Competitive Intelligence applied to Tourism Destination Management: Hotel Market Monitor

Jon Kepa Gerrikagoitia, Aukene Alzua, Igor Ibarguren, Ibai Roman, Nagore Espinosa

CICtourGUNE - Centre for Cooperative Research in Tourism
Paseo Mikeletegi 56 - 201, 20009 Donostia-San Sebastián, Spain
{jonkepagerrikagoitia, aurkenealzua,igoribarguren,ibairoman, nagoreespinosa}
@tourgune.org

Abstract

Destination competitiveness is constrained by the ability to design appropriate policies, which must be based on quality information provided by competitive intelligence tools that systematize the collection, analysis and dissemination of relevant, accurate, specific, timely, and active information about the business environment, its competitors and the organization itself (SCIP, 2005).

The tourism value chain’s changes, caused by the development of ICTs and the Internet, with an evident impact on the marketing system, transforms agents and online distribution channels into data sources which reflect the tourism market behavior. Hence, the hotel market monitor project is presented as a tool based on such online digital footprint, and is able to calculate the evolution of hotel rates, to estimate load and to predict room availability within 14 days.

Keywords: market research, competitive intelligence, web scrapping, web crawling, revenue management
1. Introduction

Competitive Intelligence is one of the emerging trends that is triggering the awakening of a growing interest in the field of strategic management and eScience. The benefits of implementing a good conceptualization based on competitive intelligence and of keeping it updated through strategic surveillance (integration of technology surveillance and markets), can provide the organization with the ability of anticipating and reacting to the changing environment. In addition, having the information organized, sorted and analyzed conveniently offers a comprehensive overview of the competitive environment. The macro trends aid forecasting the reality that tourism professionals are witnessing daily both for the obvious as to what can go unnoticed.

Thus, within the Competitive Intelligence strategy framework, market research is the process of gathering, processing and analyzing information on issues related to marketing, customers, competitors and market. The purpose of market research is to assist organizations in taking better decisions on the development of different products. A basic list of questions which can be answered through market research could be: What is happening in the market? Which are the current trends? Who are the competitors? How are our products positioned in the minds of consumers? Which needs are important to consumers? Are the needs being met by the products on the market?

Market research is a discipline that has contributed to the development of marketing and the different industries themselves over the past fifty years, being fed by the integration of multiple disciplines such as psychology, anthropology, sociology, economics, statistics, communication, among others, following the methodological approach of the social sciences. The traditional social sciences methodology strives to find the valid or appropriate data for a specific purpose, so that such data are reliable enough to eventually provide an explanation. The advent of the Internet with all its possibilities of developing e-business and e-commerce has required a completely different approach away from traditional marketing.

The key objective of the new methods of market measuring and monitoring is to combine a highly diverse range of techniques to create and maximize customer value, such as understanding customers and markets in which they reside, anticipating and responding quickly to changes in the environment and creating more value added, while remaining above the competitors. Moreover, the importance of Internet ecommerce portals in this research cannot be missed, since they are the final means which will develop and implement a marketing strategy. The research work presented takes into account this new reality, and proposes new methods based on real-time web channels to meet and monitor markets. These methods propose new scenarios and advanced prototypes as opposed to the software solutions provided by the current software industry.

Nowadays we are experiencing a mayor overload of information on the web due to the proliferation of Internet sales and the increase of marketing channels. As a result, current organizations cannot effectively manage or control the rapidly growing information utilizing traditional tools (Gretzel et al., 2006). One way to address this information overload is to use appropriate technologies for the management of all these "pieces of information" (Baumgartner,
2006), relying on semantic technologies (Davies, 2007). Research can be directed to the study of demand, supply, prices, brands, packaging, distribution etc. However, a comprehensive analysis of the various situations that may arise needs to be conducted and the main issues to be tackled need to be identified. Quota market approaches, current and potential demand, its projections, the product and its features, prices, distributing methods and manufacturing procedures should be properly treated, given that the decisions to be taken are very complicated and mistakes, if any, have disastrous results.

Currently, market research is conducted by collecting information on prices of different suppliers (or consumer enquiries) via telephone or through interviews (Wen & Wen, 2006). This process is time consuming and it is, therefore, not efficient in reacting to fluctuations in a volatile market. This research proposes a different approach: in order to automate this process, it is necessary to integrate online data collection, domain modeling, analysis technologies and information reporting related to a specific domain. The proposed research targets the field of tourism, considering that tourism is both an outstanding area in the field of electronic commerce, as well as a domain with significant challenges regarding the complexity of the information (Werthner & Klein, 1999).

Based on the assumption that the Web has become a faithful mirror of the world, in this case of the tourism industry in particular and therefore, information on market decisions can be obtained and analyzed from the web, this work is an interdisciplinary research effort that integrates web engineering and statistical methodologies applied to the subsector of hotels.

The main objective of this work is to provide a competitive intelligence tool to perform market research, fundamental to the design of tourism policies and management strategies. The scope of this research work is focused on providing indicators such as average rates by type of hotel, the occupancy rate and the estimated demand with a forecast span of 14 days ahead of a destination and its competitors. In order to do this, it is necessary to meet the following objectives:

- Definition of a complete domain model.
- Development of an extraction, transformation and loading strategy that includes a strategy for data cleansing and correction to achieve comparative analysis.
- Development of sampling strategies for Web data sources.
- Learning the empirical properties of data sources (for example, the frequency change data) and new structural features (e.g. new products). This will result in an update of the domain model by the system to adapt to market changes.
- Due to the large volume of data, it is especially relevant to design a proper data aggregation model according to the most interesting aggregation hierarchies as geographic regions, product categories, etc, taking into account a specific domain.
- Consider the development of a semantic annotation approach for product analysis and market data, for instance, the distribution of bids for a certain period of time or geographic region.
• Development and implementation of automated analysis (based on the application of statistical analysis functions) of the data collected.

As a whole, the present research core is to automate the process of analyzing market data from online Web data sources. An adaptive approach is proposed to conduct empirical analysis of Web data through feedback of information obtained.

2. Methodology

Regarding methodology, this research proposes an approach where the market research practices could be adapted to the specific online market needs of hotels. Based on the assumption that the Web has become a faithful mirror of the world, and that, therefore, information on market decisions can be obtained and analyzed from the web automatically, this task is an interdisciplinary research effort, which integrates web technologies and statistical methodologies applied to the subsector of accommodation, conducted in the following phases:

2.1. Domain model definition

This domain aims to contain data that will show the evolution of the hotel availability and prices in time. In order to do that every day a data collection will be performed. This collection will request data regarding a certain region for the following 28 days, in any case only the first 14 are publicly shown. This process will be repeated to collect data of all needed regions each time. In order to carry out this study a domain model is proposed. As mentioned it is based on availability and prices for hotels, for a specific date, obtained on a previous date, as the goal is to watch the evolution in time of the hotel availability and prices for a given date. Figure 1 shows the entities and attributes of the domain and the relationship among them.

Figure 1: Domain model diagram
2.2 Identifying information sources

In order to obtain primary information for the domain model, the Internet Distribution Systems (IDS) have been identified as the main source of information, given that the information found on those systems pertains to the public domain.

The spread of these distribution channels has opened doors for hotels that did not have the necessary infrastructure to enter the market of online sales since the platform is provided by IDSs. Consequently, nowadays almost every hotel can offer online booking. Even more, hotels that could not afford much advertising are now easily associated with a destination thanks to the search engines on Internet Distribution Systems.

The role of the IDS in the hotel market is so big that nowadays some businesses only accept reservations through them. Moreover, in situations of overbooking or rather elevated occupancy rates, on line reservations and definitively prioritized as opposed to phone bookings, which provide no assurance that such clients would eventually turn up.

2.2.1 Booking.com

With more than 135,000 hotels in the world, booking.com is arguably the largest Internet Distribution Systems. Its market share is so large that some hotels may only sell through booking.com nowadays.

The fast rate at which this distribution system grew during the first decade of the XXI century forced hotels that had not planned to sell online to join the system, even with the only purpose of keeping up with the competition.

One of the keys to booking.com’s success is that its partners are forced to offer their lowest prices through this site, allowing and rewarding their clients to report lower prices found elsewhere, punishing the hotels if such requisite is not met.

Also booking.com’s search engine offers results that by default are ordered depending on how much of the reservation commission a hotel shares with booking.com. Every hotel is requiered to share part of the income, but those willing to share more are shown first.

2.3 Knowledge generation systematic process

The knowledge generation process must be automatized so that the knowledge base is automatically updated. Figure 2 shows the systematic process that every 24 hours extracts the information from the Internet and transforms it into knowledge for decision making by the hotel sector or the administration.
The knowledge generation process can be divided into three different phases. These phases are extraction, transformation and loading and will be explained in the next paragraphs.

The extraction phase is based mainly on crawling techniques that perform searches queries on Internet Distribution System websites. The query begins by forming a special URL belonging to a certain IDS with added parameters depending on the specifics of that particular request. These searches will be for a specific place on a specific date, so these customizable parameters are the region id and the check-in date. If the IDS requires a check-out date this will be the day after the check-in date. The result of this page is a HTML page, the standard response for a web request. This response is later transformed to XHTML for reasons explained further in this section. Usually the first request will not return all possible results since the amount of viewable results per page is finite. The XHTML page can be parsed to obtain certain pieces of information, such as the total number of hotels and the maximum amount of hotels viewable per request, which leads to knowing how many additional requests need to be made to retrieve information about all the hotels on that region. Once the amount of pages is known, all of them are retrieved with HTTP requests and their HTML response is also transformed to XHTML.

The transformation phase uses screen scrapping techniques and it works by taking the XHTML phase of the previous requests and creating objects with just the required information using XPath expressions. HTML is meant to contain several data structuring and decorating tags that do not concern the scraper since only small bits of data are wanted and that’s why XPath is
used to extract some identifying information and the variables of this study, which would be an
unique hotel identification, the hotel's price and the availability for the requested date.

The reason to transform HTML pages to XHTML is that even though both are tag languages
that can describe the same, XHTML is much stricter. Whereas HTML can be dirty and
malformed and still be considered valid, XHTML, being an extension of XML, follows stricter
rules and thus it’s easier to process by standard tools because the document does not contain
unexpected elements.

The data extracted by the XPath expressions may sometimes need some cleaning or
processing because this data is the same found in the original HTML page and it's meant to be
viewed by humans, and may sometimes not be atomic enough, mixing several pieces of data in
the same tag, that need to be divided and filtered before proceeding to the next phase.

The final phase, loading, concludes by inserting the data on the objects created in the
transformation phase into the fact base designed in section 2.1. This process is repeated inside
two nested loops. Each iteration of the main loop will look for data regarding a specific region.
The secondary loop will iterate the different dates of which it wants to collect the data. The
secondary loop is conditioned by the region selected in the main loop.

2.4. Incremental Strategy

The knowledge generation capacities created from the IDS obtained data, meaning hotel
availability and rates, are not limited to the present research scope, they could easily be
extended, for instance, to a competitiveness analysis. If a specific hotel was to be chosen, its
direct competitors could be analysed, by knowing the competitors rates for a specific product, as
well as the demand curves when counting with the information of such rated for a certain time
span.

A competitive intelligence system should adapt itself nimbly and flexibly to stakeholders’
knowledge requirements. In that sense, the incremental strategy of the present approach
provides a suitable expandable and adaptable framework. Such framework emerges from a set
of determined indicators which can be increased and adapt to new indicators, such as hotel
reservation data so that more complex demand/rate elasticity simulation techniques are
conducted.

2.5. Viewpoint

The showcasing of the data obtained in the previous step will be made through a web interface,
filtering data by destination, and showing it for close-in-time dates, enhancing the rates and
hotel availability performance in a certain time span.

The development of the viewpoint is based upon the technological stack shown in Figure 3.
2.6. Assumptions

As it is mentioned in the beginning of the present methodology section, the present research is assuming that the Web has turned out to be a trustworthy reflection of reality. In this sense, the research work supposes that every hotel is linked to Internet Distribution Systems. More specifically, out of all existing Internet Distribution Systems, the research presented here focuses on one in particular, booking.com. Then, it is assumed that the hotel availability of a particular region can be extracted by dividing the hotels present in booking.com for a specific date by the total of hotels of the region.

Other aspect to consider would be that, hotels not only offer the standard and comparable individual/double/individual use double rooms, but also not so comparable products such a luxury suites using a wide range of terms to define them, which results on a complex comparison process. Therefore, the Doble Standard Room was chosen for every hotel. It should be taken into account that Spain counts with specific law, (Real Decreto 1634/1983), regulations for defining certain types of rooms per hotel category, thus the comparison is rather more reliable.
The present research work core is the automation of data market analysis process when the source is the Web online data. Hence, an adaptative focus it is proposed for the empiric analysis of web data through the feedback of the obtained information.

3. Results

The data collection process is not executed right in the moment when a system user wishes to visualize the data, rather the process is conducted once a day and it collects all the parameters needed so that all the considered queries are replied. Therefore, the data collection process implies a loop that collects data for all provinces, even though the results are shown filtered by region or city.

The entry date will also be multiple by launching different searches as part of the same loop, starting from the date when the search is launch until 28 days later, given that, even though only 14 days are shown as part of the results, these dates’ data will be compared with the previous 14 days gathered in advanced. In order to search for the hotels of a specific province, a HTTP request is launched within the booking search engine using a URL with only 5 configurable parameters:

- region: province identification of which the data wants to be obtained.
- checkin_year_month: Year and month of the entry.
- checkin_monthday: Day and month of the entry.
- checkout_year_month*: Year and month of the check out.
- checkout_monthday*: Day and month of the check out.

*Please note: the check-out date will always be the next date to the check-in date given that this allows for the rate night to be check on the daily basis.

Once the data is obtained on a hotel by hotel basis the data referred to average rate, and average availability per city an region is pre-calculated. Figure 4 shows how the user has selected a specific province, in this case Asturias (North of Spain), within the web interface of the Hotel Market Monitor, and the results are shown on the right column. The data shown refer to the hotel rates and availability in Asturias for the next 14 days. Additionally, these data are compared with the data gathered a day before, a week before, and two weeks before for the same dates.
4. Conclusions and future work

As mentioned before, changes in the value chain of tourism market require an efficient destination management to make them more competitive. The change from mass marketing to a digital playground requires enabling competitive intelligence tools. Those tools should perform data collection and analysis tasks giving the destination manager useful information for decision making.

During this work, a market monitoring tool has been developed, oriented to the accommodation subsector. This tool is based on the digital footprint given by the tourism market on the net. Thus, the market monitor is able to calculate the evolution of hotel rates, estimate load and predict room availability.

The development of this tool represents a first step for designing and shaping more efficient marketing strategies. It’s based on live market data instead of the sample information generated by statistical institutes. At the same time, performs predictive analysis allowing sector stakeholders and the public administration to conceive better strategies, as well as to improve policy making procedures.
With the development of this tool the tourism sector is one step closer to counting with a competitive intelligence system, there is future work to close the marketing circle. Future research could be focused on semantic web technologies and how to integrate them. The specific objectives of this research are summarized in three points:

- Definition of an ontological model to develop a strategy for data cleaning and correction based on the ontology. This will ensure that the system will use comparable market products and offers. In this area, there are some approaches (Walchhofer et al., 2009), that use semantic information to match different references to the same product.

- Development of a semantic annotation, a product analysis and market data, for example, the distribution of offers for a certain period of time or geographic region.

- The construction of a metadata model that captures the semantics required to allow the interpretation or analysis of market data according to application needs. The work will focus on a generic framework that can be applied to different domains.

Finally, it will be also interesting to develop a prediction tool based on the data collected by the market monitor and an event database, given that the fact that a destination hosts an event embodies diverse impacts in the tourism sector, particularly referred to service consumption (hotels, restaurants, attractions…). This database should collect events from several sources, such as, event generators, holidays...etc.

5. Bibliography


SCIP 2005 Strategic and Competitive Intelligence Professionals [URL:www.scip.org]
