

ΑΝΑΔΕΙΞΕΤΟ: Εργαλεία Ανάδειξης Τουριστικής Εμπειρίας

T1EDK-02474

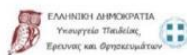
3^η Συνάντηση Εργασίας, 24 Ιανουαρίου 2019
Ενότητα Εργασίας 3 – Παραδοτέο 6



Ε.Κ. ΑΘΗΝΑ: Παύλος Εφραιμίδης, Γεώργιος Δροσάτος,
Σωτήρης Γυφτόπουλος, Γεώργιος Σταματελάτος, Ελένη
Μπριόλα



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο
Περιφερειακής Ανάπτυξης



ΕΠΑνεΚ 2014-2020
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ
ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ
ΚΑΙΝΟΤΟΜΙΑ



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Γενικά στοιχεία Ε.Ε.3

- Συμβαλλόμενοι φορείς υλοποίησης:
 - ΑΠΘ
 - Ε.Κ. ΑΘΗΝΑ
 - DOTSOFT
- Συντονιστής: **Ε.Κ. ΑΘΗΝΑ**
 - Technical manager: Γεώργιος Δροσάτος
- Περίοδος υλοποίησης:
 - Μ04 – Μ18 (~ 10/2018 – 12/2019)
- Παραδοτέο: **Π.6** (στο τέλος του Μ18)
 - Εργαλεία για υπόδειξη τουριστικής πληροφορίας από μοντέλα χρηστών, αξιολογήσεων και προσφερόμενων τιμών

Γενικά στοιχεία Ε.Ε.3 (updated)

- Συμβαλλόμενοι φορείς υλοποίησης:
 - ΑΠΘ
 - Ε.Κ. ΑΘΗΝΑ
 - DOTSOFT
- Συντονιστής: Ε.Κ. ΑΘΗΝΑ
 - Technical manager: Γεώργιος Δροσάτος
- Περίοδος υλοποίησης:
 - Μ04 – Μ18 (~ 10/2018 – 03/2020)
- Παραδοτέο: Π.6 (στο τέλος του Μ21)
 - Εργαλεία για υπόδειξη τουριστικής πληροφορίας από μοντέλα χρηστών, αξιολογήσεων και προσφερόμενων τιμών

Απαιτήσεις με βάση την περιγραφή του έργου

Αναμενόμενα αποτελέσματα της Ε.Ε.3

- Καινοτόμοι αλγόριθμοι και πιλοτική εφαρμογή για πρόβλεψη αξιολογήσεων/τιμών αναλύοντας μοντέλα χρηστών
- Καινοτόμοι αλγόριθμοι και μηχανισμοί προστασίας των προσωπικών δεδομένων των χρηστών

Επιμέρους ενέργειες εργασίας (με βάση την πρόταση)

- 1) Ανάπτυξη αλγορίθμων που υποδεικνύουν εξατομικευμένες τουριστικές συστάσεις, αξιοποιώντας μοντέλα χρηστών, χωρίς οι χρήστες να χάνουν τον έλεγχο των προσωπικών τους δεδομένων
[Υπεύθυνος φορέας: **Ε.Κ. ΑΘΗΝΑ**]
- 2) Ανάπτυξη τεχνικών προσεγγιστικής συναγωγής που επιτρέπουν πρόβλεψη σε πραγματικό χρόνο, οι οποίες θα εφαρμοστούν σε αξιολογήσεις χώρων αναψυχής και πρόβλεψη τιμών αεροπορικών εισιτηρίων ή τιμών ξενοδοχείων με κατάλληλα μοντέλα διάχυσης (π.χ. κίνηση Brown με χρονομεταβαλλόμενη παράμετρο ολίσθησης)
[Υπεύθυνος φορέας: **ΑΠΘ**]
- 3) Άντληση πληροφοριών τουριστικού ενδιαφέροντος από δίκτυα κοινωνικών μέσων, εφαρμόζοντας γραφοθεωρητική ανάλυση, αλγορίθμους και μοντέλα διάχυσης απόψεων (π.χ. DeGroot learning) στα κοινωνικά μέσα
[Υπεύθυνος φορέας: **Ε.Κ. ΑΘΗΝΑ**]
 - Η ανάλυση αυτή αποσκοπεί στην ανίχνευση τάσεων (trends, word-of-mouth) και στους μηχανισμούς που τις διαμορφώνουν, καθώς και στην αντιμετώπιση φαινομένων απάτης (πχ. fake reviews, fake news, alternative facts) που επηρεάζουν την αξιοπιστία των πληροφοριών

Ερευνητικές δραστηριότητες Ε.Ε.3 (1/3)

- Τεχνικές προσεγγιστικής συναγωγής
 - Το σημείο εκκίνησης είναι η κατανόηση του συνεργατικού φιλτραρίσματος Kalman για χρονο-εξελισσόμενες δυαδικές διαδικασίες για την πρόβλεψη αξιολογήσεων στο σύνολο δεδομένων Netflix και την πρόβλεψη τιμών ανοίγματος και κλεισίματος μετοχών 433 εταιρειών από το NASDAC και το χρηματιστήριο Νέας Υόρκης [Gultek 14]
 - Στα πλαίσια αυτής της εργασίας αναπτύχθηκαν παραλλαγές του βασικού αλγορίθμου και συγκρίθηκαν οι επιδόσεις τους με άλλες προσεγγιστικές συναγωγές (π.χ. Variational Bayesian Expectation Maximization) με ενθαρρυντικά αποτελέσματα
 - Τα μοντέλα αυτά, που επιτρέπουν πρόβλεψη σε πραγματικό χρόνο, θα εφαρμοστούν σε μια καινοτόμα προσέγγιση για αξιολογήσεις χώρων αναψυχής και πρόβλεψης τιμής αεροπορικών εισιτηρίων ή τιμών ξενοδοχείων με κατάλληλα μοντέλα διάχυσης (π.χ. κίνηση Brown με χρονομεταβαλλόμενη παράμετρο ολίσθησης)

Ερευνητικές δραστηριότητες Ε.Ε.3 (2/3)

- Άντληση πληροφοριών τουριστικού ενδιαφέροντος από κοινωνικά δίκτυα
 - Θα διερευνηθούν μηχανισμοί που επηρεάζουν ή καθορίζουν τη διαμόρφωση γνώμης, την τρέχουσα τάση (trend), την αξιοπιστία του δικτύου, και πως μπορεί αποτελεσματικά να προωθηθεί μια πληροφορία στο δίκτυο του κοινωνικού μέσου
 - Η ανάλυση στρατηγικών επιρροής της κοινής γνώμης θα αναδείξει την αντοχή του κοινωνικού δικτύου σε εξωτερικές επιρροές και τις πιθανές αλλοιώσεις που μπορούν να επιφέρουν στη “σοφία του πλήθους”
 - Η μεθοδολογία που θα χρησιμοποιηθεί έχει ως αφετηρία αποτελεί η εργασία [Golub 10] η οποία προτείνει μια ενδιαφέρουσα έννοια “σοφίας” (“wisdom”) ενός δικτύου ανθρώπων, που συσχετίζεται με τη διάχυση γνώμης του μοντέλου DeGroot Learning, και χαρακτηρίζουν τότε το δίκτυο αυτό εξασφαλίζει την επιθυμητή “σοφία” με βάση τις δομικές ιδιότητες του γραφήματος “balance” και “minimal dispersion”
 - Ένας άλλος ενδιαφέρον αλγόριθμος που θα διερευνηθεί είναι ο FRAUDAR για τον εντοπισμό “πλαστών” κόμβων σε δίκτυο follower-followee του Twitter αναλύοντας τη δομή του αντίστοιχου διμερούς γραφήματος [Hooi 16]

Ερευνητικές δραστηριότητες Ε.Ε.3 (3/3)

- Εξατομικευμένες τουριστικές συστάσεις με προστασία της ιδιωτικότητας
 - Για την προστασία της ιδιωτικότητας των χρηστών θα αξιοποιηθούν οι υπολογιστικές δυνατότητες και οι αισθητήρες έξυπνων συσκευών για την αυτόματη δημιουργία ενδιαφερόντων του χρήστη
 - Τα δεδομένα αυτά θα φυλάσσονται στη συσκευή του χρήστη όπου θα γίνεται επεξεργασία τους από κατάλληλους αλγορίθμους, όπως για παράδειγμα time-based clustering και POI matching για την αυτόματη εύρεση των ενδιαφερόντων του χρήστη
 - Θα διερευνηθεί η δυνατότητα εκτέλεσης υπολογισμών πάνω σε δεδομένα πολλών χρηστών, χωρίς όμως να αποκαλύπτονται τα δεδομένα κανενός χρήστη
 - Ενώ οι αντίστοιχοι αλγόριθμοι εκτελούνται πάνω σε προσωπικά δεδομένα, η έξοδος του αλγορίθμου είναι κάποιο συναθροιστικό αποτέλεσμα το οποίο δεν αποτελεί απειλή για την ιδιωτικότητα των χρηστών

**Υλοποιήσεις αλγορίθμων που θα
συμπεριληφθούν στο Παραδοτέο 6**

Αλγόριθμοι από Ε.Κ.Αθηνά

- **Αλγόριθμος 4:** Εξατομικευμένη σύσταση σημείων ενδιαφέροντος μέσω γραφοθεωρητικής ανάλυσης κοινωνικών δικτύων
- **Αλγόριθμος 5:** Ανίχνευση τάσεων τουριστικού ενδιαφέροντος σε κοινωνικά δίκτυα
- **Αλγόριθμος 6:** Ανάλυση μηχανισμού διαμόρφωσης τάσεων σε κοινωνικά δίκτυα
- **Αλγόριθμος 7:** Ανίχνευση αξιοπιστίας της παρεχόμενης τουριστικής πληροφορίας

Αλγόριθμος 1: Δυαδικό σωματιδιακό φίλτρο για πρόβλεψη τιμών

Περίληψη

- Χρήση του Δυαδικού Σωματιδιακού Φίλτρου (ΔΣΦ) ή Dyadic Particle Filter (DPF)
- Πρόβλεψη τιμών αεροπορικών εισιτηρίων
- Χρήση dataset από το Expedia (9 πτήσεις)
- Τα πρώτα αποτελέσματα δείχνουν ικανοποιητικά

Αλγόριθμος 2: Δυναμικό δυαδικό σωματιδιακό φίλτρο για πρόβλεψη τιμών

Περίληψη

- Χρήση του Δυναμικού Δυαδικού Σωματιδιακού Φίλτρου (ΔΔΣΦ) ή Dynamic Dyadic Particle Filter (DDPF)
- Πρόβλεψη τιμών αεροπορικών εισιτηρίων
- Χρήση dataset από το Expedia (27 πτήσεις)
- Σε σύγκριση με το ΔΣΦ και το νευρωνικό δίκτυο LSTM (Long Short term Memory), τα αποτελέσματα δείχνουν ότι σε 26 από τις 27 πτήσεις το ΔΔΣΦ υπερέχει των άλλων 2 τεχνικών

Αλγόριθμος 3: Συνεργατικό φίλτρο Kalman σταθμισμένων διατεταγμένων υποδειγμάτων για πρόβλεψη βαθμολογίας ξενοδοχείων

Περίληψη

- Χρήση του Συνεργατικού Φίλτρο Kalman Σταθμισμένων Διατεταγμένων Υποδειγμάτων (ΣΣΦΚ) ή Weighted Ordered Probit Collaborative Kalman Filter (WCKF)
- Πρόβλεψη βαθμολογιών ξενοδοχείων στο χρόνο
- Χρήση dataset από το Tripadvisor (16 Ελληνικοί προορισμοί)
- Σε σύγκριση με το απλό Συνεργατικό Φίλτρο Kalman (ΣΦΚ), το ΣΣΦΚ πραγματοποιεί καλύτερη πρόβλεψη στο χρόνο βαθμολογιών ξενοδοχείων και στις 16 προορισμούς

Αλγόριθμος 4: Εξατομικευμένη σύσταση σημείων ενδιαφέροντος μέσω γραφοθεωρητικής ανάλυσης κοινωνικών δικτύων

Περίληψη

- Γραφοθεωρητική ανάλυση λιστών POIs που έχουν δημιουργηθεί από χρήστες και ομαδοποιούν παρόμοια σημεία ενδιαφέροντος
- Εξατομικευμένη σύσταση σημείων ενδιαφέροντος με βάση το προφίλ των χρηστών
- Αμεσότητα και Ιδιωτικότητα:
 - Προεπεξεργασία λιστών στην υπηρεσία: Η σύσταση πραγματοποιείται σε πραγματικό χρόνο
 - Το προφίλ του χρήστη δεν φεύγει από τη συσκευή του
- Χρήση dataset από το Foursquare (9 περιοχές)
 - 39,780 POIs
 - 16,142 λίστες

Αλγόριθμος 5: Ανίχνευση τάσεων τουριστικού ενδιαφέροντος σε κοινωνικά δίκτυα

Περίληψη

- Χρήση Λανθάνουσας Κατανομής Dirichlet (LDA) για την ανίχνευση θεμάτων βάση κειμένου και την εξέλιξη τους στο χρόνο
- Ανίχνευση τάσεων σε θέματα τουριστικού ενδιαφέροντος
- Χρήση dataset από το Twitter
 - ~ 20,000 tweets τουριστικού ενδιαφέροντος
 - Δημοσιεύτηκαν τον Ελλαδικό χώρο
- Η ανίχνευση των θεμάτων είναι unsupervised και απαιτείται μόνο ο αριθμός των θεμάτων που χρειάζεται να εντοπιστούν

Αλγόριθμος 6: Ανάλυση μηχανισμού διαμόρφωσης τάσεων σε κοινωνικά δίκτυα

Περίληψη

- Χρήση του αλγορίθμου HITS για την άντληση πληροφορίας επιρροής από γράφους
- Εντοπισμός των κόμβων σε κοινωνικά δίκτυα που ασκούν μεγάλη επιρροή (influencers)
- Χρήση dataset από το Twitter
 - Χρήστες που έκαναν retweet τουριστικό περιεχόμενο
 - 4674 κόμβοι (χρήστες)
 - 9352 ακμές επιρροής
- Η εφαρμογή του αλγορίθμου HITS στον γράφο επιρροής ανάδειξε 33 κόμβους με ενισχυμένες τιμές authority και μπορούν να χαρακτηριστούν ως κόμβοι επιρροής

Μηχανισμοί Διαμόρφωσης Τάσεων σε Κοινωνικά Δίκτυα και Τουρισμό (1/2)

- Τα Μέσα Κοινωνικής Δικτύωσης παίζουν σημαντικό ρόλο στη σύγχρονη εποχή στη διαμόρφωση τάσεων παγκοσμίως (π.χ., πολιτική, μόδα, τρόπος ζωής).
- Καθιέρωση του όρου **Influencer** στα κοινωνικά δίκτυα: χρήστες που επηρεάζουν μεγάλο πλήθος καταναλωτών με μέσω των δημοσιεύσεών τους!
 - ZOE SUGG NAMED HIGHEST-EARNING FEMALE SOCIAL MEDIA INFLUENCER UNDER 30

Αλγόριθμος 6: Ανάλυση μηχανισμού διαμόρφωσης τάσεων σε κοινωνικά δίκτυα

Μηχανισμοί Διαμόρφωσης Τάσεων σε Κοινωνικά Δίκτυα και Τουρισμό (2/2)

- Οι τουριστικές επιχειρήσεις μπορούν να ωφεληθούν πολλαπλά από την ανάλυση τάσεων στα κοινωνικά δίκτυα. Οι τουριστικές επιχειρήσεις μπορούν:
 - να «εναρμονιστούν» με τις προωθούμενες τάσεις από τους influencers.
 - να συνεργαστούν με influencers για την προώθηση των υπηρεσιών τους.
- Το αντικείμενο έχει μελετηθεί στα πλαίσια του τουριστικού πεδίου.

Αλγόριθμος 6: Ανάλυση μηχανισμού διαμόρφωσης τάσεων σε κοινωνικά δίκτυα

Απαιτήσεις Αλγόριθμου Ανάλυσης Μηχανισμού Διαμόρφωσης Τάσεων

- Εντοπισμός κόμβων με ισχυρή επιρροή στη διαμόρφωση τουριστικών τάσεων.
- Δυνατότητα εστίασης σε συγκεκριμένη γεωγραφική περιοχή.
- Δυνατότητα παρακολούθησης της πρόσφατης δραστηριότητας στα κοινωνικά δίκτυα και άμεση ανίχνευση αλλαγής στο μηχανισμό διαμόρφωσης τάσεων.

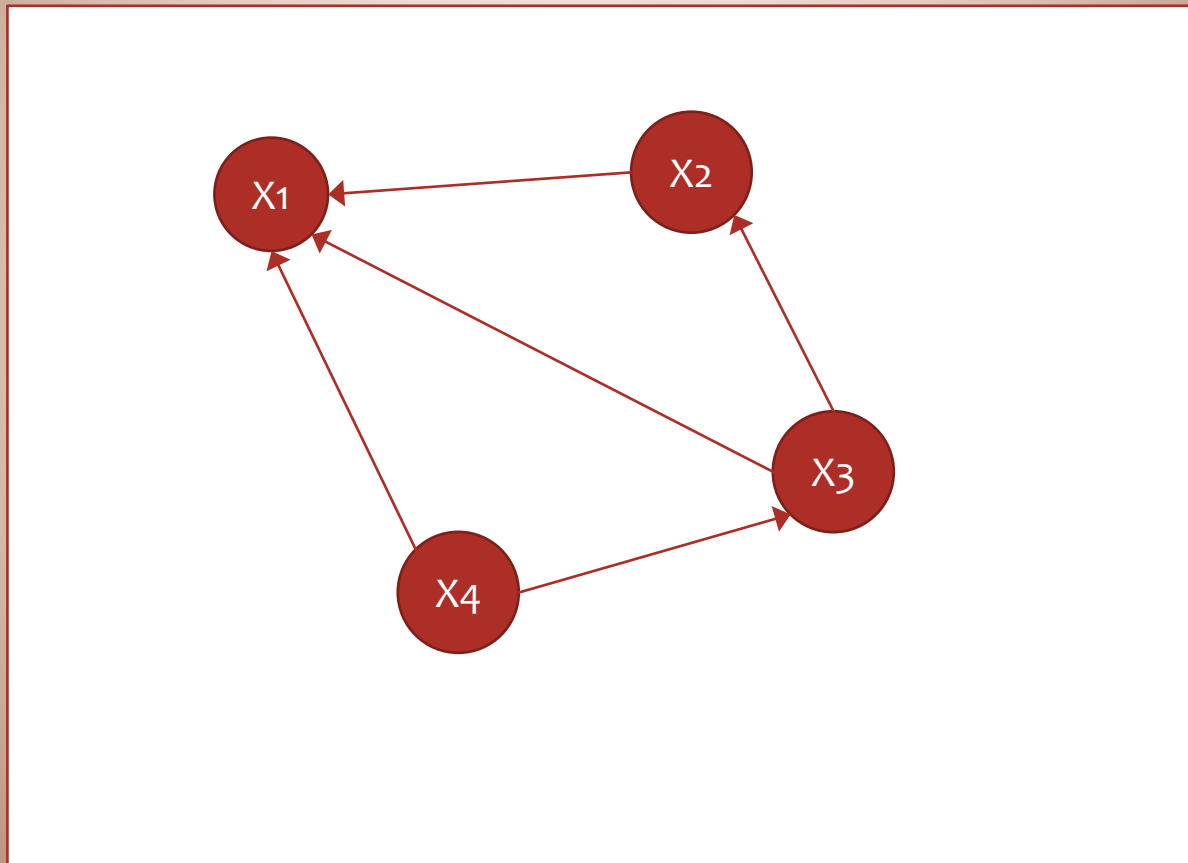
Αλγόριθμος 6: Ανάλυση μηχανισμού διαμόρφωσης τάσεων σε κοινωνικά δίκτυα

Θεωρητική Περιγραφή Αλγορίθμου (1/2)

- Βασική ιδέα: Οι χρήστες αναδημοσιεύουν μηνύματα με τα οποία συμφωνούν (ενστερνίζονται τις απόψεις που περιέχονται στα μηνύματα).
- Το πλήθος αναδημοσιεύσεων μηνυμάτων τουριστικού ενδιαφέροντος είναι η ισχυρότερη ένδειξη για την επιρροή που ασκεί ο συντάκτης ενός μηνύματος στους άλλους χρήστες.
- Αποτύπωση των αναδημοσιεύσεων με τη χρήση γράφων και αξιοποίηση αλγορίθμων ανάλυσης για την ανάδειξη των κόμβων (χρηστών) με τη μεγαλύτερη επιρροή.

Αλγόριθμος 6: Ανάλυση μηχανισμού διαμόρφωσης τάσεων σε κοινωνικά δίκτυα

Θεωρητική Περιγραφή Αλγορίθμου (2/2)



Αλγόριθμος 6: Ανάλυση μηχανισμού διαμόρφωσης τάσεων σε κοινωνικά δίκτυα

Εφαρμογή του Αλγορίθμου (1/2)

- Μέσο Κοινωνικής Δικτύωσης: Twitter
 - Παγκοσμίως γνωστό και ευρέως διαδεδομένο δίκτυο
 - Δημόσια πρόσβαση στις δημοσιεύσεις των χρηστών
 - Δομημένοι μηχανισμοί για την άντληση δεδομένων από το API
- Γράφημα Αναδημοσιεύσεων Tweets τουριστικού ενδιαφέροντος:
 - 4674 κόμβοι (χρήστες)
 - 9352 ακμές
- Εφαρμογή Αλγόριθμου HITS:
 - 33 κόμβοι με αυξημένη τιμή authority → Influencers

Αλγόριθμος 6: Ανάλυση μηχανισμού διαμόρφωσης τάσεων σε κοινωνικά δίκτυα

Εφαρμογή του Αλγορίθμου (2/2)

Rank	User	Rank	User	Rank	User
#1	PapagalosE	#12	Marina_Moscha	#23	markellashar
#2	tsoupras	#13	besttravelbook	#24	christosnezos
#3	NTINA74_	#14	agelaw_7	#25	alexkatsaiti
#4	eugeniarigopou2	#15	jmktraveller	#26	jannymrnda
#5	Laurapianist1	#16	DiscoverGRcom	#27	Athanas44229342
#6	Cyber990M	#17	BASO39604252	#28	GrecianGirly
#7	Storefortravels	#18	gefsdyeus	#29	luxury__travel
#8	Adamant19482884	#19	ArgirosKonstant	#30	USAmbPyatt
#9	giannis2221963	#20	niktop	#31	Taygetos_Katia
#10	MBerberi	#21	AlexMaragos	#32	JohnTheGreek01
#11	AstraeusHoliday	#22	Tabula_Rasa_gr	#33	clickeric

Αλγόριθμος 6: Ανάλυση μηχανισμού διαμόρφωσης τάσεων σε κοινωνικά δίκτυα

Αλγόριθμος 7: Ανίχνευση αξιοπιστίας της παρεχόμενης τουριστικής πληροφορίας

Περίληψη

- Χρήση τεχνικών μηχανικής μάθησης και αλγορίθμων classification
- Ανίχνευση ψευδών/μεροληπτικών κριτικών για σημεία ενδιαφέροντος
- Χρήση dataset από το Yelp για σημεία ενδιαφέροντος στο Λονδίνο
- Η εφαρμογή των αλγορίθμων classification και μηχανικής μάθησης για την ανίχνευση ψευδών/μεροληπτικών κριτικών είναι σε πρώιμο στάδιο λόγω προβλημάτων στον εντοπισμό του κατάλληλου dataset ειδικά στον Ελλαδικό χώρο
- Αξιοποιήθηκαν δεδομένα από το ΑΠΘ dataset.

Τέλος