

Mohammed Hamood Jabreel

AUTOMATIC ANALYSIS OF THE COMMUNICATION OF DESTINATION BRANDS THROUGH SOCIAL MEDIA

Master Thesis

Supervised By:

Dr. Antonio Moreno

School of Engineering (ETSE)

Department of Computer Science and Mathematics



Thesis submitted at the University Rovira i Virgili
in partial fulfilment of the requirements for the degree of Master on
Computer Engineering: Computer Security and Intelligent Systems

September 2015

Abstract

Destination branding has improved the management and communication of tourist destinations. Nowadays most destinations try to transmit their brands through social media and they need to evaluate their communication. This document presents a methodological framework for the automatic analysis of the communication of emotional brand values through social media using syntactic and semantic content analysis methods. The analysis of the tweets sent by the tourism offices of 12 major European destinations shows a general lack of a strategy in the definition and communication of a differentiated personality.

Acknowledgements

I would like to express my gratitude to the director of this thesis, Dr. Antonio Moreno for its guidance and advice during the elaboration of this dissertation as well as to Dra. Asunción Huertas who collaborated with us in this work.

I would also want to acknowledge the support of Hodeidah University for covering my enrolment in the Master.

A special thanks to my family and to my wife for encouraging me in the stressful moments.

Contents

CHAPTER 1: INTRODUCTION	13
1.1 Objectives	15
1.2 Overview of the methodology	16
1.3 Document structure	17
CHAPTER 2: BACKGROUND	19
2.1 Twitter	19
2.2 Knowledge repositories	20
2.2.1 Ontology basics	21
2.2.2 WordNet	25
2.3 Techniques	27
2.3.1 Preprocessing	27
2.3.2 Natural Language processing	28
2.3.2.1 Natural Language Processing parser	29
2.3.3 Ontology-based semantic similarity	31
2.4 Summary	35
CHAPTER 3: STATE OF THE ART	37
3.1 Numerical Analysis Methods	37
3.2 Content Analysis Methods	38
3.2.1 Qualitative Methods	38
3.2.2 Quantitative Methods	39
3.3 Summary	41
CHAPTER 4: METHODOLOGY	43
4.1 Emotional values	43

4.2 Selection of destinations and data retrieval	46
4.3 Syntactic and Semantic Analysis	48
4.3.1 Thematic analysis	48
4.3.2 Semantic analysis	49
4.4 Summary	52
CHAPTER 5: CASE STUDY	55
5.1 Syntactic analysis	55
5.2 Semantic analysis	58
5.3 Summary	63
CHAPTER 6: CONCLUSION	64
REFERENCES	68

List of figures

Figure 1: A sample of a tweet	20
Figure 2: Ontology expressing intended models for description of the domain	22
Figure 3: Example ontology	23
Figure 4: Wordnet entry for the term Vogue	25
Figure 5: Sentence analysis	31
Figure 6: Example of the hyponym taxonomy in WordNet	35
Figure 7: Current snapshot of @VisitBCN_EN twitter account	46
Figure 8: Histogram of the main topics (main category of the emotional value) T = 3 times and S = 0.5	61
Figure 9: Histogram of the main topics (main category of the emotional value) T = 3 times and S = 0.7	61
Figure 10: Histogram of the main topics (main category of the emotional value) T = 5 times and S = 0.6	62
Figure 11: Histogram of the main topics (main category of the emotional value) T = 10 times and S = 0.7	62

List of tables

Table 1: WordNet 3.0 database statistics	27
Table 2: Examples of recovered correct words from words with repeated letters	28
Table 3: Correlation values for each semantic measure	33
Table 4: Emotional Values	45
Table 5: Selection of Destinations	47
Table 6: Top Adjectives in Berlin Tweets	48
Table 7: Top Adjectives in London Tweets	49
Table 8: Examples of adjective to noun translation	52
Table 9: Frequently Used Adjectives	57
Table 10: Snapshot of emotional category and the 3 most similar adjectives	58
Table 11: Semantic Analysis Results	60

Chapter 1

Introduction

In recent decades, the emergence of the concept of destination branding has transformed the communication of tourist destinations and has led to the emergence of a field of academic studies on the subject, which has helped to improve the professional management and communication of destination brands.

Previous studies have proved the importance of brands to distinguish territories ([1], [2]). Brands identify the personality of destinations [3], help to create a unique identity for the region and add value to the relationships with different audiences due to their identification with the brand personality [4]. A distinctive destination brand helps to create unique and favorable associations in the memory of the public and brand equity [5], influences consumer preferences [6] and induces a greater credibility which brings loyalty to the brand [7]. All these aspects influence positively the decisions of tourists, particularly when they are choosing their destinations ([8], [9]). Thus, in the last years destinations have begun to create, manage and communicate their brands with the aim of differentiating their offer from that of their competitors, improve their image and the relationship with their customers and thus influence their decision of the tourist destination to be visited.

Moreover, the emergence of social media, with their huge communicative potential, has transformed the communication of tourist destinations [10]. Several studies have shown the importance of social media in this communication ([10], [11], [12]). The first great potential of social media is that they ease the generation of comments from users and the word of mouth effect. In the field of tourism, in which it is common that the visitor does not know the destination before deciding the visit, opinions and ratings from other users that have been in a given place are considered very valuable for potential tourists ([11], [13]) and can have a big impact on their decisions ([14], [15]).

The second important aspect of social media is interactivity. On the one hand, social media permit a permanent dialogue not only among users, but also between them and destination

managers, allowing the establishment of direct relations with the public. On the other hand, interactive media also influence heavily the brand and the image of destinations, as more interaction time increases navigation and user involvement, improves brand image and increases the user's loyalty. Despite the importance of social media in the marketing of destinations, recent studies show that the potential of interactive communication of destination brands in social media is still underused [16].

The third social media potential is the ability to communicate complex brand destinations through narratives and images [17]. With the use of narratives the meanings attributed to these destinations are perpetuated and the sense of place is communicated better [18]. Images also possess a great potential in the identification and communication of emotional values of the brand; however, most destinations use pictures to promote attractions and not to transmit emotional values of the brand.

As a result of the rise of social media, tourist destinations try to transmit their identities and their brands through these new communication tools, but due to its novelty, they lack the expertise to decide which social media to use, how to use them and what kind of strategies should be employed to communicate well their brand [19]. Thus, it is necessary to design and develop tools to analyze and evaluate how destinations communicate their identities and their brands through social media. However, there does not exist any unified methodology nor any consensus on how this analysis should be made [20], perhaps because the idea of destination brand may be conceptualised from different viewpoints and studied by diverse disciplines.

Some authors understand place branding as defining an intended identity of the destinations ([3], [4]) and they study how it can be transmitted, whereas others take it as the image perceived by the public ([21], [22], [23]). These authors analyze the image projected by destinations to different kinds of audiences and some of them analyze how images influence the decision of tourists to visit a place [24]. These two perspectives on branding are the most numerous and popular on tourist research [25].

The most prominent research field is the analysis of the destination image [20]. The works in this area analyze how tourists see a destination, and compare this external image with the

one that the destination intends to transmit ([22], [23]). These studies generally focus on user surveys, in which respondents associate the destination to a list of attributes in predefined questionnaires. Although quantitative methods of analysis are the most common in this area, sometimes qualitative methods are also applied through in-depth interviews or discussion groups.

Studies that analyze the identity of destinations are not so numerous. They generally analyze the content of the messages communicated by destinations to detect which attributes and emotional values they intend to associate with the territory [19]. They use both quantitative methods (e.g. counting the frequency of the words) and qualitative techniques (e.g. more complex analysis that try to find out the transmitted identity attributes).

Due to the increasing use of social media by tourist destination managers, the communication of their brands has also evolved. Initially, studies focused on manual content analysis. Researchers generally defined a template to analyze the association of certain attributes with the content of the publications of the destinations on their social media, and filled this template by hand. As the number of messages transmitted through social media grows daily, it is becoming impractical and expensive (and, in some cases, simply unfeasible) to make this kind of analysis. Another problem of current studies is that they usually focus on the analysis of the cognitive component of the brand, neglecting the affective or emotional dimension ([10]). For these reasons it has become necessary to devise automated methods of analysis of the large quantities of messages produced by tourist destinations, in order to discover how they transmit the affective component of brands.

1.1 Objectives

The main goal of this work is to develop a full fledged methodology that can be used to make an automatic analysis of the way in which destinations try to communicate their brands through the use of Twitter. In order to achieve this main goal, the work will focus on the following specific goals:

- Identification, with support from the Communication specialists, of the key emotional aspects that may be included in typical Tourist destination brands.
- Search of the official Twitter accounts of the most relevant European tourist destinations, and selection of those that have a high enough number of tweets written in English.
- Retrieval of the tweets sent by the selected accounts. Application of Natural Language Processing tools to obtain important components (nouns, noun phrases, adjectives, verbs, hashtags, etc.).
- Definition of syntactic analysis tools, based on the frequency of appearance of certain keywords or the co-occurrence between them.
- Definition of semantic measures that can use standard knowledge repositories like WordNet to assess the similarity between terms or between tweets.
- Use of these ontology-based measures to design tools that may be used to evaluate the effectiveness of the brand message (i.e., the relationship between the terms used in the tweets and the terms that describe the emotional aspects associated to branding).

1.2 Overview of the methodology

The main steps of the methodology that has been devised in this work to analyse the communication of destination brands through social media are the following:

- Definition of the emotional values associated to a destination brand (see Section 4.1).
- Selection of the destinations to be analysed, and retrieval of the set of tweets (see Section 4.2).
- Syntactic and semantic analysis of the content of the tweets (see Section 4.3).

1.3 Document structure

The rest of this document is organized as follows:

- Chapter 2: presents a background overview of the basic concepts, techniques, and tools used to support this work. It contains a brief discussion about Twitter, the knowledge structure WordNet, natural language processing (NLP) and ontology-based semantic similarity.
- Chapter 3: presents a brief state of the art on the methods employed in destination branding studies to study messages in social media.
- Chapter 4: explains the methodology followed in this work.
- Chapter 5: includes the experimental results of a case study and their analysis.
- Chapter 6: presents a list of conclusions of this work and comments some lines of future work.

Chapter 2

Background

In this chapter, the basic concepts and tools used to support this dissertation are presented. The first part of the chapter comments Twitter, which was considered as the source of the information to be analysed in this work. Later, we describe the knowledge structures used in this work, in particular, WordNet, the most used English lexical database. In the last part of the chapter some key techniques are introduced: natural language processing techniques (including parsing), and ontology-based semantic similarity measures.

2.1 Twitter

Twitter is an online social networking service that enables users to send and read short 140-character messages called "tweets". Each tweet is a string of up to 140 characters (Figure 1) that may basically contains text, links (marked in black rectangle), user mentions (marked in red rectangle), symbols emoticons (marked in yellow rectangle) and hashtags (strings preceded by the # symbol with which users tag their messages; marked in green rectangle). *Twitterers* (i.e., users of Twitter) may publish *tweets* to share news and opinions with the rest of the community. An important feature of Twitter that makes it different from other social networks is the fact that users do not need to give permission to the people that want to receive their messages. In fact, Twitter employs a social-networking model called —following, in which each *twitterer* can follow any other user without seeking any permission and, in consequence, he may also be followed by others without granting permission first. This is useful for users who want to receive tweets from users who they are following (i.e., *followees*) and to share their tweets with those that they are followed by (i.e. *followers*). Concerning the communication model, *tweets*, *replies* and *retweets* are the core of Twitter. Tweets are the messages published by twitterers. Any *twitterer* can reply to a tweet adding extra information or giving his impression about it creating a natural conversation among users. Finally, if a

twitterer wants to only share with his followers a tweet that he has read he may retweet it and, automatically, it will be spread among his *followers*.

Twitter is a real-time environment, which means that tweets contain the most up-to-date and inclusive stream of information and commentary on current events, people's opinions, business trends, etc.



Figure 1: A sample of a tweet

Generally, tweets are usually ungrammatical, and they contain many mood expressions, acronyms, abbreviations and symbols. These features motivate the need for systems that can aggregate and categorize all its content. All of these factors define an appealing research area of study for knowledge discovery and data mining. Twitter provides an API¹ so that researchers can retrieve sets of tweets and analyse them.

In this work, Twitter is used to exemplify the extraction of knowledge from user-generated content in a social network (in particular, *semantic analysis* of the adjectives used in the tweets to associate them to the categories of emotional values) in a very restricted and noisy setting that presents many challenges. In fact, tweets are very concise, often missing sufficient context to categorize them and moreover, there is a huge number of daily messages, which contain insignificant and redundant information, which may lead the system to an overload situation.

2.2 Knowledge repositories

This section describes the idea of the knowledge repositories. It contains two subsections. Section 2.2.1 introduces what is an ontology and lists its main characteristics. Section 2.2.2

¹ See <https://dev.twitter.com/>

briefly describes the structure of the knowledge repository WordNet and explains how to use it to obtain the synonyms, hypernyms and hyponyms of a word.

2.2.1 Ontology basics

The word “Ontology” has various definitions depending on various texts, domains and applications. In philosophy and linguistics ontology is “theory of existence” or “a taxonomy of the world”. The most famous and popular definitions of ontology in the AI community are:

- “Ontology is a formal, explicit specification of a shared conceptualization” [42]. Conceptualization refers to the construction of an abstract model of some phenomenon in the world by identifying its basic associated concepts. Explicit means that the type of concepts used, and the constraints of their use, are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private of some individual, but accepted by a group.
- “Ontology is an explicit specification of conceptualization” [43]; this definition is consistent with the usage of ontology as set of concept definitions, but more general. A conceptualization can be defined as an intensional semantic structure that encodes implicit knowledge constraining the structure of a piece of a domain. Ontology is a (partial) specification of this structure, i.e., it is usually a logical theory that expresses the conceptualization explicitly in some language. Conceptualization is language independent, while ontology is language dependent. The use can be illustrated in Figure 2 it shows how an ontology restricts (i.e., defines) possible use of constructs used in the description of the domain. Notice that ontology does not have to express all the possible constraints - the level of detail in the conceptualization depends on the requirements of the intended application and in addition expressing a conceptualization in an ontology depends on the used ontology language.
- Other approaches have defined ontologies as the shared understanding of some domain of interest [44], or as a body of knowledge describing some domain, typically common sense knowledge domain.

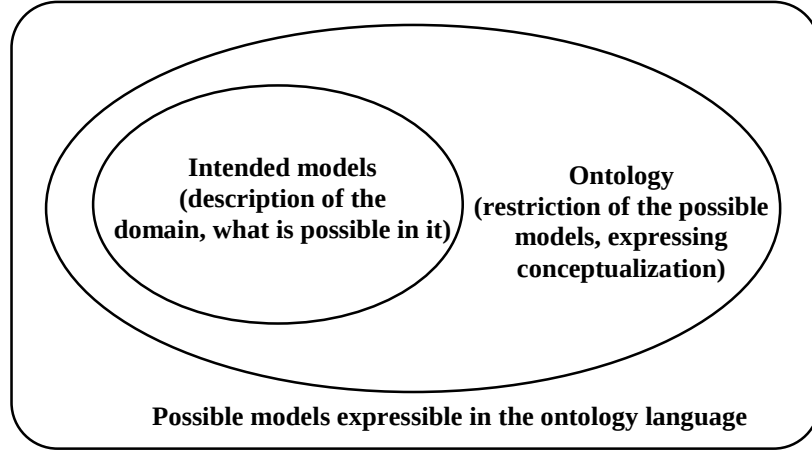


Figure 2: Ontology expressing intended models for description of the domain

- Formally, an ontology can be defined as a structure $O = (C, \leq_C, R, \sigma_R, \leq_R, A, \sigma_A, T)$ [45, 46].

where,

- C, R, A , and T represent disjoint sets of concepts, relations, attributes and data types, respectively. Concepts are sets of real world entities with common features (such as different types of diseases, treatments, actors, etc.). Relations are binary associations between concepts. There exist inter-concept relations which are common to any domain (such as hyponymy, meronymy, etc.) and domain-dependent associations (e.g. an Actor performs an Action, a Disease is treated with a certain Treatment, etc.). Attributes represent quantitative and qualitative features of particular concepts (such as the medical code of a Disease, the degree of contagiousness, etc), which take values in a given scale defined by the data type (e.g. string, integer, etc.).
- $\leq_C$ represents a concept hierarchy or taxonomy for the set C . In this taxonomy, a concept $c1$ is a subclass, specialization or subsumed concept of another concept $c2$ if and only if every instance of $c1$ is also an instance of $c2$ (which represents its superclass, generalization or subsumer). Concepts are linked by means of transitive is-a relationships (e.g. if respiratory disease is-a disorder and bronchitis is-a Respiratory_Disease, then it can be inferred that Bronchitis is-a Disorder).

Multiple inheritance (i.e. the fact that a concept may have several hierarchical subsumers) is also supported (for example, Leukaemia may be both a subclass of Cancer and Blood_Disorder).

- $\leq_R$ represents a hierarchy of relations (e.g. has-primary-cause may be a specialization of the relation has-cause, which indicates the origin of a Disorder).
- $\sigma_R: R \rightarrow C^+$ refers to the signatures of the relations, defining which concepts are involved in one specific relation of the set R . For example, the signature $\sigma_R(\text{is-treated-with}) = [\text{Disease}, \text{Treatment}]$ indicates that is-treated-with establishes a relation between the two concepts Disease and Treatment. It is worth to note that some of the concepts in C^+ correspond to the domain (the origin of the relation) and the rest to the range (the destination of the relation). In this example, Disease is the domain of the relation is-treated-with, and Treatment is the range. Those relationships may fulfil properties such as symmetry or transitivity.
- $\sigma_A: A \rightarrow C \times T$ represents the signature describing an attribute of a certain concept C , which takes values of a certain data type T (e.g. the attribute number-of-leukocytes of the concept Blood_Analysis, which must be an integer value).

The following figure shows an example of ontology.

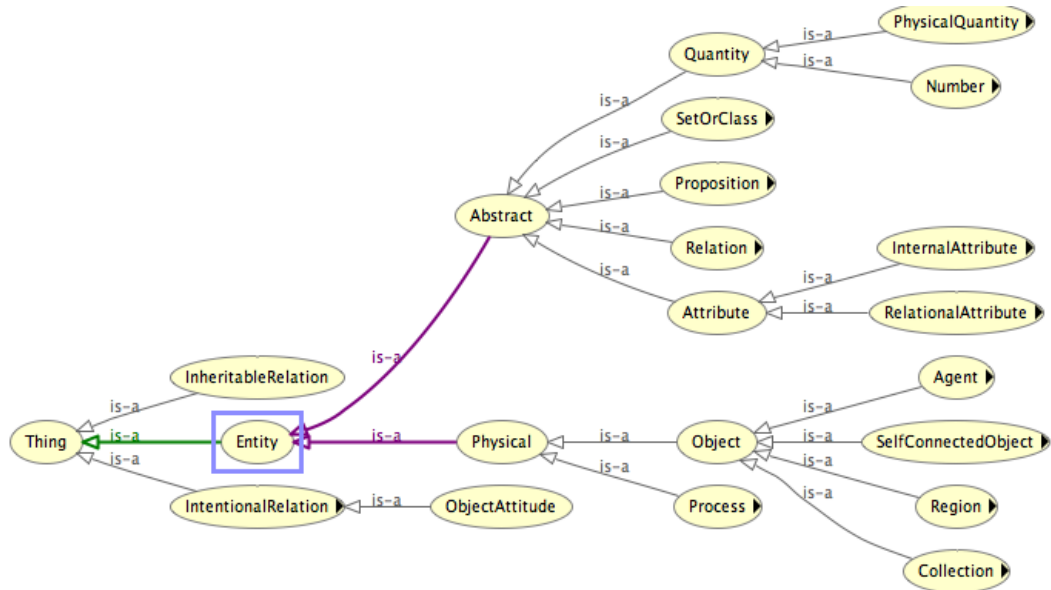


Figure 3: Example ontology

Some standard languages have been designed to construct ontologies. They are usually declarative languages based on either first-order logic or on description logics. Some examples of such ontology representation languages are KIF, RDF, KL-ONE, DAML+OIL and OWL [47]. There are some differences between them according to their supported degree of expressiveness. In particular, OWL is the most complete one, allowing one to define, in its more expressive forms (OWL-DL and OWL-Full) logical axioms representing restrictions at a class level. They are expressed with a logical language and contribute to define the meaning of the concepts, by means of specifying limitations regarding the concepts to which a given one can be related to. Several restriction types can be defined:

- Cardinality: defines that a concept's individual can be related (by means of a concrete relation type) to a minimum, maximum or exact number of other concept's instances. For example, certain types of Disease may have at minimum one Symptom.
- Universality: indicates that a concept has a local range restriction associated with it (i.e. only a given set of concepts can be the range of the relation). For example, all the Symptoms of a certain Disease must be of the same type, the same concept category.
- Existence: indicates that at least one concept must be the range of a relation. For example a Disease always presents a certain kind of Symptoms, even though other ones may also appear.

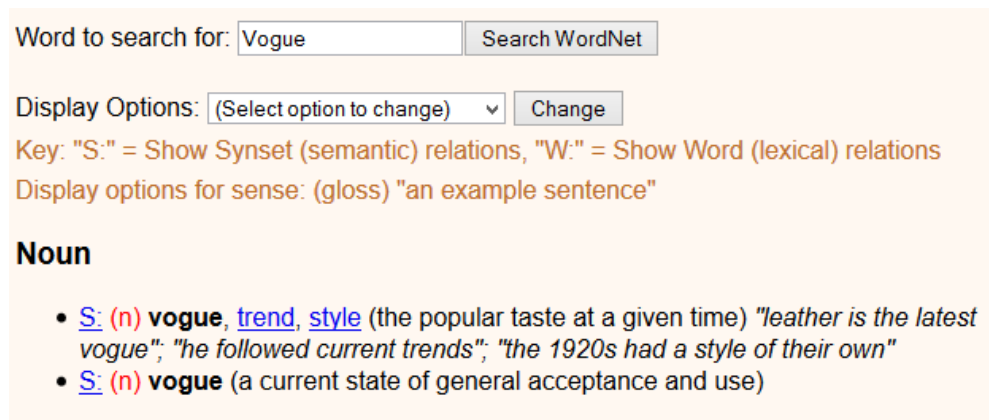
All those restrictions can be defined as Necessary (i.e. an individual should fulfil the restriction in order to be an instance of a particular class) or Necessary and Sufficient (i.e. in addition to the previous statement, an individual fulfilling the restriction is, by definition, an instance of that class). This is very useful for implementing reasoning mechanisms when dealing with unknown individuals.

In addition, OWL also permits to represent more complex restrictions by combining several axioms using standard logical operators (AND, OR, NOT, etc.). In this manner, it could be possible to define, for example, a set of Symptoms which co-occur for a particular Disease using the AND operator.

The main using of the ontologies in this work is the use of ontology-based semantic similarity measures (on the WordNet ontology) to make a semantic content analysis of the tweets of the brand destination tourists (Chapter 4).

2.2.2 WordNet

WordNet is a lexical database or semantic electronic repository for the English language. It groups English words into sets of synonyms called *synsets*, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members. WordNet can thus be seen as a combination of dictionary and thesaurus. In this section, an overview of its characteristics, structure and potential usefulness for our purposes is described.



Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) **vogue**, [trend](#), [style](#) (the popular taste at a given time) *"leather is the latest vogue"; "he followed current trends"; "the 1920s had a style of their own"*
- [S:](#) (n) **vogue** (a current state of general acceptance and use)

Figure 4: Wordnet entry for the term Vogue

WordNet is the most commonly used online lexical and semantic repository for the English language. WordNet was created in the Cognitive Science Laboratory of Princeton University. Many authors have contributed to it or used it to perform many knowledge acquisition tasks. WordNet includes the main lexical categories (nouns, verbs, adjectives and adverbs) but ignore prepositions, determiners and other kinds of words.

Concretely, it offers a lexicon, a thesaurus and semantic linkage between the majorities of English terms. It seeks to classify words into categories and to interrelate the meanings of those words. It is organized in synonym sets (*synsets*): a group of data elements that are considered semantically equivalent for the purposes of information retrieval. According to

WordNet, a *synset* or synonym set is defined as a set of one or more synonyms that are interchangeable in some context without changing the truth-value of the proposition in which they are embedded. Each word in English may have many different senses in which it may be interpreted: each of these distinct senses points to a different synset. Every word in WordNet has a pointer to at least one synset.

Each synset, in turn, must point to at least one word. Thus, we have a many-to-many mapping between English words and synsets at the lowest level of WordNet. It is useful to think of synsets as nodes in a graph. At the next level we have lexical and semantic pointers. A semantic pointer is simply a directed edge in the graph whose nodes are synsets. The pointer has one end we call a source and the other end we call a destination.

All synsets are connected to other synsets by means of semantic relations. These relations, which are not shared by all lexical categories, include:

Nouns

hypernyms: Y is a hypernym of X if every X is a (kind of) Y.

hyponyms: Y is a hyponym of X if every Y is a (kind of) X.

coordinate terms: Y is a coordinate term of X if X and Y share a hypernym.

meronym: Y is a meronym of X if Y is a part of X.

holonym: Y is a holonym of X if X is a part of Y.

Verbs

hypernym: the verb Y is a hypernym of the verb X if the activity X is a (kind of) Y.

troponym: the verb Y is a troponym of the verb X if the activity Y is doing X in some manner.

entailment: the verb Y is entailed by X if by doing X you must be doing Y.

coordinate terms: those verbs sharing a common hypernym.

Finally, each synset contains a description of its meaning, expressed in natural language as a gloss. Example sentences of typical usage of that synset are also given. All this information

summarises the meaning of a specific concept and models the knowledge available for a particular domain. Table 1 depicts the WordNet 3.0 database statistics (number of words, synsets and senses).

POS	Unique Strings	Synsets	Total Word-Sense Pairs
Noun	117798	82115	146312
Verb	11529	13767	25047
Adjective	21479	18156	30002
Adverb	4481	3621	5580
Totals	155287	117659	206941

Table 1: WordNet 3.0 database statistics

In this work, WordNet will be particularly useful to associate the extracted adjectives from the tweets to the emotional values (Table 4) based on its meaning (semantic similarity).

2.3 Techniques

This section describes the basic ideas underlying the main techniques used in this work, to provide the background necessary to understand the rest of the work. First, the main techniques in Natural Language Processing are presented in section 2.3.1. In section 2.3.2, the use of Web-based statistical measures and ontology-based semantic similarity measures to compute the relatedness of two terms are studied.

2.3.1 Preprocessing

Twitter language is casual and noisy, tweets contain numerous punctuation, spelling errors, and slang terms, thus, the tweets are pre-processed to mitigate misspelling and casual language used in Twitter as follows:

- Tweets may contain URLs, usernames, hashtags, and emoticons. All URLs, usernames, and symbols are removed from the tweet.
- To reduce the dimensionality, all tweets are converted to lowercase, and stop words have been removed.

- Words with repeated letters are corrected using the Repeated Letters Remover algorithm (algorithm1). This algorithm uses the Depth First Search (DFS) technique to get the correct word from the WordNet corpus.

Repeated letters word	Recovered word
happpppyyyy	Happy
goooooood	good, DFS will not return god.
xxxxxyyyromo	<Null>, algorithm failed

Table 2: Examples of recovered correct words from words with repeated letters

The first step in this algorithm is trying to remove one repeated letter (if there are more than two repeated letters it will be replaced by two, e.g. hhapppppppy will be hhappy) using the function *removeOneRepeatedLetter*, the second process step is to lookup for the word obtained from the previous step in WordNet corpus. If there exist some synsets of this word then WordNet contains this word and the algorithm will return it as a correct word and terminate successfully; otherwise, if there are no more repeated letters to remove then the algorithm will fail. Otherwise the algorithm will try to recover the correct word from the word obtained in the first step. Table 2 shows some examples of words with repeated letters and the corrected word returned by the algorithm.

2.3.2 Natural Language processing

In the philosophy of language, a natural language (or ordinary language) is any language which arises in an unpremeditated fashion as the result of the innate facility for language possessed by the human intellect. A natural language is typically used for communication, and may be spoken, signed, or written.

Natural Language Processing (NLP) includes the study of mathematical and computational models of various aspects of language and the development of a wide range of systems. Research in NLP is highly interdisciplinary, involving concepts in computer science, linguistics, logic, and psychology. NLP has a special role in computer science, particularly in

the sub-field of Artificial Intelligence, because many aspects of the field deal with linguistic features of computation and NLP seeks to model language computationally.

Algorithm 1: Repeated Letters Remover

repeatedLettersRemover(word) :

 word1 = removeOneRepeatedLetter(word)

 if getSynsets(word1) != {}: *// word1 has been found in WordNet*

 return word1

 else if word1 == word: *// no others repeated letters to remove*

 return word

 else:

 return repeatedLettersRemover(word1)

Concerning the analysis of text itself, this work only considers English written tweets. Therefore, a set of tools and algorithms for analysing English natural language is used for that purpose. Concretely; we have used a Natural Language Processing Parser. It is the responsible of detecting sentences, tokens and parts of speech (Text processing) and performing the syntactic analysis or Part-Of-Speech tagging. The first component is able to chunk a text in order to find its minimal parts. Once the text is chunked, the minimal pieces obtained are tagged with a Part-Of-Speech (POS) tagger. The syntactic analyser or Part-Of Speech tagger allows performing basic morphological and syntactical analyses of particular pieces of text that can contain valuable information. This will provide a way to extract Adjectives from text, in the process of the thematic analysis as explained on section 4.3.1. The following subsection explains in more detail the NLP parser.

2.3.2.1 Natural Language Processing parser

The first step of natural language parsing is to detect sentences from a text. Most natural language processing tools use as default the point (.) as a delimiter to separate sentences. Other delimiters can be used such as comma (,), question mark (?), exclamation (!) and others. Once

the sentence detector splits the whole text, the tokenization is the next step. It is in charge of separating the words for the analysis. For example, the word —don't should be separated into the words —do and —not for a good further text analysis. After that, the sentence analysis can be applied. Once the tokenization is successfully done, the part of speech tagging or word category disambiguation process is executed. POS tagging is the process of marking up the words in a text (corpus) as corresponding to a particular part of speech such as nouns, verbs, adjectives, etc. The POS tagging is very useful because it provides syntactic information of every word in the sentence structure, but it can also be useful in itself when looking for units of meaning in a sentence. The last step is text chunking, which consists of dividing a text in syntactically correlated parts of words, like noun groups or verb groups, but without specifying their internal structure or their role in the main sentence.

A simple example of sentence analysis is provided in Figure 5, with the sentence —*The red book is a black novel*. First, the words are marked as corresponding to a particular part of speech, by means of POS tagging, such as nouns, verbs, adjectives, etc. In this example the tagged components are:

- DT: Determiner
- JJ: Adjective
- NN: Common noun
- VBZ: Verb, 3rd person singular present

After POS tagging, chunking is applied in order to divide the text in syntactically correlated parts of words. In this case only in noun and verb phrases:

- NP: Noun Phrase
- VP: Verb Phrase

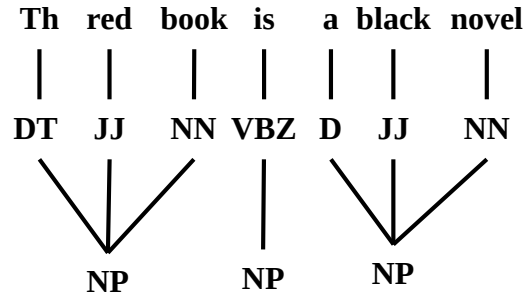


Figure 5: Sentence analysis

In this work, the Natural Language Processing toolkit “NLTK² has been used. NLTK is a leading platform for building Python programs to work with human language data, It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning.

The aim of using the parser is to extract the adjectives from the tweets (Section 4.3).

2.3.3 Ontology-based semantic similarity

The analysis and classification of textual resources implies the comparison among words. Words are labels referring to concepts, which define their semantics. In consequence, semantic similarity measures are needed. Semantic similarity is precisely the science that aims to estimate the alikeness between words or concepts by discovering, evaluating and exploiting their semantics.

In the literature, we can distinguish different approaches to compute semantic similarity according to the techniques employed and the knowledge exploited to perform the assessment. The most prominent ones are edge counting-based measures (that map terms to ontological concepts and calculate their similarity by looking at the length of the is-a path connecting them), feature-based measures (which assess the similarity between concepts as a function of their properties, i.e. taking into account synonyms, definitions, etc.) and information-content

² See <http://www.nltk.org/>

based measures (that quantify the degree of information provided by the concepts by means of the statistical information derived from a corpus).

Evaluating and comparing different semantic similarity measures is a difficult task since the notion of similarity is subjective [48]. In order to enable fair comparisons, several authors created evaluation benchmarks consisting on word pairs whose similarity was assessed by a set of humans. Rubenstein and Goodenough [49] and Miller and Charles [50] define some experiments in which native English speakers assessed the similarity of 65 word pairs selected from ordinary English nouns on a scale from 0 (semantically unrelated) to 4 (highly synonymous). The results of such experiments have become *de facto* standard benchmarks to evaluate and compare the accuracy of similarity measures. As a result, correlation values obtained against those benchmarks can be used to numerically quantify the closeness of two sets of ratings (i.e. the human judgments and the results of the computerised assessment). If the two rating sets are exactly the same, the correlation coefficient is 1 whereas 0 means that there is no relation.

In [51], the authors compared the performance of some of the most popular semantic similarity measures taking the correlation values originally reported by related works for the benchmarks of Rubenstein and Goodenough and Miller and Charles (when available). In case in which a concrete measure depended on certain parameters (such as weights or corpora selection/processing) the best correlation value reported by the authors was compiled. It is important to note that, even though some of them relied on different knowledge sources (such as tagged corpora or the Web), all ontologybased ones used WordNet (recall section 2.2.2). Table 3 summarises this comparative.

Correlation values indicate that measure accuracies are very similar through the different families. However, the applicability and generality of each type of measure depend on the principle they exploit. The main advantage of edge-based measures is their simplicity. They only rely on the geometrical model of an input ontology whose evaluation requires a low computational cost. However, several limitations hamper their performance. One of these limitations is that they rely on the notion that all links in the taxonomy represent a uniform distance [48]; thus, they require wide ontologies with a relatively homogenous distribution of

semantic links and good domain coverage to minimise this problem [52] Consistent ontologies like WordNet work properly with these measures [53]. Feature-based measures exploit more semantic evidences than edge-counting approaches, evaluating both commonalities and differences of compared concepts, but they rely on features like glosses or synsets (in addition to taxonomic and non-taxonomic relationships); therefore, those measures can only be applied to ontologies in which this information is available. Another problem is their dependence on weighting parameters that balance the contribution of each feature. Finally, information-content measures need an accurate computation of concept probabilities that requires a proper disambiguation and annotation of each noun found in the corpus. If either the taxonomy or the corpus changes, re-computations must be recursively executed for the affected concepts. So, it is necessary to perform a manual and time-consuming analysis of corpora and the resulting probabilities will depend on the size and nature of input corpora. Moreover, the background taxonomy must be as complete as possible in order to provide reliable results. All those aspects limit the scalability and applicability of those approaches.

Measure	Type	M&C	R&G
Path (Rada)	Edge	0.59	N/A
Wu & Palmer	Edge	0.74	N/A
Leacock & Chodorow	Edge	0.74	0.77
Rodriguez	Feature	0.71	N/A
Tversky	Feature	0.73	N/A
Petrakis	Feature	0.74	N/A
Resnik	IC	0.72	0.72
Lin	IC	0.7	0.72
Jiang & Conrath	IC	0.73	0.75

Table 3: Correlation values for each semantic measure

Considering all the advantages and drawbacks of each one of the types of similarities, we have chosen the Wu and Palmer similarity measure [41] to compare two different terms because it does not depend on a corpora, it does not have tuning parameters and it has a very low computational cost. The next equation shows the semantic similarity between two concepts using the Wu and Palmer distance. The main difference with the other edge-counting measures is that whereas they omit the fact that equally distant concept pairs belonging to an upper level of the taxonomy should be considered as less similar than those belonging to a lower level (as they present different degrees of generality), Wu and Palmer takes into account the depth of the concepts in the hierarchy to avoid this problem.

$$Sim_{w\&p}(C1, C2) = \frac{2*N3}{N1+N2+2*N3} \quad (1)$$

In this expression N1 and N2 are the number of hypernym links from the terms c1 and c2 to their Least Common Subsumer (LCS) in the reference ontology, and N3 is the number of hypernym links from the LCS to the root of the ontology.

This measure ranges from 1 (for identical concepts) to 0 (when the LCS is the root of the ontology, so the concepts do not have any common ancestor). The main difference of this function with respect to other edge-counting measures is that it takes into account the depth of the compared concepts in the hierarchy (given the same distance, two concepts are more similar if they are more specific). Figure 6 shows an example of the hyponym taxonomy in WordNet used for path length similarity measurement:

In the above figure, we can observe that the path length between car and auto is 0, car and truck is 2, car and bicycle is 3, car and fork is 11. A shared parent of two synsets is known as a subsumer. The least common subsumer (LCS) of two synsets is the subsumer that does not have any children that are also subsumers of two synsets. In other words, the LCS of two synsets is the most specific subsumer of the two synsets. Back to the above example, the LCS of {car, auto...} and {truck...} is {automotive, motor vehicle}, since the {automotive, motor vehicle} is more specific than the common subsumer {wheeled vehicle}, thus, the depth of the LCS of {car, auto...} and {truck...} is 7. The path length gives us a simple way to compute the distance between two word senses.

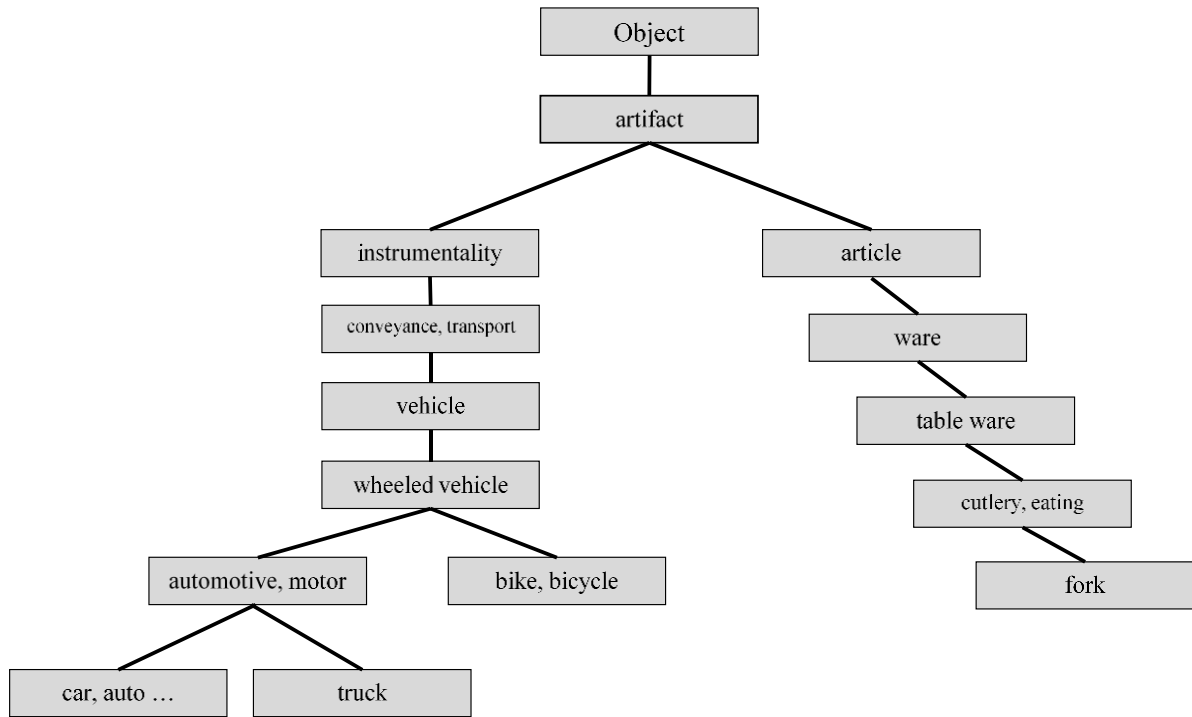


Figure 6: Example of the hyponym taxonomy in WordNet

2.4 Summary

This chapter presented a background of the techniques that have been used in this work. The aim of this overview was to make the understanding of the methodology of this work simpler and easier.

In this work the most common techniques of text pre-processing such as converting to lower-case, removing the repeated letters, .etc. have been used to mitigate the misspelling and casual language used in Twitter. The Natural Language Processing Parser is necessary in this work to extract the adjectives used in the tweets. In this work NLTK tools such as tokenizer, tagger, parser, etc. have been used.

A review of the main semantic similarity measure that can be used to estimate the resemblance between concepts, has been presented. Through the exploitation of knowledge offered by ontologies and the semantic measures reviewed in this chapter, it is possible to

compute the similarity between concepts in order to develop appropriate operators to associate the adjectives used in the tweets to the categories of emotional values shown in Table 4.

The next chapter provides a review of the state of the art on the methods that have been used to analyze and evaluate the communication through social media.

Chapter 3

State of the Art

The main challenge in the evaluation of the communication through social media lies in the large volume of information to be analysed. In recent years there have been two kinds of studies related to the evaluation of the communication of the identities and brands of territories. The first is based on the *numerical analysis*, and the second line of research, based on *content analysis*.

This chapter provides a brief survey of the two kinds of methods that have been used to analyze and evaluate the communication through social media.

3.1 Numerical Analysis Methods

This kind of evaluation methods use the numerical analysis of several communication-related attributes such as the number of followers/friends a destination has in social media, the number of comments they post, and the number of reactions generated among users (e.g. replies, retweets, number of "likes", etc.). There are already many automated tools that measure these purely quantitative aspects of communication. For instance, [26] considered various quantitative communication aspects to analyze the Spanish brand in three levels of diffusion of information. In first level (*the diffusion of messages by the State-level NTO account, @Spain*) the following quantitative measures number of messages published of Spain, and the number of times the user Spain mentions each user community has used. In the second level (*the retweets of the messages posted by Spain, and all the posts mentioning the user Spain, by the rest of the Twitter community*) two analysis measures were used: the number of times each user retweets messages posted by the user Spain and The number of times each user mentions the user Spain. The number of times each user retweets a message where Spain has been mentioned, and the number of retweets each user received has been used to analyse the third-level (*the retweeting of messages where Spain has been mentioned, in particular by the Autonomous Community-level DMOs accounts*). This kind of analysis,

although they provide interesting information about the use of social media for the communication of tourist destinations, does not provide any knowledge about the actual contents of the communication. The work reported in [19] started measuring the average reaction to top posts per destination type by calculating the average total posts, average of likes, average of comments, and average of shares for 37 Spanish destinations located in five autonomous communities (*Andalusia, Canary Islands, Catalonia, Galicia, and Madrid*).

3.2 Content Analysis Methods

Content analysis refers to a family of procedures for the systematic, replicable analysis of text. The methods of content analysis enable the researchers to include large amounts of textual information and systematically identify its properties, such as the frequencies of most used keywords by locating the more important structures of its communication content. Such amounts of textual information must be categorised to provide a meaningful reading of content under scrutiny.

These approaches that on the content analysis usually combine qualitative and quantitative methods [27].

3.2.1 Qualitative Methods

Unstructured qualitative approaches, such as in-depth interviews, open questions and focus groups, provide richer data, but they are necessarily limited to a small sample, limiting the scope of the analysis and the generality of the results [28]. For example the work introduced in [54] used a purely qualitative approach to understand the evolution of the most common themes in each destination (*Detroit, Kansas, Las Vegas, New Orleans, Orlando, Phoenix, San Francisco, and Seattle*) after the exposure to the stimuli. In that work the NVIVO version 10 qualitative research software was used to analyse the data in three stages, In the first stage, the data was arranged into two files for each city as before and after the exposure to the stimuli. The word frequency analysis was conducted in the before files and top five words were identified to find out the main themes in the files. In the second stage, concurrence analyses of the identified top five words in before and after files were conducted separately to see the

evolution of these terms after the exposure to the stimuli. In the last stage the conative component in the concurrence analysis was identified.

According to [56] there are five research purposes for which qualitative studies are especially useful:

- (1) Understanding the meaning of events, activities, situations and actions of people observed.
- (2) Understanding the particular context within which the participants act and the influence which this context has on their actions.
- (3) Identifying unanticipated phenomena and influences and generating new, “grounded” theories about the latter. This view is in stark contrast to the rather passive acceptance that all “great” theories had been discovered and that the role of research lay in testing these through quantitative “scientific” procedures.
- (4) Understanding the processes by which events and actions take place. The ability to get insights about the processes that lead to outcomes is in many cases superior to experimental and survey research that mostly capture outcomes only and are poor in investigating processes.
- (5) Developing causal explanations.

3.2.2 Quantitative Methods

In natural sciences and social sciences, quantitative research is the systematic empirical investigation of observable phenomena via statistical, mathematical or computational techniques. [57] The objective of quantitative research is to develop and employ mathematical models, theories and/or hypotheses pertaining to phenomena. The process of measurement is central to quantitative research because it provides the fundamental connection between empirical observation and mathematical expression of quantitative relationships. Quantitative data is any data that is in numerical form such as statistics, percentages, etc... [57]. The researcher analyzes the data with the help of statistics. The researcher is hoping the numbers will yield an unbiased result that can be generalized to some larger population. Qualitative

research, on the other hand, asks broad questions and collects word data from phenomena or participants. The researcher looks for themes and describes the information in themes and patterns exclusive to that set of participants.

Quantitative methods measure a series of predefined attributes of the destinations. There are two main types of quantitative content analysis: thematic and semantic [29].

Thematic analysis methods: these methods just count the themes/categories/words that appear in the text, without interpreting them; thus, they provide limited information on the content of the analysed messages. Analyses of word-counts yield inferences about the predominance of themes in texts [29]. For example the work reported in [30] tries to evaluate the personality of the destination brand from the comments of its visitors in social media and to provide useful recommendations to destination managers about how to communicate a given brand personality by using WordStat software to compare a list of words selected by the researcher (dictionary) against the text loaded into the software and returns the frequencies with which these words occur in the text. The work presented in [19] used the “Attraction factors, Activities, and Services” as themes and an adaptation of the “Brand Personality Scale” [38], to calculate the most mentioned themes and brand values, then studied the themes and brand values which generate most reactions per destination type and compared them to the most mentioned themes and values.

Semantic analysis methods: Relations among themes are encoded in a semantic text analysis, however. In a semantic text analysis the researcher begins by constructing a template (a.k.a. a semantic grammar). Themes from sampled texts are then mapped as syntactic components within this template. Semantic methods try to uncover the actual meaning of the words to make an analysis at the conceptual level, and not at the syntactic one. For instance, [31] analyzed how two popular tourist destinations (*Mexico and Brazil*) communicate the personality of their brand through Facebook, and in [32] the authors developed a system that makes a semantic analysis of the image of the Basque Country, taking into account cognitive, affective and conative aspects. In [55], a Define-Measure-Visualize (DMV) framework proposed, it is composed of two network analysis methods “social network”, and “semantic network”. The aim of the semantic network is to identifying the associations. In this work the

WordStat software used to analyse the content of tweets and calculate the frequency of the keywords and the phrases. WordStat was also used to cluster the keywords into some categories based on co-occurrence and similarity.

There are several authors that ***combine thematic and semantic analysis***, trying to overcome the limitations of plain syntactic studies (e.g. [27], [33]). The first one can identify the topics that are more related to the brand (by finding the most frequently used words) and the second one can make a more profound analysis of their degree of association with the brand and their interrelationships. This combination will also be used in the study reported in this dissertation, as will be detailed later. Firstly, an initial step will study the frequency of appearance of the adjectives used in the tweets sent by an official tourism destination account. After that, a second semantic step will analyze the relationship of those adjectives with the emotional values associated to the destination brand.

Several authors (e.g. [27]) have highlighted two models in quantitative content-based analysis. The substitution model analyzes the presence in the text of a given set of predefined categories, whereas the correlational one finds the categories of interest through the analysis of the text. Our study combines both possibilities. In the first part of the work we consider a correlational model, as we analyse which are the adjectives that are more commonly used in a set of tweets (without having any list of predefined words of interest); however, in the second part we take a substitution approach, since we associate those adjectives with a list of predefined categories of emotional values.

The authors in [33] considered the frequency of the occurrence of words as the basis of the analysis, by using word frequency and co-occurrence as a basis for analysis. Leximancer was used to conduct thematic and semantic content analysis of the 33,962 English tweets in the collected corpus.

3.3 Summary

This chapter summarized the state of the art of the most related work. Two basic types of analysis methods can be used to evaluate the communication through social media, “*Numerical Analysis Methods*” and “*Content Analysis Methods*”. The *numerical analysis*

methods use the numerical analysis of several communication-related attributes such as the number of followers/friends a destination has in social media, the number of comments they post, and etc. The methods of *content analysis* enable the researchers to include large amounts of textual information and systematically identify its properties, such as the frequencies of most used keywords by locating the more important structures of its communication content.

In this work we combined the two approaches of the content analysis methods: “Syntactic/Thematic Analysis” and “Semantic Analysis”. The next chapters (Chapter 4 and Chapter 5) explain how these methods were used and discuss the results that were obtained.

Chapter 4

Methodology

Twitter and Facebook are the social media more commonly used by tourism destination managers for their promotion [34]. In this work we have focused on the analysis of English tweets sent by official tourist destination accounts, because they provide a real-time view of how they try to communicate their brand to potential customers [33] and how they dialogue with them [35].

The steps followed in the methodology of analysis have been the following:

- Definition of the emotional values associated to a destination brand.
- Selection of the destinations to be analysed, and retrieval of the set of tweets.
- Syntactic and semantic analysis of the content of the tweets.

4.1 Emotional values

Many academics have stressed the importance of the emotional values and the personality of the destinations in their branding (e.g. [36], [37]). Destinations can be described by the values of human personality [10]. In this study we have updated and adapted the well-known Aaker's *Brand Personality Scale* (BPS) [38] to the analysis of destination brands. In fact, Ekinici and Hosany were the first authors that applied the BPS to the analysis of brand destinations [8]. They showed that personality dimensions have a positive impact on the preferences of potential tourists, as a strong and well-defined personality improves the image of the destination and the intentions to visit it. Thus, they concluded that a proper management of emotional values is vital for effective positioning and differentiation.

In the study reported in this document Aaker's BPS has been slightly adapted to the analysis of the emotional values associated to a travel destination. The five main dimensions of analysis of the personality and the emotional values of a destination are sincerity, excitement, competence, sophistication and ruggedness. Each of them has been divided into a set of

categories, which in turn have been refined to several sub-categories, represented by a set of terms. The whole template of analysis is shown on Table 4.

Emotional value	Category	Subcategory
Sincerity	Down-to-earth	Family-oriented
		Down-to-earth
		Sustainable
	Honest	Calm
		Real
		Traditional
	Wholesome	Honest
		Original
		Wholesome
		Quality of Life
	Cheerful	Happiness
		Sentimental
		Friendly
Excitement	Daring	Trendy
		Daring
		Exciting
		Exotic
		Fashionable
	Spirited	Cool
		Spirited
		Dynamic
		Vital
		Fresh
	Imaginative	Young
		Sensorial
		Unique
		Imaginative
		Creative

	Up-to-date	Up-to-date
		Independent
		Contemporary
	Cosmopolitan	Cosmopolitan
		Tolerant
		Hospitable
Competence	Reliable	Reliable
		Hard-working
		Safe
		Rigorous
	Intelligent	Intelligent
		Technical
		Corporate
		Innovative
	Successful	Successful
		Leader
		Ambitious
		Powerful
Sophistication	Luxurious	Glamorous
		Luxurious
	Charming	Seductive
		Smooth
		Romantic
		Magical
Ruggedness	Outdoorsy	Outdoorsy
		Get-away
		Recreational
	Tough	Tough
		Rugged
		Non-Conformist

Table 4: Emotional Values

4.2 Selection of destinations and data retrieval

In order to select the cities to be analysed we considered the top 25 European destinations in 2014, according to Tripadvisor³. Table 5 reports some basic data about the official Twitter accounts of the main tourism office of these destinations (the data are from mid-November 2014). In 6 cases we were not able to identify such an account, and in 2 cases we had to resort to the account of the associated region (Tuscany for Florence, and Emilia Romagna for Rimini). In 5 cases, shown in green, most of the tweets were written in the main language of the destination and not in English. Two of the destinations (Prague and Lisbon) had a very low number of tweets (a few hundreds). It was finally decided to select the 12 destinations that had more than 1,000 English tweets, which are shown in orange on Table 5.



Figure 7: Current snapshot of @VisitBCN_EN twitter account

After that, we retrieved 3,000 tweets (or all the tweets, in the case of Tuscany and Zermatt) from each account, using functions of the API provided by Twitter.

³ <http://www.tripadvisor.co.uk/TravelersChoice-Destinations-cTop-g4>.

City	Twitter account	Tweets	Followers
Istanbul	-		
Rome	@Turismoromaweb	1191	6880
London	@visitlondon	21300	239000
Prague	@PragueEU	374	471
Paris	@ParisOTC	8639	31300
Berlin	@berlintourism	4330	17600
Florence	@DiscoverTuscany	1148	4287
Barcelona	@VisitBCN_EN	3141	6058
St Petersburg	@VisitPetersburg	1359	214
Budapest	@VisitBudapest	3398	7289
Lisbon	@TurismoDeLisboa	697	973
Venice	-		
Madrid	@Visita_Madrid	6947	15400
Amsterdam	@Iamsterdam	5323	68500
Krakow	-		
Vienna	@ViennaInfoB2B	3389	3255
Milan	@turismomilano	5147	21400
Athens	@CityofAthens	4238	8271
Zermatt	@zermatt_tourism	2673	6414
Urgup	-		
Dublin	@VisitDublin	6652	44400
Moscow	-		
Dalyan	-		
Edinburgh	@edinburgh	17600	35100
Rimini	@ERTourism	14000	6682

Table 5: Selection of Destinations

4.3 Syntactic and Semantic Analysis

In previous studies on content analysis of destination brands ([27], [39]) it has been pointed out that *names* usually provide information on the particular tourist attractions, *verbs* describe actions and types of tourism, and *adjectives* communicate the emotional responses; thus, this work is focuses on the adjectives [21] used in the tweets sent by the destination managers to their followers.

4.3.1 Thematic analysis

Initially, a *thematic analysis* has been applied, in which a standard natural language parser was applied to detect the adjectives and count their frequency of use. This process gives a general idea of the main qualifications employed by the destination, without limiting the analysis to a pre-existing template. As an example, Table 6, and Table 7 show the 20 adjectives used more frequently in the tweets from “Berlin”, and “London” respectively, and the number of tweets in which they were used. Due to the size limitation of tweets it is very common to use ungrammatical sentences or expressions, which may be wrongly analysed by the parser. That is why in Table 6, and Table 7 we can see terms that are not really adjectives, such as *anniversary*, *youtube* or *id*. Chapter 5 reports the results associated to this purely syntactic analysis of the tweets of the 12 selected destinations.

(new, 183)	(great, 124)	(best, 123)	(open, 64)
(free, 51)	(top, 50)	(beatiful, 43)	(next, 36)
(many, 35)	(last, 35)	(good, 34)	(former, 26)
(international, 24)	(special, 22)	(anniversary, 22)	(German, 20)
(id, 19)	(big, 18)	(green, 15)	(old, 15)

Table 6: Top Adjectives in Berlin Tweets

(new, 1085)	(free, 362)	(best, 261)	(next, 240)
(last, 188)	(top, 150)	(great, 133)	(youtube, 120)
(good, 118)	(special, 107)	(fantastic, 103)	(open, 99)
(big, 88)	(many, 87)	(national, 67)	(live, 66)
(beautiful, 61)	(musical, 59)	(British, 55)	(classic, 48)

Table 7: Top Adjectives in London Tweets

4.3.2 Semantic analysis

This part of the study involved a *semantic analysis* of the adjectives used in the tweets. The objective was to associate them to the categories of emotional values shown in Table 4. A direct syntactic mapping is not possible, as most of the adjectives do not appear directly as categories/subcategories. The idea is to try to search for a semantic similarity measure [40] between the adjectives and the categories/subcategories. This type of measures require the use of some kind of external structured knowledge base; WordNet⁴ has been chosen to be used in this work.

Ontology-based semantic similarity measures [40] rely on the topological structure of an ontology to calculate the degree of resemblance between two terms. The length of the path between the terms (considering hyponymy and hypernymy relationships) and their position in the hierarchy of concepts (i.e. their degree of generality) are the basic points taken into account by similarity functions. In this work we have used a well-known similarity measure, defined by Wu and Palmer [41] in equation (1).

Algorithm 2 (Semantic Similarity) used to calculate semantic similarity between two words, where the `getSynsets` function return the list of synonyms of the word that exist in WordNet, and the function `pos` return the part of speech tag of the synonym of the word.

⁴ See <http://wordnet.princeton.edu>

Algorithm 2: Semantic Similarity

```
getSemanticSimilarity (word1, word2) :  
    synsets1 = getSynsets(word1) // get the set of synonyms of word1  
    synsets2 = getSynsets(word2) // get the set of synonyms of word2  
    if synsets1 = {} or synsets2 = {}:   
        return 0  
    maxSim = 0  
    foreach syns1 in synsets1:  
        foreach syns2 in synsets2:  
            if pos(syns1) = pos(syns2):  
                maxSim = max(maxSim, simw&p(syns1, syns2))  
    return maxSim
```

In the case of this study, the main problem of this approach is that WordNet uses hypernymy relationships between nouns, and we wanted to compare the adjectives found on the tweets with the subcategories associated to the emotional values (which are also mostly adjectives). Thus, before applying the Wu-Palmer semantic similarity measure we had to transform its inputs to nouns. In the case of the subcategories, they were manually translated to the equivalent nouns (e.g. *friendly* and *ambitious* were transformed into *friend* and *ambition*). Concerning the adjectives appearing in the tweets, they were automatically transformed into nouns using the algorithm 3 “Nontify”. The function getNounsDRF in this algorithm return all the possibilities of nouns sorted by probability based on the derivationally related form in WordNet, and the function getNounsAttr is based on the property attribute of synonyms in WordNet.

Algorithm 3: Nountify

nountify(adj) :

//load the Adjective => Noun dictionary

adjs_nouns = loadAdjsNounsDict()

// if adj already existed in the dictionary of adjective => nouns

if adj in adjs_nouns:

 return adjs_nouns[adj]

// else use WordNet to transform adj to closes noun

nouns = getNounsDRF(adj) *// get the closest nouns based on derivationally related form*

if nouns = {}: *// if this method failed*

 nouns = getNounsAttr(adj) *// get the closest nouns based on synsets*

attributes

if nouns = {}: *// if this method failed, assume that the noun is equal to the adjective*

 return adj

else:

// add this item to the dictionary and return the noun

updateAdjsNounsDict(adj, nouns[0], adjs_nouns)

return nouns[0]

Table 8 shows some examples of the translation of the adjectives into nouns, we can note that there is some good translation results for instance *wonderful* $\xrightarrow{\text{translated to}}$ *wonder*, and some bad results such as, *Magical* $\xrightarrow{\text{translated to}}$ *wizard*, we assume that this example is a bad example of translation because the *Magical* adjective in the context of tourist domain may means “beautiful or delightful in such a way as to seem removed from everyday life. e.g. - it

was a magical evening of pure nostalgia”, however wizard is a noun which means in general “a man who has magical powers, especially in legends and fairy tales”. In the context of computer domain “*wizard*” means “a help feature of a software package that automates complex tasks by asking the user a series of easy-to-answer questions” and both of those meanings are not related to the meaning of the adjective “*magical*” when it is used in the tourism domain.

Adjective	Noun
New	Newness
Best	Goodness
Wonderful	Wonder
Famous	Fame
Creative	Creativity
Traditional	Tradition
Magical	Wizard

Table 8: Examples of adjective to noun translation

After this pre-processing stage it was possible to apply the Wu-Palmer similarity measure to compare the emotional subcategories and the tweet adjectives.

In this way, we can assess if the destination communicates certain emotional values and if there is a strategy behind the communication of the brand (for example, a repeated use of terms like *love*, *honeymoon*, *passion* or *romance* could be interpreted as a strategy towards the promotion of a *romantic* destination). The detailed results of the semantic analysis of the 12 selected destinations are shown in the next chapter.

4.4 Summary

Three steps followed in the methodology of analysis. First “*Definition of the emotional values associated to a destination brand*” in this step an updated and adapted version of the well-known Aaker’s *Brand Personality Scale* (BPS) [38] has been used to the analysis of destination brands. Second “*Selection of the destinations to be analysed, and retrieval of the set of tweets*” in this step 12 destinations out of the top 25 European destinations in 2014,

according to Tripadvisor that had more than 1,000 English tweets have been selected (which were shown in orange on Table 5). Finally “*Syntactic and semantic analysis of the content of the tweets*”. Thematic analysis has been applied by count the frequency of the detected adjectives. The aim of this process is to give a general idea of the main qualifications employed by the destination, after that a well-known similarity measure, defined by Wu and Palmer [41] has been used to calculate the semantic similarity between the adjectives and the list of emotional values. The final output of this analysis process is associate the adjectives to the most likely emotional value.

Chapter 5

Case Study

This chapter comments the results obtained in the analysis of the tweets sent by the 12 selected destinations. First we focus on the analysis of the frequency of appearance of adjectives, and then we consider the semantic link between those adjectives and the emotional subcategories associated to a destination brand.

5.1 Syntactic analysis

In the first part of the study we focused on the adjectives that had been used at least 20 times by any of the cities in their tweets (recall that we are considering 3000 tweets from each city, so we are requiring a minimum frequency of appearance of 20/3000, i.e. once in every 150 tweets).

Table 9 shows the adjectives that were used 20 times at least by 3 different destinations. The top 8 adjectives were repeatedly used by at least 9 of the cities: *new*, *great*, *best*, *beautiful*, *free*, *good*, *top* and *open*. As we can see, they are very general and they do not provide any kind of territorial identification or distinction. In particular, *new*, *great* and *best* are heavily used by all the cities.

There are also several adjectives that are used by many destinations, like *special* (7), *gorgeous* (4), *wonderful* (4) or *big* (4). Although these adjectives are attractive, they do not perform any relevant role in the construction of a distinguishing destination brand, as they are not pointing out any particularity of the location. This fact probably shows that there is not any well-defined brand communication strategy or that the destination is trying to associate values that are too general and too similar to those of other competing cities.

Apart from the adjectives shown in Table 9, there were also 12 adjectives used at least 20 times by 2 destinations and 50 adjectives used at least 20 times by one single city. These few terms are heavily used by only 1-2 places, so they could provide a distinctive value to the territory and show the existence of a communication strategy. However, a close analysis

reveals that these terms are often associated to the location (*German, Scottish, Italian, Hungarian, Dutch, Austrian, Viennese, British, thisisedinburgh, tfloficial, pest*), to a particular event (*edrestaurantfest, edchefstable, zermatlive*) or to a very general concept (*famous, biggest, major*).

Table 9 also shows, for each city, the number of adjectives that they have used more than 20 times. London clearly stands out with 48 adjectives. 15 of them were only used more than 20 times by London, showing a communication strategy in their use (e.g. *classical, royal, exclusive, historic, festive*). However, we should keep in mind that a well-defined branding strategy should focus on a moderate number of identifying features which try to convey the personality and identity of the territory.

The use of many different adjectives may hinder the definition of a clear, unique and distinctive destination brand. On the opposite end of the spectrum, Florence (Tuscany) only uses 6 adjectives more than 20 times, and they happen to be the top 5 adjectives plus *special*; thus, this territory does not seem to follow a proper branding strategy. We can observe interesting intermediate cases, like Athens. Although it employs the top 9 generic adjectives and it does not use uniquely many specialised adjectives, the set of more frequently used adjectives (*traditional, cultural, local, national, classic*) certainly seems to follow a communication strategy coherent with its intended brand. These terms are quite interrelated and they communicate a clear identity of the territory. The relationship between the employed adjectives is a key aspect in the coherence and consistency of the communication of a distinguishing brand.

	Berl	Ath	Tus	Edi	Rim	Ams	Vie	Bar	Bud	Dub	Lon	Zer
	16	19	6	28	20	21	28	18	15	20	48	10
new												
great												
best												
beautiful												
free												
good												
top												
open												
many												
last												
special												
next												
international												
gorgeous												
full												
wonderful												
traditional												
delicious												
big												
old												
fantastic												
latest												
first												
cultural												
local												
fabulous												
national												
available												

Table 9: Frequently Used Adjectives

5.2 Semantic analysis

In the second part of the study the Wu-Palmer semantic similarity measure (equation 1, algorithm 1) was applied to each pair <tweet adjective, emotional subcategory>, after their transformation into nouns, using the Nountify (algorithm 2). This process provides a similarity value, between 0 and 1, for each pair of terms. Table 10 shows the three most similar adjectives to some emotional categories that have been found by this process.

Emotional Category	Most Similar Adjectives
Real	Real(1), Actual (0.92), True (0.86)
Independent	Independent (1), Separate (0.92), Whole (0.73)
Traditional	Traditional (1), Central (0.83), Educational (0.83)
Smooth	Gracious (0.94), Medieval (0.82), Nice(0.82)
Up-to-date	Contemporary (1), Modern (1), Current (1)
Sentimental	Mushy (1), Emotional (0.91), Dramatic (0.83)

Table 10: Snapshot of emotional category and the 3 most similar adjectives

Given a particular city we can see, for each adjective used in its tweets, which is the most similar emotional value, and then we can count how many adjectives associated to each value have been used by the city in the communication of its brand.

The results of the analysis are shown on Table 11. For each city we have only counted those adjectives that have been used at least 3 times in the tweets of the destination and which have a semantic similarity higher or equal to 0.7 with an emotional subcategory. For instance, Amsterdam used at least 3 times 5 adjectives which have been related to the concept “*traditional*”. It is possible to notice that there are several categories which have been mentioned by most of the cities (in these cases, the cells in orange show the cities which have been linked to the concept more often, whereas the cells in green show those cities that have not been associated to that value). For example Rimini, London and Berlin are the cities more linked to “*sentimental*”, whereas Florence and Barcelona do not seem to have used any

adjective related to this concept. Some values have only been associated to 1-2 cities, so they can be taken to be their distinctive features; they are shown in yellow on Table 11. For instance, Budapest is the only city linked to the “*family-oriented*” value.

Table 11 shows the following results:

- In general, there are not many differences between the values associated to each of the cities. This is not very surprising because, as we had detected in the syntactic analysis, these destinations tend to use the same adjectives.
- The first emotional value of Table 4 (*sincerity*) is the best covered by the adjectives, with many relationships to the categories *calm*, *real*, *traditional*, *honest* and *sentimental* by all the cities. *Excitement* is also well-covered, with adjectives associated to *trendy*, *daring*, *unique*, *up-to-date* and *independent*. On the other hand, the *competence* emotional values are hardly mentioned (*innovative* being the exception).
- Berlin, Rimini (Emilia Romagna) and especially London stand out as the cities that have employed more different adjectives in more subcategories. On the other side of the spectrum, Florence (Tuscany) and Zermatt show a lack of adjectives in several of the most common categories.
- London is clearly the city that has used more adjectives highly related to the emotional values (54), followed by Rimini/Emilia Romagna (39) and Athens (36). Florence/Tuscany (13) and Zermatt (12) also have the lowest positions on this respect.
- London (20) and Vienna (19) are the cities which have used adjectives linked to a higher number of subcategories. In contrast, Zermatt (9) and Florence/Tuscany (11) are the ones with a smaller variety of subcategories.

Subcategory	Ber	Flo	Vie	Zer	Ath	Bar	Lon	Dub	Bud	Ams	Rim	Edi
Family-oriented									1			
Down-to-earth												
Sustainable												
Calm	4	1	4	2	2	2	5	4	2	4	3	3
Real	2	1	2	3	3	2	3	3	2	2	4	2
Traditional	3	1	4	2	4	3	8	1	3	5	5	3
Honest	4	2	4	2	6	4	5	4	4	4	4	4
Original												
Wholesome							1					1
Quality of Life												
Happiness												
Sentimental	3	3	2	1	1	3	4	2	2	1	5	2
Friendly												
Trendy	1	1	1	3	2	1	1	1	1	1	1	3
Daring	2	3	2	1	2	3	3	2	3	2	2	3
Exciting	1						1			1	1	1
Exotic					1							
Fashionable												
Cool												
Spirited												
Dynamic												
Vital												1
Fresh												
Young	1		1			1	1			1		
Sensorial												
Unique	1	1	1	1	3	4	3	2	1	2	3	2
Imaginative												
Creative												
Up-to-date	1	1	2	3	2	2	2	1	1	2	1	3
Independent	1	1	1		1		3	2			1	
Contemporary												
Cosmopolitan												
Tolerant												
Hospitable					1						1	
Reliable												
Hard-working												
Safe												
Rigorous						1						
Intelligent												
Technical												
Corporate				1						1		
Innovative	1	3	2	3	1	1	1	1	1	1	1	1
Successful												
Leader												
Ambitious											1	
Powerful												
Glamorous	2	1	1	1	1	1	4	2	1	2	3	2
Luxurious	1		1			1	1				1	1
Seductive			1									
Smooth	2	2	2	1	3	2	2	2	2	3	3	2
Romantic												
Magical												
Outdoorsy												
Get-away	1	1	2	3	2	2	3	1	3	2	2	2
Recreational	1		1			2	1	1				
Tough			1		1	1	2			1		
Rugged												
Non-Conformist												

Table 11: Semantic Analysis Results

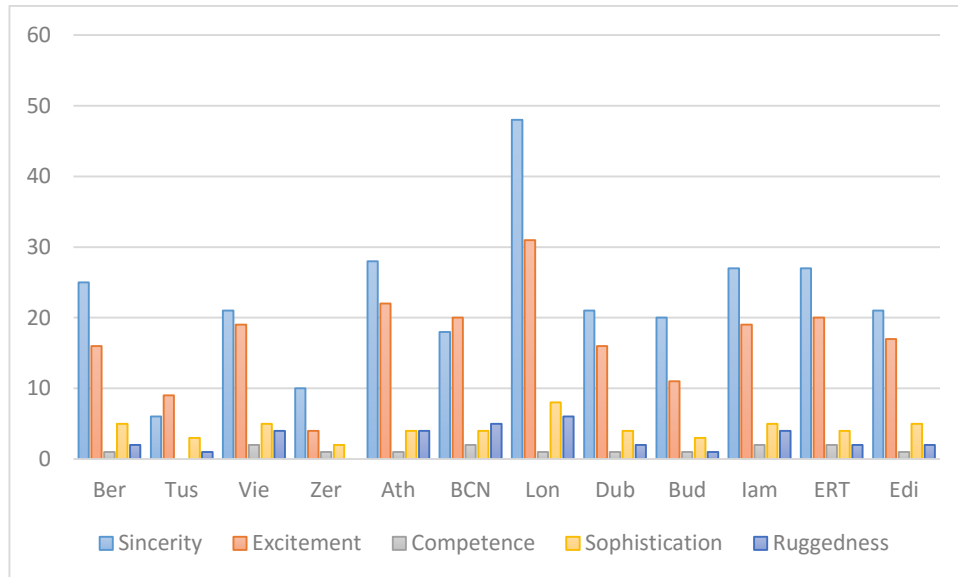


Figure 8: Histogram of the main topics (main category of the emotional value) $T = 3$ times and $S = 0.5$

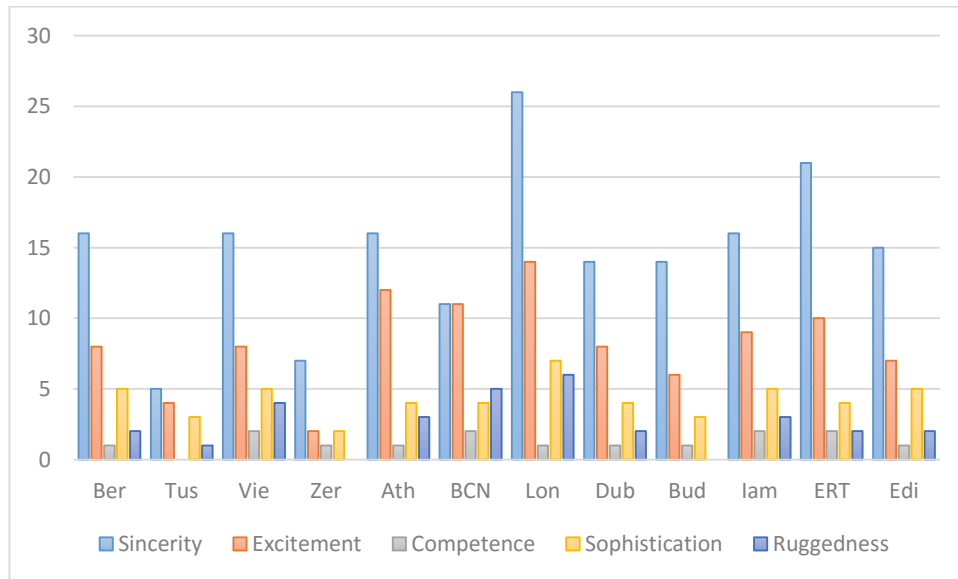


Figure 9: Histogram of the main topics (main category of the emotional value) $T = 3$ times and $S = 0.7$

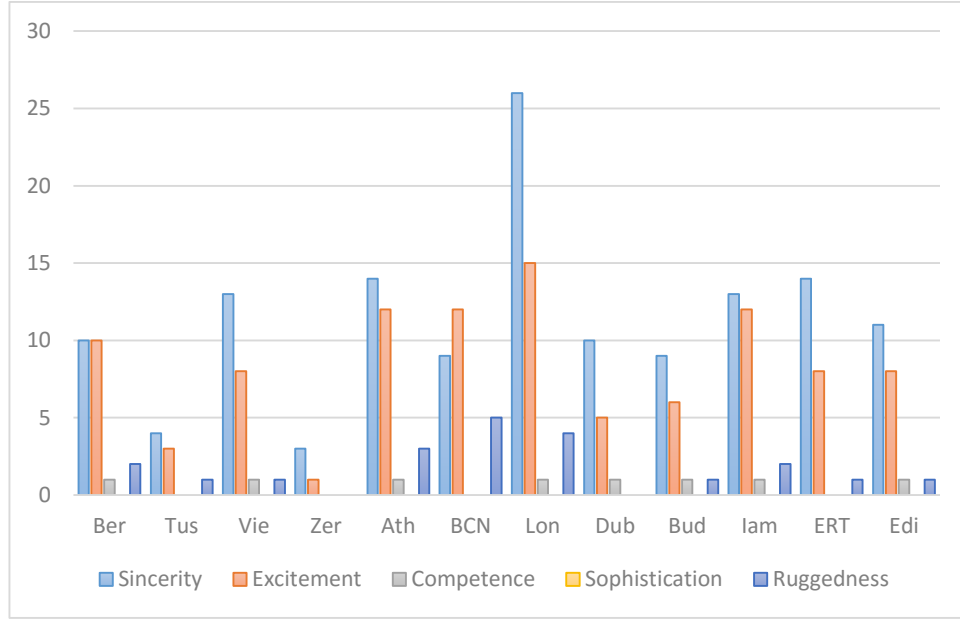


Figure 10: Histogram of the main topics (main category of the emotional value) $T = 5$ times and $S = 0.6$

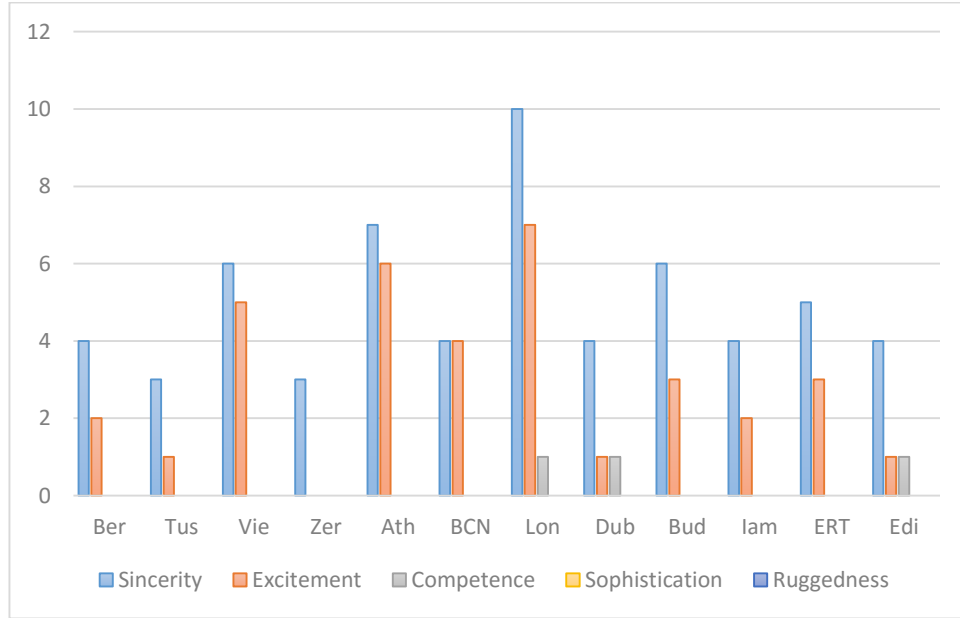


Figure 11: Histogram of the main topics (main category of the emotional value) $T = 10$ times and $S = 0.7$

Figures (8 – 11) show the histogram of the number of adjectives that have been used by the destination cities for each main topic (the main category of the emotional values) according to the number of adjectives that have been used at least T times in the tweets of the destination and which have a semantic similarity higher or equal to S with an emotional subcategory (for

instance in Figure 8 (Figure 1 we have used $T = 3$ and $S = 0.5$). Those figures show the following results:

- The first emotional value of Table 4 (sincerity) is the best covered by the adjectives, by all the cities. *Excitement* is also well-covered. On the other hand, the *competence* emotional values are hardly mentioned.
- London is the city that has used more adjectives highly related to the emotional values, followed by Rimini/Emilia Romagna and Athens. Florence/Tuscany and Zermatt also have the lowest positions on this respect.
- In general the high parameters (i.e. T and S are high) leads to low no of adjectives appears in the main categories, and vice versa.
- The parameter T has more effect than the parameter S on the results, for instance we can note that the results are almost same in case $T = 0.3$ and $S = 0.5$ or 0.7 (Figures 8, 9), however those results are seems to be different to the obtained results when $T = 5$ or 10 (Figures 9, 10).

5.3 Summary

The syntactical and semantic analysis (see Chapter 4) have been used in this case of study. The results that has been obtained showed the following aspects:

- The set of the 12 destinations (Table 5) tend to use the same adjectives.
- The first emotional value of Table 4 (sincerity) is the best covered by the adjectives.
- Berlin, Rimini (Emilia Romagna) and especially London stand out as the cities that have employed more different adjectives in more subcategories. On the other side of the spectrum, Florence (Tuscany) and Zermatt show a lack of adjectives in several of the most common categories.
- London is the city that has used more adjectives highly related to the emotional values, followed by Rimini/Emilia Romagna and Athens. Florence/Tuscany and Zermatt also have the lowest positions on this respect.

Chapter 6

Conclusion

The objective of the work was to study how the top European destinations use their Twitter accounts to communicate their brand and their emotional values with the aim to achieve a distinctive image. There is a strong match between the results of the syntactic and the semantic analysis. They both show that, in general, destinations do not seem to have a coherent communicative strategy to try to communicate emotional brand values that distinguishes them through the texts they publish in their tweets.

The syntactic analysis shows that, in general, the analysed destinations do not use specific adjectives to communicate their identity. They mostly use common and generic ones (e.g. *new*, *great*, *best*, *beautiful* or *free*). Many destinations want to be associated with these adjectives, but they do not contribute any identity or distinction. This fact supposes the wastage of the potentiality offered by Twitter, that it is an important communicative tool, to communicate the destination brands and to build their own image. The same conclusions can be drawn from the semantic analysis, as there are not many differences between the values associated to each of the cities.

The specific results of each of the different analysed destinations show that both analysis (syntactic and semantic) present quite homogeneous results. London is the city that uses more different adjectives more than 20 times, and most of them are used exclusively. The semantic analysis shows that it is also the destination that uses more different adjectives in more subcategories, and the destination that uses more adjectives highly related to the emotional values. This coincidence is not due to chance, and it shows that London certainly uses a strategy to communicate his brand and his identifying values.

The same happens to Berlin, Rimini and Athens, but in less proportion. After London, they are the destinations that use more adjectives highly related to the emotional values, and they are the cities that also communicate some identifying adjectives more than 20 times.

Therefore, it may be assumed that they have a brand strategy in their communication through Twitter.

On the contrary, Florence (Tuscany) and Zermatt only use the top adjectives, the ones more commonly employed by all the destinations. These adjectives are not identifiers of the territory. It is not a coincidence that these destinations are also the ones with a smaller variety of adjectives and with a smaller number of adjectives related to the emotional values.

It should be noted that the coincidence of the results among the syntactic and semantic analysis has lead us to draw some unexpected conclusions that are relevant for the study. Firstly, the results show that both kinds of techniques are suitable to analyse the communication of emotional values of destinations through social media. However, the semantic analysis makes a more detailed study, showing the relationships between the adjectives and the brand emotional values. Secondly, the results show that the semantic analysis actually measures the communication of brand values and, therefore, it confirms its effectiveness as a new tool in this field. Thirdly, the results show that the combination of these two kinds of analysis is complementary, because the relationships among the adjectives and the emotional values uncovered by the semantic analysis show the existence of brand communication strategies more evidently than the simple accounting of adjectives of the syntactic analysis. These findings, which were not expected at the start of the study, lead us to finally recommend the combination of syntactic and semantic analysis in the assessment of the communication of the brand values of tourist destinations. However, some limitations of this methodology should be considered: the ambiguity of natural language, the mistakes associated to the parsing errors or to the automatic translation from adjectives into nouns, and the lack of analysis of other useful information provided in the tweets (e.g. images, emoticons or linked websites).

The following scientific paper has been submitted to the conference ISKE15 <http://iske15.csie.ntust.edu.tw/>.

- A.Moreno, M.Jabreel, A.Huertas, “*Automatic analysis of the communication of tourist destination brands through social networks*”.

In our future work we aim to overcome some of these limitations. We also intend to analyse the opinion of the tourists on their tweets about a destination, and compare the emotional values they show with the ones that the destination tried to convey. The emotional values transmitted in the tweets should also be compared with the intended image projected by the website of the destination.

References

- [1] Morgan, NJ, Pritchard, A., & Piggott, R. (2003). Destination branding and the role of the stakeholders: The case of New Zealand. *Journal of Vacation Marketing*, 9 (3), 285-299.
- [2] Blain, C., Levy, SE & JRB Ritchie (2005). Destination Branding: Insights and Practices from Destination Management Organizations. *Journal of Travel Research*, 43 (4): 328-38.
- [3] Govers, R. & Go, FM (2009). Place branding: Glocal, virtual and physical identities, Constructed, imagined and experienced. Hampshire: Palgrave Macmillan.
- [4] Hankinson, G. (2004). Relational network brands: Towards a conceptual model of place brands. *Journal of Vacation Marketing*, 10: 2, 109-121.
- [5] Keller, KL (1993). Conceptualising, Measuring and Managing Customer-Based Brand Equity. *Journal of Marketing*, 57, 1-22.
- [6] Sirgy, MJ (1982). Self-concept in Consumer Behaviour: A Critical Review. *Journal of Consumer Research*, 9: 287-300.
- [7] Fournier, S. (1998). Consumers and Their Brands: Developing Relationship Theory in Consumer Research. *Journal of Consumer Research*, 24: 343-73.
- [8] Ekinci, Y. & Hosany, S. (2006). Destination Personality: An Application of Brand Personality to Tourism Destinations. *Journal of Travel Research*, 45: 127-139.
- [9] Choi, S. Lehto, XY & Morrison, AM (2007). Destination image representation on the web: content analysis of Macau travel related websites. *Tourism Management*, 28: 118-129.
- [10] Xiang, Z. & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, 31 (2), 179-188.
- [11] Leung, D., Law, R, van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing*, 30 (1-2), 3-22.
- [12] Hays, S. Page, SJ, & Buhalis, D. (2013). Social media as a marketing tool destination: its use by National Tourism Organisations. *Current Issues in Tourism* 16 (3), 211-239.

- [13] Fotis, J., Buhalis, D. & Rossides, N., 2012. Social media use and impact during the holiday travel planning process. In: Fuchs, M. Ricci, F. and Cantoni, L. (Eds.), *Communication and Information Technologies in Tourism 2012* (pp. 13-24). Vienna, Austria: Springer
- [14] Yoo, KH & Gretzel, U. (2011). Background and impacts of trust in travel-related consumer-generated media. *Information Technologies & Tourism* 12 (2), 139-152.
- [15] Zhang, L., Pan, B., Smith, W., & LI, X. (2010). Travelers' use of online reviews and recommendations: a qualitative study. *Information Technology & Tourism*, 11 (2), 157-167.
- [16] Huertas, A., Setó D. and Míguez, MI (2015). Communication of tourist destinations through the media. *Professional Information* 24 (1): 15-21.
- [17] Mackay KJ & Couldwell, CM (2004): Using visitor-employed photography to investigate destination image. *Journal of Travel Research*, 42 (4): 390-396.
- [18] Kyle G. & Chick, G. (2007). The Social Construction of a Sense of Place. *Leisure Sciences*, 29 (3): 209-25.
- [19] Huertas, A. & Mariné, E. (2015). Destination Brand Communication through the Social Media. What Contents trigger most Reactions of Users?. *Information and Communication Technologies in Tourism 2015*. Lugano, Switzerland. Springer.
- [20] Dolnicar, S. & Grün, B. (2013). Validity Measuring Destination Survey in Image Studies. *Journal of Travel Research*, 52 (1): 3-14.
- [21] Echtner, C. & Ritchie, JRB (1991). The meaning and measurement of destination image. *Journal of Tourism Studies*, 2 (2): 2-12.
- [22] Anholt, S. (2010). *Places: Identity, Image and Reputation*. Hampshire: Palgrave MacMillan.
- [23] Stepchenkova, S. & Eales, J.S. (2011). Destination Image as Quantified Media Messages. The Effect of News on Tourism Demand, *Journal of Travel Research*, 50(2): 198-212.
- [24] Chen, C.F. & Tsai, D.C. (2007). How Destination Image and Evaluative Factors Affect Behavioral Intentions? *Tourism Management*, 28(4): 115-122.
- [25] Pike, S. (2002). Destination Image Analysis- a review of 142 papers from 1973 to 2000. *Papers Tourism Management*, 23: 541-49.

- [26] Guerrero-Solé, F. & Fernández-Cavia, J. (2015). @Spain is Different: Co-branding Strategies Between Spanish National and Regional DMOs on Twitter. *Information and Communication Technologies in Tourism 2015*. Switzerland: Springer, pp. 309-320.
- [27] Stepchenkova, A., Kirilenko, AP & Morrison, AM (2009). Facilitating Content Analysis in Tourism Research. *Journal of Travel Research*, 47 (4): 454-469.
- [28] Li, X.R. & Stepchenkova, S. (2012). Chinese Outbound Tourists' Destination Image of America: Part I, *Journal of Travel Research*, 51(3): 250-266.
- [29] Roberts, CW (2000). A Conceptual Framework for Quantitative Text Analysis. *Quality & Quantity*, 34: 259-74.
- [30] Dickinger, A. & Lalicic, L. (2015). This city is Absolutely Fun and Trendy. A Destination Brand Personality Analysis in a Web 2.0 Setting. *Information and Communication Technologies in Tourism 2015*. Switzerland: Springer, pp. 321-333.
- [31] De Moya, M. & Jain, R. (2013). When tourists are your Friends: Exploring the Brand personality of Mexico and Brazil on Facebook, *Public Relations Review*, 39: 23-29.
- [32] Serna, A., Marchiori, E., Gerrikagoitia, J.K., Alzua-Sorzabal, A. & Cantoni, L. (2015). An Auto-Coding Process for Testing the Cognitive-Affective and Conative Model of Destination Image. *Information and Communication Technologies in Tourism 2015*. Switzerland: Springer, p. 111-124.
- [33] Andéhn, M., Kazemini, A., Lucarelli, A. & Sevin, E. (2014). User-generated place brand equity on Twitter: The dynamics of brand associations in social media. *Place Branding and Public Diplomacy*, 10 (2): 132-144.
- [34] Alexa Index (2014). <http://www.alexa.com/topsites> (last access April 2015).
- [35] Hvass, K.A., Munar, A.M. (2012). The takeoff of social media in tourism". *Journal of Vacation Marketing*, 18(2):93-103.
- [36] Henderson, JC (2000). Selling Places: The New Asia-Singapore Brand. *Journal of Tourism Studies*, 11 (1): 36-44.
- [37] Morgan, N., Pritchard, A. & Piggott, R. (2002). New Zealand, 100% Pure- The Creation of a Powerful Niche Destination Brand, *Journal of Brand Management*, 9: 335-354.

- [38] Aaker, J.L. (1997). Dimensions of Brand Personality. *Journal of Marketing Research*, 34: 347-356.
- [39] Stepchenkova, A. & Morrison, A.M. (2006). The destination image of Russia: From the perspective online induced. *Tourism Management*, 27: 943-956.
- [40] Sánchez, D., Batet, M., Isern, D., Valls, A. (2012) Ontology-based semantic similarity: a new feature-based approach. *Expert Systems with Applications* 39 (9), pp. 7718-7728.
- [41] Wu, Z., & Palmer, M. (1994). Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics* (pp. 133–138). Stroudsburg, PA, USA.
- [42] Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge Engineering: Principles and Methods. *Data Knowl. Eng.*, 25(1-2), 161–197. doi:10.1016/S0169-023X(97)00056-6
- [43] Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5–6), 907–928. doi:10.1006/ijhc.1995.1081.
- [44] Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11, 93–136.
- [45] Cimiano, P. (2006a). *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.
- [46] Stumme, G., Ehrig, M., Handschuh, S., Hotho, A., Maedche, A., Motik, B., ... Zacharias, V. (2003). The {K}arlsruhe view on ontologies.
- [47] Gómez-Pérez, A., Fernández-López, M., & Corcho-García, O. (2004). *Ontological Engineering*. Springer/Heidelberg.
- [48] Bollegala, D., Matsuo, Y., & Ishizuka, M. (2007). Measuring semantic similarity between words using web search engines. In *WWW '07: Proceedings of the 16th international conference on World Wide Web* (pp. 757–766). New York, NY, USA: ACM. doi:http://doi.acm.org/10.1145/1242572.1242675.
- [49] Rubenstein, H., & Goodenough, J. B. (1965). Contextual Correlates of Synonymy. *Commun. ACM*, 8(10), 627–633. doi:10.1145/365628.365657.

- [50] Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*. doi:10.1080/01690969108406936.
- [51] Sánchez, D., Batet, M., Isern, D., & Valls, A. (2012a). Ontology-based semantic similarity: A new feature-based approach. *Expert Systems with Applications*, 39(9), 7718–7728. doi:10.1016/j.eswa.2012.01.082.
- [52] Jiang, J. J., & Conrath, D. W. (1997). Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. *CoRR*, cmp-lg/970.
- [53] Pirró, G., & Seco, N. (2008). Design, Implementation and Evaluation of a New Semantic Similarity Metric Combining Features and Intrinsic Information Content. In *Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part II on On the Move to Meaningful Internet Systems* (pp. 1271–1288). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-88873-4_25.
- [54] Marchiori E. & Onder I (2015). Reframing the Image of a Destination: A Pre-Post Study on Social Media Exposure Information and Communication Technologies in Tourism 2015. Switzeland: Springer, pp. 335-347.
- [55] Sevin H. E., (2014). Understanding cities through city brands: City branding as a social and semantic network. *Cities*, 38, 47-56.
- [56] Maxwell, J.A. (1997), “Designing a qualitative study”, in Bickman, L. and Rog, D.J. (Eds), *Handbook of Applied Social Research Methods*, Sage Publications, London, pp. 69-100.
- [57] Given, Lisa M. (2008). *The Sage encyclopedia of qualitative research methods*. Los Angeles, Calif.: Sage Publications. ISBN 1-4129-4163-6.