2005, more than one out of two French people were connected (Roustan and al., 2005).

However, are consumers on the Web representative of all consumers in the market ? In practice, they tend to be.

According to Médiamétrie (www.mediametrie.fr), the number of connected French people has increased in size from 27.21 to 38.27 millions between December 2005 and December 2010. And, according to the IAB (www.iabfrance.com), Internet has become the preferred medium of French people just behind television. The public has also logged fewer specificities.

Furthermore, the Internet significantly affects the purchasing process, fostering opportunistic behaviour and encouraging cross information upstream of the buying process. Valuing the opinions of peers is becoming more important and is a "*tool of redistribution of power between consumers and producers-distributors*" (Roustan et al., 2005; p12). More than just a channel of purchase, Internet currently appears to be an information channel strengthening consumer expertise. Even if not buying online, the user willingly prepares his/her purchase through information on the Internet. Search engines are a key access point to information behind the websites of famous stores.

2.3.3 Impact of Negative Word-of-Mouth

However the operation of search engines can cause unexpected effects in the way the reputation of young companies grows. The online store DecorMyEyes (decormyeyes.com), which sells glasses, boasted at the end of 2010 of how it had really commercially taken off after creating a negative buzz around its brand. Negative opinions of disgruntled customers (and intentionally mistreated by the company) have been generating traffic and backlinks to the company website, helping the company to improve its Web positioning (Segal, 2010).

The atypical communication strategy adopted by DecorMyEyes is an opportunistic exploitation of a current weakness of search engines. They do not actually measure the semantic orientation[1] of the information published in the context of a quote link. They are mainly based on syntactic relevance and backlinks (seen as citations), thus following the principles of the PageRank algorithm (Brin and Page, 1998); (Duffez and Andrieu, 2004).

This shortcoming of search engines is a possible bias in the estimation returned by the search engine. A company would be heavily criticized and associated with a bigger estimated market share. This bias is compounded by the fact that a dissatisfied customer tends to speak more than a satisfied customer. The customer who received an effective solution speaks positively of his experience to at least five people around him (Kotler and Dubois, 2000). By contrast, an angry customer will talk about his misadventure to 11 people, heavily affecting the reputation of the company (Kotler, 1999). A communication policy based on a negative buzz may thus jeopardize the company's sales over the long term. The consumers' behaviour will depend on the final processing of the complaint. Indeed, the loyalty of a customer depends on his/her satisfaction (Kotler and Dubois, 2000).

This is particularly the case in competitive markets (Lambin, 1998). A claim, reflecting dissatisfaction, does not necessarily lead to customer defection, if it is addressed appropriately.

3 METHODOLOGY

We propose to study the relationship between market share and the media share. The media share will be estimated from the number of results (hit counts) returned by a search engine for a considered brand.

We carried out our study in the automotive market, because its sales figures are readily available and provided by recognized national organizations (FEBIAC in Belgium, CCFA in France, etc.). We will use the figures for 2010. Less volatile and well documented, this market seems appropriate for validating the usefulness of our approach.

The hit counts estimated by search engines have reliability problems (Viseur, 2012a). Our experimentation showed the Bing search engine provided the most consistent results, and we used the number of results estimated by the Microsoft Bing search engine. The operator "loc:" was used for

---

[1] A sentence is characterized by a positive semantic orientation when it has positive associations, negative otherwise (Turney, 2002).

geographic targeting ("FR", "DE", "GB"). The measurements on the search engine were made in late 2011, after preliminary tests in 2010. Note that the calculation of the media share may require a reformulation of queries (e.g. Volkswagen is often called "VW", Opel cars are sold under the Vauxhall brand in the UK, etc.) or the elimination of certain brands (e.g. "smart" is a common term in English). We will verify the following hypothesis: "Market share and media share are proportionate".

We will not test the hypothesis that the market share is negatively influenced by negative communication about a brand because of the practical difficulty in assessing the importance of positive or negative communication. Turney's method (2002) requires the execution of a large number of automated queries, and relies on linguistic tricks that are difficult to generalize in this case (use of a specific vocabulary for product reviews on the Internet, for example).

4 RESULTS

The correlation (Pearson coefficient) between the market share and media share of 25 French brands stands at 0.73 (see Table 1). The correlation can be considered as strong.

This correlation is higher when we distinguish between the premium brands (Audi, BMW, Mercedes, Alfa Romeo, Volvo, Lancia, Land Rover, Porsche and Lexus) with a distinctive character linked to exclusivity (Štrach and Everett, 2005), and the general brands (others). The correlation (Pearson coefficient) between market share and media share for generalist brands amounted to 0.83, against 0.88 for premium brands.

Table 1: Correlation coefficient.

	Germany	France	United-Kingdom
All brands	0.76	0.73	0.78
Generalist carmakers only	0.75	0.83	0.89
Premium carmakers only	0.75	0.88	0.56

The premium brands also have a greater media share than their market share.

A similar study on German data gave comparable results, with a correlation of 0.76 (Pearson coefficient) by taking the first 18 manufacturers, 0.75 taking only generalist carmakers and 0.75 with only premium manufacturers.

The results are comparable with the UK, except for premium brands. The correlation of 0.56 for premium carmakers can be explained by the absence of some popular brands in the UK (such as Jaguar), for which the market shares were not disclosed.

5 RELATED WORKS

Uncles et al. (2010) studied the impact of word-of-mouth on market share. They distinguished the effects of the total volume of word-of-mouth (WOM), the volume of positive word-of-mouth (PWOM) and the volume of negative word-of-mouth (NWOM). The authors confirm the existence of a strong correlation between market share on the one hand, and word-of-mouth (WOM) and positive word-of-mouth (PWOM) on the other hand.

Media share is strongly influenced by word-of-mouth on the Web. The results of the Uncles et al., (2010)'s study tend to confirm the validity of our approach.

Xu et al., (2010) chose a similar approach to ours. They studied the correlation between quarterly market shares of cell phone manufacturers (Nokia and Motorola) and the relative volume of Web search estimated on the basis of Google Trends (www.google.com/trends/). The goal is to detect quickly, within a turbulent market, trends and changes in the market shares of competing firms. The authors concluded that there was a strong correlation between market share and user activity measured by the volume of queries about brands.

Our conclusion is confirmed by the work of Xu et al. (2010). However their approach should provide a higher reactivity. On the other hand, it presupposes the availability of temporal data for market shares.

6 DISCUSSION AND PERSPECTIVES

The correlation between market share and media share appears strong in the automotive sector. It increases again by sorting brands, distinguishing between generalist brands and premium brands. A precise definition of the reference market is important for the proper functioning of the method. The results confirm that our approach provides an opportunity for estimating the market share and the balance of power between companies based on the media share.

This preliminary study also confirmed the limitations associated with the use of the Google search engine as a tool for webometrics, especially

when the API is used. This made the production and the communication of a study to update and deepen the potential biases of search engines API within webometrics studies: see Viseur, 2012a.

The fact that media share of premium brands seems to be greater than their market share could be explained by the additional communication about premium brands, which are able to generate passion and elicit comments in the official press or participatory media (forums, blogs, etc.). Albert et al. (2012) indeed confirm that the brand passion, i.e. "*the strong positive feeling towards a brand*", is linked to unique and prestigious brands, and in turn leads to greater positive word-of-mouth.

Several points could be studied further.

Firstly, the study could be deepened for the automotive sector. The method could be applied to models of cars, in order to assess if it could also be used to estimate the market share of products and services (and not just brands).

Secondly, although this first experiment was conducted on a mature market, our aim is to obtain a method applicable to more turbulent markets. The smartphone market is one of those markets. The application of the method presents practical difficulties related to the bias caused by the accumulation of historical data in commercial search engines. Several approaches could improve this method. They include the use of news search engines, the use of Twitter timeline content, the use of the Google "daterange" operator or the use of a custom search engine with a specific index (see e.g. Viseur, 2012b).

Thirdly, the impact of negative opinions on the market shares of brands could be further explored. The impact of negative opinions on market share is not easy to quantify. See the astonishing example of the DecorMyEyes online store. Uncles et al. (2010)'s works provide an initial insight into the effects of the total volume of word-of-mouth (WOM), the volume of positive word-of-mouth (PWOM) and the volume of negative word-of-mouth (NWOM).

REFERENCES

Albert, N., Merunka, D., Valette-Florence, P., 2012, Brand passion: Antecedents and consequences, *Journal of Business Research*, available online 5 January 2012.

Andrieu, O., Duffez O., 2004. *Google : Trucs de pros*, Editions Micro Application.

Andrieu, O., 2005. Quand les moteurs de recherche occupent l'espace média..., *Abondance.com* (retrieved April 10, 2013).

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual Web search engine, *Seventh International World-Wide Web Conference (WWW 1998)*, April 14-18, 1998, Brisbane, Australia.

Cafferky, M. E., 1995. Let Your Customers Do the Talking, *Upstart Pub Co.*

Eagle, L., Kitchen, P. J., Rose, L., 2005. Defending brand advertising's share of voice: A mature market (s) perspective, *The Journal of Brand Management*, 13(1), pp. 65-79.

Emerson, T., Ghosh, R., Smith, E., 2012. CASE STUDY: Using the Social Share of Voice to Predict Events That Are about to Happen, *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*, pp. 127-131.

Kotler, P., 1999. Le marketing selon Kotler, *Editions Village Mondial*, Paris.

Kotler, P., Dubois, B., 2000. *Marketing management*, Publi-union Editions, Paris.

Lambin, J.-J., 1998. *Le marketing stratégique*, Ediscience International, Paris.

Mayr, P., Tosques, F., 2005. Google Web APIs: an instrument for webometrics analyses, *Proceedings of the ISSI conference*.

McCown, F., Nelson, M. L., 2007. Agreeing to Disagree: Search Engines and their Public Interfaces, *ACM IEEE Joint Conference on Digital Libraries (JCDL 2007)*, June 17-23, 2007. Vancouver, BC, Canada. p. 309-318.

Millier, P., 1997, *Stratégie marketing de l'innovation technologique*, Dunod.

Porter, M., 1992. *L'avantage concurrentiel*, InterEditions.

Roustan M., Lehuede, F., Hebel, P., 2005. Qu'est-ce qu'Internet a changé aux modes d'achat des Français ?, *CREDOC (www.credoc.fr), Cahier de Recherche*, n°213, novembre 2005.

Segal, D., 2010. A Bully Finds a Pulpit on the Web, *The New York Times*, 26 novembre 2010. URL: http://www.nytimes.com/2010/11/28/business/28borker.html (retrieved April 10, 2013).

Štrach, P., Everett, A. M., 2005. Globalizing Luxury Automobiles through Mergers: Three Brands at the Crossroads», *Working Paper*, No. 5/2005.

Thelwall, M., Vaughan, L., Björneborn, L., 2005. Webometrics, *Annual Review of Information Science and Technology*, 39, 81-135.

Turney, P., 2002. Thumbs up or thumbs down ? Semantic orientation applied to unsupervised classification of reviews, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, p. 417-424, Philadelphia.

Uncles, M. D., East, R., Lomax, W., 2010, Market share is correlated with word-of-mouth volume, *Australasian Marketing Journal (AMJ)*, Volume 18, Issue 3, August 2010, pp. 145-150.

Véronis, J., 2006. Etude comparative de six moteurs de recherche, *Université de Provence*, 23 février 2006.

Viseur, R., 2012a. Les moteurs de recherche commerciaux sont-ils des outils de webométrie fiables ?, *Actes du 30ème congrès InforSID*, Montpellier (France), 29-31 mai 2012.

Viseur, R., 2012b. Create a specialized search engine: The case of an RSS search engine, *Proceedings of Data 2012 Conference*, Rome (Italy), July 25-27, 2012.

Xu, K., Xu, J., Liu, L., Ren, J. S. J., Wang, W., Liao, S. S., Song, Y., 2010. Predict Market Share with Users' Online Activities Data: An Initial Study on Market Share and Search Index of Mobile Phone. *PACIS 2010 Proceedings*.