Chapter VIII

Association between Web Semantics and Geographical Locations:
Evaluate Internet Information Consistency for Marketing Research

Kimman Lui,
The Hong Kong Polytechnic University, Hong Kong

Keith C. C. Chan,
The Hong Kong Polytechnic University, Hong Kong

Kai-Pan Mark,
City University of Hong Kong, Hong Kong

ABSTRACT

Social software completely revolutionizes the way of information sharing by allowing every individual to read, share and publish online. In terms of marketing, it is an effective way to understand consumers’ perceptions and beliefs in different local regions by analyzing and comparing the web content regarding a specific product retrieved on the Internet with respect to different locations. Interestingly, incidents originated from a location may attract more Internet discussions by individuals from remote locations. Therefore, it is difficult to measure the strength of people’s perceptions between different locations if we solely rely on the web traffic statistics. Moreover, it is difficult to compare strength of perceptions retrieved by different search engines, at different times, and on different topics. To overcome these inadequacies, the authors introduce a quantitative metric, Perceived Index on Information (PI), to measure the strength of web content over different search engines, different time intervals, and different topics with respect to geographical locations. Further visualizing PI in maps provides an instant and low-cost mean for word-of-mouth analysis that brings competitive advantages in business marketing.
INTRODUCTION

Market research is concerned with identifying the tastes, values, preferences and buying patterns and selection behavior of consumers. Nowadays, there is a considerable amount of such consumer information available on the Internet at public forums such as blogs, internet special interest communities and social networking sites. While these discussions are of interest to marketers in themselves, they are also a source of information about goods and services for consumers who are not active participants in these forums but who use search engines to search for product information and as a result will obtain numerous results of varied provenance and reliability from a large number of sites as a basis for their buying decisions. However, analyzing such consumer information published on the Internet is always challenging because: (1) the geographical origin of an individuals’ post in an online community is insufficient for identifying the geographical origin of the event; (2) consumer information is highly dynamic and is changed rapidly; (3) there exists different results when different search engines process the same query; and (4) there is a lack of an empirical way to compare different search results.

Identifying the origin of an event is important for localizing marketing campaigns. Different web content contains rich information for marketing research. Marketing information retrieved from web contents, through either reading manually, or by machine analysis, contains facts and localized peer opinions that are of high value to marketers to launch local marketing campaigns. Further knowing geographically where the peer opinions regarding a specific product are from, marketers can have a picture on the market response from different local markets. Associating different peer opinions with the source of geographical regions allows marketers to launch localized marketing campaigns easily and rapidly.

Understanding the difference in search engines results produced by the same query enables a marketer to know what consumer information are perceived by consumers through different search engines. In addition, it will be a marketer’s interest to trace the change in the strength of consumers’ perceptions by comparing two sets of search results at different times, or on different topics. The issue to be considered in this chapter is how to determine whether two sets of search results are consistent and to comparing the search results empirically.

The organization of this chapter is as follows. The next section gives the background on different approaches on information analysis research by information systems, computer science and marketing researchers. Next followed by the related works on web browsing and information sharing behaviors on information systems discipline, and on web analysis with the more technical aspect on traffic analysis and web content analysis. Then, we point out the inaccuracy in the existing systems and propose a new algorithm on measuring the strength of perception with respect to topics, time, search engines and geographical locations. Afterwards, we discuss the application and present industry cases on how this algorithm benefits marketers. Finally, we conclude this chapter by pointing out the future directions on this research and the contributions of our chapter.

BACKGROUND

Individual’s information sharing behaviors on the Internet is a popular field in information systems research (e.g. Bock et al., 2005; Wasko & Faraj, 2005). Surprisingly, people are motivated intrinsically in sharing their knowledge. Reward by money is not the primarily goal in publishing information on the Web for the public. More and more consumers prefer to publish comments on personal blogs and discussion forums. As a