Alexandros Papageorgiou

The Impact of Social Media on Travel Information Search:
The Case of Hospitality Search in Finland and Greece

Master’s Thesis in Information Systems
Supervisor: Prof. Pirkko Walden
Division of Natural Sciences and Technology
Department of Information Technologies
Åbo Akademi University

Åbo, 2010
# Table of Contents

1 INTRODUCTION .................................................................................................................. 1

2 SOCIAL MEDIA AND CENTRAL CONCEPTS......................................................................... 4

   2.1 Facts and figures about social media activity ................................................................. 4

   2.2 Interest in social media in Google Insights for Search and Alexa .............................. 5

   2.3 Definition of social media ............................................................................................. 7

       2.3.1 Social media ........................................................................................................... 8

       2.3.2 Social web .............................................................................................................. 9

       2.3.3 User-generated content (UGC) .............................................................................. 9

       2.3.4 Additional terms ..................................................................................................... 10

       2.3.5 The term and definition selected to be used in the current study ..................... 10

   2.4 Social media types ......................................................................................................... 11

       2.4.1 Blogs ....................................................................................................................... 11

       2.4.2 Wikis ....................................................................................................................... 11

       2.4.3 Micro-blogging ....................................................................................................... 12

       2.4.4 Online social networks .......................................................................................... 12

       2.4.5 Multimedia sharing sites ....................................................................................... 12

       2.4.6 Virtual worlds ....................................................................................................... 13

       2.4.7 Group-based aggregation and social bookmarking ............................................. 13
2.5 Web 2.0 ................................................................................................................. 13
2.6 Measuring social media ...................................................................................... 15
3 WEB SEARCH ........................................................................................................... 18
  3.1 How search engines work .................................................................................. 18
  3.2 How search engines rank results ...................................................................... 21
  3.3 Search engine optimization and social media .................................................. 23
  3.4 Travel search ...................................................................................................... 24
    3.4.1 Travel search process .................................................................................. 25
    3.4.2 Travel vertical (meta) search ...................................................................... 25
4 TOURISM INDUSTRY AND SOCIAL MEDIA .......................................................... 26
  4.1 From E-Travel to Travel 2.0 .............................................................................. 27
  4.2 Importance of social media for tourism domain .............................................. 28
  4.3 Credibility of social media in travel ................................................................. 30
5 STATE OF THE ART RESEARCH ......................................................................... 32
  5.1 Review of relevant literature .......................................................................... 32
  5.2 Critical literature review .................................................................................. 35
6 RESEARCH METHODOLOGY AND DESIGN ...................................................... 37
  6.1 Research design ............................................................................................... 37
  6.2 Tools and methodology used ............................................................................ 42
7 DATA ANALYSIS AND FINDINGS ..................................................................... 48
  7.1 Proportion of social media for hotel brand searches ...................................... 49
7.2 The main types of social media represented on the top three search pages...50
7.3 Top social media sites represented in the search results...............................52
7.4 Destination specific social media distribution..............................................53
7.5 Social media content on the first three search pages....................................55
7.6 Hotel brands and number of social media results........................................58
7.7 Sub-study analysis.......................................................................................60
  7.7.1 Sub-study 1.............................................................................................60
  7.7.2 Sub-study 2:............................................................................................61
  7.7.3 Sub-study 3.............................................................................................63
  7.7.4 Sub-study 4.............................................................................................64
7.8 Comparing social media with non-social media results..............................65
7.9 Data mining analysis....................................................................................67
8 CONCLUSIONS..............................................................................................71
  8.1 General conclusions....................................................................................71
  8.2 Hotels, social media sites and destinations that stood out...........................73
    8.2.1 Social media websites..........................................................................74
    8.2.2 Additional analysis on Tripadvisor......................................................75
    8.2.3 Hotels.....................................................................................................76
    8.2.4 Destinations...........................................................................................76
  8.3 Relating to previous research......................................................................77
  8.4 Recommendations for marketers...............................................................79
8.4.1 Owned media........................................................................................................... 80
8.4.2 Bought media ......................................................................................................... 81
8.4.3 Earned media ......................................................................................................... 81
8.5 Search engine evaluation ............................................................................................. 85
8.6 Assessment of research: limitations & future research ............................................. 86
8.7 Future web, social media and tourism ........................................................................ 87
9 REFERENCES .................................................................................................................. 90
10 APPENDICES ............................................................................................................... 98

APPENDIX A: Types of web 2.0 marketing initiatives hoteliers worldwide plan to implement ................................................................................................................................. 98

APPENDIX B: US natural search visits by search engine results page ......................... 99

APPENDIX C: Technology that will have the greatest impact on the way the next generation researches travel .................................................................................................................. 99

APPENDIX D: Top 10 search engines in the US ................................................................. 100

APPENDIX E: Top 10 upstream search engines visited by US internet users before visiting travel websites ................................................................................................................. 100

APPENDIX F: Metrics used to monitor and measure the impact of social media according to US marketers ........................................................................................................................ 101

APPENDIX G: Google Docs interface where students entered and coded the collected data ........................................................................................................................................ 102

APPENDIX E: Web video created to provide step by step instructions for the students ........................................................................................................................................ 103
List of Tables

Table 7.3.1 Frequency, frequency (%) and cumulative (%) of SM sites ...................... 52
Table 7.6.1 Number of social media results for the top 24 hotels of the sample .......... 59
Table 7.8.1 Number of observations and % of the top commercial and SM websites. 66
Table 7.9.1 Association rules linking city, hotel, website and search page variables... 69
Table 8.3.1 Comparison of top SM sites between Xiang and Gretzel’s and present study ................................................................................................................................. 78
List of Figures

Figure 2.2.1 Search activity for the term "social media", 2004-2009............................... 6
Figure 2.2.2: The daily reach of the most popular social media applications ................. 7
Figure 2.5.1 Relationship between social media and Web 2.0 (Gould, 2008) ............ 14
Figure 3.1.1 From Web crawling to data encoding (Smith et al., 2005) .................... 19
Figure 6.2.1 A typical Google search result including title, snippet and URL .......... 43
Figure 7.5.2 Percentage of SM distributed by type for search pages 1, 2, 3 ............ 57
Figure 7.1.1 Social media vs. non-social media (%)............................................. 49
Figure 7.2.1 Social media results distribution by type (%)........................................ 50
Figure 7.3.1 Cumulative % of social media representation and unique domains ...... 53
Figure 7.4.1 SM and non-SM frequency & SM/non-SM ratio by location ............... 54
Figure 7.4.2 VCS by destination ............................................................................. 55
Figure 7.4.3 CRS by destination ............................................................................. 55
Figure 7.5.1 Percentage of SM, non-SM and SM/non-SM ratio across search pages... 56
Figure 7.7.1 SM generated by hotel generic queries, distribution by type ............. 63
Figure 7.7.2 No of SM and non-SM results & SM/non-SM ratio, Athens and Helsinki
combined .................................................................................................................. 64
Figure 7.9.1 Data mining output (XLMiner) linking social media type, city and search
page .............................................................................................................................. 67
Figure 8.4.2 Social media strategy & tactics tree.................................................... 84
List of Abbreviations

AJAX: Asynchronous JavaScript and XML
API: Application Programming Interface
CGC: Consumer Generated Content
CRS: Consumer Review Sites
CSS: Cascading Style Sheets
HTML: Hyper Text Markup Language
IAB: Interactive Advertising Bureau
IMC: Integrated Marketing Strategy
MMOG: Massively Multiplayer Online Game
MSS: Media Sharing Sites
OECD: Organization for Economic Cooperation and Development
RDF: Resource Description Framework
RSS: Really Simple Syndication
SEM: Search Engine Marketing
SEO: Search Engine Optimization
SM: Social Media
SNS: Social Networking Sites
VCS: Virtual Community Sites
WTM: World Travel Market
1 INTRODUCTION

If one looks at the top websites worldwide based on web traffic patterns provided by Alexa (www.alexa.com/topsites), it will not be hard to notice that in essence this represents a mix of popular social media sites together with widely used search engines. The present study will investigate the role and significance of these two rapidly growing types of online properties with a view to the travel industry.

The tourism field with its established strong presence on the Internet is no exception to the struggle for visibility among the Internet users and particularly online travelers. At the same time we are witnessing the rise and rapid growth in popularity of social media websites, with user-generated content represented in social media increasing in volume exponentially. For travel in particular this content has high impact on travel decision making, as it is regarded by travelers as a highly trustworthy source of information (Litvin et al., 2008).

As social media is increasing both in volume and significance, this trend emerges as an opportunity for players in the travel industry to achieve valuable additional exposure, which can translate into incremental business. At the same time, this could also prove a potential challenge for travel businesses, considering that social media content, while highly relevant and influential, cannot be easily controlled or manipulated by the travel industry players.

On the other hand, there are web search engines, which are often described as the gateway for Internet users to reach online content. Travel related online content constitutes no exception to that. According to a recent study published by Google and OTX, nearly two thirds of the online travelers use search engines for travel planning (Sullivan, 2009). Moreover an estimated 70% of all online bookings funnel through a search engine (Donogue, 2006).

Search engines deploy sophisticated algorithms to rank and present search results, based on a high number of relevant factors. Increasing popularity and significance of user-generated content represented in social media is not being ignored by search engines. Social media content is frequently updated and extremely well hyperlinked, thus creating hard-to-miss signals for the search engines. Thanks to its search engine friendliness, this content tends to be prominently featured in the returned results.
It has been suggested, however, that search engine users rarely go past the third result page when searching for content, while the majority will actually stop at the first page (iCrossing, 2010). This underlines the importance for any piece of online content, including social media, to appear with a good search ranking, or otherwise risk being overlooked by users. This struggle for content visibility through search engines has given rise to new areas of research and professional practice such as Search Engine Optimization (SEO) and Search Engine Marketing (SEM), both of which have found wide adoption within the online travel industry.

Based on the above observations it appears that both web search and social media are emerging trends, which have an impact on the distribution and accessibility of online travel information to consumers. A critical question that arises based on the framework described above, relates to the extent and the precise ways in which travel related social media content is represented through the major search engines.

Understanding how prominently (or not) social media content is displayed alongside other content on the search results pages can be particularly useful for marketers in the travel industry in terms of helping to evaluate:

- What is the proportion of social media among all the other search results retrieved by a search engine?
- Which are the social media websites and their corresponding general types that are represented most prominently on the first pages of the search results?
- How does the social media proportion in search results vary from one page to the other?

These are the three fundamental research questions that this study will investigate. The objective of this research, by providing answers to these questions, is to enable travel marketers to make informed decisions on potential courses of action they can develop, in order to support their search engine marketing as well as their search engine and social media optimization strategy.

The three generic type questions mentioned above will be supplemented by additional ones which are relevant to the specific domains that this study will investigate and which will be analyzed in the following chapters. Namely, these domains are hospitality, in terms of the industry in focus and Finland and Greece in terms of geography. Once these domains have been examined, all the research questions will be analytically discussed (chapter 6).

The basic definitions and concepts, including also a relevant classification of social media and the associated technologies, will be discussed in chapter two. This chapter
also provides evidence pertaining to the increasing popularity and importance of social media.

Chapter three focuses on search. The chapter examines the ways in which search engines rank results and discusses the relationship between social media and search.

In chapter four, the evolution from Tourism to eTourism and Tourism 2.0 is discussed and examples that highlight the importance of social media for tourism are provided.

Chapter five provides an overview of relevant literature that focuses on travel search, travel social media as well as the intersection of these fields and critically reviews it.

This leads to chapter six that discusses the rationale of the present study and explains the research methodology deployed. In this chapter the research questions are analyzed and a number of complementary sub-studies discussed.

Chapter seven contains the data analysis part, which covers in detail the most meaningful co-relations identified between the collected data and drives to the discussion section, where the key findings are presented.

The final chapter analyzes further some of the key conclusions and provides recommendations for travel marketers as well as ideas for future research. The study concludes with a view on the future web, social media and tourism.
2 SOCIAL MEDIA AND CENTRAL CONCEPTS

2006 can be regarded as the break-through year for social media. At that point, popular early applications like Wikipedia (www.wikipedia.com) and MySpace (www.myspace.com) had already gathered a significant number of users. Furthermore, social media applications like You Tube (www.youtube.com) and Facebook (www.facebook.com) were gaining popularity at high speed.

Acknowledgment of social media’s emergence to the realm of mainstream, is the Time magazine’s article about the Person of the Year 2006, that figured prominently on the December 2006 magazine’s cover. The Person of the Year 2006 was “You” meaning the millions of anonymous contributors of user-generated content on various social media websites. Grossman’s article (2006) highlights the proliferation of social media and their high impact on society.

“But look at 2006 through a different lens and you’ll see another story, one that isn’t about conflict or great men. It’s a story about community and collaboration on a scale never seen before. It’s about the cosmic compendium of knowledge Wikipedia and the million-channel people’s network YouTube and the online metropolis MySpace. It’s about the many wrestling power from the few and helping one another for nothing and how that will not only change the world, but also change the way the world changes. The tool that makes this possible is the World Wide Web...The new Web is a very different thing. It’s a tool for bringing together the small contributions of millions of people and making them matter”

From a European perspective on the other hand, EU acknowledges that social media have caused significant repercussions leading to disruptive effects on industry citizens, identity, social inclusion, education, health and public governance (Lindmark, 2009). For EU commissioner Reding (2008), the new phenomenon is all about connecting minds and creativity never seen before.

2.1 Facts and figures about social media activity

According to the official You Tube blog, the site receives over 1 billion of video views on a daily basis as of 2009 (Hurley, 2009). Users who upload videos include US president Barack Obama, Britain’s Queen Elisabeth and the Pope. Facebook has accelerated its growth too, with user base reaching 400 million worldwide and increasing with an estimated rate of 800.000 users per day as of 2010. Additionally, more than 1.5 million pieces of content (web links, news stories, blog posts, notes, photos, etc.) are shared on Facebook on a daily basis (Facebook, 2009). Wikipedia contains over 13 million articles and studies show that it is more accurate than Encyclopedia Britannica, with 500 times more traffic compared to the latter (Press,
Twenty-five per cent of search results for the world’s top 20 largest brands are links to user-generated content (Aarons et al., 2009). Moreover there are over 130 million blogs in existence as of 2009 (Singer, 2009). These are just a few examples that can illustrate the high impact of social media on today’s world, while social media activity continues to grow on exponential rates across geographies and demographics.

From a business perspective, social media platforms are transforming the way businesses are producing, marketing and selling products. The one-to-many communication of the past is now developing into a many-to-many open channel that empowers consumers, allowing them to benefit from the social media’s collective intelligence. This is marking a major socio-economic shift where businesses need to learn to behave, communicate and interact with their customers in brand new ways (Qualman, 2009).

### 2.2 Interest in social media in Google Insights for Search and Alexa

A more graphic illustration of the social media growth is represented in figure 2.2.1 and Figure 2.2.2 through the use of two online tools, Google Insights for Search (www.google.com/insights/search) and Alexa top site (www.alexa.com) application.
Google Insights for Search is a tool that analyzes a portion of worldwide Google web searches across all Google domains, in order to compute the number of searches for the terms entered. The numbers on the graph are normalized and presented on a scale from 0-100, reflecting the number of searches for the term “social media”, relative to the total number of Google searches over time. Figure 2.2.1 illustrates the popularity of the term social media on Google for the period 2004-2009, underscoring the steep surge of interest in social media, as expressed through search activity, especially from 2007 and onwards.

Figure 2.2.1 Search activity for the term “social media”, 2004-2009
Figure 2.2.2: The daily reach of the most popular social media applications

Figure 2.2.2 represents some of the most popular types of social media applications: Wikipedia, the web’s open encyclopedia, MySpace and Facebook from the social networking space, Flickr, the photo sharing site and YouTube for video sharing. Most of the websites experience high growth rates and no signs of slowing down (with the exception of MySpace). Considering that these are the most popular among social media, their growth has direct implications for the social media industry as a whole.

2.3 Definition of social media

In the recent times along with social media, a number of other related terms have emerged too. Examples include the social web, user-generated content, participative web and more, which are often used to describe the same concept. It frequently occurs that certain type of media, organizations or professional groups prefer to adopt one of these terms over the others and attempt to use it in order to describe the whole phenomenon (EU for instance calls it social computing, OECD refers to it as participative web, several technology groups tend to prefer Web 2.0 etc.). Many of these terms are often used interchangeably. Nevertheless, their meaning and essence are not necessarily identical. Thus, there is an essential need to define accurately and separate these terms, before any further analysis follows.
2.3.1 Social media

For the term social media itself there is a plethora of definitions available. According to Blackshaw (2006)

“Social media can be generally understood as Internet-based applications that carry consumer-generated content which encompasses media impressions created by consumers, typically informed by relevant experience, and archived or shared online for easy access by other impressionable consumers”

Safko and Brake (2009, p.6) focus on the conversational element of the social media.

“Social media are linked to activities, practices and behaviors among communities of people who gather online to share information, knowledge and opinions using conversational media, such as web based applications that make it possible to create and easily transmit content in the form of words, pictures, videos and audio”

Wikipedia (Wikipedia contributors, 2009) on the other hand defines social media as:

“Media designed to be disseminated through social interaction, created using highly accessible and scalable publishing techniques. Social media supports the human need for social interaction, using Internet- and web-based technologies to transform broadcast media monologues (one to many) into social media dialogues (many to many). It supports the democratization of knowledge and information, transforming people from content consumers into content producers”

The Wikipedia definition, therefore, marks the new attributes (highly accessible, scalable, from monologue to dialogue) of social media against the so-called traditional or industrial media which are rather one-way. Wikipedia also highlights a number of elements that make the two distinct, including:

- Accessibility, with social media being practically available to anyone with an Internet connection, without need for any high investment or possession of relevant assets.

- Usability, with social media being easy to use in relation to the specialized knowledge required to operate traditional media.

- Recency, with traditional media in need of significantly more time to broadcast the news and with social media having the ability to be critically more up to date.

Social media can take many alternative forms, including blogs, wikis, media sharing and social networking sites.

To enable more accurate identification of social media sites Lietsala and Sirkkunen (2008, p.24) have gone beyond the mere definition of social media, to suggest the
following five common characteristics that they share. This can serve as a compass in terms of determining if a site is of social media type or not.

1. There is space to share the content
2. Participants in this space create, share, or evaluate the content themselves
3. It is based on social interactions
4. All content has a URL to link it to external networks
5. All actively participating member of the site have their own profile page to link to other people, to the content, to the platform itself and to the possible applications

Some other terms such as new media or digital media are often used in a similar sense like social media. Even though there is some overlap among these terms, the social component is highlighted in social media, while this is not the case with the other two terms.

2.3.2 Social web
According to Wikipedia contributors (2009), the social web is used to describe how people socialize or interact with each other on the World Wide Web, based on a variety of shared interests, which could be either people or activity focused. Social web as a term is often used in combination with others within the context of the web evolution chain. These other terms include: semantic web, real-time web and web of things.

2.3.3 User-generated content (UGC)
UCG also known as consumer-generated media (CGM) or user-created content (UCC) refers to various kinds of media content, publicly available, which is produced by end-users.

OECD (2007) defines it as content:

-Made publicly available over the Internet

-Reflecting a certain amount of creative effort

-Created outside of professional routines and practices

Based on the above definition, it can be argued that user generated content, as a fundamental element, is like the lifeblood of the social media system and this is what effectively distinguishes it from other media types.
A variation of user generated content is the user-driven content, which stresses that not all content is created by the users (and not all of the users anyway), instead content is frequently just copied and distributed from other sources to another location (Lietsala and Sirkkunen, 2008). Fernando (2007) suggests that UGC is the opposite to traditional forms of media and marketing, since content is generated by the consumers rather than the marketers. Ahuja et al. (2007) point out that UGC sites equate to electronic word of mouth marketing, whereby a user who has an opinion about a product or service can share views and experiences with peers.

2.3.4 Additional terms
Along with the terms presented above, there are several others in existence that relate closely to the social media concept. Even though it is outside the scope of this study to cover extensively all existing terms, three of them i.e. social computing, participative web and Web 2.0 have been identified and will be discussed briefly as they can offer interesting additional perspectives to the social media phenomenon.

-Social computing (also referred to as social software)
It principally refers to the code, software and technologies utilized for social media implementations. Lindmark (2008) defines it as:

“The set of open-web based and user-friendly applications that enable users to network, share data, collaborate and co-produce content”

-Participative web
According to OECD (2007) the term represents an Internet increasingly influenced by intelligent web services, based on new technologies empowering the user to be an increasing contributor to developing rating, collaborating and distributing Internet content and developing and customizing Internet applications.

-Web 2.0
One relevant, frequently used term used in the same context is Web 2.0. Due to its high significance as well as relative complexity as a concept, the term will be analyzed in more detail in a following section. For the purpose of this section, Web 2.0 refers to the technological, entrepreneurial and ideological aspects of the new phase of development of the web that has contributed to the high levels of impact and success of social media (O’Reilly, 2005). The relationship between social media and Web 2.0 will be covered in the respective section to follow.

2.3.5 The term and definition selected to be used in the current study
Even though each of the expressions discussed captures some dimensions of the phenomenon and misses others, in this report (unless there is strong reason to do otherwise) the term social media is used more prominently as it is broader, more neutral and more widely used in media, industry and literature.
The definition of social media that will be used for the scope of this study is the following:

“Social media are emerging sources of online information that are created, initiated, circulated, and used by consumers with the intent of educating each other about products, brands, services and issues. In contrast, to content provided by marketers and suppliers, social media are produced by consumers to be shared among themselves” (Blackshaw, 2006)

This definition, instead of underscoring the contrast between social and industrial media or any socio-psychological factors, it rather highlights the differentiation between the media and content created by the marketers against the ones generated by users for users. This definition was deemed the fittest since the study investigates the prominence of social media in relation to content that has been developed by industry professionals such as marketers, PR agencies or any other paid employees.

2.4 Social media types

There are numerous classifications of social media, often containing several subcategories for each main type. The importance of classification for this study is high, since one of the desired outcomes is to define the most prominent social media types in the search results. Understanding of social media classification will also be essential, with reference to the analysis part where relationships between social media types, search pages and destinations will be examined. In this section the most prominent social media types are described.

2.4.1 Blogs
Websites where entries (posts) are displayed in reverse chronological order. They are date stamped and content expanded to include links, texts and images as well as rich media formats. Blogs are often utilized as online personal diaries and are open to the Internet public to place comments and tags on them. It is a practical way for anyone to publish content online including individuals, groups or organizations etc., without the need to have programming or technical skills. The large number of users engaged in blogging has given rise to the term blogosphere, which refers to the ecosystem created by bloggers’ posts, permalinks, trackbacks, blog rolls etc. In addition, blogging platforms allow for rapid syndication of content to interested audiences using opt-in protocols such as Really Simple Syndication (RSS) (IAB, 2009).

2.4.2 Wikis
A set of web pages that allow users with access to easily create, edit and link content, collectively, using a standard browser. There is a history and a rollback function that enable users to revert to previous versions. Wikis enable and promote collaborative work and collective intelligence. In a wiki projects like Wikipedia there are distinct roles such as reader editor, contributor, administration, policy, subject matter expert,
software developer, patroller, content maintainer, system operator and more (Lytras et al., 2009). Wikis can be internal or open to anyone, or with gradual levels of access and rights. Simplicity and flexibility are among the top advantages of wikis, which started to gain traction with the success of Wikipedia. The latter is currently the top fifth most visited site on the web containing over 13 million articles in 200 languages (Ala-Mutka et al., 2009).

2.4.3 Micro-blogging
Micro-blogging is a relatively new approach to blogging, as a compact form of it, which has rapidly become popular in the social media community. The concept of micro-blogging is commonly represented by Twitter, which is an application allowing users to communicate status updates (tweets) using a maximum of 140 characters and then sharing those with their followers (the network of their friends and contacts). Thanks to its real-time nature, Twitter, has been referred to as the heart pulse of the web. It has been gaining rapid user base growth since its 2006 launch, while as of November 2009 Twitter had over 27 million subscribed users (TIG Global, 2009). The micro-blogging application was among the key technologies on the peak of Gartner Hype cycle in 2009.

2.4.4 Online social networks
These are sites that provide an online virtual community for people with similar interests or social groups, to network and share online digital assets such as images, video, content tags, lists of friends, etc. Social networks are usually based on the users creating profiles with biographical information and their users can connect with each other forming friendships based on interests and likes. Moreover, they can have access to their friends’ profile (public or semi-public), see when they are online, chat and message them.

Social networking sites’ user base has experienced an exponential growth and spending time on social networks is considered among the most popular activities on the Internet for users, not only the young ones, but also reaching increasingly older demographics. There is a multitude of social networks available on the web, from mainstream ones, such as Facebook, MySpace etc. to niche ones addressing a wide range of interests and demographics. Closely related to social networks are the online or virtual communities. Even though classification types can vary according to source, for the purpose of this study social networking sites encompass virtual communities as a subdivision.

2.4.5 Multimedia sharing sites
Multimedia sharing services facilitate the storage and sharing of audio and video content produced by the users themselves or by others. This functionality is provided
through websites and applications that facilitate the upload and display of various media. The majority of multimedia sharing sites provide multiple views such as thumbnails, slideshows for images and ability to classify and add annotations and comments. Multimedia sharing sites have certain similarities with online social networks, but their main characteristic is that these sites focus on the sharing of content. Top photo sharing site, Flickr reports one million images uploaded daily to the site (Ala-Mutka et al., 2009), while video sharing You Tube is turning into a disruptive factor in media industry as it reports over one billion daily views as of 2009 (Hurley, 2009).

2.4.6 Virtual worlds
A virtual world is a simulated three dimensional environment which allows users to interact using graphical models of themselves, the so-called avatars. This type also includes the category of massively multi player online games (MMOG) like World of Warcraft. In such games there is a purpose; it is usually about winning. However, in other virtual worlds such as Second Life or Habbo Hotel, it is mainly about the experience of entertainment and interaction. Moreover, some virtual environments such as Second Life provide users with a scripting language and integrated development tools which empower users to create objects on which they are allowed to keep the associated intellectual property rights. As a result of that, virtual economies have emerged, such as the one in Second Life, where more than 1 billion USD per year has been contracted between players since the virtual world’s beginning. As of March, 2009 over 1.4 million residents were active users (OECD, 2007 and Lindmark, 2009).

2.4.7 Group-based aggregation and social bookmarking
This is a type of social media the objective of which is to build on views and knowledge from across the web community. It primarily consists of group-based collection of links or group based rating of such links. Users collecting the links are also the ones to tag them, rate them and comment on the related pages. An example of this model is Digg (www.digg.com), users of which store fresh links to the site to be evaluated by other users. Delicious (www.delicious.com) on the other hand provides a method for users to store, organize and search bookmarks of web pages on the Internet, with their bookmark files usually publicly available or shared with specific individuals or groups (OECD, 2007).

2.5 Web 2.0

Web 2.0 as a concept is tightly linked to the evolution and success of social media. Indeed to such an extent that often the two terms are used interchangeably. Despite how interrelated the two concepts are, they do have a distinct meaning. In the previous section, social media as a concept was defined and a framework of social
media-related technologies was presented. Web 2.0 as opposed to social media does not refer to any specific type of online media, rather it relates to the technology, economics and attitudes that surround this new phase of development for the web (see Figure 2.5.1). The latter is seen as more socially oriented compared to the previous, so called “version 1” of the web or “Web 1.0”.

![Figure 2.5.1 Relationship between social media and Web 2.0 (Gould, 2008)](image)

Several definitions of Web 2.0 are available. Dippelreighter et al. (2008) consider Web 2.0 as a technological and sociological paradigm shift, that has taken place on the Internet, by creating new standards and principles. Kaplan and Haenlein (2010) consider Web 2.0 the ideological and technological foundation based on which social media are built and thanks to which user generated content can effectively circulate.
The coining of the term as well as the development of key concepts around it, are associated with entrepreneur and publisher Tim O’Reilly. Web 2.0 as term first emerged in 2004 just a few years after the “dot com” collapse and while the need for a new model appeared to be a necessity for the Internet industry to regain its lost momentum. Web 2.0 as a concept was triggered through a series of brainstorming sessions taking place in Silicon Valley at the time, to describe the trend that seemed to emerge which could lead to the next step in the evolution of the Web. Here’s how O’Reilly (2007 p.17) defined Web 2.0 on a widely cited white paper he posted on his corporate blog.

“Web 2.0 is the network as platform, spanning all connected devices; Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an “architecture of participation” and going beyond the page metaphor of Web 1.0 to deliver rich user experiences”

2.6 Measuring social media

With the increasing popularity and impact of social media, one of the most discussed topics both in literature and online forums relates to the need of developing methods to measure the effectiveness of social media engagement from the industry’s end.

Social media metrics can be used in a number of ways: from campaign planning and valuation, market research and trend spotting, to increasing customer loyalty by leveraging communities, reducing churn and marketing costs, even develop and launch products in the market at greater speed, lower cost, often co-designed together with the consumers (Jaokar et al., 2009). Social media offer both opportunities and challenges in the establishment of metrics, as they add a level of qualitative information to the quantitative data which are traditionally made available through web analytics (Stuart, 2009).

Concerns, nonetheless, are expressed with regard to how efficient is the current way social media activity is measured either from marketers individually or from research companies or from IAB (Internet Advertising Bureau), given the release of reports (Macnamara, 2009) that emphasize the lack in quality on online media metrics.

It is the lack of measurement quality that makes it challenging to answer questions such as: are 1,000 blog page views worth more or less than 100 re-tweets? Or Are 100 You Tube subscribers better than 200 LinkedIn group members? How much attention should be paid to the quality of interactions against quantity?
To partially address this Macnamara (2009) suggests key metrics that can be productively measured:

1. Unique visitors and duration of visits (avoiding duplications and short time stays)
2. Number and quality of incoming links to the social media site, to measure site credibility and influence
3. Downloads of documents and video views, user ratings and comments, as a measure of engagement of site visitors and site
4. Return visits, to measure visitor frequency an stickiness
5. Next clicks or where visitors go next, can provide useful information too, in terms of determining intentions

Social media can of course be measured quantitatively by page views or video views, application downloads etc., however since the goal of participating in social media is to make a conversation, there might be more meaningful way to measure the engagement with them. Beyond the quantitative part, Macnamara (2009) also stresses the importance of conducting content analysis in order to understand its availability and the message it communicates using qualitative methods, drawing on techniques like text analysis, semiotics and narrative analysis.

This analysis will result in the creation of a framework based on which effective metrics and measurement method can be defined. In any case, as regularly iterated in relevant literature (Joogar, 2009) before attempting to measure social media impact and success and defining relevant metrics, it is wise to think first about what the actual goal or the business or organization is and what would the desired outcome of a given social media campaign be.

Considering blog as an example, it would be meaningful along with the received by a particular blog post to also consider the reaction to it, as expressed by the amount of comments posted and the new links created pointing to that new blog entry. On a more general level it will be meaningful to track how many times a blog is listed on a blog roll (list of recommended blogs by third bloggers) and how the number and location of RSS feeds subscribers is trending. This is mainly because that data do not merely reflect a number of users who randomly e.g. driven by search engines, stumbled upon the blog, but rather users that value the blog and want to be continuously updated about the authors views and experiences.
Similarly for social networking sites, even though it is possible for certain groups or individuals to have a great amount of friends, contacts or followers, this obviously doesn’t mean that everyone will be actively participating in the dialogue. It is therefore recommended to equally take into account the number of comments, feedbacks, tags etc. that will be involved and by which people (obviously when influential people join the discussion this could offer added value)

The same ideas hold true for measuring the success of micro-blogging activity. What is most important is eventually the users who are actively engaged with one's micro-content (such as tweets) and who could spread the message over by sharing it with others (for example by re-tweeting). It is therefore important to refrain from taking the number of followers at face value (Stuart, 2009), but instead try to consider how the followers interact with one’s stream.

To make this analysis more complete it is important to examine factors such as who are members of the community participating in the discussion and how they can be classified in separate categories based on engagement, influence factors and overall position in the network. This takes into account their propensity to start using new products, the propensity to attract new member in a network etc., and it couples it with dynamic demographic and behavioral profile data (Jaokar et al., 2009).

Beyond the above mentioned metrics used in the examples, there is a number of additional ways which are tailored to marketing initiatives taken on social media sites, through developing social media widgets or applications. Such metrics as defined by IAB include interaction rate (proportion of users who interact with an ad or application), time spent on section, microsite or community, the medium’s conversation reach and conversational density, content freshness and relevance where the message will appear and more. The characteristic, however, of this type of metrics is that they refer to paid inclusion on social media sites, rather than participation in the actual conversation.

To conclude, in this chapter the main concepts pertaining to social media were discussed; from definition, to classification, importance and measurement of social media. As this is the central theme of the study, it was deemed essential to provide this theoretical background that will enable a better understanding of the methodology and research design which are discussed in the following chapters. Particularly the classification of social media is of prime importance in relation to the way this study was carried out and also with reference to its defined objectives. The next chapters discuss the other two pivotal concepts in this study, eTourism and search engines. When examining the common ground between those two and social media, several of the concepts analyzed in chapter two will emerge in the discussion again.
3  WEB SEARCH

With the abundance of information, navigation within the Internet environment has become a challenging task for many users. Often overwhelmed by the enormous amount of information available, Internet users can become frustrated at the difficulty of locating relevant information. Search engines have naturally emerged as an essential tool to help searchers locate the most relevant information rapidly and effectively and facilitate the information overload. A recent study ranks web search as the top activity performed by users on the Internet (European Interactive Association, 2007).

Search engines are the practical applications of information retrieval techniques to large scale text collections, which enable users to seek for and find relevant information sources online based on keyword searches they conduct (Croft, 2010).

Commercial search engines such as Google, Yahoo and Bing, index a great number of webpages on the Internet and thus serve as the “Hubble” telescope with which people learn about the entire virtual “galaxy” (Castells 2001; Spink et al. 2002 cited in Xiang and Gretzel, 2009). With the explosion of content created on the web, these search engines can now crawl terabytes of data, as they have indexes that contain billions of pages and have the capacity to return results in fractions of a second.

Modern search engines use complex algorithms and hundreds of ranking criteria to produce their results. Among the data sources is the feedback loop generated by the frequency of search terms, the number of user clicks on search results and users’ personal search and browsing history. Even though this study focuses on web search engines, it is worth to mention that search engines can be found in many other applications including desktop, vertical, enterprise search and more.

3.1 How search engines work

There is a high number of commercial search engines available online, with different characteristics each, alternative technologies deployed and often distinct target group of users. Nevertheless, they all perform some common basic tasks (Strickland, 2009).

Search engines fundamentally consist of two parts. The offline part gathers web pages and builds an internal representation of them called an index. The online part serves users requests by finding matching documents, ordering and ranking them with the goal of presenting the most relevant documents on top (Henzinger, 2007). In
the next section the operating principles and basic architecture of a search engine are discussed in more detail.

Figure 3.1.1 From Web crawling to data encoding (Smith et al., 2005)

The two main goals for a search engine are effectiveness i.e. to be able and retrieve the most relevant documents for each given query and efficiency, practically speed. The architecture of a search engine aims to maximize these two qualities.

The evaluation method regarding whether these goals are met, is characterized by four critical metrics. Precision is the proportion of retrieved documents which are relevant. Recall is the proportion of relevant documents that are retrieved as compared to the total number of documents retrieved. Another important measure is coverage, which examines how much of the total existing information can be retrieved. Recency or freshness, measures the age of the stored information.

There are three standard tasks that search engines have to complete in order to provide their services and be able to process queries as fast as possible, as illustrated in figure 3.1.1.
The first step is to crawl the web. Before a search engine is able to return results corresponding to a query, it has to locate the relevant documents and pages on the web. With the billions of web pages in existence, this work can only be performed by special software robots, called spiders or web-crawlers. The process is referred to as web-crawling.

In order to build and maintain useful lists of words, the spiders need to look at as many places as possible. In order to achieve this, there must be a starting page which is usually contains some lists of servers including a mix of popular pages (called seeds). From that point and onwards the spider starts to travel through the web going from one link to another, scanning the content that comes its way. Regardless of the original file format, it converts content to HTML for quick and easy retrieval at a later stage.

All documents are stored in a repository. As a natural result of the way spiders travel the web i.e. following links, it is expected that the most popular or well interlinked properties of the web will be prioritized. Obviously in the extreme case that a page has no links pointing at it, the spider will never discover it -unless the owner of the page manually submits it to the search engine for inclusion. Consequently a significant amount of information is hidden in the so-called “deep web”, the part of the web that is not indexed. There are pages practically out of reach for the crawlers, such as password protected, unlinked or protected with robots.txt extension (Holmberg, 2009).

Once crawling is completed (or close to completed, given that the web by its nature is ever changing) the search engine must store information in ways that render this information useful and possible to retrieve rapidly and accurately.

There are two components based on which the information gathered with the spider can become available to searchers: (a) The type of information stored with the data and (b) the method by which the information is indexed.

With regard to the way information is stored, the easiest way to achieve this is to link the words located in the text to specific web pages. This, however, is not sufficient to provide users with a relevant answer to their query. This is because there are many other elements that in effect determine the special weight that each word bears, or its relative importance in a document. These elements include the position of the keyword in the text, the number of occurrences, the size, whether the text is bold, underlined or not, whether it carries a link to another page in the document or external, meta-tags included in the html code of the page etc. Every search engine has its proprietary method to assign weight to these values which is one of the reasons why different search engines produce diversified results against a given query.
No matter how a search engine deals with this task however, the next step is to encode the data to save storage space. Once the information is stored in a compact way, the next step is to create an index, which in many ways is similar to a book index (it can reveal which words contained in the index match any particular query terms). The indexing process includes: text transformation, that transforms documents into parts or index terms and index creation which takes the output of the text transformation component and creates the indexes and data structures which enable fast searching. This process is essential so that the information can be accessed, retrieved and returned to the user rapidly. In fact most modern search engines are in position to return results in fraction of a second.

The combination of effective storage and indexing accelerates the process of returning information to the user, even when the queries are complicated, involving a long string of terms.

The last part of the process is the building of a search, which involves a user typing in the search box one or more terms or even submitting a more complex query involving the Boolean operators (AND, OR, NOT etc.) to refine the search. A last step is the ranking which after receiving the user’s query, it processes it and returns a list of ranked documents based on scores and retrieval methods.

These principles are common to all search engines. Commercial search engines, however, strive to differentiate themselves and provide improved results as well as more intuitive user interfaces, ergonomic presentation and organization of returned results, based among other factors, on personal preferences and search history (Hearst, 2009).

Search engines also apply machine learning and artificial intelligence techniques to improve performance automatically, by learning relationships and associations within the stored data. Search engines re-crawl popular frequently changing pages at a rate that could be proportional roughly to the frequency with which pages are updated. Those are referred to as “fresh crawls” (Franklin, 2009).

### 3.2 How search engines rank results

There is much speculation around how top search engines rank their results. In fact there is a whole growing industry, SEO (Search Engine Optimization) that deals with ways to optimizing websites for favorable rankings on search engines. On the other hand, the search engines themselves remain secretive about their ranking algorithms for fear of having hackers game the system (Lornali, 2009).
One of the most well established search engine optimization blogs, SEOmoz (2008) has conducted a broad survey asking some of the most eminent professionals of the industry to give their views on the Google search ranking criteria and their respective importance. The results unsurprisingly have not been validated by Google. They have, however, been favorably received by the international SEO community.

Here's a summary of the top 10 factors that influence the Google search engine results

1. Keyword use in the title tag - placing the target keyword on the title tag of the web page's HTML header
2. Anchor text of inbound link - the exact text that accompanies the link from other external pages to the target page
3. Global link popularity of site – the overall authority that the target site appears to have as measured by the sites across the web, i.e. click quantity and quality.
4. Age of site, which refers to the date indexable content was first discovered by the search engine. Google places more value on pages with an established history.
5. Link popularity within the site’s internal structure, refers to the number and importance of internal links pointing to the target page.
6. Topical relevance of inbound links to site, the subject specific relationship between the sites/pages linking to the target page and target keyword.
7. Link popularity of site in topical community, refers to the link weight/authority of the target website amongst its topical peers in the online world
8. Keyword use in the body text, using the targeted search term in the visible html text of the page.
9. Global link popularity of linking site, referring to the importance of receiving links from popular sites
10. Topical relationship of linking page, while all links help, the ones from topically related sites help more.

It is claimed that Google has over 200 factors influencing the results (Eustace, 2006). Even though Google does not validate the results of surveys like the above, it does recognize that central importance in the ranking of results is played by PageRank, its trademarked algorithm which assigns each page a relevancy score. Due to its high importance as one of the main reasons that have made Google the most successful search engine, PageRank is analyzed further below.
The formula and definition of Page Rank is defined by Brin and Page (1998) as follows:

“We assume page A has pages T1...Tn point to it (i.e., are citations). Parameter d is a damping factor which can be set between 0 and 1. We usually set d to 0.85. Also C(A) is defined as the number of links going out of page A. The PageRank of a page A is given as follows:

\[ PR(A) = (1-d) + \frac{d}{C(T1)} \times PR(T1) + \ldots + \frac{d}{C(Tn)} \times PR(Tn) \]

(The PageRanks form a probability distribution over web pages, so the sum of all web pages Page Ranks will be 1)

Practically, the PageRank of a page reflects the "importance" of that page, giving higher ratings to pages with more pages linking to them, particularly when those third pages are of high quality. In other words PageRank reflects how much Google "trusts" a page.

3.3 Search engine optimization and social media

When a user searches online for a brand, it is not uncommon that among the first results there will be a Facebook or Twitter page or a You Tube video and so on. Given the increasing popularity of both search engines and social media applications, it is interesting to investigate the relationship between the two.

Social media can drive search engine results by providing strong signals to them. The more comments, tags, blog posts, bookmarks etc. a site has and the more recent those are, the higher the site will rank when consumers search for specific terms on search engines. For example in 2007 Google launched its Universal Search application which besides text, provides results of various other content types, including image, video, news, blogs etc., all integrated into one page of relevant information.

Research is also carried out aiming to combine data from social networks and documents reference networks like PageRank, to create a dual layer of trust enhanced (or socially enhanced) search result rankings, taking into account the so-called collective intelligence or “wisdom of crowds”. As a result social search provides more validated and verified answers than that of an automated search engine algorithm. According to Grehan (2010):

“Signals from end users who previously couldn't vote for content via links from web pages are now able to vote for content with their clicks bookmarks and tags and ratings. Those are strong signals to Search engines and best of all they do not rely on the elitism of one website owner linking to another or the often mediocre crawl of a dumb bot”
It is not uncommon that pages with high PageRank score are ranked below social media site results (Baker, 2007). This is evidence of the possibility that thousands of video views, coupled with multiple comments on YouTube videos or the high rating of a Flicker photo album can strongly influence the search ranking order. Baker (2007) claims that Google could easily determine the quality comments on news sites or blogs which could make a difference on how search results are ordered.

Another example comes from Yahoo local, where the ranking in the lists returned when a user for instance searches for hotels in a particular area, directly depends on the number of reviews and the corresponding rating that these qualified hotels have received (Slawski, 2009).

There are patents filed by major search engines that provide supportive evidence that this trend will continue, as social media is becoming a major source of ranking signals for the search engines in the future. A patent application from Yahoo titled “Method and apparatus for rating user generated content in search results” explores an approach for indexing UGC and including it in search results (Slawski, 2009).

The patent application introduces three concepts that will be helpful in ranking UGC, with the aim of showing up in web search results when useful. The main concepts are:

- Document goodness, possibly looking at user ratings among other things.
- Author rank, a measure of the expertise of the author in the given area
- Location rank, i.e. where the content is located (message board, forum, group etc.).

At the same time, there is speculation that Google and Bing have been also exploring how to rank UGC and could be examining similar signals (Slawski, 2009).

Catalytic can also be the role of bookmarking services such as Delicious and StumbleUpon, Digg and Reddit. In many ways, when a user decides to bookmark a site, this action can be considered as a positive vote towards that page. Moreover, the keyword and description tagging process associated with social bookmarking enables the content to be more accurately categorized and therefore returned as a relevant result for corresponding searches (Baker, 2007)

3.4 Travel search

Given the significance of search engines in the general use of the Internet (e.g., Spink and Jansen, 2004) and particularly travel information, search is emerging as a significant part of the purchase decision process for travelers worldwide (Buhalis and Law, 2008). In fact search engines to a great extent define the practical
boundaries of the online tourism domain and dictate the way it is represented (Henzinger, 2007). According to a study by the Travel Industry Association of America (TIA, 2005) nearly two thirds of the online travelers use search engines for travel planning. Moreover Hitwise (2010) reports the critical role of web search in terms of generating upstream traffic to tourism websites, with Google and Yahoo alone accounting for over 30% of the total traffic.

3.4.1 Travel search process
Typically, the interaction between a traveler and a search application starts when a traveler with an initial information need and a particular mental model types in a keyword into a search engine box. The traveler’s mental model consists of a set of factors including the traveler’s perception of how the search system works, specialized domain knowledge as well as the search task itself (Gretzel et al., 2006 cited in Xiang and Gretzel, 2009). According to the query input the search application will return relevant results matching the nature of the keyword and will present the results in a predefined way. The traveler can then make a judgment on whether the results are satisfying and if this is not the case, the query will have to be refined and re-entered to begin a new round of search.

3.4.2 Travel vertical (meta) search
The importance of search engines for travel has already been highlighted. In relation to travel search the three popular commercial search engines, Google, Yahoo and Bing, are not always the obvious choice.

Travel meta search is a new generation of travel vertical search engines, also referred to as travelbots, that combine, with a single click from one site, data from a multitude of travel sources to match the requirements of the user, based on the given budget range and destination. The convenience that these sites provide for the travel consumer is that they help eliminate the need to visit multiple websites separately, before making the comparison needed to make a decision, which can be time consuming and frustrating. Meta sites (such as Kayak, Sidestep and Mobilissimo) use alliances with major hotel chains, airlines, car rental companies etc., in order to have access to their inventory data, even though they typically do not process any transactions, but instead refer or link customers directly with the suppliers. (Christodoulidou, 2007) Unlike the three dominant search engines, these search applications do not typically include social media content but rather commercial results.
Tourism is considered a top industry worldwide. It generates over 10% of global GDP and it is a top industry provider of employment (Buhalis, 2003). Its impact has a multiplication effect across many other industries. Several countries and regions of the world rely on tourism as a vital source of income.

Tourism is defined by the World Tourism Organization (WTO cited in Buhalis, 2003) as the activity of traveling and staying in places outside one’s usual environment for more than twenty-four hours and not more than one consecutive year for leisure, business and other purposes not related with the exercise of an activity remunerated from within the place visited. The present study has a special focus on the impact of Internet and technology on Tourism. The field of tourism that incorporates the impact of ICTs in the tourism product, especially with focus on the digitalization of processes and value chains, is often referred to as eTourism (Buhalis, 2003).

From an eTourism point of view it is worth to note that technological progress and tourism have been advancing hand in hand for years. Increasingly, ICTs play a critical role for the competitiveness of tourism organizations and destinations as well as for the entire industry as a whole (UNWTO, 2001). The establishment of the Computer Reservation Systems in the 1970s and Global Distribution Systems in the late 1980s, followed by the development of the Internet in the late 1990s, has brought about disruptive changes to tourism. The latter is now ranked as the foremost industry in terms of volume of online transactions (eBusiness W@tch 2006, cited in Buhalis and Law, 2008).

Moreover the Internet has changed tourism consumer behavior dramatically. Developments in carrying capacity, speed of networks and search engines have influenced the number of travelers around the world who use technologies for planning and experiencing their travels (Buhalis and Law, 2008). From information search, to destination consumption and post experience engagement, ICTs offer a range of tools to facilitate and improve the process for consumers. The latter search the web for travel-related information, make online air-ticket bookings, online room reservations and other online purchases themselves, instead of relying on travel agencies to undertake this task on their behalf. A Eurostat survey confirms that 71% of Internet users use the Internet for these activities (eMarketer, 2007).

Prospective travelers have direct access to a much greater wealth of information provided by tourism organizations, private enterprises and increasingly by other fellow consumers. It is evident that the future of eTourism will be focused on
consumer centric technologies that will support organizations to interact with their customers dynamically, becoming incredibly powerful, sophisticated and increasingly capable to determine elements of their tourism products (Buhalis and Law, 2008).

Technology has also impacted the way Travel businesses market their products and services. Due to the popularity of Internet applications, most tourism organizations including hotels, airlines, and travel agencies have embraced Internet technologies as part of their marketing and communication strategies. This is also because Internet marketing can provide cost benefits and improved reach to consumers. As a subdivision of online marketing, social media marketing has emerged as a new, direct and effective way of reaching consumers by actually participating in conversations with them.

According to Mangold and Faulds (2009) social media should be included in the promotion mix as a hybrid element when developing integrated marketing communication (IMC) strategies. Considering the nature of the tourism product, which is impossible for consumers to experience or evaluate prior to purchase, both Internet marketing in general as well as social media marketing in particular, have gained substantial traction among Travel marketers as persuasive means of reaching and interacting with consumers. At the same time social media and the other advances in web technology in general are already challenging established marketing practices (Xiang and Gretzel, 2009).

4.1 From E-Travel to Travel 2.0

Sigala (2007) points out that the content and information generated by users of social media technologies are having a tremendous impact not only on the profile, expectations and decision making behavior of Internet users, but also on the e-business models that businesses need to develop or adapt to. When travel industry and social media technology unite, the outcome can be referred to as travel 2.0. According to the WTM Goal Trends report the Travel 2.0 concept was first coined in 2003, referring to the importance of the second wave of travel information websites for developing online travel communities and forums. These sites were uniquely driven by travellers interested in the opinions of fellow travellers instead of professional travel companies or guidebooks.

With Travel 2.0 travellers are turning towards direct forms of interaction, in search of more authentic experiences in travel. The evolution of Travel 2.0 concept and travel networking is also behind the rapid development of innovative forms of travel such as hospitality tourism and home exchanges. Travel 2.0 is also behind the success of travel networking websites where travellers exchange their experiences and in virtual context meet the people who live where they plan to travel.
What actually differentiates Travel 2.0 from the first, booking-oriented wave (e.g. Expedia, Kayak, AA.com) is that it is fully interactive and expands via user-generated content. A Travel 2.0 site is structured to allow users to easily contribute words and images, reviews and travelogues. In the words of Philip Wolf, President of PhoCusWright (2007):

“Travel 2.0, our industry’s collective fulfillment of Web 2.0, embodies how companies can differentiate themselves in a vast, dynamic space. New travel researching and planning approaches are empowering consumers in unprecedented ways. As was the case with Travel 1.0, this phenomenon is not entirely about technology; rather, 2.0 is focused on solving big problems for all types of travel customers by exploiting the latest advancements”

4.2 Importance of social media for tourism domain

As made evident from the previous sections, developments in social media space are having a disruptive impact across multiple industries, and travel –the industry of focus in this study- is no exception. In fact, travel by its very nature, as an activity centered on the notion of experience, it has been in the epicenter of conversation in a multitude of social media.

Flourishing travel blogs are a testimony to this. Travel is an activity that inspires people and motivates them to write about their experience in visited destinations. Thanks to this, travel blogs have gained high popularity. Blog posts can themselves travel across the web providing tips and recommendations that can reach and influence numerous other travelers through the so-called e-word of mouth process. At the same time blogs are not a privilege of the travelers only. Companies and organizations are not ignoring the merits of blogging. As an evidence of this numerous hotels, ranging from small local ones to multinational chains, now have their own blog where they communicate on an informal tone with their past and prospective customers. In the same space they can also solicit feedback, gain insights into traveler profiles, conduct free on line marketing research and enhance loyalty (Sigala, 2007).

Micro-blogging on the other hand, even though by nature it is not adequate for giving full details about travel experience, it can play an important role for travel too. Typically it is the smart travel businesses, the early adopters who use the micro-medium, to inform their followers about their limited last minute offers, which can then be re-tweeted repeatedly. Twitter can be thought of as opt-in marketing mechanism, where users select to receive updates on things they like, attracted by the real time and spontaneous nature of Twitter. More often than not this content is delivered to one of the travelers’ favorite gadgets, their mobile phone (Wheeler, 2009).
Social networking sites and virtual communities are gaining momentum for travel too. General purpose sites such as Facebook contain numerous conversations happening every day, in groups dedicated to travel. Moreover there are several travel Facebook applications like Tripadvisor’s “Where I’ve been” counting multimillion user base, that keep the users engaged. Travel companies take advantage of the link, photo sharing and conversational aspect of Facebook, by establishing fun pages and trying to attract new friends. On these pages information can be provided about current events, happenings etc. Arguably the highest impact on travel comes from travel dedicated communities such as Tripadvisor and Virtual Tourist where travelers can meet and build relationships with other users, regardless of geographic or other restriction (Chung and Buhalis, 2009). Those are communities where consumers exchange information, opinions and recommendations about destinations as well as tourism products and services.

Dedicated Travel networks can foster richer travel experience and the gaining of a local perspective. Couchsurfing (www.couchsurfing.com) and Hospitality Club (www.hospitalityclub.org) have jumped in popularity, reaching hundreds of thousands of members in a short period of time. Couchsurfing in particular has a global reach by being present in over 200 countries. These online communities strive to bring together travelers and locals willing to offer free hospitality and socializing opportunities. Likewise, Wayn (www.wayn.com), the largest travel network with over 13 millions of users, is built on the idea of visualizing on a world map, the places where friends are currently located. While travelling, it allows travelers to share info and opinions about trips and destinations and meet people living locally (WTM, 2008).

Based on a social component too, wiki-type applications like Travel Wiki (www.wikitravel.org) have been attracting high numbers of visitors not only to utilize the site as their travel guide, but also to contribute by making additions and improvements. Travel Wiki represents the efforts of Internet users to collaboratively create and continuously update an online global travel guide (Sigala, 2007).

Another interesting type of application usually associated with Google maps is travel mash-ups where users can see for example the exact hiking trail that other travelers have followed along with comments, pictures and videos from the most worthy spots. Applications such as RoomAtlas (www.roomatlas.com) provide an effective tool for travelers to visualize their hotel options on a map, based on their preferences, and combine this with user-generated reviews retrieved from Expedia.

Multimedia sharing applications are on the rise for travel, too. Travelers make use of the possibility to upload their pictures and share them with the world, while other users can comment, tag and rate them. It is similar for travel related videos. Travelers
not uncommonly, besides making videos of the monuments and natural beauties, may also produce video recordings of the hotel areas and hotel rooms where they have resided. Once this content becomes public every potential customer can watch these videos in order to ensure the information received from the hotel website or brochure is accurate enough. Video sharing in extreme cases can be used for the consumers either to praise to criticize a service.

From a more commercial standpoint, video sharing can be a new channel of communication for the travel suppliers themselves, who can set up a branded You Tube channel for instance, where they can share videos associated with the service they offer. In fact on You Tube there is a separate video category which is exclusively for travel related videos. Some of these videos can become viral virtually overnight, resulting in a large amount of positive publicity for usually a low production cost.

Other tools for travelers enable them to organize trips with others from their network online. With tools such as Realtravel (www.realtravel.com) and Travelpost (www.travelpost.com) there is possibility to create an itinerary and share it with friends. They can edit it, propose changes, synchronize with calendars etc., until consensus is achieved to enable a group booking.

The use of virtual worlds like Second Life as an alternative channel for market research, new product testing and promotion of new advertising campaigns can be of great potential (Sigala, 2007). For example Starwood has built a three-dimensional virtual hotel which is targeting young consumers in Second Life. In this way its marketers were able to promote Starwood’s new hotel brand and obtain consumer feedback about the future design of the hotel (Hanyoung, 2009). Many other tourism related enterprises have created their representative offices and virtual headquarters in Second Life, such as destination marketing organizations, museums, hotels and event organizers (Sigala, 2007).

4.3 Credibility of social media in travel

Arguably one of the main reasons why social media content has such a high impact on travel is its perceived trustworthiness by the travelers, often taking the form of a new, digital word of mouth.

It is clear that there is a high perceived credibility of opinions expressed in user generated content through social media, compared with traditional tourism information sources such as travel agents or accommodation operators, which are operated by a business with a vested interest. On the other hand, online reviews and word of mouth recommendations are growing in importance as information sources, because of the perceived independence of the message source. This is because the
provider of the information is not generally aiming to make a financial gain from sharing their experiences and views with others (Litvin et al., 2008)

In a recently published survey from eMarketer.com (2009), it was found that in the UK considerably more consumers trusted sites with amateur reviews than professionally written guides or travel agencies. Gretzel et al. (2007) reported in a study conducted with users of Tripadvisor (www.tripadvisor.com), that the most popular online activities carried out during the trip planning phase were looking at other tourists’ comments and materials as well as reading travel related blogs.

This trend has important monetary implications for the travel industry. A recent study conducted by Compete Inc. (2009) found that over twenty per cent of consumers rely on UGC when planning their trips, while UGC already influences about US $10 billion a year in online travel bookings.
5 STATE OF THE ART RESEARCH

The present study has its research focus on the intersection of social media, online information search and the travel industry—all three of these fields thoroughly discussed in previous sections, including also their joint impact and relatedness. Travel information search and social media have been thoroughly analyzed in several published studies. Nevertheless, the two fields are typically researched separately, especially given the fact that social media is quite a recent phenomenon and therefore its study is far from comprehensive.

On the other hand, extensive and multi-dimensional research has been carried out on the travel information search field, which has traditionally captured the attention of travel researchers. Only a single case of research has been identified to cover both areas and it will be discussed more thoroughly towards the end of this section.

5.1 Review of relevant literature

Xiang and Gretzel (2009) worked on the area of the representation of the online tourism domain in search engines. The researchers attempted to assess the visibility of destination-related information on the Internet, examining also the visibility of the various tourism-industry sectors relevant to the destinations. An interesting finding was that despite the huge amount of information indexed, only a minor fraction of the travel domain can actually be accessed by Internet users. There is in effect a low number of websites dominating the search results. Even though the study does not discriminate between social and non-social media sites, it does stress the importance of social media and Web 2.0 technologies in facilitating and enhancing the travel search experience. The authors called for the development of new, advanced search technologies to improve the organization and overall representation of the tourism domain.

Pan et al. (2007) conducted research on the topic of accommodation search query formulation. The research offers insights as to how search engine travel queries are performed. That was possible with the processing of a significant amount of actual search server log data. Findings reveal that most travelers search by looking for specific hotels in conjunction with the destination city. The study also confirms earlier research findings that travelers’ web searches are highly functional in nature and rarely hedonical.
Jansen et al. (2009) in their research have similarly looked at the process of searching for travel related information on the Internet. They utilized a pool of over 1.5 million search queries by over 500 thousand users, analyzed both quantitatively and qualitatively. Some of their findings suggest that the proportion of travel related searches in comparison with the total is quite sizable, at the level of 6.5%. The geographical impact appears strong as geographical information accounted for 50% of the queries, whereas general travel information accounted for less than 10%.

Individual terms analysis revealed a high number of searches for travel specific websites like Orbitz, Travelocity and Mapquest, while term pairs most frequently occurring were location-related such as city, state and specific location (representing roughly 60% of searches). With regard to repeated frequency, the terms were quite diversified with even the most frequently occurring ones representing a low percentage of overall usage (0.2% of all terms usage). Those were clustered around location and general information topics. This level of diversity was claimed to show a level of uncertainty in the information retrieval process, something which is also supported by previous studies (Toms et al., 2003, Urbany et al., 1989 cited in Jansen, Ciamacca et al. 2008).

Finally the researchers observed a dichotomy among queries of very specific nature, such as specific hotels, car hire agencies and travel websites, while at the same time there was no shortage of travel queries of a generic nature, like general travel information, information about countries etc. This according to the writer implies that there is a lack of technology linkage between the two extremes, and searchers might have difficulty to express their travel information needs.

Several studies have focused on the role of social media and user generated content in the travel domain, such as blogs and travel review sites or travel virtual communities (Buhalis 2008, Dippelreiter et al., 2008, Carson, 2007). These research show how user generated content contributes in creating and sharing new experiences among travelers. Conclusions also include the high credibility and significance in the decision making processes of this type of content for the travel community. Of particular interest is the research conducted by Dippelreiter et al. (2008). The evaluation benchmarks a selected set of popular travel community websites based on a number of core Web 2.0 and social media criteria as well as a number of other technology and community evaluation frameworks, adapted for use in the domain of online travel communities. Their analysis is then based on the three phases of a trip, pre-trip, on-site and after-trip. The results provide a multivariate and multidimensional evaluation map that includes some of the prime online travel properties (many of which are also prominently featured in the data produced by the present study).
Despite all the relevant research efforts discussed above, the extent to which social media constitute the online tourism domain was not touched upon until recently. Neither was there any empirical data to describe the relation between the two; social media and travel domain.

The only research that examines this area, combining both search and social media, to the author’s knowledge, has been conducted by Xiang and Gretzel (2009). That was part of their research efforts in studying and analyzing the ways in which the tourism domain is represented on the Internet, especially through the filters of search engines. In one of their studies (2009) they examined the role of social media for the comprehensive understanding of the online tourism domain. They examined the likelihood of an online traveler coming across social media content during the web search process. In this way they contributed to a better understanding of the market intelligence that social media can add to the marketing mix.

Xiang and Gretel (2009) focused on US locations and attempted to provide answers to the above questions by choosing a targeted set of keywords that could sufficiently represent the tourism domain for the purposes of the study. Namely the ten selected keywords were: accommodation, hotel activities, attractions, park, events, tourism, restaurant, shopping and nightlife. The selection of those keywords was based upon a number of past studies and was intended to reflect the generic terms or broad categories that represent tourism (Pan et al., 2007 cited in Xiang and Gretzel, 2009). They defined search terms based on the combination of locations and tourism domain keywords. The study demonstrated that the social media websites are ubiquitous in online travel information search since they occur everywhere, on various search results pages and for alternative tourist destinations, no matter what search keywords a traveler uses. The principal finding was that approximately 11% of the search results across the first 10 pages represented social media, which was considered to represent quite a substantial part of the online tourism domain.

The most prominent social media domain names included: tripadvisor.com, virtualltourist.com, igougo.com, mytravelguide.com, yelp.com and meetup.com, which on aggregate accounted for over 30% of the total unique domain names. As it is evident from the most prominent domain names, the strongest type of social media represented was virtual communities, followed by consumer review sites and blogs. Video and audio media sharing sites were not well represented among the most prominently displayed websites. As for the spread of the social media results among the search pages, the study discovered that with low fluctuation there was approximately one social media site present on every page including 10 ten search results; in other words a very homogenous distribution.
The authors of the study also examined the relationship between specific search queries and social media representation. The proportion of social media content against other types of content, with very few exceptions (smaller cities), remains similar across the range of geographic locations. On the other hand, across the alternative tourism domain keywords the picture seemed differentiated. The study pointed out that some keywords can trigger larger number of social media results (such as nightlife, with over 20% social media representation) compared to others (like attractions with just over 5%) indicating the size and structure of the sub-domains represented by those keywords might be substantially different from one another. It seems that certain keywords are more likely to generate certain types of social media within a trip planning context. It is interesting to note that the terms including the keyword “hotel” were among the ones with the lowest representation rate of social media (6-7%, which is the second lowest percentage after attractions).

As far as hotels are concerned, the study reported a trend for social media of the consumer review and virtual community type to occur the most frequently. On the other hand, keywords such as events, nightlife and park were mainly associated with social networking, video and photo sharing websites. The final conclusion drawn from the study was that marketers can no longer ignore the role of social media in distributing travel related information. Two suggested ways to leverage this would be search engine and social media optimization, in order to improve their rankings, advertise or provide content on those sites and finally integrate social media and Web 2.0 technologies on their own site.

5.2 Critical literature review

The combined knowledge of all the above studies offers a rich background for the present study. There are, however, some limitations in the past studies that will be discussed here. First and foremost past research has focused on either social media or travel information search. Most of the social media-related studies focused on the socio-psychological aspects of social media. They were mainly based on data collected through self-reported questionnaires or controlled experimental settings (for example by asking subjects to conduct a trip planning task online) and thus the degree of objectivity is considered quite limited (Gretzel & Yoo 2008 cited in Xiang and Gretzel 2009).

On the other hand, there is research about travel information search. Even though the research is quite extensive, the main focus is on the interaction between the online traveler and the “tourism industry”, therefore the impact of social media was hardly touched upon. As mentioned already, to the author’s knowledge, the only piece of research that meaningfully covers both areas comes from Xiang and Gretzel (2009).
However, given that the research space is quite wide, there are several limitations as well as open areas in the field which can be fit for further research.

In particular, Xiang and Gretzel (2009) who aimed to cover the whole travel domain, followed quite a generic approach as regards the type of keywords used for the studies. As will be discussed in the next section, however, a marketer would be equally interested -if not more- to see the impact of social media on some more specific areas, such as the brand they represent, which in practice would be a more realistic scenario. Therefore the brand-related component despite being a significant aspect of the travel product buying process, is clearly missing from the previous study.

Moreover, for the same reason, marketers would be likely to desire more insight into how influential social media are within their own specific travel sector, for example accommodation, flights, car rental sector etc. With respect to the keywords deployed, those were all preselected based on previous studies, with the intention of reflecting broad categories that represent tourism. As a result of this pre-setting, the keywords were not chosen freely and naturally by the users, as would have happened in a real-life situation.

Another limitation, acknowledged also by the authors themselves, is the fact that only one search engine (Google) was exclusively used for the study, which surely leaves space for more pluralism. Moreover, the past study only focused on the USA market and the only language used for the queries was English. To the author's knowledge, there is no other study to apply this research in European destinations and in languages other than English.
For the purpose of this study a quantitative research approach has been employed. That was chosen because the objective of quantitative research is to develop a hypothesis pertaining to “natural type” phenomena-in this case the activity of users searching for travel information on the Internet. In this study the observations collected about the impact of social media in search results could lead to some general conclusions regarding the online travel information search domain. The main goal is to provide specific facts which business decision makers can use to make accurate predictions about relationships between market factors and behavior as well as to gain meaningful insights into these relationships (Walden lecture notes, 2009).

The quantitative approach deployed for this study is based on registering, coding, evaluating and then statistically analyzing a large set of search engine data. The core part for the analysis aims to draw conclusions based on a large sample of data. A number of lower-scale complementary sub-studies will also take place with the aim of contributing to further exploration and comparative analysis.

6.1 Research design

Following from the conceptual framework, the defined goal in the present study is to gain an understanding of the impact of social media on travel-related search engine results, as seen from a travel marketing perspective. The general subject area (i.e. social media representation through search engines) matches that of Xiang and Gretzel's research (2009), which was discussed in the literature review section. That research serves as a guiding compass, as many of the concepts which are used in the present study are built on the foundations of the Xiang and Gretzel's work. The topic, however, will be approached from a new dimension with a distinct focus, which will be explained below, along with the research rationale. There are several questions that will be addressed with respect to how the research design is developed and implemented.

The research design essentially simulates the travelers’ information search process by using a set of travel-related keywords relevant to a number of city destinations, in two European countries, to query a search engine, by simulating user behavior in a trip planning scenario. Content analysis will be used to understand the data in order to provide insights to questions regarding: the proportion of social media among other type of search results, the most prominently featured social media site, their respective types and more. Results will be presented in detail at a later stage. The research design is split into a number of sequential steps which are discussed below.

First it will be essential to choose the specific travel terms to represent the travel domain in this research. Obviously there are infinite travel terms that users can use
to query a search engine, starting from very generic such as holidays in X destination or attractions in Y destination (this is usually how online travelers start their pre-trip research), or something more specific such as flight from point A to B, car hire in Destination X etc. Typically towards the final stage of the trip planning process and when they already have a rough idea of what they are interested in booking, travelers search for something more brand-specific, as relevant research reveals (ComScore, 2007).

The ideal scenario would be to combine a sample of query terms across the whole travel research and booking cycle spectrum, in order to achieve a good overview of the travel domain. Nevertheless, due to the limitations of this study and the fact that prior research has been carried out focusing on the first stage of the travel buying cycle (i.e. the generic travel terms), the present study will instead focus on the brand-related aspect of the research area. The emphasis on brand-related terms has been selected not just to differentiate this study from previous ones, but also in order to examine the more specific challenges that travel industry marketers are faced with. It is quite important from a marketing perspective to have a solid idea of what type of content is displayed on search engines, when potential customers are searching for one’s specific brand. There are several reasons why this is the case.

First, it can be meaningful to examine the opposite scenario, i.e. when a user searches for a generic keyword such accommodation in destination X. Then based on Xiang’s and Gretzel’s analysis there is on average 11% probability for the user to meet social media content displayed in the search results. This 11% of results would typically lead to a third-party web page, such as tripadvisor.com or virtualtourist.com, which present reviews of a high number of destination-specific hotels. From a hotel marketer’s perspective the chances that a user will (a) choose a social media web site out of all the available search results and (b) by navigating on that website the user will actually find and read the specific review for that hotel, are rather slim.

On the other hand, when a user searches for the specific hotel brand name, e.g. Hilton Helsinki, then if a significant number of the results are social media related, the chances that this content will have a direct impact on the hotel’s business can grow substantially. This suggests that hotel marketers would be more inclined to monitor for web search results (including social media results), which are specific to their brand and location rather than for keywords of more generic type. Additionally ComScore analysis (ComScore, 2007) suggests that even though travel consumers tend to use generic terms at the start, as they move through the purchase channel, their searches tend to become increasingly brand-specific.

This conclusion is also supported by Pan et al. (2007). Their study indicated, through a sample of travel related query data from Excite search engine, that the specific hotel
name was the most frequently used accommodation term utilized in the sample (rather than the name of the city only or the city name in combination with the generic term “hotels” or “accommodation” etc.)

The next step is to decide what would the actual brands, for inclusion in the study, be and which travel industry sector would be selected. This is in order to make the study more focused. Among airlines, cars hire agencies, hotels, tour operators, restaurants etc. the industry of choice for this study is hospitality. Accommodation is typically the first query type made by a traveler, once a destination has been chosen and often the largest single trip expenditure (Pan and Fesenmaier, 2006). Additionally, the hotel industry is, together with air travel, the highest in turnover within the tourism industry domain.

Hotels are ubiquitous and the vast majority of them have their own website, where in most cases the option of direct booking is offered. Moreover, a hotel reservation is often a more brand sensitive choice compared to the booking of a flight or holiday vehicle, which is considered by travelers as more price sensitive rather than brand sensitive. Therefore it is expected that hotel brand marketers will potentially have more reasons to closely monitor the web about what is being said and commented on their brand, especially when this opinion is exposed among the top ranked search engine results. By choosing hotel brands it will be easier to insert a geographical component, by examining specific hotel brands across a number of alternative destinations.

The next step is to determine the target destinations. The previous study in the area focused on USA destinations. For this study the region of focus will be Europe. Besides avoiding duplication, this will also serve as a stimulus for a comparative analysis between the US and European results. Twelve city destinations from two European countries are selected, one from the South and one from the North of Europe. The Southern country is Greece, which is traditionally a travel-dependent economy and one of the most visited travel destinations worldwide. The Northern country chosen is Finland, which has a special element of exoticism to offer as a top Nordic destination.

Despite the popularity of both countries as travel destinations, the respective touristic products are quite diversified. Greece is promoted as a summer destination and Finland as a winter one. Moreover the two countries have several contrasts in terms of how their economy and business sector function. In particular, when it comes to Internet readiness Finland is ranked among the top countries both in Europe and the world, whereas Greece is usually at the bottom in the developed countries group. This aspect should make the comparison of the results based on the current selection of the two countries more interesting.
From the two countries a total of 12 city destinations have been chosen. These correspond to the top five cities of Finland and top seven cities of Greece, ranked based on the population criterion. The two additional cities of Greece come as a result of the fact that the country has a relatively high number of destinations of high touristic demand. The respective cities are: Helsinki, Vantaa, Espoo, Tampere and Turku for Finland and Athens, Thessaloniki, Patras, Larissa, Rhodes and Pireaus for Greece. In all destinations there is an abundance of hotels and accommodations of various types, quality etc.

These 12 city destinations will be matched with ten hotel brands each, to create relevant query strings to be tested and analyzed in a web search. The hotel brand selection originally was planned to be based on a set constituted of the top global hotel brands. However, it would be unlikely that all of these hotel brands have presence in each and every one of the selected destinations. Instead, another method was chosen to be employed.

In every destination the top 10 hotels featured upon a Google maps search following the formula “city name + hotels”, will be the ones to be included in the test. This not only makes the choice more practical and realistic, but also gives the possibility of having a varied mix of hotels included in the study, ranging from the top world chains to small city hotels, reflecting the variety of results that a Google Maps search typically returns. This also makes the scenario more realistic given that Google Maps is ranked as the number one travel application online and one of the most common activities performed on it is hotel search (eMarketer, 2010). The study will therefore be based on combinations of destinations and hotel brands or names. For example when City = athens and Hotel Brand= parthenon hotel, the respective query term will be “athens parthenon hotel”.

The results generated by the search engine will be registered and documented. A search engine typically indexes a high number of results. However, according to the information retrieval literature (e.g. Spink et al., 2002), most search engine users (>85%) do not go past the third page to view search results. In fact, it has been suggested that searchers rarely go beyond the first results page (Pass et al.. cited in Bing et al., 2008). Based on those findings and in order to simulate how most of the real users examine and evaluate the search results, the results of the first three search pages will be registered and analyzed in this study.

The analysis will answer six research questions. Some of them (the general questions, i.e. Q1, Q2, Q3, Q5) were already revealed briefly in the introduction chapter and in this section they are further analyzed. Moreover, some additional questions (Q4, Q6), are listed and discussed in the paragraph below, with focus on the specific industry and geography domains, which were examined earlier.
• Q1: What is the proportion of social media related results for hotel name and brand searches?

This will be a straightforward way to evaluate right away the visibility and therefore relevant importance of social media within a travel search setting. Given the broad spectrum of websites from which the search engines return their results it will be interesting to see how social media are positioned.

• Q2: Which are the main types of social media that are represented in the top three pages?

This is to provide additional insights by “digging” deeper into the returned social media results. For hotel marketers is not enough just to know if a search on a hotel brand will return many social media websites, but also which are the dominant types, on which obviously more emphasis should be placed.

• Q3: Which are the top social media sites represented in the search results?

The question examines the key social media properties within the context of a hospitality search. Knowing with precision the social media websites that are most likely to bring visibility to a hotel brand is a highly prioritized deliverable for this study.

• Q4: Do destinations (cities and countries) differ from each other in terms of social media coverage? i.e. are larger destinations more likely to generate more social media content

This question places the focus on location, both on country and on city level, trying to examine deviations in the way social media cover the specific destinations included in the study.

• Q5: How does the social media content found on search results vary from one page to the other across the first 3 pages?

Obviously appearing on the first pages of search results is of critical importance and the role of this question is to examine whether moving from one search page to the next plays a significant role regarding the observed average proportion of social media. This could also offer insights with respect to the general representation of social media within the travel domain.

• Q6: Which are the hotel names and brands that trigger the highest number of social media results?
This question has the role of discovering unique cases of hotels that are attracting the largest amount of social media attention. It can also assist in identifying particular characteristics that these hotels have, for instance if they are associated with an international or local brand.

Based on the findings of this study, a number of recommendations will be presented in a later section that could enable travel (hotel in particular) marketers to form strategies that will enhance the positive visibility of their respective websites on the Internet.

Along with the study, four sub-studies will be carried out in parallel, but on a lower scale, compared to the basic one. These additional low scale studies will be deployed in order to understand the social media representation of the travel domain through additional dimensions.

Sub-study 1: Ten hotel brands will be tested in one of the original ten destinations using number two search engine, Yahoo, as well as Microsoft's search engine, Bing. This will enable comparisons between the results generated from three major search engines.

Sub-study 2: Ten hotels will be tested in the Finnish and Greek search domains for corresponding local destinations. The search engine used will be configured in a way that will return results written only in Finnish or Greek. This way it will be possible to better understand how strong the social media presence is when searching in non-English languages.

Sub-study 3: A more generic type keyword, i.e. “hotel” will be used for one of the destinations. This will enable comparison of the results between the brand-related keywords, such as “hilton helsinki”, against generic terms such as “hotel helsinki”.

Sub-study 4: For a set of keywords the experiment will go past the first three results pages, to include results up to the tenth page. This could provide insight as to whether additional results pages include additional types of social media or if it is a mere reproduction of roughly the same social media sites across all pages.

Due to limitations in the scope of this study, for these sub-studies only a limited amount of data is documented, registered and analyzed, basically corresponding to only one of the destinations involved in the study.

6.2 Tools and methodology used

The search engine that is used in the study is Google. Google is widely considered as the undisputed champion in the online desktop web search industry, by providing
highly sophisticated search technology. Moreover based on relevant research, Google’s worldwide market share is over 67% (Maher, 2009). With relevance to Europe, Google’s market share figure becomes even higher. Google has an index of around 35 billion web pages and serves 250 million queries per day (Brooks 2004, Burns 2007 cited in Xiang and Gretzel, 2009).

In particular, in relation to the travel sector, Google is among the top ten websites generating the most upstream traffic to travel websites (Hopkins, 2008). eMarketer (2009) also confirms Google as a search engine that brings the most traffic to websites in the tourism industry comparing to its competitors. Google therefore will be queried on ten hotel specific keywords for the range of 12 preselected destinations which equals to 120 hotel-related queries (without considering here the sub-studies). It is expected that for a query of this type a search engine like Google will return a large amount of results, possibly amounting to several thousands. Managing this amount of data would not only be unpractical but also unrealistic, given that search engine users rarely go past the third result page when searching for content, while the majority will actually stop at the first page (iProspect, 2006).

The first three pages of the search results will be considered, with every search results page containing a standard of 10 results. As an exception, one of the sub-studies will consider search results pages one through ten, to enable additional comparison. Each search result consists of a title line, a few lines of summary (or snippet as often called) and a URL, as illustrated in Figure 6.2.1. For the purpose of this study, part of the search results that is used for analysis is the URL, leading on the page of a site which could either be of social media type or not.

Figure 6.2.1 A typical Google search result including title, snippet and URL

For the purpose of this study the travel information search model adopted is based on the conceptual framework developed by Xiang and Gretzel, which focuses on the interactions between an online traveler, a search engine and the tourism domain (2008). The framework includes three core parts.

1. The online travel searcher who is driven by his/her information needs and is interested in performing a specific search task. The searcher and his/her behavior can be influenced by a number of variables including his/her personal characteristics, the characteristics of the trip and the nature of the
information need, as expressed through the specific query formulation and potential reformulation.

2. The search engine. Its indexed content represents the online tourism domain from the searchers standpoint. A number of elements can influence the output of a search engine (and consequently the traveler’s perception and decision making) including the technology deployed to rank results, the users interface, the way results are presented and the depth of the index.

3. The online tourism domain which is a space that combines and remixes content coming from a variety of sources including travel industry players (such as travel agencies), state controlled institutions (such as destination marketing organizations), mass media (newspapers, TV programs etc.) and content originating from other travelers.

A group of 36 university students was deployed to review the results, assess them i.e. whether they are of social media nature or not, and if social media, code them based on the specific type of social media they belong to. The students thus first categorized the search results in social and non-social media websites. This was feasible by accessing the corresponding URLs of each result provided by the search engine. The second stage was about assigning the social media websites to respective categories.

There is no formal typology to use as a guide in classifying social media. One can think of several types and sub-types which can vary according to industry and research domain. Many of the most frequently used types were discussed in chapter one. However, the categorization developed by Xiang and Gretzel (2009) was utilized for the purpose of this research, in order to make comparability of results feasible. Even though the typology is not comprehensive, it is quite a good fit for the travel industry, since the types of sites that it entails are the ones which are most commonly found in the online tourism domain. Based on that framework social media website types are divided in the following six general categories:

- Virtual community sites such as Lonely Planet (www.lonelyplanet.com) and IgoUgo (www.igougo.com). Key features in this category are the interactive forums of discussion and the possibility for social connections to be established between the members of the community based on common interests.

- Consumer review sites such as Tripadvisor (www.tripadvisor.com) and Virtual Tourist (www.virtualtourist), where the main activity on the website is to post reviews and ratings or comment on other traveler’s reviews about accommodations, attractions etc.

- Personal blogs and blog aggregators such as Blogspot (www.blogspot.com) and Wordpress (www.wordpress.com). A key feature of blogs is that the
content is always tagged and structured in a reverse chronological order and that there is always an option for the readers to write comments.

- Social networking sites such as Facebook and MySpace. These are general purpose networks of people whose main components are the possibility for creation of detailed member profiles, who can connect to each other and then interact socially, thus forming a social network.

- Media sharing sites such as YouTube and Flickr. The main function of these websites is to share media related content, particularly videos and photos while at the same time giving the option to other users to tag and comment on them.

- “Other” category that includes any other social media types that cannot be categorized based on the above classification. A typical example that would belong in this category is Wiki-type of sites.

A clarification to be made here is that based on most other social media typologies that exist in literature, virtual communities and consumer review sites do not typically constitute a category of their own. In fact they are often placed under the umbrella of social networking sites. Despite the common ground with social networking, for this study these two types of sites constitute independent categories given how prominently featured they are within the travel domain.

Additionally, it is worth to note that among the six social media types defined above there is certainly some common ground. For example a virtual community may have an integrated media sharing application in it, or a media sharing site can also have social networking elements. Obviously in all these “hybrid” cases the emphasis is put on the dominant element of a website, which enables the most accurate categorization.

As a result of the coding process:

- 6150 URLs were retrieved and reviewed from over six hundred search results pages corresponding to well over a hundred hotels
- 12 touristic cities in Finland and Greece were reviewed.
- 3 languages were used to review landing pages: English, Finnish, Greek
- 4 commercial search engines were deployed

Below is the data collection process, described step by step, as performed by the students and based on the framework discussed above.
1. Students start by making a Google search on "assigned location + hotels", e.g. "helsinki hotels", "athens hotels" etc.

2. Then they click on the link "Local business results for hotels near ..." above the Google Maps box and register the 10 hotels that appear on the 1st page.

3. Students search Google for each one of these ten hotels. The search terms are in the form "hotel name + location", e.g. "hotel kamp helsinki".

4. Students work with the URLs of the search engine result pages 1, 2 and 3 (30 results total) ignoring sponsored links both on the top of the page and on the right hand side as well as any maps-related results. The total number of page URLs to be reviewed is 300 for each destination.

5. Students register the following information on the database: The URL of the website they review, the website name, whether it is social media or not and if it is, then assign a social media type based on the provided framework. Moreover for every hotel reviewed additional information is stored regarding the type of the hotel, i.e. if it is local or if it belongs to an international chain.

There were some additional guidelines that were given to students. If the page is written in a language that the students do not understand, then a translation tool could be used to help students pass a judgment. Moreover, commercial websites whose main purpose is to drive reservations like Expedia (www.expedia.com) and Booking (www.booking.com) were not considered as social media for the purpose of this research (even if they included reviews).

Even though every effort was made to make the data as consistent as possible, a number of data consistency issues came up that had to be addressed. The two major ones pertained to instances of mis-classification of websites. Some websites that were not social media, contained consumer reviews, and this tempted some of the students to register them as social media. To address that a comprehensive list of “faux” social media websites was supplied to the students, who then made the necessary adjustments. For this study, as social media were considered sites where the social component was the essential basis for the value the site would add for its users. If the social element only had a secondary importance, (when the principal focus is on the bookings for example in websites like Expedia.com and Booking.com) then the respective sites were not classified as social media.

This rule has been the basis for judgment in several border-line cases. As a result there were a number of sites classified as social media even though they did offer possibilities for bookings. In these cases, however, booking option was an addition to
the service and by no means the core part of it. When the students were faced with such border-line cases communication with the author was established until agreement was reached.

In addition to this, all the students shared a common online workbook hosted on Google Docs servers. This cloud computing solution employed not only offered the students the possibility to share and compare their work on real time, but also enabled them to utilize the presence and chat applications in order to discuss their findings and help each other resolve any issues (note: this is a practice that was followed intentionally in order for the study itself to benefit by harnessing the power of social media. This enhanced the effectiveness of the human factor involved and helped to provide improved results accuracy).

In another case, some web sites were correctly classified as social media but then not every coder had the same opinion about the specific social media type they belong to. In the spirit of Web 2.0 and social media, this issue was resolved by applying a “collective intelligence” rule. Practically the verdict of the majority was applied across the whole set of data specific to a website. This was essential to ensure data consistency and accuracy. Interestingly the logic seemed to make sense in all reviewed by the author cases of mis-classification.

Once results were documented, coded and then updated to ensure consistency, the analysis part followed suit. In order to answer the research questions this part entailed calculating the average number of social media results recorded for every city and similarly the average number for every search results page. Then additional analysis was conducted with the aim of specifying the hotel names and brands that appear to attract the highest number of social media results as well as the specific social media websites that appear as the most dominant. Consideration was also given to the distribution of social media across the first three Google result pages and also the distribution of social media based on the destination criterion.

In the following chapter the results of the data analysis will be discussed in detail. This will provide answers to all the given research questions and based on these answers conclusions will be drawn and suggestions will be proposed for the travel marketers.
7 DATA ANALYSIS AND FINDINGS

The present part of the study contains the analysis results of the collected data, including also relevant visualization and commentary to address the six basic research questions.

In the following pages there is a section dedicated in the analysis of each of the research questions. Wherever meaningful the analysis expands in scope to include additional observations and commentary to supplement the basic research questions. The principal method deployed for the analysis is based on pivot table analysis performed with Microsoft Excel. This pivot table analysis practically examined every possible meaningful correlation among the variables. The tables and illustrations included in this part highlight the key relationships among the data, which provide meaningful answers to the research questions. Additionally, complementary analysis was performed using XLMiner, a data mining software program, which is used as an add-on application for Excel.

The data analysis process adopted in this study is described below:

• Step 1: Problem formulation- Corresponds to the definition of the research questions which are listed above.
• Step 2: Data collection- The process of collecting data through the search engines and registering them i.e. a process performed by the students.
• Step 3: Pre-processing (cleaning of data)- There were several instances of “noise” in the data (such as inconsistencies and missing values) that had to be manually cleaned, either by the students or the author.
• Step 4: Transformation- Categorical values (such as social media or not) were transformed to binary values (0 or 1) to enable better data manipulation.
• Step 5: Choice of data analysis method – Excel pivot table analysis and data mining.
• Step 6: Result evaluation and visualization. This is the part that follows suit.
7.1 Proportion of social media for hotel brand searches

Providing answer to the first research question requires a high-level overview of the representation of social media in the search results. This can be considered as the fundamental question which will shed the first light in the research and will set expectations for the questions to follow. Figure 7.1.1 gives a clear answer to this question by illustrating that a significant proportion of the total number of search results are classified under the social media label. In specific, from the 3600 observations available, 996 are social media related, leaving the remaining 2604 in the non-social media category.

Transformed into a percentage, this corresponds to 27.7% of the sample (as social media positive). Even though social media do not constitute the absolute majority, a representation approaching 30% of the sample can be considered as a very significant proportion. It is difficult to think of any other category of websites that could capture such a high level of visibility. It is even harder to imagine that social media could have achieved such a strong online presence in the recent past.

The first research question therefore solidly confirms the high importance of social media within the travel search context. Social media appear to be ubiquitous and surely impossible to ignore by travelers or by marketers. In this context, the above result can be considered as a landmark. There is still however no answer as to how prominently social media results are featured alongside other types of search results and neither which type they belong to. These questions are covered in detail in the sections following suit.
7.2 The main types of social media represented on the top three search pages

Given the high volume of social media results contained in the sample, it is vital to examine their classification. Six categories have been defined, namely: Customer Review Site (CRS), Virtual Community Site (VCS), Blog (BL), Media Sharing Site (MSS), Social Networking Site (SNS) and Other (OT). Based on the analysis as demonstrated in figure 7.2.1 the dominant category without doubt is Customer Review Site. This category includes websites such as Tripadvisor and Real Travel, the main objective of which is to enable travelers to read and write travel reviews. These sites appear to be prominent enough to capture over 50% of the results classified as social media. Despite the high margin, this result is not considered surprising, given the essential need of travelers to read relevant reviews before finalizing their travel plans. Among all the other types of social media sites, CRS are the ones containing the most comprehensive databases with reviews and ratings, and therefore the most likely ones to be visited by the travelers (even though similar information can exist in Blogs or Virtual Communities etc.).

At the same time, the result of this question has been influenced by the high popularity of a small set of websites such as Tripadvisor, Real Travel and Yahoo Travel, which by themselves alone have captured 75% of the CRS class and approximately 50% of the whole sample. Virtual Community Sites (VCS) are ranked second, with representative sites like Virtual Tourist, TravelBuddy and Lonely Planet. VCS correspond to 21.3% of the sample. A possible reason for this is the community features that they offer, which enable travelers to connect directly with each other, through profiles, forum discussions, integrated blogging and more.

It is noteworthy that CRS and VCS grouped together account for roughly 80% of the social media sites of the sample. This is particularly interesting, given the substantial degree of similarity between these two types. Testimony to this is the fact that the

![Figure 7.2.1 Social media results distribution by type (%)]
most frequent conflict in the data collected by the students, was the classification of
the same type of sites as VCS by some and CRS by others (in that case the collective
intelligence rule, described earlier on was employed to resolve the conflict).

Blogging at 15.9% holds a strong position in the social media map and this is rather
unexpected result, if one considers that blogs have to complete with a high number
of established, high-traffic websites for the limited space in the search engine results
pages. One way to justify this might be the highly relevant content that some of these
blogs have to offer, particularly in cases where a blog is maintained by the hotel
company itself and therefore appears as a very relevant site (and therefore well
ranked) on the search results. Another reason for the significant presence of blogs can
be the influence of blog aggregators, such as the general purpose Wikio and Blogspot
or the travel focused TravelPod and Hotel Chatter. These sites have an established
presence on the web, which can partly explain their good visibility on search.

The Other category (OT) is in the fourth position with 36 instances of it recorded,
corresponding to a 3.6% of the sample. There was a range of website types classified
by the students as OT. In many cases those were Wiki type websites. Overall the
impact of these websites appears to be low.

Seventeen and four were the results belonging to the Media Sharing Site (MSS) and
Social Networking Site (SNS) category respectively. The percentages are extremely
low, 1.7% and 0.4%, which in many ways emerges as an unexpected outcome, given
that these two categories contain social media mega-sites such as Facebook, You
Tube and Twitter. Even though all of them were present (and most notably of all You
Tube), there was just a minimal number of associated results.

One possible explanation for this is that in order for the hotel firms to be visible in
this type of media, they need to take steps and proactively build their presence
themselves. This could be achieved by establishing a Facebook fun page, a Twitter
account, a dedicated You Tube channel among other possible initiatives, and then
promote these on the web, in order to gain traction, followers and ultimately
improved search rankings. There is therefore a degree of contrast between that
proactive behavior and the “passive” inclusion in customer review sites, virtual
communities etc. that, in fact, the travelers rather than the marketers are in charge of.
Thus it appears that for the present time there is only a limited number of hotels
proactively trying to build their social media presence using social networking and
media sharing sites. This could partly explain the striking absence of these sites in the
sample.

A thought-provoking finding that applies across all the social media types is that
typically a small number of websites, usually three, four, or five depending on the
exact category, account for the vast majority of the associated observations (similar to
the case of the consumer review sites mentioned at the start of this section). On the other hand, there are numerous minor websites represented in the data set by one or two instances only.

### 7.3 Top social media sites represented in the search results

In this section the analysis will dive one level deeper so as to examine which are the most dominant social media websites in the data set. Interesting correlations with the other variables will be presented as well. There was a total of 168 unique social media domains registered in the sample, which represented 996 social media results in total. There was a high number of unique websites registered just one, two or three times. In fact this group of low visibility sites constitutes 86% of the total number of unique domains recorded. Nevertheless, due to low frequency, the relative importance of these sites is quite low, representing just 20% of the social media results in the sample. This means that the relatively small number of the remaining sites account for the majority of the social media results. Table 7.3.1 illustrates this group of sites whose frequency in the sample is 12 or more. One can easily observe the strong presence of Tripadvisor which is by far the site with the greatest visibility in the search results.

<table>
<thead>
<tr>
<th>Social media site</th>
<th>Frequency</th>
<th>Frequency %</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tripadvisor</td>
<td>286</td>
<td>28.7%</td>
<td>28.7%</td>
</tr>
<tr>
<td>Real travel</td>
<td>94</td>
<td>9.4%</td>
<td>38.2%</td>
</tr>
<tr>
<td>Wikio</td>
<td>73</td>
<td>7.3%</td>
<td>45.5%</td>
</tr>
<tr>
<td>Virtual Tourist</td>
<td>61</td>
<td>6.1%</td>
<td>51.6%</td>
</tr>
<tr>
<td>TravBuddy</td>
<td>59</td>
<td>5.9%</td>
<td>57.5%</td>
</tr>
<tr>
<td>Travel Yahoo</td>
<td>49</td>
<td>4.9%</td>
<td>62.4%</td>
</tr>
<tr>
<td>TravelPod</td>
<td>40</td>
<td>4.0%</td>
<td>66.5%</td>
</tr>
<tr>
<td>Lonely Planet</td>
<td>30</td>
<td>3.0%</td>
<td>69.5%</td>
</tr>
<tr>
<td>Ciao</td>
<td>14</td>
<td>1.4%</td>
<td>70.9%</td>
</tr>
<tr>
<td>Hotel Chatter</td>
<td>12</td>
<td>1.2%</td>
<td>72.1%</td>
</tr>
</tbody>
</table>

Table 7.3.1 Frequency, frequency (%) and cumulative (%) of SM sites

In fact, it accounts for roughly 30% of all social media results documented. Tripadvisor is one of the most frequently visited travel websites and with reference to hotel search its popularity becomes evident through the favorable search rankings. In the second and third place, but in long distance from Tripadvisor there is Real Travel, another customer review site and Wikio, a blog aggregator.
The solid presence of a limited number of social media sites which appear in the search results with high frequencies provide motivation to examine the distribution of frequencies in different parts of the sample. The 10 first sites whose frequency is 12 or more, represent over 70% of the retrieved social media results, as demonstrated in Table 7.3.1, while they only account for 7% of the registered 168 unique domains. These relationships are illustrated in figure 7.3.1 which highlights how prominent the presence of the top social media websites is. At the same time, it demonstrates how the impact diminishes as we move on the horizontal axis and the corresponding part of the blue line starts to flatten. Conversely, we have a long list of over hundred websites appearing just one time each, but their cumulative impact is not more than 10% with reference to the total number of social media results found in the sample.

There is also a middle layer containing 57 sites with occurrences between 2 and 8 times, which account for 15% of the whole sample. Provided the results of the analysis above, a possible conclusion is that the distribution of frequencies among the unique domains follows a (common to the Web) head-torso-long tail pattern. The head is by far the strongest part, considering the top four sites (Tripadvisor, Real Travel, Wikio and Virtual Tourist) represent over 50% of all the instances of social media found in the search results.

### 7.4 Destination specific social media distribution

In this section the social media results are distributed based on the destination of the hotel query that triggered them. Finnish destinations accumulated 465 social media results, whereas Greek destinations 531. Given that there were seven Greek destinations against five from Finland, it would be more meaningful to examine the average amount of social media results associated with a Greek and a Finnish destination. The result is that destinations in Finland average 93 social media related
results whereas the Greek ones 75. Converting this on a per hotel basis, means that when users searches through the first three search pages for a hotel based in Finland they are likely to find five additional social media results, comparing with Greece-based hotels. This is acknowledged as a significant deviation, suggesting that Finnish hotels are achieving substantially improved social media exposure.

In figure 7.4.1 the results for each of the tested destinations are illustrated, from which several interesting points can be made. The first observation is that the size of a destination population-wise seems to play a negative role with reference to the number of associated social media results. Two Finnish cities, Turku and Tampere, appear to be the ones to enjoy the greatest visibility on social media with 129 and 128 cases respectively, as illustrated in figure 7.4.1. Rhodes follows in the third position, but relatively close to the other two, with a score of 115. Then there seems to be a gap comparing to the other destinations, whose score range between 80 (Patras) and 58 (Heraklion). It is noteworthy that the largest cities, population-wise, like Helsinki, Athens, Espoo and Heraklion are left at the bottom of the list. One possible explanation for this is that due to their size and role as tourist hubs, the social media results for these locations are overshadowed by search results which are rather commercial in nature, associated with sites such Expedia, Booking, Priceline and others for which these destinations can prove quite lucrative. If Finnish and Greek cities are considered as one group, then the average volume of social media results per destination equals to 83, out of a total of 300 results, which suggests high exposure of social media across all the destinations.

Within this context, the scores of the two Finnish cities remain impressive considering that almost every second search result for hotels based on these two areas is expected to be of social media nature. For Turku the presence of blogs
appears to be the key for its top position in the list (it is the city which has by far the highest number of blog related social media). Tampere on the other hand tops the lists of virtual communities, media sharing and other sites.

Given the high influence of customer review and virtual community sites this section also contains a city break down for these specific categories.

Figure 7.4.3 reveals that Rhodes has, by a margin, gathered the most customer review sites, with Turku and Tampere following. It is impressive that two very touristic Greek cities, Athens and Herakion, are considerably lacking exposure through customer review sites. With reference to virtual communities, Tampere (top position) and Turku (third position) maintain their strong presence, this time joined also by Patras, which has a remarkably good score in this category, as shown in figure 7.4.2. Analyzing in higher degree of granularity enables to examine the locations through the perspective of the websites with key presence. The high impact of Tripadvisor and Real Travel, as the sites with the highest ratio of social media in the sample, has already been emphasized. Turku is the location that triggered the most results from Tripadvisor, 33 in total. This means that over 10% of the total sample from Turku, including also non-social media results, is attributed to Tripadvisor. Tampere and Larissa follow closely with a score of 28. Real Travel for the first time provides a somehow diverse picture, with cities other than Tampere and Turku topping the list. Two Greek cities are the ones triggering the highest number of results from Real Travel, followed by Turku, which stands in the third position.

7.5 Social media content on the first three search pages

Understanding the way social media results are distributed across the search engine results pages is vital, since searchers tend to prioritize top results in top pages. The
importance of first page display is particularly high. Interestingly the distribution appears to be quite balanced and homogenous, with approximately one third of social media results corresponding to each one of the first three results pages.

As shown in figure 7.5.1 there is a light trend for more social media results in latter pages, but this trend is very mild. What is indeed captivating based on this result, is that social media appear to be ubiquitous across all three first pages and clearly not “buried” at the bottom, nor in any way disappearing after the first page. Examining further the composition of these numbers, it is possible to analyze the results further. Breaking down the data based on the social media type, reveals the dominance of customer review sites in the first page, corresponding to over 70% of the total number of social media located on the first page. This could be a sign of the high weight these sites have for the results provided by the Google search engine. This seems to be true particularly for the website Tripadvisor that alone accounts for 55% of the customer review sites on the first page, 43% on the second and 48% on the third.

Figure 7.5.1 Percentage of SM, non-SM and SM/non-SM ratio across search pages
Another key site in this category, Real Travel follows a declining trend, as it starts strong with 49 observations on the first page, 36 on the second and just 9 on the third one. The distribution landscape, however, changes as we move on to the second and third page. The impact of customer review sites is still high (48% for the second and 51% for the third page), but nowhere as close as in the first page. This is illustrated on figure 7.5.2. With the exception of virtual communities (which are subject to some sharp fluctuations across the three pages), all the other social media types clearly gain ground moving on to the second and third page.

![Figure 7.5.2 Percentage of SM distributed by type for search pages 1, 2, 3](image)

Blog type in particular is the one to gain the most. This whole behavior could be correlated to the combined decline in the number of social media results labeled under the category of customer review and virtual community sites, in the pages following the first one.

Even though numbers are relatively low, of particular interest is the rise of media sharing sites, which despite being almost non-existing in the first page, they clearly gain ground on the following pages. In most cases these media sharing sites are YouTube videos created and uploaded by users. Another interesting conclusion with reference to the presence of individual sites, is the lack of any underlying pattern with regard to the frequency of these sites across the three pages. In most cases there
are fluctuations from one page to the other, that appear random, particularly when examining the key websites. For example, Yahoo Travel has 19, 10 and 20 occurrences respectively on each one of the three pages. Travel Buddy 7, 33 and then 19. Lonely Planet scores 7, 17 and 6 respectively.

We can therefore conclude that, as far as ranking is concerned, what counts the most is likely the actual query itself, rather than the existence of a rule that relates websites with specific search results pages. With reference to locations, Rhodes is the one with the highest number of social media results identified on its first page, with a total of 41. Turku is the location scoring the highest number of social media on the second page with 44 and third one with an impressive 56, which means that over 50% of all results in the third page for Turku are social media related. This is the top score across all the search pages and cities. On the other hand, Vantaa (18 for the first page), Helsinki (15 for the third page) and Espoo (15 for the second page) are the locations with the lowest scores.

Obviously given the way visibility is shared among the pages, having a low score on the first page (locations such as Vantaa, Heraklion, Espoo, Helsinki and Larissa) could be a more alarming signal compared to a low score on the third page. A low score on the third page typically implies the dominance of commercial nature results. This, however, could be equal to a missed opportunity for the concerned destinations, given how credible and influential for travel decision making social media are.

**7.6 Hotel brands and number of social media results**

This section gives a brand related dimension to the analysis of the results. In the sample there was a total of 120 hotels searched for, from 12 locations. There was a wide range of social media occurrences identified across the various hotels, ranging from 2 all the way up to 27, with a median of 7 and an average of 8.44. The majority of the hotels and accommodations (with social media results) belong to the Local category with 837 associated social media observations in total, while the ones belonging to the International chain type number 160. Table 7.6.1 contains the 24 hotels, corresponding to the top 20% of the sample based on the frequency of social media results associated with the respective hotel searches. On the very top, Omena hotel has by a significant margin the highest number of associated results, 27 in total. It is followed by Kauppi Hotel (19), Park Hotel (17) and Scandic Julia (17).
Interestingly the top of the list is exclusively dominated by hotels based either in Turku or Tampere. Additional hotels from these two cities are also very prominently featured in the list. On a country level, most hotels are Finland-based and relatively few from Greece, with the island of Rhodes being the top Greek contributor.

The Omena hotel case is worth to be examined more closely. It is remarkable that there is a considerably balanced distribution of social media types, including 5 from Blogs, 7 from customer review sites, 8 other (OT) and 7 virtual community sites. Comparing with other top hotels, it is the number of blogs and other (OT) sites that bring Omena to the very top. In fact it is no surprise that across all top hotels the customer review sites and virtual community sites constitute the highest proportion of associated social media results. Similarly to Omena, the amount of related blog results is often the main differentiator that brings the hotels higher or lower in the list.

<table>
<thead>
<tr>
<th>Hotel name</th>
<th>City</th>
<th>Country</th>
<th>Social media results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omenahotelli</td>
<td>Tampere</td>
<td>Finland</td>
<td>27</td>
</tr>
<tr>
<td>Hotelli Kauppi</td>
<td>Tampere</td>
<td>Finland</td>
<td>19</td>
</tr>
<tr>
<td>Park Hotel</td>
<td>Turku</td>
<td>Finland</td>
<td>17</td>
</tr>
<tr>
<td>Scandic Julia</td>
<td>Turku</td>
<td>Finland</td>
<td>17</td>
</tr>
<tr>
<td>Best Western Plaza</td>
<td>Rhodes</td>
<td>Greece</td>
<td>16</td>
</tr>
<tr>
<td>Cosmopolitan</td>
<td>Rhodes</td>
<td>Greece</td>
<td>15</td>
</tr>
<tr>
<td>Faliraki SA</td>
<td>Rhodes</td>
<td>Greece</td>
<td>15</td>
</tr>
<tr>
<td>Hotelli Korpilampi</td>
<td>Espoo</td>
<td>Finland</td>
<td>15</td>
</tr>
<tr>
<td>Best Western Hotel</td>
<td>Turku</td>
<td>Finland</td>
<td>13</td>
</tr>
<tr>
<td>Seaport</td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Sokos Hotel Seurahuone</td>
<td>Turku</td>
<td>Finland</td>
<td>14</td>
</tr>
<tr>
<td>Aldemar Paradise</td>
<td>Rhodes</td>
<td>Greece</td>
<td>13</td>
</tr>
<tr>
<td>Holiday Club Caribia</td>
<td>Turku</td>
<td>Finland</td>
<td>13</td>
</tr>
<tr>
<td>Rhodos Beach</td>
<td>Rhodes</td>
<td>Greece</td>
<td>13</td>
</tr>
<tr>
<td>Sokos Hotel City Bös</td>
<td>Turku</td>
<td>Finland</td>
<td>13</td>
</tr>
<tr>
<td>Sokos Hotel Ilves</td>
<td>Tampere</td>
<td>Finland</td>
<td>13</td>
</tr>
<tr>
<td>Sokos Hotel Tammer</td>
<td>Tampere</td>
<td>Finland</td>
<td>13</td>
</tr>
<tr>
<td>Cumulus Hotel Koskikatu</td>
<td>Tampere</td>
<td>Finland</td>
<td>12</td>
</tr>
<tr>
<td>Holiday Inn</td>
<td>Turku</td>
<td>Finland</td>
<td>12</td>
</tr>
<tr>
<td>Hotel Siikaranta</td>
<td>Espoo</td>
<td>Finland</td>
<td>12</td>
</tr>
<tr>
<td>Scandic Gateway</td>
<td>Helsinki</td>
<td>Finland</td>
<td>12</td>
</tr>
<tr>
<td>Cumulus Hotel</td>
<td>Turku</td>
<td>Finland</td>
<td>11</td>
</tr>
<tr>
<td>Mediterrane Hotel</td>
<td>Patras</td>
<td>Greece</td>
<td>11</td>
</tr>
<tr>
<td>Olympic Star hotel</td>
<td>Patras</td>
<td>Greece</td>
<td>11</td>
</tr>
<tr>
<td>Semiramis</td>
<td>Rhodes</td>
<td>Greece</td>
<td>11</td>
</tr>
<tr>
<td>Sokos Hotel Villa</td>
<td>Tampere</td>
<td>Finland</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 7.6.1 Number of social media results for the top 24 hotels of the sample
With regard to search results page analysis, a first observation is that the distribution of social media for the top hotels seems to imply that the strongest social media search pages are the ones that follow page one. Omena hotel has 8 occurrences on page 1, 10 for page 2 and 9 for page 3. Kauppi hotel has 5 for page 1 and 7 for pages 2 and 3. Park Hotel has 4, 7 and 6 respectively for the first three pages.

It was also worth examining which hotels top the various social media categories. Best Western in Turku has the top spot in the list of hotels with the highest number of Blog related results, followed by Hotel Palace Sello and Park Hotel Turku. In the customer review type Hotel Siikaranta is the most prominent in the list, followed by Aldemar of Rhodes and Hotel Korpilampi of Espoo. In the virtual community category Omena leads and it is followed by Cosmopolitan in Rhodes and Park Hotel of Turku. For the other three categories the numbers are quite low, so hotels can be identified on top of the list by having just two or three associated observations.

Another parameter associated with the hotels is their type, namely International or Local. The local hotels were higher in number in the whole sample; nevertheless, proportionally speaking there was a lower percentage of related results that had social media content, precisely 27.1%. International hotels corresponded to approximately one sixth of the total number of search results associated with local hotels and the rate of social against non-social media was 31.4%. This is not a vast dissimilarity percentage-wise but given the significant size of the sample, it does suggest an improved probability for hotels belonging to international chains to trigger additional social media results. Investigating the types of social media websites, that are generated upon relevant searches, the observation is that only minimal differences exist when considering the proportional composition of the numbers. In other words, the percentage of customer review, virtual community sites etc. in both local and international categories are quite similar, thus it cannot be suggested that this is a critical factor influencing the results.

7.7 Sub-study analysis

7.7.1 Sub-study 1

In this sub-study 10 Helsinki hotel brands were tested using, besides Google, three additional search engines, namely Yahoo, Bing and Ask. Overall all four search engines are very close to each other, concerning the number of social media results they return. Google provides 64 social media results for Helsinki, 65 for Ask.com, 66 for Bing and 67 for Yahoo.

This produces evidence that social media have strong presence across all major commercial search engines on the web. In this study Yahoo was the search engine to return the highest number of social media related results and Google the one with the fewest. The margin, however, is so small that it cannot be considered significant.
By looking deeper, into the specific social media types, some additional assumptions can be made. For the blogs category for example, Google has by a margin provided the most results, five in total, while Yahoo has only one and Ask none. The distribution is more balanced in the customer review site category where there is a range from 41 (Google) to 52 returned results (Ask).

Media sharing sites seem to have been largely ignored by all search engines, since across the sample there is only one instance, coming from Google. As regards the other (OT) type, Bing is the search engine with the highest score (7) and Yahoo is the second and last search engine in this list with only one observation. Social networking sites are also scarce. Yahoo and Bing have two and Google just one. With reference to the virtual community type there is more balance, with Google and Yahoo holding 16, Ask 13 and Bing 9. The conclusion is that for the major types, there is more balance in the results, in comparison with the “niche” ones.

Examining the individual site distribution, it should not be considered a surprise that the major sites such as TripAdvisor, Real Travel, Yahoo Travel, and Virtual Tourist. are ubiquitous, even though separate search engines seem to give alternative weight to each one of them. Then there is the long tail of sites that can be quite dissimilar from one search engine to another. What is quite remarkable though is that some search engines like Bing and Yahoo, seem to be more pluralistic in what regards the variety of results they provide. Yahoo’s social media results originate from 18 unique domains and Bing’s from 20. Conversely Google returns 12 unique domains and Ask just 9.

Providing results from additional domains does not necessarily mean enhanced results, since what is most important is the relevance for the user. However, this sub-study demonstrates that potentially the travel domain is better represented in some search engines, at least within the context of the first three result pages in a hotel search setting.

7.7.2 Sub-study 2:
Ten hotels were tested in the Finnish and Greek search domains for corresponding local destinations. Google was configured appropriately so as to return results written only in Finnish or Greek language.

This approach has been employed in order to provide a better understanding on how prominent the travel social media landscape is in Finland and Greece. The results provide a significantly different picture for the two languages. In Finnish there are a total of 51 social media associated results. This is certainly lower if we compare with the English language results, but if we consider that Finnish language is only used by a relatively small population, it still shows that there is a high level of social media activity around the travel and particularly hotel sector.
It is worth to note also that there is a high level of pluralism observed as the results come from a total of 35 unique domain names. Therefore, contrary to the searches in English language, there is no high concentration of results around three or four sites; instead there is a wide range of websites contributing one, two or three observations each. The website with the highest frequency (5) is helsinki-hotelli.info. Remarkable is also the presence of a high number of other (OT) sites in the social media results mix.

The travel social media landscape for the Greek language is quite different. In fact there were extremely few social media related results identified. In total there were just 11 results registered, most of them attributed to Blogs. In several cases these social media results were rather random since the actual theme of the social media websites represented was not relevant to the query. Thus those were practically inaccurate search results returned by the search engine. This suggests that Greek language is in shortage of social media sites that discuss travel. It might also mean that Greek users prefer to use foreign language social media sites. Another possibility could be that given the particularity of the Greek language, search engines are not capable of finding and providing adequate number of relevant results. Due to the limited data used for the present sub-study it is impossible to provide a good answer on this question. The fact of the matter nonetheless is that depending on the language, the social media landscape can vary considerably.
7.7.3 Sub-study 3

The more generic-type keyword “hotel” was used in combination with the name of the destination. Examples of queries used for this study are: “hotels Athens” “hotels Helsinki” etc., for a total of 10 locations, 5 from Greece and 5 from Finland. There was a total of 47 social media search results identified. This is considerably lower than the average of 83 related results which were registered when providing queries for specific hotels in the same locations.

![Figure 7.7.1 SM generated by hotel generic queries, distribution by type](image)

Even though the number of social media results for the hotel generic type is roughly half of the average in the study, the main contributors are practically the same, namely, Tripadvisor, Yahoo Travel, Virtual Tourist, Real Travel etc. The order of the social media types based on contribution is also similar, with customer review sites leading with a high margin (by providing over 70% of the total) and followed by virtual communities (20%) and the remaining types, which have a minimal presence as shown in figure 7.7.1. The main assumption from this study is that the more specific the search term used is (always within the travel context), the higher the likelihood to meet social media results.
7.7.4 Sub-study 4

For a set of keywords the sub-study 4 goes past the first three results pages, to include results up to the tenth page, seeking for insights as to whether additional results pages include additional types of social media. In this sub-study, search queries were supplied to Google, for a total of six hotels, three for each of the two destination cities, Helsinki and Athens. Therefore, for each one of the destinations 300 results were collected. This is the same number per destination as with the main study, where searches were made for 10 hotels per destination and results were collected for the first three pages, rather than the first ten. Even though not all factors are equal, this study can still enable comparisons, based on the total number of social media related results collected with the two methods. When search results were collected via the main study method for Helsinki there were 64 social media results out of 300, whereas with the sub-study 58. Similarly for Athens, the first method provided 65 social media results and the second 50. It seems therefore that the proportion of social media results is denser in the first pages, still without this implying any critical dissimilarity, at least based on the data collected for this study.

Figure 7.7.2 demonstrates that while the first three pages are relatively stable, as far as social media to non-social media ratio is concerned, from the fourth page and onwards there is a considerably higher degree of fluctuation. Moreover, considering the social media results depend also on the specific hotels chosen and not just the location, it is also likely that there is some bias in the comparison (given that the hotels are not equal in number).

Examining the social media types and the social media specific websites it is similarly hard to draw conclusions or assumptions about how the additional search
pages might be affecting the social media distribution. The main travel social media websites are still present. In particular Tripadvisor has presence across all the ten pages and it is the only social media site to achieve this. The proportional contribution, however, of these key social media sites, seems to be dissimilar when comparing the two methodologies, with the second one providing more diversified results.

Another observation is that sites like You Tube and Flickr (which had very low visibility with the main study) are now displayed a number of times and this occurs mainly on pages 7, 8 or 9. This could provide some evidence that media sharing sites are more visible as we move on deeper in the results provided by the search engines. A suggestion for further research would be to repeat this type of study by including more destinations. In any case, however, the result would be rather of theoretical and not practical value, given that users rarely visit search results pages beyond the first three, as it has been noted earlier on.

7.8 Comparing social media with non-social media results.

The focus of this study is on social media content and the way it is represented within search results. However, the data that have been collected for this study allow for some analysis of the non-social media results, as well as a useful comparison between social media and non-social media results. The non-social media observations were grouped by site and table 7.8.1 illustrates the ten websites that provided the highest number of non-social media results.

All of these sites are of commercial nature, most of them belonging to the online travel agency type, with prime examples Booking (www.booking.com), Hotels Combined (www.hotelscombined.com) and Agoda (www.agoda.com). Interesting is also the presence of travel search engine, Kayak (www.kayak.com) at the sixth position and Maplandia (www.maplandia.com), a service that utilizes extensively the Google maps application, at the eighth place. Booking is by far the most dominant site, but after it there is a balance with regard to the proportion of non-social media results that the other top sites capture. The top 50 non-social media unique websites account for 50% the non-social media observations.
It is interesting to compare this with a similar analysis concerning social media results. As already revealed, a relatively low number of the social media websites contributes a disproportionately high number of social media results. Tripadvisor for example, top site in the social media category, contributes double the number of observations that Booking does for the non-social media category. Real Travel and Wikio are ahead of Hotels-Combined and Agoda in the second and third position respectively.

The difference is even wider if we consider the corresponding relative proportions in the third and sixth column of table 7.8.1. After this the numbers of the following social media sites are rapidly diminishing in comparison with the non-social media. Nonetheless, it is still valid to conclude that comparing the two categories of sites, social media are much more saturated by a small number of sites, especially in the top.

<table>
<thead>
<tr>
<th>Website (Commercial)</th>
<th>Observations</th>
<th>Non-SM %</th>
<th>Website (Social media)</th>
<th>Observations</th>
<th>SM %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booking</td>
<td>118</td>
<td>4.3%</td>
<td>Tripadvisor</td>
<td>286</td>
<td>28.7%</td>
</tr>
<tr>
<td>Hotelscombined</td>
<td>83</td>
<td>3.1%</td>
<td>Real travel</td>
<td>94</td>
<td>9.4%</td>
</tr>
<tr>
<td>Agoda</td>
<td>68</td>
<td>2.5%</td>
<td>Wikio</td>
<td>73</td>
<td>7.3%</td>
</tr>
<tr>
<td>Activehotels</td>
<td>60</td>
<td>2.2%</td>
<td>Virtual Tourist</td>
<td>61</td>
<td>6.1%</td>
</tr>
<tr>
<td>Expedia</td>
<td>58</td>
<td>2.1%</td>
<td>TravBuddy</td>
<td>59</td>
<td>5.9%</td>
</tr>
<tr>
<td>Kayak</td>
<td>53</td>
<td>2.0%</td>
<td>Travel Yahoo</td>
<td>49</td>
<td>4.9%</td>
</tr>
<tr>
<td>Wego</td>
<td>52</td>
<td>1.9%</td>
<td>TravelPod</td>
<td>40</td>
<td>4.0%</td>
</tr>
<tr>
<td>Maplandia</td>
<td>49</td>
<td>1.8%</td>
<td>Lonely Planet</td>
<td>30</td>
<td>3.0%</td>
</tr>
<tr>
<td>Hotelclub</td>
<td>49</td>
<td>1.8%</td>
<td>Ciao</td>
<td>14</td>
<td>1.4%</td>
</tr>
<tr>
<td>Yeego</td>
<td>46</td>
<td>1.7%</td>
<td>Hotel Chatter</td>
<td>12</td>
<td>1.2%</td>
</tr>
</tbody>
</table>

*Table 7.8.1 Number of observations and % of the top commercial and SM websites*
7.9 Data mining analysis

Data mining, also known as Knowledge Discovery in Databases, is a technology-enabled field that aims at inductively extracting novel implicit and actionable knowledge from large datasets.

Figure 7.9.1 Data mining output (XLMiner) linking social media type, city and search page
It can be considered as a convergence of fields like statistics, machine learning and artificial intelligence. There are various methods that can be used in data mining, others predictive such as regression, classification and collaborative filtering and others descriptive like clustering, association rules and deviation detection.

The aim of this section is to identify interesting associations or correlation relationships among the set of data items that were collected and the methodology deemed the most fit to perform this task is association rules. Figure 7.9.1 illustrates the output of XLMiner’s association rule mining. XLMiner is a data mining add-in for Excel used for identifying patterns and relationships in large datasets. The association rules listed above display attribute value conditions that occur frequently together in the dataset.

Key metrics from the table include: A) Support, which is the number of entries that include all items in the antecedent and consequent parts of the rule. B) Confidence, which expresses the ratio of the number of entries that include all items in the consequent, as well as the antecedent to the number of entries that include all items in the antecedent. C) Lift, which is the ratio of Confidence to Expected Confidence (the number of entries that include the consequent divided by the total number of entries).

For the present test the minimum confidence was set to 50% and the minimum support to 10, which provided the flexibility to the XLMiner tool to return an extensive number of results, which could then be available for evaluation. Figure 7.9.1 lists the top 20 rules that express probabilistic relationships among the data. Intentionally variables with high variance (the website URLs, website names and search queries) or low variance (type of hotel i.e. international or local) have been excluded from the current test. This is in order to allow the analysis to focus on fewer and more relevant associations among social media types, cities and search pages, that would be suitable to the nature of the study.

For example rule 1 shows that when search page=3 and social media type Other (OT), then chances are that the city in question is Tampere. In other words, the rule emphasizes the fact the there is a significant proportion of OT websites on the third page for the city of Tampere (always in relation to the other destinations). The high lift of 5.7 implies that the probability that “if the result is from page 3 and classified as OT, then the city is Tampere” is roughly 6 times stronger. The 73% confidence expresses the conditional probability that a randomly selected item in the data set will include “Tampere” in the consequent part, provided that the antecedent contains “search page 3” and “OT”.

This way of “reading” the rule interpretation of this methodology might seem counter-intuitive to some extent, given the nature of the study; however it refers to
real relationships between the data. Rule 2 generalizes further and suggests that the proportion of OT websites is high in Tampere, regardless of page number (even though we know that page 3 is the one with the strongest presence of OT sites). In practice, it wouldn’t be meaningful to examine all the rules one by one, since some of them may appear self-evident or others simply suggest some of the points covered already with the pivot table analysis earlier in the chapter.

For example several of the rules are of the type: (1, destination) as the antecedent and (CRS) as the antecedent. These rules support the suggestion that for most of the cities, the first search results page is dominated by social media of the CRS type. This is also implied in a more general form with rule 17. Other interesting rules highlight the relationship between VCS in page 2 and BL in page 3 for Turku.

<table>
<thead>
<tr>
<th>Rule #</th>
<th>Conf. %</th>
<th>Antecedent (a)</th>
<th>Consequent (c)</th>
<th>Support(a)</th>
<th>Support(c)</th>
<th>Support(a U c)</th>
<th>Lift Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>Omenahotelli?&gt;</td>
<td>Tampere</td>
<td>27</td>
<td>128</td>
<td>27</td>
<td>7.8</td>
</tr>
<tr>
<td>2</td>
<td>57.89</td>
<td>TravelPod =&gt;</td>
<td>3</td>
<td>38</td>
<td>336</td>
<td>22</td>
<td>1.7</td>
</tr>
<tr>
<td>3</td>
<td>54.39</td>
<td>TravelBuddy =&gt;</td>
<td>2</td>
<td>57</td>
<td>335</td>
<td>31</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>51.85</td>
<td>Real travel =&gt;</td>
<td>1</td>
<td>81</td>
<td>325</td>
<td>42</td>
<td>1.6</td>
</tr>
</tbody>
</table>

*Table 7.9.1 Association rules linking city, hotel, website and search page variables*

There are several other similar “hidden relationships” for other cities, which would have been challenging to identify with pivot table analysis only, such as CRS for Thessaloniki in page 1 and VCS for Heraklion in page 2. One of the drawbacks, nonetheless, of this particular analysis (given the chosen settings relating to minimum support and minimum confidence) is that it tends to focus more on very specific relationships between data, rather than offering the greater picture.

Table 7.9.1 displays rules based on search page, city and domain name variables. Rule 1 is self-evident given that we already know that Omena hotel is based in Tampere. The following three rules suggest that RealTravel is associated with the first page of search results, whereas TravelBuddy with the second and TravelPod with the third.

Several other variable sets were tested with alternative combinations of minimum support and confidence, without however revealing any novel insights. The main reason for this is that the Excel-based descriptive analysis that preceded, already covered multiple combinations of relationships among the data. This came as a result of the relatively low volume of variables and data collected for the purpose of this study.

Data mining could have been of more use in the case that the data set would contain a much higher number of observations, which would render knowledge discovery difficult to achieve manually. Moreover association rules analysis would have been more relevant if the variables had a higher degree of independency. This is the
reason why many of the rules that the program produced were self-evident, for example when a given hotel is associated with a particular city, something which is already known in advance. Despite these shortcomings, analysing the data set with a data mining program like XLMiner was successful in emerging subtle relationships that would have been unlikely to identify otherwise.
8 CONCLUSIONS

8.1 General conclusions
Given the high significance and increasing use of web search and social media sites in the context of travel information search, this study has attempted to analyze the way social media are represented in the search results. This study contributes to the online travel marketing domain by “connecting the dots” between two major forces underlying the travel (and particularly the hotel domain), search and social media. This was achieved by examining the relationship between the two within the hotel search context, through the analysis of empirical data. This is an area which has not been widely researched previously and this study supplements previous research which has focused more on the sociological and psychological aspects of social media use.

Having as a starting point Xiang and Gretzel’s (2009) research, this study builds on that work but shifts the focus from the generic type of travel search to specific hotel brand search. The study has also tested a number of additional assumptions in the form of sub-studies which were analyzed on a lower scale. The sub-studies included local language results, deeper search results, alternative search engine results and also a test on a generic term i.e. hotel.

The main finding is that, despite the fact their distributed nature and newness (Xiang and Gretzel, 2009), social media indeed constitute a very significant portion of the search results. As much as 27.7% of the search results on average, upon a hotel search, lead to a social media website, a finding that confirms the high importance of social media for the travel industry. This is indeed a proportion that cannot be ignored by the travel searchers. In fact, considering the high credibility of consumer-generated content, it is highly likely that travelers will be exposed to social media sites. On its own right this is a revelation of high importance for the hotel industry, especially with reference to understanding the ongoing shift of media attention to emerging media, not just in the online channel but on an overall level, too. This finding could constitute a challenge for the established marketing practices through the traditional channels of communication. Social media results are omni-present across all search pages (1-3) analyzed and across all the destinations involved, both in Finland and in Greece.

What is also interesting is the high degree of concentration of the websites that contribute the highest number of social media results. A handful of sites, led by Tripadvisor are responsible for the overwhelming majority of social media presence
in the sample. Most of these sites belong to the consumer review type (57%), followed by the virtual community type (21%) and with blogs having a decent presence too (16%). Not surprisingly first in the consumer review category ranks TripAdvisor, which is a powerhouse of hotel-related social media results.

Travel marketers are recommended to pay special attention to this site and it will be discussed further in a following section. It is not surprising that sites like TripAdvisor have strong influence on the results (considering the size and relevancy of their content). However, from another angle this seems to be limiting the choice of online travelers, by effectively confining the travel domain, as it is accessible to them, and possibly “pushing” some other sources of information, deeper in the search results. It is in fact not uncommon to have several instances of TripAdvisor on the first page of the search results, which is a signal of how dominant this website is.

In general there is a clear head-torso-long tail distribution pattern in the 996 social media results linked to the sample. Over 50% of the social media results originate from four sites, namely, TripAdvisor, Real Travel, Wikio and Virtual Tourist. On the other hand, a striking absence was observed for the mainstream social media such as Facebook, Twitter and YouTube and other popular sites, most notably belonging to the social networking and media sharing site categories.

A possible reason for that is that travel marketers have not yet exploited fully the opportunity of interacting with social media users through these channels. The analysis confirms in practice the assumption that social media sites are particularly search engine friendly. Even though search engines do not reveal such information on this topic, the general assumption is that this search engine friendliness is supported by the way social media sites are interlinked, their up-to-date content as well as the variety of content available.

Another notable conclusion relates to the way social media results are spread over hotels in cities of various sizes. Contrary to what would possibly be expected, i.e. more social media results for hotels in larger, touristic destinations, the picture is rather the opposite. Medium-sized destinations such as Turku and Tampere in Finland prove to be the social media champions, leaving far behind popular, touristic destinations such as Athens and Heraklion.

A possible explanation for this is the higher availability of commercial search results covering the most popular destinations. In general, Finnish hotels seem to be more exposed to social media in comparison with the ones based in Greece. The most remarkable example of a Finnish hotel is Omena, which out of the 120 hotel sample is the one with the highest number of related social media results. Moreover, even though most hotels in the sample have been classified as local, it appears that searches of international chains have more chances to return social media content.
Along with the main research, a number of sub-studies were carried out. The main conclusions are as follows: The first study utilized, with the exception of Google, a number of other popular commercial search engines. The number of retrieved social media results were almost the same across all search engines, which is one more sign of the social media ubiquity, not just on Google but in practically all other search engines. The principal difference when comparing alternative search engine results relates to the fact that Bing and Yahoo returned a more diversified set of results. Conversely, Google and Ask were not as pluralistic in terms of the number of alternative sites generating the social media results.

The second sub-study involved the application of the methodology to searches that would return results exclusively in the local language, Greek or Finnish. The conclusion is that social media content appears to be much more prevalent in English. Finnish provides a fair number of social media results, considering also the relatively small number of population using the language. On the other hand, Greek seems to be deprived almost completely of travel social media content, which could reflect the relatively limited (at least by European standards) use of the web in general by the Greek population. By contrast, the more active online part of the population might be more inclined to consume or produce content in English.

The third sub-study simulated the methodology of Xiang and Gretzel’s research, by testing out the generic type of query “hotels + {destination}”, just to provide one more piece of evidence that generic keywords tend to return a substantially lower amount of social media content.

The fourth sub-study analyzed search pages beyond the first three. Even though there was a degree of fluctuation regarding the number of results which were found following the third page, the size of the sample does not suffice to draw general conclusions. Overall social media demonstrated a strong presence throughout the page sample. One assumption that can be made based on this analysis is that social media content of the infrequent types (based on the results of the main study) such as social networking sites and most notably media sharing sites, seem to emerge in higher frequencies in deeper pages. In any case the findings of this last study are rather theoretical given that very seldom do users take the time to search deeper than the third search results page.

8.2 Hotels, social media sites and destinations that stood out

According to the results of this experimental study, a number of hotels, web sites and destinations emerged that could be considered as the social media “champions”. It is thus worth to examine more thoroughly this area and discuss implications.
8.2.1 Social media websites

With respect to the key websites identified through the analysis process, four are the ones standing out, with a combined share of social media contribution exceeding 50% of the total. These four sites are in order of contribution: TripAdvisor, Real Travel, Wikio and Virtual Tourist. All of them have consistently had a very strong presence, independently of destination. The one site, however, that has by far been the most frequently displayed in the search results is TripAdvisor. The social media site not only has captured a striking 29% of all the social media results, but it is often displayed multiple times in the same search page (often the first one). It would therefore be worth to discuss further this website and its significance for travel marketing.

Even though search engines do not offer the tools for search ranking evaluation, it is possible to examine some of the critical factors that may impact the ranking of a site, based on the content of the earlier chapter about search engines. These factors include Page Rank score and the number of inbound links.

Another, more practical way to examine the link popularity of TripAdvisor in the search results is through a basic Webometric analysis. Webometrics is the research field that studies the hyperlink structure of the Web and attempts to create new knowledge from hidden information embedded in it (Holmberg, 2009).

The tool deployed for this analysis was Yahoo Site Explorer. By entering a website URL into the tool, it returns the list of indexed sites that provide outbound links to the provided website, as well as a figure that represents the total number of sites supplementing the links. As discussed in a previous chapter the number of links that point to a website (in combination with their quality, which however cannot be assessed accurately in the context of this study) can impact the way the given site ranks in the search engine results. For the purpose of this study some of the top social media websites were examined with the Site Explorer tool.

The results show that TripAdvisor has 14,100,918 inbound links, Wikio 3,672,633 and Virtual Tourist 3,550,749 (excluding links that originate from within the websites themselves). This result is an indication of the search power that TripAdvisor has when compared to the other top websites. This is also reflected on the high Page Rank score that the travel website has achieved (8 out of 10).

Additional reason for TripAdvisor’s high visibility is the fact that it has been long established (since 2000) and it is considered that search engines value pages with an established history, as discussed in the previous chapter. Moreover TripAdvisor shares its review databases with several other sites and this naturally creates a flow of inbound links from sites which are both relevant and popular. There have been more studies suggesting the strong position of TripAdvisor in search results. Search
Engine Marketing agency Greenlight reported that TripAdvisor was the most effective site in the hotel sector in terms of its natural search visibility (comScore Media Metrix, 2010).

8.2.2 Additional analysis on TripAdvisor
As already noted TripAdvisor is primarily a consumer review site, facilitating the independent reviewing of hotels and other travel services in numerous destinations around the world. Its brand is for many travelers synonymous with hotel reviews. It has greatly benefited from network effects since travelers wish to write reviews where the most readers are, and travelers also prefer to read reviews where the highest variety is present. The group of TripAdvisor branded sites receives over 30 million unique visitors per month and contain over 30 million reviews on more than 500,000 hotels and attractions around the world, making it the largest travel community on the web at present (TripAdvisor, 2010). Besides its principal function, the website has also incorporated several Web 2.0 features transforming it into a social media rich website. TripAdvisor enables users to interact with each other in theme-specific discussion forums, create friends lists, contribute to topical Wikis, upload media content or browse mashups with Google maps.

Even though anyone can post reviews for any hotel, TripAdvisor claims that all reviews are screened and special algorithms are used to detect possible cases of abuse. The fact is that TripAdvisor offers a very comprehensive database with reviews and ratings. Considering the high visibility it enjoys on search engines, it makes it an imperative for hotel marketers to pay special attention to it and ensure that it has a positive impact for their hotel business.

This could be achieved by performing a few simple tasks, such as ensuring that all the hotel information listed on the site is accurate and up-to-date, improving the hotel’s appeal by posting videos and images of it as well as monitoring closely the incoming user reviews, in order to address potential issues and receive constructive feedback. Overall the importance of receiving positive reviews by real travelers cannot be underestimated as it can often be much more effective than any other paid marketing initiative in the eyes of the consumers.

To place additional stress in its importance, the same company which TripAdvisor belongs to, the Expedia group, also owns Virtual Tourist, Travel Pod and Travel Library within a portfolio that includes a total of 15 popular Internet travel brands. In this study TripAdvisor, Virtual Tourist and Travel Pod hold approximately a combined 40% share of the total number of social media results. It is worth to note that TripAdvisor operates a number of local websites targeting over 15 major travel markets.
8.2.3 Hotels
Another basic component of the analysis relates to the specific hotels whose names were searched for on Google and the other search engines. Even though to a higher or lower degree all hotels demonstrated a strong link with social media sites, there were a number of them that stood by far beyond the average. The following four Finnish hotels were on the top of the list: Omena hotelli, Hotelli Kauppi, Park Hotel, Scandic Julia.

Most notably it was Omena hotel, the one to stand out, considering that almost all the search results returned were references to social media sites (27 out of 30). Omena is a hotel chain operating in four cities in Finland. When examining the social media results which search for Omena hotel has generated, one can notice that there is a high degree of pluralism. Social media content from many of the major contributor sites are present, together also with some local Finnish sources. There is a lack of commercial results and the reason for this could be that the hotel seems to have completely opted out of these channels of distribution.

Even though it is not easy to conclude which are the reasons travelers are keen on discussing about Omena hotel online, it is quite likely that the reason behind the “buzz” is the distinct business and operational model that the hotel has adopted. Omena business model is based 100% on digital interaction between the enterprise and its customers, from reservation to check out. This is also combined with the low level of rates that as claimed on the website are significantly lower than the general rate level, thanks to their “no-frills” policy and the fact that no human personnel needs to be employed.

8.2.4 Destinations
With regard to the destinations, the Finnish city of Turku is the “champion” topping the list with 129 out of the 300 results related to social media. Most of the results are associated with the category of consumer review sites, but what actually brings Turku to the top is the high amount of blog type results in comparison with the other cities. Even though the analysis itself cannot validate any hypothesis about the reasons why some particular cities trigger more social media results than others, a number of assumptions can be made.

It seems that Turku (as well as Tampere which follows in the list with just one social media result less) is the city whose size and travel industry dynamics provide the highest ratio between social media and commercial or other non-social media results. It seems that the international touristic hubs (Athens, Helsinki, Heraklion) when searched for hotels and accommodations they tend to return an overwhelming amount of commercial results coming from sites such as Expedia, Booking etc., due to their popularity and high commercial value in this respect.
By contrast, one can suggest that smaller destinations where traffic is limited would naturally not generate many commercial results. The same would hold true for social media results since the “buzz” around these places would be relatively low. Therefore there will not be numerous reviews and ratings for local hotels and possibly those destinations would not be included in some of the social media sites whatsoever. It appears that destinations like Turku are able to “gain” from the fact that they are sizable enough to generate interest, but still do not fall under the tourist “hotspot”, that would cause inevitably a high number of commercial results competing with social media for the top positions in search engines.

8.3 Relating to previous research

The current study has its foundations in the work of Xiang and Gretzel (2009), which is also the research to be compared to, with reference to the drawn conclusions, since there are no other known studies in this particular area. The main difference between the two studies is that Xiang and Gretzel’s approach is in a sense more theoretical, since the main objective is to evaluate the role of social media and search engines in representing the tourism domain as a whole.

This is possibly the reason why Xiang and Gretzel (2009) preferred to use generic search terms and looked deeper into the results, up to the tenth search page. In both studies the general outcome is that social media are ubiquitous and have a strong presence in the search results pages. It emerges, however, that in Xiang and Gretzel’s research, the proportion of social media is substantially lower compared with the finding of the presence study, where brand/name-specific terms were used and the study was focused on the hotel industry.

In Xiang and Gretzel’s research it was approximately 11% of the search results representing social media, whereas in the present study the rate is over two times as much, at 27%. It is interesting to note that for the term “hotel + destination” the results from Xiang and Gretzel’s study were even lower in terms of social media contained, at just 7% and for accommodation at 8%. Common in both studies was the relative homogeneity in the distribution of social media across the search results pages.

A key contrast relates to the social media types and specific sites that contributed the most social media results. Xing and Gretzel’s study identified virtual community sites as the ones occurring with the highest frequency (40%). In the present research, consumer review sites were by far the most frequent ones (57.1%). The divergences in composition of the social media results as regards the social media types involved is illustrated in more detail in Table 8.3.1.
Tripadvisor in both studies was the top social media results generator, but in Xiang and Gretzel’s study the site contributed just a fraction of what it did in the present study (8.3% to be compared with 27.7 %). The rest of the lists containing the key social media sites, despite some common ground with sites like Virtual Tourist and Real Travel, are quite diverse as illustrated in Table 8.3.1. This can be explained by the fact that the two studies used dissimilar keywords on separate geo-locations and in a different point in time.

Generally there seems to be a much higher concentration of social media results across a limited number of key sites on the current study, even though this is to some extent the case also in Xiang and Gretzel’s study, where a head and long tail pattern was also identified. A common conclusion in both studies is that media sharing sites are not so well represented in the sample, even more so in the current study. The other two minor categories Other (OT) and social networking sites have also gathered lower numbers in the present study, as indicated in Table 8.3.1.

Overall the current study takes the research a step further by showing that the number of social media results highly depends on the nature of the travel keywords, i.e. whether they are generic or brand specific. By taking a brand-specific approach in this study it was possible to induce that the relative importance of social media in travel is significantly higher compared to what Xiang and Gretzel had proposed. The current study also proposed new assumptions about the role of social media based on additional factors including language, location, hotel type and search engine used.

<table>
<thead>
<tr>
<th>Site (Xiang and Gretzel’s study)</th>
<th>Cumulative %</th>
<th>Site (present study)</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tripadvisor</td>
<td>8.30%</td>
<td>Tripadvisor</td>
<td>28.7%</td>
</tr>
<tr>
<td>Virtual Tourist</td>
<td>15.10%</td>
<td>Real travel</td>
<td>38.2%</td>
</tr>
<tr>
<td>igougo</td>
<td>20.20%</td>
<td>Wikio</td>
<td>45.5%</td>
</tr>
<tr>
<td>Mytravelguide</td>
<td>24.90%</td>
<td>Virtual Tourist</td>
<td>51.6%</td>
</tr>
<tr>
<td>Yelp</td>
<td>28.10%</td>
<td>TravBuddy</td>
<td>57.5%</td>
</tr>
<tr>
<td>Meetup</td>
<td>30.70%</td>
<td>Travel Yahoo</td>
<td>62.4%</td>
</tr>
<tr>
<td>Travelpost</td>
<td>33.00%</td>
<td>TravelPod</td>
<td>66.5%</td>
</tr>
<tr>
<td>Insiderpages</td>
<td>35.10%</td>
<td>Lonely Planet</td>
<td>69.5%</td>
</tr>
<tr>
<td>Associatedcontent</td>
<td>36.60%</td>
<td>Ciao</td>
<td>70.9%</td>
</tr>
<tr>
<td>Yellowbot</td>
<td>38.10%</td>
<td>Hotel Chatter</td>
<td>72.1%</td>
</tr>
</tbody>
</table>

*Table 8.3.1 Comparison of top SM sites between Xiang and Gretzel’s and present study*
8.4 Recommendations for marketers

All the findings and conclusions detailed in the previous section of this chapter provide several insights to the travel marketers. This is to assist them in better evaluating how the advent and prominence of social media on the web can have an impact on their digital marketing strategy. Travel companies that understand how to use social media services as part of a broader marketing strategy can gain an advantage over their competition. This can bring some immediate benefits in terms of market exposure now and it can also bring the significant advantage of improved search engine rankings in the future.

The core message is that, given the before mentioned attributes, social media cannot be ignored nor can they be controlled by the industry players. The fact that social media are so ubiquitous in the search results means that this phenomenon is having a real impact on hotel business. Being aware of information about the way social media results are distributed across the search pages, knowing the hotels which have adopted best practices and understanding the key trends and relationships between size of cities, languages, locations, search engines and associated amount of social media content can help marketers work on the right strategies and tactics.

Overall this study recommends marketers to consider social media as a key component of the integrated marketing communications mix, in line with Mangold et al. (2009) suggestion. Marketers therefore are only left with the option of participating in social media, engaging with their users and in this way attempt to bring a positive influence in their business as well as build a healthy relationship with current and potential customers. In fact, content on the key travel social media is indexed by the search engines regularly because of their authority in the travel and tourism domain. This means that any user contribution about an experience, positive or negative, with any tourism related business will be indexed relatively quickly.

Most hotels and other travel businesses will not have the sources to outrank the top social media sites in the search engines, meaning that these user reviews, advice and opinions will typically appear earlier in the search engine results and will be found by more people.

The real estate of the first page or pages on Google and the other search engines is limited yet very valuable for any travel business. Travel marketers can start by analyzing the given search space, its components and how each of them gets ranked or featured. There are primarily three ways to achieve high visibility, by using three types of media: Owned, bought or earned media (Erkkola, 2010) which within a hotel
business context would correspond to a hotel’s website (and possibly other proprietary microsites), online advertisements, and social media that feature the hotel respectively.

The three concepts described above can be transferred effectively to a web search context where owned media would mean the hotel website, bought media would correspond to the sponsored links, featured alongside the search results and earned media would represent the social media results, as illustrated in figure 8.4.1. These media types, which are discussed more thoroughly below, match the three ways marketers have at their disposal to achieve search engine visibility. All of them can be deployed in the media mix for optimal results. A number of different strategies and tactics which can assist in achieving this are represented in figure 8.4.2 and analyzed in detail in the following sections.

Figure 8.4.1 An example of bought, owned and earned media on Google

8.4.1 Owned media
The importance of a hotel’s official web presence is naturally very high, since this is the way for the company to directly communicate with customers and control the content provided to a maximum degree. This is typically where the most profitable reservations are made. In addition, the travel marketers can avail of a wealth of web traffic data from the visitors that can be used to better understand the customers and their behavior on the site, including their interests, demographic information, ways of navigating the site, making reservations, etc.
In order to maximize the benefits from the owned website hotel marketers can experiment with integrating Web 2.0 and social media features such as forums, chat, image, video sharing applications, even establishing an exclusive social network for their most loyal customers. These technologies are not only beneficial for the reasons described above, but also relatively easy to implement and monitor. An example could be the addition of reviews to the site which can have a positive impact in conversions, since consumers are willing to pay a considerably higher rate for better rated hotels (Mashable, 2010). They can also be used in order to augment the content of the website and the update frequency, which can in turn have a positive impact on search rankings.

Beyond this, there are several other techniques available mainly classified under the label of search engine optimization which are critical in terms of helping boost the performance of the website in search rankings. Other techniques focus on optimizing metatags and description tags of the html pages, in order to display a persuasive message when displayed as part of the search result entry, in the form of a text snippet. Search engine optimization is usually carried out by professionals and specialized agencies, or even in-house (particularly for larger enterprises).

8.4.2 Bought media
In the context of web search, bought media take the form of sponsored text ads which are placed alongside the search results. This is an additional way for hotels and travel companies to be featured prominently in the search results pages, even though this comes at a cost (cost per click or cost per thousand impressions). The benefit is the flexibility that the sponsored links programs offer, which enables marketers to choose the exact keywords that will have their ad triggered. Additionally, marketers can easily make changes and customizations to the ad text. By taking part in search engine marketing programs, as they are called, such as Google AdWords or Yahoo’s Search Marketing, advertisers can often feature their ads “above the fold” i.e. in a space which is above the natural search results, thus gaining critical visibility.

8.4.3 Earned media
These typically correspond to social media and in general constitute references to the brand which are neither owned nor paid for. Instead they are a result of consumer interaction with the brand, which motivates consumers to discuss their experience with it and share it with their peers. Earned social media could be divided in two types, the ones where the brand can exercise some control over and the ones where the brand has virtually no control of.

Example of social media that the brands cannot control are some of the key consumer review sites included in the study, such as Tripadvisor and Virtual Tourist. These are websites whose content is created entirely by users/travelers and therefore the only
actions that a hotel business can take is to monitor the activity and make sure to follow all constructive feedback (and only very selectively intervene in the event of a malicious review being posted or some inaccuracy mentioned). In other words the role of the hotel business is more reactive and the fact is that regardless of whether the hotel marketer takes any action about it or not, the entries referring to the hotel from sites such as TripAdvisor will be made anyway.

In this study the vast majority of social media results were completely user-generated, meaning that the associated hotels didn’t have to take any initiative on their own. In order to counter-balance this “passive” role, marketers have the option to advertise on social media. This can be achieved in a number of ways, for example by displaying banner ads, deploying social gadgets, sponsoring different sections of the website, or organizing co-promotions and co-marketing initiatives together with the social media website. As research has shown, when a brand is exposed on a social media website, chances are higher that the user will have better brand-recall and more likely to click on a relevant result after a search (Copeland, 2009).

There is also the type of social media that even though they cannot be completely controlled by the travel business, it is rather up to the hotel marketer to take the initiative and create this social media presence. Afterwards content can be created there by both the enterprise and the consumers. This is in fact a method of creating social media that has proven to be utilized in a very limited way, based on the results of the study. This can interpret why some of the social media types of the study received very low scores, such as social networking and media sharing sites.

Examples of how a marketer can initiate social media actions include: The creation of a Facebook fun page with frequent updates, offers etc. (users can become fans and receive all this info in their feeds stream). Creation of a Twitter account where followers can receive up-to-the-minute updates about relevant travel info, events adjacent to the hotel areas, twitter exclusive offers on unfilled room inventory and more. These messages can then be re-tweeted multiple times and thus, thanks to the power of word of mouth, the original message can achieve a broad reach. At the same time there is possibility to interact real-time with customers, by soliciting their opinion or receiving their feedback, requests etc.

Another worthy initiative could be the creation of a dedicated YouTube Channel, where videos will be posted featuring the hotel itself, as well as highlighting the nearby attractions or festivals, events etc. which can supplement the hotel product or videos including real customers in real situations that would add credibility to the video. Often the creation of a video that can “go viral” can have impact that no other paid initiative or bought medium can bring.

Other initiatives could include the creation of a virtual property in a virtual world
such as Second Life, or the creation of a regularly updated blog that could also serve as a venue for conversation with consumers. At the same time hotel marketers can also participate in other individual or corporate blogs, Facebook pages, discussion groups etc. in order to build a genuine presence, as well as a circle of friends and followers. All these initiatives that are relatively easy to implement and no specialized technical knowledge is required to make them work effectively. In fact these initiatives frequently result in enhanced linkability which in turn helps rankings on search engines.

This could be an unexploited opportunity for many of the hotels featured in the study, as they could have more social media, partly controlled by them, displaying on the search results, rather than additional commercial entries from sites such as Expedia and Booking. This is in effect what would represent the proactive approach to social media, as opposed to the passive or reactive one that was described earlier and which seems to be the norm as far as the majority of hotels in the study.

It is critical for all travel businesses to experiment first, then explore the social media arena and follow up with social media optimization actions, until they find the niche that will enable them to most effectively harness the power of social media for their type of business and target market. In addition it is also crucial to monitor effectively the social media activity, which can be achieved by implementing social media analytics solutions that several market research firms offer. Moreover data and text mining techniques can also be employed in order to provide useful insights, such as sentiment analysis, based on social media activity on the Web. Monitoring activity especially from travel related social media such as TripIT or Dopplr, often sent by means of feeds to other more mainstream applications, such as Facebook and LinkedIn, can provide useful information about the intentions of perspective customers (given the open APIs most of these applications have, hotels can pull the data submitted and build custom applications using it).
Figure 8.4.2 Social media strategy & tactics tree
In addition to its usefulness for individual hotel marketers this study also provides practical knowledge for a number of other industry professionals. Destination management organizations will find the analysis and conclusions of this study useful, given that it provides a high level overview of how social media can impact not just individual hotels but also the whole range of the destinations involved. Moreover advertising agency professionals, especially the ones who are specialized and work with travel clients, will also be able to extract useful pieces of information from the report, given that the findings and recommendations can apply to the hotel industry as a whole.

Similarly, professionals working in local or national hotels, travel associations and industry chambers will be able to gain an improved understanding of the impact of social media in their business. They could later on be the agents of change in promoting the message to their respective associations and helping their members benefit from the new opportunities that are arising.

8.5 Search engine evaluation

Search engines have made tremendous progress in their accuracy, relevancy and user friendliness in recent years. This study, however, has caused a number of their weak points to emerge.

The most critical is the lack of pluralism that has been observed as far as the results provided by Google and other commercial search engines are concerned. A small number of websites (both of commercial and social media nature) dominate the search results, often without adding any value to the searchers (for example when multiple language versions of a website are included in the results, referring to the same hotel). Search engines can be considered as the filter through which searchers access the online travel domain. However, in many ways the returned received are far from comprehensive. It is indeed challenging for the majority of other websites to compete against the small group of established travel sites containing millions of reviews and having thousands of inbound links pointing at them. The chances of a new travel website to gain some visibility on search results would be extremely low given the current status quo.

This issue is expected to last and deteriorate as the amount of information that becomes available in a multitude of types and formats is on the rise. Prioritizing and making the search results relevant is a serious issue. Another challenge is related to offering users the possibility to determine by their own behavioral signals the popularity of web pages and thus the ranking on web pages, which would then act as a vote of confidence from the users end.

Another possibility could be the emergence of travel vertical search, that would enable more accurate prioritization of results and sources, exclusively for the hotel
industry. Optimally this would be accomplished by making the whole process more social, for instance by taking into consideration friends’ travel experiences as well as the searcher’s own personal preferences. These are ideas that are not brand new, but there has been no known application incorporating some or all of the above features, to gain momentum as of yet.

8.6 Assessment of research: limitations & future research

The results of this study are based on the evaluation of over 5000 search results. This has been a catalyst, having enabled the collection and analysis of a large amount of data based on queries in multiple languages, locations and search engines.

The primary conclusions of the study could be characterized as describing the overall social media landscape in the researched field. Key conclusions include: the substantial proportion of social media sites in the search results, the dominance of a relatively small group of social media sites, the homogenous distribution across the first three pages and the absence of some popular types of social media.

The analysis was based on a comprehensive manipulation of pivot tables, practically covering all possible combinations between the variables, in search of meaningful correlations. Additionally, data mining tools were used to examine hidden patterns. The assumptions induced from the sub-studies are surely not sufficient to be generalized, but they do provide an interesting hypothesis which unquestionably needs to be tested by future research.

This study has attempted to cover some of the limitations of previous relevant studies, by examining more specific keyword terms, in multiple languages, locations and search engines. There were certainly a number of limitations and challenges on the way as well as some new questions which arose. A serious challenge was the task of coordinating the data coming from the evaluation performed by the students. The data had to be double-checked and cleaned to avoid instances of misclassification wherever possible. There were certain cases, however, where results were ambiguous or border-line and therefore the classification had to be based on subjective criteria.

Another unavoidable limitation pertains to the way search engines, particularly Google, generate their results. To an extent the displayed results are linked to the personal preferences and history of the searcher. In the past this was possible to overcome by signing out of a Google account, but in 2009 Google started providing personalized results for users regardless of whether they sign in or not. In any case, the impact of this, even though hard to quantify, is not expected to have played a significant role, given the relatively neutral nature of the search terms.
The choice of hotels was based on Google maps results. Therefore, had an alternative set of hotels been used in the study, results could have been diversified. Even though the study was based on more than one language, location and search engine, there are obviously many other combinations of locations and languages that could be tested by future researchers. It would be beneficial to test alternative keywords too such as “accommodations”, “bed and breakfast”, “lodgings” etc. that are also used widely as synonyms to hotels. This research used multiple search engines, but their use was limited, so future research can attempt to more effectively incorporate the study of additional search engines, for a more representative results range.

Another idea would be to incorporate a behavioral component in the research by attempting not only to evaluate the various results that are displayed on the search engine pages, but also to study the way a group of subjects would interact with these results i.e. how eyes move on the search pages so as to create heat maps, how the mouse moves and where it clicks etc., to achieve a better understanding of the significance of ranking and snippet text of the search results.

It would also be of interest to repeat the study for travel products and services other than hotels. Finally, this study was carried out at a given point in time and considering the rapid pace with which the web is evolving, it is an imperative for marketers and researchers to keep monitoring continuously the way social media content evolves and how it impacts the travel industry.

8.7 Future web, social media and tourism

This study has focused on the intersection of travel, search and social media, which are all fields critically affected by the development of new technologies. The study provides a good understanding of these three areas in their current state for the travel marketers.

Nevertheless, in order to stay at the cutting edge and benefit the most from the new applications of technology, marketers are recommended to research what is expected to emerge in the short and long term future and how it would affect the current state of things in the field under discussion.

In the area of search a number of innovations are under way. Social search promises to return and prioritize search results coming from sources that belong to the user’s social graph, thus making results highly relevant to one’s social context (Heymans, 2009). A travel query about a destination could trigger results from the blog or status update of a friend who has already visited the location.

Another innovation is real-time search that captures content which is extremely up-to-date, such as Facebook and Twitter status updates. Providing time-sensitive
information could surely be of good use for travelers while being on a trip (Singhal, 2009).

At the same time, the introduction of rich snippets signifies the advent of semantically marked (in RDF language) content that enables search engines to incorporate a social component in their results by integrating user generated ratings of hotels, restaurants etc. within the text snippet that appears on the search pages (Goel et al., 2009). As such, these new applications have the potential to expand the horizon for developing search technologies and to improve the organization and representation of the tourism product.

In the field of social media there is high amount of activity too, with new social media applications emerging on a frequent basis. Even if not all will finally become sustainable, there are several promising applications currently making their way into the market.

By many accounts the most promising new applications are the ones supporting location-sharing functionalities (prominent examples include Foursquare and Gowalla). These applications are offering a new dimension to social networking by integrating it with real life interaction, giving incentives to the rapidly growing user base.

Location-aware apps are naturally very critical for the travel industry and the combination with the social networking concept could provide an invaluable source of information about customer profiles and their travel habits. At the same time it can offer incentives (such as game points, special offers, opportunities to meet old or new friends, opportunities to engage in interesting activities) for additional stays or other use of services in a hotel business.

In the area of travel, a new, highly anticipated evolution is the one related to the advent of Tourism 3.0, i.e. the outcome of the combination of travel with Web 3.0 (also known as Semantic Web) technology. Web 3.0 pre-supposes that the content found on the web gradually becomes more structured, acquiring a machine readable format that will enable computers and software to understand the real meaning (or semantics) of it, based on metadata descriptions.

In fact, there are currently several research initiatives trying to make compatible the usability and flexibility of the user-generated tagging and folksonomy data with the precision of semantic ontologies (Herzog et al., 2007), practically attempting to unite the Social with the Semantic Web. In the future, Web 3.0 with intelligent travel agent technology could deliver rapid, comprehensive answers and complete tasks on behalf of the traveler, relating to many aspects of travel research and booking process.
REFERENCES


Erkkola, J. (2010). *Social media at nokia, presentation in Åbo Akademi, 27/272010*


## APPENDIX A: Types of web 2.0 marketing initiatives hoteliers worldwide plan to implement

<table>
<thead>
<tr>
<th>Initiative</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create profiles for my hotel(s) on social networks (Facebook, Twitter, Flickr, etc.)</td>
<td>13.30%</td>
<td>14.00%</td>
<td>50.00%</td>
</tr>
<tr>
<td>Create and post videos on YouTube</td>
<td>na</td>
<td>na</td>
<td>46.60%</td>
</tr>
<tr>
<td>Advertise on social media sites (e.g., TripAdvisor, Facebook, etc.)</td>
<td>8.10%</td>
<td>15.10%</td>
<td>39.70%</td>
</tr>
<tr>
<td>A blog on the hotel Website</td>
<td>14.50%</td>
<td>14.00%</td>
<td>37.90%</td>
</tr>
<tr>
<td>Sweepstakes and contests on the hotel Website</td>
<td>9.00%</td>
<td>3.50%</td>
<td>36.20%</td>
</tr>
<tr>
<td>A photo-sharing functionality on the hotel Website</td>
<td>12.70%</td>
<td>4.70%</td>
<td>32.80%</td>
</tr>
<tr>
<td>Survey and comment card on the hotel Website</td>
<td>18.40%</td>
<td>14.00%</td>
<td>31.00%</td>
</tr>
<tr>
<td>Share this site' and RSS on the Website</td>
<td>na</td>
<td>na</td>
<td>24.10%</td>
</tr>
<tr>
<td>Actively participate in blogs that concern my hotel</td>
<td>12.70%</td>
<td>5.80%</td>
<td>24.10%</td>
</tr>
<tr>
<td>Subscribe to a reputation monitoring service</td>
<td>8.40%</td>
<td>2.30%</td>
<td>19.00%</td>
</tr>
<tr>
<td>I am not planning on Web 2.0 and social media initiatives for 2010</td>
<td>na</td>
<td>15.10%</td>
<td>6.90%</td>
</tr>
</tbody>
</table>

Source: Hospitality eBusiness Strategies (HeBS) as cited in company blog, March 10, 2010
APPENDIX B: US natural search visits by search engine results page

<table>
<thead>
<tr>
<th></th>
<th>Page 1</th>
<th>Page 2</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>95.80%</td>
<td>2.50%</td>
<td>1.70%</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>95.20%</td>
<td>2.80%</td>
<td>1.90%</td>
</tr>
<tr>
<td>Bing</td>
<td>95.00%</td>
<td>3.40%</td>
<td>1.60%</td>
</tr>
</tbody>
</table>

Source: iCrossing, "The Importance of Page-One Visibility," February 11, 2010

APPENDIX C: Technology that will have the greatest impact on the way the next generation researches travel

<table>
<thead>
<tr>
<th>Technology</th>
<th>% of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social networking (Facebook, etc.)</td>
<td>47.50%</td>
</tr>
<tr>
<td>User-generated reviews</td>
<td>19.80%</td>
</tr>
<tr>
<td>Mobile devices</td>
<td>16.70%</td>
</tr>
<tr>
<td>Video-sharing (YouTube)</td>
<td>9.80%</td>
</tr>
<tr>
<td>Visualization tools</td>
<td>6.20%</td>
</tr>
</tbody>
</table>

Note: n=786
### APPENDIX D: Top 10 search engines in the US

**Top 10 Search Engines in the US, Ranked by Searches, January 2010**

<table>
<thead>
<tr>
<th></th>
<th>Searches (millions)</th>
<th>% share</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Google</td>
<td>6,805.40</td>
<td>66.30%</td>
</tr>
<tr>
<td>2. Yahoo!</td>
<td>1,488.50</td>
<td>14.50%</td>
</tr>
<tr>
<td>3. Bing</td>
<td>1,116.50</td>
<td>10.90%</td>
</tr>
<tr>
<td>4. AOL</td>
<td>251.8</td>
<td>2.50%</td>
</tr>
<tr>
<td>5. Ask.com</td>
<td>194.2</td>
<td>1.90%</td>
</tr>
<tr>
<td>6. My Web</td>
<td>112.4</td>
<td>1.10%</td>
</tr>
<tr>
<td>7. Comcast</td>
<td>59.6</td>
<td>0.60%</td>
</tr>
<tr>
<td>8. Yellow Pages</td>
<td>35.1</td>
<td>0.30%</td>
</tr>
<tr>
<td>9. NexTag</td>
<td>34.7</td>
<td>0.30%</td>
</tr>
<tr>
<td>10. BizRate</td>
<td>20.1</td>
<td>0.20%</td>
</tr>
</tbody>
</table>

Source: Nielsen Online, "MegaView Search" as cited in the nielsenwire blog, February 16, 2010

### APPENDIX E: Top 10 upstream search engines visited by US internet users before visiting travel websites

**Top 10 Upstream Search Engines Visited by US Internet Users Before Visiting Travel Websites, January 2010 (% of total visits)**

<table>
<thead>
<tr>
<th></th>
<th>% of total visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Google</td>
<td>29.50%</td>
</tr>
<tr>
<td>2. Yahoo! Search</td>
<td>4.10%</td>
</tr>
<tr>
<td>3. Bing</td>
<td>2.89%</td>
</tr>
<tr>
<td>4. Ask.com</td>
<td>0.51%</td>
</tr>
<tr>
<td>5. Google Images</td>
<td>0.48%</td>
</tr>
<tr>
<td>6. AOL Search</td>
<td>0.36%</td>
</tr>
<tr>
<td>7. Google Canada</td>
<td>0.09%</td>
</tr>
<tr>
<td>8. MyWebSearch</td>
<td>0.08%</td>
</tr>
<tr>
<td>9. Yahoo! Everything</td>
<td>0.08%</td>
</tr>
<tr>
<td>10. Dogpile</td>
<td>0.08%</td>
</tr>
</tbody>
</table>

### APPENDIX F: Metrics used to monitor and measure the impact of social media according to US marketers

<table>
<thead>
<tr>
<th>Metric</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visitors and sources of traffic</td>
<td>72%</td>
</tr>
<tr>
<td>Network site in terms of followers, fans, members, etc.</td>
<td>63%</td>
</tr>
<tr>
<td>Quantity of commentary about your brand or product</td>
<td>56%</td>
</tr>
<tr>
<td>Sentiment or quality of commentary about your brand or product</td>
<td>50%</td>
</tr>
<tr>
<td>Search engine ranking position</td>
<td>48%</td>
</tr>
<tr>
<td>Leads generated</td>
<td>48%</td>
</tr>
<tr>
<td>Progress toward achieving your social media objectives</td>
<td>38%</td>
</tr>
<tr>
<td>Engagement with influential bloggers, journalists, Twitterers, etc.</td>
<td>38%</td>
</tr>
<tr>
<td>Sales conversions or other ROI metrics</td>
<td>35%</td>
</tr>
<tr>
<td>Competitive share of social media coverage</td>
<td>19%</td>
</tr>
<tr>
<td>Criteria to identify and profile audiences</td>
<td>17%</td>
</tr>
</tbody>
</table>

Note: n=2,317

APPENDIX G: Google Docs interface where students entered and coded the collected data

<table>
<thead>
<tr>
<th>Query</th>
<th>In hotel an Int'l chain (L) or local (R)</th>
<th>Search Page no</th>
<th>URL</th>
<th>Social Media? (1=Yes 0=No)</th>
<th>Name of site?</th>
<th>Social Media Type?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sokos Hotel City Birs</td>
<td>L</td>
<td>1</td>
<td><a href="http://www.hotels.com/hotel-rate/">http://www.hotels.com/hotel-rate/</a></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sokos Hotel City Birs</td>
<td>L</td>
<td>1</td>
<td><a href="http://www.hotelscombined.com">http://www.hotelscombined.com</a></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sokos Hotel City Birs</td>
<td>L</td>
<td>1</td>
<td><a href="http://www.hotelscombined.com/hotel-rate/">http://www.hotelscombined.com/hotel-rate/</a></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sokos Hotel City Birs</td>
<td>L</td>
<td>1</td>
<td><a href="http://www.hotelscombined.com/hotel-rate/">http://www.hotelscombined.com/hotel-rate/</a></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sokos Hotel City Birs</td>
<td>L</td>
<td>1</td>
<td><a href="http://www.travel.com/hotel-rate/">http://www.travel.com/hotel-rate/</a></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sokos Hotel City Birs</td>
<td>L</td>
<td>1</td>
<td><a href="http://www.maplandia.com/finnland/varkaus/hotels/sokos-hotel-hamburg-square-city-borns/">http://www.maplandia.com/finnland/varkaus/hotels/sokos-hotel-hamburg-square-city-borns/</a></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sokos Hotel City Birs</td>
<td>L</td>
<td>1</td>
<td><a href="http://www.hoteltarvel.com%E8%8A%AC%E5%85%B0">http://www.hoteltarvel.com芬兰</a></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sokos Hotel City Birs</td>
<td>L</td>
<td>1</td>
<td><a href="http://www.tripbuddy.com/sokos-hotel-city-birs/Sokos-Hotel-City-Birs-Turku-54689">http://www.tripbuddy.com/sokos-hotel-city-birs/Sokos-Hotel-City-Birs-Turku-54689</a></td>
<td>1</td>
<td>TrueBuddy</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E: Web video created to provide step by step instructions for the students