
ΠΡΟΛΟΓΟΣ	11
-----------------	-----------

ΚΕΦΑΛΑΙΟ 1	
ΕΙΣΑΓΩΓΗ	13

ΚΕΦΑΛΑΙΟ 2	
ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΜΕΤΡΗΣΕΙΣ	17

2.1	Σύντομη εισαγωγή στην εξαγωγή και ανάλυση κειμένου στην Python	17
2.2	Ανάγνωση και Εξαγωγή Κειμένου από Αρχεία	19
2.2.1	Ανάγνωση Απλών Αρχείων Κειμένου (.txt)	19
2.2.2	Ανάγνωση Αρχείων PDF	21
2.2.3	Σύγκριση εξαγωγής κειμένου από αρχεία .txt και .pdf	22
2.2.4	Εφαρμογή σε Πραγματικό Παράδειγμα: Ανάλυση Αναλυτικών Προγραμμάτων Σπουδών στο περιβάλλον Google Colab	23
2.2.5	Εξαγωγή κειμένου των αρχείων	25
2.3	Υπολογισμός βασικών μετρήσεων κειμένου	26
2.3.1	Εμφάνιση και σύγκριση μετρήσεων	28
2.4	Ανάλυση Συχνοτήτων Λέξεων	30
2.5	Προεπεξεργασία (preprocessing)	32
2.6	Σύγκριση πριν και μετά την προεπεξεργασία	34

ΚΕΦΑΛΑΙΟ 3	
ΒΑΣΙΚΕΣ ΤΕΧΝΙΚΕΣ ΕΦΓ	37

3.1	Οπτικοποίηση δεδομένων	37
3.1.1	Ραβδογράμματα συχνότητας	37
3.1.2	Συννεφόμελα (Word Clouds)	39
3.2	N-γράμματα	42
3.3	Cosine similarity (ομοιότητα συνημιτόνου)	46
3.4	Κατάτμηση Κειμένου (Tokenization) με τη βιβλιοθήκη spaCy	49
3.5	Γραμματική Ανάλυση (POS Tagging) με το spaCy	50

ΚΕΦΑΛΑΙΟ 4

ΑΝΑΓΝΩΡΙΣΗ ΟΝΤΟΤΗΤΩΝ: ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ **53**

4.1	Εισαγωγή	53
4.2	Αναγνώριση Οντοτήτων με το NLTK	55
4.3	Αναγνώριση Οντοτήτων με το Flair	58
4.4	Σύγκριση των μεθόδων NER: NLTK vs Flair	61

ΚΕΦΑΛΑΙΟ 5

ΕΝΤΟΠΙΣΜΟΣ ΘΕΜΑΤΙΚΩΝ ΠΕΔΙΩΝ **63**

5.1	Εισαγωγή	63
5.2	Διευκρινίσεις σχετικά με τα θεματικά πεδία	64
5.3	Topic Modeling στην πράξη	65
5.4	Αξιολόγηση: Perplexity και Coherence	68

ΚΕΦΑΛΑΙΟ 6

ΑΝΑΛΥΣΗ ΣΥΝΑΙΣΘΗΜΑΤΟΣ **69**

6.1	Εισαγωγή	69
6.2	Το πρόβλημα της Ανάλυσης Συναισθήματος	72
6.3	Ανάλυση συναισθήματος στην πράξη I: αλγόριθμοι	103
6.3.1	Naive Bayes	103
6.3.2	N-γράμματα μοντέλα	105
6.3.3	Support Vector Machines	105
6.3.4	Τεχνητά Νευρωνικά Δίκτυα	106
6.3.4.1	Νευρωνικά δίκτυα συνέλιξης και επαναλαμβανόμενα νευρωνικά δίκτυα	109
6.3.4.2	Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης	109
6.3.5	Το σχήμα στάθμισης tf-idf	110
6.3.6	Λεξικές ενθέσεις	111
6.3.6.1	GloVe	111
6.3.6.2	FastText	112
6.3.7	BERT	113
6.4	Ανάλυση συναισθήματος στην πράξη II: αξιολόγηση	114
6.5	Ανάλυση συναισθήματος στην πράξη III: ο κώδικας	115
6.5.1	Ανάλυση συναισθήματος με το VADER	115
6.5.2	Εποπτευόμενη ανάλυση συναισθήματος	117

ΚΕΦΑΛΑΙΟ 7	
ΕΠΕΞΕΡΓΑΣΙΑ ΚΕΙΜΕΝΟΥ ΣΤΙΣ	
ΚΛΑΣΙΚΕΣ ΓΛΩΣΣΕΣ	121

ΚΕΦΑΛΑΙΟ 8	
ΗΛΕΚΤΡΟΝΙΚΗ ΛΕΞΙΚΟΓΡΑΦΙΑ	135

8.1	Εισαγωγή	135
8.2	Θετικά και αρνητικά ηλεκτρονικής λεξικογραφίας	136
8.3	Ηλεκτρονική λεξικογραφία στην πράξη	141

ΑΝΤΙ ΕΠΙΛΟΓΟΥ	143
---------------	------------

ΒΙΒΛΙΟΓΡΑΦΙΑ	145
--------------	------------
