



e-ΧΝΗΛΑΤΗΣ

εξυπνη ΧωροχροΝικΗ ολοκλήρωση περιβαλλοντικών πολιτιστικών και
τουριστικών διαδρομών περιήγΗΣης
«ΕΡΕΥΝΩ – ΔΗΜΙΟΥΡΓΩ – ΚΑΙΝΟΤΟΜΩ»

Π2.3.2

Τεχνικές Προσωποποίησης Εμπειρίας Χρήστη

Επίπεδο διάδοσης:	Δημόσιο
Ημερομηνία παράδοσης σύμφωνα με το έργο:	Μήνας 45, 5 Ιουνίου 2022
Πραγματική ημερομηνία παράδοσης:	Μήνας 45, 5 Ιουνίου 2022
Ενότητα εργασίας:	ΕΕ2: Ανάπτυξη τεχνολογιών χωρο-χρονικής διασύνδεσης πολλαπλών χώρων και γεγονότων περιβαλλοντικού, πολιτιστικού και τουριστικού ενδιαφέροντος
Υποενότητα:	ΥΕ2.4 Τεχνικές προσωποποίησης εμπειρίας χρήστη
Είδος:	Έκθεση
Κατάσταση έγκρισης:	Εγκεκριμένο
Έκδοση:	2.0
Αριθμός σελίδων:	34
Όνομα αρχείου:	Π2.3.2-Personalization_v1.pdf
Περίληψη	

Στην αναφορά πραγματοποιείται εκτενής περιγραφή των τεχνικών που χρησιμοποιήθηκαν για να προσφέρουν υψηλό επίπεδο προσωποποίησης στην εμπειρία του χρήστη. Στο πλαίσιο αυτό αναπτύχθηκαν κατάλληλες τεχνικές αναζήτησης ή επιλογής σημείων ενδιαφέροντος, βάσει διαφορετικών κριτηρίων που θα επιτρέπουν στο σύστημα να προσωποποιεί τα αποτελέσματα της επιλογής σημείων ενδιαφέροντος βάσει των προσωπικών αναγκών και ενδιαφερόντων του χρήστη μια δεδομένη στιγμή.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο
Περιφερειακής Ανάπτυξης



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ
ΟΙΚΟΝΟΜΙΑΣ & ΑΝΑΠΤΥΞΗΣ
ΕΙΔΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΕΠΠΑ & ΤΣ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΕΠΑΝΕΚ



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ Παιδείας,
Ερευνας και Θρησκευμάτων



ΓΓΕΤ
ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ
ΕΡΕΥΝΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ



ΕΠΑΝΕΚ 2014-2020
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ
ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ
ΚΑΙΝΟΤΟΜΙΑ



ΕΣΠΑ
2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Ιστορικό

Έκδοση	Ημερομηνία	Αιτιολογία	Αναθεωρήθηκε από	Εγκρίθηκε από
0.1	13/11/2019	Βιβλιογραφική Ανασκόπηση	VARLab	VARLab
0.2	11/02/2020	Αρχιτεκτονική Συστήματος Εξατομικευμένων Προτάσεων	VARLab	VARLab
0.3	01/06/2020	Οντολογικό Μοντέλο	VARLab	VARLab
0.4	20/01/2021	Μοντέλο Χρήστη	VARLab	VARLab
0.8	04/10/2021	Συστήμα Εξατομικευμένων Προτάσεων, Συμπεράσματα	VARLab	VARLab
0.9	15/05/2022	Εσωτερική αναθεώρηση	VARLab	VARLab
1.0	05/06/2022	Τελική έκδοση	VARLab	VARLab

Λίστα συγγραφέων

Οργανισμός	Όνομα	Στοιχεία επικοινωνίας
ΕΚΕΤΑ	Ειρήνη Μητσοπούλου	emitsopou@iti.gr
ΕΚΕΤΑ	Αθανάσιος Πατενίδης	apatenidis@iti.gr
ΕΚΕΤΑ	Ιωάννης Παλιόκας	ipaliokas@iti.gr
ΕΚΕΤΑ	Κωνσταντίνος Βότης	kvotis@iti.gr

Περίληψη κυριότερων σημείων

Τα κυριότερα σημεία του παραδοτέου περιλαμβάνουν μια αρχική βιβλιογραφική ανασκόπηση συστημάτων παραγωγής εξατομικευμένων προτάσεων στον τομέα του τουρισμού καθώς και συστημάτων παραγωγής εξατομικευμένων προτάσεων με επίγνωση πλαισίου (Context-Aware Recommender Systems). Στη συνέχεια, αναλύεται το γενικό σύστημα παραγωγής προτάσεων της διαδικτυακής πλατφόρμας του e-ΧΝΗΛΑΤΗ και τα βήματα που ακολουθούνται για την επιλογή των κατάλληλων τοποθεσιών δυνητικού ενδιαφέροντος, που θα προταθούν στον χρήστη-τουρίστα, στην περιοχή της Βόρειας Ελλάδας.

Η διαχείριση της ετερογένειας, σε μη δομημένες πληροφορίες που προέρχονται από τον παγκόσμιο ιστό, γίνεται σύμφωνα με ένα ενοποιημένο οντολογικό μοντέλο. Η μηχανή παραγωγής συστάσεων ενσωματώνει μηχανισμούς συλλογιστικής που βασίζονται σε κανόνες, ενώ λαμβάνονται υπόψη ρητά καθορισμένοι περιορισμοί καθώς και παράμετροι χωροχρονικού και περιβαλλοντικού πλαισίου. Επιπρόσθετα, χρησιμοποιείται ένας συνδυασμός τεχνικών παραγωγής συστάσεων (φιλτράρισμα που βασίζεται στο περιεχόμενο και συνεργατικό φιλτράρισμα) που εκμεταλλεύεται ένα σύνολο μετρικών ομοιότητας και τεχνικών εξόρυξης δεδομένων για τη δημιουργία ταξινομημένων προτάσεων, προσαρμοσμένων στα δυναμικά χαρακτηριστικά των προφίλ των τουριστών. Τέλος, με κριτήριο τα αποτελέσματα ενός πειράματος αξιολόγησης της ποιότητας των παραγόμενων συστάσεων, προσδιορίζονται οι τιμές των βαρών της σταθμισμένης υβριδικής τεχνική που χρησιμοποιεί η μηχανή παραγωγής συστάσεων του e-ΧΝΗΛΑΤΗ.

Συντομογραφίες και ακρωνύμια

POI	Point of Interest
TRS	Travel Recommender System
LARS	Location-Aware Recommender System
CARS	Context-Aware Recommender System
<i>k</i>-NN	<i>k</i> -Nearest Neighbors
OWA	Ordered Weighted Averaging Aggregation
CF	Collaborative Filtering
CBF	Content Based Filtering
GPS	Global Positioning System
Max	Maximum
Min	Minimum
NDCG	Normalized Discounted Cumulative Gain
DCG	Discounted Cumulative Gain
IDCG	Ideal Discounted Cumulative Gain
KB	Knowledge Base
Η/Υ	Ηλεκτρονικός Υπολογιστής

Πίνακας περιεχομένων

1	ΕΙΣΑΓΩΓΗ	7
2	ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ	8
2.1	Συστήματα Παραγωγής Εξατομικευμένων Προτάσεων στον Τομέα του Τουρισμού	8
2.2	Συστήματα Παραγωγής Εξατομικευμένων Προτάσεων με Επίγνωση Πλαισίου	10
3	ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΕΞΑΤΟΜΙΚΕΥΜΕΝΩΝ ΠΡΟΤΑΣΕΩΝ	12
3.1	Γενική Αρχιτεκτονική	12
3.2	Οντολογικό Μοντέλο	14
3.3	Μοντέλο Χρήστη	15
3.3.1	Δημιουργία Δυναμικών Προφίλ Χρηστών	15
3.3.2	Δημιουργία Γειτόνων	18
3.4	Το Σύστημα Εξατομικευμένων Προτάσεων e-ΧΝΗΛΑΤΗΣ	18
3.4.1	Φιλτράρισμα που Βασίζεται στη Γνώση	19
3.4.2	Φιλτράρισμα που Βασίζεται στο Περιεχόμενο	20
3.4.3	Συνεργατικό Φιλτράρισμα	21
3.4.4	Υβριδική Προσέγγιση	22
3.4.5	Προσδιορισμός Βαρών Σταθμισμένης Υβριδικής Τεχνικής	24
4	ΣΥΜΠΕΡΑΣΜΑΤΑ	32
5	ΑΝΑΦΟΡΕΣ	33

1 ΕΙΣΑΓΩΓΗ

Σήμερα, υπάρχει πληθώρα πληροφοριών που είναι διαθέσιμες σε ιστοσελίδες, ταξιδιωτικούς οδηγούς και περιοδικά σχετικά με τουριστικά αξιοθέατα. Η μεγάλη πλειονότητα αυτών των πόρων αναφέρεται σε διακριτά σημεία ενδιαφέροντος (Points of Interest / POIs) που θεωρούνται δημοφιλή, χωρίς να λαμβάνονται υπόψη οι προτιμήσεις των τουριστών ή οι περιορισμοί που θέτει. Επιπλέον, τα μεγάλα δεδομένα (big data) δυσκολεύουν τους τουρίστες στην επεξεργασία και την αφομοίωση όλων των πιθανών επιλογών. Έτσι, η δημιουργία πιο εξελιγμένων λύσεων, η ενίσχυση της χρήσης τέτοιων πληροφοριών και η βοήθεια των χρηστών στη λήψη αποφάσεων με βάση τα χαρακτηριστικά και τους περιορισμούς που θέτουν, θεωρείται απαραίτητη.

Εκτεταμένη έρευνα έχει διεξαχθεί για τη βελτίωση των τουριστικών εμπειριών κατά την τελευταία δεκαετία [1][2], εν μέρει λόγω του γεγονότος ότι τα smartphones έχουν μετατραπεί σε ισχυρά υπολογιστικά εργαλεία που παρέχουν εξατομικευμένες συστάσεις, περιεχόμενο και υπηρεσίες, περιορίζοντας επομένως τις αρνητικές επιπτώσεις της υπερφόρτωσης πληροφοριών [3]. Η εξατομίκευση (personalization) βρήκε αρχικά επιτυχία στο ηλεκτρονικό εμπόριο, παρέχοντας συστάσεις για εμπορεύματα και παρέχοντας πληροφορίες στους καταναλωτές για τη διευκόλυνση των διαδικασιών λήψης αποφάσεων [4]. Αυτά τα συστήματα, κοινώς γνωστά ως συστήματα παραγωγής συστάσεων, βασίζονται στο φιλτράρισμα πληροφοριών προκειμένου να προβλέψουν τις προτιμήσεις των χρηστών και να προτείνουν σχετικό περιεχόμενο στους χρήστες [5] (π.χ. ταινίες, βιβλία, ειδήσεις και ερευνητικά άρθρα).

Στον τομέα του τουρισμού, τα συστήματα παραγωγής συστάσεων θα μπορούσαν να χρησιμεύσουν ως πολύτιμα εργαλεία κατά την προετοιμασία ενός ταξιδιού μεταξύ διαφορετικών προορισμών, πολλών POI ή δραστηριοτήτων. Θα μπορούσαν να βοηθήσουν τους τουρίστες που αναζητούν ένα ταξίδι ή υπηρεσίες που ταιριάζουν με τα ατομικά γούστα και τους περιορισμούς τους, εξοικονομώντας πολύτιμο χρόνο και ενέργεια [6][7] ή ακόμη και να δημιουργήσουν συστάσεις που σκοπεύουν να ικανοποιήσουν μια ομάδα τουριστών συνολικά [8][9]. Οι τεχνικές εξατομίκευσης στα Travel Recommender Systems (TRS) [10] στοχεύουν στο να λάβουν υπόψη το πλαίσιο με την έννοια των ιδιαιτεροτήτων, των προτιμήσεων και του ιστορικού πλοήγησης των χρηστών, προκειμένου να παρέχουν σχετικά αποτελέσματα από ένα συντριπτικό σύνολο επιλογών. Συνήθως, ένα TRS χρησιμοποιείται για τη δημιουργία μιας λίστας με τα πιο κατάλληλα POIs. Ορισμένα συστήματα χρησιμοποιούν τεχνητή νοημοσύνη, όπως διάφορους τύπους τεχνικών φιλτραρίσματος (δημογραφικό, βασισμένο στο περιεχόμενο, βασισμένο στη γνώση κ.λπ.), αλγόριθμους αυτόματης ομαδοποίησης και μεθοδολογίες που βασίζονται στην συλλογιστική (reasoning), όπως είναι οι οντολογίες [11].

2 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

Αυτή η ενότητα περιγράφει βασικές έννοιες σχετικά με τα συστήματα παραγωγής συστάσεων, με ιδιαίτερη εστίαση στον τομέα του τουρισμού.

2.1 Συστήματα Παραγωγής Εξατομικευμένων Προτάσεων στον Τομέα του Τουρισμού

Ένα σύστημα συστάσεων είναι ένα εργαλείο εξατομίκευσης που προσπαθεί να παρέχει στους ανθρώπους ένα σύνολο αντικειμένων που ικανοποιούν καλύτερα τις ατομικές τους ανάγκες [12]. Στον τομέα του τουρισμού, τα TRS στοχεύουν στην αντιστοίχιση των χαρακτηριστικών των τουριστικών ειδών με τις προτιμήσεις των χρηστών. Τα υπάρχοντα συστήματα συστάσεων στον ηλεκτρονικό τουρισμό λαμβάνουν ρητά τις ανάγκες, τα μοναδικά χαρακτηριστικά και τις απαιτήσεις του χρήστη και στη συνέχεια προτείνουν προορισμούς, εκδηλώσεις, δραστηριότητες ή πλήρη τουριστικά πακέτα. Ο ταξιδιωτικός οδηγός City Trip Planner [13] είναι ένα τέτοιο παράδειγμα που καταγράφει το πλαίσιο του χρήστη, δηλαδή τα τοπικά ενδιαφέροντα και τους περιορισμούς ταξιδιού, μέσω ενός σύντομου ερωτηματολογίου, προκειμένου να προγραμματιστεί ένα τουριστικό ταξίδι. Ομοίως, στο [14], οι τουρίστες δηλώνουν άμεσα τους περιορισμούς τους ως έκφραση προτίμησης για να λάβουν πιο εξατομικευμένες υπηρεσίες ταξιδιωτικού σχεδιασμού. Ως επέκταση των προαναφερθέντων συστημάτων, τα συστήματα παραγωγής συστάσεων που προσπαθούν να συλλάβουν με έμμεσο τρόπο πρόσθετα δεδομένα για τις προτιμήσεις των τουριστών αναλύοντας τη διαδικτυακή τους δραστηριότητα, δεν μπορούν να παραλειφθούν. Για παράδειγμα, το σύστημα ταξιδιωτικών συστάσεων CRUMBS [15] προσφέρει στους χρήστες τη δυνατότητα να αξιολογούν προορισμούς που έχουν ήδη επισκεφτεί και επωφελούνται από αυτές τις αξιολογήσεις ενημερώνοντας σιωπηρά τις προτιμήσεις των χρηστών. Ομοίως, στην εφαρμογή e-Tourism [16], οι ενημερώσεις των προτιμήσεων των χρηστών γίνονται με δυναμικό τρόπο, μετά την ανάλυση της συνάφειας των αξιολογήσεων που λαμβάνονται ρητά από τον χρήστη.

Με βάση τις παραπάνω αναφορές, παρατηρείται ότι τα συστήματα συστάσεων επεξεργάζονται ένα ευρύ φάσμα δεδομένων πριν προβάλουν προτάσεις στους χρήστες. Τόσο η ποσότητα όσο και το είδος των πληροφοριών επηρεάζουν άμεσα την επάρκεια και την ορθότητα των συστάσεων. Με τα χρόνια, ανάλογα με τα υπάρχοντα πληροφοριακά δεδομένα και τις τεχνικές που χρησιμοποιούνται για τη σύσταση στοιχείων, στη βιβλιογραφία έχουν προταθεί διαφορετικές προσεγγίσεις για την ανάπτυξη συστημάτων παραγωγής συστάσεων [17][18]. Αυτές μπορούν να ταξινομηθούν ως: φιλτράρισμα με βάση τα δημογραφικά χαρακτηριστικά, συνεργατικό φιλτράρισμα, φιλτράρισμα που βασίζεται στο περιεχόμενο και φιλτράρισμα που βασίζεται στη γνώση. Η κύρια ιδέα της πρώτης προσέγγισης είναι η διαίρεση των χρηστών σε κατηγορίες λαμβάνοντας υπόψη τα

προσωπικά χαρακτηριστικά. Ως εκ τούτου, δίνονται συστάσεις ανάλογα με την κατηγορία στην οποία ταξινομούνται οι χρήστες [17]. Το συνεργατικό φιλτράρισμα είναι μια τεχνική που εφαρμόζεται ευρέως όταν πρόκειται για συστήματα παραγωγής συστάσεων, η οποία δημιουργεί συστάσεις για έναν συγκεκριμένο χρήστη με βάση τις αξιολογήσεις άλλων χρηστών. Συνήθως, αυτά τα συστήματα εντοπίζουν ομοιότητες μεταξύ των χρηστών, όσον αφορά τη συμπεριφορά αξιολόγησης τους και δημιουργούν νέες συστάσεις που βασίζονται σε συγκρίσεις μεταξύ των χρηστών [17]. Το φιλτράρισμα που βασίζεται στο περιεχόμενο αναλύει το περιεχόμενο των αντικειμένων προκειμένου να εντοπίσει εκείνα που παρουσιάζουν ιδιαίτερο ενδιαφέρον για τον χρήστη [17]. Τέλος, το φιλτράρισμα που βασίζεται στη γνώση επιχειρεί να δημιουργήσει προτάσεις χρησιμοποιώντας συγκεκριμένη γνώση για να ανακαλύψει τα αντικείμενα που ταιριάζουν καλύτερα στις ανάγκες και τις απαιτήσεις του χρήστη [17].

Οι παραδοσιακές λύσεις συστάσεων συχνά υποφέρουν από αρκετές αδυναμίες και περιορισμούς. Για παράδειγμα, η ακρίβεια ενός συστήματος που χρησιμοποιεί συνεργατικό φιλτράρισμα είναι πολύ ευαίσθητη ως προς το πλήθος των αξιολογημένων αντικειμένων που σχετίζονται με έναν συγκεκριμένο χρήστη. Συγκεκριμένα, η τεχνική συνεργατικού φιλτραρίσματος μπορεί να υποφέρει από το λεγόμενο πρόβλημα «ψυχρής εκκίνησης (cold start)» [19], επομένως απαιτείται μεγάλος όγκος ιστορικού αξιολογήσεων για να ληφθεί μια ακριβής σύσταση. Επιπλέον, οι αλγόριθμοι συστάσεων που βασίζονται στο περιεχόμενο συχνά υποφέρουν από το πρόβλημα της «ανάλυσης περιορισμένου περιεχομένου» [20], ελλείπει διακριτών μεταδεδομένων στη λίστα των υποψήφιων αντικειμένων προς σύσταση. Δεδομένου ότι η αποκλειστική χρήση μιας τεχνικής φιλτραρίσματος μπορεί να αποτύχει όταν πρόκειται για προτάσεις που αφορούν σύνθετα αντικείμενα, οι υβριδικές λύσεις που συνδυάζουν πολλούς αλγόριθμους παραγωγής συστάσεων είναι πολύ πιθανό να επιτύχουν και να ξεπεράσουν τις αδυναμίες των παραδοσιακών αλγορίθμων. Οι υβριδικές προσεγγίσεις αξιοποιούν πλήρως τα πλεονεκτήματα των διαφόρων τεχνολογιών για την άμβλυνση των περιορισμών μιας συγκεκριμένης τεχνολογίας και τη βελτίωση της συνιστάμενης απόδοσης και ακρίβειας [17].

Με βάση τα παραπάνω, διάφορες εργασίες έχουν χρησιμοποιήσει υβριδικές προσεγγίσεις για να μοντελοποιήσουν διαφορετικές παραλλαγές του προβλήματος της παραγωγής συστάσεων στον τουρισμό τα τελευταία χρόνια. Ένα σύστημα προγραμματισμού περιηγήσεων που αναπτύχθηκε για να βοηθήσει τους τουρίστες να βρουν ένα εξατομικευμένο πρόγραμμα περιήγησης σε μια πόλη της Πορτογαλίας, παρουσιάζεται στο [21]. Η προτεινόμενη υβριδική προσέγγιση παραγωγής συστάσεων χρησιμοποιεί συνεργατικό φιλτράρισμα και φιλτράρισμα που βασίζεται στο περιεχόμενο, συνδυάζοντας μια τεχνική ομαδοποίησης και έναν αλγόριθμο ταξινόμησης για τη βελτίωση της ποιότητας των προτάσεων. Το Buddy@Move [22] είναι ένα άλλο υβριδικό TRS που παρέχει εξατομικευμένες προτάσεις σε έναν ταξιδιώτη εν κινήσει, χρησιμοποιώντας μεμονωμένες

δημογραφικές πληροφορίες και συνδυάζοντας τις δυνατότητες τριών τεχνικών φιλτραρίσματος (συνεργατικό φιλτράρισμα, φιλτράρισμα που βασίζεται στο περιεχόμενο και δημογραφικό φιλτράρισμα). Η web και mobile εφαρμογή e-Turist [11], η οποία αναπτύχθηκε αρχικά για τουριστικές περιοχές στην περιοχή της Σλοβενίας, χρησιμοποιεί μια υβριδική προσέγγιση παραγωγής συστάσεων που αποτελείται από ένα συνδυασμό συνεργατικού φιλτραρίσματος, φιλτραρίσματος που βασίζεται σε περιορισμούς και φιλτραρίσματος που βασίζεται στο περιεχόμενο για την επιλογή POI σύμφωνα με τις ιδιαιτερότητες και τις προτιμήσεις του χρήστη. Επιπλέον, στο [23], ένα υβριδικό μοντέλο παραγωγής συστάσεων ενσωματώνει συνεργατικό φιλτράρισμα και φιλτράρισμα που βασίζεται στο περιεχόμενο, και παρέχει εξατομικευμένες και ποικίλες προτάσεις για εστιατόρια, μπαρ, ξενοδοχεία και αθλητικές δραστηριότητες για έναν τουριστικό προορισμό των Άλπεων.

2.2 Συστήματα Παραγωγής Εξατομικευμένων Προτάσεων με Επίγνωση Πλαισίου

Τα τελευταία χρόνια, τα συστήματα συστάσεων έχουν βελτιώσει την ανταγωνιστικότητά τους ενσωματώνοντας τη επίγνωση πλαισίου (context-awareness) στους αλγόριθμους φιλτραρίσματος για να μοντελοποιήσουν και να προβλέψουν κατάλληλα τις προτιμήσεις των χρηστών. Η επίγνωση πλαισίου έχει εξελιχθεί στα συστήματα παραγωγής συστάσεων για την παροχή συστάσεων με υψηλή ακρίβεια σύμφωνα με διαφορετικά στοιχεία πλαισίου. Το πλαίσιο μπορεί να αναφέρεται σε οποιαδήποτε πρόσθετη πληροφορία που μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει τις ιδιότητες μιας δεδομένης οντότητας. Μια οντότητα μπορεί να είναι ένα άτομο, μέρος ή αντικείμενο που θεωρείται σχετικό με την αλληλεπίδραση μεταξύ ενός χρήστη και μιας εφαρμογής (συμπεριλαμβανομένου του χρήστη και των ίδιων των εφαρμογών) [24]. Η έννοια του πλαισίου αρχικά αναφερόταν σε γεωγραφικές πληροφορίες λαμβάνοντας υπόψη τη γεωγραφική θέση του χρήστη, γεγονός που έκανε τα Location-Aware Recommender Systems (LARSs) το πιο διαδεδομένο σύστημα συστάσεων με επίγνωση πλαισίου (Context-Aware Recommender System / CARS) [25]. Οι πληροφορίες που σχετίζονται με το πλαίσιο (context) μπορούν να οριστούν ως στατικές ή δυναμικές και μπορούν να ληφθούν ρητά, σιωπηρά, χρησιμοποιώντας εξόρυξη δεδομένων ή μέσω ενός συνδυασμού αυτών των μεθόδων.

Αρκετά CARS έχουν αναπτυχθεί μέχρι στιγμής στον τομέα των τουριστικών εφαρμογών. Συγκεκριμένα, ο οδηγός Magitti [26] λαμβάνει υπόψη την τρέχουσα ώρα, τοποθεσία και πρόσθετα δεδομένα περιβάλλοντος, όπως ο καιρός και οι ώρες λειτουργίας των χώρων. Το PSiS Mobile [27] είναι ένα σύστημα συστάσεων για κινητά που αξιοποιεί πληροφορίες που σχετίζονται με το περιβάλλον για τη δημιουργία συστάσεων σχετικά με μέρη ενδιαφέροντος. Οι παράμετροι που λαμβάνει υπόψη περιλαμβάνουν την τρέχουσα ώρα και τοποθεσία και προγνώσεις καιρού σύμφωνα με τη θέση του χρήστη. Το Turist@ [28] είναι ένα άλλο σύστημα συστάσεων που γνωρίζει τη θέση του χρήστη (με παρακολούθηση συντεταγμένων με χρήση

GPS), καθώς και χωρικές πληροφορίες για κάθε δραστηριότητα, προκειμένου να παρέχει συστάσεις με βάση τη θέση. Επιπλέον, το [29] πρότεινε ένα άλλο CARS, το οποίο λαμβάνει υπόψη τις τοποθεσίες των POI, τις ώρες ανοίγματος και κλεισίματος των POI, το οδικό δίκτυο, το ιστορικό κινητικότητας του χρήστη και την απόσταση που διένυσε ο χρήστης στο οδικό δίκτυο. Τέλος, η Android εφαρμογή LOOKER [30] χρησιμοποιεί βασικές πληροφορίες όπως την τοποθεσία του χρήστη και τον διαθέσιμο χρόνο.

3 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΕΞΑΤΟΜΙΚΕΥΜΕΝΩΝ ΠΡΟΤΑΣΕΩΝ

Η πλατφόρμα e-ΧΝΗΛΑΤΗΣ είναι μια διαδικτυακή υπηρεσία που είναι προσβάσιμη μέσω οποιουδήποτε προγράμματος περιήγησης ιστού μέσω επιτραπέζιων ή κινητών συσκευών. Περιλαμβάνει ένα ενσωματωμένο υβριδικό εξατομικευμένο σύστημα συστάσεων που συνδυάζει τρεις διαφορετικούς τύπους τεχνικών φιλτραρίσματος:

1. Τεχνική φιλτραρίσματος που βασίζεται στη γνώση και σε συλλογισμό βάσει κανόνων (rule-based reasoning),
2. Τεχνική φιλτραρίσματος που βασίζεται στο περιεχόμενο και
3. Τεχνική συνεργατικού φιλτραρίσματος με βάση τη συνεργατική προσέγγιση φιλτραρίσματος που βασίζεται στη μνήμη (memory-based).

Η μηχανή παραγωγής προτάσεων σχεδιάστηκε και αναπτύχθηκε για να εκμεταλλεύεται τις αλληλεπιδράσεις των χρηστών με την πλατφόρμα και να συλλέγει δεδομένα με ρητό και έμμεσο τρόπο, με απώτερο στόχο την προσφορά εξατομικευμένων συστάσεων προς τους τουρίστες σχετικά με αξιοθέατα δυνητικού ενδιαφέροντος.

3.1 Γενική Αρχιτεκτονική

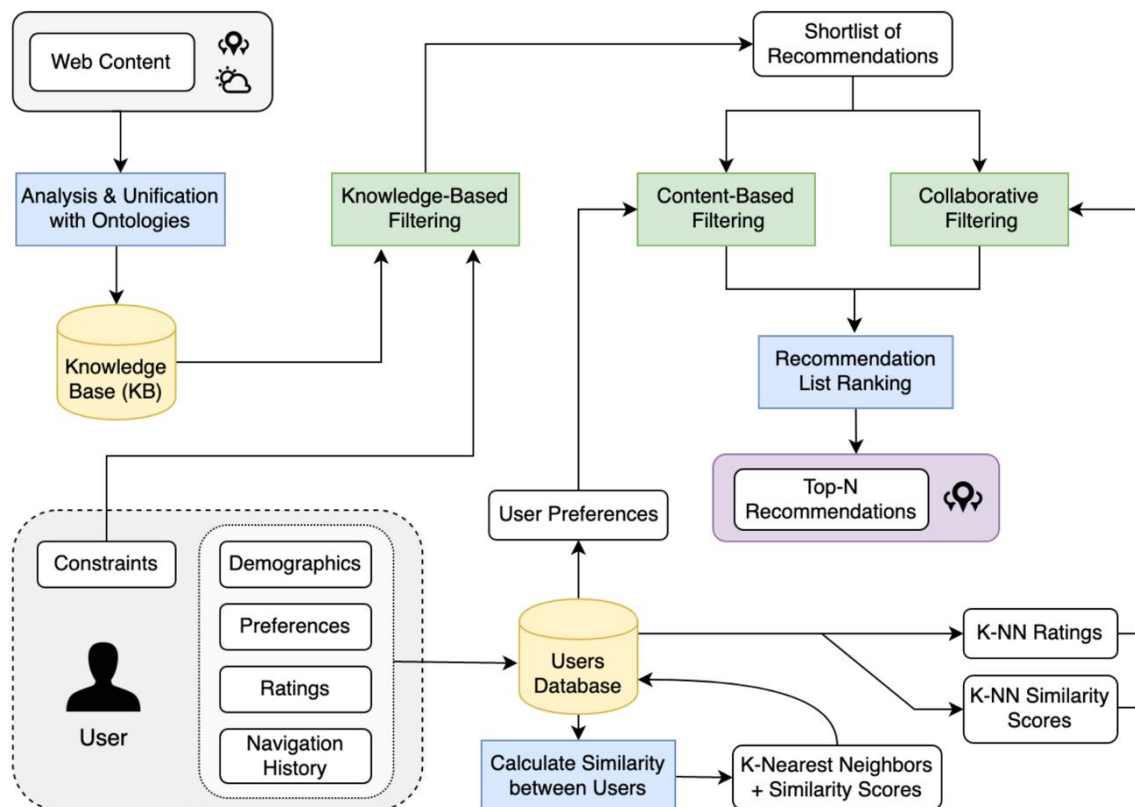
Το γενικό σύστημα παραγωγής προτάσεων που απεικονίζεται στην Εικόνα 1, είναι ένα πλήρως λειτουργικό και αυτόνομο στοιχείο της αρχιτεκτονικής του e-ΧΝΗΛΑΤΗ. Όσον αφορά τις πληροφορίες που είναι αποθηκευμένες στην πλευρά του διακομιστή, χρησιμοποιούνται δύο βάσεις δεδομένων:

1. Η «**Βάση Γνώσης (Knowledge Base / KB)**» που αποτελείται από γενικές πληροφορίες που σχετίζονται με τουριστικά αξιοθέατα, όπως ταξιδιωτικούς προορισμούς, δραστηριότητες και εκδηλώσεις.
2. Η «**Βάση Δεδομένων Χρηστών (User Database)**» που αποθηκεύει τα προφίλ των χρηστών —που δημιουργούνται από προσωπικές πληροφορίες που παρέχονται τόσο ρητά όσο και έμμεσα από τους χρήστες— καθώς και σκορ ομοιότητας μεταξύ των πιο όμοιων χρηστών.

Όσον αφορά την πρώτη βάση, το περιεχόμενο που είναι αποθηκευμένο στη KB συλλέγεται από διάφορες πηγές Ιστού. Η χρήση σημασιολογικών τεχνολογιών, και ειδικότερα οντολογιών, διευκολύνει την ερμηνεία και την ενοποίηση των ετερογενών δεδομένων που συλλέγονται, με αποτέλεσμα να ομογενοποιείται ανόμοιο περιεχόμενο και η γνώση να δομείται με τέτοιο τρόπο όπως ορίζεται από το οντολογικό μοντέλο (Π.2.2.1, Π.2.2.2).

Εκτός από τη συγκέντρωση πληροφοριών σχετικά με τουριστικά αξιοθέατα, μεγάλος όγκος δεδομένων των χρηστών συλλέγεται και διατηρείται στη βάση δεδομένων, όπου αποθηκεύεται σε δομές δεδομένων, γνωστές ως προφίλ χρηστών. Ένα προφίλ χρήστη βασίζεται σε πληροφορίες που παρέχονται είτε ρητά είτε σχετίζονται με ενέργειες του

χρήστη. Για να αποφευχθεί το πρόβλημα της ύπαρξης στατικών προφίλ, παρέχεται δυνατότητα δυναμικής ενημέρωσης μετά την αλληλεπίδραση κάθε χρήστη με την πλατφόρμα του e-ΧΝΗΛΑΤΗ. Κάθε χρήστης μπορεί προαιρετικά — με την εγγραφή του — να υποδείξει τις προτιμήσεις του για ταξιδιωτικές κατηγορίες καθώς και να δηλώσει μια σειρά από προκαθορισμένα δημογραφικά χαρακτηριστικά. Επιπλέον, η βάση δεδομένων καταγράφει τις αξιολογήσεις που δίνονται στους ταξιδιωτικούς προορισμούς, ενώ ταυτόχρονα, το σύστημα ενημερώνει τα ταξιδιωτικά αρχεία (ιστορικό πλοήγησης), κάθε φορά που ένας χρήστης παρέχει βαθμολογία για έναν συγκεκριμένο προορισμό, υποθέτοντας ότι η επίσκεψή του σε αυτόν τον προορισμό έχει ολοκληρωθεί.



Εικόνα 1: Γενική αρχιτεκτονική του συστήματος παραγωγής προτάσεων

Κατά τη διαδικασία παραγωγής προτάσεων, το σύστημα του e-ΧΝΗΛΑΤΗ δεν λαμβάνει υπόψη μόνο δεδομένα που προέρχονται από προφίλ χρηστών, αλλά και διάφορες παραμέτρους χωρικού, χρονικού και περιβαλλοντικού πλαισίου. Προκειμένου να επιτευχθεί η λεγόμενη ικανοποίηση των περιορισμών, πραγματοποιείται η πρώτη φάση φιλτραρίσματος στη λίστα των υποψηφίων προορισμών προς πρόταση, όπου ρητά καθορισμένοι περιορισμοί λαμβάνονται υπόψη από μία τεχνική φιλτραρίσματος που βασίζεται στη γνώση. Η τεχνική αυτή εφαρμόζει συλλογιστική (reasoning) στην ΚΒ, εκμεταλλευόμενη κανόνες που αναφέρονται στον τρόπο με τον οποίο οι απαιτήσεις των χρηστών σχετίζονται με τα χαρακτηριστικά των αντικειμένων (δηλ. των αξιοθέατων),

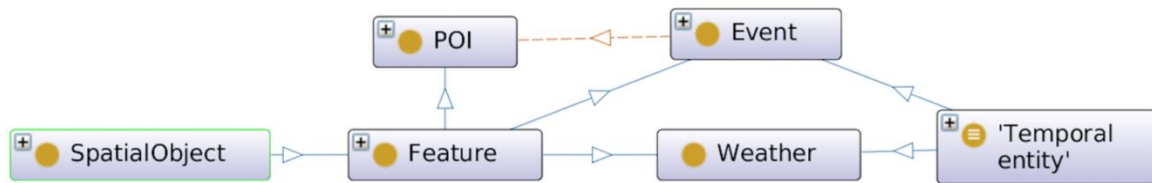
αποκλείοντας έτσι προορισμούς που δεν πληρούν πλήρως αυτά τα κριτήρια. Αυτό οδηγεί σε πολύ αποτελεσματική μείωση του μεγέθους της λίστας των αντικειμένων, από το πρώτο στάδιο της διαδικασίας παραγωγής συστάσεων.

Η δεύτερη φάση του φιλτραρίσματος περιλαμβάνει έναν συνδυασμό αλγορίθμων φιλτραρίσματος (φιλτράρισμα που βασίζεται στο περιεχόμενο και συνεργατικό φιλτράρισμα). Προκειμένου να εκτιμηθεί ένας βαθμός ενδιαφέροντος (ή απλούστερα μια βαθμολογία) για κάθε ένα από τα αντικείμενα της λίστας, το φιλτράρισμα που βασίζεται στο περιεχόμενο, λαμβάνει υπόψη τις ταξιδιωτικές προτιμήσεις του τουρίστα μαζί με τις κατηγορίες που αποδίδονται σε κάθε αξιοθέατο, ενώ το συνεργατικό φιλτράρισμα βασίζεται στις βαθμολογίες και τις ομοιότητες μεταξύ των προφίλ των χρηστών. Βασικά, οι πληροφορίες που είναι αποθηκευμένες σε κάθε προφίλ χρήστη αξιοποιούνται για τον υπολογισμό ενός συνόλου πλησιέστερων γειτόνων για τον χρήστη-στόχο. Επιπλέον, το σύστημα υπολογίζει με ασύγχρονο τρόπο τους πλησιέστερους γείτονες, σε τακτά χρονικά διαστήματα και με κάθε ενημέρωση του προφίλ κάθε χρήστη για να αυξήσει την αποτελεσματικότητά του, ενώ στο μεταξύ αυτή η πρόσφατα αποκτηθείσα γνώση καταγράφεται. Οι βαθμολογίες που προκύπτουν από τους προαναφερθέντες αλγόριθμους συγχωνεύονται σε μια λίστα, χρησιμοποιώντας μια συνάρτηση σταθμισμένου αθροίσματος. Η διαδικασία ολοκληρώνεται με την ταξινόμηση της λίστας με τα υποψήφια αξιοθέατα προς πρόταση με βάση τον βαθμό ενδιαφέροντος.

Στο τελευταίο βήμα μια γεννήτρια ψευδοτυχαίων αριθμών χρησιμοποιείται για να επιλέξει τέσσερα στοιχεία από τα δέκα σημεία που έχουν συγκεντρώσει τον μεγαλύτερο βαθμό ενδιαφέροντος, τα οποία τελικά επιλέγονται για να εκπροσωπήσουν τα επιλεχθέντα σημεία ενδιαφέροντος στον τελικό χρήστη. Αυτό το βήμα είναι πολύ σημαντικό για να υπάρχει η κατάλληλη ανακύκλωση του διαθέσιμου υλικού και η απαιτούμενη ανανέωση στις προτάσεις που παράγει το σύστημα και να αποφευχθεί έτσι η βαρετή επανάληψη των ίδιων αποτελεσμάτων. Περισσότερες πληροφορίες για τις μεθόδους δειγματοληψίας που χρησιμοποιούνται αναφέρονται στο Π2.3.1 στην υποενότητα 4.2.

3.2 Οντολογικό Μοντέλο

Το πλήρες οντολογικό μοντέλο του e-ΧΝΗΛΑΤΗ περιγράφεται αναλυτικά στα Π.2.2.1, Π.2.2.2. Οι βασικές έννοιες που επαρκούν για το υβριδικό εξατομικευμένο σύστημα παραγωγής συστάσεων εκτείνονται κυρίως σε λίγες έννοιες, του οποίου η ιεραρχία παρουσιάζεται στην Εικόνα 2. μαζί με τις σχέσεις τους.



Εικόνα 2:Ελαχιστοποιημένη οντολογία του e-ΧΝΗΛΑΤΗ με σχετικές έννοιες και σχέσεις

Οι κλάσεις «SpatialObject» και «Feature» είναι οι έννοιες που παρέχουν στις οντότητες φυσική έκταση, με τη μορφή γεωσυντεταγμένων. Επιπλέον, η κλάση «Temporal entity» εμποτίζει μια χρονική διάσταση σε στιγμιότυπα των υποκατηγοριών της. Συνδυαστικά, οι προαναφερθείσες κλάσεις επιτρέπουν τη σταδιακή παρακολούθηση αναφορικά με το χώρο και το χρόνο, μια διαδικασία που μπορεί να περιγράψει σημασιολογικά εξελικτικά καιρικά φαινόμενα (κλάση «Weather») ή γεγονότα με χρονικούς περιορισμούς. Τα μπλε βέλη αντιπροσωπεύουν τη σχέση "rdfs:subClassOf" [31], ενώ το ένα πορτοκαλί βέλος μεταξύ «POI» και «Event» αντιπροσωπεύει την σχέση "etr:associatedWith", που σημαίνει ότι ένα «Event» μπορεί να λάβει χώρα σε ένα «POI», αν συμπίπτουν χωρικά στο ίδιο γεωγραφικό μήκος και πλάτος.

3.3 Μοντέλο Χρήστη

Ο όρος "μοντέλο χρήστη" μπορεί να χρησιμοποιηθεί για να περιγράψει ένα ευρύ φάσμα γνώσεων σχετικά με ένα άτομο [32]. Τα TRS συλλέγουν συνήθως μεγάλο όγκο δεδομένων για κάθε εγγεγραμμένο τουρίστα και αποθηκεύουν αυτές τις γνώσεις σε δομές προσωπικών δεδομένων, γνωστές ως προφίλ χρηστών.

3.3.1 Δημιουργία Δυναμικών Προφίλ Χρηστών

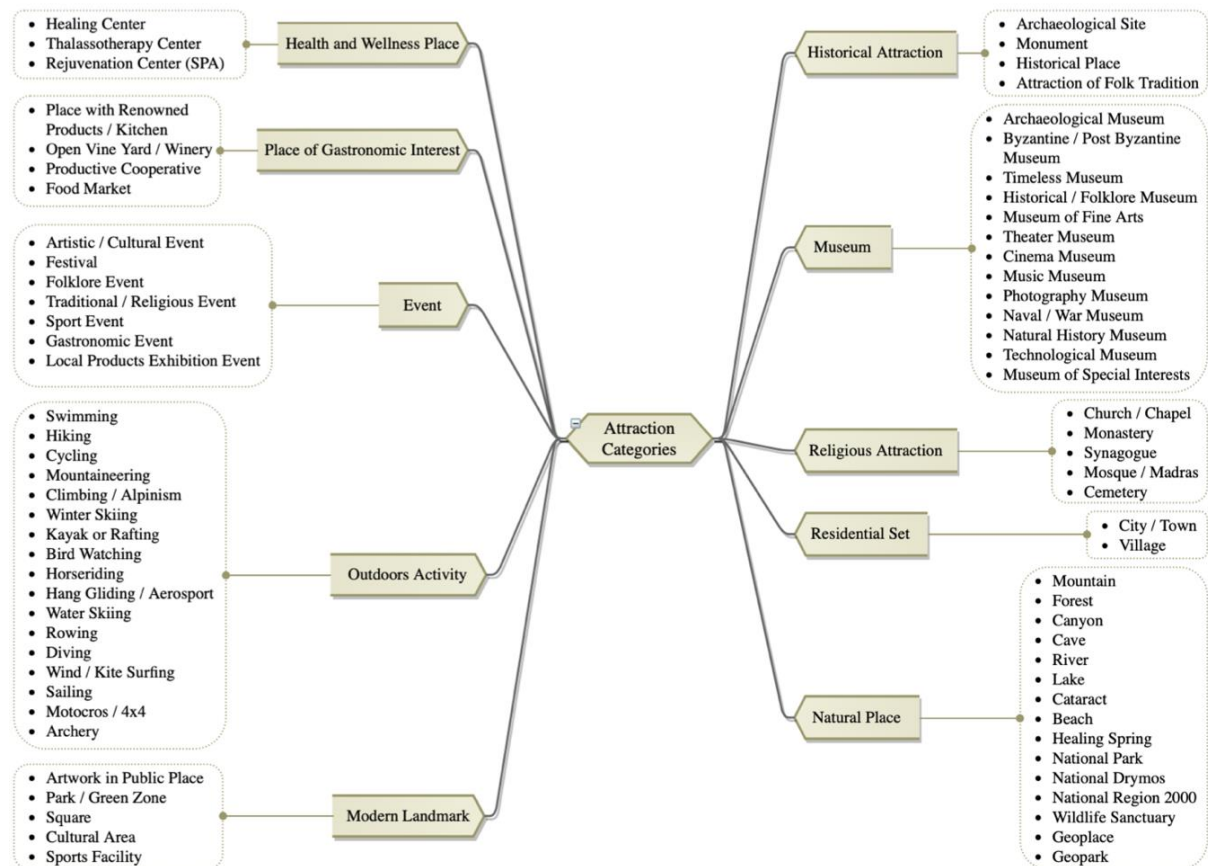
Η σύνθεση ενός προφίλ χρήστη λειτουργεί ως καταλύτης στην εξατομικευμένη επιλογή αντικειμένων, έτσι ώστε το σύστημα παραγωγής συστάσεων να μπορεί να χρησιμοποιηθεί με επιτυχία για τη δημιουργία προτάσεων που ταιριάζουν στα ενδιαφέροντα του τουρίστα. Βασική προϋπόθεση για τη δημιουργία προφίλ χρήστη είναι η ύπαρξη διαθέσιμων δεδομένων για εξαγωγή. Γίνεται συχνά διάκριση μεταξύ δύο τύπων προσωπικών πληροφοριών στα συστήματα συστάσεων [33]:

- Ρητά παρεχόμενες πληροφορίες, είτε μέσω ηλεκτρονικών φορμών και ερωτηματολογίων είτε μέσω ρητών σχολίων, όπως κριτικές και αξιολογήσεις.
- Έμμεσα παρεχόμενες πληροφορίες αναλύοντας τη συμπεριφορά των χρηστών και τις ενέργειες που εκτελούνται κατά την αλληλεπίδραση με το σύστημα συστάσεων.

Το σύστημα συστάσεων e-ΧΝΗΛΑΤΗΣ παρακολουθεί τις διαδικτυακές αλληλεπιδράσεις του χρήστη και τροφοδοτεί το αντίστοιχο προφίλ χρήστη με πληροφορίες που παρέχονται είτε ρητά είτε έμμεσα από ενέργειες του χρήστη. Κατά την εγγραφή, ο χρήστης καλείται να

παρέχει ρητές προσωπικές πληροφορίες. Πιο συγκεκριμένα, δημογραφικά χαρακτηριστικά όπως ηλικία, φύλο, χώρα διαμονής, επίπεδο εκπαίδευσης (πρωτοβάθμια, δευτεροβάθμια, τριτοβάθμια, μεταπτυχιακή), επίπεδο δεξιοτήτων Η/Υ (αρχάριος, μέσο επίπεδο, προχωρημένος, ειδικός) και πιθανές αναπηρίες (όραση, ακοή, κίνηση), καθώς και μια αρχική επιλογή ενδιαφερόντων από μια λίστα 10 βασικών κατηγοριών που σχετίζονται με την τυπολογία των τουριστικών αξιοθέατων (βλ. Εικόνα 3), καταγράφονται μέσω ενός σύντομου ερωτηματολογίου που συμπληρώνεται από τον χρήστη. Αν και η συμπλήρωση του ερωτηματολογίου είναι προαιρετική, συνιστάται, προκειμένου να είναι δυνατές οι συγκρίσεις μεταξύ των χρηστών και να ενεργοποιηθεί η δυνατότητα προσωποποίησης κατά τη διαδικασία δημιουργίας προτάσεων.

Εκτός από τις ρητές πληροφορίες που δίνονται κατά την εγγραφή, το TRS μπορεί επίσης να λάβει σαφείς πληροφορίες από τις αξιολογήσεις των χρηστών για αξιοθέατα που έχουν ήδη επισκεφτεί, ως τρόπο έκφρασης του βαθμού ικανοποίησής τους. Η κλίμακα αξιολόγησης κυμαίνεται μεταξύ 0.5 (χαμηλότερος συντελεστής) και 5.0 (υψηλότερος ρυθμός) και περιλαμβάνει 10 διακριτές τιμές που αυξάνονται κατά μισή μονάδα, με σημείο εκκίνησης το 0.5. Καθώς το προφίλ χρήστη «παρακολουθεί» τα σχόλια που δίνονται στους ταξιδιωτικούς προορισμούς, το TRS ενημερώνει έμμεσα το ιστορικό πλοήγησης του χρήστη, υποθέτοντας ο χρήστης έχει ολοκληρώσει την επίσκεψή του στα αξιοθέατα που έχει αξιολογήσει.



Εικόνα 3: Τυπολογία τουριστικών αξιοθέατων χωρισμένη σε 10 κύριες κατηγορίες και τις αντίστοιχες υποκατηγορίες τους

Οι χρήστες έχουν τη δυνατότητα να τροποποιούν τις αρχικές προτιμήσεις ή καταχωρήσεις τους όταν το κρίνουν απαραίτητο και χωρίς κανένα περιορισμό. Ταυτόχρονα, το σύστημα e-ΧΝΗΛΑΤΗΣ είναι σε θέση να ανακαλύπτει και να λαμβάνει υπόψη οποιαδήποτε νέα πληροφορία και ως εκ τούτου, να εμπλουτίζει ή να ενημερώνει το προφίλ χρήστη καθώς ο χρήστης αλληλεπιδρά με την πλατφόρμα. Έτσι, το προφίλ χρήστη δεν παραμένει στατικό, καθώς μπορεί να ενσωματώσει δυναμικά στο περιεχόμενό του κάθε είδους πληροφορία που προέρχεται από τις αλληλεπιδράσεις του χρήστη-τουρίστα.

Όταν οι πληροφορίες που σχετίζονται με τα προφίλ χρηστών δεν είναι διαθέσιμες στο σύστημα ή είναι ημιτελείς όπως συμβαίνει στις περισσότερες περιπτώσεις νέων χρηστών (χρηστών που έχουν δημιουργήσει προφίλ πρόσφατα), τότε το σύστημα δεν είναι σε θέση να παράγει εξειδικευμένες προτάσεις λόγω έλλειψης στοιχείων. Μία από τις τεχνικές που προτείνονται (η οποία περιγράφεται αναλυτικά και στο Π2.3.1) είναι η ομαδοποίηση των χρηστών υπό μια ταξινόμια χρηστών με γνωστά χαρακτηριστικά που έχουν προσδιοριστεί εκ των προτέρων, οι οποίοι ονομάζονται «προσωπικότητες» ή αλλιώς «Personas». Η τεχνική αυτή χρησιμοποιείται για όσο διάστημα συλλέγονται δεδομένα για τους νέους χρήστες προκειμένου να παρέχονται υπηρεσίες προσωποποίησης προς αυτούς. Όταν συλλεχθούν

ικανοποιητικά αρκετές πληροφορίες για πραγματικούς χρήστες της πλατφόρμας, τότε ο ρόλος των «Personas» περιορίζεται δραστικά.

3.3.2 Δημιουργία Γειτόνων

Η δημιουργία ενός προφίλ χρήστη ακολουθείται από τη διαδικασία αναγνώρισης ενός συνόλου χρηστών με υψηλό βαθμό ομοιότητας, γνωστοί ως πλησιέστεροι γείτονες. Ο υπολογισμός των πλησιέστερων γειτόνων είναι ο πυρήνας του συνεργατικού φιλτραρίσματος που βασίζεται στη μνήμη στη φάση της σύστασης και μπορεί να επιτευχθεί με την εκμετάλλευση όλων των πληροφοριών που είναι αποθηκευμένες στα προφίλ χρηστών. Το TRS του e-ΧΝΗΛΑΤΗ βασίζεται στα διαφορετικά χαρακτηριστικά προφίλ χρήστη, όπως δημογραφικά χαρακτηριστικά, προσωπικές προτιμήσεις και το ιστορικό επισκέψεων και αξιολογήσεων με σκοπό τη δημιουργία γειτόνων.

Για να υπολογίσουμε την ομοιότητα σε κάθε εξεταζόμενο χαρακτηριστικό του προφίλ του χρήστη, κατηγοριοποιούμε τα χαρακτηριστικά ανάλογα με τον τύπο τους σε κατηγορίες και για καθεμία από αυτές ορίζουμε το αντίστοιχο μέτρο ομοιότητας. Στην υποενότητα 3.4 του Π2.3.1 παραγράφονται αναλυτικά οι αλγόριθμοι υπολογισμού της ομοιότητας μεταξύ οντοτήτων, ανάλογα με τη φύση της διαθέσιμη πληροφορίας.

Εφόσον υπολογιστούν τα σκορ ομοιότητας από τις επιμέρους μεταβλητές για κάθε ένα χαρακτηριστικό των ψηφιακών οντοτήτων (βασιζόμενοι στους αλγορίθμους υπολογισμού ομοιότητας), τότε χρησιμοποιείται ένας τελεστής OWA για τη συγκέντρωση των βαθμολογιών ομοιότητας που προέκυψαν από όλα αυτά τα χαρακτηριστικά με σκοπό τον υπολογισμό του συνολικού σκορ ομοιότητας των δύο οντοτήτων (χρήστες).

Για κάθε χρήστη, η διαδικασία δημιουργίας γειτόνων ολοκληρώνεται με την κατασκευή μιας λίστας που περιέχει τις βαθμολογίες ομοιότητας που προέκυψαν μετά από συγκρίσεις με όλα τα προφίλ χρηστών του συστήματος και αποθηκεύοντας στη βάση δεδομένων τους k γείτονες με τον υψηλότερο βαθμό ομοιότητας (k -Nearest Neighbors/ k -NN). Αξίζει να σημειωθεί ότι για να αυξήσει την αποτελεσματικότητά του, το σύστημα του e-ΧΝΗΛΑΤΗ κατά την ενημέρωση του προφίλ κάθε χρήστη, υπολογίζει με ασύγχρονο τρόπο τους k -NN και ενημερώνει ανάλογα τη βάση δεδομένων, ενώ σε τακτά χρονικά διαστήματα (για κάθε 10 χρήστες που καταγράφονται στην πλατφόρμα) ενεργοποιείται μια μαζική ενημέρωση σχετικά με τους k -NN όλων των χρηστών, μετά από τις αντίστοιχες υπολογιστικές διαδικασίες.

3.4 Το Σύστημα Εξατομικευμένων Προτάσεων e-ΧΝΗΛΑΤΗΣ

Ο σκοπός ενός συστήματος παραγωγής προτάσεων σχετικά με ταξιδιωτικούς προορισμούς είναι συνήθως να δημιουργήσει ένα σύνολο από τα πιο σχετικά αξιοθέατα σύμφωνα με τις ιδιαιτερότητες, τις προτιμήσεις και τους περιορισμούς του χρήστη. Η μηχανή συστάσεων e-

ΧΝΗΛΑΤΗΣ χωρίζει τη διαδικασία παραγωγής προτάσεων κατάλληλων αξιοθέατων στις ακόλουθες δύο φάσεις:

- Πρώτη φάση: Εφαρμόζεται φιλτράρισμα βασισμένο στη γνώση στους υποψήφιους προορισμούς, λαμβάνοντας υπόψη περιορισμούς, προκειμένου να φιλτράρονται τα μη λειτουργικά αξιοθέατα ανά χρήστη και να δημιουργηθεί μια μειωμένη λίστα προτάσεων.
- Δεύτερη φάση: Η μειωμένη λίστα προτάσεων τροφοδοτεί διαδοχικά μία υβριδική τεχνική που συνδυάζει έναν αλγόριθμο φιλτραρίσματος που βασίζεται στο περιεχόμενο και έναν συνεργατικό αλγόριθμο φιλτραρίσματος και στοχεύει στην παροχή εξατομικευμένων προτάσεων αναφορικά με νέους προορισμούς που έχουν επιλεγεί από όμοιους χρήστες του συστήματος και είναι κοντά στις προτιμήσεις του τελικού χρήστη.

3.4.1 Φιλτράρισμα που Βασίζεται στη Γνώση

Τα συστήματα συστάσεων που εφαρμόζουν φιλτράρισμα που βασίζεται στη γνώση (knowledge-based filtering), δημιουργούν προτάσεις χρησιμοποιώντας συγκεκριμένες γνώσεις σχετικά με τον τρόπο με τον οποίο ορισμένα χαρακτηριστικά (attributes) ικανοποιούν τις απαιτήσεις του χρήστη. Ένας συγκεκριμένος τύπος σύστασης που βασίζεται στη γνώση είναι εκείνος που βασίζεται σε περιορισμούς (constraint-based), ο οποίος συμβουλευτεί μια γνωσιακή βάση προκαθορισμένων κανόνων, προκειμένου να συσχετίσει ένα σύνολο ρητά καθορισμένων περιορισμών με τα χαρακτηριστικά των αντικειμένων [18].

Η πρώτη φάση της διαδικασίας παραγωγής προτάσεων του e-ΧΝΗΛΑΤΗ περιλαμβάνει φιλτράρισμα βασισμένο στη γνώση (με βάση σκληρούς περιορισμούς) και χρησιμοποιεί χωροχρονικούς μηχανισμούς συλλογιστικής βάσει κανόνων (rule-based reasoning) για την εξαίρεση αξιοθέατων σε περιοχές δυνητικού ενδιαφέροντος, που δεν ικανοποιούν τους τους περιορισμούς του ταξιδιώτη [35]. Πιο συγκεκριμένα, η διεπαφή της πλατφόρμας του e-ΧΝΗΛΑΤΗ ωθεί τον χρήστη να καθορίσει ταξιδιωτικούς περιορισμούς. Οι περιορισμοί περιλαμβάνουν την απόλυτη χιλιομετρική Ευκλείδεια απόσταση που είναι διατεθειμένος να διανύσει ο χρήστης, την διαθεσιμότητά του σε ώρες, καθώς και πιθανές αναπηρίες που μπορεί να έχει. Εκτός από τις ρητές απαιτήσεις του χρήστη, λαμβάνονται υπόψη πρόσθετες παράμετροι χωρικού, χρονικού και περιβαλλοντικού πλαισίου, όπως η τρέχουσα τοποθεσία του χρήστη (καθώς οι συντεταγμένες GPS προωθούνται μέσω μιας έξυπνης συσκευής), ο χρόνος, η μέση εκτιμώμενη διάρκεια μιας επίσκεψης και πιθανά ακραία καιρικά φαινόμενα. Εκτενέστερη αναφορά για την παραπάνω διαδικασία γίνεται στα Π.2.2.1, Π.2.2.2.

Αυτό θεωρείται το πρώτο βήμα μιας προεπεξεργασίας που έχει ως αποτέλεσμα σημαντική μείωση της λίστας των υποψήφιων αξιοθέατων προς πρόταση. Η προκύπτουσα σύντομη λίστα προωθείται σε επόμενους αλγόριθμους για περαιτέρω φιλτράρισμα.

3.4.2 Φιλτράρισμα που Βασίζεται στο Περιεχόμενο

Στο φιλτράρισμα με βάση το περιεχόμενο οι προτιμήσεις του χρήστη μαζί με τα χαρακτηριστικά των αξιοθέατων παίζουν καθοριστικό ρόλο στη διαδικασία δημιουργίας προτάσεων. Μια τεχνική φιλτραρίσματος βάσει περιεχομένου εφαρμόζεται αποτελεσματικά στη δεύτερη φάση της διαδικασίας παραγωγής συστάσεων για να προτείνει αξιοθέατα σύμφωνα με τα ενδιαφέροντα του χρήστη.

Ο βασικός στόχος αυτής της τεχνικής είναι ο υπολογισμός του βαθμού ενδιαφέροντος του χρήστη, που κυμαίνεται στο διάστημα $[0, 1]$, για κάθε αξιοθέατο στη λίστα των υποψηφίων αξιοθέατων. Απαραίτητη προϋπόθεση για μια τέτοια λειτουργία είναι η ρητή δήλωση των προτιμήσεων του χρήστη και ο καθορισμός των χαρακτηριστικών που σχετίζονται με κάθε αξιοθέατο. Στην πλατφόρμα του e-ΧΝΗΛΑΤΗ, οι χρήστες μπορούν να δηλώσουν ρητά το ενδιαφέρον τους για 10 διαφορετικές κατηγορίες αξιοθέατων κατά την εγγραφή τους ή σε μεταγενέστερο στάδιο, ενώ οι κατηγορίες (μία ή περισσότερες) που σχετίζονται με κάθε αξιοθέατο σημειώνονται σημασιολογικά.

Η συσχέτιση μεταξύ των κατηγοριών των αξιοθέατων και των προτιμήσεων του χρήστη υπολογίζεται χρησιμοποιώντας ένα μέτρο ομοιότητας που λειτουργεί καλά σε σύνολα συμβολοσειρών. Για να γίνουμε πιο συγκεκριμένοι, ας υποθέσουμε ότι υπάρχουν δύο σετ συμβολοσειρών, όπου το πρώτο περιλαμβάνει τις κατηγορίες αξιοθέατων που προτιμά ο χρήστης U και το δεύτερο, τις κατηγορίες που αποδίδονται στο αξιοθέατο A . Εάν τα μεγέθη των δύο συνόλων είναι $size_U$ και $size_A$ αντίστοιχα, και $size_{UA}$ το μέγεθος της τομής των δύο συνόλων, τότε ο τύπος τομής των δύο συνόλων είναι [34]:

Εξ. 1

$$CBF_{(U,A)} = \frac{2 * size_{UA}}{size_U * size_A}$$

όπου όλες οι πιθανές τιμές του $CBF_{(U,A)}$ κυμαίνονται μεταξύ 0 και 1.

Για παράδειγμα, ας υποθέσουμε ότι οι προτιμώμενες κατηγορίες του χρήστη U είναι ένα σύνολο {Φυσικό τοπίο, Event, Υπαίθρια δραστηριότητα} ($size_U = 3$) και υπάρχουν τρία υποψήφια αξιοθέατα A, B, C . Οι κατηγορίες του αξιοθέατου A είναι ένα σύνολο {Event, Υπαίθρια δραστηριότητα} ($size_A = 2$), του B ένα σύνολο {Event} ($size_B = 1$) και του C ένα σύνολο {Μουσείο} ($size_C = 1$). Επιπλέον, το σύνολο τομής των U και A είναι {Event, Υπαίθρια δραστηριότητα} ($size_{UA} = 2$), του U και B είναι {Event} ($size_{UB} = 1$) και του U και C {Μουσείο} ($size_{UC} = 1$). Έπειτα, χρησιμοποιώντας την εξίσωση 1, ο βαθμός ενδιαφέροντος του χρήστη U για καθένα από τα αξιοθέατα υπολογίζεται ως εξής:

$$- CBF_{(U,A)} = \frac{2*2}{3+2} = 0.80 (80\%),$$

- $CBF_{(U,B)} = \frac{2*1}{3+1} = 0.50$ (50%) και
- $CBF_{(U,C)} = \frac{2*0}{3+1} = 0$ (0%).

Με βάση τα παραπάνω, είναι προφανές ότι το καλύτερο ταίριασμα για τον χρήστη U είναι το αξιοθέατο A γιατί σχετίζεται με δύο κατηγορίες που έχει προτιμήσει ο χρήστης και στη συνέχεια ακολουθεί το αξιοθέατο B με χαμηλότερο βαθμό ενδιαφέροντος, καθώς περιλαμβάνει μόνο μία από τις προτιμήσεις του χρήστη. Ο βαθμός ενδιαφέροντος σχετικά με το αξιοθέατο C είναι ίσος με 0, λόγω του ότι η κατηγορία του δεν περιλαμβάνεται στις προτιμήσεις του χρήστη.

3.4.3 Συνεργατικό Φιλτράρισμα

Το συνεργατικό φιλτράρισμα, που συχνά αναφέρεται ως κοινωνικό φιλτράρισμα (social filtering) [36], βασίζεται κυρίως στην ομοιότητα μεταξύ των χρηστών. Αυτή η τεχνική φιλτραρίσματος προτείνει ένα σύνολο νέων αντικειμένων που έχουν αξιολογηθεί θετικά από χρήστες οι οποίοι στο παρελθόν έχουν κάνει επιλογές παρόμοιες με αυτές του χρήστη-στόχου. Αναλυτικά, μετά τον προσδιορισμό της ομάδας χρηστών που μοιράζονται παρόμοια ενδιαφέροντα, το συνεργατικό φιλτράρισμα προβλέπει την αξιολόγηση του χρήστη αναφορικά με ένα αντικείμενο, με βάση τις αξιολογήσεις των πλησιέστερων γειτόνων για το συγκεκριμένο αντικείμενο.

Στον e-ΧΝΗΛΑΤΗ, η τεχνική συνεργατικού φιλτραρίσματος, που εφαρμόζεται στη δεύτερη φάση της διαδικασίας παραγωγής προτάσεων, βασίζεται σε μια προσέγγιση συνεργατικού φιλτραρίσματος που βασίζεται στη μνήμη. Σύμφωνα με αυτή την προσέγγιση, οι προβλέψεις δημιουργούνται σταθμίζοντας τις αξιολογήσεις κάθε πλησιέστερου γείτονα αναλογικά με την ομοιότητα με τον χρήστη-στόχο. Ο κύριος στόχος είναι ο υπολογισμός του βαθμού ενδιαφέροντος του χρήστη, που προκύπτει από την κανονικοποίηση της προβλεπόμενης τιμής στο διάστημα $[0, 1]$, για κάθε αξιοθέατο της λίστας των υποψηφίων αξιοθέατων για σύσταση. Η διαδικασία υπολογισμού των πλησιέστερων γειτόνων καθώς και η ομοιότητα μεταξύ των χρηστών, συζητήθηκε στην υποενότητα 3.3.2 αυτού του παραδοτέου.

Με βάση τα παραπάνω, ο αλγόριθμος συνεργατικού φιλτραρίσματος προβλέπει τη βαθμολογία $P_{(U,A)}$ του χρήστη U για ένα αξιοθέατο A υπολογίζοντας το σταθμισμένο άθροισμα των αξιολογήσεων που δίνονται στο αξιοθέατο A από τους k -NN του χρήστη U . Ουσιαστικά, η βαθμολογία κάθε γείτονα $r_{(i,A)}$ για το αξιοθέατο A πολλαπλασιάζεται με την αντίστοιχη ομοιότητα $s_{(U,i)}$ μεταξύ του χρήστη U και του γείτονα i . Επιπλέον, το σταθμισμένο άθροισμα διαιρείται με το άθροισμα των όρων ομοιότητας για να περιοριστεί η πρόβλεψη στο προκαθορισμένο εύρος $[0.5, 5.0]$. Έτσι, η προβλεπόμενη βαθμολογία του χρήστη U στο αξιοθέατο A ορίζεται από τον τύπο [37]:

$$P_{(U,A)} = \frac{\sum_{i \in N_U} r_{(i,A)} * S_{(U,i)}}{\sum_{i \in N_U} S_{(U,i)}}$$

Εξ. 2

όπου N_U είναι το σύνολο των k -NN του χρήστη U .

Όπως αναφέρθηκε, η τιμή της προβλεπόμενης βαθμολογίας $P_{(U,A)}$ κυμαίνεται μεταξύ 0.5 και 5.0. Επομένως, μια μέθοδος ελάχιστης-μέγιστης κλίμακας (min-max scaling) [38] θα πρέπει να χρησιμοποιηθεί για την αντιστοίχιση αυτής της τιμής στην περιοχή [0, 1]. Έτσι, ο γενικός τύπος που κανονικοποιεί το εύρος του $P_{(U,A)}$ δίνεται παρακάτω:

$$CF_{(U,A)} = \frac{P_{(U,A)} - P_{min}}{P_{max} - P_{min}}$$

Εξ. 3

όπου P_{min} είναι η ελάχιστη τιμή που μπορεί να λάβει η προβλεπόμενη βαθμολογία ($P_{min} = 0.5$) και P_{max} η μέγιστη αντίστοιχα ($P_{max} = 5.0$). Η τιμή $CF_{(U,A)}$ αντιπροσωπεύει τον βαθμό ενδιαφέροντος του χρήστη U για το αξιοθέατο A και κυμαίνεται μεταξύ 0 και 1.

3.4.4 Υβριδική Προσέγγιση

Με στόχο να συνδυαστούν τα αποτελέσματα των προαναφερθέντων αυτόνομων τεχνικών φιλτραρίσματος και να παραχθούν ικανοποιητικές συστάσεις στους ταξιδιώτες, χρησιμοποιείται σε συνδυασμό η τεχνική του καταρράκτη και ένα σταθμισμένο υβριδικό μοντέλο παραγωγής προτάσεων. Η τεχνική του καταρράκτη είναι μία αυστηρά ιεραρχική υβριδική τεχνική, στην οποία οι «ασθενέστεροι» αλγόριθμοι παραγωγής προτάσεων δεν μπορούν να ανακαλέσουν αποφάσεις που ελήφθησαν από έναν «ισχυρότερο», αλλά μπορούν απλώς να τις βελτιώσουν [39]. Στον e-ΧΝΗΛΑΤΗ, το φιλτράρισμα που βασίζεται στη γνώση παίζει το ρόλο του «ισχυρότερου» αλγορίθμου, όπου κατά την πρώτη φάση της διαδικασίας παραγωγής συστάσεων φιλτράρει τα άσχετα αντικείμενα (αξιοθέατα). Η προκύπτουσα μειωμένη λίστα από αξιοθέατα χρησιμοποιείται, στη δεύτερη φάση της διαδικασίας παραγωγής συστάσεων, ως είσοδος σε «ασθενέστερους» αλγόριθμους παραγωγής προτάσεων (φιλτράρισμα που βασίζεται στο περιεχόμενο και συνεργατικό φιλτράρισμα) για περαιτέρω βελτιώσεις. Στη συνέχεια, οι «ασθενέστεροι» αλγόριθμοι υπολογίζουν μια βαθμολογία (βαθμό ενδιαφέροντος) για κάθε αξιοθέατο στη λίστα και στη συνέχεια, χρησιμοποιώντας μια σταθμισμένη υβριδική τεχνική, αυτές οι βαθμολογίες συνδυάζονται για να παράγουν μια τελική (ενιαία) βαθμολογία για κάθε υποψήφιο προορισμό.

Αναλυτικά, η σταθμισμένη υβριδική τεχνική συνδυάζει δύο αλγόριθμους φιλτραρίσματος (φιλτράρισμα που βασίζεται στο περιεχόμενο και συνεργατικό φιλτράρισμα), μέσω ενός

γραμμικού συνδυασμού για τη βελτίωση της ακρίβειας των προτάσεων. Ως εκ τούτου, η τελική βαθμολογία κάθε αξιοθέατου βασίζεται σε ένα σταθμισμένο μέσο όρο των ξεχωριστά υπολογισμένων βαθμολογιών. Υποθέτοντας ότι η βαθμολογία (βαθμός ενδιαφέροντος) του χρήστη U για το αξιοθέατο A , που λαμβάνεται από το φιλτράρισμα που βασίζεται στο περιεχόμενο είναι $CBF_{(U,A)}$ και αυτή που λαμβάνεται από το συνεργατικό φιλτράρισμα είναι $CF_{(U,A)}$, τότε η τελική βαθμολογία του χρήστη U στο αξιοθέατο A ορίζεται από τον τύπο:

$$Score_{(U,A)} = W_{CBF} * CBF_{(U,A)} + W_{CF} * CF_{(U,A)} \quad \text{Εξ. 4}$$

όπου W_{CBF} είναι το βάρος (weight) της βαθμολογίας του φιλτραρίσματος που βασίζεται στο περιεχόμενο, και W_{CF} το βάρος της βαθμολογίας του συνεργατικού φιλτραρίσματος, οι τιμές των οποίων θα καθοριστούν στη παρακάτω υποενότητα.

Επιπροσθέτως,

$$W_{CBF} + W_{CF} = 1 \quad \text{Εξ. 5}$$

Με τη βοήθεια της εξίσωσης 4, υπολογίζεται η τελική βαθμολογία για κάθε αξιοθέατο στη λίστα. Σε περίπτωση που ο χρήστης U δεν έχει προσδιορίσει τα ενδιαφέροντά του, η τελική βαθμολογία βασίζεται αποκλειστικά στο συνεργατικό φιλτράρισμα. Επιπλέον, αν υποθέσουμε ότι ένα αξιοθέατο A δεν έχει βαθμολογηθεί από κανέναν από τους πλησιέστερους γείτονες του χρήστη, ο υπολογισμός της προβλεπόμενης βαθμολογίας $P_{(U,A)}$ κατά το συνεργατικό φιλτράρισμα δεν είναι εφικτός και επομένως η μέση τιμή 2.5 ορίζεται από προεπιλογή ($P_{(U,A)} = 2.5$) για να αυξηθεί η πιθανότητα να προταθεί, έναντι άλλων αξιοθέατων με χαμηλή βαθμολογία.

Ως τελικό βήμα, τα αξιοθέατα ταξινομούνται με βάση την κατάταξη της λίστας που δημιουργήθηκε και από αυτά που βρίσκονται στην κορυφή της λίστας (αυτά με την υψηλότερη τελική βαθμολογία), αφού περάσουν από ένα αλγόριθμο παραγωγής ψευδοτυχαίων αριθμών, επιλέγονται εκείνα που θα προταθούν στον τελικό χρήστη. Τέλος, είναι σημαντικό να σημειωθεί ότι εάν ο τελικός χρήστης είναι νέος στην πλατφόρμα και δεν έχει δηλώσει προτιμήσεις ή δεν έχει πραγματοποιήσει αξιολογήσεις, τότε οι τεχνικές φιλτραρίσματος (συνεργατικό και βασισμένο στο περιεχόμενο) δεν έχουν επαρκείς πληροφορίες για να υπολογίσουν τον βαθμό ενδιαφέροντος για τα υποψήφια αξιοθέατα και έτσι, η μηχανή συστάσεων απλώς προτείνει τα πιο δημοφιλή (με τον υψηλότερο βαθμό αξιολόγησης) αξιοθέατα της λίστας.

3.4.5 Προσδιορισμός Βαρών Σταθμισμένης Υβριδικής Τεχνικής

Προκυμμένου να προσδιοριστούν τα βάρη της σταθμισμένης υβριδικής τεχνική της εξίσωσης 4, πραγματοποιήθηκε ένα πείραμα αξιολόγησης της ποιότητα των συστάσεων που παράγονται από το σύστημα παραγωγής συστάσεων e-ΧΝΗΛΑΤΗΣ.

Για την εκτέλεση του πειράματος και την αξιολόγηση του προτεινόμενου συστήματος συστάσεων, κρίθηκε απαραίτητο να συλλεχθεί ένα σύνολο προφίλ χρηστών, τα οποία θα χρησιμοποιηθούν κατά τη διαδικασία της αξιολόγησης. Δεδομένου ότι το προτεινόμενο TRS χρησιμοποιεί εκτενώς το στοιχείο εξατομίκευσης, εκμεταλλεύεται πληροφορίες που προκύπτουν από τα διάφορα χαρακτηριστικά των προφίλ χρηστών, σε αντίθεση με την πλειονότητα των βάσεων δεδομένων ανοιχτού κώδικα που περιλαμβάνουν δεδομένα που αφορούν αποκλειστικά αξιολογήσεις χρηστών.

Έτσι, λόγω ανεπαρκών πληροφοριών στις υπάρχουσες βάσεις δεδομένων, αποφασίστηκε η συλλογή πειραματικών δεδομένων. Δημιουργήθηκε ένα διαδικτυακό ερωτηματολόγιο που χορηγήθηκε μέσω Google Forms¹ και κλήθηκαν ορισμένοι κάτοικοι που βρίσκονται στον άξονα της Εγνατίας Οδού να συμμετάσχουν εθελοντικά. Σύμφωνα με τις οδηγίες που δόθηκαν, ζητήθηκε από τους συμμετέχοντες να βαθμολογήσουν αποκλειστικά τα αξιοθέατα με τα οποία είναι εξοικειωμένοι και έχουν επισκεφτεί τουλάχιστον μία φορά, από μια συγκεκριμένη λίστα, χρησιμοποιώντας μια κλίμακα βαθμολογίας δέκα βαθμών. Η λίστα των αξιοθέατων προς βαθμολόγηση περιορίστηκε στα 100 σημεία και αφορά αξιοθέατα που βρίσκονται κατά μήκος της Εγνατίας Οδού στη Βόρεια Ελλάδα. Η επιλογή έγινε με τέτοιο τρόπο ώστε να καλύπτει καθεμία από τις 10 διαφορετικές κατηγορίες αξιοθέατων (βλ. Εικόνα. 3). Εκτός από τις βαθμολογίες, ζητήθηκε από τους συμμετέχοντες να δηλώσουν ρητά τα δημογραφικά τους χαρακτηριστικά (ηλικία, φύλο, χώρα διαμονής, επίπεδο εκπαίδευσης, επίπεδο δεξιοτήτων Η/Υ και πιθανές αναπηρίες), καθώς και να επιλέξουν τα ενδιαφέροντά τους σχετικά με την τυπολογία των τουριστικών αξιοθέατων. Συμπερασματικά, περίπου το 80% των ατόμων που προσκλήθηκαν, ανταποκρίθηκαν στο κάλεσμα και συμμετείχαν στο ερωτηματολόγιο. Έτσι, το σύνολο δεδομένων που προέκυψε περιέχει τελικά 3.274 βαθμολογίες που δόθηκαν από 72 συμμετέχοντες (55 άνδρες και 17 γυναίκες), ηλικίας 19 έως 55 ετών (με μέση ηλικία τα 32,85 έτη), με διαφορετικά ενδιαφέροντα, εκπαιδευτικό και τεχνολογικό υπόβαθρο (βλ. **Error! Reference source not found.**).

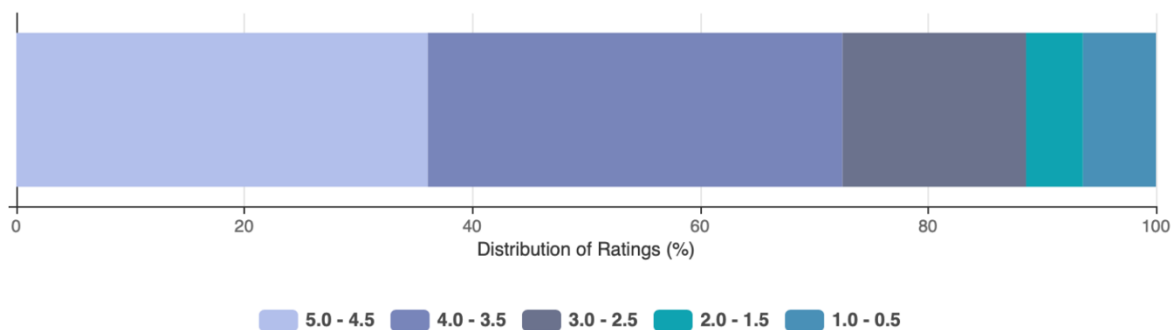
¹ <https://www.google.com/forms/about/>

Μεταβλητή	Κατηγορία	Απόλυτη συχνότητα (N)	Σχετική συχνότητα (%)
Ηλικία	18-25	18	25.0
	26-35	28	38.9
	36-50	24	33.3
	50+	2	2.8
Φύλο	Άνδρας	55	76.4
	Γυναίκα	17	23.7
Επίπεδο εκπαίδευσης	Πρωτοβάθμια	0	0.0
	Δευτεροβάθμια	10	13.9
	Κάτοχος πτυχίου	32	44.4
	Κάτοχος μεταπτυχιακού	30	41.7
Επίπεδο δεξιοτήτων Η/Υ	Αρχάριος	4	5.5
	Μέσο επίπεδο	27	37.5
	Προχωρημένος	20	27.8
	Ειδήμονας	21	29.2
Αναπηρία	Όραση	4	5.5
	Ακοή	0	0.0
	Κίνηση	0	0.0
Ενδιαφέροντα	Ιστορικό αξιοθέατο	48	66.7
	Μουσείο	33	45.8
	Θρησκευτικό αξιοθέατο	31	43.1

	Οικιστικό σύνολο	12	16.7
	Φυσικός τόπος	54	75.0
	Χώρος παροχής υγείας και ευεξίας	14	19.4
	Τόπος γαστρονομικού ενδιαφέροντος	25	33.3
	Εκδήλωση	27	37.5
	Υπαίθρια δραστηριότητα	14	19.4
	Σύγχρονο τοπόσημο	19	26.4

Πίνακας 1: Απόλυτες και σχετικές συχνότητες δειγματοληπτικών μεταβλητών με βάση τις απαντήσεις του ερωτηματολογίου.

Η Εικόνα 4. παρέχει περισσότερες λεπτομέρειες σχετικά με την κατανομή των 3.274 αξιολογήσεων που δόθηκαν από τους 72 συμμετέχοντες στο ερωτηματολόγιο (κατά μέσο όρο 46 αξιολογήσεις ανά χρήστη). Όπως βλέπουμε, πάνω από το 30% των συνολικών αξιολογήσεων έχουν τιμή πάνω από 4.0, ενώ ταυτόχρονα, το ποσοστό των αξιολογήσεων που έχουν τιμή μικρότερη ή ίση με 3.0 είναι υψηλότερο από 25%.



Εικόνα 4: Κατανομή των βαθμολογιών των συμμετεχόντων του ερωτηματολογίου

Η βασική μέθοδος που χρησιμοποιείται για την αξιολόγηση ενός συστήματος προτάσεων βασίζεται στη διαίρεση του κύριου συνόλου δεδομένων σε ένα σύνολο δεδομένων εκπαίδευσης (training) και δοκιμής (testing). Η δημιουργία μοντέλων παραγωγής συστάσεων πραγματοποιείται στο σύνολο δεδομένων εκπαίδευσης και για τη δοκιμή της απόδοσής τους χρησιμοποιείται το σύνολο δεδομένων δοκιμής. Η μέθοδος δοκιμής που χρησιμοποιείται συνήθως είναι η μέθοδος διασταυρούμενης επικύρωσης k-fold cross validation. Στο πείραμα, το μικρό σύνολο δεδομένων και ο περιορισμένος αριθμός στοιχείων υποδηλώνουν ότι η χρήση της k-fold cross validation δεν είναι κατάλληλη, λόγω του γεγονότος ότι δεν υπάρχουν

αρκετά στοιχεία για να τροφοδοτήσουν κατάλληλα τα στάδια εκπαίδευσης και επικύρωσης. Έτσι, η μέθοδος διασταυρούμενης επικύρωσης που επιλέχθηκε για τη διεξαγωγή του πειράματος είναι η leave-one-out cross validation, για να διασφαλιστεί ο μεγαλύτερος δυνατός αριθμός στοιχείων εκπαίδευσης σε κάθε διαδικασία επικύρωσης. Η leave-one-out cross validation είναι μια υπολογιστικά δαπανηρή διαδικασία, αλλά παρ' όλα αυτά οδηγεί σε μια αξιόπιστη και αμερόληπτη εκτίμηση της απόδοσης του μοντέλου προς αξιολόγηση [40].

Σύμφωνα με αυτήν τη μέθοδο επικύρωσης, ένα μεμονωμένο στοιχείο χρησιμοποιείται από το αρχικό δείγμα ως δεδομένο επικύρωσης-δοκιμής και τα υπόλοιπα στοιχεία ως δεδομένα εκπαίδευσης. Έτσι, στο πείραμα, σε κάθε επανάληψη της μεθόδου, τα χαρακτηριστικά των προφίλ χρηστών, που είναι διαθέσιμα στο σύνολο δεδομένων της εκπαίδευσης, χρησιμοποιούνται για τον υπολογισμό της ομοιότητας των χρηστών και την εκτίμηση του βαθμού ενδιαφέροντος του χρήστη για το στοιχείο της επικύρωσης. Οι Bobadilla et al. [41] παρουσιάζουν λεπτομερώς τη γενική λογική πίσω από τη μέθοδο leave-one-out cross validation και παρέχουν ένα παράδειγμα για το πώς δημιουργούνται κορυφαίες προτάσεις για κάθε χρήστη σε ένα σύνολο δεδομένων. Στη συνέχεια, η λίστα κατάταξης των προτάσεων για κάθε χρήστη μπορεί να αξιολογηθεί με βάση τις διάφορες μετρικές.

Η μετρική Normalized Discounted Cumulative Gain (NDCG) [42] χρησιμοποιείται για να προσδιορίσει την αποτελεσματικότητα της προτεινόμενης τεχνικής παραγωγής συστάσεων αναφορικά με την κατάταξη (ταξινόμηση) των συστάσεων, σε σύγκριση με μια βέλτιστη κατάταξη. Όπως ήδη αναφέρθηκε, οι χρήστες κλήθηκαν να βαθμολογήσουν αποκλειστικά τα αξιοθέατα χρησιμοποιώντας μια κλίμακα αξιολόγησης δέκα βαθμών. Στη διαδικασία αξιολόγησης, αυτές οι αξιολογήσεις θεωρήθηκαν ως βαθμολογημένες κρίσεις συνάφειας που μετρούν το κέρδος (gain) ή αλλιώς τη σχετικότητα κάθε αξιοθέατου στην προτεινόμενη λίστα συστάσεων, με βάση τη θέση τους στη λίστα. Όσο χαμηλότερη είναι η θέση ενός αξιοθέατου στη λίστα, τόσο λιγότερο σχετικό είναι για τον χρήστη [42]. Είναι σημαντικό να σημειωθεί ότι σε κάθε προτεινόμενη λίστα συστάσεων, τα αξιοθέατα που δεν βαθμολογήθηκαν ρητά από τον χρήστη-στόχο, θεωρήθηκαν άσχετα και εξαιρέθηκαν από τη λίστα.

Με βάση τα παραπάνω, εφαρμόστηκε η μετρική NDCG για την αξιολόγηση της ποιότητας κατάταξης της σταθμισμένης υβριδικής τεχνικής παραγωγής συστάσεων. Υποθέτοντας ότι κάθε χρήστης U έχει ένα κέρδος g_{U,A_j} εάν το αξιοθέατο A παρουσιάζεται στην προτεινόμενη λίστα συστάσεων στην κατάταξη j , τότε το μέσο μειωμένο αθροιστικό κέρδος (Discounted Cumulative Gain /DCG) για όλους τους χρήστες ορίζεται από τον τύπο:

$$DCG = \frac{1}{n} \sum_{U=1}^n \sum_{j=1}^m \frac{2^{g_{U,A_j}} - 1}{\log_2(j+1)} \quad \text{Εξ. 6}$$

όπου n είναι ο συνολικός αριθμός χρηστών και η παράμετρος m καθορίζει τον αριθμό των σχετικών αξιοθέατων στην προτεινόμενη λίστα συστάσεων για κάθε χρήστη U .

Το NDCG είναι η κανονικοποιημένη έκδοση του DCG και ορίζεται από τον τύπο:

$$NDCG = \frac{DCG}{IDCG} \quad \text{Εξ. 7}$$

όπου το IDCG είναι το ιδανικό DCG, που προέκυψε ταξινομώντας όλα τα αξιothέατα στην προτεινόμενη λίστα με βάση τη βαθμολογημένη συνάφειά τους, παράγοντας το μέγιστο δυνατό DCG (βέλτιστη κατάταξη).

Η έκδοση cut-off top- l του NDCG ορίζει το discount ίσο με το μηδέν για τάξεις μεγαλύτερες του l . Αυτό αναφέρεται ως NDCG@ l και ορίζεται από τον τύπο:

$$NDCG@l = \frac{DCG@l}{IDCG@l} \quad \text{Εξ. 8}$$

όπου l είναι το μήκος της λίστας συστάσεων (top- l συστάσεις).

Το IDCG@ l είναι το ιδανικό DCG@ l που προέκυψε ταξινομώντας όλα τα αξιothέατα στην προτεινόμενη λίστα με βάση τη βαθμολογημένη συνάφειά τους, παράγοντας το μέγιστο δυνατό DCG@ l και έτσι το DCG@ l ορίζεται από τον τύπο:

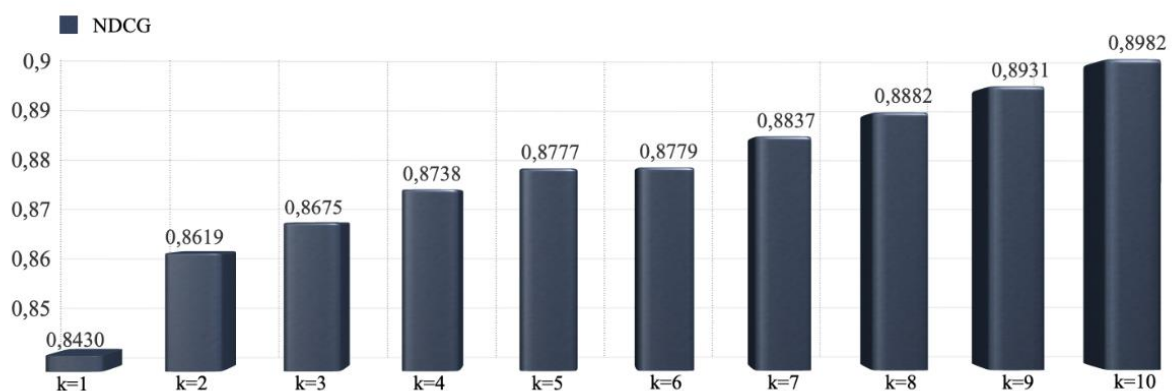
$$DCG@l = \frac{1}{n} \sum_{U=1}^n \sum_{j=1}^l \frac{2^{g_{U,A_j} - 1}}{\log_2(j+1)} \quad \text{Εξ. 9}$$

Όπως αναφέρθηκε, η ποιότητα των ταξινομημένων προτάσεων μπορεί να υπολογιστεί χρησιμοποιώντας τη μέτρηση NDCG, όπου $0 \leq NDCG \leq 1$, με το 0 και το 1 να είναι η χειρότερη και καλύτερη περίπτωση, αντίστοιχα. Αρχικά, η ποιότητα κατάταξης όλων των σχετικών στοιχείων για κάθε χρήστη (αξιothέατα που έχουν βαθμολογηθεί) θα αξιολογηθεί χρησιμοποιώντας τη μέτρηση NDCG και, στη συνέχεια, αφού απομονωθούν οι top- l συστάσεις για κάθε χρήστη για διαφορετικές τιμές l , θα εφαρμοστεί η cut-off top- l έκδοση του NDCG.

Κατά τη διαδικασία αξιολόγησης του συστήματος, τόσο μεμονωμένες προσεγγίσεις παραγωγής συστάσεων (φιλτράρισμα που βασίζεται στο περιεχόμενο και συνεργατικό φιλτράρισμα), η προτεινόμενη υβριδική προσέγγιση (σταθμισμένη υβριδική προσέγγιση e-ΧΝΗΛΑΤΗ), όσο και μια baseline προσέγγιση που απλώς προτείνει τα πιο δημοφιλή αξιothέατα σε κάθε χρήστη (προορισμοί που έχουν την υψηλότερη μέση βαθμολογία με βάση τις αξιολογήσεις όλων των χρηστών), ελέγχθηκαν όσον αφορά το πόσο αποτελεσματικά ταξινομούνται οι παραγόμενες συστάσεις.

Με βάση τα αποτελέσματα (βλ. **Error! Reference source not found.**), η τιμή NDCG της baseline προσέγγισης υπολογίστηκε στο 0.8586 και στη συνέχεια, η τιμή NDCG της

προσέγγισης που χρησιμοποιεί φιλτράρισμα που βασίζεται στο περιεχόμενο υπολογίστηκε στο 0.8270. Όσον αφορά την τεχνική του συνεργατικού φιλτραρίσματος που αποτελεί κομμάτι της υβριδικής προσέγγισης παραγωγής συστάσεων του e-ΧΝΗΛΑΤΗ, και βασίζεται σε αξιολογήσεις των k -NN, ήταν απαραίτητο να προσδιοριστεί ο ιδανικός αριθμός των πλησιέστερων γειτόνων για τον αλγόριθμο. Δεδομένου του περιορισμένου μεγέθους δείγματος, ο αλγόριθμος δοκιμάστηκε για $k = 1$ έως $k = 10$ πλησιέστερους γείτονες (βλ. Εικόνα 5). Μπορούμε ξεκάθαρα να παρατηρήσουμε ότι όσο μεγαλύτερος είναι ο αριθμός των γειτόνων, τόσο υψηλότερη είναι η τιμή NDCG και επομένως τόσο πιο αποτελεσματική είναι η κατάταξη των συστάσεων. Έτσι, η τιμή k ορίστηκε στο 10 και η τιμή NDCG της προσέγγισης συνεργατικού φιλτραρίσματος, χρησιμοποιώντας 10-NN, υπολογίστηκε στο 0.8982.



Εικόνα 5: Ποιότητα κατάταξης της τεχνικής συνεργατικού φιλτραρίσματος για ποικίλο αριθμό (k) πλησιέστερων γειτόνων

Κρίνοντας από τα μέχρι τώρα αποτελέσματα, μπορούμε να συμπεράνουμε ότι το συνεργατικό φιλτράρισμα κατατάσσει τις προτάσεις πιο αποτελεσματικά από την προσέγγιση φιλτραρίσματος που βασίζεται στο περιεχόμενο, καθώς η τιμή NDCG είναι υψηλότερη. Με βάση αυτή τη γνώση, το βάρος του συνεργατικού φιλτραρίσματος (W_{CF}) θα πρέπει να είναι μεγαλύτερο από το βάρος του φιλτραρίσματος που βασίζεται στο περιεχόμενο (W_{CBF}) στην σταθμισμένη υβριδική προσέγγιση του e-ΧΝΗΛΑΤΗ. Επομένως, για την επιλογή των κατάλληλων βαρών, δοκιμάστηκαν αρκετοί συνδυασμοί βαρών (σύμφωνα με την εξίσωση 4), με το βάρος του φιλτραρίσματος που βασίζεται στο περιεχόμενο να έχει πάντα τη χαμηλότερη τιμή (βλ. **Error! Reference source not found.**).

W_{CBF}	W_{CF}	$NDCG$
0.20	0.80	0.90658
0.25	0.75	0.90703
0.30	0.70	0.90824
0.35	0.65	0.90826
0.40	0.60	0.90825
0.45	0.55	0.90804

Πίνακας 2: Ποιότητα κατάταξης της σταθμισμένης υβριδικής προσέγγισης του e-ΧΝΗΛΑΤΗ για διαφορετικούς συνδυασμούς βαρών

Όπως φαίνεται στον **Error! Reference source not found.**, η σταθμισμένη υβριδική προσέγγιση του e-ΧΝΗΛΑΤΗ αποδίδει καλύτερα όταν το βάρος φιλτραρίσματος που βασίζεται στο περιεχόμενο ισούται με 0.35 ($W_{CBF} = 0.35$) και το βάρος του συνεργατικού φιλτραρίσματος ισούται με 0.65 ($W_{CF} = 0.65$). Έτσι, η εξίσωση 4 παίρνει την ακόλουθη μορφή:

$$Score_{(U,A)} = 0.35 * CBF_{(U,A)} + 0.65 * CF_{(U,A)}$$

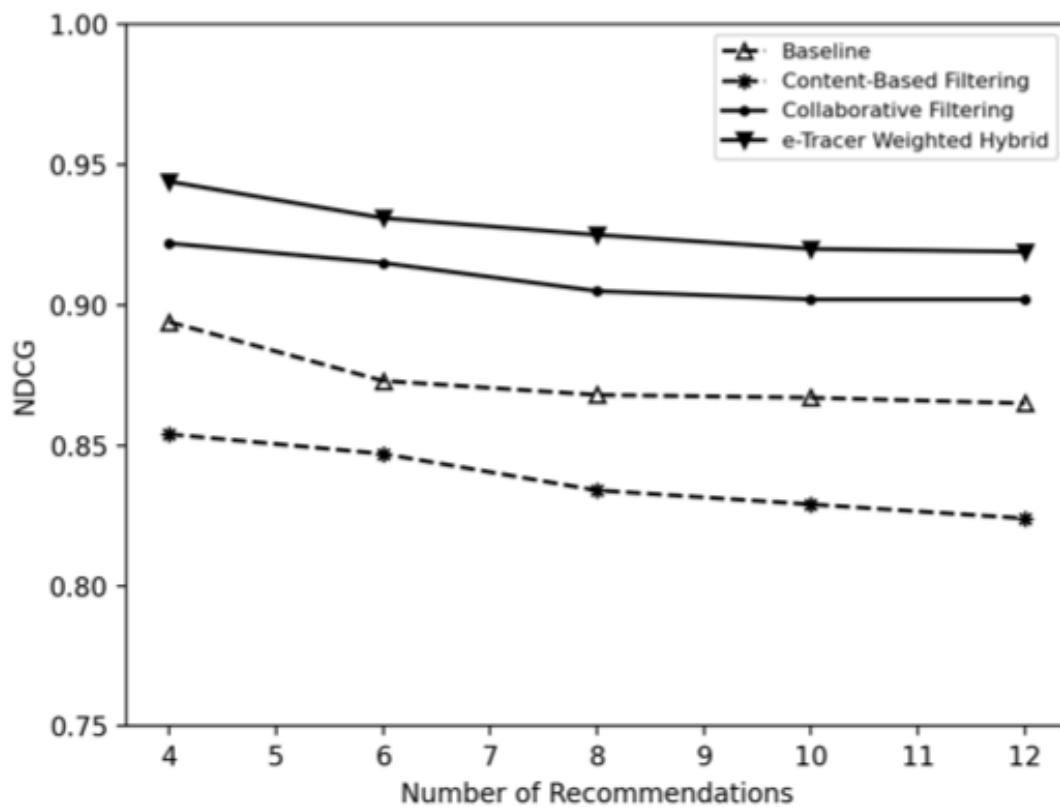
Εξ. 10

Τέλος, η τιμή NDCG της σταθμισμένης υβριδικής προσέγγισης του e-ΧΝΗΛΑΤΗ υπολογίστηκε στο 0.9083 και επομένως, η προτεινόμενη προσέγγιση είναι βέλτιστη σε σύγκριση με τις αυτόνομες προσεγγίσεις (φιλτράρισμα που βασίζεται στο περιεχόμενο και συνεργατικό φιλτράρισμα) και την baseline προσέγγιση, όσον αφορά την ποιότητα των παραγόμενων ταξινομημένων συστάσεων (βλ. **Error! Reference source not found.**).

Προσέγγιση Παραγωγής Συστάσεων	$NDCG$
Baseline	0.8586
Φιλτράρισμα που Βασίζεται στο Περιεχόμενο	0.8270
Συνεργατικό Φιλτράρισμα	0.8982
Σταθμισμένη Υβριδική Προσέγγιση e-ΧΝΗΛΑΤΗ	0.9083

Πίνακας 3: Σύγκριση προσεγγίσεων παραγωγής συστάσεων ως προς την ποιότητα κατάταξης με χρήση της μετρικής NDCG

Εκτός αυτού, λαμβάνοντας υπόψη τις top- l συστάσεις για όλους τους χρήστες του dataset, μπορούμε να συμπεράνουμε εξετάζοντας την Εικόνα 6, ότι η σταθμισμένη υβριδική προσέγγιση του e-ΧΝΗΛΑΤΗ επιτυγχάνει την καλύτερη απόδοση, δεδομένου ότι η προτεινόμενη προσέγγιση έχει τη μεγαλύτερη τιμή $NDCG@l$ για l ίσο με 4, 6, 8, 10 και 12.



Εικόνα 6: Σύγκριση των αποτελεσμάτων $NDCG@l$ για προσεγγίσεις παραγωγής συστάσεων

4 ΣΥΜΠΕΡΑΣΜΑΤΑ

Το παρόν παραδοτέο περιέγραψε τις βασικές έννοιες σχετικά με τα συστήματα παραγωγής συστάσεων, με ιδιαίτερη εστίαση στον τομέα του τουρισμού. Λαμβάνοντας υπόψη τα υπάρχοντα συστήματα συστάσεων και ειδικότερα τις μεθοδολογίες που χρησιμοποιούνται, το σύστημα παραγωγής συστάσεων e-ΧΝΗΛΑΤΗΣ στόχευσε στην προσφορά μιας ολοκληρωμένης υπηρεσίας που παρέχει εξατομικευμένες συστάσεις για σημεία ενδιαφέροντος, εκδηλώσεις και δραστηριότητες που συνδυάζουν φιλτράρισμα που βασίζεται στη γνώση, συνεργατικό φιλτράρισμα και φιλτράρισμα που βασίζεται στο περιεχόμενο, σε μια προσπάθεια καταπολέμησης των περιορισμών που έχει η αποκλειστική χρήση μιας μεμονωμένης τεχνικής φιλτραρίσματος. Πιο συγκεκριμένα, εισήγαγε και ανέλυσε ένα υβριδικό σύστημα παραγωγής προτάσεων που παρέχει εξατομικευμένες συστάσεις σε πραγματικό χρόνο, διευκολύνοντας τη διαδικασία εντοπισμού των πιο κατάλληλων ανεξερεύνητων προορισμών στην διαδικτυακή πλατφόρμα του e-ΧΝΗΛΑΤΗ.

Το έγγραφο αυτό παρέχει τις τεχνικές και μεθοδολογικές λεπτομέρειες που χρειάζεται η υλοποίηση των αντίστοιχων υπομονάδων λογισμικού και μπορεί να χρησιμοποιηθεί και ως αναφορά σε μελλοντικές καταστάσεις όπως για παράδειγμα κατά την δεύτερη φάση των πιλοτικών δοκιμών όπου θα δοκιμαστεί και η συνολική αποδοτικότητα του συστήματος.

Για την χειροκίνητη (εκτός της εφαρμογής e-ΧΝΗΛΑΤΗΣ) δοκιμή των υπηρεσιών που παρουσιάζονται στο παρόν παραδοτέο, υπάρχει διαθέσιμη συλλογή “Postman” στον σύνδεσμο που ακολουθεί:

https://drive.google.com/drive/folders/1dqBOuGQZuYqkslqS_DrX_BfdIJ8MDZ7b?usp=drive_link

5 ΑΝΑΦΟΡΕΣ

- [1] Gavalas, D., Konstantopoulos, C., Mastakas, K., & Pantziou, G. (2014). Mobile recommender systems in tourism. *Journal of Network and Computer Applications*, 39, 319–333. <https://doi.org/10.1016/j.jnca.2013.04.006>
- [2] Borràs, J., Moreno, A., & Valls, A. (2014). Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, 41(16), 7370–7389. <https://doi.org/10.1016/j.eswa.2014.06.007>
- [3] Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. In *Recommender systems handbook* (pp. 1–34). Springer. https://doi.org/10.1007/978-1-4899-7637-6_1
- [4] Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5(1–2), 115–153. <https://doi.org/10.1023/A:1009804230409>
- [5] Gavalas, D., & Kenteris, M. (2011). A web-based pervasive recommendation system for mobile tourist guides. *Personal and Ubiquitous Computing*, 15(7), 759–770. <https://doi.org/10.1007/s00779-011-0389-x>
- [6] Meehan, K., Lunney, T., Curran, K., & McCaughey, A. (2013). Context-aware intelligent recommendation system for tourism. 328–331. <https://doi.org/10.1109/PerComW.2013.6529508>
- [7] Kzaz, L., Dakhchoune, D., Dahab, D. et al. (2018). Tourism recommender systems: An overview of recommendation approaches. *Int. J. Comput. Appl.*, 180(20), 9–13. <https://doi.org/10.5120/ijca2018916458>
- [8] Christensen, I., Schiaffino, S., & Armentano, M. (2016). Social group recommendation in the tourism domain. *Journal of Intelligent Information Systems*, 47(2), 209–231. <https://doi.org/10.1007/s10844-016-0400-0>
- [9] Delic, A., Neidhardt, J., Nguyen, T. N., & Ricci, F. (2018). An observational user study for group recommender systems in the tourism domain. *Information Technology & Tourism*, 19(1-4), 87-116. <https://doi.org/10.1007/s40558-018-0106-y>
- [10] Ricci, F. (2002). Travel recommender systems. *IEEE Intelligent Systems*, 17(6), 55–57.
- [11] Cvetković, B., Gjoreski, H., Janko, V. et al. (2016). e-Turist: An intelligent personalised trip guide. *Informatica J. Comput. Inform.*, 40(4), 447-455.
- [12] Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56–58. <https://doi.org/10.1145/245108.245121>
- [13] Vansteenwegen, P., Souffriau, W., Berghe, G. V., & Van Oudheusden, D. (2011). The city trip planner: An expert system for tourists. *Expert Systems with Applications*, 38(6), 6540–6546. <https://doi.org/10.1016/j.eswa.2010.11.085>
- [14] Choi, I. Y., Ryu, Y. U., & Kim, J. K. (2021). A recommender system based on personal constraints for smart tourism city. *Asia Pacific Journal of Tourism Research*, 26(4), 440-453. <https://doi.org/10.1080/10941665.2019.1592765>
- [15] Garcia, A., Torre, I., & Linaza, M. T. (2013). Mobile social travel recommender system. In *Information and communication technologies in tourism 2014* (pp. 3–16). Springer. https://doi.org/10.1007/978-3-319-03973-2_1
- [16] Sebastia, L., Garcia, I., Onaindia, E., & Guzman, C. (2009). e-Tourism: A tourist recommendation and planning application. *International Journal on Artificial Intelligence Tools*, 18(05), 717–738. <https://doi.org/10.1142/S0218213009000378>
- [17] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331–370. <https://doi.org/10.1023/A:1021240730564>

-
- [18] Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook* (pp. 1–35). Springer. https://doi.org/10.1007/978-0-387-85820-3_1
- [19] Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132. <https://doi.org/10.1016/j.knosys.2013.03.012>
- [20] Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender systems handbook* (pp. 73–105). Springer. https://doi.org/10.1007/978-0-387-85820-3_3
- [21] Lucas, J. P., Luz, N., Moreno, M. N., Anacleto, R., Figueiredo, A. A., & Martins, C. (2013). A hybrid recommendation approach for a tourism system. *Expert systems with applications*, 40(9), 3532–3550. <https://doi.org/10.1016/j.eswa.2012.12.061>
- [22] Renjith, S., & Anjali, C. (2014, December). A personalized mobile travel recommender system using hybrid algorithm. In *2014 First International Conference on Computational Systems and Communications (ICCS)* (pp. 12–17). IEEE.
- [23] Majeed, T., Stämpfli, A. E., Liebrich, A., & Meier, R. (2018, June). Personalized hybrid recommendations for daily activities in a tourist destination. In *International Symposium on Ambient Intelligence* (pp. 155–165). Springer, Cham. https://doi.org/10.1007/978-3-030-01746-0_18
- [24] Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing*, 5(1), 4–7. <https://doi.org/10.1007/s007790170019>
- [25] Levandoski, J. J., Sarwat, M., Eldawy, A., & Mokbel, M. F. (2012, April). Lars: A location-aware recommender system. In *2012 IEEE 28th international conference on data engineering* (pp. 450–461). IEEE. <https://doi.org/10.1109/ICDE.2012.54>
- [26] Bellotti, V., Begole, B., Chi, E. H. et al. (2008, April). Activity-based serendipitous recommendations with the Magitti mobile leisure guide. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1157–1166).
- [27] Anacleto, R., Figueiredo, L., Luz, N., Almeida, A., & Novais, P. (2011). Recommendation and planning through mobile devices in tourism context. In *Ambient Intelligence-Software and Applications* (pp. 133–140). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-19937-0_17
- [28] Batet, M., Moreno, A., Sánchez, D., Isern, D., & Valls, A. (2012). Turist@: Agent-based personalised recommendation of tourist activities. *Expert Systems with Applications*, 39(8), 7319–7329. <https://doi.org/10.1016/j.eswa.2012.01.086>
- [29] Bahramian, Z., Ali Abbaspour, R., & Claramunt, C. (2017). A cold start context-aware recommender system for tour planning using artificial neural network and case based reasoning. *Mobile Information Systems*, 2017. <https://doi.org/10.1155/2017/9364903>
- [30] Missaoui, S., Kassem, F., Viviani, M., Agostini, A., Faiz, R., & Pasi, G. (2019). LOOKER: a mobile, personalized recommender system in the tourism domain based on social media user-generated content. *Personal and Ubiquitous Computing*, 23(2), 181–197. <https://doi.org/10.1007/s00779-018-01194-w>
- [31] McBride, B. (2004). The resource description framework (RDF) and its vocabulary description language RDFS. In *Handbook on ontologies* (pp. 51–65). Springer. https://doi.org/10.1007/978-3-540-24750-0_3
- [32] Rich, E. (1983). Users are individuals: Individualizing user models. *International Journal of Man-Machine Studies*, 18(3), 199–214. [https://doi.org/10.1016/S0020-7373\(83\)80007-8](https://doi.org/10.1016/S0020-7373(83)80007-8)
- [33] Eirinaki, M., & Vazirgiannis, M. (2003). Web mining for web personalization. *ACM Transactions on Internet Technology (TOIT)*, 3(1), 1–27. <https://doi.org/10.1145/643477.643478>
- [34] Nguyen, M. H., & Nguyen, T. H. (2015). A general model for similarity measurement between objects. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, 6(2), 235–239. <https://doi.org/10.14569/IJACSA.2015.060234>
-

-
- [35] Stathopoulos, E. A., Kokkalas, A., Mitsopoulou, E. E. et al. (2020). Knowledge-Based Management and Reasoning on Cultural and Natural Touristic Routes. 355–367. https://doi.org/10.1007/978-3-030-49161-1_30
- [36] Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence, 43-52.
- [37] Candillier, L., Meyer, F., & Fessant, F. (2008). Designing specific weighted similarity measures to improve collaborative filtering systems. 242–255. https://doi.org/10.1007/978-3-540-70720-2_19
- [38] Aggarwal, C. C. (2015). Data mining: The textbook. Springer. <https://doi.org/10.1007/978-3-319-14142-8>
- [39] Burke, R. (2007). Hybrid web recommender systems. In The adaptive web (pp. 377–408). Springer. https://doi.org/10.1007/978-3-540-72079-9_12
- [40] Chapelle, O., Vapnik, V., Bousquet, O., & Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. Machine learning, 46(1-3), 131-159. <https://doi.org/10.1023/A:1012450327387>
- [41] Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). A collaborative filtering approach to mitigate the new user cold start problem. Knowledge-based systems, 26, 225-238. <https://doi.org/10.1016/j.knosys.2011.07.021>
- [42] Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems (TOIS), 20(4), 422–446. <https://doi.org/10.1145/582415.582418>