

simakis@ceid.upatras.gr

17 Φεβρουαρίου 2017

- 1 Εισαγωγή
- 2 Βασικές Έννοιες
- 3 Εργαλεία Υλοποίησης
- 4 Μεθοδολογία
- 5 Συλλογή Δεδομένων
- 6 Προεπεξεργασία/Επισημείωση Δεδομένων
- 7 Εξαγωγή Χαρακτηριστικών
- 8 Επιλογή Χαρακτηριστικών
- 9 Αποτελέσματα Κατηγοριοποίησης
- 10 Συμπεράσματα

. . .
.
.

-
-
-

(Text Categorization)

- (Document Indexing)
-

- (Authorship Attribution)
-

Python

- NLTK (Natural Language ToolKit)
- WEKA (Waikato Environment for Knowledge Analysis)

:

1

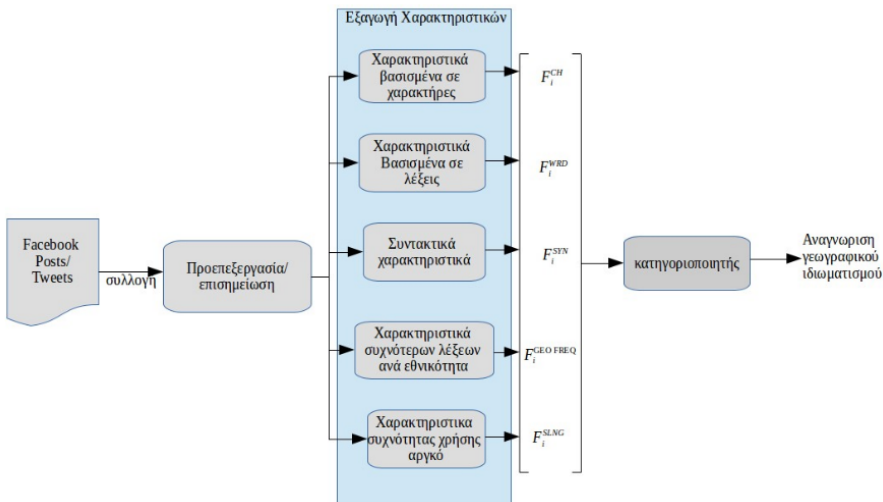
2 /

3

4

5

6



Facebook Twitter

- Facepager:
- 252.112 , 357
- 34.969.115 5.372.512

id							
1		F/M	A/B/C/ D/E/F*	> 14	Facebook/ Twitter	US/CAN/ UK/AUS/ NNS	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- (46)

- (8)

- (13)

- (9)

-

-

(Feature Selection)

/ ReliefF score	
1	0.00329917
2	0.00263219
3	0.00190155
4	0.00181929
5	0.00175353
6	0.00149185
7	0.00132278
8	0.00108097
9	0.00095305
10	0.00086441

Πίνακας: 10

	top10	top20	top30	top40	top50	top60	all
J48	52.28	53.88	52.24	51.23	50.31	51.26	51.30
MLP	47.17	51.16	51.20	51.11	50.07	49.36	45.01
RandomTree	55.58	55.20	50.66	46.01	41.53	40.62	40.52
REPTree	50.45	53.18	53.22	52.87	52.01	52.87	52.88
RBFNetwork	46.70	49.30	48.89	48.00	48.53	47.91	48.44
Bagging	57.96	61.87	59.61	58.34	57.50	58.34	58.35
Boosting	46.71	47.62	47.62	47.62	47.78	47.62	47.62
IBk	56.34	49.08	46.52	41.57	39.08	39.05	38.88

Πίνακας:

ῥεατιε ὁμμονς , 4.01 , .

