



Σχολή Μηχανικών

Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

## Διπλωματική Εργασία

Π.Μ.Σ. «Επιστήμη και Τεχνολογία της Πληροφορικής και των Υπολογιστών»

Ειδίκευση: Λογισμικού και Πληροφοριακών Συστημάτων

# *Εξόρυξη Δεδομένων, Κοινωνικά Δίκτυα & Τουρισμός*



Εκπόνηση διπλωματικής εργασίας: **Ιωάννα Μόκα (mcse19024)**

Επιβλέποντες καθηγητές: **Χρήστος Σκουρλάς**

**Αικατερίνη Μαρινάγη**

Αθήνα

Δεκέμβριος 2020

**ΙΩΑΝΝΑ ΜΟΚΑ**

**ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ, ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ  
ΤΟΥΡΙΣΜΟΣ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Εξεταστική Επιτροπή**

Χρήστος Σκουρλάς, Καθηγητής, Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών, Πανεπιστήμιο Δυτικής Αττικής, Επιβλέπον Μέλος

Αικατερίνη Μαρινάγη, Καθηγήτρια, Γενικό Τμήμα, Γεωπονικό Πανεπιστήμιο Αθηνών, Συνεπιβλέπον Μέλος

Βασίλειος Μάμαλης, Καθηγητής, Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών, Πανεπιστήμιο Δυτικής Αττικής, Μέλος Εξεταστικής Επιτροπής



Σχολή Μηχανικών  
Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

---

**ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ**

Η κάτωθι υπογεγραμμένη Ιωάννα Μόκα του Αθανασίου, με αριθμό μητρώου mcse19024 φοιτήτρια του Προγράμματος Μεταπτυχιακών Σπουδών “Επιστήμη και Τεχνολογία της Πληροφορικής και των Υπολογιστών” του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Η Δηλούσα

## ΠΡΟΛΟΓΟΣ

Τα μέσα κοινωνικής δικτύωσης την τελευταία δεκαετία κέρδισαν αξιοσημείωτη προσοχή. Αυτό αποδίδεται κυρίως στο γεγονός ότι η πρόσβαση στα κοινωνικά δίκτυα, όπως Twitter, Google+, Facebook, Instagram κ.λπ., είναι οικονομικά προσιτή. Όλο και περισσότεροι άνθρωποι ενδιαφέρονται και βασίζονται στα κοινωνικά δίκτυα για να αντλήσουν πληροφορίες και τη γνώμη άλλων χρηστών για μεγάλη ποικιλία θεμάτων.

Επιπλέον, ο διαρκώς αυξανόμενος ανταγωνισμός στον τομέα της παροχής τουριστικών υπηρεσιών αναγκάζει εταιρείες και ιδρύματα, οι οποίες παρέχουν αυτές τις υπηρεσίες, να σκέφτονται συνεχώς τις δυνατότητες αύξησης της ανταγωνιστικότητας, της αποδοτικότητας και της παραγωγικότητάς τους. Απώτερος στόχος αυτής της πρωτοβουλίας είναι να αυξήσει την ικανοποίηση του πελάτη αναφορικά με τις παρεχόμενες υπηρεσίες, οι οποίες ασφαλώς σχετίζονται άμεσα με την αύξηση του μεριδίου αγοράς και τη βιωσιμότητα της εταιρείας. Ένας από τους σημαντικότερους τρόπους ενίσχυσης της ανταγωνιστικότητας είναι η συλλογή, ανάλυση και εφαρμογή δεδομένων μεγάλης κλίμακας. Η πηγή αυτών των δεδομένων είναι κυρίως το διαδίκτυο καθώς και υπηρεσίες και λύσεις που βασίζονται στον Παγκόσμιο Ιστό, όπως το Google Trends, η μηχανή αναζήτησης Google, το Google Analytics, το Flickr, τα κοινωνικά δίκτυα, κ.λπ. Ως εκ τούτου, κρίνεται ιδιαίτερα σημαντικό να μεταφράσουμε τα συναισθήματα που εκφράζουν οι χρήστες στο διαδίκτυο σε χρήσιμες πληροφορίες χρησιμοποιώντας τεχνικές εξόρυξης δεδομένων μεγάλης κλίμακας. Επομένως, οι εν λόγω τεχνικές οφείλουν να είναι ικανές να χειριστούν τα τρία κυρίαρχα ζητήματα, κατά την εξόρυξη διαδικτυακών δεδομένων, τα οποία είναι το μέγεθος, ο θόρυβος και ο δυναμισμός των εν λόγω δεδομένων.

Στην παρούσα εργασία με τίτλο «Εξόρυξη Δεδομένων, Κοινωνικά Δίκτυα και Τουρισμός» επιχειρείτε στο πρώτο μέρος μια πλήρης εννοιολογική προσέγγιση των μεγάλων δεδομένων, εξόρυξη δεδομένων, κοινωνικών δικτύων και του τουρισμού. Στο δεύτερο μέρος παρουσιάζονται οι τεχνικές εξόρυξης δεδομένων που χρησιμοποιούνται επί του παρόντος για να αντληθεί χρήσιμη πληροφορία από τα διαδικτυακά δεδομένα, όπως τα κείμενα και οι φωτογραφίες, τα οποία αναρτούν οι χρήστες στα κοινωνικά δίκτυα αναφορικά με τον τουρισμό. Επιπροσθέτως, γίνεται

μα απόπειρα να εντοπιστούν οι προκλήσεις και οι μελλοντικές κατευθύνσεις της εξόρυξης δεδομένων από τα κοινωνικά δίκτυα για την βελτιστοποίηση του τουριστικού κλάδου. Η παρούσα εργασία καταλήγει παραθέτοντας συμπεράσματα για το ως άνω θέμα.

**Λέξεις κλειδιά:** Μεγάλα Δεδομένα ή Μεγάλης Κλίμακας Δεδομένα, Εξόρυξη Δεδομένων, Κοινωνικά Δίκτυα, Τουρισμός, Τεχνικές Εξόρυξης Δεδομένων, Προκλήσεις, Μελλοντικές Κατευθύνσεις

## **ΕΥΧΑΡΙΣΤΙΕΣ**

Θα ήθελα να εκφράσω τις θερμές και ειλικρινείς ευχαριστίες μου στους επιβλέποντες καθηγητές της παρούσας διπλωματικής εργασίας και συγκεκριμένα τον κ. **Σκουρλά Χρήστο** και την κα. **Μαρινάγη Αικατερίνη** για την εμπιστοσύνη που μου έδειξαν και μου ανέθεσαν το τόσο ενδιαφέρον και επίκαιρο θέμα εργασίας. Η καθοδήγησή τους, τόσο κατά τη διάρκεια των μεταπτυχιακών μου σπουδών όσο και κατά την εκπόνηση της διπλωματικής μου εργασίας, ήταν πολύτιμη και συνετέλεσε καίρια στην ολοκλήρωση της.

## ΠΕΡΙΛΗΨΗ

Όλοι οι φορείς που καταπιάνονται με την τουριστική βιομηχανία και παρέχουν πολιτιστικές και τουριστικές πληροφορίες καθώς και άλλες συναφείς υπηρεσίες πρέπει να είναι σε θέση να καθορίσουν σχέσεις μεταξύ τουριστικών δραστηριοτήτων και προτιμήσεων των τουριστών έτσι ώστε να σχεδιάσουν και να προσφέρουν ελκυστικότερα τουριστικά πακέτα. Πόσο μάλλον στην Ελλάδα όπου ο τουρισμός είναι ο πιο σημαντικός οικονομικός κλάδος της και η συνολική συνεισφορά του στην οικονομία ανέρχεται μεταξύ 37 και 45 δις, δηλαδή περισσότερο από το 20% του ΑΕΠ της χώρας. Χρειάζεται επομένως μια λεπτομερή ανάλυση της λειτουργίας του τουριστικού κλάδου ώστε να ληφθούν αποφάσεις στρατηγικού σχεδιασμού στο μέλλον. Για να καταστεί αυτό εφικτό, είναι επιτακτική, περισσότερο από ποτέ, η ανάγκη εφαρμογής της εξόρυξης δεδομένων. Σύμφωνα με τους Bose και Mahapatra (2001) από το Πανεπιστήμιο του Χονγκ Κονγκ, η εξόρυξη δεδομένων μπορεί να χρησιμοποιηθεί σε τρεις βασικούς τομείς του τουρισμού, συγκεκριμένα:

- ✦ Πρόβλεψη εξόδων για τον τουρισμό
- ✦ Ανάλυση του τουριστικού προφίλ σε μια ομάδα στόχο και
- ✦ Πρόβλεψη του αριθμού των τουριστικών αφίξεων

Τα κοινωνικά δίκτυα, τα οποία την τελευταία δεκαετία έχουν εισχωρήσει πολύ ενεργά στη ζωή μας, αποτελούν ένα πλούσιο «ορυχείο» εξόρυξης δεδομένων μεγάλης κλίμακας για τον τουρισμό. Ο όρος κοινωνικά δίκτυα θα μπορούσε γενικά να γίνει κατανοητός ως εφαρμογές Διαδικτύου, οι οποίες εμπεριέχουν το περιεχόμενο που δημιουργεί ο χρήστης αναφορικά με τις εμπειρίες που αποκόμισε από έναν ταξιδιωτικό προορισμό έτσι ώστε να μπορούν εύκολα και άλλοι χρήστες να έχουν πρόσβαση σε αυτό. Τα κοινωνικά δίκτυα περιλαμβάνουν έτσι μια μυριάδα εφαρμογών που επιτρέπουν σε μεμονωμένους χρήστες να δημοσιεύουν και να επισημαίνουν τις εντυπώσεις, τις εμπειρίες, τις περιπέτειες τους αλλά και τις ανεπιβεβαίωτες φήμες στο Διαδίκτυο. Αυτή η μεγάλη συλλογή πληροφοριών ωφελεί όλο και περισσότερους τους μελλοντικούς τουρίστες, καθιστώντας παράλληλα όλο και πιο δύσκολο για τις επιχειρήσεις να εμπορεύονται τον τουρισμό.

Η παρούσα εργασία έχει ως στόχο να αναδείξει την ιδιαίτερη πρόκληση μα και ευκαιρία να αναλυθεί ένας τεράστιος όγκος δεδομένων, που προσφέρεται μέσω των κοινωνικών δικτύων, και ο οποίος μοντελοποιείται, επιλέγεται και διερευνάται για να προσδιοριστούν κατανοητές και χρήσιμες πληροφορίες για τη βελτιστοποίηση του τουριστικού κλάδου. Χωρίζεται σε δύο μέρη, όπου στο πρώτο μέρος φέρνει τον αναγνώστη σε μια πρώτη επαφή με την επιστήμη των μεγάλων δεδομένων, την Εξόρυξη Δεδομένων (Data Mining), τα κοινωνικά δίκτυα και τον τουρισμό. Παρέχεται μια πλήρης εννοιολογική προσέγγιση των παραπάνω όρων, καθώς επίσης αναλύονται τα χαρακτηριστικά των μεγάλων δεδομένων, τα στάδια της εξόρυξης δεδομένων και οι τεχνικές που χρησιμοποιούνται για να εξορύξουμε χρήσιμη πληροφορία από αυτά. Επιπροσθέτως, παρουσιάζεται η εξέλιξη και η σημασία των κοινωνικών δικτύων ως προς την εξόρυξη δεδομένων και οι προκλήσεις που καλούμαστε να αντιμετωπίσουμε. Το πρώτο μέρος ολοκληρώνεται με την εξέλιξη της ανάλυσης του τουριστικού κλάδου, καθώς αρχίζει να γίνεται σημαντικό τμήμα των εθνικών οικονομιών μετά τον Β' Παγκόσμιο Πόλεμο, και πως οδηγηθήκαμε στην ανάλυσή του βάσει δεδομένων.

Στο δεύτερο μέρος αναλύεται η εφαρμογή των μεγάλων δεδομένων στον τουρισμό και τα πλεονεκτήματα που απορρέουν από τη χρήση τους. Εξετάζεται ο ρόλος και η σημασία της εξόρυξης δεδομένων μεγάλης κλίμακας μέσα από τα κοινωνικά δίκτυα για την αύξηση της αποδοτικότητας και της αποτελεσματικότητας του τουριστικού κλάδου. Επιπλέον, παρουσιάζονται οι τεχνικές εξόρυξης δεδομένων που χρησιμοποιούνται για την εξόρυξη χρήσιμης γνώσης από το περιεχόμενο που δημιουργούν/αναρτούν οι χρήστες στα κοινωνικά δίκτυα αναφορικά με τον τουρισμό, επί παραδείγματι κείμενα και φωτογραφίες. Τέλος, γίνεται μια απόπειρα να εντοπιστούν οι προκλήσεις και οι μελλοντικές κατευθύνσεις της εξόρυξης δεδομένων από τα κοινωνικά δίκτυα για την βελτιστοποίηση του τουριστικού κλάδου.

Η παρούσα εργασία καταλήγει παραθέτοντας συμπεράσματα για το ως άνω θέμα. Εν κατακλείδι, η αξιοποίηση των μεγάλων διαδικτυακών δεδομένων για την βελτιστοποίηση του προσφερόμενου τουριστικού πακέτου βρίσκεται ακόμη σε πρώιμο στάδιο, παρόλα αυτά μεγάλες ταξιδιωτικές εταιρίες επιχειρούν να εκμεταλλευτούν όλο και περισσότερο τα οφέλη που προκύπτουν μέσω της εξόρυξης



των δεδομένων μεγάλης κλίμακας που απλόχερα προσφέρουν οι χρήστες των κοινωνικών δικτύων και της αναδυόμενης πληροφορίας που απορρέει από αυτά.

## ABSTRACT

All actors involved in the tourism industry and providing cultural and tourism information as well as other related services should be able to establish relationships between tourism activities and tourist preferences in order to design and offer more attractive tourism packages. Especially in Greece, where tourism is its most important economic sector and its total contribution to the economy amounts to between 37 and 45 billion, ie more than 20% of the country's GDP. A detailed analysis of the operation of the tourism industry is therefore needed in order to make strategic planning decisions in the future. To make this possible, the need for data mining is more urgent than ever. According to Bose and Mahapatra (2001) from the University of Hong Kong, data mining can be used in three key areas of tourism, specifically:

- ⊕ Expenditure forecast for tourism
- ⊕ Analysis of the tourism profile in a target group and
- ⊕ Forecast of the number of tourist arrivals

Social media, which in the last decade have very actively penetrated into our lives, are a rich "mine" of large-scale data mining for tourism. The term social media could generally be understood as Internet applications, which include user-generated content related to the experiences gained from a travel destination so that other users can easily access it. Social media thus include a myriad of applications that allow individual users to post and highlight their impressions, experiences, adventures and even unconfirmed rumors on the Internet. This large collection of information is increasingly benefiting prospective tourists, while making it increasingly difficult for businesses to market tourism.

This paper aims to highlight the particular challenge and opportunity to analyze a huge amount of data, offered through social media, and which is modeled, selected and explored to identify understandable and useful information for the optimization of the tourism industry. It is divided into two parts, where in the first part it brings the reader into a first contact with the science of big data, Data Mining, social media and tourism. A complete conceptual approach to the above terms is provided, as well as the characteristics of big data, the stages of data mining and the techniques used to

extract useful information from them. In addition, the evolution and importance of social media in terms of data mining and the challenges we are called to face are presented. The first part concludes with the evolution of the analysis of the tourism industry, as it begins to become an important part of national economies after World War II, and how we were led to its analysis based on data.

The second part analyzes the application of big data in tourism and the benefits arising from their use. The role and importance of large-scale data mining through social media to increase the efficiency and effectiveness of the tourism industry is examined. In addition, the data mining techniques used to extract useful knowledge from the content created / posted by users on social media related to tourism are presented, for example texts and photos. Finally, an attempt is made to identify the challenges and future directions of data mining from social media to optimize the tourism industry.

This paper concludes by presenting conclusions on the above issue. In conclusion, the utilization of large online data to optimize the tourism package offered is still in its infancy, however large travel companies are increasingly trying to take advantage of the large-scale data mining generously offered by social media users and the information that emerges from them.

## ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ.....	4
ΕΥΧΑΡΙΣΤΙΕΣ.....	6
ΠΕΡΙΛΗΨΗ.....	7
ABSTRACT.....	10
ΕΙΣΑΓΩΓΗ.....	16
ΜΕΘΟΔΟΣ ΚΑΙ ΔΟΜΗ ΕΡΓΑΣΙΑΣ.....	20

### ΜΕΡΟΣ Ι

#### Δεδομένα Μεγάλης Κλίμακας – Εξόρυξη Δεδομένων – Κοινωνικά Δίκτυα - Τουρισμός

##### Κεφάλαιο 1

Δεδομένα Μεγάλης Κλίμακας – Μεγάλα Δεδομένα.....	24
1.1 Ορισμός.....	24
1.2 Χαρακτηριστικά.....	26

##### Κεφάλαιο 2

Εξόρυξη Δεδομένων.....	29
2.1 Ορισμός.....	29
2.2 Στάδια Εξόρυξης Δεδομένων.....	29
1. Συλλογή Δεδομένων (Data Collection).....	30
2. Προεπεξεργασία Δεδομένων (Preprocessing).....	30
3. Μετασχηματισμός Δεδομένων (Transformation).....	31
4. Εξόρυξη Δεδομένων (Data Mining).....	31
5. Διερμηνεία και Αξιολόγηση (Interpretation/Evaluation).....	31
2.3 Τεχνικές Εξόρυξης Δεδομένων.....	31

### **Κεφάλαιο 3**

<b>Κοινωνικά Δίκτυα.....</b>	<b>34</b>
3.1 Εξέλιξη Κοινωνικών Δικτύων.....	34
3.2 Τα Κοινωνικά Δίκτυα και η Σημασία τους.....	37
3.3 Εξόρυξη Δεδομένων στα Κοινωνικά Δίκτυα.....	39
3.4 Προκλήσεις Εξόρυξης Δεδομένων από τα Κοινωνικά Δίκτυα.....	40

### **Κεφάλαιο 4**

<b>Τουρισμός.....</b>	<b>43</b>
4.1 Εξέλιξη Ανάλυσης Τουριστικού Κλάδου.....	43
4.2 Ανάλυση Τουριστικού Κλάδου Βάσει Δεδομένων.....	48

## **ΜΕΡΟΣ II**

### **Σύγχρονες Τεχνικές Εξόρυξης Δεδομένων Μεγάλης Κλίμακας στον Τομέα του Τουρισμού**

### **Κεφάλαιο 5**

<b>Εφαρμογή Δεδομένων Μεγάλης Κλίμακας στον Τουρισμό.....</b>	<b>52</b>
5.1 Εισαγωγή.....	52
5.2 Η Έννοια των Δεδομένων Μεγάλης Κλίμακας στον Τουρισμό.....	54
5.3 Τα Πλεονεκτήματα Χρήσης Δεδομένων Μεγάλης Κλίμακας στον Τουρισμό.....	55
5.4 Η Συμπεριφορά των Καταναλωτών.....	58
5.5 Μηχανισμοί Ανατροφοδότησης.....	60

### **Κεφάλαιο 6**

<b>Εξόρυξη Δεδομένων Μεγάλης Κλίμακας από τα Κοινωνικά Δίκτυα για τον Τουρισμό.....</b>	<b>62</b>
6.1 Ο Ρόλος της Εξόρυξης Δεδομένων στα Κοινωνικά Δίκτυα.....	62

6.2 Η Σημασία των Κοινωνικών Δικτύων για τον Τουρισμό.....	63
--	----

## **Κεφάλαιο 7**

<b>Τεχνικές Εξόρυξης Δεδομένων από τα Κοινωνικά Δίκτυα για τον Τουρισμό.....</b>	<b>68</b>
--	-----------

7.1 Εισαγωγή.....	68
-------------------	----

7.2 Περιεχόμενο Προερχόμενο από τον Χρήστη (UGC).....	69
---	----

7.2.1 Διαδικτυακά Δεδομένα Κειμένου και Τεχνικές Εξόρυξης.....	70
--	----

7.2.1.1 Προεπεξεργασία Δεδομένων.....	71
---------------------------------------	----

1) Data cleaning/cleansing.....	71
---------------------------------	----

2) Tokenization.....	72
----------------------	----

3) Word Stemming.....	72
-----------------------	----

4) POST.....	72
--------------	----

7.2.1.2 Ανακάλυψη Προτύπων.....	73
---------------------------------	----

1) Latent Dirichlet Allocation (LDA).....	73
---	----

2) Sentiment analysis.....	74
----------------------------	----

3) Statistical analysis.....	75
------------------------------	----

4) Clustering και Categorization.....	76
---------------------------------------	----

5) Text summarization.....	77
----------------------------	----

6) Dependency modeling.....	77
-----------------------------	----

7.2.1.3 Προκλήσεις και Μελλοντικές Κατευθύνσεις.....	79
--	----

7.2.2 Διαδικτυακά Δεδομένα Φωτογραφιών και Τεχνικές Εξόρυξης.....	80
---	----

7.2.2.1 Τεχνικές Εξόρυξης Διαδικτυακών Φωτογραφιών.....	81
---	----

7.2.2.2 Προκλήσεις και Μελλοντικές Κατευθύνσεις.....	84
--	----

## **Κεφάλαιο 8**

<b>Συμπεράσματα.....</b>	<b>85</b>
--------------------------	-----------

<b>Βιβλιογραφία/Αναφορές.....</b>	<b>87</b>
-----------------------------------	-----------

## ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

<b>Εικόνα 1.</b> Μέσα κοινωνικής δικτύωσης.....	16
<b>Εικόνα 2.</b> Εκτιμώμενα δεδομένα που παράγονται στα κοινωνικά δίκτυα κάθε λεπτό.....	18
<b>Εικόνα 3.</b> Διεθνείς αφίξεις τουριστών το 2015.....	20
<b>Εικόνα 4.</b> Βασικά στάδια Εξόρυξης Δεδομένων.....	30
<b>Εικόνα 5.</b> Δένδρο Αποφάσεων.....	32
<b>Εικόνα 6.</b> Συσταδοποίηση.....	33
<b>Εικόνα 7.</b> Το σύμπαν τεχνολογιών του Web 2.0.....	34
<b>Εικόνα 8.</b> Χαρακτηριστικά και δυνατότητες εφαρμογών κοινωνικής δικτύωσης.....	36
<b>Εικόνα 9.</b> Διαδικασία εξόρυξης διαδικτυακών δεδομένων κειμένου για τον τουρισμό.....	70
<b>Εικόνα 10.</b> Στάδια καθαρισμού δεδομένων.....	71
<b>Εικόνα 11.</b> Παράδειγμα Tokenization.....	72
<b>Εικόνα 12.</b> Παράδειγμα Word Stemming.....	72
<b>Εικόνα 13.</b> Παράδειγμα τεχνικής POST.....	73
<b>Εικόνα 14.</b> Ανάλυση συναισθήματος.....	75
<b>Εικόνα 15.</b> Παράδειγμα συσταδοποίησης με τρεις συστάδες.....	76
<b>Εικόνα 16.</b> Παράδειγμα περίληψης κειμένου.....	77
<b>Εικόνα 17.</b> Μοντελοποίηση εξάρτησης μεταξύ τουριστών, κατοίκων και παρόχων τουριστικών προϊόντων.....	78
<b>Εικόνα 18.</b> Διαδικασία χρήσης διαδικτυακών φωτογραφικών δεδομένων στην έρευνα της τουριστικής βιομηχανίας.....	82

## ΕΙΣΑΓΩΓΗ

Το διαδίκτυο έχει επιφέρει σημαντικές αλλαγές ως προς την αναζήτηση ταξιδιωτικών πληροφοριών από τους τουρίστες, τον σχεδιασμό των δικών τους ταξιδιών και τον διαμοιρασμό των ταξιδιωτικών τους εμπειριών με άλλους (Buhalis & Law, 2008; Hays, Page & Buhalis, 2013; Xiang & Gretzel, 2010). Η έλευση των τεχνολογιών Web 2.0 ανέπτυξε νέα κανάλια για τους χρήστες του διαδικτύου ως προς το να επικοινωνούν και να συνεργάζονται μεταξύ τους καθώς και να μοιράζονται περιεχόμενο που δημιουργούν οι ίδιοι (UGC) με άλλους χρήστες όπως διαδικτυακές κριτικές, φωτογραφίες και βίντεο σε διάφορες πλατφόρμες κοινωνικών δικτύων. Το εν λόγω διαδικτυακό περιεχόμενο χρησιμοποιείται ευρέως στη βιομηχανία του τουρισμού και της φιλοξενίας. Οι τουρίστες τείνουν να μοιράζονται τις δικές τους ταξιδιωτικές εμπειρίες μέσω των κοινωνικών δικτύων, τα σημαντικότερα εκ των οποίων αποτυπώνονται στην εικόνα 1.



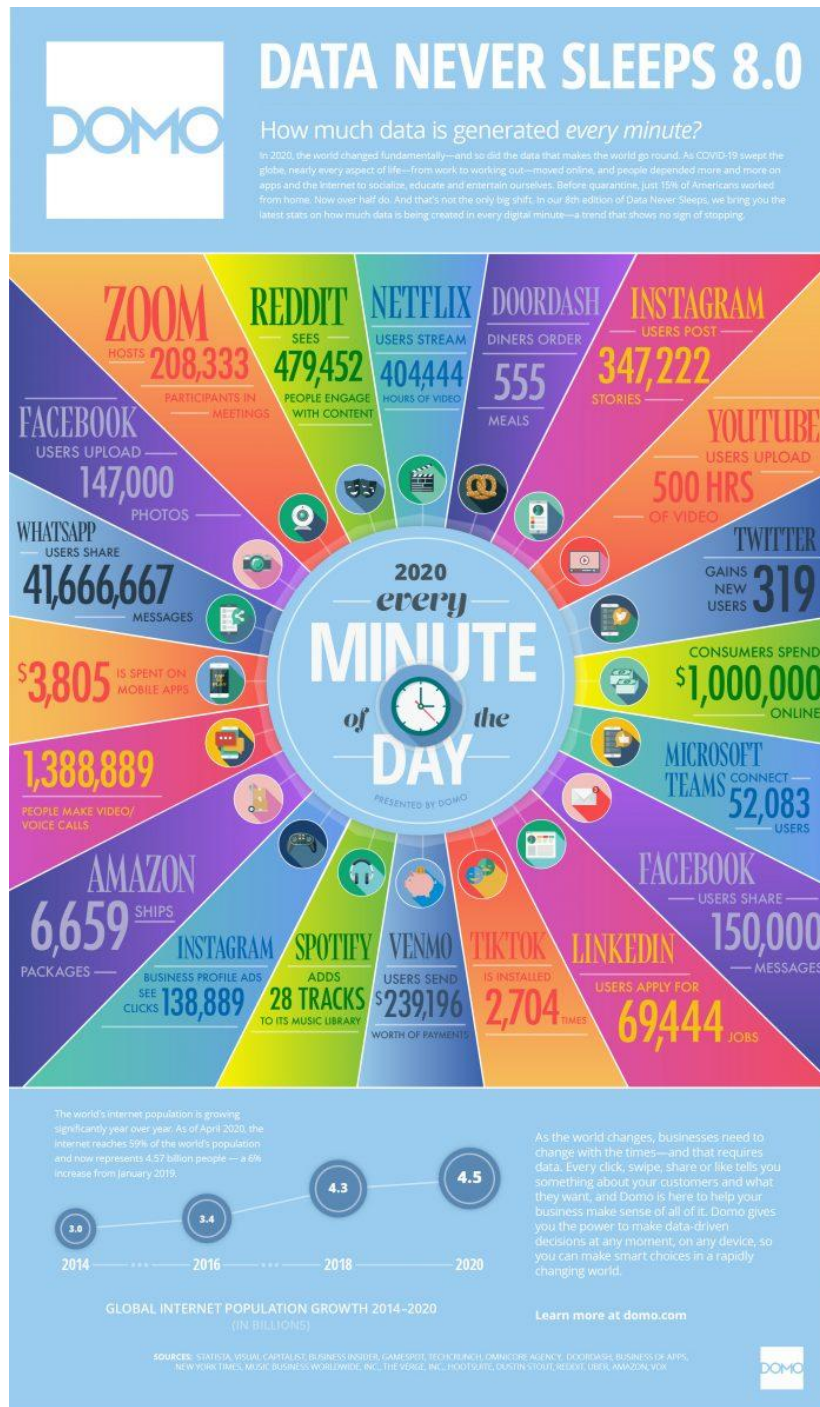
*Εικόνα 1. Μέσα κοινωνικής δικτύωσης (Adedoyin-Olowe, Gaber & Stahl)*



Οι εν λόγω περιγραφές των τουριστικών εμπειριών, τις οποίες μοιράζονται ενεργά οι χρήστες, έχουν ληφθεί ευρέως υπόψη ως διαδικτυακές ταξιδιωτικές κριτικές και περιλαμβάνουν κυρίως κριτικές ξενοδοχείων, κριτικές για εστιατόρια και κριτικές για αξιοθέατα (Xiang, Schwartz, Gerdes & Uysal, 2015). Οι διαδικτυακές κριτικές είναι οι πιο δημοφιλείς πηγές πληροφοριών για τους τουρίστες προκειμένου να λάβουν ταξιδιωτικές πληροφορίες και να σχεδιάσουν το ταξίδι τους, αναφορικά κυρίως με τη κράτηση εισιτηρίων και ξενοδοχείων (Li & Yi, 2014; Zhang, Song, Cai, Liu, Pujol & Feng, 2016a, 2016b). Σε αντίθεση με τα περισσότερα καταναλωτικά προϊόντα, η οργάνωση του ταξιδιού περιλαμβάνει κάτι περισσότερο από μια απλή απόφαση αγοράς, και η τουριστική κατανάλωση είναι συνυφασμένη με την προθυμία του κοινού. Οι καταναλωτές επηρεάζονται εύκολα από τις απόψεις άλλων ανθρώπων και η προθυμία τους να αναζητήσουν τις απόψεις και τις εμπειρίες των ομότιμων καταναλωτών πριν από την αγορά ενός προϊόντος είναι απόλυτα συναφής (García-Pablos, Cuadros & Linaza, 2016; Guo, Barnes & Jia, 2017; Hu, Chen & Chou, 2017; Schuckert, Liu & Law, 2015a). Τα δεδομένα της τουριστικής βιομηχανίας δείχνουν ότι περίπου το 77% των υποψήφιων ταξιδιωτών είτε «πάντα» είτε «συνήθως» δεν λαμβάνουν αποφάσεις μέχρι να διαβάσουν διαδικτυακές κριτικές άλλων χρηστών (Ye, Law & Gu, 2009). Οι ταξιδιώτες μπορούν να ελαχιστοποιήσουν το κόστος ενός ταξιδιού και να αποκτήσουν έμμεσες εμπειρίες αγορών με την περιήγηση σε ταξιδιωτικές πληροφορίες (π.χ. κυβερνητικοί ιστότοποι τουρισμού, διαδικτυακές πύλες ταξιδιών κ.λπ.), μειώνοντας με αυτόν τον τρόπο την αβεβαιότητα και επιτυγχάνοντας μια ευχάριστη ψυχολογική εμπειρία (Lian & Yu, 2017; Xiang, Schwartz, Gerdes & Uysal, 2015; Ye, Law, Gu & Chen, 2011). Τα κοινωνικά δίκτυα είναι επίσης ένα μέρος για τους φορείς που ασχολούνται με τον τουρισμό να ξεκινήσουν συνομιλία με τους καταναλωτές τους, να οραματιστούν και να χαράξουν μακροπρόθεσμες τουριστικές στρατηγικές, να βελτιώσουν και να προωθήσουν το τουριστικό προϊόν τους και γενικότερα να λάβουν καλύτερες επιχειρηματικές αποφάσεις.

Ωστόσο, τα δεδομένα που δημιουργούνται από τους χρήστες των κοινωνικών δικτύων εξελίσσονται πολύ γρήγορα και είναι τόσο περίπλοκα που χρειάζονται ειδικές αναλύσεις, τεχνολογικές υποδομές και εργαλεία για την αντιμετώπιση αυτών των λεγόμενων «μεγάλων δεδομένων». Το περιοδικό Forbes (Newman, 2016) δημοσίευσε τον Αύγουστο του 2016 τις 10 τάσεις ψηφιακής μεταμόρφωσης για το 2017 και μέσα

σε αυτά ήταν φυσικά και τα Μεγάλα Δεδομένα. Γενικότερα όμως, τα τελευταία χρόνια θα λέγαμε ότι σε όλες τις λίστες που αφορούν σαρωτικές τεχνολογίες που θα αλλάξουν την πραγματικότητα του επιχειρηματικού κόσμου προβάλλονται σε περίοπτη θέση τα Μεγάλα Δεδομένα. Μαζικά δεδομένα δημιουργούνται κάθε λεπτό στα κοινωνικά δίκτυα όπως μας αποκαλύπτει η εικόνα 2.



Εικόνα 2. Εκτιμώμενα δεδομένα που παράγονται στα κοινωνικά δίκτυα κάθε λεπτό (<https://www.domo.com/learn/data-never-sleeps-8>, 2020)

Τα τεράστια αυτά δεδομένα συλλέγονται συνεχώς σε αυτούς τους ιστότοπους και η χρήση παραδοσιακών μεθόδων για την ερμηνεία τους είναι κάτι παραπάνω από απαγορευτική. Κρίνεται επιτακτικό να χρησιμοποιηθούν εργαλεία ικανά να αναλύσουν τα δεδομένα μεγάλης κλίμακας όπως για παράδειγμα η έκφραση απόψεων/συναισθημάτων που είναι βασικό χαρακτηριστικό των κοινωνικών δικτύων. Οι τεχνικές εξόρυξης δεδομένων έχουν αποδειχθεί ικανές να χρησιμοποιηθούν για την εξόρυξη μεγάλων δεδομένων που δημιουργούνται στα κοινωνικά δίκτυα. Αυτό γίνεται εφικτό μέσω της εξαγωγής πληροφοριών από μεγάλα σύνολα δεδομένων που δημιουργούνται στα εν λόγω δίκτυα και τη μετατροπή τους σε κατανοητή δομή για περαιτέρω χρήση. Σύμφωνα με τον Berry (1997), η εξόρυξη δεδομένων είναι η διαδικασία αναζήτησης και ανάλυσης δεδομένων προκειμένου να βρεθούν σιωπηρές αλλά δυνητικά χρήσιμες πληροφορίες. Περιλαμβάνει την επιλογή, τη διερεύνηση και τη μοντελοποίηση μεγάλων ποσοτήτων δεδομένων για την αποκάλυψη προηγουμένως άγνωστων προτύπων και τελικά κατανοητών πληροφοριών από μεγάλες βάσεις δεδομένων. Χρησιμοποιεί μια ευρεία οικογένεια υπολογιστικών μεθόδων που περιλαμβάνουν τη στατιστική ανάλυση, τα δέντρα αποφάσεων, τα νευρωνικά δίκτυα, την πρόκληση και βελτίωση των κανόνων και την απεικόνιση γραφικών (Brachman, 1996). Η εξόρυξη δεδομένων καθορίστηκε από τους Turban, Aronson, Liang και Sharda (2007) ως μια διαδικασία που χρησιμοποιεί τεχνικές στατιστικής, μαθηματικής, τεχνητής νοημοσύνης και μηχανικής μάθησης για την εξαγωγή και αναγνώριση χρήσιμων πληροφοριών και στη συνέχεια τη συγκέντρωση γνώσεων από μεγάλες βάσεις δεδομένων. Στα κοινωνικά δίκτυα χρησιμοποιούνται προηγμένες αναλυτικές τεχνικές (π.χ. ισχυρή επεξεργασία φυσικής γλώσσας (NLP), μηχανική μάθηση κ.λπ.), οι οποίες αποκαλύπτουν κρυμμένα μοτίβα, μη καθορισμένους συσχετισμούς και άλλες χρήσιμες πληροφορίες, καθιστώντας τις παραδοσιακές μεθόδους μη υλοποιήσιμες (Elgendy & Elragal, 2014).

Εν κατακλείδι, τα κοινωνικά δίκτυα αποτελούν ένα πλούσιο «ορυχείο» εξόρυξης δεδομένων μεγάλης κλίμακας για τον τουρισμό. Χρησιμοποιώντας σύνθετες τεχνικές εξόρυξης δεδομένων θα είμαστε σε θέση να φέρουμε στην επιφάνεια όλη αυτή την πολύτιμη και χρήσιμη πληροφορία που εσωκλείουν, βελτιστοποιώντας με αυτόν τον τρόπο τις παρεχόμενες τουριστικές υπηρεσίες.

## ΜΕΘΟΔΟΣ ΚΑΙ ΔΟΜΗ ΕΡΓΑΣΙΑΣ

Αναμφισβήτητα, ο τουρισμός αποτελεί έναν από τους σημαντικότερους οικονομικούς τομείς σε παγκόσμια κλίμακα. Ο τουρισμός, ο οποίος αντιπροσωπεύει το 10-12% του Παγκόσμιου Ακαθάριστου Προϊόντος, αποτελεί ίσως το σημαντικότερο οικονομικό τομέα, με πολυεπίπεδες επιπτώσεις στην κοινωνία, στο φυσικό και πολιτιστικό περιβάλλον (UNWTO, 2015). Οι διεθνείς αφίξεις τουριστών αυξήθηκαν κατά 4,4% το 2015 και έφτασαν συνολικά τα 1.184 εκατομμύρια, σύμφωνα με την τελευταία μέτρηση World Tourism Barometer, του Παγκόσμιου Οργανισμού Τουρισμού.



*Εικόνα 3. Διεθνείς αφίξεις τουριστών το 2015  
(UNWTO, 2016)*

Για τον Nikodem Szumilo, επίκουρο καθηγητή των οικονομικών στο UCL σημαντικό ρόλο στην ανάπτυξη της τουριστικής βιομηχανίας παίζει η επέκταση των ψηφιακών υπηρεσιών, που επιτρέπουν να διαβάζει κανείς κριτικές στο Ίντερνετ ή να κάνει διαδικτυακές κρατήσεις για δωμάτια και εισιτήρια σε πολύ χαμηλότερες τιμές σε σχέση με το παρελθόν.

Αποδεδειγμένα, η νέα γενιά του Παγκόσμιου Ιστού αποτέλεσε ένα κύριο μέσο για την προώθηση της ψηφιακής πληροφορίας, δίνοντας κίνητρα και τεχνολογίες στους χρήστες για εύκολη καταχώρηση δεδομένων και μετα-δεδομένων από αυτούς. Έτσι,

οι χρήστες παύουν, πλέον, να έχουν έναν στατικό/παθητικό ρόλο, όπως στην προηγούμενη γενιά του Ιστού όπου απλά επισκέπτονταν σελίδες χωρίς μεγάλες δυνατότητες δικής τους συνεισφοράς, και γίνονται οι ίδιοι ενεργοί, δυναμικοί ρυθμιστές των διαδικτυακών δεδομένων και εφαρμογών, δρώντας στον Παγκόσμιο Ιστό, όπως δρούσαν παλιά στον υπολογιστή τους. Η νέα, αυτή, γενιά του Παγκόσμιου Ιστού είναι γνωστή με τον όρο Web 2.0. και χρησιμοποιείται για να περιγράψει μια ομάδα τεχνολογιών και ιστοχώρων, που επιτρέπουν την εφαρμογή συνεργατικών μεθόδων δημιουργίας και οργάνωσης περιεχομένου (O'Reilly, 2007). Οι βασικοί παράγοντες για την επιτυχία των εφαρμογών Web 2.0 είναι η μαζική, μη κατευθυνόμενη συμμετοχή και η συνεχής εξέλιξη. Πράγματι, ο δυναμικός χαρακτήρας των εν λόγω εφαρμογών προκύπτει από χρήστες που προσθέτουν νέο περιεχόμενο, διαρκώς, χωρίς καθοδήγηση ή πλάνο. Η πληθώρα αυτή του περιεχομένου οδηγούν σε έναν μεγάλο όγκο δεδομένων και μετα-δεδομένων που παράγονται από μη προβλέψιμη δραστηριότητα και συνιστούν πολύτιμη πηγή για εξόρυξη γνώσης.

Συνεπώς, η παρούσα εργασία αναπτύχθηκε σε δύο βασικά στάδια:

- 1) Σύλληψη της ιδέας – Καθορισμός γενικού στόχου και επιλογή τρόπου εργασίας.
- 2) Ανάλυση βασικής ορολογίας και σημαντικών εννοιών μέσω μελέτης της βιβλιογραφίας.

Το πρώτο στάδιο αφορούσε τη σύλληψη της ιδέας η οποία αναφέρονταν στην εξόρυξη δεδομένων, τα κοινωνικά δίκτυα και τον τουρισμό. Καθορίστηκε επίσης και ο συνολικός τρόπος προσέγγισης του εν λόγω θέματος καθώς και ο γενικός στόχος της εργασίας. Όπως έχει ήδη αναφερθεί, η παρούσα εργασία έχει ως στόχο να αναδείξει την ιδιαίτερη πρόκληση μα και ευκαιρία να αναλυθεί ένας τεράστιος όγκος δεδομένων, που προσφέρεται μέσω των κοινωνικών δικτύων, και ο οποίος μοντελοποιείται, επιλέγεται και διερευνάται για να προσδιοριστούν κατανοητές και χρήσιμες πληροφορίες για τη βελτιστοποίηση του τουριστικού κλάδου. Τα κοινωνικά δίκτυα, τα οποία αποτελούν ένα κομμάτι των εφαρμογών Web 2.0, έχουν εισχωρήσει πολύ ενεργά στη ζωή μας την τελευταία δεκαετία και θεωρούνται ένα πλούσιο «ορυχείο» δεδομένων και μετα-δεδομένων μεγάλης κλίμακας. Τα εν λόγω δίκτυα επιτρέπουν τη δημιουργία και την ανταλλαγή περιεχομένου που παράγεται από χρήστες. Είτε μεμονωμένα είτε σε επίπεδο κοινότητας συν-δημιουργούν, συζητούν,

και τροποποιούν περιεχόμενο που παράγεται από τους ίδιους. Αυτό το πολύτιμο «ορυχείο» δεδομένων είναι πολύ χρήσιμο, εκτός των άλλων, και για τον τουρισμό, καθώς οι χρήστες αναρτούν ταξιδιωτικές κριτικές, φωτογραφίες και βίντεο. Ο τουριστικός κλάδος θα πρέπει να χρησιμοποιήσει υπέρ του όλη αυτή τη γνώση, ούτως ώστε να βελτιστοποιήσει το προσφερόμενο ταξιδιωτικό προϊόν. Για να καταστεί αυτό εφικτό κρίνεται απαραίτητο να εφαρμοστούν τεχνικές εξόρυξης δεδομένων μεγάλης κλίμακας για την αυτόματη εξαγωγή χρήσιμων και πολύτιμων πληροφοριών και σχέσεων από τεράστιες ποσότητες δεδομένων και μετα-δεδομένων.

Στο δεύτερο στάδιο αναλύονται βασικές έννοιες και ορολογίες μέσω της εκτενούς μελέτης της βιβλιογραφίας. Η παρούσα εργασία χωρίζεται σε δύο μέρη, όπου στο πρώτο μέρος επιχειρείται η εννοιολογική προσέγγιση των βασικών θεμάτων που πραγματεύεται η εν λόγω διπλωματική και φέρνει τον αναγνώστη σε μια πρώτη επαφή με την επιστήμη των μεγάλων δεδομένων ή δεδομένων μεγάλης κλίμακας, την Εξόρυξη Δεδομένων (Data Mining), τα κοινωνικά δίκτυα και τον τουρισμό. Επίσης, αναλύονται τα χαρακτηριστικά των μεγάλων δεδομένων, τα στάδια της εξόρυξης δεδομένων καθώς και οι τεχνικές που χρησιμοποιούνται για να εξορύξουμε χρήσιμη πληροφορία από αυτά. Επιπροσθέτως, παρουσιάζεται η εξέλιξη και η σημασία των κοινωνικών δικτύων ως προς την εξόρυξη χρήσιμης πληροφορίας από αυτά, όπως και τις προκλήσεις που καλούμαστε να αντιμετωπίσουμε, δεδομένης της πολυπλοκότητας του όλου εγχειρήματος. Το πρώτο μέρος ολοκληρώνεται με την εξέλιξη της ανάλυσης του τουριστικού κλάδου, καθώς αρχίζει να γίνεται σημαντικό τμήμα των εθνικών οικονομιών μετά τον Β' Παγκόσμιο Πόλεμο, και πως οδηγηθήκαμε στην ανάλυσή του βάσει δεδομένων, αποσκοπώντας στην κατανόηση του ταξιδιωτικού κοινού και στην προώθηση και βελτιστοποίηση της τουριστικής βιομηχανίας.

Στο δεύτερο μέρος αναλύεται η εφαρμογή των μεγάλων δεδομένων στον τουρισμό και τα πλεονεκτήματα που απορρέουν από τη χρήση τους σε έναν τόσο σημαντικό τομέα. Επίσης, μελετάται η συμπεριφορά του καταναλωτή-τουρίστα, από την σκοπιά εκείνου, που διαμοιράζεται π.χ. κριτικές, εμπειρίες, φωτογραφίες, βίντεο και από την άλλη πλευρά, εκείνου που χρησιμοποιεί τα προαναφερθέντα για να σχεδιάσει το δικό του ταξίδι. Στην συνέχεια, εξετάζεται ο ρόλος και η σημασία της εξόρυξης δεδομένων μεγάλης κλίμακας μέσα από τα κοινωνικά δίκτυα για την αύξηση της αποδοτικότητας και της αποτελεσματικότητας του τουριστικού κλάδου. Επιπλέον,

παρουσιάζονται οι σημαντικότερες τεχνικές εξόρυξης δεδομένων που χρησιμοποιούνται για την εξόρυξη χρήσιμης πληροφορίας, μέσω του περιεχομένου που δημοσιεύουν οι χρήστες στα κοινωνικά δίκτυα, αναφορικά με τον τουρισμό, επί παραδείγματι κείμενα και φωτογραφίες. Τέλος, γίνεται μια προσπάθεια να εντοπιστούν οι προκλήσεις και οι μελλοντικές κατευθύνσεις της εξόρυξης δεδομένων από τα κοινωνικά δίκτυα για την βελτιστοποίηση της τουριστικής βιομηχανίας.

Η παρούσα εργασία καταλήγει παραθέτοντας συμπεράσματα για το ως άνω θέμα. Εν κατακλείδι, αν και η αξιοποίηση των μεγάλων διαδικτυακών δεδομένων, για την βελτιστοποίηση του προσφερόμενου τουριστικού πακέτου, βρίσκεται ακόμη σε πρώιμο στάδιο, παρόλα αυτά, μεγάλες ταξιδιωτικές εταιρίες επιχειρούν να εκμεταλλευτούν όλο και περισσότερο τα οφέλη που προκύπτουν, μέσω της εξόρυξης των δεδομένων μεγάλης κλίμακας που απλόχερα προσφέρουν οι χρήστες των κοινωνικών δικτύων και της αναδυόμενης πληροφορίας που απορρέει από αυτά. Ο Βρετανός μαθηματικός Κλάιβ Χάμπι εξέφρασε για πρώτη φορά, το 2006, την άποψη ότι «Τα δεδομένα είναι το νέο πετρέλαιο!», δεκατέσσερα χρόνια μετά, τα δεδομένα αναδεικνύονται σε υπερπολύτιμο πόρο και η αποτελεσματική αξιοποίησή τους, ξεχωρίζει τους νικητές από τους χαμένους, σε ένα άκρως ανταγωνιστικό περιβάλλον.

# ΜΕΡΟΣ Ι

## Δεδομένα Μεγάλης Κλίμακας – Εξόρυξη Δεδομένων – Κοινωνικά Δίκτυα - Τουρισμός

### Κεφάλαιο 1

#### Δεδομένα Μεγάλης Κλίμακας – Μεγάλα Δεδομένα

##### 1.1 Ορισμός

Τα δεδομένα μεγάλης κλίμακας αποτελούν συλλογές συνόλων δεδομένων. Ο όγκος, η ποικιλία και η ταχύτητά τους είναι πολύ μεγάλα, ως εκ τούτου η χρήση παραδοσιακών βάσεων δεδομένων και εργαλείων επεξεργασίας δεδομένων για να αποθηκευτούν, να διαχειριστούν, να επεξεργαστούν και να αναλυθούν είναι κάτι παραπάνω από ανέφικτη. Η ραγδαία εξέλιξη της πληροφορικής, της βιομηχανίας, της υγείας, του Διαδικτύου των Πραγμάτων (Internet of Things) καθώς και άλλων τομέων οδήγησε στην αλματώδη αύξηση του όγκου τόσο των δομημένων όσο και των μη δομημένων δεδομένων που παράγονται από τους εν λόγω τομείς.

Σύμφωνα με την εκτίμηση της IBM, καθημερινά παράγονται 2,5 πεντάκις εκατομμύρια bytes δεδομένων (Taft, Vartak, Satish, Sundaram, Madden & Stonebraker, 2014). Η αναφορά του DOMO περιγράφει ζοφερά τον όγκο των δεδομένων που παράγεται από τις δημοφιλέστερες διαδικτυακές πλατφόρμες ανά λεπτό (DOMO, 2020). Ορισμένα βασικά στοιχεία της εν λόγω αναφοράς παρατίθενται παρακάτω:

- ⊕ Οι χρήστες του Facebook μοιράζονται περί τις 147.000 φωτογραφίες.
- ⊕ Οι χρήστες του Twitter αυξάνονται κατά 319 κάθε λεπτό.
- ⊕ Οι χρήστες του Instagram ανεβάζουν 347.222 ιστορίες.
- ⊕ Οι χρήστες του Youtube ανεβάζουν βίντεο διάρκειας 500 ωρών.
- ⊕ Οι χρήστες του Whatsapp μοιράζονται 41.666.667 μηνύματα.
- ⊕ Οι χρήστες δημιουργούν 1.388.889 βιντεοκλήσεις και φωνητικές κλήσεις.
- ⊕ Οι χρήστες του Facebook μοιράζονται 150.000 μηνύματα.



Κρίνεται απαραίτητο η επόμενη γενιά εφαρμογών να αξιοποιήσει τη δύναμη των δεδομένων μεγάλης κλίμακας, καθιστώντας τες πιο έξυπνες και αποτελεσματικές. Τομείς όπως ο τραπεζικός, ο χρηματοπιστωτικός, ο υγειονομικός, ο βιομηχανικός, ο περιβαλλοντικός, το μάρκετινγκ, το λιανικό εμπόριο, το διαδίκτυο καθώς και τα κυβερνο-φυσικά συστήματα (cyber-physical systems) αποτελούν γόνιμο έδαφος για την εφαρμογή μεγάλων δεδομένων.

Η συλλογή, αποθήκευση, επεξεργασία, ανάλυση και εξαγωγή χρήσιμων πληροφοριών από τα δεδομένα μαζικής κλίμακας απαιτεί την χρήση ειδικών εργαλείων και πλαισίων στις περιπτώσεις όπου:

A. Η αποθήκευση, η επεξεργασία και η ανάλυσή τους με ένα μηχάνημα μόνο είναι δύσκολο να επιχειρηθεί εξαιτίας του μεγάλου **όγκου** των εν λόγω δεδομένων.

B. Η ανάλυση των δεδομένων πρέπει να πραγματοποιείται σε πραγματικό χρόνο καθώς η **ταχύτητά** τους είναι πολύ υψηλή.

Γ. Η συλλογή δεδομένων από διαφορετικές πηγές δημιουργεί μια **ποικιλία δεδομένων** τα οποία μπορούν να δομηθούν, να ημι-δομηθούν ή να αποδομηθούν.

Δ. Πληθώρα διαφορετικών **αναλύσεων**, όπως η διαγνωστική ανάλυση, η περιγραφική ανάλυση, η συμβουλευτική ανάλυση και η προγνωστική ανάλυση, πρέπει να εκτελεστούν προκειμένου να εξαχθεί χρήσιμη πληροφορία από τα μεγάλα δεδομένα.

Οι δυνατότητες των μεγάλων δεδομένων, η δυναμική τους και η αρχιτεκτονική της παράλληλης επεξεργασίας των εργαλείων και πλαισίων μπορούν χαρακτηριστικά να συγκριθούν με αυτές ενός τεράστιου συνόλου υπολογιστικών συστημάτων.

Ο καθαρισμός δεδομένων, ο μετασχηματισμός δεδομένων (data munging), η επεξεργασία και η οπτική απεικόνισή τους αποτελούν ορισμένα μόνο βήματα της ανάλυσης μεγάλων δεδομένων. Η συλλογή δεδομένων από πολλαπλές πηγές αποτελεί το πρώτο βήμα του κύκλου ζωής της ανάλυσης μεγάλων δεδομένων. Σε αυτή τη φάση απαιτείται η χρήση εξειδικευμένων εργαλείων και πλαισίων προκειμένου να συγκεντρωθούν στη βάση δεδομένα από διαφορετικές πηγές. Με τη βοήθεια εξειδικευμένων υπηρεσιών αποθήκευσης δεδομένων, όπως είναι οι μη σχεσιακές βάσεις δεδομένων και τα κατανεμημένα συστήματα αρχείων, τα δεδομένα συλλέγονται και αποθηκεύονται.

Η επιλογή και χρήση εξειδικευμένων πλαισίων επηρεάζεται από τον τύπο της ανάλυσης που θα πραγματοποιήσουμε (διαγνωστική, περιγραφική, συμβουλευτική ή προγνωστική) καθώς και από τις απαιτήσεις της ανάλυσης (σε πραγματικό χρόνο ή παρτίδα). Τεχνολογίες όπως τα παράλληλα και κατανεμημένα πλαίσια επεξεργασίας, η υπολογιστική νέφους (cloud computing), η in-memory computing και οι μη σχεσιακές βάσεις δεδομένων βοηθούν σε μεγάλο βαθμό την ανάλυση μεγάλων δεδομένων.

Ορισμένα παραδείγματα δεδομένων μεγάλης κλίμακας παρατίθενται παρακάτω:

- ⊕ Τα δεδομένα που παράγονται από κοινωνικά δίκτυα, όπως κείμενα, εικόνες, ηχητικά και οπτικοακουστικά δεδομένα.
- ⊕ Δεδομένα χρήσης (clickstream data) που παράγονται από διαδικτυακές εφαρμογές όπως το e-Commerce, για την ανάλυση της συμπεριφοράς των χρηστών.
- ⊕ Δεδομένα που συλλέγονται από αισθητήρες σε ενσωματωμένες μηχανές βιομηχανικών και ενεργειακών συστημάτων για την παρακολούθηση της κατάστασής τους και την ανίχνευση σφαλμάτων.
- ⊕ Δεδομένα υγειονομικής περίθαλψης που συλλέγονται στα συστήματα ηλεκτρονικής καταγραφής της υγείας (EHR).
- ⊕ Αρχεία καταγραφής που δημιουργούνται από διαδικτυακές εφαρμογές.
- ⊕ Στοιχεία χρηματιστηριακών αγορών.
- ⊕ Δεδομένα συναλλαγών που παράγονται από τραπεζικές και χρηματοοικονομικές εφαρμογές.

## 1.2 Χαρακτηριστικά

Τα πιο σημαντικά χαρακτηριστικά των δεδομένων μεγάλης κλίμακας, τα οποία αναφέρονται στην βιβλιογραφία και ως 5Vs, είναι τα εξής (Erl Thomas, 2016):

- Όγκος

Τα δεδομένα μεγάλης κλίμακας είναι δεδομένα που χαρακτηρίζονται από τον μεγάλο τους όγκο, ως εκ τούτου είναι ανέφικτο να χωράν σε έναν μόνο υπολογιστή αλλά σε

σύνολα υπολογιστικών συστημάτων. Επομένως, κρίνεται επιβεβλημένη η χρήση εξειδικευμένων εργαλείων και πλαισίων για την διεκπεραίωση των απαραίτητων διεργασιών και την ανάλυση των εν λόγω δεδομένων. Για παράδειγμα, παράγονται καθημερινά Terabytes δεδομένων μέσω βιομηχανικών και ενεργειακών συστημάτων, επίσης, μέσω των εφαρμογών κοινωνικής δικτύωσης στέλνονται δισεκατομμύρια μηνύματα σε καθημερινή βάση κλπ. Αυξάνεται εκθετικά ο όγκος των δεδομένων που παράγεται από τους τομείς της υγείας, της πληροφορικής, της βιομηχανίας, του Διαδικτύου των Πραγμάτων και των λοιπών συστημάτων. Το γεγονός αυτό οφείλεται κυρίως στην ανάγκη εξαγωγής χρήσιμης πληροφορίας για την βελτιστοποίηση των επιχειρηματικών διεργασιών, την αποτελεσματικότητα και την καλύτερη εξυπηρέτηση των καταναλωτών καθώς και λόγω του μειωμένου κόστους των τεχνολογιών επεξεργασίας και αποθήκευσης δεδομένων. Στην βιβλιογραφία δεν αναφέρεται κάποιο συγκεκριμένο όριο, αναφορικά με τον όγκο των δεδομένων, ώστε να θεωρούνται δεδομένα μεγάλης κλίμακας. Ωστόσο ο όρος έχει επικρατήσει να χρησιμοποιείται για να περιγράψει τα μεγάλα δεδομένα, τα οποία είναι δύσκολο, χρησιμοποιώντας παραδοσιακές μεθόδους επεξεργασίας και βάσεις δεδομένων, να αποθηκευτούν, να επεξεργαστούν και να διαχειριστούν παρά μόνο με τη χρήση σύγχρονων συστημάτων διαχείρισης.

#### ▪ Ταχύτητα

Ο βασικός λόγος της ραγδαίας ανάπτυξης των μεγάλων δεδομένων είναι η ταχύτητά τους και αναφέρεται στο πόσο γρήγορα παράγονται τα δεδομένα. Αποτελεί ένα από τα σημαντικότερα χαρακτηριστικά των δεδομένων μεγάλης κλίμακας. Είναι σημαντικό να αναφερθεί ότι από ορισμένες πηγές παράγονται δεδομένα σε πολύ υψηλές ταχύτητες, χαρακτηριστικό παράδειγμα αποτελούν τα κοινωνικά δίκτυα ή τα δεδομένα που παράγονται από αισθητήρες μηχανημάτων. Η υψηλή τους ταχύτητα, με την οποία παράγονται, οδηγεί από τη μία πλευρά να συλλέγονται πολύ γρήγορα και από την άλλη μεριά, σε σύντομο χρονικό διάστημα, να καθίσταται ο όγκος τους πολύ μεγάλος. Σε ορισμένες περιπτώσεις δε, τα δεδομένα πρέπει να επεξεργάζονται και να αναλύονται σε πραγματικό χρόνο, όπως στον τομέα των πωλήσεων ή στις εφαρμογές ανίχνευσης ηλεκτρονικής απάτης, καθώς η ανάλυση των εν λόγω δεδομένων υπόκεινται σε αυστηρές χρονικές προθεσμίες. Ως εκ τούτου απαιτούνται εξειδικευμένα εργαλεία, ούτως ώστε να ανταπεξέλθουν στις τόσο υψηλές ταχύτητες λήψης των δεδομένων στο σύστημα.

- Ποικιλία

Με τον όρο ποικιλία αναφερόμαστε στα διάφορα είδη των δεδομένων. Σε ότι αφορά τα δεδομένα μεγάλης κλίμακας, αυτά μπορούν να έχουν διάφορες μορφές, όπως για παράδειγμα ημι-δομημένα, μη δομημένα ή δομημένα. Επιπλέον, εισρέουν διάφορα είδη μεγάλων δεδομένων όπως κείμενα, εικόνα, ήχος και βίντεο. Έτσι έχουμε να κάνουμε τόσο με τα διάφορα είδη δεδομένων όσο και με τη διαφορετική τους δομή. Σήμερα καλούμαστε να αποθηκεύσουμε, να επεξεργαστούμε και να αναλύσουμε δεδομένα που προέρχονται από πολλές και διαφορετικές πηγές π.χ. κοινωνικά δίκτυα, δίκτυα κινητής τηλεφωνίας, κάμερες κ.λπ. Ως εκ τούτου τα συστήματα που διαχειρίζονται μεγάλα δεδομένα θα πρέπει να είναι ευέλικτα ώστε να ανταπεξέλθουν στη μεγάλη ποικιλομορφία των προαναφερθέντων δεδομένων.

- Εγκυρότητα

Μια ακόμη πτυχή των μεγάλων δεδομένων είναι η εγκυρότητά τους, το πόσο ακριβή είναι τα εν λόγω δεδομένα. Μόνο όταν τα δεδομένα είναι ουσιαστικά και ακριβή μπορούν να επωφεληθούν οι εφαρμογές που βασίζονται σε αυτά. Προκειμένου να εξαχθεί χρήσιμη πληροφορία και αξία από τα δεδομένα, θα πρέπει να φιλτραριστούν κατάλληλα ούτως ώστε να αφαιρεθούν λανθασμένες τιμές και περιττές πληροφορίες. Ως εκ τούτου, η διαδικασία του καθαρισμού των δεδομένων είναι πολύ σημαντική για την αφαίρεση ελαττωματικών τιμών και την εξαγωγή χρήσιμης πληροφορίας.

- Αξία

Μια ακόμη παράμετρος κατά την εξέταση μεγάλων δεδομένων είναι η αξία των δεδομένων και αναφέρεται στην χρησιμότητά τους ως προς τον επιδιωκόμενο σκοπό. Η εξαγωγή αξίας από τα δεδομένα είναι ο απώτερος στόχος ενός συστήματος ανάλυσης δεδομένων μεγάλης κλίμακας. Η εν λόγω αξία σχετίζεται άμεσα με την ακρίβεια και την εγκυρότητα των δεδομένων. Επιπροσθέτως, σε ορισμένες εφαρμογές, η αξία είναι συνυφασμένη με την ταχύτητα επεξεργασίας των δεδομένων. Συμπερασματικά λοιπόν, οι επιχειρήσεις καλούνται να προχωρήσουν στην επιλογή της πιο αποτελεσματικής, από πλευράς κόστους, λύσης επιδιώκοντας την αξιοποίηση της πληροφορίας, η οποία με τη σειρά της θα οδηγήσει στην έγκαιρη και όσο το δυνατό πιο σωστή λήψη αποφάσεων, δίνοντας μεγαλύτερη αξία στην επιχείρηση.

## Κεφάλαιο 2

### Εξόρυξη Δεδομένων

#### 2.1 Ορισμός

Ο όγκος των δεδομένων που παράγεται καθημερινά και αυτός που έχει ήδη παραχθεί και βρίσκεται αποθηκευμένος σε βάσεις δεδομένων είναι πρωτοφανής στην ανθρώπινη ιστορία. Ο βασικός στόχος της εξόρυξης δεδομένων έχει να κάνει με την εξαγωγή χρήσιμων και κατανοητών ποιοτικών γνώσεων από αυτά τα δεδομένα, ή αλλιώς η ανακάλυψη των "κρυμμένων γνώσεων" που υπάρχουν μέσα σε αυτό το πλήθος δεδομένων. Σύμφωνα με τους Sang Jun Lee και Keng Siau (2001) στόχος της εξόρυξης δεδομένων (Data mining) είναι η πληροφορία που θα εξαχθεί και τα πρότυπα που θα προκύψουν να έχουν δομή κατανοητή προς τον άνθρωπο έτσι ώστε να τον βοηθήσουν να πάρει τις κατάλληλες αποφάσεις. Γι' αυτό και ένα από τα ονόματα με το οποίο έγινε γνωστός αυτός ο κλάδος της πληροφορικής είναι η Ανακάλυψη Γνώσης από βάσεις δεδομένων (Knowledge discovery in databases - KDD). Ο γενικός ορισμός της Ανακάλυψης Γνώσης από βάσεις δεδομένων όπως δόθηκε από τους Frawley, Piatetsky-Shapiro και Matheus (1992) είναι:

*"Ανακάλυψη Γνώσης από βάσεις δεδομένων είναι η ντετερμινιστική διαδικασία αναγνώρισης έγκυρων, καινοτόμων, ενδεχομένως χρήσιμων και εν τέλει κατανοητών προτύπων στα δεδομένα".*

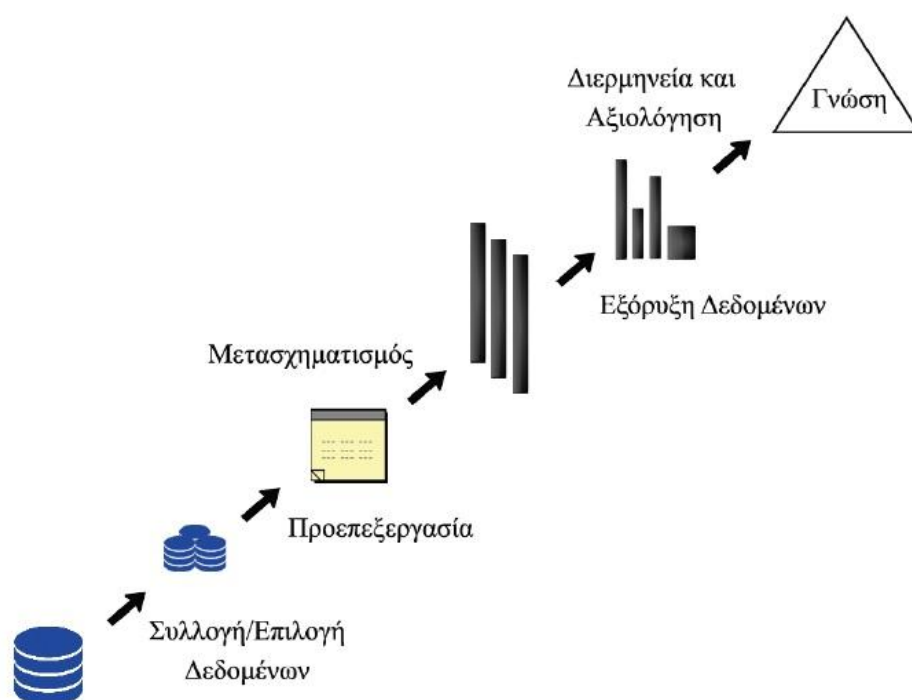
Ως εκ τούτου, η ανακάλυψη γνώσης από βάσεις δεδομένων είναι η εξεύρεση μιας ενδιαφέρουσας, αυτονόητης, μη προφανούς και πιθανόν χρήσιμης πληροφορίας ή προτύπου από μεγάλες βάσεις δεδομένων με χρήση αλγορίθμων ομαδοποίησης ή κατηγοριοποίησης και των αρχών της στατιστικής, της τεχνητής νοημοσύνης, της μηχανικής μάθησης και των συστημάτων βάσεων δεδομένων.

#### 2.2 Στάδια Εξόρυξης Δεδομένων

Η διαδικασία Εξόρυξης Δεδομένων αποτελείται από συγκεκριμένα στάδια. Πρόκειται για την αποκάλυψη ή παραγωγή λειτουργικής γνώσης μέσα από την ανάλυση των δεδομένων. Αναφέρεται σε ολόκληρη τη διαδικασία, από τη συλλογή δεδομένων μέχρι την αξιοποίηση των αποτελεσμάτων σε πιο πρακτικό επίπεδο (Βερούκιος,

Καγκλής & Σταυρόπουλος, 2015). Τα βασικά στάδια της Εξόρυξης Δεδομένων (Εικόνα 4) είναι τα παρακάτω:

1. Συλλογή Δεδομένων (Data Collection)
2. Προεπεξεργασία Δεδομένων (Preprocessing)
3. Μετασχηματισμός Δεδομένων (Transformation)
4. Εξόρυξη Δεδομένων (Data Mining)
5. Διερμηνεία και Αξιολόγηση (Interpretation/Evaluation)



*Εικόνα 4. Βασικά στάδια Εξόρυξης Δεδομένων (Βερούκιος, Καγκλής & Σταυρόπουλος, 2015)*

Ακολουθεί μια σύντομη περιγραφή για το κάθε στάδιο της Εξόρυξης Δεδομένων.

### 2.2.1 Συλλογή Δεδομένων

Το πρώτο βήμα της Εξόρυξης Δεδομένων είναι η συλλογή και η αποθήκευση των δεδομένων. Η συλλογή των δεδομένων συνήθως γίνεται είτε αυτόματα, π.χ. με χρήση αισθητήρων, είτε μη αυτόματα, π.χ. με χρήση ερωτηματολογίων.

### 2.2.2 Προεπεξεργασία Δεδομένων

Το δεύτερο και πιο σημαντικό στάδιο της Εξόρυξης Δεδομένων είναι η προεπεξεργασία του συνόλου δεδομένων, η οποία γίνεται με στόχο τον καθαρισμό

τους, δηλαδή την τακτοποίηση εσφαλμένων, προβληματικών ή ελλειπόντων δεδομένων. Μπορεί να απαιτήσει έως και το 60% της συνολικής προσπάθειας και αυτό διότι, αν τα δεδομένα δεν είναι «καθαρά» και στην κατάλληλη μορφή, δεν έχει νόημα να μιλάμε για ποιότητα αποτελεσμάτων.

### **2.2.3 Μετασχηματισμός Δεδομένων**

Ο μετασχηματισμός των δεδομένων αποτελεί το τρίτο στάδιο της Εξόρυξης Δεδομένων. Ουσιαστικά, πρόκειται για τη μετατροπή των δεδομένων κάτω από ένα κοινό πλαίσιο, για επεξεργασία. Χρησιμοποιείται κυρίως για την εξομάλυνση των δεδομένων και απομάκρυνση θορύβου, για τη συνάθροιση των δεδομένων, δηλαδή για την παραγωγή σύνοψης τους, για την κανονικοποίησή τους, δηλαδή την κλιμάκωση των χαρακτηριστικών του συνόλου δεδομένων σε ένα συγκεκριμένο και περιορισμένο εύρος τιμών ή τέλος για τη δημιουργία νέων χαρακτηριστικών από τα ήδη υπάρχοντα.

### **2.2.4 Εξόρυξη Δεδομένων**

Σε αυτό το στάδιο εφαρμόζεται κάποιος αλγόριθμος για την παραγωγή ενός μοντέλου. Έχοντας καθαρίσει και μετασχηματίσει τα δεδομένα, είναι έτοιμα να χρησιμοποιηθούν από κάποιον αλγόριθμο, ώστε να δημιουργηθεί κάποιο μοντέλο, συνήθως κατηγοριοποίησης ή πρόβλεψης.

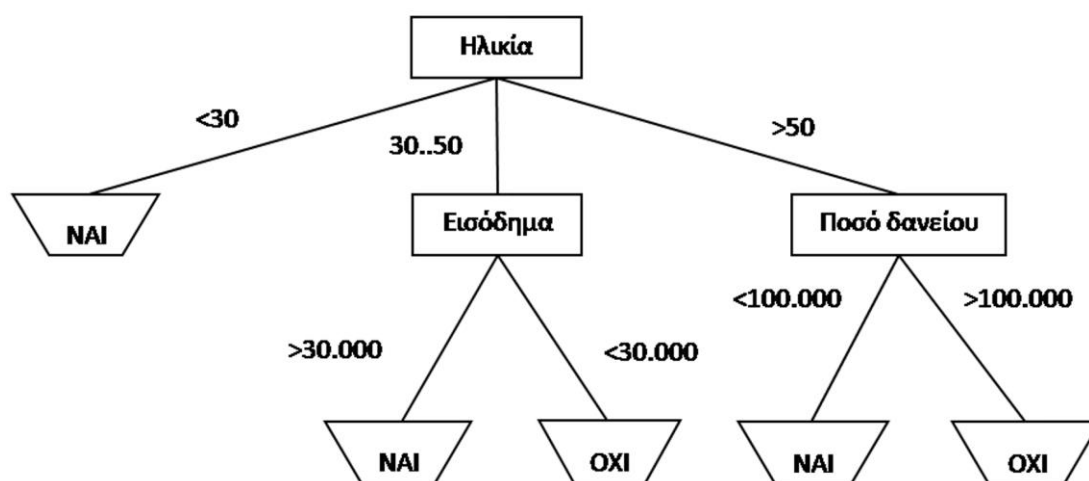
### **2.2.5 Διερμηνεία και Αξιολόγηση**

Στο τελευταίο στάδιο πραγματοποιείται η διερμηνεία και η αξιολόγηση των αποτελεσμάτων από τον αναλυτή. Μπορεί να απορρίψει πρότυπα που δεν του είναι χρήσιμα και αν το θεωρήσει αναγκαίο, μπορεί να επιστρέψει στα προηγούμενα στάδια ώστε να τροποποιήσει το σύνολο των δεδομένων ή να επιλέξει διαφορετική μέθοδο εξόρυξης αν δεν τον ικανοποιούν τα αποτελέσματα με βασικό κριτήριο την αξιοπιστία.

## **2.3 Τεχνικές Εξόρυξης Δεδομένων**

Η εξόρυξη δεδομένων χωρίζεται χονδρικά σε δύο μεγάλες κατηγορίες τεχνικών, την μάθηση με επίβλεψη (supervised learning) και τη μάθηση χωρίς επίβλεψη (unsupervised learning).

Η μάθηση με επίβλεψη περιλαμβάνει τη διαδικασία κατηγοριοποίησης (classification) των δεδομένων, δηλαδή του ταιριάσματος του συνόλου δεδομένων σε προκαθορισμένες κατηγορίες. Βασίζεται στην εξέταση των χαρακτηριστικών ενός νέου αντικειμένου (μη κατηγοριοποιημένου) το οποίο με βάση τα χαρακτηριστικά αυτά αντιστοιχίζεται σε ένα προκαθορισμένο σύνολο κλάσεων. Η εργασία της κατηγοριοποίησης χαρακτηρίζεται από έναν καλά καθορισμένο ορισμό των κατηγοριών και το σύνολο που χρησιμοποιείται για την εκπαίδευση του μοντέλου αποτελείται από προκατηγοριοποιημένα παραδείγματα. Η βασική εργασία είναι να δημιουργηθεί ένα μοντέλο το οποίο θα μπορούσε να εφαρμοστεί για να κατηγοριοποιήσει δεδομένα που δεν έχουν ακόμα κατηγοριοποιηθεί (ανατεθεί σε κάποια από τις κατηγορίες). Βασικά παραδείγματα τεχνικών κατηγοριοποίησης είναι η επαγωγή από δέντρα απόφασης, τα δίκτυα Bayes, οι ταξινομητές κοντινότερου γείτονα και τα νευρωνικά δίκτυα. Η εικόνα 5 απεικονίζει ένα Δένδρο Αποφάσεων για την έγκριση τραπεζικών δανείων.



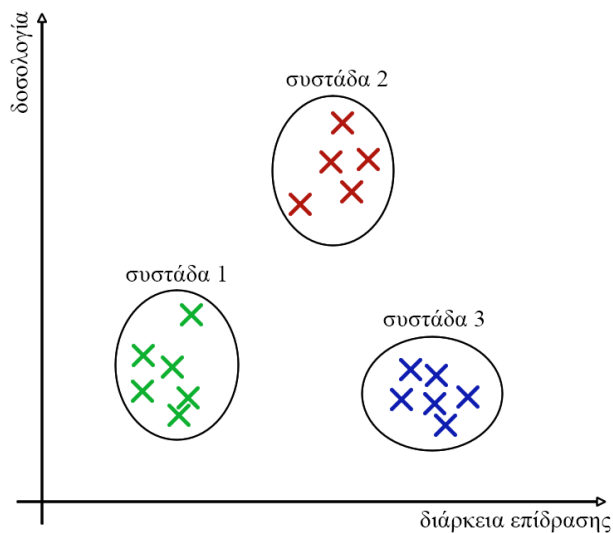
*Εικόνα 5. Δένδρο Αποφάσεων (Κύρκος, 2015)*

Η μάθηση χωρίς επίβλεψη αναφέρεται κυρίως στην τεχνική της συσταδοποίησης (clustering) που είναι η εργασία του καταμερισμού ενός ετερογενούς πληθυσμού σε ένα σύνολο περισσότερων ετερογενών συστάδων (clusters). Αυτό που διαφοροποιεί τη συσταδοποίηση από την κατηγοριοποίηση είναι ότι η συσταδοποίηση δε βασίζεται σε προκαθορισμένες κατηγορίες. Οι εγγραφές ομαδοποιούνται σε σύνολα με βάση την ομοιότητα που παρουσιάζουν μεταξύ τους. Επαφίεται σε εμάς να καθορίσουμε τη σημασία που θα έχει η κάθε μια από τις ομάδες που προκύπτουν. Βασικές τεχνικές



συσταδοποίησης είναι οι ιεραρχικοί αλγόριθμοι, οι αλγόριθμοι διαίρεσης (partitioning algorithms), οι βασισμένοι σε πλέγμα αλγόριθμοι κλπ.

Στην εικόνα 6 βλέπουμε το αποτέλεσμα της συσταδοποίησης φαρμακευτικών δεδομένων. Έχουν δημιουργηθεί τρεις συστάδες με βάση τα χαρακτηριστικά «δοσολογία» και «διάρκεια επίδρασης».



*Εικόνα 6. Συσταδοποίηση (Βερύκιος, Καγκλής & Σταυρόπουλος, 2015)*

## Κεφάλαιο 3

### Κοινωνικά Δίκτυα

#### 3.1 Εξέλιξη Κοινωνικών Δικτύων

Μετά την κρίση του 1999 στις εταιρίες νέων τεχνολογιών, που ονομάστηκε "dot com bubble", αναδείχτηκε μια νέα αντίληψη για το διαδίκτυο σαν έναν πρωτότυπο συνδυασμό τεχνολογιών και συμπεριφορών τόσο από πλευράς χρήστη όσο και από πλευράς επιχειρήσεων, οργανισμών και εξυπηρετητών. Η απαίτηση για δυναμικό περιεχόμενο οδήγησε στην άνθηση τεχνολογιών όπως η ασύγχρονη JavaScript (AJAX) και η XML και στη δημιουργία πλατφορμών όπως τα blogs και τα wikis. Στο περιβάλλον του Web 2.0 και των δυναμικών τεχνολογιών που περιλαμβάνει, έγινε δυνατό να αναπτυχθούν πλατφόρμες ψηφιακής κοινωνικής δικτύωσης και διαμοιρασμού περιεχομένου που διευκολύνουν και επιταχύνουν την κοινοποίηση περιεχομένου από δεκάδες εκατομμύρια χρήστες. Η στροφή σε μια ακόμα πιο οριζόντια και κοινωνική χρήση του παγκόσμιου ιστού έστρωσε το έδαφος για την εμφάνιση των ψηφιακών κοινωνικών δικτύων όπως το Facebook, το Twitter, τις κοινότητες απάντησης ερωτήσεων όπως το Yahoo! Answers και το stack overflow και πολλά άλλα λιγότερο ή περισσότερο γνωστά κοινωνικά δίκτυα.



*Εικόνα 7. Το σύμπαν τεχνολογιών του Web 2.0*  
([https://en.wikipedia.org/wiki/Web\\_2.0#/media/File:Web\\_2.0\\_Map.svg](https://en.wikipedia.org/wiki/Web_2.0#/media/File:Web_2.0_Map.svg))

Ο ιστός τώρα είναι πιο κοινωνικός και η συνεργασία, η συνεισφορά και η έννοια της κοινότητας είναι κεντρικής σημασίας. Οι «έξι μεγάλες ιδέες» όπως αυτές διατυπώθηκαν από τον O'Reilly το 2005 είναι:

- ⊕ Παραγωγή ατομικού περιεχομένου χρήστη (user generated content).
- ⊕ Εκμετάλλευση της συλλογικής ευφυΐας που γεννιέται στο διαδίκτυο (collective intelligence).
- ⊕ Δεδομένα σε επική κλίμακα (data on epic scale).
- ⊕ Αρχιτεκτονική που βοηθάει την συμμετοχή (Architecture of participation).
- ⊕ Φαινόμενα δικτύου (network effects).
- ⊕ Ο ιστός είναι πλέον ανοικτός σε όλους όπως είναι και οι εφαρμογές του (openness).

Το περιεχόμενο του νέου παγκόσμιου ιστού και των εφαρμογών του είναι ανοικτό και εξελίσσεται χάρις στη συνεχή διαχείριση των δεδομένων του από τους χρήστες. Το Web 2.0 χαμήλωσε το φράγμα της συμμετοχής και επέτρεψε στους «πρώην ακροατές» να γίνουν «ενεργοί συμμετοχοί» στη σημερινή διαδικτυακή κοινωνία και να εκφράσουν τα συναισθήματά τους, τη γνώμη τους, τις εμπειρίες τους και τις γνώσεις τους.

Οι χρήστες των σημερινών εφαρμογών του Παγκόσμιου Ιστού (ΠΙ) έχουν τη δυνατότητα να δημοσιοποιούν, να μοιράζονται, να οργανώνουν και να αξιολογούν το διαδικτυακά διαθέσιμο περιεχόμενο. Επιπλέον, μέσα από δραστηριότητες με επίκεντρο το περιεχόμενο, αλληλεπιδρούν και σχηματίζουν κοινωνικά δίκτυα μεταξύ τους. Οι δυνατότητες αυτές προσφέρονται μέσα από ποικιλία διαδικτυακών εφαρμογών, οι οποίες είναι γνωστές ως εφαρμογές κοινωνικού περιεχομένου (social media applications) ή ως εφαρμογές κοινωνικής δικτύωσης και χαρακτηρίζονται από έμφαση στο περιεχόμενο που δημοσιεύεται στο περιβάλλον τους και στις δυνατότητες κοινωνικής δικτύωσης μεταξύ των χρηστών τους. Η εικόνα 8 παρουσιάζει ένα δείγμα από αντιπροσωπευτικές εφαρμογές κοινωνικής δικτύωσης που συγκεντρώνουν πολύ μεγάλο μέρος των διαδικτυακών δραστηριοτήτων στον Παγκόσμιο Ιστό κατά τα τελευταία χρόνια. Οι περισσότερες από αυτές έχουν επίκεντρο το περιεχόμενο. Το πολύ δημοφιλές Facebook υποστηρίζει σε μεγάλο βαθμό την αλληλεπίδραση με ποικιλία περιεχομένου (π.χ. like σε εξωτερικούς συνδέσμους). Στην περίπτωση του Twitter, η αλληλεπίδραση με το περιεχόμενο

επιτυγχάνεται μέσα από την επισύναψη υπερσυνδέσμων και τη χρήση hashtags για την περιγραφή του.

Εφαρμογή	Περιεχόμενο	Χαρακτηριστικά Επισημείωσης		Δυνατότητες Κοινωνικής Δικτύωσης	
		Ετικέτες	Κατηγορίες	Φίλοι	Ομάδες
Delicious.com	Ιστοσελίδες	✓		✓	
Flickr.com	Εικόνες	✓		✓	✓
Youtube.com	Βίντεο	✓	✓	✓	✓
Last.fm	Μουσική	✓		✓	✓
Digg.com	Ειδήσεις		✓	✓	
Reddit.com	Ειδήσεις		✓	✓	
Citeulike.org	Βιβλιογραφία	✓			✓
Bibsonomy.org	Βιβλιογραφία	✓			✓
Facebook.com	Διάφορα	✓		✓	✓
Twitter.com	Διάφορα	✓		✓	

*Εικόνα 8. Χαρακτηριστικά και δυνατότητες εφαρμογών κοινωνικής δικτύωσης*

Η μεγάλη ποικιλία τέτοιων εφαρμογών δημιουργεί ένα σύνθετο πλαίσιο μελέτης, στο οποίο εμπλέκονται ετερογενείς μηχανισμοί δημιουργίας και διαμοιρασμού περιεχόμενου διαφορετικών τύπων, με το οποίο αλληλεπιδρούν οι χρήστες αξιοποιώντας τις ποικίλες δυνατότητες που αυτές προσφέρουν. Ανεξάρτητα όμως από τα επιμέρους χαρακτηριστικά και τις δυνατότητές τους, οι σύγχρονες εφαρμογές κοινωνικής δικτύωσης βασίζονται στις ακόλουθες τρεις θεμελιώδεις λειτουργίες:

- **Επισήμανση (bookmarking):** Η λειτουργία αυτή συνίσταται στην αποθήκευση της δικτυακής τοποθεσίας (URL) ενός στοιχείου περιεχόμενου ή στην παροχή ψήφου με σκοπό την εκδήλωση της προτίμησης ή του ενδιαφέροντος του χρήστη προς το εν λόγω στοιχείο. Χαρακτηριστικά παραδείγματα επισήμανσης είναι τα like στην εφαρμογή Facebook, τα bookmarks στην εφαρμογή Delicious, τα diggs στην εφαρμογή Digg και τα retweets στην εφαρμογή Twitter.
- **Επισημείωση (tagging):** Η επισημείωση περιλαμβάνει τη συσχέτιση ενός στοιχείου περιεχόμενου με ένα σύνολο χαρακτηριστικών λέξεων-κλειδιών (keywords) ή ετικετών (tags) όπως αποκαλούνται συχνά, καθώς και την ταξινόμησή του σύμφωνα με ένα προκαθορισμένο σχήμα κατηγοριοποίησης. Συνεπώς, η επισημείωση επικεντρώνεται στην απόδοση σημασιολογικής ή/και

ταξινομικής πληροφορίας στο περιεχόμενο. Η πλειοψηφία των εφαρμογών κοινωνικής δικτύωσης προσφέρουν δυνατότητα συσχέτισης του περιεχόμενου τους με ετικέτες, ενώ συχνά δίνουν τη δυνατότητα στους χρήστες να περιγράψουν τα στοιχεία περιεχόμενου με τίτλους και σύντομες περιγραφές. Σε κάποιες περιπτώσεις (π.χ. στις εφαρμογές YouTube και Digg) προσφέρεται ένα προκαθορισμένο σύνολο κατηγοριών, με χρήση του οποίου οι χρήστες μπορούν να ταξινομήσουν το περιεχόμενό τους.

- **Κοινωνική δικτύωση (social networking):** Η λειτουργία αυτή υποστηρίζει τη δημιουργία σχέσεων μεταξύ των χρηστών της εφαρμογής. Στην περίπτωση που οι σχέσεις συνάπτονται μεταξύ δύο χρηστών, αυτές αναφέρονται ως σχέσεις 'φίλων', όπως καθιερώθηκε στην εφαρμογή Facebook, ή σχέσεις 'επαφών' (contacts) όπως αναφέρονται στην εφαρμογή Flickr. Επιπρόσθετα, πολλές εφαρμογές προσφέρουν τη δημιουργία σχέσεων μεταξύ πολλών χρηστών, που επικεντρώνονται γύρω από ένα θέμα. Αυτού του είδους οι πολλαπλές σχέσεις αποκαλούνται 'ομάδες' (groups).

Οι τρεις αυτές λειτουργίες μπορούν να υλοποιηθούν ανεξάρτητα ή σε συνδυασμό μεταξύ τους.

### 3.2 Τα Κοινωνικά Δίκτυα και η Σημασία τους

Αν, ήδη από την πρώτη εμφάνιση και διάδοση του, ο παγκόσμιος ιστός είχε δημιουργήσει τεράστιες προσδοκίες όσον αφορά τις οικονομικές και κοινωνικές δυνατότητες που προσφέρει, τόσο ώστε να οδηγήσει σε έκρηξη της "νέας οικονομίας", με την ανάπτυξη των κοινωνικών ιδιοτήτων του Web 2.0 και των online μέσων κοινωνικής δικτύωσης, αυτές οι δυνατότητες έχουν ξεπεράσει κάθε προσδοκία. Η ανάλυση των δεδομένων που προκύπτουν από τη χρήση των online κοινωνικών δικτύων έχει ωφελήσει ήδη μια σειρά από τομείς οικονομικής, κοινωνικής και πολιτικής δραστηριότητας (Schmidt & Cohen, 2014):

- ✦ Στο μάρκετινγκ και τη διαφήμιση, ο όρος "ιογενές" μάρκετινγκ (viral marketing) οφείλει την ύπαρξή του στην εξάπλωση των online εφαρμογών κοινωνικής δικτύωσης και στην αντίστοιχη χρήση τους για τη διαφήμιση προϊόντων και

υπηρεσιών μέσα από αυτά. Χαρακτηριστική περίπτωση είναι το μάρκετινγκ μέσω Facebook και η ανάλυση των προτιμήσεων των χρηστών ώστε να γίνει στοχευμένη και προσωποποιημένη διαφήμιση (π.χ. Google Adwords, Facebook recommended pages).

- ⊕ Στη δημοσιογραφία, η διάδοση των νέων μορφών παραγωγής, αποθήκευσης και διάδοσης πληροφορίας που σχετίζονται με τις τεχνολογίες του Web 2.0 έχει ανατρέψει τις αντιλήψεις για το ρόλο και τα μέσα δραστηριοποίησης της δημοσιογραφίας, έχει επιταχύνει αφάνταστα τη διάδοση των ειδήσεων, έχει επηρεάσει το στυλ και τη θεματολογία της και έχει οδηγήσει τα μεγάλα διεθνή πρακτορεία να στραφούν αποφασιστικά σε online συνεργατικές μορφές δημοσιογραφίας.
- ⊕ Στην πολιτική και τη διακυβέρνηση, με την καθημερινή και άμεση συμμετοχή στο δημόσιο διάλογο των πολιτών, η ανάπτυξη των εφαρμογών κοινωνικής δικτύωσης έχει οδηγήσει σε μεγαλύτερη διαφάνεια, σε αλλαγή προσέγγισης μεταξύ πολιτικών και πολιτών (π.χ. η καμπάνια επανεκλογής του Ομπάμα το 2012 βασίστηκε ιδιαίτερα σε τεχνικές εξόρυξης δεδομένων από τον παγκόσμιο ιστό), στην ανάπτυξη και διάδοση ιδεών και κοινωνικών κινήματων (π.χ. η περίπτωση των αναταραχών στις αραβικές μεσογειακές χώρες που ονομάστηκε "Αραβική Άνοιξη" χαρακτηρίστηκε εν πολλοίς μια εξέγερση των social media).
- ⊕ Στην αγορά εργασίας, online εφαρμογές όπως το LinkedIn έχουν τροποποιήσει το μάντζιμεντ ανθρώπινων πόρων και προσφέρουν ευρύτερες δυνατότητες για άμεση σύνδεση των διαφορετικών εταίρων στην αγορά εργασίας σε διεθνή κλίμακα.
- ⊕ Στον τουριστικό τομέα, οι εφαρμογές κοινωνικής δικτύωσης κερδίζουν μεγαλύτερη προσοχή και διαδραματίζουν σημαντικό ρόλο στη διαδικασία λήψης αποφάσεων των πελατών. Από τη μία πλευρά οι τουρίστες χρησιμοποιούν τις εφαρμογές κοινωνικής δικτύωσης προκειμένου να εκφράσουν τη γνώμη τους για τα προϊόντα και τις υπηρεσίες που χρησιμοποίησαν και από την άλλη πλευρά για να ενημερωθούν για την ποιότητα και την καταλληλότητα του προϊόντος πριν από την κατανάλωση.

Ενδεικτικά, σύμφωνα με μια έρευνα της Deloitte UK (2015) όπου εξετάζεται η επιρροή του Facebook μόνο στην παγκόσμια οικονομία, η πλατφόρμα του Facebook συνδέει 1,35 δισεκατομμύρια ανθρώπους με ετήσιο οικονομικό αποτέλεσμα για το 2014 τα 227 δισεκατομμύρια δολάρια και τη δημιουργία 4,5 εκατομμυρίων θέσεων εργασίας παγκοσμίως την ίδια χρονιά.

### 3.3 Εξόρυξη Δεδομένων στα Κοινωνικά Δίκτυα

Η εξόρυξη δεδομένων στον παγκόσμιο ιστό, λόγω της δικτυακής δομής των πληροφοριακών οντοτήτων του ιστού, θέτει νέα προβλήματα στη δημιουργία του συνόλου δεδομένων προς ανάλυση, στις τεχνικές ανάλυσης και στο εύρος των ερωτημάτων που μπορούν να απαντηθούν (Liu, 2011).

Η πιο βασική δραστηριότητα του παγκόσμιου ιστού, η διαδικτυακή αναζήτηση βασίζεται στις τεχνικές ανάκτησης πληροφορίας (Information Retrieval). Η βασική ιδέα πίσω από την ανάκτηση πληροφορίας στο διαδίκτυο είναι ότι αναζητούνται αποτελέσματα που έχουν τη μορφή εγγράφων (documents), βάσει ενός διατυπωμένου ερωτήματος (query) που αποτελείται από όρους (terms), το οποίο ερώτημα καλείται να ικανοποιήσει ένα σύστημα ανάκτησης πληροφορίας (IR system). Στην περίπτωση του παγκόσμιου ιστού, τα έγγραφα είναι σελίδες ιστού που εκτός από κείμενο, περιέχουν και υπερσυνδέσμους που τις διασυνδέουν δημιουργώντας ένα ιδεατό δίκτυο εγγράφων.

Η εξέλιξη της αντίληψης για τον παγκόσμιο ιστό και η άνοδος της έννοιας των κοινωνικών δικτύων οδήγησε την εξόρυξη δεδομένων προς την κατεύθυνση της ανάλυσης κοινωνικών δικτύων (social network analysis). Στη συγκεκριμένη περίπτωση, οι πληροφοριακές οντότητες που αφορούν την έρευνα αποτελούνται από δρώντες (actors) οι οποίοι αναπαρίστανται από τους κόμβους του δικτύου και οι οποίοι με τη σειρά τους δημιουργούν μεταξύ τους σχέσεις που αναπαρίστανται από τις ακμές του δικτύου. Οι βασικοί αλγόριθμοι αναφοράς για την αναζήτηση και ανάκτηση πληροφορίας στα κοινωνικά δίκτυα είναι ο αλγόριθμος HITS (Kleinberg, 1999) και ο αλγόριθμος PageRank (Page, Brin, Motwani & Winograd, 1999).

Μια ειδική εφαρμογή των τεχνικών ανάλυσης κοινωνικών δικτύων είναι η ανακάλυψη κοινοτήτων (community discovery). Τα ερωτήματα που περιλαμβάνει

αυτή η εργασία έχουν να κάνουν με την ανακάλυψη των κόμβων επιρροής (influential nodes), των υπο-κοινοτήτων που διαμορφώνονται σε ευρύτερα κοινωνικά δίκτυα, των ομαδοποιήσεων στα τηλεπικοινωνιακά δίκτυα και την πρόβλεψη των εκτιμήσεων των χρηστών, ερωτήματα που έχουν ιδιαίτερη σημασία στο ιογενές μάρκετινγκ και στην πρόβλεψη δραστηριότητας των χρηστών στα κοινωνικά δίκτυα. Ένα ανερχόμενο πεδίο ανάλυσης έχει να κάνει με την εξόρυξη γνώμης και συναισθημάτων (opinion and sentiment mining) στα κοινωνικά δίκτυα. Ο τρόπος που οι άνθρωποι αντιλαμβάνονται την πραγματικότητα έχει να κάνει ιδιαίτερα με τις απόψεις και κοσμοαντιλήψεις των ανθρώπων στο στενό τους περιβάλλον και επηρεάζουν τον τρόπο με τον οποίο οδηγούνται σε λήψεις αποφάσεων. Σε ένα διαδικτυακό περιβάλλον υψηλά διασυνδεδεμένο, ο όγκος των πληροφοριών γνώμης και συναισθήματος στο άμεσα προσβάσιμο περιβάλλον των χρηστών είναι τεράστιος και οδηγεί στην ανάγκη για λειτουργίες σύνοψης και εξόρυξης τέτοιων πληροφοριών. Οι σχετικές τεχνικές έχουν ιδιαίτερη εφαρμογή στα συστήματα κριτικών προϊόντων και υπηρεσιών που προσφέρονται για την ανατροφοδότηση από τους χρήστες και τη βελτίωση των παρεχόμενων υπηρεσιών.

### 3.4 Προκλήσεις Εξόρυξης Δεδομένων από τα Κοινωνικά Δίκτυα

Η εξόρυξη γνώσης από τα κοινωνικά δίκτυα αντιμετωπίζει σημαντικές εγγενείς προκλήσεις που οφείλονται στη φύση των κοινωνικών μέσων και στον τρόπο δημιουργίας των δεδομένων και μεταδεδομένων τους. Συνοπτικά, αυτές οι προκλήσεις έχουν να κάνουν με την ποιότητα και την ποσότητα των τελευταίων, για αυτό και στη συνέχεια τις διακρίνουμε σε ποιοτικές και ποσοτικές. Οι ποιοτικές προκλήσεις σχετίζονται με την οργάνωση των δεδομένων προς επεξεργασία, τη δομή και την ποσότητα θορύβου σε αυτά. Στα κοινωνικά δίκτυα οι **ποιοτικές** προκλήσεις όσον αφορά στην ανάλυση μεταδεδομένων συνοψίζονται στις ακόλουθες:

#### Πρόκληση 1: Ασαφές επίπεδο αξιοπιστίας και εγκυρότητας των μεταδεδομένων.

Εφόσον οι καταχωρήσεις προέρχονται από τους τελικούς χρήστες, χωρίς να μεσολαβεί κάποιου είδους ελέγχου, το γεγονός ότι οι χρήστες είναι επιρρεπείς σε λάθη (π.χ. συντακτικά, γραμματικά, κλπ.), σε συνδυασμό με την κακόβουλη διάθεση ορισμένων για εσκεμμένα λάθη, μπορεί να οδηγήσει σε μη έγκυρα μεταδεδομένα, που χαρακτηρίζονται ως "θόρυβος".



Πρόκληση 2: Έλλειψη (ιεραρχικής) δομής των πληροφοριών. Εκ κατασκευής, τα κοινωνικά δίκτυα δεν περιέχουν εσωτερικές δομές αναπαράστασης των δεδομένων με βάση σημασιολογικές σχέσεις μεταξύ τους. Ως εκ τούτου, τα δεδομένα δεν έχουν μεταξύ τους κάποιου είδους σύνδεση (με εξαίρεση τις συνδέσεις που ορίζονται από τους χρήστες), καταλήγοντας σε μία επίπεδη δομή.

Πρόκληση 3: Φαινόμενα φυσικών γλωσσών. Τα φαινόμενα αυτά προκύπτουν από τη μη απόλυτη ευθυγράμμιση μεταξύ γλωσσικών και εξιδανικευμένων νοηματικών χώρων (conceptual spaces) και τα πιο γνωστά είναι: i) η ασάφεια όρων (term ambiguity), δηλαδή ένας όρος μπορεί να έχει πολλές ερμηνείες, ii) η συνωνυμία όρων (synonymy), δηλαδή δύο διαφορετικές λέξεις μπορεί να έχουν την ίδια σημασία και iii) η διακύμανση της διακριτότητας των όρων (basic level variation), δηλαδή οι χρήστες δεν χρησιμοποιούν το ίδιο επίπεδο περιγραφής, όταν αναφέρονται σε μια έννοια π.χ. ορισμένοι άνθρωποι χρησιμοποιούν τη λέξη “ζώο”, ενώ άλλοι χρησιμοποιούν τον όρο “γάτα” για να περιγράψουν έναν ψηφιακό πόρο που απεικονίζει μια γάτα.

Πρόκληση 4: Πλεονασμός μεταδεδομένων. Πολλοί χρήστες, για να αντιμετωπίσουν τις δύο παραπάνω προκλήσεις (2 και 3) συνηθίζουν να χρησιμοποιούν πλεονάζουσες ετικέτες για τη βελτίωση της συλλογής σχετικού περιεχομένου. Αυτό όμως, επιδεινώνει την ακρίβεια των συστημάτων, καθώς με αυτόν τον τρόπο επιστρέφονται αρκετά μη σχετικά αντικείμενα στο χρήστη.

Όσον αφορά στις **ποσοτικές** προκλήσεις, διακρίνουμε, μεταξύ άλλων, δύο παραμέτρους που χρήζουν εξέτασης και μελέτης στη σχεδίαση τεχνικών ανάλυσης κοινωνικών δεδομένων για εξαγωγή γνώσης από αυτά: τον όγκο και τη δυναμικότητα των δεδομένων. Για να μπορέσει κανείς να συλλάβει τις τάξεις μεγέθους των δεδομένων στα οποία αναφερόμαστε, όπως επίσης και τους ρυθμούς παραγωγής και τη χρονική μεταβλητότητά τους ενδεικτικά αναφέρουμε:

Πρόκληση 5: Όγκος δεδομένων. Τα στατιστικά, όπως έχουν αποτυπωθεί σε επίσημες πηγές, είναι χαρακτηριστικά για το μέγεθος και το ρυθμό εξέλιξης των δεδομένων στα κοινωνικά δίκτυα. Στο Flickr υπάρχουν 87 εκατομμύρια εγγεγραμμένα μέλη και ήδη φιλοξενούνται πάνω από 6 δισεκατομμύρια εικόνες (<http://en.wikipedia.org/wiki/>

[Flickr](#), 2013). Επιπροσθέτως, στο Facebook έχουν καταγραφεί 2.22 δις χρήστες παγκοσμίως με 68.5% διείσδυση στις ΗΠΑ και 47.4% διείσδυση στην Ευρώπη (<http://www.internetworldstats.com/facebook.htm>, 2020). Για να δώσουμε μια ένδειξη του μεγέθους του περιεχομένου στο Facebook, αναφέρουμε ότι τον τελευταίο χρόνο κατά μέσο όρο, κάθε 60 δευτερόλεπτα, οι χρήστες ανέβασαν 147.000 φωτογραφίες και έστειλαν 150.000 μηνύματα. Στο Twitter αυξάνονται οι χρήστες κατά 319 κάθε λεπτό. Αντίστοιχα, στο YouTube τα μεγέθη των αριθμών σε ώρες βίντεο και δράσεις χρηστών είναι πολύ μεγάλα. Συγκεκριμένα, έχει καταγραφεί ότι οι χρήστες ανέβασαν συνολικά 500 ώρες βίντεο (<https://www.domo.com/learn/data-never-sleeps-8>, 2020).

Πρόκληση 6: Δυναμικότητα δεδομένων. Με τον όρο δυναμικότητα αναφερόμαστε στους ρυθμούς με τους οποίους προστίθεται και "καταναλώνεται" νέο περιεχόμενο στα κοινωνικά μέσα και, γενικά στη δυναμική τους, όπως αυτή έχει καταγραφεί με στατιστικές μελέτες τα τελευταία χρόνια. Συγκεκριμένα, το YouTube έχει πάνω από 2 δισεκατομμύρια μηνιαίους συνδεδεμένους χρήστες. Επιπλέον, κάθε μέρα οι άνθρωποι παρακολουθούν πάνω από ένα δισεκατομμύριο ώρες βίντεο και δημιουργούν δισεκατομμύρια προβολές (views). Πάνω από 500 ώρες περιεχομένου μεταφορτώνονται στο YouTube κάθε λεπτό (<http://www.youtube.com/yt/press/statistics.html>, 2020). Στο Flickr 3.5 εκατομμύρια νέες εικόνες μεταφορτώνονται καθημερινά. Στο Facebook πάνω από 100 δισεκατομμύρια μηνύματα και 1 δισεκατομμύριο ιστορίες (stories) κοινοποιούνται καθημερινά (<https://newsroom.fb.com/Key-Facts>, 2020).

Θέλοντας λοιπόν να αξιοποιήσουμε τη Σοφία του Πλήθους και να εντοπιστούν ενδιαφέροντα μοτίβα και σχέσεις μέσα από τις διαδικτυακές αλληλεπιδράσεις εκατομμυρίων χρηστών στα κοινωνικά δίκτυα, είναι αναγκαίο να βρεθούν τρόποι αντιμετώπισης των παραπάνω προκλήσεων, ώστε να γίνει εφικτή η επεξεργασία τους.

## Κεφάλαιο 4

### Τουρισμός

#### 4.1 Εξέλιξη Ανάλυσης Τουριστικού Κλάδου

Ο τουρισμός αποτελεί ένα πολυδιάστατο φαινόμενο με οικονομικές, κοινωνικές, πολιτιστικές και περιβαλλοντικές διαστάσεις και ο οποίος αρχίζει να γίνεται σημαντικό τμήμα των εθνικών οικονομιών μετά τον Β' Παγκόσμιο Πόλεμο. Παρόλα αυτά, η επιχειρηματική ανάλυση του κλάδου (business analysis) δεν γνώρισε παρόμοια ακμή. Οι σχετικές μελέτες αφορούσαν κυρίως συγκεκριμένους προορισμούς, ήταν ποιοτικού χαρακτήρα και δεν στόχευαν στην εξαγωγή ευρύτερων συμπερασμάτων.

Η πρώτη σημαντική απόπειρα δημιουργίας ενός γενικού μοντέλου ερμηνείας της τουριστικής αγοράς έγινε από τον Plog το 1972. Εκείνο το άρθρο επικαιροποιήθηκε το 2001 για να συμπεριλάβει τις αλλαγές στις στατιστικές μεθόδους και τα προφίλ των τουριστικών προορισμών που έγιναν μετά την αρχική δημοσίευση (Plog, 2001). Ακολουθώντας το μεταπολεμικό ρεύμα του συμπεριφορισμού και την πρόσφατη ανάπτυξη της νευροεπιστήμης από τα μέσα της δεκαετίας του 1990, ο Plog χρησιμοποιεί την “ψυχογραφική μέθοδο” ανάλυσης για την κατάτμηση και ανάλυση της τουριστικής αγοράς. Αρχικά, το δείγμα της ζήτησης (ταξιδιώτες) κατανέμεται ανάλογα με την προτίμησή του σε κοντινούς, γνωστούς, οικείους προορισμούς ή σε μακρινούς, άγνωστους προορισμούς. Η κατανομή είναι προσεγγιστικά κανονική. Στην αριστερή ουρά της κατανομής βρίσκονται οι λεγόμενοι ψυχοκεντρικοί ή εξαρτώμενοι ταξιδιώτες, που δείχνουν πολύ μεγάλη ή αποκλειστική προτίμηση σε σχετικά κοντινούς προορισμούς μαζικού τουρισμού που εξασφαλίζουν προβλέψιμο πρόγραμμα διακοπών. Στη δεξιά ουρά της κατανομής βρίσκονται οι λεγόμενοι αλλοκεντρικοί ταξιδιώτες ή εξερευνητές. Οι εξερευνητές είναι απρόβλεπτοι ή επιδιώκουν το απρόβλεπτο στα ταξίδια τους και δοκιμάζουν νέους προορισμούς. Σε απόσταση μεταξύ μίας και δύο τυπικών αποκλίσεων από τον μέσο βρίσκονται οι προσεγγιστικά ψυχοκεντρικοί και οι προσεγγιστικά αλλοκεντρικοί ταξιδιώτες αντίστοιχα, ενώ σε απόσταση το πολύ μίας τυπικής απόκλισης είναι περίπου τα 2/3 του δείγματος, οι ενδιάμεσοι ταξιδιώτες που προτιμούν σημαντική εναλλαγή μεταξύ γνώριμων και καινούριων προορισμών. Στη συνέχεια, οι τουριστικοί προορισμοί κατανέμονται ανάλογα με το ποιο από τα παραπάνω πέντε ψυχολογικά προφίλ τούς

προτιμά και σε τι βαθμό. Η κατανομή των τουριστικών προορισμών έδειξε ότι πολλοί προορισμοί που προσέλκυαν “εξερευνητές” στην αρχική μελέτη έτειναν να προσελκύουν ενδιάμεσους ή κι “εξαρτώμενους” στην επικαιροποιημένη.

Με βάση τα παραπάνω, προτάθηκε ένα γενικό μοντέλο ερμηνείας της ανόδου και πτώσης της δημοφιλίας των τουριστικών προορισμών. Αρχικά, οι τουριστικοί προορισμοί γίνονται δημοφιλείς όταν αρχίζουν να ελκύουν ολοένα και περισσότερους αλλοκεντρικούς ή προσεγγιστικά αλλοκεντρικούς ταξιδιώτες. Με τις θετικές τους εμπειρίες, οι “εξερευνητές” καθιερώνουν έναν τόπο ως έγκυρο τουριστικό προορισμό για όλους τους ταξιδιώτες, δίνουν κίνητρα στους τοπικούς συμμετοχούς να επενδύσουν σε τουριστική υλική υποδομή (π.χ. καταλύματα) αλλά και άυλη (π.χ. σχετική εκπαίδευση του πληθυσμού). Συνεπώς, οι “εξερευνητές” προκαλούν αύξηση της δημοφιλίας ενός τουριστικού προορισμού μέσω της προσέλκυσης ταξιδιωτών ενδιάμεσου προφίλ. Όμως, όταν τα χαρακτηριστικά του τουριστικού προορισμού αλλοιωθούν (π.χ. μαζικοποίηση περιηγήσεων, υπερδότηση), η τάση είναι να προσελκύονται σε αυτόν προσεγγιστικά ψυχοκεντρικοί ή ψυχοκεντρικοί ταξιδιώτες. Αυτοί οι ταξιδιώτες είναι λιγότεροι σε αριθμό από τους ενδιάμεσους κι επιλέγουν προορισμό με βάση την ευκολία πρόσβασης και την προβλεψιμότητα του προγράμματος του ταξιδιού τους, εξουδετερώνοντας με αυτόν τον τρόπο ό,τι συγκριτικό πλεονέκτημα του τουριστικού προορισμού δεν έχει ήδη αλλοιωθεί. Το αποτέλεσμα είναι ο τουριστικός προορισμός να πέφτει σε δημοφιλία και τελικά να χάνει την ίδια την υπόστασή του ως τουριστικός προορισμός.

Το μοντέλο του Plog έδειξε τη σημασία της στοχευμένης προώθησης ενός τουριστικού προορισμού η οποία, αν γίνει σωστά κι έγκαιρα, μπορεί να ανατρέψει την παραπάνω πτωτική τάση. Ο Plog αναφέρει τέτοια αντιπαραδείγματα στη μελέτη του αλλά εκτιμά ότι, αν δεν αλλάξει η στρατηγική προώθησης τουριστικών προορισμών προς την κατεύθυνση της προσέλκυσης “εξερευνητών”, ο τουριστικός κλάδος θα συρρικνωθεί μόνιμα υπέρ εναλλακτικών μεθόδων ψυχαγωγίας. Μέχρι στιγμής, αυτός ο φόβος αποδεικνύεται αβάσιμος, ενώ η ίδια η ανάλυση δεν ήταν δυνατόν να προβλέψει την τάση εξατομίκευσης της προώθησης προϊόντων, τη ραγδαία τεχνολογική βελτίωση των μέσων μεταφοράς και τη μεγαλύτερη ευκολία προγραμματισμού ενός ταξιδιού μέσω διαδικτύου.

Η έρευνα των Bigné και Sanchez (Bigné & Sanchez, 2001) δείχνει ότι η στοχευμένη προσέγγιση των δυνητικών τουριστών που προτείνει ο Plog θα είναι αποτελεσματική αν έχει ως αντικείμενο την εικόνα που έχει ο δυνητικός τουρίστας για τον τόπο. Αυτή είναι μία γενικότερη τάση στην προώθηση προϊόντων, και όχι μόνο τουριστικών τόπων, η οποία στοχεύει στην κοινωνική – συμβολική αξία των καταναλωτικών προτιμήσεων και όχι τόσο στα αντικειμενικά χαρακτηριστικά τους. Σύμφωνα με αυτή την έρευνα, η τουριστική εικόνα ενός τόπου επηρεάζεται από την υποκειμενική ποιότητα, την ικανοποίηση, την πρόθεση επιστροφής και την προθυμία θετικής σύστασης του τόπου.

Σε επιχειρησιακό επίπεδο, η προσέλκυση “εξερευνητών” και η βελτίωση της εικόνας ενός τουριστικού τόπου είναι πιθανό ότι μπορούν να υλοποιηθούν μέσα από την προώθηση συγκεκριμένων δραστηριοτήτων ή συγκεκριμένων κατηγοριών εξειδικευμένου τουρισμού (οικοτουρισμός, πολιτισμικός τουρισμός, θρησκευτικός κτλ), όπως προτείνουν οι Johns και Gyimothy (Johns & Gyimóthy, 2002). Η ίδια έρευνα αναδεικνύει έμμεσα το βασικότερο πρόβλημα της τουριστικής αγοράς, την εποχικότητά της. Η μεγάλη πλειοψηφία των τουριστικών προορισμών προσελκύει τουρίστες σε συγκεκριμένες εποχές του χρόνου, με αποτέλεσμα να υπάρχει μη αποδοτική χρήση πόρων τους υπόλοιπους μήνες. Ο βασικός στόχος των περισσότερων τουριστικών επιχειρηματικών σχεδίων είναι πλέον η προσέλκυση τουριστών σε όλη τη διάρκεια του έτους. Μία άλλη έρευνα (Zhou, 2014) αναλύει ποιοτικά μία συγκεκριμένη εικόνα τουριστικού προορισμού, αυτή της “αγροτικότητας” (rurality), κι επισημαίνει την ανάγκη προώθησης συγκεκριμένων σχετικών δραστηριοτήτων για επιβεβαίωση κι ενίσχυση αυτής της εικόνας. Μία επαναλαμβανόμενη χρήση γνώριμων περιγραφών και οπτικοακουστικών ερεθισμάτων μπορεί να οδηγήσει σε “αισθητική κόπωση” (aesthetic fatigue) και να έχει αρνητική καθαρή επίπτωση στην εικόνα ενός τουριστικού προορισμού.

Τα τελευταία χρόνια, η έρευνα στον τουριστικό κλάδο έχει στραφεί στην ανάλυση συμμετόχων (stakeholder analysis). Ο λόγος είναι ότι στον τουριστικό κλάδο έχει παρατηρηθεί ότι ο ανταγωνισμός είναι μεταξύ τόπων και όχι μεταξύ επιχειρήσεων ή ατόμων: ο δυνητικός τουρίστας πρώτα θα επιλέξει πού θα πάει και μετά θα κανονίσει τις υπόλοιπες λεπτομέρειες του ταξιδιού του. Οι Elbe κι Emmoth (Elbe & Emmoth, 2014) ασχολούνται με την ανάλυση συμμετόχων στον τουριστικό κλάδο από

νεοθεσμική σκοπιά, με τη χρήση της έννοιας της νομιμοποίησης (legitimacy). Η βασική υπόθεση είναι ότι κάθε οργανισμός παροχής τουριστικών υπηρεσιών καλείται να πείσει τους πελάτες του, τους προμηθευτές του, τους ανταγωνιστές τους, τις αρχές και την τοπική κοινωνία ότι καλώς υπάρχει και δραστηριοποιείται στον συγκεκριμένο κλάδο (ηθική νομιμοποίηση), τον οποίο γνωρίζει εκτενώς (γνωσιακή νομιμοποίηση) κι επομένως μπορεί να είναι ένας καλός συνétairos, συνεργάτης ή έστω συνομιλητής (πραγματική νομιμοποίηση). Σε αυτή την προσπάθειά του, κάθε οργανισμός επιστρατεύει τη ρητορική πειθώ των μελών του. Οι Elbe κι Emmoth (Elbe and Emmoth, 2014) κατηγοριοποιούν την πειθώ σε επίκληση στη λογική (logos), επίκληση στο συναίσθημα (pathos) κι επίκληση στο ήθος του συνομιλητή (ethos) και στη συνέχεια ισχυρίζονται ότι κάθε συμμετόχος πείθει και πείθεται από διαφορετικά είδη ρητορικής, από τα οποία αντλεί ή δίνει διαφορετικά είδη νομιμοποίησης. Συγκεκριμένα, η επίκληση στη λογική είναι το συνηθέστερο ρητορικό μέσο για την προσπάθεια ηθικής και πραγματικής νομιμοποίησης του οργανισμού. Η γνωσιακή νομιμοποίηση θεωρείται περίπου δεδομένη. Η επίκληση στο συναίσθημα χρησιμοποιείται κυρίως από τοπικούς ή περιφερειακούς οργανισμούς παροχής τουριστικών υπηρεσιών και αναφέρεται κυρίως στις εξωτερικές απειλές και τις αρνητικές επιπτώσεις της μη νομιμοποίησής τους στην τοπική κοινωνία (π.χ. υψηλότερο ποσοστό ανεργίας). Η επίκληση στο ήθος του συνομιλητή δεν χρησιμοποιείται σχεδόν καθόλου γιατί δεν υπάρχει κάποια τυπική, επίσημη δέσμευση των συμμετόχων να “νομιμοποιήσουν” έναν οργανισμό παροχής τουριστικών υπηρεσιών.

Εκτός από τη νομιμοποίηση, η ανάλυση συμμετόχων μπορεί να αναδείξει την πιθανότητα ένας ταξιδιώτης να επισκεφτεί επανειλημμένα τον ίδιο τόπο ως τουρίστας και μακροπρόθεσμα να σκεφτεί τη μετεγκατάστασή του σε αυτόν, κυρίως στην περίπτωση των μεγάλων αστικών κέντρων. Ο τουρισμός ως μέσο προσέλκυσης ταλέντου από την εγχώρια και την ξένη αγορά εργασίας και όχι ως αυτοσκοπός δεν είναι καινούρια πρακτική και συχνά έχει τη μορφή οργανωμένων εκδρομών ή εκπαιδευτικών ταξιδιών, που οργανώνονται από τους ίδιους τους συμμετόχους ενός αστικού τουριστικού τόπου. Οι Noni, Orsi και Zanderighi (Noni, Orsi & Zanderighi, 2014) εντοπίζουν ως σημαντικότερους παράγοντες μεγαλύτερης αυτοδέσμευσης ενός ταξιδιώτη με έναν τόπο, τις επιλογές εκπαίδευσης και τον πολιτισμό που ο τόπος αυτός προσφέρει. Οι ευκαιρίες απασχόλησης είναι σημαντικός παράγοντας αλλά όχι

ο πιο σημαντικός, δεδομένου ότι η σύγχρονη αγορά εργασίας απαιτεί ολοένα και μεγαλύτερη κινητικότητα από τους εργαζομένους. Είναι αξιοσημείωτο ότι οι δημόσιες παροχές ενός αστικού κέντρου δεν παίζουν μεγάλο ρόλο ούτε στην τουριστική του εικόνα ούτε στην προοπτική του ως πιθανό τόπο μετεγκατάστασης.

Σε αντίθεση με τα παραπάνω, ο Zemla (Zemla, 2008) υποστηρίζει ότι η συναίνεση και συνεργασία των συμμετόχων σε έναν τουριστικό τόπο δεν είναι αρκετή για να αυξήσει την ανταγωνιστικότητά του και να του δώσει συγκριτικό πλεονέκτημα. Συχνά οι τουρίστες θα χρησιμοποιήσουν έναν τουριστικό τόπο ως βάση για να εξερευνήσουν μία ευρύτερη περιοχή. Επίσης, ο αυξημένος ανταγωνισμός μεταξύ γειτονικών και ομοειδών τουριστικών προορισμών μειώνει την αποδοτικότητα της προώθησης. Μία διαφημιστική καμπάνια που αφορά ένα σύνολο τόπων, π.χ. της ίδιας χώρας, θα μπορούσε να πετύχει το ίδιο ή και καλύτερο αποτέλεσμα με μικρότερο κόστος. Ακόμα, η συναίνεση και συνεργασία των συμμετόχων ενός συγκεκριμένου τόπου δεν είναι πάντα σαφώς στοχευμένη προς παροχή συγκεκριμένου είδους τουριστικών υπηρεσιών: ένας τουριστικός τόπος μπορεί να ανήκει σε περισσότερες από μία κατηγορίες ή δίκτυα τουριστικών τόπων (ομοειδείς, ομοεθνείς, γειτονικοί κτλ).

Αν και η συνεργασία μεταξύ ομοεθνών τόπων φαίνεται να είναι δύσκολη αλλά πιο εφικτή από τη συνεργασία μεταξύ τόπων με μεταξύ τους πολιτισμικές διαφορές, η έρευνα του Zemla δεν δίνει σαφή απάντηση αν η διατοπική συνεργασία πρέπει να γίνεται με επίσημο ή ανεπίσημο τρόπο. Η βέλτιστη απόφαση φαίνεται να εξαρτάται από τη διάθεση συνεργασίας των τοπικών αρχών, ιδίως μέσα από σχήματα συμμετοχής και συνδιαμόρφωσης πολιτικής όπως οι ομάδες δομημένης συζήτησης (focus groups) και τα πάνελ πολιτών (citizens' panels). Σε κάθε περίπτωση, ως μεγαλύτερος ανασχετικός παράγοντας διατοπικής συνεργασίας αναδεικνύεται η έλλειψη εμπιστοσύνης μεταξύ των συμμετόχων των διαφορετικών τόπων, καθώς επίσης και το κόστος μετακίνησης από έναν τόπο στον άλλο, περιορισμός που επιβάλλεται τόσο από το γεωμορφολογικό τοπίο όσο και από την εκάστοτε τεχνολογία μεταφορών.

Με εξαίρεση ορισμένες από τις πιο πρόσφατες, οι περισσότερες από τις παραπάνω έρευνες έγιναν με τη χρήση κλασικών ποιοτικών και ποσοτικών μεθόδων. Η συνεχής ανάπτυξη των στατιστικών πακέτων, η εξέλιξη της στατιστικής επιστήμης με την

επινόηση κι εφαρμογή νέων μεθόδων επεξεργασίας δεδομένων (π.χ. ανάλυση συστάδων) και η δυνατότητα εξόρυξης και χρήσης ολοένα και περισσότερων δεδομένων συνιστούν σημαντική τομή στην ανάλυση του τουριστικού κλάδου.

#### 4.2 Ανάλυση Τουριστικού Κλάδου Βάσει Δεδομένων

Η βασικότερη συνεισφορά σε αυτό το πρόσφατο αντικείμενο προέρχεται από τη Sara Dolnicar (Dolnicar, 2002 και Dolnicar, 2014). Πριν τη μεταέρευνα του 2002, η ανάλυση του τουριστικού κλάδου βάσει δεδομένων με τη χρήση συστάδων ακολουθούσε άτυπα ορισμένες παραδοχές σε μεγάλο βαθμό: τα δεδομένα ήταν συνήθως τακτικά (η αριθμητική τιμή δεν έχει φυσική σημασία και κατά συνέπεια το ίδιο ισχύει για την ευκλείδεια απόστασή τους) και προεπεξεργασμένα, ο διαχωρισμός σε συστάδες γινόταν με βάση τον αλγόριθμο του Ward, ο προσδιορισμός του πλήθους τους γινόταν με καθαρά υποκειμενικά κριτήρια του ερευνητή και η ευστάθεια της λύσης δεν εξεταζόταν.

Η Dolnicar (Dolnicar, 2002) αρχικά παρατήρησε ότι η χρήση εκ των προτέρων χαρακτηριστικών ή τυπολογιών για τη δημιουργία συστάδων (π.χ. φύλο) τείνει να μειωθεί προς όφελος της χρήσης εκ των υστέρων χαρακτηριστικών ή ταξινομιών (π.χ. ικανοποίηση). Στην πράξη, τα εκ των υστέρων στοιχεία αποδείχτηκαν χρησιμότερα στη δημιουργία συστάδων και τον εντοπισμό του συγκριτικού πλεονεκτήματος ή του σημείου επικέντρωσης. Στη συνέχεια του άρθρου, διαπιστώνεται ότι το μέγεθος του δείγματος και το πλήθος των συστάδων δεν συσχετίζεται θετικά, αντίθετα απ' ό,τι αναμενόταν. Αυτό το αντιδιαισθητικό συμπέρασμα δείχνει μία τάση συμπερίληψης υπερβολικά πολλών μεταβλητών για το μέγεθος του εκάστοτε δείγματος, η οποία κατά κανόνα δυσκολεύει τον εντοπισμό συστάδων ακόμα και με το απλό κριτήριο της ελάχιστης συνδιακύμανσης, υπονομεύοντας έτσι την ανθεκτικότητα (robustness) της “προτεινόμενης” δομής δεδομένων.

Στην ίδια έρευνα εντοπίστηκε σημαντικό πρόβλημα στην παρουσίαση της προεπεξεργασίας των δεδομένων από τους ερευνητές. Οι περισσότεροι επιλέγουν με ανεπαρκή αιτιολόγηση την παραγοντική ανάλυση, με σκοπό την εύρεση συσχετίσεων μεταξύ των μεταβλητών και κατά συνέπεια την απάλειψη ορισμένων από αυτές. Παρόλα αυτά, η παραγοντική ανάλυση αλλοιώνει τη δομή των αρχικών δεδομένων,



ιδίως όταν αυτά είναι τακτικά όπως συμβαίνει στην πλειοψηφία των περιπτώσεων ανάλυσης τουριστικής αγοράς, και οδηγεί σε μη βέλτιστη δημιουργία συστάδων. Παρόμοιο πρόβλημα απώλειας κρίσιμης πληροφορίας υπάρχει και στην τυποποίηση των τιμών ως μέθοδο προεπεξεργασίας. Είναι αξιοσημείωτο ότι το 38% των ερευνητών δεν προεπεξεργάζεται καθόλου τα δεδομένα.

Επίσης, ο αλγόριθμος διαχωρισμού σε συστάδες δεν συσχετίζεται σημαντικά με το μέγεθος του δείγματος. Οι ιεραρχικοί αλγόριθμοι<sup>1</sup> όπως αυτός του Ward τείνουν να είναι καταλληλότεροι για μικρότερα δείγματα γιατί η εφαρμογή τους απαιτεί σε κάθε βήμα τον υπολογισμό των ανά δύο αποστάσεων των παρατηρήσεων. Αντίθετα, οι επαναληπτικοί αλγόριθμοι όπως των k-μέσων δεν έχουν παρόμοιους υπολογιστικούς περιορισμούς και μπορούν να χρησιμοποιηθούν αποδοτικά σε μεγαλύτερα δείγματα. Παρόλα αυτά, η μέση τιμή του πλήθους παρατηρήσεων των ερευνών που χρησιμοποιούν ιεραρχικούς αλγόριθμους είναι 1077 και η αντίστοιχη μέση τιμή για τις έρευνες που χρησιμοποιούν επαναληπτικούς αλγόριθμους είναι 1245, μία διαφορά που δεν είναι στατιστικά σημαντική.

Άλλα σημαντικά προβλήματα στη δημιουργία συστάδων για την ανάλυση του τουριστικού κλάδου βάσει μεγάλων δεδομένων εντοπίστηκαν στο μέτρο της απόστασης μεταξύ των παρατηρήσεων και το πλήθος των συστάδων. Η συνήθης ευκλείδεια απόσταση<sup>2</sup> είναι κατάλληλη για δεδομένα μετρικού χώρου, όπου η απόσταση έχει φυσική σημασία, καθώς και για δυαδικά δεδομένα. Αντίθετα, η ευκλείδεια απόσταση μεταξύ τακτικών δεδομένων εξορισμού δεν έχει νόημα. Αν και η έρευνα δεν συστήνει συγκεκριμένη μέθοδο, τα τακτικά δεδομένα συνήθως περιγράφονται από μέτρα θέσης (πχ διάμεσος, εκατοστημόρια) και μη παραμετρικές αποστάσεις. Η ίδια έλλειψη συγκεκριμένης σύστασης ισχύει και στο πλήθος των συστάδων. Αν και υπάρχουν διάφορες εμπειρικές ερευνητικές μέθοδοι του βέλτιστου πλήθους, στην πράξη συνήθως το πλήθος επιλέγεται καθαρά υποκειμενικά από τον ερευνητή.

---

<sup>1</sup> Ιεραρχικοί αλγόριθμοι: Η ιεραρχική μέθοδος αναφέρεται στη μέθοδο της ανάλυσης κατά συστάδες και λειτουργεί ιεραρχικά με την έννοια ότι ξεκινά εντάσσοντας κάθε παρατήρηση σε μια ομάδα και σε κάθε βήμα ενώνει ομάδες, όπου οι παρατηρήσεις τους βρίσκονται πιο κοντά.

<sup>2</sup> Ο όρος ευκλείδεια απόσταση για την ανάλυση κατά συστάδες σημαίνει ότι θα γίνει σύγκριση των τιμών κάθε μεταβλητής, κάθε παρατήρησης. Οι παρατηρήσεις αναπαριστούνται πάνω σε μια γραφική παράσταση και οι μεταβλητές καθορίζουν ένα σημείο. Ευκλείδεια συγκρίνουμε αυτό το σημείο με το κέντρο της ομάδας που είναι και αυτό ένα σημείο.

Το αποτέλεσμα είναι ότι το πλήθος των συστάδων (συνήθως τρεις ή τέσσερις) είναι στατιστικά ανεξάρτητο από το μέγεθος του δείγματος και τα άλλα χαρακτηριστικά της έρευνας. Συμπερασματικά, το κριτήριο διαχωρισμού σε συστάδες π.χ. ανά δραστηριότητα, ανά κατά κεφαλή δαπάνη κτλ, είναι λιγότερο σημαντικό από την καθαυτή διαδικασία δημιουργίας συστάδων, η οποία δεν αποτελεί τόσο μέθοδο όσο “εργαλειοθήκη” που πρέπει να προσαρμόζεται στο εκάστοτε μέγεθος του δείγματος και δομή του πληθυσμού.

Την ίδια χρονιά, οι Johns και Gyimothy (Johns & Gyimóthy, 2002) δημοσιεύουν ένα άρθρο κατάτμησης τουριστικής αγοράς βάσει δεδομένων με τη χρήση συστάδων που σχεδόν στερεοτυπικά ανταποκρίνεται στο προφίλ που σκιαγράφησε η Dolnicar (Dolnicar, 2002). Τα δεδομένα είναι τακτικά, χρησιμοποιείται παραγοντική ανάλυση και ιεραρχικός αλγόριθμος για τον προσδιορισμό των συστάδων, ενώ το τελικό αποτέλεσμα επαληθεύεται βάσει εξωτερικής πληροφορίας (αποτελέσματα παρόμοιων ερευνών).

Μία μεταγενέστερη έρευνα προσομοίωσης (Dolnicar, 2014) προσπαθεί να εντοπίσει έναν γενικό εμπειρικό κανόνα επιλογής μεγέθους δείγματος ως προς το πλήθος των μεταβλητών. Με τη χρήση του προσαρμοσμένου δείκτη Ραντ, που υπολογίζει κατά πόσο η δομή των δεδομένων είναι συμβατή με τον επιλεγμένο διαχωρισμό σε συστάδες, καταλήγει σε έναν εμπειρικό κανόνα μεγέθους δείγματος τουλάχιστον 70 φορές μεγαλύτερο από τον αριθμό των μεταβλητών. Δεδομένου ότι στα ίδια άρθρα συνιστάται η χρήση λίγων μεταβλητών και η εκ των προτέρων αποκοπή των μεταβλητών θορύβου, συμπεραίνεται ότι ένα καλό μέγεθος δείγματος για ένα μοντέλο τριών ερμηνευτικών μεταβλητών, όπως αυτό της παρούσας εργασίας, είναι τουλάχιστον 210.

Ο Akerkar (Fuchs, Horken & Lexhagen, 2014) εστιάζει περισσότερο στις μελλοντικές προοπτικές χρήσης μεγάλων δεδομένων στον τουριστικό τομέα. Στην αναφορά του, επισημαίνει ότι οι εμπλεκόμενοι στην προσφορά τουριστικών υπηρεσιών έχουν εδώ και χρόνια πρόσβαση σε πολλά δεδομένα αλλά δεν τα αξιοποιούν κατάλληλα. Οι λόγοι είναι τόσο τεχνικοί (έλλειψη υπολογιστικής ισχύος της υπάρχουσας υποδομής τους) όσο και ουσιαστικοί: τα μεγάλα δεδομένα είναι συχνά αδόμητα και περιλαμβάνουν πολλή περιττή για τους σκοπούς του αναλυτή πληροφορία. Οι Fuchs, Horken και Lexhagen (2014) εντοπίζουν το ίδιο πρόβλημα

της ανεπαρκούς αξιοποίησης των μεγάλων δεδομένων και προσπαθούν να το λύσουν με τη χρήση πληροφοριακών συστημάτων διοίκησης προορισμών (Destination management information systems – DMIS). Τα DMIS χρησιμοποιούνται ήδη σε άλλες εφαρμογές ανάλυσης επιχειρηματικής ευφυΐας. Η συγκεκριμένη έρευνα είναι η πρώτη που χρησιμοποιεί DMIS στον τουριστικό κλάδο, αντλώντας δεδομένα από τις διαδικτυακές αναζητήσεις των επίδοξων τουριστών, τα στοιχεία των κρατήσεών τους και την ανατροφοδότηση που δίνουν σχετικά με την ποιότητα των παρεχόμενων υπηρεσιών.

Οι Hjalager και Nordin (2011) στη δική τους έρευνα εξετάζουν τις διάφορες μεθόδους ανάλυσης της πελατοκεντρικής ή βασισμένης στον χρήστη καινοτομίας, η οποία τέμνεται με την ανάλυση αγοράς βάσει δεδομένων στον βαθμό που πολλά από αυτά τα δεδομένα προέρχονται ακριβώς από τη δραστηριότητα των πελατών. Ορισμένες από αυτές τις μεθόδους απαιτούν την ενεργή συμμετοχή των χρηστών και κάποιες άλλες όχι, ενώ ορισμένες μέθοδοι επιλέγουν να παίρνουν “λίγα από πολλούς” (λίγες μεταβλητές – μεγάλο δείγμα) για περισσότερες δυνατότητες στατιστικού χειρισμού των δεδομένων και άλλες “πολλά από λίγους” (πολλές μεταβλητές – μικρό δείγμα) ώστε να εξασφαλίσουν την καλύτερη δυνατή εξέταση συγκεκριμένων μόνο χρηστών.

Όμως, με την έλευση και ανάπτυξη των Νέων Τεχνολογιών και του Διαδικτύου άλλαξαν ριζικά και σημαντικά τα κανάλια επικοινωνίας και οι δυνατότητες, με αποτέλεσμα να επαναπροσδιοριστούν ξανά οι ρόλοι και οι αλληλεξαρτήσεις. Ο κύριος λόγος που οι Νέες Τεχνολογίες και το Διαδίκτυο έχει επαναπροσδιορίσει τις σχέσεις αλληλεξάρτησης και το ρόλο της κάθε επιχείρησης στην τουριστική αλυσίδα διανομής και αξίας αποτελεί το γεγονός ότι το Διαδίκτυο προσδίδει δυνατότητες οικονομικής και αποτελεσματικής άμεσης επικοινωνίας με τον διεθνή τουρίστα σε κάθε είδους τουριστική επιχείρηση και οργανισμό.

## ΜΕΡΟΣ II

### Σύγχρονες Τεχνικές Εξόρυξης Δεδομένων Μεγάλης Κλίμακας στον Τομέα του Τουρισμού

#### Κεφάλαιο 5

#### Εφαρμογή Δεδομένων Μεγάλης Κλίμακας στον Τουρισμό

##### 5.1 Εισαγωγή

Η τεχνολογία των πληροφοριών οδηγεί στην δημιουργία νέων μοτίβων στη συμπεριφορά των ταξιδιωτών κατά την λήψη αποφάσεων. Γενικά νοείται ότι η διαδικασία ταξιδιού περιλαμβάνει τρία στάδια, στο προ-ταξιδιωτικό, στο καθ' οδόν και επί τόπου, και στο μετά-ταξιδιωτικό, κατά τα οποία ο ταξιδιώτης αναλαμβάνει διαφορετικές δραστηριότητες χρήσης πληροφοριών και αλληλεπίδρασης με το περιβάλλον. Πρόσφατες μελέτες έδειξαν ότι η χρήση των smartphones κατά τη διάρκεια του ταξιδιού είναι δυνατό να αλλάξουν την ταξιδιωτική εμπειρία "ξεκλειδώνοντας" την διαδικασία των τριών σταδίων (Wang, Xiang & Fesenmaier, 2014). Οι δραστηριότητες στις οποίες επιβάλλονταν οι τουρίστες στο προ-ταξιδιωτικό και στο μετά-ταξιδιωτικό στάδιο, αρχίζουν ολοένα και περισσότερο να εκπληρώνονται στο καθ' οδόν και επί τόπου στάδιο λόγω της γενικευμένης σύνδεσης στο διαδίκτυο οδηγώντας σε σημαντικές αλλαγές συμπεριφοράς. Το επίπεδο της ευελιξίας λήψης αποφάσεων κατά τη διάρκεια του ταξιδιού είναι πιθανό να αυξηθεί μιας και ο ταξιδιώτης θα μπορεί εύκολα να αλλάξει το αρχικό του σχέδιο λόγω της διαθεσιμότητας νέων πληροφοριών. Τελικά οι ταξιδιωτικές δραστηριότητες μπορεί να γίνουν πιο αυθόρμητες, με συνέπεια να πραγματοποιούνται περισσότερα απρογραμματίστα ταξίδια ή δραστηριότητες. Η διαδικασία λήψης αποφάσεων στο καθ' οδόν στάδιο είναι πιο δυναμική δεδομένου ότι περιλαμβάνει μια σειρά από αλληλεξαρτώμενες αποφάσεις εντός των οποίων τα πλαίσια μεταγενέστερων αποφάσεων εξαρτώνται από προηγούμενες αποφάσεις. Συνεπώς η χρήση έξυπνων συσκευών όπως smartphones αλλάζει τα δεδομένα σε επίπεδο λήψης αποφάσεων για το καθ' οδόν και επί τόπου στάδιο ειδικά λαμβάνοντας υπόψη την ευκολία πρόσβασης σε μηχανές αναζήτησης και μέσω κοινωνικής δικτύωσης οπουδήποτε και οποτεδήποτε. Αυτό προκαλεί τις ως τώρα γνώσεις μας αναφορικά με τη

διαδικασία λήψης αποφάσεων ταξιδιών και δημιουργεί την ανάγκη για μια νέα σειρά εργαλείων ανάλυσης που μπορούν να αποθηκεύουν και να μετρούν τη διαδικασία και τη δομή της ταξιδιωτικής συμπεριφοράς.

Νέες τεχνολογικές καινοτομίες συνεχίζουν να αναδύονται και απαιτούν νέα οράματα για την ανάπτυξη του τουρισμού (Gretzel, Sigala, Xiang & Koo, 2015). Ένα τέτοιο παράδειγμα αποτελεί το Διαδίκτυο των Πραγμάτων το οποίο δηλώνει την παρουσία του γύρω μας μέσω ποικιλίας αντικειμένων όπως είναι οι αισθητήρες, οι ενεργοποιητές, οι κινητές συσκευές κλπ. Τα αντικείμενα αυτά είναι σε θέση να αλληλεπιδρούν μεταξύ τους και να συνεργάζονται με κοντινά τους αντικείμενα για την διεκπεραίωση κοινού έργου (Atzori, Iera & Morabito, 2010). Επίσης μέσω της δυνατότητας σύνδεσης στο διαδίκτυο γεφυρώνεται το χάσμα μεταξύ του πραγματικού κόσμου και του ψηφιακού χώρου. Επιπλέον, η ανάπτυξη φορητών υπολογιστών υποστηρίζει μια πληθώρα εφαρμογών συμπεριλαμβανομένης και της οπτικής σήμανσης των αντικειμένων και των συσκευών Near Field Communication (NFC) που συμβάλλουν στην ανάπτυξη του Διαδικτύου των Πραγμάτων (Borrego, Ruiz & Gomez, 2011). Το σημαντικό είναι ότι η εμφάνιση του Διαδικτύου των Πραγμάτων μετατόπισε την παροχή υπηρεσιών από το σημερινό όραμα των υπηρεσιών συνεχούς σύνδεσης στις πάντα διαθέσιμες υπηρεσίες, οι οποίες δημιουργούνται και συγκροτούνται κατά την εκτέλεση ώστε να ανταποκρίνονται στις αυξημένες ανάγκες των χρηστών. Έτσι, προβλέπεται ότι μέσα στην επόμενη δεκαετία το διαδίκτυο θα μπορεί να πραγματοποιήσει το όραμα πολλών ετών με το να επιτύχει την σύνδεση αντικειμένων που μπορούν να αναγνωριστούν, να εντοπιστούν, να παρακολουθηθούν και να διαχειριστούν οποιαδήποτε στιγμή και οπουδήποτε.

Αυτή η νέα τεχνολογική δομή δημιουργεί νέες δυνατότητες σύνδεσης και τρόπους αλληλεπίδρασης εντός και εκτός ταξιδιού καθιστώντας δυνατό να επηρεαστεί ο τρόπος με τον οποίο αντιλαμβανόμαστε τη διαδικασία ταξιδιού. Ως εκ τούτου, είναι σαφές ότι οι τεχνολογικές εξελίξεις στον τομέα των κινητών, κοινωνικών, επικοινωνιακών και τεχνολογιών εντοπισμού θέσης έχουν επαυξήσει τις αισθήσεις και τις εμπειρίες των τουριστών μέσω της συναισθηματικής, αισθητικής, ενημερωτικής, παιχνιδιάρικης και κοινωνικής τους εμπλοκής, προκαλώντας τους να είναι πιο δημιουργικοί και αυθόρμητοι (Richards, 2011). Αυτές οι πρόσφατες εξελίξεις απαιτούν τη δημιουργία νέων μοντέλων συμπεριφοράς ταξιδιών, νέων

μοντέλων σχεδιασμού προϊόντων και νέων μοντέλων έρευνας και αξιολόγησης, που με τη σειρά τους θα θεσπίσουν ένα νέο πρότυπο τουριστικής διοίκησης.

## 5.2 Η Έννοια των Δεδομένων Μεγάλης Κλίμακας στον Τουρισμό

Η πληροφορία ευδοκimei στην τουριστική βιομηχανία. Οι τεράστιες νέες αποθήκες δεδομένων (πολύ μεγαλύτερες από αυτές που συλλέγονται σε παραδοσιακές βάσεις δεδομένων) μπορούν να προσφέρουν στους ερευνητές, τους διευθυντές και τους υπεύθυνους χάραξης πολιτικής, τα στοιχεία που απαιτούνται για τη λήψη αποφάσεων βάσει αριθμών και αναλύσεων αντί διαισθήσεων ή προηγούμενων εμπειριών (Frederiksen, 2012). Η αφθονία των δεδομένων μεγάλης κλίμακας στον τουρισμό έχει τη δυνατότητα να παράσχει νέα - άκρως ενημερωμένα συμπεράσματα σχετικά με την ανθρώπινη δραστηριότητα και συμπεριφορά που θα προσδώσει μεγάλη ώθηση στην τουριστική βιομηχανία και θα ωφελήσει όχι μόνο τους καταναλωτές αλλά και επαγγελματίες που δραστηριοποιούνται στην εν λόγω βιομηχανία (Fuchs, Hopken & Lexhagen, 2014).

Οι ταξιδιώτες αφήνουν διαφορετικά ψηφιακά ίχνη στο διαδίκτυο καθώς χρησιμοποιούν κινητές τεχνολογίες. Ο κάθε ταξιδιώτης μπορεί να διαθέσει μεγάλες ποσότητες δεδομένων για οποιοδήποτε από τα τρία ταξιδιωτικά στάδια: προ-ταξιδιωτικό, καθ' οδόν και επί τόπου και μετά-ταξιδιωτικό (Hendrik & Perdana, 2014). Τα περισσότερα από αυτά τα δεδομένα είναι εξωτερικής φύσης όπως είναι τα tweets ή τροφοδοσίες (feeds) άλλων μέσων κοινωνικής δικτύωσης.

Λόγω του μεγάλου όγκου αποθηκευμένων ακατέργαστων δεδομένων στο νέφος (cloud) και προκειμένου να εξαχθεί αξία και χρήσιμη πληροφορία από αυτά, απαραίτητη είναι η επεξεργασία και ανάλυσή τους. Πολλοί από τους δυνητικούς τουρίστες που προγραμματίζουν ένα ταξίδι δεν περιορίζονται σε μια απλή έρευνα στο διαδίκτυο κατά την αναζήτηση προορισμού, εισιτηρίων και διαμονής. Οι καταναλωτές των τουριστικών υπηρεσιών στρέφονται ολοένα και περισσότερο στα δεδομένα μεγάλης κλίμακας για την βελτίωση της λήψης αποφάσεων, την ανεύρεση ευκαιριών και την βελτίωση των εν γένει επιδόσεων τους (Irudeen & Samaraweera, 2013). Για παράδειγμα, τα Δεδομένα Μεγάλης Κλίμακας μπορούν να

χρησιμοποιηθούν για τη διασύνδεση των διάχυτων πληροφοριών από διαφορετικά συστήματα ώστε να βελτιωθεί η δυνατότητα λήψης αποφάσεων.

Τα δεδομένα μεγάλης κλίμακας παρέχουν πρωτοφανείς πληροφορίες για τον τρόπο λήψης αποφάσεων των πελατών, επιτρέποντας στις εταιρείες να εντοπίζουν και να αναλύουν τα καταναλωτικά πρότυπα, τις συστάσεις, την αγοραστική συμπεριφορά και άλλα κίνητρα που επηρεάζουν τις πωλήσεις. Οι εταιρείες που εμπλέκονται με τον τουρισμό μπορούν να βρουν καινοτόμους τρόπους χρήσης των ποικίλων πηγών δεδομένων ώστε να συνδεθούν με τους πιθανούς επισκέπτες σε κάθε στάδιο του ταξιδιού και να χρησιμοποιήσουν τις αποκτηθείσες πληροφορίες για την καλύτερη και έγκαιρη κατανόηση των δημογραφικών στοιχείων των επισκεπτών. Επίσης μπορούν να στοχεύσουν διαφημίσεις σε δυνητικούς αγοραστές που έχουν αναζητήσει κάποιον συγκεκριμένο προορισμό σε ιστοσελίδα ταξιδιωτικού πρακτορείου (Sust, Plera, Berengue, Garcia, Alonso & Torres, 2014). Με τη χρήση δεδομένων μεγάλης κλίμακας οι βιομηχανίες γίνονται πιο αποτελεσματικές. Όλο και περισσότερες εταιρείες αρχίζουν να εξειδικεύονται στην αποθήκευση και αξιολόγηση μεγάλου όγκου δεδομένων που σχετίζονται με διαμονές τουριστών σε ξενοδοχεία, τις συναλλαγές αγοράς και τις πληροφορίες πελατών προκειμένου να παρέχουν αποτελεσματικότερες και υψηλότερης ποιότητας υπηρεσίες.

### 5.3 Τα Πλεονεκτήματα Χρήσης Δεδομένων Μεγάλης Κλίμακας στον Τουρισμό

Το εξατομικευμένο μάρκετινγκ και η στοχευμένη σχεδίαση προϊόντος παρουσιάζουν εξαιρετικά ισχυρές ευκαιρίες τόσο για τους προμηθευτές τουριστικών προϊόντων όσο και για τους καταναλωτές. Είναι ξεκάθαρο ότι τα δεδομένα μεγάλης κλίμακας δίνουν τη δυνατότητα στις επιχειρήσεις να παρέχουν καλύτερες, στοχευμένες και πιο κερδοφόρες υπηρεσίες και προϊόντα στους καταναλωτές (Pries & Dunnigan, 2015). Για παράδειγμα οι αναλυτές δεδομένων μεγάλης κλίμακας μπορούν να ανιχνεύσουν πληροφορίες σχετικά με τα ενδιαφέροντα κάποιου καταναλωτή από φωτογραφίες που έχει αναρτήσει στο Facebook ή σε άλλα μέσα κοινωνικής δικτύωσης (πχ. ένας πάροχος τουριστικών υπηρεσιών που βλέποντας αναρτήσεις σχετικές με ορεινή ποδηλασία θα μπορούσε να προωθήσει τοπικούς προορισμούς ποδηλασίας ή λέσχες ποδηλασίας).

Προηγούμενες μελέτες με θέμα τον τουρισμό βασίστηκαν σε έρευνες ή απόψεις εμπειρογνομόνων πράγμα που σημαίνει ότι τα δείγματα λήφθηκαν από το σύνολο του πληθυσμού και όχι αμιγώς από το τουριστικό πλήθος. Αντίθετα, σε άλλη μελέτη με θέμα τα δεδομένα μεγάλης κλίμακας στον τουρισμό έγινε προσπάθεια παρουσίασης δεδομένων που βασίζονταν σε πραγματικές δραστηριότητες από όλους τους χρήστες τουριστικών προϊόντων αντί να αντληθούν πληροφορίες από δείγματα ερευνών (Irudeen & Samaraweera, 2013).

Όπως παρουσιάζεται παρακάτω τα δεδομένα μεγάλης κλίμακας στον τουρισμό με την χρήση καινοτόμων μεθόδων έχουν πολλά πλεονεκτήματα έναντι των παραδοσιακών μεθόδων:

### 1. Αξιοπιστία

Τα δεδομένα μεγάλης κλίμακας βασίζονται σε πραγματικά στοιχεία που προκύπτουν από τις ενέργειες των χρηστών και όχι σε έρευνες. Δηλαδή, έχουν αναλυθεί πραγματικές ενέργειες και όχι δηλωμένες προθέσεις ή απαντήσεις σε ερωτήσεις. Λαμβάνοντας υπόψη όλες τις πηγές πληροφόρησης μπορεί να πιστοποιηθεί ότι τα δεδομένα μεγάλης κλίμακας αυξάνουν το πλήθος δειγμάτων στο οποίο η συμβατική έρευνα τείνει να βασίζεται σε αρκετές τάξεις μεγέθους (Meeker & Hong 2014). Η αξιοπιστία της ανάλυσης μεγάλων δεδομένων μας επιτρέπει να εξετάσουμε όλες τις πτυχές των πληροφοριών προκειμένου να παρέχουμε ολοκληρωμένα αποτελέσματα αντί για μεροληπτικά συμπεράσματα λόγω απώλειας πληροφοριών στο πλήθος δειγμάτων.

### 2. Νέες ροές πληροφοριών

Τα τουριστικά δεδομένα μεγάλης κλίμακας αποτελούν ένα είδος πληροφοριών που παράγονται από τους ίδιους τους ταξιδιώτες, εμπλουτίζουν τη γνώση της τουριστικής αγοράς στόχου και είναι πολύ χρήσιμα για την ανάλυση της καταναλωτικής ζήτησης για διάφορα τουριστικά προϊόντα και υπηρεσίες (Hendrik & Perdana, 2014). Δεδομένου ότι τα τουριστικά δεδομένα μεγάλης κλίμακας είναι δομημένα και επανατοποθετημένα είναι δυνατόν να τα παραπέμψουμε σε άλλες πηγές όπως αυτές των μέσων κοινωνικής δικτύωσης και των ανοικτών δεδομένων, είτε πρόκειται για πηγές που δομούνται επί του παρόντος, είτε για πηγές πληροφοριών που ενδέχεται να δημιουργηθούν ή να δημοσιοποιηθούν στο μέλλον. Η ανάλυση των τουριστικών



δεδομένων μεγάλης κλίμακας μπορεί να συγκριθεί με τα εσωτερικά δεδομένα της κάθε τουριστικής επιχείρησης προκειμένου να καθοριστεί εάν η προσφορά τουριστικών προϊόντων / υπηρεσιών στην περιοχή μιας πόλης είναι σε αρμονία με την τουριστική ζήτηση.

### 3. Δεδομένα σε πραγματικό χρόνο και η άμεση πρόβλεψη (nowcasting)

Μία από τις καινοτόμες χρήσεις των δεδομένων μεγάλης κλίμακας είναι η άμεση πρόβλεψη (nowcasting), το οποίο ουσιαστικά είναι η χρήση των δεδομένων σε πραγματικό χρόνο για να περιγραφούν οι πρόσφατες δραστηριότητες προτού δημοσιοποιηθούν οι επίσημες πηγές δεδομένων (Bollier & Firestone, 2010). Σύμφωνα με τον Varian (2014) ένα καλό παράδειγμα άμεσης πρόβλεψης καταναλωτικής συμπεριφοράς αποτελούν οι αναζητήσεις της Google σε πραγματικό χρόνο, καθώς η ανάλυση ταυτόχρονου συσχετισμού που προσφέρεται από την Google Correlate εξακολουθεί να είναι κατά έξι εβδομάδες μπροστά από τις αναφερόμενες τιμές (reported values).

Πολλές μελέτες χρησιμοποίησαν δομημένα δεδομένα μηχανών αναζήτησης για την άμεση πρόβλεψη (nowcasting) και την πρόβλεψη στον τουρισμό (Artola, Pinto & Pedraza, 2015; Bangwayo & Skeete, 2015; Yang, Pan, Evans & Lv, 2015). Εκτός από τα δεδομένα μηχανών αναζήτησης υπάρχουν και άλλοι τύποι πραγματικών δεδομένων που μπορούν να συγκεντρωθούν και να αναλυθούν, όπως για παράδειγμα είναι τα δεδομένα που αντλούνται από τις αγορές πιστωτικών καρτών, η αποστολή και ο εντοπισμός δεμάτων και η χρήση κινητών τηλεφώνων. Πολλά από αυτά τα δεδομένα γίνονται διαθέσιμα σχεδόν σε πραγματικό χρόνο και τα οποία μπορούν να χρησιμοποιηθούν για την πρόβλεψη των μακροδεδομένων που θα συγκεντρωθούν κάποια στιγμή στο μέλλον (Jeng & Fesenmaier, 2002).

Ο απώτερος στόχος της χρήσης δεδομένων μεγάλης κλίμακας σε πραγματικό χρόνο είναι η ανάπτυξη εφαρμογών που θα είναι σε θέση να παράσχουν προτάσεις μόλις ληφθεί ο οικονομικός παλμός της επιχείρησης, αυτό βέβαια πρέπει να γίνεται υπό ελεγχόμενες συνθήκες και να υπάρχει η δυνατότητα διακοπής τους ανά πάσα στιγμή.

Η δυναμική της ανάλυσης δεδομένων μεγάλης κλίμακας αλλάζει τον κόσμο της φιλοξενίας και του ταξιδιού καθώς οι τουριστικές επιχειρήσεις έχουν να κάνουν με πληθώρα δεδομένων που καλύπτουν όλες τις πτυχές των πληροφοριών όπως είναι οι

επιβεβαιώσεις πτήσεων ή οι προτιμήσεις δωματίου πελατών, δημιουργώντας έτσι δυνατότητες συσχετισμού δεδομένων για την εύρεση πληροφοριών που άλλοτε ήταν άγνωστες (Turner, 2017). Ουσιαστικά τα δεδομένα μεγάλης κλίμακας μπορούν να χρησιμοποιηθούν για την προσαρμογή εκστρατειών μάρκετινγκ στην πραγματικότητα της εκάστοτε αγοράς και στην εύρεση ανεπαρκειών στα επιχειρηματικά μοντέλα. Η ανάλυση δεδομένων μεγάλης κλίμακας μπορεί να προσφέρει κομβικής σημασίας επιχειρηματικές πληροφορίες και να αποτελέσει την πηγή καινοτομίας για τους τουριστικούς οργανισμούς και τη βιομηχανία εν γένει. Οι δυνατότητες των μεγάλων δεδομένων στον τουρισμό είναι τεράστιες και οι τουριστικοί οργανισμοί δεν πρέπει να υποτιμούν τη σημασία τους (Pries & Dunnigan, 2015).

Με τη σωστή προσέγγιση ο τουριστικός κλάδος μπορεί να μάθει πολλά για τις προτιμήσεις των καταναλωτών και να χρησιμοποιεί αυτές τις πληροφορίες για την οικοδόμηση σχέσεων με τους μεμονωμένους τουρίστες. Η ικανότητα να προσφέρεται η σωστή υπηρεσία ή προϊόν τη σωστή στιγμή είναι καίριας σημασίας. Δίχως τις σωστές πληροφορίες και μια καλή στοχευμένη τουριστική στρατηγική, η διαφήμιση δεν θα αποδίδει και δεν θα μπορεί να παράξει αξία. Ο τουρισμός είναι μια βιομηχανία που κινείται σε υψηλούς ρυθμούς και αυτό γεννά την ανάγκη ταχείας ανάλυσης δεδομένων και γρήγορων αποφάσεων. Στον τουρισμό η οποιαδήποτε ζήτηση πρέπει να ικανοποιείται άμεσα προκειμένου ο πάροχος του προϊόντος / υπηρεσίας να παραμείνει προσιτός στους πελάτες, πράγμα που καθιστά τα δεδομένα μεγάλης κλίμακας εξαιρετικά σημαντικά.

Με την ραγδαία αύξηση του αριθμού των εφαρμογών κοινωνικής δικτύωσης, ανακύπτει η ανάγκη της αλληλεπίδρασης των παραδοσιακών τουριστικών δεδομένων με τα νέα. Για παράδειγμα τα τηλεφωνικά κέντρα θα αλληλεπιδρούν με τις ηλεκτρονικές κριτικές των πελατών, τα προγράμματα αφοσίωσης-ανταμοιβής θα συνδέονται με το ιστορικό κρατήσεων και οι προτιμήσεις διαμονής θα συνδυάζονται με τα chat των εφαρμογών κοινωνικής δικτύωσης.

#### 5.4 Η Συμπεριφορά των Καταναλωτών

Βρισκόμαστε σε μια εποχή πρωτοφανούς ρευστότητας καταναλωτικής συμπεριφοράς, προσδοκιών πελατών και επιχειρηματικών μοντέλων που

δημιουργήθηκαν από τις ίδιες τεχνολογίες, οι οποίες διαταράσσουν την λειτουργία καταξιωμένων επιχειρήσεων και ταυτόχρονα δημιουργούν νέες (Marko, 2015; McAfee, Brynjolfsson, Davenport, Patil & Barton, 2012). Ωστόσο τα στοιχεία των τουριστικών μεγάλων δεδομένων καταδεικνύουν σημαντικές αλλαγές στη σχέση μεταξύ των επιχειρήσεων και των πελατών τους. Έτσι τα μεγάλα δεδομένα μπορούν να χρησιμοποιηθούν για την παροχή ανώτερων υπηρεσιών αγοράς και υποστήριξης με σκοπό την ενίσχυση της επιλογής και των προσδοκιών των πελατών. Καταλυτικό ρόλο για την επανα-αναγνώριση της καταναλωτικής συμπεριφοράς με την χρήση μεγάλων δεδομένων, παίζει η εκτεταμένη χρήση κινητών συσκευών, εφαρμογών και άλλων μέσων κοινωνικής δικτύωσης, τα οποία διαδραματίζουν ολοένα και μεγαλύτερο ρόλο στη συλλογή ακατέργαστων δεδομένων.

Τα μεγάλα δεδομένα μπορούν να παρέχουν πολλές πληροφορίες σχετικά με τη καταναλωτική συμπεριφορά, ορισμένες από τις οποίες ήδη αποδίδουν ενώ άλλες δεν έχουν ανακαλυφθεί ακόμη. Οι δυνατότητες που δημιουργούνται από τη χρήση μεγάλων δεδομένων είναι μεγάλες ιδιαίτερα στο λιανικό εμπόριο καθώς οι βιομηχανίες και οι επιχειρησιακές διεργασίες μπορούν να εκμεταλλευτούν επιτυχώς τα νέα κανάλια επικοινωνίας και τις επιλογές παροχής υπηρεσιών (Marko, 2015). Η συλλογή, η συσχέτιση και η ανάλυση όλων των καναλιών των μεγάλων δεδομένων του τουρισμού από την αλληλεπίδραση με τους πελάτες είναι το κλειδί για τη μετατροπή της εμπειρίας των πελατών από εφιάλτη σε νιρβάνα (Chase, 2013). Ο σύνδεσμος μεταξύ των μεγάλων δεδομένων και της μηχανικής μάθησης σε όλες τις εκφάνσεις, συμπεριλαμβανομένης της προγνωστικής ανάλυσης όπως και των νευρωνικών δικτύων βαθιάς μάθησης, αποτελεί το θεμέλιο μιας καλά ενημερωμένης, άκρως αποτελεσματικής και ιδιαίτερα ικανοποιητικής αλληλεπίδρασης που ωφελεί τόσο τους πελάτες όσο και τις επιχειρήσεις.

Ο στόχος της χρήσης των τουριστικών μεγάλων δεδομένων είναι η δημιουργία μιας αυθεντικής συναισθηματικής σχέσης μεταξύ των πελατών και των συμμετεχόντων στην τουριστική βιομηχανία προκειμένου να βελτιωθεί σημαντικά ο τομέας της εξυπηρέτησης και υποστήριξης πελατών. Η εξερεύνηση των μεγάλων δεδομένων στον τουρισμό έχει τεράστιες επιπτώσεις και παρέχει ευκαιρίες για την απρόσκοπτη δικτύωση των καταναλωτικών εμπειριών κατά τη χρήση κινητών συσκευών, ιστοσελίδων και επικοινωνίας μέσω διάφορων καναλιών (τηλέφωνο, μηνύματα,

ηλεκτρονικό ταχυδρομείο, chat καθώς και τα μέσα κοινωνικής δικτύωσης). Το κλειδί για την χρήση των τουριστικών μεγάλων δεδομένων δεν είναι μόνο η παροχή ολοκληρωμένων υπηρεσιών αλλά η προορατικότητα. Πρέπει να μπορούμε να προβλέψουμε τις ανάγκες των πελατών και να προλαμβάνουμε τα προβλήματα. Με άλλα λόγια, μπορούμε να προβλέψουμε προβλήματα και απορίες με τη χρήση στατιστικών μοντέλων πρόβλεψης πριν μας γίνει η ίδια ερώτηση ή μας ζητηθεί να εξηγήσουμε επανειλημμένα την ίδια κατάσταση.

## 5.5 Μηχανισμοί Ανατροφοδότησης

Η ανατροφοδότηση ή η κριτική στην τουριστική βιομηχανία είναι σημαντική για τον εντοπισμό των προτιμήσεων των πελατών και για την μετέπειτα παροχή θετικών εμπειριών. Η επεξεργασία κριτικών είναι ένα από τα πιο σημαντικά στοιχεία για την επίτευξη ραγδαίας ανάπτυξης της επιχείρησης και για την οικοδόμηση μιας στρατηγικής που θα καλύπτει καλύτερα της ανάγκες των πελατών. Ανατροφοδότηση βάσει τουριστικών μεγάλων δεδομένων από πελάτες, υπαλλήλους, συνεργάτες, προμηθευτές και κοινότητες έχει επίσης βελτιώσει τις δυνατότητες της ανάλυσης μεγάλων δεδομένων. Οι επιχειρηματικές και καταναλωτικές εφαρμογές που βασίζονται στα δεδομένα αποτελούν τις πιο διαδεδομένες μεθόδους συλλογής κριτικών οποτεδήποτε και οπουδήποτε. Ολοένα και περισσότερες υπηρεσίες νεφών (cloud) μας δίνουν την άμεση και καθολική ικανότητα αλληλεπίδρασης μέσω έξυπνων τηλεφώνων, τάμπλετς (tablets) ή ακόμα και ρολογιών.

Η αυξημένη χρήση σύγχρονων τεχνικών για την συλλογή κριτικών έχει οδηγήσει το παραδοσιακό μάρκετινγκ κριτικών να αντικατασταθεί σταδιακά από εμπορικά μηνύματα που είναι άμεσα, κοφτά, μοναδικά και προσαρμοσμένα στον κάθε πελάτη. Μία από τις εφαρμογές των μηχανισμών ανατροφοδότησης που εφαρμόζουν οι τουριστικοί οργανισμοί είναι αυτή της προσαρμογής τιμών, κατά την οποία μια μεταβολή στην ζήτηση ταξιδιών που προκύπτει από προγνωστική ανάλυση μεγάλων δεδομένων μπορεί να παρέχει χρήσιμες πληροφορίες για την γρήγορη και αποτελεσματική προσαρμογή των τιμών.

Η μηχανική μάθηση είναι μία από τις βασικές μεθόδους που χρησιμοποιείται στον τουριστικό κλάδο για την κατασκευή του μηχανισμού ανατροφοδότησης μεταξύ

πελατών και τουριστικών πρακτορείων (Bajari, Nekipelov, Ryan & Yang, 2015). Παραδείγματος χάρη, μέσω συνέργειας μεταξύ τουριστικών οργανισμών, πιστωτικών ιδρυμάτων και οργανισμών τηλεπικοινωνιών, η μηχανική μάθηση μπορεί να προσδιορίσει εάν ένα άτομο έχει αλλάξει την διεύθυνση κατοικίας του ή ταξιδεύει διεθνώς, μέσω ελέγχου για ασυνήθιστες χρεώσεις. Η χρήση της μηχανικής μάθησης μεγάλων δεδομένων σε εμπειρίες πελατών δίνει την δυνατότητα στις τουριστικές επιχειρήσεις και τους ταξιδιωτικούς πράκτορες να προωθούν στοχευμένες προσφορές μέσω μηνυμάτων ή τηλεφωνικών κλήσεων. Πιο συγκεκριμένα, η μηχανική μάθηση θα μπορούσε να τροποποιήσει το σύστημα ανάδρασης προσδιορίζοντας παλαιότερες κινήσεις του χρήστη και υπολογίζοντας το ποσοστό επιτυχίας τους. Με την χρήση αυτών των πληροφοριών οι τουριστικοί οργανισμοί θα μπορούσαν να παρέχουν λύσεις για την αναποτελεσματικότητα κάποιων διεργασιών.

Οι προγνωστικές αναλύσεις παρουσιάζονται συχνά ως πανάκεια για τις επιχειρήσεις και μπορούν όντως να φανούν εξαιρετικά αποτελεσματικές. Η προγνωστική ανάλυση στα τουριστικά μεγάλα δεδομένα που χρησιμοποιείται στους σύγχρονους μηχανισμούς ανατροφοδότησης είναι σημαντική βελτίωση από τους απαρχαιωμένους τρόπους συλλογής πληροφοριών. Η προγνωστική ανάλυση μπορεί να δώσει στα διευθυντικά στελέχη του μάρκετινγκ περισσότερες πληροφορίες για τις προτιμήσεις των πελατών, κάτι που μπορεί να τους βοηθήσει να κατανοήσουν καλύτερα τις ανάγκες των πελατών και να οδηγήσει στην αύξηση των πωλήσεων (LaValle, Lesser, Shockley, Hopkins & Kruschwitz, 2011). Ωστόσο εδώ αξίζει να σημειωθεί ότι η επιτυχία των προγνωστικών αναλύσεων εξαρτάται τόσο από την ποιότητα των μεγάλων δεδομένων όσο και από τους μηχανισμούς ανατροφοδότησης των πελατών.

Οι μηχανισμοί ανατροφοδότησης των πελατών πρέπει να είναι καλά σχεδιασμένοι και κατανοητοί ώστε να αποδίδουν έγκαιρα χρήσιμα δεδομένα. Τα έγκαιρα και αξιόπιστα τουριστικά μεγάλα δεδομένα μπορεί να παρέχουν πλήρη εικόνα των πελατών και των δυνητικών πελατών, που ακολούθως μεταφράζονται σε υψηλής ακρίβειας επενδύσεις μάρκετινγκ σε πιο παραγωγικά κανάλια.

## Κεφάλαιο 6

### Εξόρυξη Δεδομένων Μεγάλης Κλίμακας από τα Κοινωνικά Δίκτυα για τον Τουρισμό

#### 6.1 Ο Ρόλος της Εξόρυξης Δεδομένων στα Κοινωνικά Δίκτυα

Ο κατεξοχήν τρόπος για εξαγωγή χρήσιμης πληροφορίας από πρωτογενή δεδομένα είναι η εφαρμογή τεχνικών εξόρυξης δεδομένων. Η εξόρυξη στα Κοινωνικά Δίκτυα αφορά στην εφαρμογή τεχνικών, μεθοδολογιών και μοντέλων της εξόρυξης δεδομένων σε δεδομένα που προέρχονται από αυτά.

Στα παραπάνω συμπεριλαμβάνονται προσεγγίσεις για εργασίες πρόβλεψης (π.χ. προβλέψεις συνδέσμων, συστάσεις ετικετών, προτιμήσεις χρηστών) και ανίχνευσης κοινοτήτων (ή ομάδων). Στις εργασίες πρόβλεψης εξετάζονται και αναλύονται οι σχέσεις αντικειμένων στα κοινωνικά δίκτυα, όπως αυτές διαμορφώνονται από τις αλληλεπιδράσεις των χρηστών, για να προβλεφθούν μελλοντικές καταστάσεις. Όσον αφορά στην ανίχνευση κοινοτήτων, ή εναλλακτικά ομαδοποίηση, στόχος είναι η ανάθεση δεδομένων σε ομάδες (κοινότητες), έτσι ώστε αυτά που είναι σε μία ομάδα να είναι όσο το δυνατόν όμοια μεταξύ τους, ενώ αυτά που είναι σε διαφορετικές ομάδες να είναι αρκετά ανόμοια, με βάση κάποια μετρική ομοιότητας. Μία άλλη κατηγορία μεθοδολογιών είναι αυτές που στοχεύουν στη μοντελοποίηση θέματος των δραστηριοτήτων που λαμβάνουν χώρα στα κοινωνικά δίκτυα, συμπεριλαμβανομένων μοντέλων που εξετάζουν παράλληλα θέματα και συναισθήματα σε κοινωνικά δίκτυα, λαμβάνοντας υπόψη την υποκείμενη δυναμική, και εξατομικευμένες μεθόδους ανίχνευσης θέματος ανά χρήστη. Η δυναμικότητα των θεμάτων, ως προς το χρόνο, αποτελεί ένα από τα καυτά ερευνητικά ζητήματα των κοινωνικών δικτύων. Έτσι, ιδιαίτερο ενδιαφέρον έχουν προσεγγίσεις που όχι μόνο εντοπίζουν τα θέματα και τις απόψεις των χρηστών σε κάθε δεδομένη στιγμή, αλλά παρακολουθούν και περιγράφουν την εξέλιξή τους στο χρόνο με έναν ερμηνεύσιμο τρόπο.

Οι μέθοδοι εξόρυξης δεδομένων στα κοινωνικά δίκτυα χρειάζεται να λάβουν υπόψη τις προκλήσεις 1-6 που αναφέρθηκαν στην Ενότητα 3.4. Συχνά, η ομαδοποίηση χρησιμοποιείται ως μια τεχνική αντιμετώπισης των ποιοτικών προκλήσεων 1-4 και βελτίωσης της αποτελεσματικότητας στην εξαγωγή γνώσης από τα κοινωνικά δίκτυα,

για αυτό και εφαρμόζεται, συνήθως, ως προκαταρκτικό βήμα στις προσεγγίσεις εξόρυξης δεδομένων στα κοινωνικά δίκτυα.

Πράγματι, οι εγγενείς αδυναμίες των κοινωνικών δικτύων προέρχονται από τη διαφορετικότητα των χρηστών (διαφορετικές κουλτούρες, γλώσσες, ενδιαφέροντα, κλπ) και τις παραλλαγές όρων που μπορεί να χρησιμοποιεί ο κάθε χρήστης. Αυτά τα θέματα αναμένεται ότι μπορούν να αμβλυνθούν μέσω της εξέτασης της συλλογικής δραστηριότητας των χρηστών πάνω σε ομάδες αντικειμένων (clusters, communities), και όχι εστιάζοντας σε κάθε αντικείμενο ξεχωριστά. Η κάθε ομάδα περιέχει σχετιζόμενους όρους, σύμφωνα με κάποια μετρική ομοιότητας, και οι διαφορετικές έννοιες της ίδιας (ασαφούς) λέξης αναμένεται να είναι σε διαφορετικές ομάδες, ενώ αυτές οι λέξεις που, για κάποιο λόγο, οι άνθρωποι τείνουν να χρησιμοποιούν μαζί ομαδοποιούνται. Έτσι, η κάθε ομάδα αντιστοιχεί σε ένα συγκεκριμένο θέμα και το σύνολο των όρων-ετικετών που περιέχει αντικατοπτρίζει τον τρόπο που σκέφτονται και περιγράφουν οι χρήστες το συγκεκριμένο θέμα. Περαιτέρω, μελετώντας τη δομή της κάθε ομάδας, ενδιαφέρουσες και πιθανώς μη αναμενόμενες σχέσεις μεταξύ των μελών της μπορεί να αποκαλυφθούν.

Συνοψίζοντας, η ομαδοποίηση συχνά χρησιμοποιείται ως ένα πρώτο στάδιο, το οποίο είναι πολύ σημαντικό, μιας και έχει μεγάλη συνεισφορά στην υπόλοιπη διαδικασία της εξόρυξης γνώσης και στην εξαγωγή λαοταξινομιών.

Στο πλαίσιο της παρούσας διπλωματικής εργασίας εξετάζονται στη συνέχεια σύγχρονες τάσεις στις τεχνικές εξόρυξης δεδομένων μεγάλης κλίμακας των κοινωνικών δικτύων που εφαρμόζονται στον τομέα του τουρισμού.

## 6.2 Η Σημασία των Κοινωνικών Δικτύων για τον Τουρισμό

Τα κοινωνικά δίκτυα και οι διαδικτυακές κοινότητες ενισχύουν την διάσταση της αλληλεπίδρασης από πελάτη σε πελάτη και έχουν τροποποιήσει τις υφιστάμενες ιεραρχίες και δομές των αγορών, αλλάζοντας διακριτά τους κανόνες της σύγχρονης τουριστικής αγοράς (Munar, Gyimóthy & Cai, 2013). Πρόσφατα παρατηρείται έντονη αύξηση του ενδιαφέροντος για την ανάλυση περιεχομένου προερχόμενου από χρήστες στην τουριστική βιομηχανία (Kim, Park & Yun, 2017) και παρατηρείται ότι η άνοδος του περιεχομένου προερχόμενου από χρήστες έχει μειώσει το μέγεθος της

επιρροής του μάρκετινγκ των προορισμών (Pan, MacLaurin & Crofts, 2007). Το περιεχόμενο προερχόμενο από χρήστες στο πλαίσιο του τουρισμού χαρακτηρίζεται από το ότι αποτελεί μία αντιπροσωπευτική πηγή των εμπειριών, των σκέψεων, αξιολογήσεων και συναισθημάτων των τουριστών (Carson, 2008), ενώ θεωρείται ότι επικοινωνεί τις εμπειρίες των τουριστών, συμπεριλαμβανομένων των βιωματικών, υποκειμενικών εμπειριών με πιο προσωπικό χαρακτήρα, όπως αυτές δημιουργούνται κατά την εναλλαγή των τουριστικών δραστηριοτήτων (Wang, 1999). Το περιεχόμενο προερχόμενο από χρήστες συνδέεται ως πηγή πληροφορίας με σημαντικούς παράγοντες διαμόρφωσης της φήμης ενός προορισμού, όπως τα προϊόντα, οι υπηρεσίες, η καινοτομία και η κοινωνία (Inversini, Marchiori, Dedekind & Cantoni, 2010) και ο διαμοιρασμός του μέσα από τις πλατφόρμες των κοινωνικών δικτύων ουσιαστικά συμβάλλει στην κοινοποίηση των υποκειμενικών κοινωνικών πραγματικοτήτων που βιώνονται από τους τουρίστες και στην διαμόρφωση της εικόνας των τουριστικών προορισμών (Stangl, Reino, Wang & Yuan, 2016). Οι Xiang και Gretzel (2010) τονίζουν ότι τα κοινωνικά δίκτυα ως μια ισχυρή συνιστώσα του τουρισμού στο διαδίκτυο παίζουν πολύ σημαντικό ρόλο στο πλαίσιο του σχεδιασμού ενός ταξιδιού με χρήση των καθιερωμένων μηχανών αναζήτησης, στις οποίες τα κοινωνικά δίκτυα αντιπροσωπεύονται ως πηγές αναζήτησης ταξιδιωτικής πληροφορίας σε μεγάλο βαθμό. Στο πλαίσιο των μεγάλων δεδομένων που προσφέρουν πληροφόρηση σε δυνητικούς ή υφιστάμενους πελάτες, το περιεχόμενο προερχόμενο από χρήστες παράγει προστιθέμενη αξία και θεωρείται πιο αξιόπιστο από τις πηγές πληροφόρησης των τουριστικών οργανισμών, στο πλαίσιο της λήψης αποφάσεων από τους καταναλωτές (Kim, Park & Yun, 2017). Το περιεχόμενο προερχόμενο από χρήστες στον τουρισμό (που έχει οριστεί και ως περιεχόμενο προερχόμενο από τουρίστες) αποτελεί μία επικοινωνιακή και προσωπική δίοδο έκφρασης των τουριστών, που ενθαρρύνει τόσο τον διαμοιρασμό γνώσης αναφορικά με τουριστικά προϊόντα όσο και στην κοινωνικοποίηση της τουριστικής εμπειρίας. Επίσης, έχει καταδειχθεί ότι οι τουρίστες δεν ενσωματώνουν στοιχεία των επίσημων πηγών πληροφόρησης και προώθησης (όπως π.χ., τα σλόγκαν) και ακολουθούν διαφορετικές λογικές έκφρασης και περιεχομένου στις συνεισφορές τους (Munar, 2011).

Από την πλευρά της τουριστικής προσφοράς, οι συνθήκες αυτές δημιουργούν μια σειρά από προκλήσεις. Η ανάλυση του περιεχομένου των ταξιδιωτών συμβάλλει στην



κατανόηση της συμπεριφοράς των τουριστών και της εικόνας του τουριστικού προορισμού όπως αυτή διαμορφώνεται από τις καταγραφές τους, ενώ δίνεται η δυνατότητα ανάπτυξης προσαρμοσμένων στρατηγικών προώθησης των προορισμών που θα αξιοποιούν τις ροές πληροφόρησης από την πλευρά των ταξιδιωτών (Li, Lin, Tsai & Wang, 2015). Επίσης, έχει αναγνωριστεί η ανάγκη ανάπτυξης νέων μοντέλων ανάλυσης του αδόμητου περιεχομένου των χρηστών που θα είναι πιο αποτελεσματικά. Οι επιχειρήσεις καλούνται να μεταβούν από τις παραδοσιακές μορφές ανάλυσης των δομημένων δεδομένων που προέρχονται από πελάτες προς τεχνικές και μεθοδολογίες που θα αντιμετωπίζουν την ανάγκη ανάλυσης περιεχομένου σε αδόμητη μορφή, που χαρακτηρίζει το περιεχόμενο προερχόμενο από τουρίστες, με αυτό να συνεπάγεται και την ανάπτυξη των ανάλογων τεχνολογικών υποδομών (Zhong, Li & Wu, 2012). Σύμφωνα με την María Munar (2011), το προερχόμενο από τουρίστες διαδικτυακό περιεχόμενο λαμβάνει χώρα σε διαφορετικά στάδια της τουριστικής εμπειρίας και έχει διαφορετικών τύπων επιπτώσεις στην ταυτότητα του τουριστικού προορισμού, ενώ η διαχείρισή του φέρει τους Οργανισμούς Διαχείρισης Προορισμών αντιμέτωπους με την πρόκληση τριών διαφορετικών στρατηγικών:

- ⊕ **Μιμητική (mimetic)**, στην οποία οι Οργανισμοί Διαχείρισης Προορισμών αντιγράφουν τις δομές και τις πρακτικές των κοινωνικών δικτύων και γενικότερα των ιστοσελίδων με τεχνολογικά στοιχεία που επιτρέπουν την παραγωγή και διάδοση περιεχομένου από χρήστες, επεκτείνοντας ουσιαστικά τους επίσημους ιστοτόπους τους με ενσωμάτωση εργαλείων που ενθαρρύνουν και επιτρέπουν την ανάρτηση και πρόσβαση περιεχομένου από χρήστες. Η συγκεκριμένη στρατηγική θεωρείται εύκολη στην εφαρμογή της και μικρού κόστους, ενώ επιτρέπει την διατήρηση του ελέγχου του περιεχομένου από την πλευρά των Οργανισμών Διαχείρισης Προορισμών, ιδιαίτερα του μη επιθυμητού (ή μη επιτρεπτού). Εμφανίζει όμως το μειονέκτημα ότι η λογική μιας ιστοσελίδας που αντιπροσωπεύει έναν οργανισμό είναι πολύ διαφορετική, με διαφορετικά στοιχεία κουλτούρας και ελευθερίας του περιεχομένου.
- ⊕ **Διαφημιστική (advertising)**, που αντιπροσωπεύει τις παραδοσιακές μεθόδους αντιμετώπισης του φαινομένου του περιεχομένου προερχόμενου από τουρίστες, που θεωρούν τους τύπους αυτών των μέσων ως εναλλακτικές διαφημιστικές

πλατφόρμες, συνεπώς ωθούνται προς την κατεύθυνση της ενίσχυσης της παρουσίας των Οργανισμών Διαχείρισης Προορισμών και των προορισμών στα κοινωνικά δίκτυα και τις συναφείς ιστοσελίδες. Παρ' όλα αυτά, στην στρατηγική αυτή προσέγγιση οι οργανισμοί δεν εκμεταλλεύονται τους όγκους πληροφορίας που παρέχονται από το περιεχόμενο των χρηστών σε όλες τις διαθέσιμες πηγές.

✦ **Αναλυτική (analytic)**, που συνίσταται σε δύο διαστάσεις, αυτές της πρόληψης και της γνώσης. Οι Οργανισμοί Διαχείρισης Προορισμών κάνουν προληπτικές ενέργειες προκειμένου να παρακολουθούν την ανάπτυξη του περιεχομένου προερχόμενου από τουρίστες αναφορικά με την ταυτότητα του προορισμού και δραστηριοποιούνται με στόχο την αποφυγή επιβλαβών εξελίξεων ή την βελτίωση των στρατηγικών προώθησης. Αναφορικά με την διάσταση γνώσης, ακολουθείται σπανιότερα καθώς προϋποθέτει σημαντικές επενδύσεις στην ανάπτυξη και εκμάθηση εξειδικευμένων τεχνολογιών πληροφορίας και επικοινωνιών, με στόχο την μετατροπή της χαοτικής και ογκώδους μορφής του περιεχομένου προερχόμενου από τουρίστες σε στρατηγική γνώση.

Η έρευνα στο μάρκετινγκ προορισμών παρουσιάζει επίσης την ανάλυση και διαχείριση του περιεχομένου από τουρίστες ως μια σημαντική πτυχή του «έξυπνου» τουρισμού (smart tourism) (Kim, Park & Yun, 2017). Ο «έξυπνος τουρισμός» προϋποθέτει τη συμμετοχή των χρηστών ως ενεργών παραγωγών περιεχομένου, την επέκταση των βιωματικών εμπειριών του τουρισμού μέσω της τεχνολογίας και την παροχή περισσότερων ευκαιριών αλληλεπίδρασης με τον τοπικό πληθυσμό. Η έκθεση των χρηστών σε διαδικτυακό περιεχόμενο και δη σε περιεχόμενο χρηστών στο πλαίσιο του «έξυπνου τουρισμού» χαρακτηρίζεται από στενότητα χρόνου και κινητικότητα των χρηστών, συνθήκες στις οποίες οι μορφές παρουσίασης του περιεχομένου σε γραφικό και λεκτικό επίπεδο παίζουν μεγάλο ρόλο, ιδιαίτερα σε στάδια διαμόρφωσης των νοητικών εικόνων του προορισμού από τους τουρίστες (Marchiori & Cantoni, 2015).

Τέλος, στο πλαίσιο της αντιμετώπισης του φαινομένου διαχείρισης του περιεχομένου προερχόμενου από τουρίστες, δημιουργούνται παράλληλα και ανάγκες ανάλυσης των χαρακτηριστικών των χρηστών σε συνδυασμό με το περιεχόμενο των τουριστών, είτε ως δέκτες είτε ως παραλήπτες. Οι Lange-Faria και Elliot (2012) επισημαίνουν ότι ενώ οι τάσεις χρήσης περιεχομένου των κοινωνικών δικτύων από τους ταξιδιώτες για τον

σχεδιασμό των ταξιδιών τους αυξάνεται, ο τρόπος χρήσης του περιεχομένου επηρεάζεται από τα στοιχεία ηλικίας, φύλου και λοιπών δημογραφικών χαρακτηριστικών των χρηστών. Επίσης, έχει καταδειχθεί (Simms, 2012) ότι οι συχνοί ταξιδιώτες είναι πιο πιθανό να στραφούν προς το διαδικτυακό περιεχόμενο, από χρήστες που δεν ταξιδεύουν συχνά, ενώ είναι επίσης πιο πιθανό να δημιουργήσουν περιεχόμενο που να επηρεάσει τον σχεδιασμό ταξιδιού για τους άλλους ταξιδιώτες. Ακόμη, ότι οι δυνητικοί ταξιδιώτες τείνουν να χρησιμοποιούν σε υψηλότερο βαθμό το περιεχόμενο που δημιουργείται από χρήστες όταν αφορά έναν προορισμό που θα επισκεφτούν για πρώτη φορά. Τέλος, αναφορικά με τις ανάγκες των χρηστών στο πλαίσιο άντλησης πληροφόρησης από διαδικτυακές κριτικές, επισημαίνεται ότι εκτός από πληροφόρηση για τα προϊόντα και υπηρεσίες οι χρήστες απαιτούν σχετικές πληροφορίες αναφορικά με αυτούς που έχουν αναρτήσει την κριτική, αλλά και την ύπαρξη ικανού αριθμού κριτικών (Williams, Wiele, Van Der, Iwaarden, Van & Eldridge, 2010).

## Κεφάλαιο 7

### Τεχνικές Εξόρυξης Δεδομένων από τα Κοινωνικά Δίκτυα για τον Τουρισμό

#### 7.1 Εισαγωγή

Αν και σε πρώιμο στάδιο, ένα πλούσιο "ορυχείο" μεγάλων δεδομένων που σχετίζονται με τον τουρισμό έχει δημιουργηθεί από τρεις κύριες πηγές: χρήστες, συσκευές και λειτουργίες. Ως προς αυτό είναι σημαντικό να τονίσουμε ότι πρώτον, το διαδίκτυο έχει προωθήσει μια ραγδαία άνοδο στα κοινωνικά δίκτυα, προσφέροντας μια μεγάλη πλατφόρμα για τη διάδοση περιεχομένου που δημιουργείται από τους ίδιους τους χρήστες (User Generated Content, UGC) (από άποψη κειμένων, φωτογραφιών κ.λπ.) (Xiang, Du, Ma & Fan, 2017). Δεύτερον, λόγω της έντονης ανάπτυξης του διαδικτύου των πραγμάτων (IoT), έχουν αναπτυχθεί διάφορες συσκευές που λειτουργούν με αισθητήρες και απασχολούνται για την παρακολούθηση τουριστικών κινήσεων και περιβαλλοντικών συνθηκών, παρέχοντας σημαντικά χωρικά-χρονικά μεγάλα δεδομένα (όπως παγκόσμια δεδομένα συστήματος θέσης (GPS), δεδομένα περιαγωγής κινητής τηλεφωνίας, δεδομένα Bluetooth, κ.λπ.) (Shoval & Ahas, 2016). Τρίτον, ο τουρισμός είναι ένα πολύπλοκο σύστημα που καλύπτει μια σειρά λειτουργιών όπως αναζήτηση ιστού, επίσκεψη ιστοσελίδων, διαδικτυακή κράτηση και αγορά κ.λπ., παράγοντας έτσι τα αντίστοιχα δεδομένα συναλλαγών, δεδομένα αναζήτησης ιστού, δεδομένα επίσκεψης ιστοσελίδας, διαδικτυακά δεδομένα κράτησης κ.λπ., για την κατανόηση της τουριστικής συμπεριφοράς και τη βελτίωση του τουριστικού μάρκετινγκ. Με βάση τα μεγάλα δεδομένα από αυτές τις τρεις κύριες πηγές, η τουριστική συμπεριφορά και η τουριστική αγορά μπορούν να διερευνηθούν περαιτέρω και να γίνουν κατανοητές τόσο από τον ακαδημαϊκό όσο και από τον τουριστικό κλάδο.

Χρησιμοποιώντας τα προαναφερθέντα πολύτιμα μεγάλα δεδομένα, η τουριστική έρευνα έχει προχωρήσει στην εποχή των μεγάλων δεδομένων και το γεγονός αυτό έχει επιφέρει εκπληκτικές βελτιώσεις. Για παράδειγμα, σύμφωνα με τους Yang, Pan, Evans και Lv (2015) η μεγάλη κλίμακα δεδομένων θα μπορούσε να αντισταθμίσει το ζήτημα του περιορισμένου μεγέθους δείγματος που αντιμετωπίζουν οι χρήστες κατά την έρευνα δεδομένων, παρέχοντας έναν νέο τρόπο κατανόησης της τουριστικής συμπεριφοράς. Ομοίως, οι Li, Pan, Law και Huang (2017) υποστήριξαν ότι τα μεγάλα

δεδομένα θα μπορούσαν να παρέχουν επαρκή δεδομένα χωρίς δειγματοληψία, βοηθώντας και τους οργανισμούς και τις τουριστικές βιομηχανίες να κατανοήσουν καλύτερα την τουριστική συμπεριφορά. Κατά τους Xiang, Schwartz, Gerdes και Uysal (2015) τα μεγάλα αναλυτικά δεδομένα θα μπορούσαν να βοηθήσουν στην ανάπτυξη νέων γνώσεων οι οποίες θα αναμορφώσουν την κατανόηση της τουριστικής βιομηχανίας. Συμπερασματικά, με βάση τα όσα ειπώθηκαν παραπάνω, τα μεγάλα δεδομένα επιτρέπουν την καλύτερη κατανόηση της τουριστικής ζήτησης, της τουριστικής συμπεριφοράς και της ικανοποίησης του τουρίστα.

Στη παρούσα εργασία θα καταπιαστούμε με την πρώτη πηγή δημιουργίας μεγάλων δεδομένων, τους χρήστες. Όπως έχουμε ήδη αναφέρει, η ταχύτατη ανάπτυξη των εφαρμογών web 2.0 και των κοινωνικών δικτύων έχει προσφέρει μια ευρύχωρη πλατφόρμα για τους χρήστες να μοιράζονται τις τουριστικές τους εμπειρίες, σε ότι αφορά τα διαδικτυακά δεδομένα κειμένου (συμπεριλαμβανομένων των κριτικών και ιστολογίων) και φωτογραφιών. Τα εν λόγω δεδομένα που δημοσιεύουν οι χρήστες είναι ευρέως γνωστά ως δεδομένα UGC (περιεχόμενο προερχόμενο από χρήστες) σε σχετικές μελέτες (Hu, Chen & Chou, 2017; Lu, Wu & Sang, 2017; Shi, Serdyukov, Hanjalic & Larson, 2011; Xiang, Du, Ma & Fan, 2017).

## 7.2 Περιεχόμενο Προερχόμενο από τον Χρήστη (UGC)

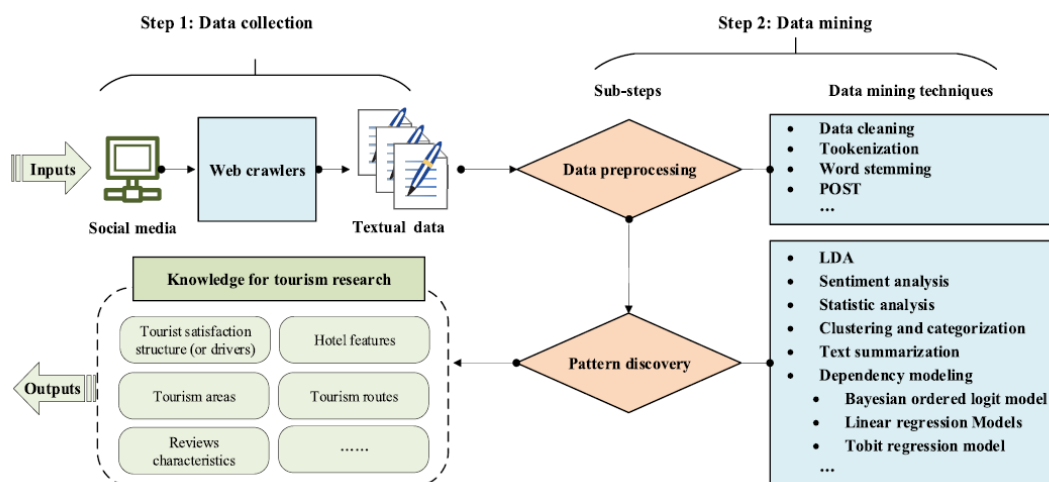
Στην ψηφιακή εποχή, η ραγδαία άνοδος του Παγκόσμιου Ιστού και των κοινωνικών δικτύων έχει επηρεάσει σημαντικά τον τουριστικό κλάδο, παρέχοντας μία ευρύχωρη πλατφόρμα για κοινή χρήση δεδομένων από τους χρήστες. Ως μια μεγάλη κατηγορία μεγάλων δεδομένων, τα δεδομένα που παράγονται από τους χρήστες έχουν χρησιμοποιηθεί για την προώθηση της τουριστικής έρευνας και διακρίνονται σε δύο τύπους:

- ⊕ Διαδικτυακά δεδομένα κειμένου, όπως οι κριτικές προϊόντων και τα ιστολόγια
- ⊕ Διαδικτυακά δεδομένα φωτογραφιών, τα οποία δημοσιεύονται στα κοινωνικά δίκτυα, στις ιστοσελίδες κ.λπ.

### 7.2.1 Διαδικτυακά Δεδομένα Κειμένου και Τεχνικές Εξόρυξης

Τα κοινωνικά μέσα τα τελευταία χρόνια ευδοκούν με την έντονη ανάπτυξη του διαδικτύου, το οποίο προσφέρει άφθονο χώρο για τους τουρίστες να διαδώσουν μια ποικιλία τουριστικών πληροφοριών αναφορικά με ταξιδιωτικές κριτικές και εμπειρίες. Για παράδειγμα, οι ταξιδιώτες μπορούν πλέον εύκολα να εκφράσουν την ικανοποίηση ή/και τη δυσαρέσκειά τους για τα τουριστικά προϊόντα, δημιουργώντας ένα πλούσιο «ορυχείο» με online τουριστικές κριτικές. Μπορούν επίσης να μοιραστούν τις προοπτικές και τις εμπειρίες του ταξιδιού τους σε ευρείας χρήσης ιστολόγια, όπως το Twitter, παρέχοντας πολύτιμες πληροφορίες για πιθανούς μελλοντικούς τουρίστες. Αυτές οι διαδικτυακές κριτικές και εμπειρίες σε μορφή κειμένου αποτελούν τον πρώτο τύπο μεγάλων δεδομένων στην τουριστική έρευνα και κρίνονται πολύ σημαντικές καθώς μεταφέρουν τα συναισθήματα και τις διαθέσεις των τουριστών.

Αναφορικά με την εξόρυξη διαδικτυακών δεδομένων κειμένου, τα συλλεγόμενα διαδικτυακά κείμενα αναλύονται για την εξαγωγή χρήσιμων γνώσεων για την τουριστική έρευνα, μέσω δύο επιμέρους σταδίων: προεπεξεργασία δεδομένων και ανακάλυψη προτύπων.



**Εικόνα 9.** Διαδικασία εξόρυξης διαδικτυακών δεδομένων κειμένου για τον τουρισμό (Li, Xu, Tang & Wang, 2018)

### 7.2.1.1 Προεπεξεργασία Δεδομένων

Στο πρώτο στάδιο της προεπεξεργασίας των δεδομένων λαμβάνει χώρα η προετοιμασία των δεδομένων, η οποία εκτελείται πριν την καθαυτό εξόρυξη διαδικτυακών δεδομένων κειμένου. Η προεπεξεργασία των δεδομένων είναι απαραίτητη, καθώς τα αρχικά δεδομένα «πάσχουν» από διαφόρων ειδών προβλήματα. Σε αυτά συγκαταλέγονται η ύπαρξη αλληλοσυγκρουόμενων πληροφοριών, η ύπαρξη ασυνεπειών ως προς την κωδικοποίηση, την ονοματοδοσία πεδίων καθώς και η ύπαρξη χαμένων τιμών και θορύβου, τυχαία δηλαδή κυμαινόμενων δεδομένων χωρίς ουσιαστικό περιεχόμενο. Τα προβληματικά αυτά δεδομένα καλούνται «ακάθαρτα» και η διαδικασία αντιμετώπισης των προβλημάτων τους καλείται «καθαρισμός δεδομένων». Ναι μεν, η προεπεξεργασία των δεδομένων περιλαμβάνει τον καθαρισμό τους, αλλά δεν περιορίζεται μόνο σε αυτόν. Ειδικές απαιτήσεις των μεθόδων επεξεργασίας συχνά επιβάλλουν τον μετασχηματισμό των δεδομένων. Στο εν λόγω στάδιο χρησιμοποιούνται διάφορες τεχνικές εξόρυξης διαδικτυακών δεδομένων κειμένου με πιο δημοφιλείς τις παρακάτω:

#### 1) Data cleaning/cleansing

Η εν λόγω τεχνική χρησιμοποιείται για την ανίχνευση και αφαίρεση ανακριβών ή άχρηστων εγγραφών από διαδικτυακά δεδομένα κειμένου, όπως για παράδειγμα διπλοκαταχωρημένες εγγραφές, εγγραφές με αντικρουόμενο περιεχόμενο, χρήση συνωνύμων τιμών σε κάποιο πεδίο (π.χ. η «Θεσσαλονίκη» μπορεί να καταχωρείται ολογράφως ή ως «Θεσ/νικη»), τα ορθογραφικά λάθη (Xiang, Schwartz, Gerdes & Uysal, 2015), γλώσσα χωρίς στόχους και χαμηλή συχνότητα εμφάνισης λέξεων (Guo, Barnes & Jia, 2017), προκειμένου να παραμείνουν οι πολύτιμες και σχετικές πληροφορίες με τον τουρισμό.



*Εικόνα 10. Στάδια καθαρισμού δεδομένων (<https://mtd.media/data-cleansing/>, 2020)*

## 2) Tokenization

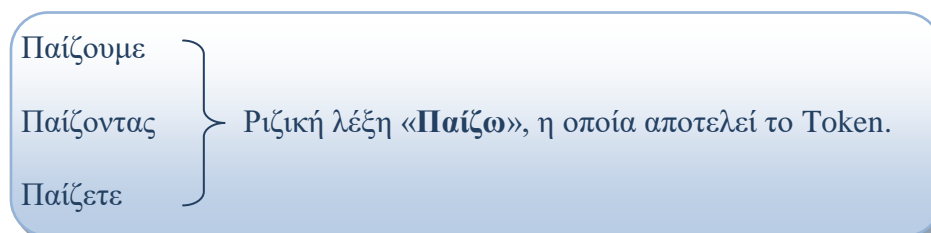
Η συγκεκριμένη τεχνική έχει ως σκοπό τη διάσπαση μιας ακολουθίας συμβολοσειρών, στη προκειμένη περίπτωση ενός διαδικτυακού κειμένου σχετικό με τα ταξίδια, σε κομμάτια όπως λέξεις, λέξεις-κλειδιά, φράσεις, σύμβολα και άλλα σημαντικά στοιχεία που ονομάζονται Tokens. Κατά τη διάρκεια της συγκεκριμένης διαδικασίας απορρίπτονται ορισμένοι χαρακτήρες όπως τα σημεία στίξης. Ως εκ τούτου, λέξεις-κλειδιά σχετικές με τουριστικά σημεία, συναισθήματα ως προς τα ταξίδια κ.λπ. μπορούν να φιλτραριστούν από μαζικές προτάσεις (Guo, Barnes & Jia, 2017; Xiang, Du, Ma & Fan, 2017; Xu & Li, 2016).



*Εικόνα 11. Παράδειγμα Tokenization*

## 3) Word Stemming

Η τεχνική Word stemming χρησιμοποιείται για τον εντοπισμό των ριζών μιας λέξης. Περιλαμβάνει σε ένα Token όλες εκείνες τις λέξεις που προέρχονται από την ίδια ρίζα. Επομένως το μοντέλο μας είναι πιο απλοποιημένο και εύκολα διαχειρίσιμο (Xu & Li, 2016).



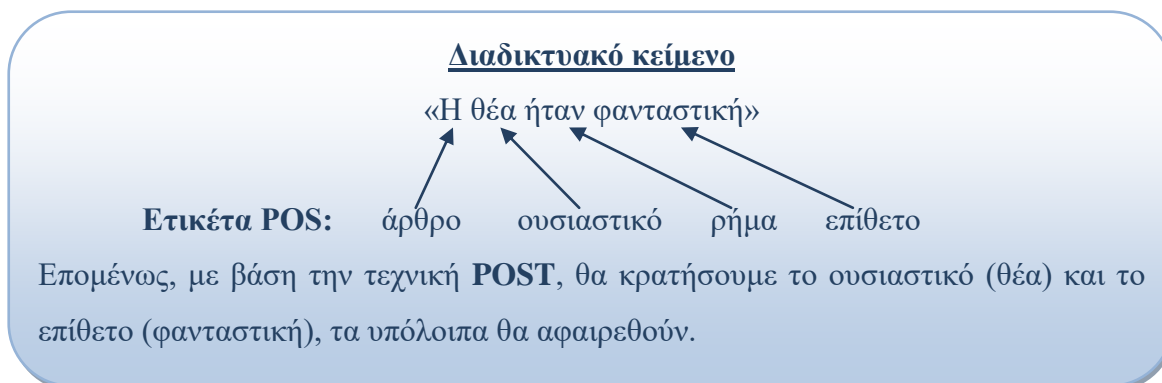
*Εικόνα 12. Παράδειγμα Word Stemming*

## 4) POST

Η εν λόγω τεχνική χρησιμοποιείται για την επισήμανση κάθε χρήσιμης λέξης που απαντάται σε μια πρόταση με ετικέτα POS, όπως για παράδειγμα ουσιαστικό, επίθετο



ή επίρρημα. Επομένως, δεδομένου ότι οι κριτικές σε ένα ξενοδοχείο εκφράζονται κυρίως με ουσιαστικά, επίθετα και αρνητικά επιρρήματα, μπορούν εύκολα να αφαιρεθούν ασήμαντες λέξεις που έχουν σημειωθεί με άλλες ετικέτες (Guo, Barnes & Jia, 2017; Hu, Chen & Chou, 2017).



*Εικόνα 13. Παράδειγμα τεχνικής POST*

#### 7.2.1.2 Ανακάλυψη Προτύπων

Η ανακάλυψη προτύπων, είναι ένα άλλο κρίσιμο στάδιο της εξόρυξης διαδικτυακών δεδομένων κειμένου, το οποίο στοχεύει στην αναζήτηση σημαντικών πληροφοριών στα διαδικτυακά κείμενα. Συγκεκριμένα, στόχος του εν λόγω σταδίου είναι να ανακαλύψει ενδιαφέροντα μοτίβα, τα οποία περιλαμβάνουν περιοδικά και μη φυσιολογικά μοτίβα, από χρονικά δεδομένα. Η ανακάλυψη προτύπων, που έχουν ενδιαφέρον, έχει γίνει ένα από τα πιο σημαντικά καθήκοντα της εξόρυξης δεδομένων και μπορεί να εφαρμοστεί σε πολλούς τομείς όπως και στον τουρισμό. Ιδιαίτερα όταν τα πρότυπα που προέρχονται από τον ειδικό τομέα δεν υπάρχουν ή δεν είναι ολοκληρωμένα, είναι απαραίτητος ένας αλγόριθμος για την αυτόματη εύρεση συγκεκριμένων μοτίβων ή σχημάτων από τα χρονικά δεδομένα. Ένας τέτοιος αλγόριθμος είναι αξιοσημείωτος στο ότι δεν προϋποθέτει προηγούμενη γνώση του αριθμού των ενδιαφέρων δομών, ούτε απαιτεί εξαντλητική εξήγηση των περιγραφόμενων προτύπων. Παρακάτω παρουσιάζουμε τις πιο σημαντικές τεχνικές εξόρυξης του εν λόγω σταδίου:

##### 1) Latent Dirichlet Allocation (LDA)

Είναι ένα μοντέλο θέματος το οποίο χρησιμοποιείται για τον προσδιορισμό των αφηρημένων «θεμάτων» σε ένα συγκεκριμένο διαδικτυακό κείμενο. Συγκεκριμένα,

φαντάζεται ένα σταθερό σύνολο θεμάτων. Κάθε θέμα αντιπροσωπεύεται από ένα σύνολο λέξεων. Στόχος της συγκεκριμένης τεχνικής είναι να χαρτογραφήσει/ταξινομήσει όλα τα έγγραφα στα θέματα κατά τέτοιο τρόπο, έτσι ώστε οι λέξεις σε κάθε έγγραφο να συλλαμβάνονται κυρίως από αυτά τα φανταστικά θέματα. Οι Guo, Barnes και Jia (2017) χρησιμοποίησαν τη τεχνική LDA για να ανακαλύψουν γρήγορα ένα μείγμα θεμάτων π.χ. τους παράγοντες που επηρεάζουν την ικανοποίηση των πελατών ενός συγκεκριμένου ξενοδοχείου, από έναν τεράστιο αριθμό σχολίων.

Για παράδειγμα, έχουμε τα παρακάτω έγγραφα (διαδικτυακά κείμενα):

Έγγραφο 1: Τρώω σάντουιτς με φυστικοβούτυρο για πρωινό.

Έγγραφο 2: Μου αρέσει να τρώω αμύγδαλα, φυστίκια και καρύδια.

Έγγραφο 3: Ο γείτονάς μου πήρε ένα μικρό σκυλί χθες.

Έγγραφο 4: Οι γάτες και οι σκύλοι είναι οι καλύτεροι φίλοι.

Έγγραφο 5: Δεν πρέπει να ταΐζετε φυστίκια το σκυλί σας.

Το μοντέλο LDA ανακαλύπτει τα διαφορετικά θέματα που αντιπροσωπεύουν τα παραπάνω έγγραφα και σε τι ποσοστό το κάθε θέμα εμφανίζεται σε ένα έγγραφο. Ως εκ τούτου, το μοντέλο LDA μπορεί να παράξει τα ακόλουθα αποτελέσματα:

Θέμα 1: 30% φυστίκια, 15% αμύγδαλα, 10% πρωινό κ.ο.κ. (εύλογα αυτό το θέμα αφορά τα τρόφιμα).

Θέμα 2: 20% σκύλοι, 10% γάτες, 5% φυστίκια κ.ο.κ (εύλογα αυτό το θέμα ασχολείται με κατοικίδια ζώα ή ζώα γενικά).

Επομένως, τα Έγγραφα 1 και 2 ανήκουν 100% στο Θέμα 1

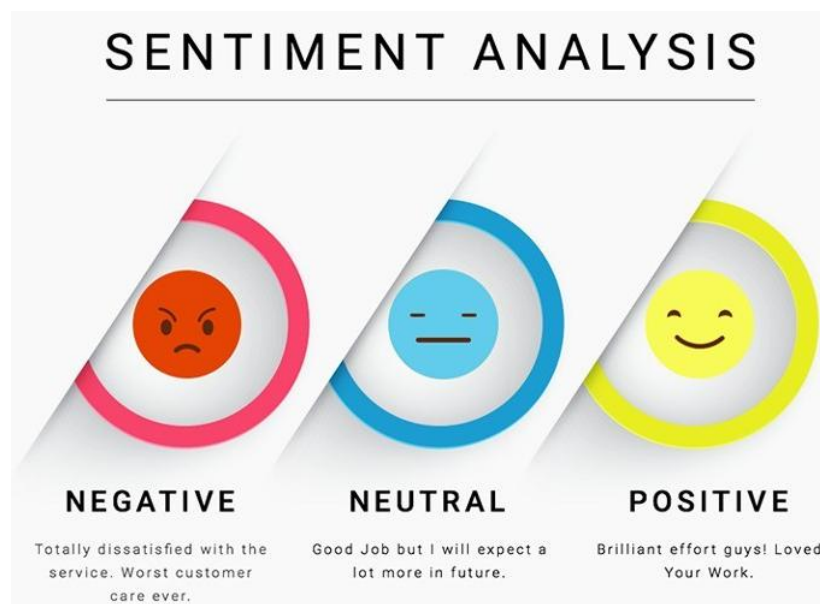
τα Έγγραφα 3 και 4 ανήκουν 100% στο Θέμα 2

το Έγγραφο 5 ανήκει 70% στο Θέμα 1 και 30% στο Θέμα 2

## 2) Sentiment analysis

Ανάλυση συναισθήματος καλείται η αυτοματοποιημένη διαδικασία εξαγωγής πληροφοριών για την συναισθηματική πολικότητα ενός σώματος κειμένου και συχνά αναφέρεται εναλλακτικά ως εξόρυξη γνώμης. Αποτελεί ένα πεδίο έρευνας, που προσελκύει έντονο ενδιαφέρον τα τελευταία χρόνια εξαιτίας της μεγάλης επιρροής των κοινωνικών δικτύων στην καθημερινότητά μας, του αυτοματοποιημένου τρόπου

που παρέχει για την ανάλυση της γραπτής πληροφορίας που αφθονεί στις διαδικτυακές πηγές αλλά και της σημαντικής προόδου που σημειώνεται τελευταία στα πεδία της μηχανικής μάθησης, της τεχνητής νοημοσύνης και της βαθιάς μάθησης. Σε ότι αφορά τον τουρισμό, μέσω της ανάλυσης συναισθήματος είμαστε σε θέση να προσδιορίσουμε τη στάση των τουριστών απέναντι στα τουριστικά προϊόντα ή σε γραφικά σημεία ενδιαφέροντος, ταξινομώντας τα διαδικτυακά δεδομένα κειμένου σε κατηγορίες συναισθημάτων: θετικά, αρνητικά ή ουδέτερα. Για παράδειγμα, πρόσφατες μελέτες τείνουν να χρησιμοποιούν την ανάλυση συναισθημάτων ως ένα χρήσιμο εργαλείο για τη διερεύνηση των απόψεων των ταξιδιωτών συγκεκριμένης ξενοδοχειακής υπηρεσίας (Hu, Chen & Chou, 2017; Philander & Zhong, 2016) και τοποθεσίες τουριστικού ενδιαφέροντος (Xu, Yuan, Ma & Qian, 2015).



*Εικόνα 14. Ανάλυση συναισθήματος*  
(<https://www.kdnuggets.com/images/sentiment-fig-1-689.jpg>, 2020)

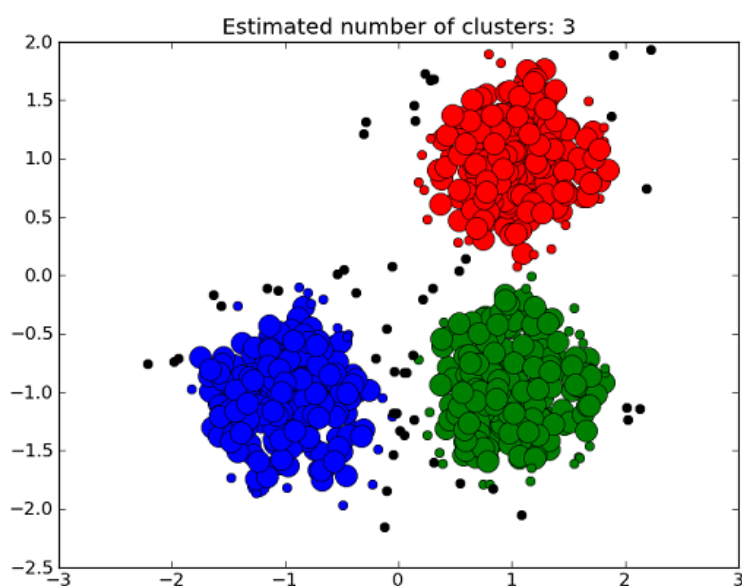
### 3) Statistical analysis

Η στατιστική ανάλυση θεωρείται η πιο βασική τεχνική ανάλυσης διαφόρων τύπων δεδομένων, συμπεριλαμβανομένων διαδικτυακών δεδομένων κειμένου προερχόμενα από τα κοινωνικά δίκτυα. Στην τουριστική έρευνα, περιγραφικά στατιστικά στοιχεία (π.χ. μέσος όρος, διακύμανση κ.λπ.) (Racherla & Friske, 2012), t-test (Schuckert, Liu & Law, 2015a), correlation matrix (Racherla & Friske, 2012), Kruskal-Wallis test, Mann-Whitney U test (Lu & Stepchenkova, 2012) και ανάλυση ανταπόκρισης (Költringer & Dickinger, 2015), χρησιμοποιούνται ευρέως για την περιγραφή

διαφορετικών πληροφοριών που περιέχονται στα διαδικτυακά δεδομένα κειμένου, όπως ο αριθμός των γνωστοποιήσεων ταυτότητας, η φήμη κριτικών και οι βαθμολογίες ξενοδοχείων.

#### 4) Clustering και Categorization

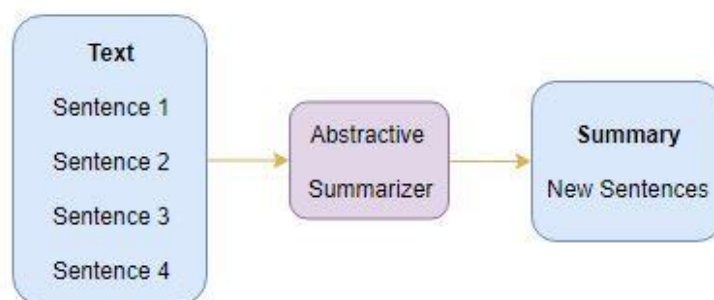
Η τεχνική της συσταδοποίησης ή αλλιώς κατηγοριοποίησης εντοπίζει ομάδες αντικειμένων στα διαδικτυακά δεδομένα βάσει των κοινών χαρακτηριστικών τους και της απόστασης μεταξύ τους. Οι ομάδες των αντικειμένων διαμορφώνονται με βάση συγκεκριμένο μέτρο απόστασης, έτσι ώστε τα αντικείμενα της ίδιας ομάδας να είναι όσο το δυνατόν πλησιέστερα μεταξύ τους, ενώ τα αντικείμενα διαφορετικών ομάδων να είναι όσο το δυνατόν πιο απομακρυσμένα. Καθώς δεν υπάρχει καθολικός ορισμός για την αποδεκτή ομάδα, η επιλογή εξαρτάται κάθε φορά από το συγκεκριμένο πρόβλημα και τα δεδομένα που καλούμαστε να εξορύξουμε. Για παράδειγμα, οι Bordogna, Frigerio, Cuzzocrea και Psaila (2016) χρησιμοποίησαν την τεχνική της συσταδοποίησης για να κατηγοριοποιήσουν τα ταξίδια, των οποίων οι σειρές των αναγνωριστικών geoslot ήταν παρόμοιες σε μια κλάση, με βάση τις γεωγραφικές ετικέτες των μηνυμάτων στο Twitter. Οι Marine-Roig και Clavé (2015) χρησιμοποίησαν την τεχνική της κατηγοριοποίησης για να ομαδοποιήσουν σε σημαντικές κατηγορίες τις κρυφές πληροφορίες ταξιδιωτικών ιστολογίων και κριτικών για τη Βαρκελώνη σύμφωνα με τις λέξεις-κλειδιά.



**Εικόνα 15.** Παράδειγμα συσταδοποίησης με τρεις συστάδες (<https://www.datanami.com/2017/06/14/8-concrete-data-mining-techniques-will-deliver-best-results/>, 2020)

## 5) Text summarization

Η τεχνική της περίληψης κειμένου αποσκοπεί στην παρουσίαση των κύριων σημείων του, σε μια δηλαδή περιεκτική μορφή. Μια πραγματική περίληψη, θα πρέπει να εκφράζει την ουσία του κειμένου, αποκαλύπτοντας το βαθύτερο νόημα του περιεχομένου του. Σκοπός της είναι η ανακάλυψη ενδιαφέρουσας και απροσδόκητης πληροφορίας. Στην εξόρυξη διαδικτυακών δεδομένων κειμένου χρησιμοποιείται για την αυτόματη παραγωγή μιας σύντομης περίληψης ενός ή περισσότερων κειμένων, για πιο αποτελεσματική ανακάλυψη βασικών πληροφοριών από πρωτότυπα κείμενα. Ως προς την συγκεκριμένη τεχνική, βασική συνεισφορά στον τουριστικό κλάδο αποτελεί το ερευνητικό έργο των Hu, Chen & Chou (2017), στο οποίο προτάθηκε η τεχνική περίληψης ενός πολλαπλού κειμένου (multitext) για να εντοπιστούν οι πιο ενημερωτικές κριτικές ξενοδοχείων.



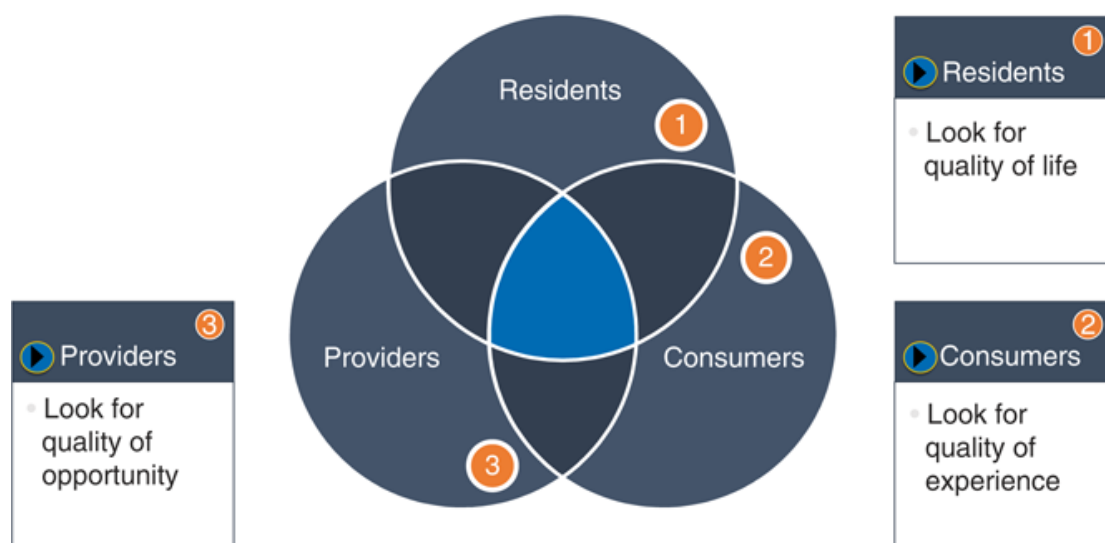
*Εικόνα 16. Παράδειγμα περίληψης κειμένου*

(<https://www.analyticsvidhya.com/blog/2019/06/comprehensive-guide-text-summarization-using-deep-learning-python/>, 2020)

## 6) Dependency modeling

Μοντελοποίηση εξάρτησης είναι η τεχνική της εύρεσης σχέσεων μεταξύ των μεταβλητών ενός συνόλου δεδομένων. Στον τουριστικό κλάδο, η συγκεκριμένη τεχνική δύναται να χρησιμοποιηθεί για την εύρεση της σχέσης μεταξύ των διαδικτυακών δεδομένων ενός κειμένου (π.χ. διαδικτυακές κριτικές ξενοδοχείων) και τουριστικών παραγόντων, όπως η επίδοση (βαθμολογία) του ξενοδοχείου (Xie & Zhang, 2014; Ye, Law & Gu, 2009), η επίδοση (βαθμολογία) του εστιατορίου (Zhang, Ye, Law & Li, 2010) και οι ταξιδιωτικές συμπεριφορές (Zhang & Cole, 2016). Μια ποικιλία μοντέλων παλινδρόμησης μπορούν να εφαρμοστούν, όπως το μοντέλο Bayesian Ordered Logit (Zhang & Cole, 2016), το μοντέλο γραμμικής

παλινδρόμησης (Xie & Zhang, 2014; Ye, Law & Gu, 2009; Zhang, Ye, Law & Li, 2010) και το μοντέλο παλινδρόμησης Tobit (Fang, Ye, Kucukusta & Law, 2016).



*Εικόνα 17. Μοντελοποίηση εξάρτησης μεταξύ τουριστών, κατοίκων και παρόχων τουριστικών προϊόντων (Postma, 2003)*

Είναι αξιοσημείωτο ότι πολλά χρήσιμα εργαλεία εξόρυξης δεδομένων και πακέτα λογισμικού έχουν αναπτυχθεί για την επεξεργασία κειμένου. Τέτοια εργαλεία πολλαπλών λειτουργιών είναι το Waikato Environment for Knowledge Analysis (WEKA), το LingPipe και το TextBlob. Συγκεκριμένα, το WEKA, που αναπτύχθηκε από το Πανεπιστήμιο του Waikato, μπορεί να χειριστεί τα περισσότερα προβλήματα επεξεργασίας κειμένου, όπως καθαρισμός δεδομένων, Tokenization, ανάλυση συχνότητας λέξεων, ομαδοποίηση και κατηγοριοποίηση, εξόρυξη προτύπων κ.λπ. Το LingPipe, το οποίο αναπτύχθηκε από την Alias, είναι μια ισχυρή εργαλειοθήκη για τον καθαρισμό δεδομένων κειμένου, Tokenization, τεχνική POST, ανάλυση συχνότητας λέξεων, ανάλυση συναισθημάτων, ομαδοποίηση και κατηγοριοποίηση, εξόρυξη προτύπων κ.λπ. Το TextBlob προσφέρει μια διεπαφή προγραμματισμού εφαρμογών (API) για επεξεργασία φυσικής γλώσσας και μπορεί να χρησιμοποιηθεί για ουσιαστική εξαγωγή φράσεων, τεχνική POST, ανάλυση συναισθημάτων, ταξινόμηση, μετάφραση κ.λπ. (Hu, Chen & Chou, 2017).

### 7.2.1.3 Προκλήσεις και Μελλοντικές Κατευθύνσεις

Η διαρκώς αυξανόμενη χρήση των κοινωνικών δικτύων οδήγησε το ενδιαφέρον της τουριστικής βιομηχανίας στην έρευνα τουριστικού περιεχομένου μέσα από τα διαδικτυακά δεδομένα κειμένου μεγάλης κλίμακας, όπως είναι οι διαδικτυακές κριτικές και τα blogs. Ωστόσο, υπάρχει ακόμη σημαντικός χώρος για να επεκταθεί και να αναπτυχθεί η τουριστική έρευνα, καθώς βρίσκεται ακόμη σε πρώιμο στάδιο. Έχει εκτιμηθεί ότι η πραγματική εξόρυξη δεδομένων καλύπτει μόνο το 10% του χρόνου που απαιτείται για την πλήρη διαδικασία της ανακάλυψη γνώσης μέσα από τα δεδομένα μεγάλης κλίμακας (Pyle, 1999). Κατά τη γνώμη μας, αν και χρονοβόρο το βήμα της προεπεξεργασίας των δεδομένων είναι ουσιαστικής σημασίας για την εξόρυξη δεδομένων (Han & Kamber, 2001). Είναι κάτι παραπάνω από μια κουραστική αναγκαιότητα. Οι τεχνικές που χρησιμοποιούνται στο στάδιο της προεπεξεργασίας μπορούν να επηρεάσουν σε βάθος τα αποτελέσματα του επόμενου βήματος, την πραγματική δηλαδή εφαρμογή ενός αλγορίθμου εξόρυξης δεδομένων. Ως εκ τούτου, λόγω της κρισιμότητας του εν λόγω σταδίου θα πρέπει να αναπτυχθούν αποτελεσματικότεροι και αξιόπιστοι αλγόριθμοι προεπεξεργασίας των δεδομένων.

Το μέγεθος του δείγματος και τα σύνολα δεδομένων που χρησιμοποιούνται στις περισσότερες υπάρχουσες μελέτες είναι σε σχετικά μικρή κλίμακα. Επιπροσθέτως, τα μικρά δείγματα είναι επιρρεπή σε προκαταλήψεις επιλογής και μεροληψίες εκτίμησης, που ενδέχεται να οδηγήσουν σε λανθασμένα αποτελέσματα ανάλυσης και σε αντίθεση με τις πραγματικές καταστάσεις και τα γενικευμένα ευρήματα. Ως εκ τούτου, μελλοντική έρευνα με ένα μέγεθος δείγματος μεγαλύτερης κλίμακας είναι επιθυμητό και πιο αποτελεσματικό. Έχουν πραγματοποιηθεί αρκετές έρευνες που εστιάζουν στην εξαγωγή πολύτιμων γνώσεων από μεγάλα διαδικτυακά δεδομένα κειμένου, όπως για παράδειγμα τα τουριστικά συναισθήματα ή οι απόψεις για ένα τουριστικό προϊόν ή προορισμό. Ωστόσο, πώς να χρησιμοποιήσουμε όλη αυτή την αναδυόμενη γνώση, η οποία προκύπτει μέσω της τουριστικής έρευνας, για τον σχεδιασμό ελκυστικότερων τουριστικών προϊόντων και πακέτων για τον τουρίστα είναι ένα σημείο που χρήζει περαιτέρω διερεύνησης. Επιπλέον, άλλο ένα ενδιαφέρον ζήτημα, αναφορικά με τα διαδικτυακά δεδομένα κειμένου που σχετίζονται με τα ταξίδια, το οποίο μπορεί να εξεταστεί μελλοντικά, είναι η αξιοπιστία των δεδομένων. Για παράδειγμα, μερικές φορές οι επισκέπτες ενδέχεται να κάνουν μια ψεύτικη θετική κριτική, για να αποφύγουν πιθανά προβλήματα ή για να κερδίσουν κάποιο δώρο

(Schuckert, Liu & Law, 2015b). Η προαναφερθείσα τουριστική διαδικτυακή συμπεριφορά αξίζει μια βαθύτερη μελλοντική μελέτη.

Τέλος, λόγω των μη δομημένων και περίπλοκων χαρακτηριστικών που συναντάμε στα διαδικτυακά σχόλια/κριτικές και στα δεδομένα των ιστολογίων (blogs) που σχετίζονται με τον τουριστικό κλάδο, οι τεχνικές επεξεργασίας και ανάλυσης αυτών έρχονται αντιμέτωπες με σημαντικές προκλήσεις. Για παράδειγμα, οι μελέτες που έχουν πραγματοποιηθεί έως σήμερα αφορούσαν κυρίως τα κείμενα σε μία μόνο γλώσσα, ενώ οι κριτικές σε άλλες γλώσσες αγνοήθηκαν (Guo, Barnes & Jia, 2017). Ωστόσο, οι ταξιδιώτες από ένα διαφορετικό πολιτιστικό και γλωσσικό υπόβαθρο αναζητούν ένα διαφορετικό ταξιδιωτικό όφελος και μπορεί να έχουν διαφορετικές τουριστικές προτιμήσεις (Schuckert, Liu & Law, 2015a). Επομένως, η αποτελεσματική αναλυτική μέθοδος πολυγλωσσικών κειμένων είναι σίγουρα επιτακτική ανάγκη ούτως ώστε να αποφευχθεί η απώλεια πληροφοριών. Επιπλέον, στην ανάλυση συναισθήματος για τουριστικές κριτικές λαμβάνονται κυρίως υπόψη, ως σημαντικές λέξεις, τα ουσιαστικά, τα επίθετα και τα αρνητικά επιρρήματα. Ωστόσο, το συναίσθημα μπορεί να εκφράζεται και με τη χρήση βαθμού π.χ. «πολύ», «περισσότερο», «υπερβολικά» κ.ά. (Hu, Chen & Chou, 2017). Επομένως, και οι παραπάνω λέξεις πρέπει να ληφθούν υπόψη ως σημαντικές για την αποτελεσματικότερη ανάλυση των τουριστικών συναισθημάτων.

#### 7.2.2 Διαδικτυακά Δεδομένα Φωτογραφιών και Τεχνικές Εξόρυξης

Εκτός από τα διαδικτυακά δεδομένα κειμένου, οι χρήστες δημοσιεύουν και φωτογραφίες στα κοινωνικά δίκτυα. Οι φωτογραφίες που ανεβαίνουν από τους τουρίστες περιέχουν πολλές χρήσιμες πληροφορίες σε σχέση με τους ίδιους, τις τουριστικές τοποθεσίες και το χρόνο ανάρτησης, παρέχοντας μια νέα προοπτική για τη μελέτη της τουριστικής συμπεριφοράς, τις ταξιδιωτικές προτάσεις π.χ. τουριστικά σημεία ενδιαφέροντος και ταξιδιωτικά σχέδια, και το τουριστικό μάρκετινγκ. Επομένως, τα διαδικτυακά δεδομένα φωτογραφιών σε έναν μη δομημένο τύπο είναι φυσικό να κεντρίσουν τη προσοχή της τουριστικής βιομηχανίας.

Τα διαδικτυακά δεδομένα φωτογραφιών μεταφέρουν πολλά χρήσιμα μηνύματα σε μορφή μεταδεδομένων, δηλαδή, μεταδεδομένα είναι οι ετερογενείς πληροφορίες που



είναι ενσωματωμένες σε φωτογραφίες. Τα πολύτιμα μεταδεδομένα για την τουριστική έρευνα εμπίπτουν σε τέσσερις κύριες κατηγορίες:

- ⊕ Πληροφορίες σχετικά με το χρήστη, όπως το ID της φωτογραφίας για την οποία πραγματοποιήθηκε λήψη από μια ιστοσελίδα και το ID του χρήστη που μεταφόρτωσε μια φωτογραφία.
- ⊕ Χρονικές πληροφορίες, όπως η ημερομηνία και ώρα λήψης της φωτογραφίας καθώς και η ημερομηνία και ώρα μεταφόρτωσής της σε κάποια ιστοσελίδα.
- ⊕ Γεωγραφικές πληροφορίες, το γεωγραφικό πλάτος και μήκος που πραγματοποιήθηκε η λήψη της φωτογραφίας.
- ⊕ Πληροφορίες κειμένου, τις οποίες καθόρισαν οι τουρίστες, όπως ο τίτλος, η περιγραφή και οι ετικέτες της φωτογραφίας.

#### 7.2.2.1 Τεχνικές Εξόρυξης Διαδικτυακών Φωτογραφιών

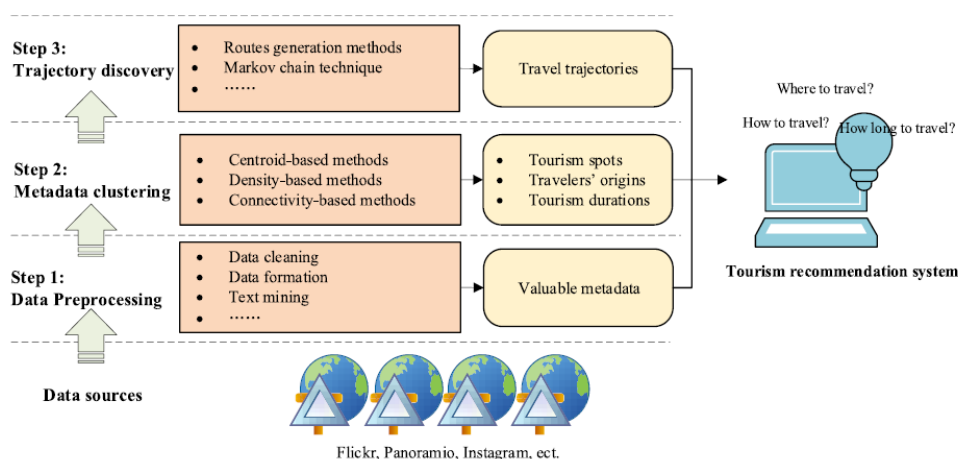
Για να εξερευνήσουμε τις πολύτιμες πληροφορίες που κρύβονται στα διαδικτυακά δεδομένα φωτογραφιών, εφαρμόζεται μια ποικιλία τεχνικών εξόρυξης διαδικτυακών φωτογραφιών για την κατασκευή ενός συστήματος τουριστικών προτάσεων. Οι εν λόγω τεχνικές εξόρυξης ομαδοποιούνται σε τρία σημαντικά βήματα:

- ⊕ Την προεπεξεργασία δεδομένων
- ⊕ Την ομαδοποίηση μεταδεδομένων και
- ⊕ Την ανακάλυψη προτύπων

Η εικόνα 18 απεικονίζει μια τυπική διαδικασία της χρήσης διαδικτυακών φωτογραφικών δεδομένων στην τουριστική έρευνα μαζί με τις αντίστοιχες τεχνικές εξόρυξης που χρησιμοποιούνται στο κάθε βήμα.

Αρχικά, τα πρωτογενή δεδομένα που συλλέγονται από ιστότοπους κοινής χρήσης φωτογραφιών (Flickr, Instagram κ.λπ.) επιδέχονται μια προεπεξεργασία όσον αφορά τον καθαρισμό δεδομένων, τη διαμόρφωση και την εξόρυξη κειμένου, έτσι ώστε να παραμείνουν τα πολύτιμα μεταδεδομένα για περαιτέρω αναλύσεις. Για παράδειγμα, οι Lee και Cai, (2014) κατήγησαν τις διπλότυπες φωτογραφίες και μετέτρεψαν τα δεδομένα σε μορφοποιημένο τύπο για ανάλυση ομαδοποίησης. Ο Önder (2017) χρησιμοποίησε μια μέθοδο καθαρισμού δεδομένων για να προσδιοριστεί αν οι

φωτογραφίες δημοσιεύτηκαν από κατοίκους του συγκεκριμένου προορισμού ή από τουρίστες. Όσον αφορά τα μεταδεδομένα κειμένου των φωτογραφιών, χρησιμοποιήθηκαν μέθοδοι εξόρυξης κειμένου για να ερευνηθούν τα ενδιαφέροντα και τα κίνητρα των τουριστών με βάση τις φωτογραφίες που ανήρτησαν (Miah, Vu, Gammack & McGrath, 2017). Όπως εύκολα μπορούμε να αντιληφθούμε, μέσω της προεπεξεργασίας των δεδομένων μπορούν να εξαχθούν πολύτιμα μεταδεδομένα για περαιτέρω μελέτες στα ακόλουθα δύο βήματα.



**Εικόνα 18.** Διαδικασία χρήσης διαδικτυακών φωτογραφικών δεδομένων στην έρευνα της τουριστικής βιομηχανίας (Li, Xu, Tang & Wang, 2018)

Στη συνέχεια, διεξάγεται ανάλυση ομαδοποίησης στα εξαγόμενα μεταδεδομένα του πρώτου βήματος με βάση τρεις κύριες προοπτικές: τη **χωρική** διάσταση για να ανακαλύψουμε τουριστικά σημεία ενδιαφέροντος (π.χ. Kisilevich, Krstajic, Keim & Andrienko, 2010; Lee & Cai, 2014; Mamei, Rosi & Zambonelli, 2010; Okuyama & Yanai, 2013; Zhou, Xu & Kimmons, 2015), τη **γεωγραφική** διάσταση του χρήστη για να ανακαλύψουμε την προέλευση των ταξιδιωτών (Da Rugna, Chareyron & Branchet, 2012) και τέλος τη **χρονική** διάσταση για να εντοπίσουμε τη τουριστική διάρκεια ενός ταξιδιού (Lu, Wang, Yang, Pang & Zhang, 2010; Popescu, Grefenstette & Moëllic, 2009). Η υφιστάμενη μέθοδος ομαδοποίησης για την ανάλυση των μεταδεδομένων στην τουριστική έρευνα μπορεί να ταξινομηθεί σε τρεις κύριες κατηγορίες: με βάση το κεντροειδές, με βάση την πυκνότητα και με βάση τις μεθόδους συνδεσιμότητας.

Όσον αφορά τη κεντροειδή μέθοδο, ο k-means είναι ένας δημοφιλής αλγόριθμος στην τουριστική έρευνα. Ωστόσο, όπως υποστηρίζουν οι Lee και Cai (2014), ο k-means αλγόριθμος που βασίζεται στη κεντροειδή μέθοδο έχει ένα προφανές μειονέκτημα, ότι δηλαδή απαιτεί μια εξαντλητική αναζήτηση για να βρεθεί το καλύτερο κέντρο του δείγματος και το γεγονός αυτό τον καθιστά αναποτελεσματικό σε περιβάλλοντα με μεγάλη κλίμακα δεδομένων. Αντίθετα, η ομαδοποίηση βάσει πυκνότητας μπορεί να είναι πιο κατάλληλη για μεγάλα διαδικτυακά δεδομένα φωτογραφιών, καθώς δεν απαιτείται η πολύ καλή γνώση του τομέα παρά μόνο ένα αποτελεσματικό φιλτράρισμα των ακραίων σημείων του δείγματος. Επιπλέον, λειτουργεί πιο αποτελεσματικά ακόμη και στη παρουσία σημείων θορύβου στο δείγμα μας. Ως εκ τούτου, η χωρική ομαδοποίηση με βάση την πυκνότητα έχει αναπτυχθεί και εφαρμοστεί εκτενώς στην ομαδοποίηση φωτογραφιών για τις ανάγκες της τουριστικής έρευνας (Lee & Cai, 2014; Majid, Chen, Mirza, Hussain & Woodward, 2013; Miah, Vu, Gammack & McGrath, 2017; Oku, Hattori & Kawagoe, 2015; Vu, Li, Law & Ye, 2015; Zhou, Xu & Kimmons, 2015). Για να οικοδομήσουμε μια ιεραρχία ομάδων, η ομαδοποίηση βάσει συνδεσιμότητας (επίσης γνωστή ως ιεραρχική ομαδοποίηση), είναι ένας ευέλικτος και γρήγορος αλγόριθμος που βασίζεται στην ανομοιότητα του δείγματος ανεξάρτητα από τους μετρικούς χώρους (Quack, Leibe & Van Gool, 2008). Έχει επίσης χρησιμοποιηθεί για τον εντοπισμό τουριστικών θέσεων και την ομαδοποίηση των πόλεων που επισκέφθηκαν οι τουρίστες (Okuyama & Yanai, 2013; Önder, 2017).

Τέλος, οι διαδρομές του ταξιδιού, δηλαδή η ακολουθία των τουριστικών σημείων και τα χρονικά διαστήματα μεταξύ τους, διερευνώνται για τη δημιουργία κατάλληλων προσφερόμενων ταξιδιωτικών πακέτων. Αναφορικά με τις τεχνικές για την ανακάλυψη προτύπων, έχει δημιουργηθεί και χρησιμοποιηθεί μια ποικιλία μεθόδων δημιουργίας διαδρομών (Lu, Wang, Yang, Pang & Zhang, 2010; Okuyama & Yanai, 2013). Επιπλέον, η τεχνική Markov chain είναι ένα μοντέλο πρόβλεψης που χειρίζεται διαδοχικές πληροφορίες. Επιπροσθέτως, είναι ένας πολλά υποσχόμενος αλγόριθμος ο οποίος λαμβάνει πιο λεπτομερείς πληροφορίες των ταξιδιωτικών διαδρομών και μπορεί να προβλέψει τα επόμενα τουριστικά σημεία με βάση τις τρέχουσες τοποθεσίες των τουριστών (Kurashima, Iwata, Irie & Fujimura, 2013, Lu, Wang, Yang, Pang & Zhang, 2010; Vu, Li, Law & Ye, 2015). Μέσω αυτού του

βήματος, οι διαδρομές ενός ταξιδιού μπορούν να εξαχθούν από μαζικές κινήσεις ταξιδιωτών (Okuyama & Yanai, 2013).

Συμπερασματικά, μέσα από τα παραπάνω τρία βήματα, μπορεί να σχηματισθεί ένα ισχυρό σύστημα τουριστικών προτάσεων, το οποίο θα χρησιμοποιηθεί όχι μόνο για την εξαγωγή πολύτιμων πληροφοριών από διαδικτυακές φωτογραφίες για την τουριστική συμπεριφορά, τα τουριστικά σημεία ενδιαφέροντος, τη διάρκεια και τις διαδρομές του ταξιδιού αλλά θα παρέχει χρήσιμες και ολοκληρωμένες προτάσεις για να ταξιδέψετε, όπως επί παραδείγματι πού, πως και πόσο καιρό να ταξιδέψετε.

#### 7.2.2.2 Προκλήσεις και Μελλοντικές Κατευθύνσεις

Σύμφωνα με τα όσα ειπώθηκαν παραπάνω, η τουριστική έρευνα χρησιμοποιώντας τα διαδικτυακά δεδομένα φωτογραφιών μπορεί να βελτιωθεί αφενός μεν επεκτείνοντας την ερευνητική της περιοχή και αφετέρου δε τελειοποιώντας τις αναλυτικές προς χρήση τεχνικές. Όσον αφορά την ερευνητική περιοχή, το να ανακαλύψεις τουριστικά σημεία ενδιαφέροντος (δηλαδή πού να ταξιδέψεις) από τη χωρική διάσταση των διαδικτυακών φωτογραφιών είναι ένα κομμάτι που έχει μελετηθεί επαρκώς. Εν αντιθέσει, η έρευνα για τη χρονική διάσταση, δηλαδή το πότε να ταξιδέψεις, χρήζει περεταίρω διερεύνησης καθώς λίγες έρευνες έχουν καταπιαστεί με αυτό το κομμάτι. Επιπλέον, εξίσου σημαντικοί χρονικοί παράγοντες για διερεύνηση είναι το πότε θεωρείται ο καλύτερος χρόνος για να ταξιδέψεις καθώς και η διάρκεια του ταξιδιού.

Αναφορικά με τις αναλυτικές τεχνικές, η ομαδοποίηση και η ανάλυση αλληλουχιών έχουν χρησιμοποιηθεί ευρέως για τη διερεύνηση του τουρισμού με βάση τα διαδικτυακά δεδομένα φωτογραφιών. Επίσης, και άλλες σημαντικές τεχνικές εξόρυξης δεδομένων έχουν χρησιμοποιηθεί στην τουριστική έρευνα για την εκτίμηση, την ταξινόμηση, την πρόβλεψη, τους κανόνες συγγένειας και συσχέτισης. Κρίνουμε σημαντικό να αναφέρουμε ότι οι περισσότερες υπάρχουσες μελέτες διεξάγουν αναλύσεις έμμεσα, πάνω στα μεταδεδομένα που έχουν ενσωματωθεί στις φωτογραφίες. Ωστόσο, τα ίδια τα δεδομένα φωτογραφιών περιλαμβάνουν πληθώρα ενδιαφέρουσας πληροφορίας εκτός από τα μεταδεδομένα, ως εκ τούτου απαιτούνται ισχυρές τεχνικές εξόρυξης διαδικτυακών φωτογραφιών οι οποίες θα δρουν απευθείας στις εικόνες.

## **Κεφάλαιο 8**

### **Συμπεράσματα**

Στη νέα εποχή των μεγάλων δεδομένων, δεδομένα μεγάλης κλίμακας και διαφορετικών τύπων (κείμενα και φωτογραφίες) έχουν χρησιμοποιηθεί για την κατανόηση και την προώθηση του τουριστικού κλάδου. Στη παρούσα εργασία γίνεται μια διεξοδική αναφορά της εξόρυξης διαδικτυακών δεδομένων, τα οποία αναρτούν οι χρήστες στα κοινωνικά δίκτυα, αναφορικά με τον τουρισμό. Επιπλέον, γίνεται ειδική μνεία για τις τεχνικές εξόρυξης διαδικτυακών δεδομένων καθώς και τις προκλήσεις που θα κληθεί να αντιμετωπίσει η έρευνα στον τομέα του τουρισμού για να βελτιστοποιήσει τα ευρήματα ως προς το προσφερόμενο τουριστικό πακέτο. Τέλος, προτείνονται μελλοντικές κατευθύνσεις για την αποτελεσματικότερη ανάλυση και ερμηνεία του τουριστικού κλάδου.

Δεδομένου ότι η αξιοποίηση των μεγάλων διαδικτυακών δεδομένων για την βελτιστοποίηση του προσφερόμενου τουριστικού πακέτου βρίσκεται ακόμη σε πρώιμο στάδιο, παρόλα αυτά μεγάλες ταξιδιωτικές εταιρίες επιχειρούν να εκμεταλλευτούν όλο και περισσότερο τα οφέλη που προκύπτουν μέσω της εξόρυξης των δεδομένων μεγάλης κλίμακας που απλόχερα προσφέρουν οι χρήστες των κοινωνικών δικτύων και της αναδύμενης πληροφορίας που απορρέει από αυτά.

Το περιεχόμενο που δημιουργείται από τους χρήστες (κείμενα και εικόνες) στα κοινωνικά δίκτυα εξυπηρετεί την ανάλυση του τουριστικού συναισθήματος, την ανάλυση της τουριστικής συμπεριφοράς, το τουριστικό μάρκετινγκ και εν γένει τη βελτιστοποίηση των τουριστικών προτάσεων και προσφορών. Οι τουρίστες μεταφέρουν οικειοθελώς τα συναισθήματά τους από έναν ταξιδιωτικό προορισμό και αυτό αποτελεί από μόνο του σημαντικό πλεονέκτημα για την αναβάθμιση του τουριστικού κλάδου. Η αναδύμενη γνώση που προκύπτει από τις διαδικτυακές φωτογραφίες είναι εξίσου σημαντική για να ανθίσει μια νέα εποχή στον εν λόγω τομέα. Από την άλλη πλευρά όμως, στην περίπτωση του διαδικτυακού κειμένου, προκύπτουν και θέματα αξιοπιστίας, κατά πόσο δηλαδή οι τουρίστες μεταφέρουν επακριβώς τη γνώμη τους ή είναι ωραιοποιημένη λόγω κάποιας, παραδείγματος χάριν, αμοιβής ή δώρου. Επίσης, παρατηρείται χαμηλή ακρίβεια στην πρόβλεψη της τοποθεσίας όταν τα διαδικτυακά δεδομένα είναι φωτογραφίες.

Τέλος, αν και έχουν υπάρξει σημαντικές βελτιώσεις και καινοτομίες, υπάρχει ακόμη άφθονος χώρος για την ανάπτυξη της τουριστικής έρευνας χρησιμοποιώντας μεγάλα διαδικτυακά δεδομένα, ειδικά από την προοπτική της επέκτασης του ερευνητικού χώρου και την ανάπτυξη αναλυτικών τεχνικών. Συγκεκριμένα, η αξιοποίηση των δεδομένων ήχου, βίντεο και πολλαπλών τύπων από τα κοινωνικά δίκτυα βρίσκεται σε εμβρυακό στάδιο παρόλου που αποτελούν ένα πλούσιο «ορυχείο» χρήσιμων πληροφοριών, το οποίο θα μας βοηθήσει να κατανοήσουμε καλύτερα την τουριστική συμπεριφορά και γενικότερα τον τουριστικό τομέα. Επίσης, περιφερειακά δεδομένα ή δεδομένα μεταξύ τομέων π.χ. υγεία, ασφάλιση, εκπαίδευση είναι εξίσου πολύτιμα καθώς μπορούν να επηρεάσουν σημαντικά την τουριστική προτίμηση. Όσον αφορά τις αναλυτικές τεχνικές, δεδομένου ότι τα μεγάλα δεδομένα είναι συνήθως σε μαζική κλίμακα, μη δομημένα (ή ημι-δομημένα) και έχουν πολύπλοκη μορφή, είναι απαραίτητη η ανάπτυξη κατάλληλων αναλυτικών τεχνικών, οι οποίες θα εγγυώνται την αποτελεσματικότητα και την εγκυρότητα των αναδυόμενων πληροφοριών. Υποσχόμενα παραδείγματα αποτελούν η ανάλυση ομιλίας για δεδομένα ήχου, η ανάλυση περιεχομένου βίντεο για δεδομένα βίντεο (Gandomi & Haider, 2015) και οι υβριδικές τεχνικές για πολλαπλούς τύπους ή / και πολύ-χαρακτηριστικά δεδομένα (Tang, Yu, Liu & Xu, 2013). Ως εκ τούτου, παραδείγματα προηγμένης αναλυτικής μεθόδου όπως η μηχανική μάθηση και η βαθιά μάθηση, που χρησιμοποιούνται σε δεδομένα μεγάλης κλίμακας, είναι πραγματικά πολλά υποσχόμενες τεχνικές.

## Βιβλιογραφία/Αναφορές

- Adedoyin-Olowe M., Mohamed Medhat G., Stahl F., “A Survey of Data Mining Techniques for Social Media Analysis”, School of Computing Science and Digital Media, Robert Gordon University Aberdeen, AB10 7QB, UK, School of Systems Engineering, University of Reading, Whiteknights, Reading, RG6 6AY, UK.
- AnalyticsVidhya, (2020), [Ηλεκτρονικός σύνδεσμος] <https://www.analyticsvidhya.com/blog/2019/06/comprehensive-guide-text-summarization-using-deep-learning-python/> [Ανακτήθηκε 15-12-2020].
- Artola C., Pinto F., Pedraza P., (2015), “Can internet searches forecast tourism inflows?” International Journal of Manpower, vol. 36(1), Pages 103-116.
- Bajari P., Nekipelov D., Ryan S. P., Yang M., (2015), “Machine Learning Methods for Demand Estimation”, American Economic Review, vol. 105, no. 5, Pages 481-485.
- Bangwayo-Skeete, P. F., Skeete R. W., (2015), “Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach”, Tourism Management, vol. 46, Pages 454-464.
- Berry M. J. A., (1997), “Data Mining Techniques: For Marketing, Sales, and Customer Support”.
- Βερούκιος Β., Καγκλής Β., Σταυρόπουλος Η., (2015), “Η επιστήμη των δεδομένων μέσα από τη γλώσσα R”, Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, κεφ 1, [Ηλεκτρονικός σύνδεσμος]<https://repository.kallipos.gr/handle/11419/2966> [Ανακτήθηκε 15-12-2020].
- Bigné J. E., Sánchez M. I., Sánchez J., (2001), “Tourism image, evaluation variables and after purchase behaviour: inter-relationship. Tour. Manag”, vol. 22, Pages 607-616, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/abs/pii/S0261517701000358?via%3Dihub> [Ανακτήθηκε 15-12-2020].
- Bollier D., Firestone C. M., (2010), “The Promise and Peril of Big Data”, The Aspen Institute.
- Bordogna G., Frigerio L., Cuzzocrea A., Psaila G., (2016), “Clustering geo-tagged tweets for advanced big data analytics”, In 2016 IEEE international congress on big data, San Francisco, USA.
- Borrego-Jaraba F., Miguel A., Irene Luque R., Gomez N., (2011), “A NFC-based pervasive solution for city touristic surfing”, Published online: 18 January 2011, Pers Ubiquit Comput vol. 15, Pages 731-742, [Ηλεκτρονικός σύνδεσμος] <https://link.springer.com/article/10.1007/s00779-010-0364-y> [Ανακτήθηκε 15-12-2020].
- Bose I., Mahapatra R., (2001), “Business Data Mining –A Machine Learning Perspective”, Information Management 39, Pages 211-225.
- Brachman R. J., Anand T., (1996), “The Process of Knowledge Discovery in Databases”, Intelligent Information Management, vol. 7, no.3.

- Buhalis D., Law R., (2008), “*Progress in information technology and tourism management: 20 years on and 10 years after the internet – the state of eTourism research*”, *Tourism Management*, vol. 29, No. 4, Pages 609-623.
- Carson D., (2008), “*The ‘blogosphere’ as a market research tool for tourism destinations: A case study of Australia’s Northern Territory*”, *Journal of Vacation Marketing*, vol. 14(2), Pages 111–119, [Ηλεκτρονικός σύνδεσμος] <https://journals.sagepub.com/doi/10.1177/1356766707087518> [Ανακτήθηκε 15-12-2020].
- Chase C. W., (2013), “*Using Big Data to Enhance Demand-Driven Forecasting and Planning*”, *The Journal of Business Forecasting*, vol. 32, no. 2, Pages 27-32.
- Da Rugna J., Chareyron G., Branchet B., (2012), “*Tourist behavior analysis through geotagged photographs: A method to identify the country of origin*”, In 2012 IEEE 13th international symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary.
- Dan W., Zheng X., Fesenmaier D. R., (2014), “*Smartphone Use in Everyday Life and Travel*”, First Published May 19, Research Article, [Ηλεκτρονικός σύνδεσμος] <https://journals.sagepub.com/doi/10.1177/0047287514535847> [Ανακτήθηκε 15-12-2020].
- Datanami, (2020), [Ηλεκτρονικός σύνδεσμος] <https://www.datanami.com/2017/06/14/8-concrete-data-mining-techniques-will-deliver-best-results/> [Ανακτήθηκε 15-12-2020].
- De Noni I., Orsi L., Zanderighi L., (2014), “*Attributes of Milan influencing city brand attractiveness*”, *J. Destin. Mark. Manag.* 3, Pages 218–226, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/pii/S2212571X14000304?via%3Dihub> [Ανακτήθηκε 15-12-2020].
- Deloitte UK, (2015), “*Global Economic Impact of Facebook*”, [Ηλεκτρονικός σύνδεσμος] <https://www2.deloitte.com/uk/en/pages/technology-media-and-telecommunications/articles/the-global-economic-impact-of-facebook.html> [Ανακτήθηκε 15-12-2020].
- Dheddy H., Husada Fadjar P., (2014), “*Trip Guidance: A Linked Data Based Mobile Tourists Guide*”, *Advanced Science Letters* vol. 20(1), Pages 75-79.
- Dolnicar S., (2002), “*A Review of Data-Driven Market Segmentation in Tourism*”, *J. Travel Tour. Mark.* 12, Pages 1–22, [Ηλεκτρονικός σύνδεσμος] [https://www.tandfonline.com/doi/abs/10.1300/J073v12n01\\_01](https://www.tandfonline.com/doi/abs/10.1300/J073v12n01_01) [Ανακτήθηκε 15-12-2020].
- Dolnicar S., Grün B., Leisch F., Schmidt K., (2014), “*Required Sample Sizes for Data-Driven Market Segmentation Analyses in Tourism*”, *J. Travel Res.*, vol. 53, Pages 296–306, [Ηλεκτρονικός σύνδεσμος] <https://journals.sagepub.com/doi/10.1177/0047287513496475> [Ανακτήθηκε 15-12-2020].
- Domo, (2020), “*Data never sleeps*”, [Ηλεκτρονικός σύνδεσμος] <https://www.domo.com/learn/data-never-sleeps-8> [Ανακτήθηκε 15-12-2020].
- Elbe J., Emmoth A., (2014), “*The use of rhetoric in legitimation strategies when mobilizing destination stakeholders*”, *J. Destin. Mark. Manag.*, vol. 3, Pages 210–217, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/pii/S2212571X14000341?via%3Dihub> [Ανακτήθηκε 15-12-2020].



- Elgendy N., Elragal A., (2014), “*Big data analytics: a literature review paper*”, In Industrial Conference on Data Mining, Springer, Cham, Pages 214-227.
- Erl Thomas K. W., (2016), “*Big Data Fundamentals Concepts, Drivers & Techniques*”, Service, USA: Prentice Hall, Service Tech Press.
- Facebook, (2020), [Ηλεκτρονικός σύνδεσμος] <https://about.fb.com/company-info/> [Ανακτήθηκε 15-12-2020].
- Fang B., Ye Q., Kucukusta D., Law R., (2016), “*Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics*”, Tourism Management, vol. 52, Pages 498-506.
- Flickr, (2013), [Ηλεκτρονικός σύνδεσμος] <https://en.wikipedia.org/wiki/Flickr> [Ανακτήθηκε 15-12-2020].
- Frawley W. J., Piatetsky-Shapiro G., Matheus C. J., (1992), “*Knowledge Discovery in Databases: An Overview*”, vol. 13, no.3.
- Frederiksen L., (2012), “*Big Data*”, Pages 345-349, Published online: 20 Nov 2012, [Ηλεκτρονικός σύνδεσμος] <https://www.tandfonline.com/doi/abs/10.1080/15228959.2012.730411> [Ανακτήθηκε 15-12-2020].
- Fuchs M., Höpken W., Lexhagen M., (2014), “*Big data analytics for knowledge generation in tourism destinations – A case from Sweden*”, J. Destin. Mark. Manag., vol. 3, Pages 198–209, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/pii/S2212571X14000353?via%3Dihub> [Ανακτήθηκε 15-12-2020].
- Fuchs M., Höpken W., Lexhagen M., (2014), “*Sentiment analysis—extracting decision-relevant knowledge from UGC*”, in Z. Xiang & I. Tussyadiah (Eds.), Information and communication technologies in tourism, Pages 249–270.
- Gandomi A., Haider M., (2015), “*Beyond the hype: Big data concepts, methods, and Analytics*”, International Journal of Information Management, vol 35(2), Pages 137-144.
- García-Pablos A., Cuadros M., Linaza M. T., (2016), “*Automatic analysis of textual hotel reviews*”, Information Technology & Tourism, vol. 16(1), Pages 45–69.
- Greg R., (2011), “*Creativity and tourism: The State of the Art*”, Annals of Tourism Research vol. 38, Issue 4, October 2011, Pages 1225-1253, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/abs/pii/S0160738311001204?via%3Dihub> [Ανακτήθηκε 15-12-2020].
- Gretzel U., Sigala M., Xiang Z., Koo C., (2015), “*Smart tourism: Foundations and developments*”, Electronic Markets.
- Guo Y., Barnes S. J., Jia Q., (2017), “*Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichl et al location*”, Tourism Management, vol. 59, Pages 467–483.
- Hays S., Page S. J., Buhalis D., (2013), “*Social media as a destination marketing tool: its use by national tourism organisations*”, Current Issues in Tourism, vol. 16, no. 3, Pages 211-239.

- Hjalager A. M., Nordin S., (2011), “*User-driven Innovation in Tourism—A Review of Methodologies*”, J. Qual. Assur. Hosp. Tour. 12, Pages 289–315. [Ηλεκτρονικός σύνδεσμος] <https://www.tandfonline.com/doi/abs/10.1080/1528008X.2011.541837> [Ανακτήθηκε 15-12-2020].
- Hu Y. H., Chen Y. L., Chou H. L., (2017), “*Opinion mining from online hotel reviews – a text summarization approach*”, Information Processing & Management, vol. 53(2), Pages 436–449.
- Internet World Stats, (2020), [Ηλεκτρονικός σύνδεσμος] <https://www.internetworldstats.com/facebook.htm> [Ανακτήθηκε 15-12-2020].
- Inversini A., Marchiori E., Dedekind C., Cantoni L., (2010), “*Applying a Conceptual Framework to Analyze Online Reputation of Tourism Destinations*”, Information and Communication Technologies in Tourism 2010, Pages 321–332, [Ηλεκτρονικός σύνδεσμος] [https://link.springer.com/chapter/10.1007%2F978-3-211-99407-8\\_27](https://link.springer.com/chapter/10.1007%2F978-3-211-99407-8_27) [Ανακτήθηκε 15-12-2020].
- Jeng J., Fesenmaier D. R., (2002), “*Conceptualizing the Travel Decision-Making Hierarchy: A Review of Recent Developments*”, [Ηλεκτρονικός σύνδεσμος] <https://www.ingentaconnect.com/content/cog/ta/2002/00000007/00000001/art00002;jsessionid=58adsp163uvm7.x-ic-live-03> [Ανακτήθηκε 15-12-2020].
- Jiawei H., Micheline K., (2001), “*Data Mining: Concepts and Techniques*”, Intelligent Database Systems Research Lab School of Computing Science, Simon Fraser University, Canada.
- Jingjing L., Lizhi X., Ling T., Shouyang W., Ling L., (2018), “*Big data in tourism research: A literature review Tourism Management*”, vol. 68, Pages 301-323.
- Johns N., Gyimóthy S., (2002), “*Market Segmentation and the Prediction of Tourist Behavior: The Case of Bornholm*”, Denmark. J. Travel Res., vol. 40, Pages 316–327, [Ηλεκτρονικός σύνδεσμος] <https://journals.sagepub.com/doi/10.1177/0047287502040003009> [Ανακτήθηκε 15-12-2020].
- Kdnuggets, (2020), [Ηλεκτρονικός σύνδεσμος] <https://www.kdnuggets.com/images/sentiment-fig-1-689.jpg> [Ανακτήθηκε 15-12-2020].
- Kim K., Park O., Yun S., Yun H., (2017), “*What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management*”, Technological Forecasting and Social Change, vol. 123, Pages 362–369, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/abs/pii/S0040162517300112?via%3Dihub> [Ανακτήθηκε 15-12-2020].
- Kisilevich S., Krstajic M., Keim D., Andrienko N., Andrienko G., (2010), “*Eventbased analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections*”, In 2010 14th international conference on Information Visualisation (IV), London, UK.
- Kleinberg J., (1999), “*Authoritative sources in a hyperlinked environment*”, Journal of the ACM, vol. 46 (5), Pages 604–632.

- Költringer C., Dickinger A., (2015), “*Analyzing destination branding and image from online sources: A web content mining approach*”, *Journal of Business Research*, vol. 68(9), Pages 1836-1843.
- Kurashima T., Iwata T., Irie G., Fujimura K., (2013), “*Travel route recommendation using geotagged photos*”, *Knowledge and Information Systems*, vol. 37(1), Pages 37-60.
- Κύρκος Ε., (2015), “*Επιχειρηματική ευφορία και εξόρυξη δεδομένων*”, Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, [Ηλεκτρονικός σύνδεσμος] <https://repository.kallipos.gr/handle/11419/1226> [Ανακτήθηκε 15-12-2020].
- Lange-Faria W., Elliot S., (2012), “*Understanding the role of social media in destination marketing*”, *Tourismos*, vol. 7(1), Pages 193–211.
- LaValle S., Lesser E., Shockley R., Hopkins M. S., Kruschwitz N., (2011), “*Big Data, Analytics and the Path From Insights to Value*”, *MIT Sloan Management Review*, vol. 52, no. 2, Pages 21-31.
- Lee I., Cai G., Lee K., (2014), “*Exploration of geo-tagged photos through data mining approaches*”, *Expert Systems with Applications*, vol. 41(2), Pages 397-405.
- Li M., Yi L., (2014), “*Application and optimization of customer relationship management in tongcheng website*”, *Business Economy*, vol. 36(11), Pages 136–141.
- Li Y. R., Lin Y. C., Tsai P. H., Wang Y. Y., (2015), “*Traveller-Generated Contents for Destination Image Formation: Mainland China Travellers to Taiwan as a Case Study*”, *Journal of Travel and Tourism Marketing*, vol. 32(5), Pages 518–533, [Ηλεκτρονικός σύνδεσμος] <https://www.tandfonline.com/doi/full/10.1080/10548408.2014.918924> [Ανακτήθηκε 15-12-2020].
- Li X., Pan B., Law R., Huang X., (2017), “*Forecasting tourism demand with composite search index*”, *Tourism Management*, vol. 59, Pages 57-66.
- Lian T., Yu C., (2017), “*Representation of online image of tourist destination: A content analysis of huangshan*”, *Asia Pacific Journal of Tourism Research*, vol. 22(2), Pages 1–20.
- Liu B., (2011), “*Web Data Mining, Exploring Hyperlinks, Contents, and Usage Data*”, 2nd Edition, Springer – Verlag.
- Lu X., Wang C., Yang J. M., Pang Y., Zhang L., (2010), “*Photo2trip: Generating travel routes from geo-tagged photos for trip planning*”, In *Proceedings of the 18th ACM international conference on multimedia*, Firenze, Italy.
- Lu W., Stepchenkova S., (2012), “*Ecotourism experiences reported online: Classification of satisfaction attributes*”, *Tourism Management*, vol. 33(3), Pages 702-712.
- Lu D., Wu R., Sang J., (2017), “*Overlapped user-based comparative study on photosharing Websites*”, *Information Sciences*, vol. 376, Pages 54-70.
- Luigi A., Antonio I., Giacomo M., (2010), “*The Internet of Things: A survey*”, *Computer Networks*, vol. 54, Issue 15, 28 October 2010, Pages 2787-2805, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/abs/pii/S1389128610001568?via%3Dihub> [Ανακτήθηκε 15-12-2020].

- Majid A., Chen L., Chen G., Mirza H. T., Hussain I., Woodward J., (2013), “A context-aware personalized travel recommendation system based on geotagged social media data mining”, *International Journal of Geographical Information Science*, vol. 27(4), Pages 662-684.
- Mamei M., Rosi A., Zambonelli F., (2010), “Automatic analysis of geotagged photos for intelligent tourist services”, In 2010 6th international conference on Intelligent Environments (IE), Kuala Lumpur, Malaysia.
- Marchiori E., Cantoni L., (2015), “The role of prior experience in the perception of a tourism destination in user-generated content”, *Journal of Destination Marketing and Management*, vol. 4(3), Pages 194–201, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/pii/S2212571X15000323?via%3Dihub> [Ανακτήθηκε 15-12-2020].
- Marine-Roig E., Clavé S. A., (2015), “Tourism analytics with massive user generated content: A case study of Barcelona”, *Journal of Destination Marketing & Management*, vol. 4(3), Pages 162-172.
- Marko K., (2015), “Using Big Data And Machine Learning To Enrich Customer Experiences”, *Forbes*.
- McAfee A., Brynjolfsson E., Davenport T. H., Patil D. J., Barton D., (2012), “Big data: the management revolution”, *Harvard business review*, vol. 90(10), Pages 60-68.
- Meeker W. Q., Hong Y., (2014), “Reliability Meets Big Data: Opportunities and Challenges”, *Quality Engineering*, vol. 26, Pages 102-116.
- Miah S. J., Vu H. Q., Gammack J., McGrath M., (2017), “A big data analytics method for tourist behaviour analysis”, *Information & Management*, vol. 54(6), Pages 771-785.
- MTD, (2020), [Ηλεκτρονικός σύνδεσμος] <https://mtd.media/data-cleansing/> [Ανακτήθηκε 15-12-2020].
- Munar M. A., (2011), “Tourist-created content: Rethinking destination branding”, *International Journal of Culture, Tourism and Hospitality Research*, vol. 5(3), Pages 291–305, [Ηλεκτρονικός σύνδεσμος] <https://www.emerald.com/insight/content/doi/10.1108/17506181111156989/full/html> [Ανακτήθηκε 15-12-2020].
- Munar M. A., Gyimóthy S., Cai L., (2013), “Tourism Social Media. In *Tourism Social Science*”, vol. 18, Pages 1–15, [Ηλεκτρονικός σύνδεσμος] [https://www.emerald.com/insight/content/doi/10.1108/S1571-5043\(2013\)0000018003/full/html](https://www.emerald.com/insight/content/doi/10.1108/S1571-5043(2013)0000018003/full/html) [Ανακτήθηκε 15-12-2020].
- Newman D., (2016), *Forbes*, [Ηλεκτρονικός σύνδεσμος] <https://www.forbes.com/sites/danielnewman/2016/08/30/top-10-trends-for-digital-transformation-in-2017/#5f37dfe647a5> [Ανακτήθηκε 06-12-2020].
- Oku K., Hattori F., Kawagoe K., (2015), “Tweet-mapping method for tourist spots based on now-tweets and spot-photos”, *Procedia Computer Science*, vol. 60, Pages 1318-1327.
- Okuyama K., Yanai K., (2013), “A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the web”, In *The Era of interactive media*, Pages 657-670, New York: Springer.

- Önder I., (2017), “*Classifying multi-destination trips in Austria with big data*”, *Tourism Management Perspectives*, vol. 21, Pages 54-58.
- O'Reilly T., (2005), "What is Web 2.0", [Ηλεκτρονικός σύνδεσμος] <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html> [Ανακτήθηκε 15-12-2020].
- O'Reilly T., (2007), “*What is web 2.0: Design patterns and business models for the next generation of software*”, *Communications & strategies*, vol. 1, Page 17.
- Page L., Brin S., Motwani R., Winograd T., (1999), "*The PageRank Citation Ranking: Bringing Order to the Web*", Technical Report, Stanford InfoLab.
- Pan B., MacLaurin T., Crotts J. C., (2007), “*Travel Blogs and the Implications for Destination Marketing*”, *Journal of Travel Research*, vol. 46(1), Pages 35–45, [Ηλεκτρονικός σύνδεσμος] <https://journals.sagepub.com/doi/10.1177/0047287507302378> [Ανακτήθηκε 15-12-2020].
- Philander K., Zhong Y., (2016), “*Twitter sentiment analysis: Capturing sentiment from integrated resort tweets*”, *International Journal of Hospitality Management*, vol. 55(2016), Pages 16-24.
- Plog S., (2001), “*Why destination areas rise and fall in popularity: an update of a Cornell Quarterly classic*”, *Cornell Hotel Restaur. Adm. Q.* 42, Pages 13–24.
- Popescu A., Grefenstette G., Moëllic P. A., (2009), “*Mining tourist information from user-supplied collections*”, In *Proceedings of the 18th ACM conference on information and knowledge management*, Paris, France.
- Postma A., (2003), “*Quality of Life, competing value perspectives in leisure and tourism*”, Conference: ATLAS 10th International Conference, “Quality of Life, competing value perspectives in leisure and tourism” At: Leeuwarden, Netherlands.
- Pries K. H., Dunnigan R., (2015), “*Big Data Analytics: A Practical Guide for Managers*”.
- Pyle D., (1999), “*Data Preparation for Data Mining*”, Morgan Kaufmann Publishers, California.
- Quack T., Leibe B., Van Gool L., (2008), “*World-scale mining of objects and events from community photo collections*”, In *Proceedings of the 2008 international conference on content-based image and video retrieval*, Niagara falls, Canada.
- Racherla P., Friske W., (2012), “*Perceived ‘usefulness’ of online consumer reviews: An exploratory investigation across three services categories*”, *Electronic Commerce Research and Applications*, vol. 11(6), Pages 548-559.
- Rinusha I., Sanjeeva S., (2013), “*Big data solution for Sri Lankan development: A case study from travel and tourism*”, Published Computer Science International Conference on Advances in ICT for Emerging Regions (ICTer), [Ηλεκτρονικός σύνδεσμος] <https://ieeexplore.ieee.org/document/6761180> [Ανακτήθηκε 15-12-2020].
- Sang J. L., Keng S., (2001), “*A review of data mining techniques. Industrial Management & Data Systems*”, 101/1, Pages 41-46.

- Schmidt E., Cohen J., (2014), “*Η Νέα Ψηφιακή Εποχή*”, Εκδόσεις Διάλογος.
- Schuckert M., Liu X., Law R., (2015a), “*A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently*”, *International Journal of Hospitality Management*, vol. 48, Pages 143-149.
- Schuckert M., Liu X., Law R., (2015b), “*Hospitality and tourism online reviews: Recent trends and future directions*”, *Journal of Travel & Tourism Marketing*, vol. 32(5), Pages 608-621.
- Shi Y., Serdyukov P., Hanjalic A., Larson M., (2011), “*Personalized landmark recommendation based on geotags from photo sharing sites*”, In *Proceedings of the 5th international AAAI conference on weblogs and social media*, Barcelona, Spain.
- Shoval N., Ahas R., (2016), “*The use of tracking technologies in tourism research: The first decade*”, *Tourism Geographies*, vol. 18(5), Pages 587-606.
- Simms A., (2012), “*Online user-generated content for travel planning - different for different kinds of trips?*”, *E-Review of Tourism Research*, vol. 10(3), Pages 76–85.
- Stangl B., Reino S., Wang D., Yuan Y., (2016), “*ENTER 2016 PhD Workshop Research Proposals*”.
- Sust V. O., Illera E. G., Berengué A. S., García R. G., Alonso M., Torres M., (2014), “*Big Data and Tourism: New indicators for tourism management*”.
- Taft R., Manasi V., Nadathur R. S., Narayanan S., Madden S., Stonebraker M., (2014), “*Genbase: A complex analytics genomics benchmark*”, *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, Pages 177-188.
- Tang L., Yu L., Liu F., Xu W., (2013), “*An integrated data characteristic testing scheme for complex time series data exploration*”, *International Journal of Information Technology and Decision Making*, vol. 12(3), Pages 491-521.
- Turban A., Liang S., (2007), “*Decision Support and Business Intelligence Systems*”, 8th Edition.
- Turner E., (2017), “*How Big Data Is Changing the Travel Industry*”.
- UNWTO, (2015), “*It’s not sustainable tourism...just tourism!*”, [Ηλεκτρονικός σύνδεσμος] <https://medium.com/@WTTC/it-s-not-sustainable-tourism-just-tourism-e480e3e1f3fe#.i2hn7 1i9t> [Ανακτήθηκε 06-12-2020].
- UNWTO, (2016), “*World Tourism Barometer*”, vol. 14, Advance Release.
- Varian H. R., (2014), “*Big Data: New Tricks for Econometrics*”, *Journal of Economic Perspectives*, vol. 28, Pages 3-27.
- Vu H. Q., Li G., Law R., Ye B. H., (2015), “*Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos*”, *Tourism Management*, vol. 46, Pages 222-232.
- Wang N., (1999), “*Rethinking authenticity in tourism experience*”, *Annals of tourism research*, vol. 26(2), Pages 349-370.

- Wikipedia, “*Το σύμπαν τεχνολογιών του Web 2.0*”, [Ηλεκτρονικός σύνδεσμος] [https://en.wikipedia.org/wiki/Web\\_2.0#/media/File:Web\\_2.0\\_Map.svg](https://en.wikipedia.org/wiki/Web_2.0#/media/File:Web_2.0_Map.svg) [Ανακτήθηκε 15-12-2020].
- Williams R., Wiele T., Iwaarden J., Eldridge S., (2010), “*The importance of user-generated content: the case of hotels*”, The TQM Journal, vol. 22(2), Pages 117–128, [Ηλεκτρονικός σύνδεσμος] <https://www.emerald.com/insight/content/doi/10.1108/17542731011024246/full/html> [Ανακτήθηκε 15-12-2020].
- Xiang Z., Gretzel U., (2010), “*Role of social media in online travel information search*”, Tourism Management, vol. 31(2), Pages 179–188, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/abs/pii/S0261517709000387?via%3Dihub> [Ανακτήθηκε 15-12-2020].
- Xiang Z., Schwartz Z., Gerdes J. H., Uysal M., (2015), “*What can big data and text analytics tell us about hotel guest experience and satisfaction?*”, International Journal of Hospitality Management, vol. 44(4), Pages 120–130.
- Xiang Z., Du Q., Ma Y., Fan W., (2017), “*A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism*”, Tourism Management, vol. 58, Pages 51-65.
- Xie K. L., Zhang Z., (2014), “*The business value of online consumer reviews and management response to hotel performance*”, International Journal of Hospitality Management, vol. 43, Pages 1-12.
- Xu H., Yuan H., Ma B., Qian Y., (2015), “*Where to go and what to play: Towards summarizing popular information from massive tourism blogs*”, Journal of Information Science, vol. 41(6), Pages 830-854.
- Xu X., Li Y., (2016), “*The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach*”, International Journal of Hospitality Management, vol. 55, Pages 57-69.
- Yang X., Pan B., Evans J. A., Lv B., (2015), “*Forecasting Chinese tourist volume with search engine data*”, Tourism Management, vol. 46, Pages 386–397.
- Ye Q., Law R., Gu B., (2009), “*The impact of online user reviews on hotel room sales*”. International Journal of Hospitality Management, vol. 28(1), Pages 180–182.
- Ye Q., Law R., Gu B., Chen W., (2011), “*The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings*”. Computers in Human Behavior, vol. 27(2), Pages 634–639.
- Youtube, (2020), [Ηλεκτρονικός σύνδεσμος] <https://blog.youtube/press/> [Ανακτήθηκε 15-12-2020].
- Zemla M., (2008), “*The product quality of Polish ski-resorts: A case study of Silesian skiers' requirements, satisfaction and complaints*”, Journal of Tourism.
- Zhang Z., Ye Q., Law R., Li Y., (2010), “*The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews*”, International Journal of Hospitality Management, vol. 29(4), Pages 694-700.

- Zhang Y., Cole S. T., (2016), “*Dimensions of lodging guest satisfaction among guests with mobility challenges: A mixed-method analysis of web-based texts*”, *Tourism Management*, vol. 53, Pages 13-27.
- Zhang F., Song Y., Cai W., Liu S., Pujol S., Feng D. D., (2016a), “*Pairwise latent semantic association for similarity computation in medical imaging*”, *IEEE Transactions on Biomedical Engineering*, vol. 63(5), Pages 1058–1069.
- Zhang Z., Zhang Z., Yang Y., (2016b), “*The power of expert identity: How web site recognized expert reviews influence travelers' online rating behavior*”, *Tourism Management*, vol. 55, Pages 15–24.
- Zhong N., Li Y., Wu S. T., (2012), “*Effective pattern discovery for text mining*”, *IEEE transactions on knowledge and data engineering*, vol. 24(1), Pages 30-44.
- Zhou L., (2014), “*Online rural destination images: Tourism and rurality*”, *J. Destin. Mark. Manag.* 3, Pages 227–240, [Ηλεκτρονικός σύνδεσμος] <https://www.sciencedirect.com/science/article/pii/S2212571X14000195?via%3Dihub> [Ανακτήθηκε 15-12-2020].
- Zhou X., Xu C., Kimmons B., (2015), “*Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform*”, *Computers, Environment and Urban Systems*, vol. 54, Pages 144-153.