



e-ΧΝΗΛΑΤΗΣ

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

Π2.1.2

Τεχνικές αυτόματης αναζήτησης και εξόρυξης γνώσης από κοινωνικά δίκτυα και ΤΟΝ ΙΣΤΟ

Επίπεδο διάδοσης:	Δημόσιο
Ημερομηνία παράδοσης σύμφωνα με το έργο:	Μήνας 46, 05/06/2022
Πραγματική ημερομηνία παράδοσης:	Μήνας 46, 05/06/2022
Ενότητα εργασίας:	ΕΕ2 Ανάπτυξη τεχνολογιών χωρο-χρονικής διασύνδεσης πολλαπλών χώρων και γεγονότων περιβαλλοντικού, πολιτιστικού και τουριστικού ενδιαφέροντος
Υποενότητα:	ΥΕΕ2.1 Αυτόματη αναζήτηση και γνώση από κοινωνικά δίκτυα και το διαδίκτυο σε πραγματικό χρόνο
Είδος:	Έκθεση + Λογισμικό
Κατάσταση έγκρισης:	Εγκεκριμένο
Έκδοση:	1.0

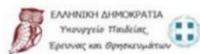
Αριθμός σελίδων:	35
Όνομα αρχείου:	Π2.1.2_final.pdf
Περίληψη	
<p>Αυτό το παραδοτέο αποτελεί τη δεύτερη έκδοση της υποενότητας 2.1 περιγράφει εκτενώς τα μοντέλα δεδομένων τα οποία χρησιμοποιήθηκαν και διαμορφώθηκαν για τις ανάγκες του e-ΧΝΗΛΑΤΗ. Ακόμη παρουσιάζει τις νέες πηγές δεδομένων που χρησιμοποιήθηκαν για εξόρυξη δεδομένων καθώς και τα σύνολα δεδομένων δημιουργήσαν. Το παραδοτέο κλείνει με τα συμπεράσματα και ακολουθούν παραρτήματα με τις αναφορές, κάποια παραδείγματα πηγαίου κώδικα και αποκρίσεις APIs από χρήσιμες πλατφόρμες.</p>	



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο
Περιφερειακής Ανάπτυξης



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ
ΟΙΚΟΝΟΜΙΑΣ & ΑΝΑΠΤΥΞΗΣ
ΕΙΔΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΕΠΑ & ΤΣ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΡΙΣΗΣ ΕΠΑ-ΕΚ



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Παιδείας,
Έρευνας και Θρησκευμάτων



ΓΓΕΤ
ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ
ΕΡΕΥΝΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ



ΕΥΣΕ
ΕΙΔΙΚΟ



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
V0.1	30/4/2021	Πίνακας περιεχομένων	Ευάγγελος Σταθόπουλος Αλέξανδρος Κοκκάλας
V0.3	5/10/2021	Ανανεωμένη ανάλυση τεχνολογιών αιχμής	Αλέξανδρος Κοκκάλας
V0.8	25/12/2021	Κυκλοφορία πρώτης έκδοσης για σχόλια	Αλέξανδρος Κοκκάλας
V0.9	30/2/2022	Εσωτερική αναθεώρηση	Ευάγγελος Σταθόπουλος
V0.99	2/4/2022	Ενσωμάτωση σχολίων	Αλέξανδρος Κοκκάλας
V1.0	5/6/2022	Ετοιμασία τελικής έκδοσης	Αλέξανδρος Κοκκάλας

Λίστα συγγραφέων

Οργανισμός	Όνομα	Στοιχεία επικοινωνίας
ΕΚΕΤΑ	Αλέξανδρος Κοκκάλας	akokkalas@iti.gr
ΕΚΕΤΑ	Ευάγγελος Σταθόπουλος	estathop@iti.gr
ΕΚΕΤΑ	Ιωάννης Κομπατσιάρης	ikom@iti.gr

Περίληψη κυριότερων σημείων

Το παρόν παραδοτέο εμπεριέχει και αναλύει κυρίως τις εργασίες που πραγματοποιήθηκαν στο πλαίσιο της υποενότητας εργασίας ΥΕ2.1, τις σχετικές με την ανάπτυξη εργαλείων για τη διεκπεραίωση των διαδικασιών της ανάκτησης πληροφορίας από το διαδίκτυο και τη διαμόρφωση συνόλων δεδομένων προς εξυπηρέτηση των σκοπών του e-ΧΝΗΛΑΤΗ.

Πιο συγκεκριμένα παρουσιάζεται η τελική έκδοση των εργαλείων για τη συλλογή δεδομένων από κοινωνικά δίκτυα και δικτυακές πηγές. Παρουσιάζονται τα μοντέλα δεδομένων τα οποία υιοθετήθηκαν για να αποθηκεύσουν την εξόρυξη γνώσης σε σύνολα δεδομένων από το διαδίκτυο είτε δομικά σε πλούσιες και χρήσιμες μορφές (GEOSON), είτε σημασιολογικά για καλύτερη αναπαράσταση γνώσης και περαιτέρω εφαρμογή κανόνων συλλογιστικής για εξόρυξη υποβόσκουσας γνώσης (Οντολογίες OWL/RDF). Τα σύνολα δεδομένων που ακολουθούν είναι παραδείγματα άντλησης και ενσωμάτωσης πληροφοριών από 3 μεγάλες ιστοσελίδες συναφείς με το αντικείμενο του e-ΧΝΗΛΑΤΗ.

Τέλος, αναφέρονται ρητά τα συμπεράσματα που εξήχθησαν κατά τις εργασίες της υποενότητας ΥΕ2.1 Το παραδοτέο κλείνει με τις αναφορές και τα παραρτήματα που απεικονίζουν παραδείγματα πηγαίου κώδικα και αποκρίσεις των εξωγενών APIs.

Συντομογραφίες και ακρωνύμια

API	Application Programming Interface
CSS	Cascading Style Sheets
DOM	Document Object Model
HTML	Hyper Text Markup Language
OWL	Web Ontology Language
REGEX	Regular Expressions
REST	Representational State Transfer
SHACL	Shapes Constraint Language
SPARQL	SPARQL Protocol and RDF Query Language
UML	Unified Modelling Language
URL	Uniform Resource Locator
AMEA	Άτομα με ειδικές ανάγκες

Πίνακας περιεχομένων

1. ΕΙΣΑΓΩΓΗ	8
2. ΜΟΝΤΕΛΟ ΔΕΔΟΜΕΝΩΝ	9
2.1 GeoJSON	9
2.2 Οντολογία	12
3. ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ	13
3.1 Σύνολα δεδομένων από epirusforallseasons.gr	13
3.2 Σύνολα δεδομένων από Πολιτιστικό Ίδρυμα Ομίλου Πειραιώς (ΠΙΟΠ)	19
3.3 Σύνολα δεδομένων από OpenStreetMaps	19
4. ΣΥΜΠΕΡΑΣΜΑΤΑ & ΕΠΟΜΕΝΑ ΒΗΜΑΤΑ	22
Αναφορές	23
Παράρτημα	24

Πίνακας εικόνων

Εικόνα 1: Σημεία ενδιαφέροντος ανά ενότητα σε ιστοσελίδα epirusforallseasons.gr	17
Εικόνα 2: Λεπτομέρειες σημείου ενδιαφέροντος με χωρικές πληροφορίες σε ιστοσελίδα epirusforallseasons.gr	17
Εικόνα 3: Θεματικά μουσεία ΠΙΟΠ	18
Εικόνα 4: Συλλεγμένα σημεία ενδιαφέροντος απο OpenStreetMaps	20

1. ΕΙΣΑΓΩΓΗ

Η δεύτερη ενότητα εργασίας είναι υπεύθυνη για την αυτόματη εξόρυξη γνώσης από κοινωνικά δίκτυα και τον ιστό και να προσφέρει το βασικό πλαίσιο για την κωδικοποίηση των πληροφοριών και την ενσωμάτωση τους σε όλα τα πεδία εφαρμογής του e-ΧΝΗΛΑΤΗ.

Στο παρόν παραδοτέο Π2.1.2 παρουσιάζονται οι αλλαγές και τροποποιήσεις που υλοποιήθηκαν στα εργαλεία συλλογής και ανάλυσης δεδομένων από κοινωνικά δίκτυα και της στοχευμένης εξαγωγής δεδομένων από δικτυακές πηγές. Ακόμη παρουσιάζονται και αναλύονται νέες πηγές πληροφορίας που ενσωματώθηκαν στο σύστημα του e-ΧΝΗΛΑΤΗ για την εξόρυξη νέων δεδομένων και σημείων ενδιαφέροντος.

Για άντληση της πληροφορίας αναπτύχθηκαν crawlers με κατάλληλα φίλτρα για τη συλλογή μόνο υποσυνόλων ενδιαφέροντος σχετικών πληροφοριών με το έργο. Τρεις τύποι δεδομένων ανιχνεύονται και ανακτώνται στα πλαίσια του e-ΧΝΗΛΑΤΗ: προσφερόμενα ελεύθερα APIs από κοινωνικές πλατφόρμες που σερβίρουν χρήσιμα δεδομένα και μεταδεδομένα, επίσημες στοχευμένες ιστοσελίδες από έγκυρους φορείς με πλούσιο περιεχόμενο, όπως επίσης και γεωπύλες από αντίστοιχα έγκυρους φορείς.

Το συλλεγόμενο υλικό αποθηκεύεται και χρησιμοποιείται από τις υπόλοιπες λειτουργικές μονάδες του έργου.

2. ΜΟΝΤΕΛΟ ΔΕΔΟΜΕΝΩΝ

Καθώς υπάρχει πληθώρα σχετικών πόρων που αντλούνται από ετερογενείς πηγές, κρίνεται αναγκαία η διερεύνηση μεθόδου ομογενοποίησης των ληφθέντων δεδομένων και μεταδεδομένων από το διαδίκτυο. Από την πρώτη έκδοση του παραδοτέου Π2.1.1 τα δεδομένα κρατούνται σε μία ενοποιημένη, καθολική, ειδικού τύπου μορφή JSON, το GEOJSON που αναλύεται στην υποενότητα 6.1. Εν συνεχεία, σειριακά ακολουθεί η αντιστοίχιση (mapping) των πεδίων GEOJSON στις οντότητες της ήδη διαμορφωμένης οντολογίας (όπως παρουσιάζεται εκτενώς στο παραδοτέο Π2.2.1 και Π2.2.2) μέσω αυτοματοποιημένων μεθόδων με σκοπό τον πληθυσμό (population) της γνωσιακής βάσης (Knowledge Base).

2.1 GeoJSON

Για τη μεταφορά και την απεικόνιση των δεδομένων έχει επιλεγθεί η μορφή του geoJSON. Το geoJSON βασίζεται στο JavaScript Object Notation (JSON) και το έχει επεκτείνει ώστε να αναπαριστά γεωχωρικά δεδομένα μαζί με τα μη χωρικά στοιχεία τους. Μπορεί να αναπαραστήσει σημεία (points), γραμμές (lines), πολύγωνα (polygons) αλλά και τον συνδυασμό των παραπάνω χωρικών τύπων. Όπως και το JSON είναι γλωσσικά ανεξάρτητο και το καθιστά ιδανικό για επικοινωνία διαφορετικών συστημάτων.

Στη παρούσα υλοποίηση που αναπαριστά σημεία ενδιαφέροντος ο χωρικός τύπος που χρησιμοποιήθηκε είναι του σημείου (point) και για τη πλήρη περιγραφή των δεδομένων του σημείου, επιλέχθηκαν τα κατάλληλα πεδία (attributes).

Στη τελική έκδοση του συστήματος το geoJSON έχει κρατήσει την αρχική του γενική δομή και μετά τις πιλοτικές δοκιμές και την αξιολόγηση του συστήματος τροποποιήθηκε κατάλληλα και προστέθηκαν πεδία για την καλύτερη απεικόνιση και αποθήκευση της πληροφορίας.

Παρακάτω ακολουθεί ένα παράδειγμα σημείων ενδιαφέροντος σε μορφή geoJSON.

```
{
  "type": "FeatureCollection",
  "name": "points_of_interest",
  "features": [
    {
      "id": null,
      "type": "Feature",
      "geometry": {
        "type": "Point",
        "coordinates": [21.158333, 39.785833]
      },
      "properties": {
        "wc": null,
        "food": null,
        "name": "Χιονοδρομικό Κέντρο Προφήτη Ηλία Ιωάννινα",
        "dimos": null,
        "email": null,
        "image": "https://epirusforallseasons.gr/wp-content/uploads/2019/12/05_02_003_2-768x509.jpg",
        "is_ar": null,
        "awards": null,
        "refuge": null,
        "routes": null,
        "unesco": null,
        "address": null,
        "parking": null,
        "s_descr": null,
        "website": "https://epirusforallseasons.gr/aksiotheata/peripetia/chionodromiko-kentro-pr-ilia/"
      }
    }
  ]
}
```

```

"category": ["05"],
"cultural": null,
"name_eng": null,
"shopping": null,
"periferia": null,
"protected": null,
"published": null,
"exhibition": null,
"full_descr": "Διαθέτει συνολικά 6 πίστες. Ειδικότερα: 2 πράσινες, μήκους 150 μ. και 350 μ., αντίστοιχα, 2 μπλε: η μία συνολικού μήκους 1200 μ., μπλε για 800 μ. & πράσινη για 400 μ. και η άλλη συνολικού μήκους 900 μ., μπλε για 600 μ. & κόκκινη για 300 μ.1 κόκκινη, κυρίως, συνολικού μήκους 750 μ. (κόκκινη για 550 μ. & μπλε για 200 μ.)1 πίστα δρόμου αντοχής (Lang-Lauf), μήκους 15 χλμ. ενώ υπάρχει και ένας επίπεδος χώρος 50 στρεμμάτων ιδανικός για snowmobile. Το Χιονοδρομικό Κέντρο Προφήτης Ηλίας διαθέτει σαλέ με εστιατόριο και καφέ, σχολές σκι, μεγάλο χώρο στάθμευσης, καταστήματα ενοικίασης σκι και snow truck (ένα όχημα σαν λεωφορείο, που αποτελείται από 2 βαγόνια και κάνει βόλτες τους επιβάτες στα χιόνια). Χιονοδρομικό Κέντρο – Πολιτιστές: Δ-Κ, όλο το χρόνο, Τηλ.: 26560-41095, 42630, website: http://www.metsovo-ski.gr, e-mail: info@metsovo-ski.gr. Τις καθημερινές το lift κοστίζει 7€ και τα Σαββατοκύριακα 10€.",
"restaurant": null,
"source_cat": "https://epirusforallseasons.gr",
"time_spent": null,
"amea_access": null,
"per_enotita": null,
"s_descr_eng": null,
"subcategory": [],
"ticket_cost": null,
"interchanges": null,
"accommodation": null,
"certification": null,
"info_material": null,
"observatories": null,
"difficulty_lvl": null,
"full_descr_eng": null,
"interchange_id": [],
"kallikratis_id": null,
"historic_period": null,
"outdoor_company": null,
"nearest_neighbour": null,
"outdoor_activities": null,
"recreational_areas": null
}}, {
"id": null,
"type": "Feature",
"geometry": {
"type": "Point",
"coordinates": [21.094235, 39.829983]
},
"properties": {
"wc": null,
"food": null,
"name": "Ορεινή Ποδηλασία στη Λίμνη Πηγών Αωού Ιωάννινα",
"dimos": null,
"email": null,
"image": "https://epirusforallseasons.gr/wp-content/plugins/wp-fastest-cache-premium/pro/images/blank.gif",
"is_ar": null,
"awards": null,
"refuge": null,
"routes": null,
"unesco": null,
"address": null,
"parking": null,
"s_descr": null,
"website": "https://epirusforallseasons.gr/aksiotheata/peripetia/orini-podilasia-sti-limni-pigon-aou/",
"category": ["05"],
"cultural": null,
"name_eng": null,
"shopping": null,
"periferia": null,
"protected": null,
"published": null,
"exhibition": null,

```

```

"full_descr": "Άρκει ένα ποδήλατο βουνού για να μυηθεί κανείς στα μυστικά της άγριας ζωής μέσα στα πιο όμορφα δάση της περιοχής γύρω από την λίμνη του Αώου. Η ποδηλατική περιήγηση περιμετρικά της λίμνης πηγών Αώου προσφέρει μοναδικές δυνατότητες οικοξενάγησης. Στη διαδρομή, περίπου 32 χλμ, ξεχωρίζουν τα γέρικα μαυρόπευκα, αλλά κι η πλούσια πανίδα και χλωρίδα, που περιλαμβάνει αντίστοιχα από αγριόκοτες, βασιλαετούς κι αρκούδες έως τα φαρμακευτικά φυτά και τα μανιτάρια της περιοχής. Video",
"restaurant": null,
"source_cat": "https://epirusforallseasons.gr",
"time_spent": null,
"amea_access": null,
"per_enotita": null,
"s_descr_eng": null,
"subcategory": [],
"ticket_cost": null,
"interchanges": null,
"accommodation": null,
"certification": null,
"info_material": null,
"observatories": null,
"difficulty_lvl": null,
"full_descr_eng": null,
"interchange_id": [],
"kallikratis_id": null,
"historic_period": null,
"outdoor_company": null,
"nearest_neighbour": null,
"outdoor_activities": null,
"recreational_areas": null
}
}, {
"id": null,
"type": "Feature",
"geometry": {
"type": "Point",
"coordinates": [20.745055, 39.973103]
},
"properties": {
"wc": null,
"food": null,
"name": "Διασχίζοντας το Εθνικό Πάρκο Βίκου - Αώου Ιωάννινα",
"dimos": null,
"email": null,
"image": "https://epirusforallseasons.gr/wp-content/plugins/wp-fastest-cache-premium/pro/images/blank.gif",
"is_ar": null,
"awards": null,
"refuge": null,
"routes": null,
"unesco": null,
"address": null,
"parking": null,
"s_descr": null,
"website": "https://epirusforallseasons.gr/aksiotheata/peripetia/pezoporia-sto-ethniko-parko-vikou-aou/",
"category": ["05"],
"cultural": null,
"name_eng": null,
"shopping": null,
"periferia": null,
"protected": null,
"published": null,
"exhibition": null,
"full_descr": "Το φαράγγι του Βίκου "σχιζει" κέθετα τα ασβεστολιθικά υψίπεδα της Τύμφης για 20 χιλιόμετρα, διαχωρίζοντας τα χωριά της Δυτικού και Κεντρικού Ζαγορίου. Τα τοιχώματα του φαραγγιού φθάνουν τα 1000μ σε ύψος. Η διάσχιση του φαραγγιού με τα πόδια είναι ίσως το αποκορύφωμα της επίσκεψης σας στο Ζαγόρι. Το φαράγγι είναι ιδιαίτερα γνωστό στους λάτρεις της πεζοπορίας, αλλά παρά τις τακτικές προσπάθειες συντήρησης των μονοπατιών του, η διάσχισή του δεν είναι τόσο απλή όσο ακούγεται. Χρειάζεται γι'αυτό το λόγο να είστε εξοπλισμένοι με σωστή υπόδηση (μπότες πάνω από τον αστράγαλο) και κατάλληλη ένδυση. Χρειάζεται επίσης να έχετε μαζί σας δοχεία γεμάτα με νερό και πιθανά νερού και ορειβατικά μπατόν. Από τις εκδόσεις Ανάβαση μπορείτε να προμηθευτείτε τους καλύτερους χάρτες για την περιοχή. Ελέγξτε τις κλιματικές συνθήκες πριν την επίσκεψή σας. Κατά τη διάρκεια του Απριλίου ή στις αρχές Μαΐου, τα χιόνια που λιώνουν κάνει συχνά το πέρασμα στο Μονοδένδρι

```

```
αδιάβατο, και κατά τη διάρκεια βροχοπτώσεων οι πλευρές του φαραγγιού είναι επιρρεπείς σε
κατολισθήσεις. Video",
  "restaurant": null,
  "source_cat": "https://epirusforallseasons.gr",
  "time_spent": null,
  "amea_access": null,
  "per_enotita": null,
  "s_descr_eng": null,
  "subcategory": [],
  "ticket_cost": null,
  "interchanges": null,
  "accommodation": null,
  "certification": null,
  "info_material": null,
  "observatories": null,
  "difficulty_lvl": null,
  "full_descr_eng": null,
  "interchange_id": [],
  "kallikratis_id": null,
  "historic_period": null,
  "outdoor_company": null,
  "nearest_neighbour": null,
  "outdoor_activities": null,
  "recreational_areas": null
}
```

Σημεία ενδιαφέροντος μορφής geoJSON

2.2 Οντολογία

Η χρήση οντολογιών στα πλαίσια του e-ΧΝΗΛΑΤΗ εξυπηρετεί σκοπούς που απαιτούν αποθήκευση, διαχείριση, επεξεργασία και συλλογιστική δεδομένων και ευρύτερης γνώσης με πλούσια σημασιολογία. Στα πλαίσια αυτών των αναγκών, αναπτύσσεται κι ανανεώνεται σε τακτά χρονικά διαστήματα η οντολογία που υποστηρίζει όλα τα προαναφερθέντα.

Περαιτέρω πληροφορίες, ανάλυση και επεξήγηση αναφορικά με τα σημασιολογικά μοντέλα και τις τεχνικές σημασιολογικού συσχετισμού για τη δημιουργία διαδρομών ενδείκνυται το παραδοτέο Π2.2.2 «Σημασιολογικά μοντέλα και τεχνικές σημασιολογικού συσχετισμού για την δημιουργία διαδρομών».

3. ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ

Σε αυτήν την ενότητα παρουσιάζεται η συλλογή των νέων συνόλων δεδομένων που δημιουργήθηκαν χρησιμοποιώντας τους crawlers του e-ΧΝΗΛΑΤΗ. Οι πόροι που συλλέχθηκαν καλύπτουν τις ανάγκες συλλογής δεδομένων που έχουν καθοριστεί από την αρχή του έργου.

3.1 Σύνολα δεδομένων από epirusforallseasons.gr

Η ιστοσελίδα του “epirusforallseasons” πρόκειται για μια επίσημη πηγή πληροφοριών από τη περιφέρεια Ηπείρου. Περιέχει υλικό και πόρους για πολλά μνημεία, δραστηριότητες και σημεία ενδιαφέροντος από τη περιφερειακή ενότητα Ιωαννίνων, Άρτας, Πρέβεζας και Θεσπρωτίας. Τα σημεία ενδιαφέροντος αποτελούνται από:

- Φυσικού ενδιαφέροντος
- Πολιτισμικού ενδιαφέροντος
- Ήλιος και θάλασσα
- Θρησκευτική κληρονομιά
- Πόλεις και χωριά
- Τέχνη και πολιτισμός

Πολύ σημαντικό κριτήριο επιλογής για την ανάκτηση πληροφορίας από τον ιστότοπο και τη παραγωγή συνόλου δεδομένων είναι η παροχή γεωχωρικών συντεταγμένων, γραπτών πληροφοριών και πολυμεσικού υλικού που μπορεί να χρησιμοποιηθεί για τη δημιουργία και την προβολή σημείων ενδιαφέροντος στην ακτίνα δράσης του e-ΧΝΗΛΑΤΗ κατά την πρόταση εξατομικευμένων μοναδικών δυναμικών διαδρομών. Ακόμη από τον τρόπο που είναι δομημένη η ιστοσελίδα και τα σημεία ενδιαφέροντος μπορούμε να ανακτήσουμε αυτόματα και την θεματική ενότητα που ανήκουν τα σημεία και να την αντιστοιχίσουμε με την αντίστοιχη που χρησιμοποιείται από την πλατφόρμα.

Για τη συλλογή των δεδομένων από την ιστοσελίδα “epirusforallseasons.gr”, χρησιμοποιήθηκε ως βάση το εργαλείο του website crawler που παρουσιάστηκε στη πρώτη έκδοση του παραδοτέου (Π2.1.1) και τροποποιήθηκε καταλλήλα για την υλοποίηση ενός εξειδικευμένου crawler του ιστοτόπου. Αρχικά αναλύθηκε η δομή της ιστοσελίδας και ξεχωρίστηκαν οι θεματικές ενότητες για τη στοχευμένη εξόρυξη πληροφορίας (focused crawling). Στη συνέχεια βρέθηκαν οι συνδέσεις μεταξύ των σελίδων και ορίστηκαν οι διευθύνσεις (URL’s) και οι επιλογείς (CSS selectors).

Παρακάτω παρουσιάζεται παράδειγμα του κώδικα που αναπτύχθηκε για τη συλλογή των δεδομένων από την ιστοσελίδα:

- εγγραφής για ένα συγκεκριμένο σημείο ενδιαφέροντος όπως περισυλλέχθηκε απευθείας με την ημι-αυτόματη χρήση του λογισμικού
- τελικής μορφής του σημείου ενδιαφέροντος συμβατής με τη βάση δεδομένων και τη γνωσιακή βάση

```

{
  "url": {
    "base_url": "https://epirusforallseasons.gr"
  },
  "group_of_urls": ["/aksiotheata/fysi/",
"/aksiotheata/politismos/", "/aksiotheata/peripetia/", "/aksiotheata/ilios-
thalassa/", "/aksiotheata/thriskia/"],
  "source_name": "epirus",
  "table_selector": "#attraction-type-wrapper div.single-attarction-wrapper div div
div.route-block-wrapper div.search-cards",
  "entity_name": "sights",
  "entity_info": [{
    "label": "sight_title",
    "value": {
      "selector": "div > a > h3",
      "type": "text"
    }
  },
  {
    "label": "img",
    "value": {
      "selector": "div > a > figure > img",
      "type": "image"
    }
  }
],
  "metrics": [
    {
      "label": "crawl_to",
      "value": {
        "selector": "div > a",
        "type": "link"
      }
    }
  ],
  "crawl": {
    "metrics": [
      {
        "label": "full_title",
        "value": {
          "selector": "div > div > div > div.area-banner > div >h1",
          "type": "text"
        }
      },
      {
        "label": "full_content",
        "value": {
          "selector": "div > div > div > div.single-attarction-wrapper.area-article-
wrapper > div > div",
          "type": "text"
        }
      }
    ]
  },
  "store": {
    "format": "json",
    "hd_path": "/home/tassoskarag/projects/crawler_inxilatis/results/epirus_result.json"
  },
  "dynamic_page": true
}

```

Αρχείο συστήματος για συλλογή δεδομένων από epirusforallseasons.gr

```

"sights": {
  "sight_title": "Χιονοδρομικό Κέντρο Προφήτη Ηλία",
  "img": "https://epirusforallseasons.gr/wp-content/uploads/2019/12/05_02_003_2-768x509.jpg"
},
"metrics": [
  {
    "name": "full_title",
    "citeyear": 2022,
    "type": "textual",
    "value": "Χιονοδρομικό Κέντρο Προφήτη Ηλία Ιωάννινα",
    "source": "https://epirusforallseasons.gr/aksiotheata/peripetia/chionodromiko-kentro-pr-ilia/"
  },
  {
    "name": "full_content",
    "citeyear": 2022,
    "type": "textual",
    "value": "Διαθέτει συνολικά 6 πίστες. Ειδικότερα: 2 πράσινες, μήκους 150 μ. και 350 μ., αντίστοιχα, 2 μπλε: η μία συνολικού μήκους 1200 μ., μπλε για 800 μ. & πράσινη για 400 μ. και η άλλη συνολικού μήκους 900 μ., μπλε για 600 μ. & κόκκινη για 300 μ.1 κόκκινη, κυρίως, συνολικού μήκους 750 μ. (κόκκινη για 550 μ. & μπλε για 200 μ.)1 πίστα δρόμου αντοχής (Lang-Lauf), μήκους 15 χλμ. ενώ υπάρχει και ένας επίπεδος χώρος 50 στρεμμάτων ιδανικός για snowmobile. Το Χιονοδρομικό Κέντρο Προφήτη Ηλίας διαθέτει σαλέ με εστιατόριο και καφέ, σχολές σκι, μεγάλο χώρο στάθμευσης, καταστήματα ενοικίασης σκι και snow truck (ένα όχημα σαν λεωφορείο, που αποτελείται από 2 βαγόνια και κάνει βόλτες τους επιβάτες στα χιόνια). Χιονοδρομικό Κέντρο – Πολιτιστές: Δ-Κ, όλο το χρόνο, Τηλ.: 26560-41095, 42630, website: http://www.metsovo-ski.gr, e-mail: info@metsovo-ski.gr. Τις καθημερινές το lift κοστίζει 7€ και τα Σαββατοκύριακα 10€.",
    "source": "https://epirusforallseasons.gr/aksiotheata/peripetia/chionodromiko-kentro-pr-ilia/"
  },
  {
    "name": "lat",
    "type": "textual",
    "value": 39.785833
  },
  {
    "name": "long",
    "type": "textual",
    "value": 21.158333
  }
]

```

Αρχικά δεδομένα από epirusforallseasos.gr

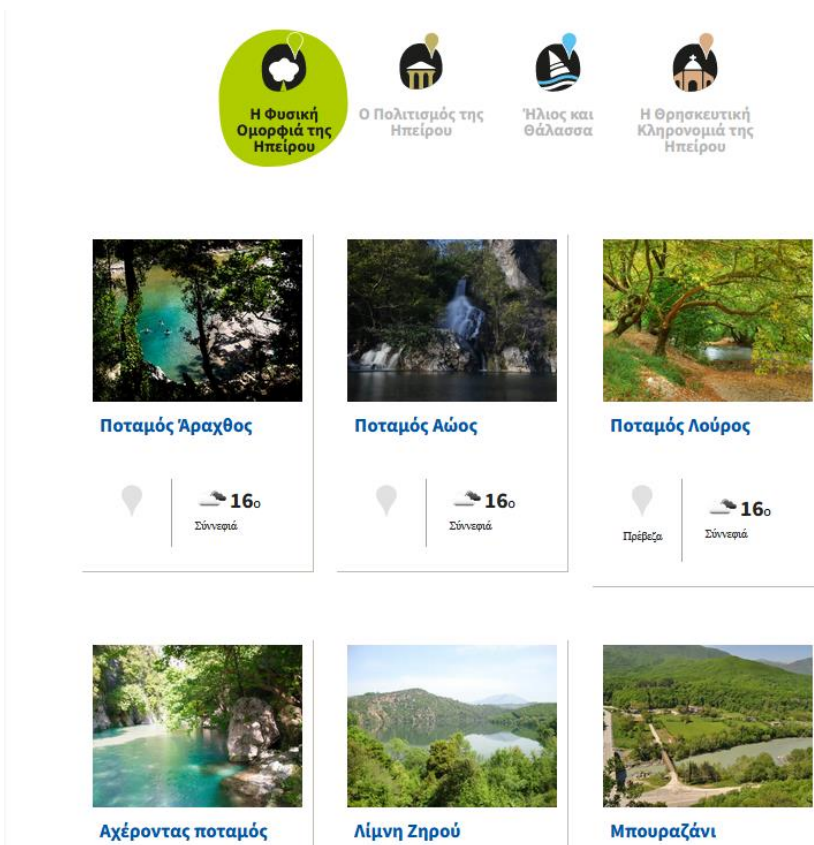
```

{
  "id": null,
  "type": "Feature",
  "geometry": {
    "type": "Point",
    "coordinates": [20.273927, 39.386663]
  },
  "properties": {
    "wc": null,
    "food": null,
    "name": "Παραλία Αγίας Παρασκευής Θεσπρωτία",
    "dimos": null,
    "email": null,
    "image": "https://epirusforallseasons.gr/wp-content/plugins/wp-fastest-cache-premium/pro/images/blank.gif",
    "is_ar": null,
    "awards": null,
    "refuge": null,
    "routes": null,
    "unesco": null,
    "address": null,

```


```
"parking": null,
"s_descr": null,
"website": "https://epirusforallseasons.gr/aksiotheata/ilios-thalassa/paralia-agias-
paraskevis/",
"category": ["08"],
"cultural": null,
"name_eng": null,
"shopping": null,
"periferia": null,
"protected": null,
"published": null,
"exhibition": null,
"full_descr": "Απέχει 9 χλμ. από το χωριό της Πέρδικας και βρίσκεται στον δρόμο προς
τα Σύβοτα. Γραφική και γαλήνια, με χοντρή άμμο και βότσαλα, καθαρά κρυστάλλινα και
βαθιά νερά, προσφέρει εικόνες «καρτιποσταλικής» ομορφιάς, καθώς την περιβάλλουν
ελαιόδεντρα, ενώ μπροστά της υπάρχει το καταπράσινο νησάκι της Αγίας Παρασκευής, όπου
μπορείτε να φτιάσετε κολυμπώντας. Διαθέτει καφέ και μπαρ.",
"restaurant": null,
"source_cat": "https://epirusforallseasons.gr",
"time_spent": null,
"amea_access": null,
"per_enotita": null,
"s_descr_eng": null,
"subcategory": [],
"ticket_cost": null,
"interchanges": null,
"accommodation": null,
"certification": null,
"info_material": null,
"observatories": null,
"difficulty_lvl": null,
"full_descr_eng": null,
"interchange_id": [],
"kallikratis_id": null,
"historic_period": null,
"outdoor_company": null,
"nearest_neighbour": null,
"outdoor_activities": null,
"recreational_areas": null
}
```

Προσαρμοσμένα δεδομένα στη δομή της βάσης δεδομένων




Εικόνα 1: Σημεία ενδιαφέροντος ανά ενότητα σε ιστοσελίδα epirusforallseasons.gr

Στην Ήπειρο > Αξιοθέατα > Ποταμός Άραχθος




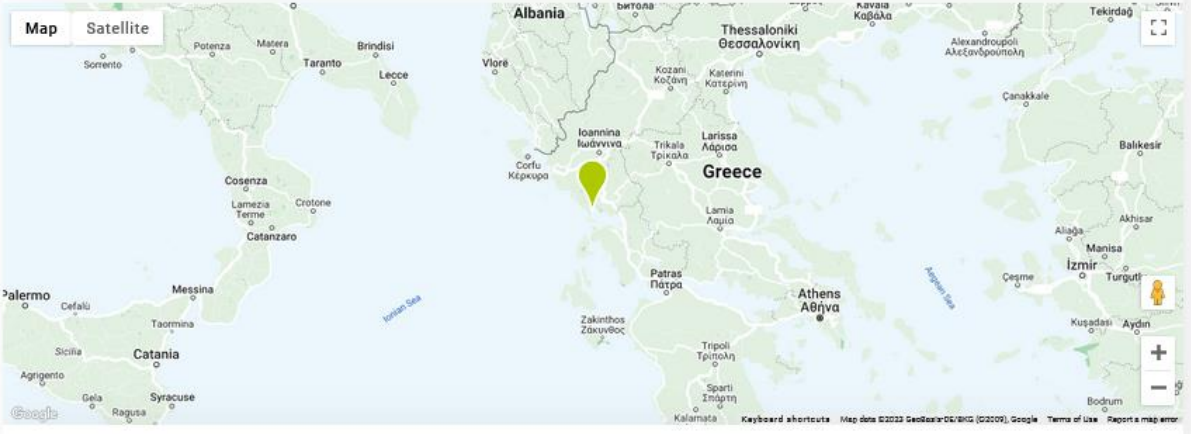
Ποταμός Άραχθος



Ο Άραχθος είναι ένας από τους εντυπωσιακότερους ποταμούς της Ηπείρου και ο όγδοος μεγαλύτερος στην Ελλάδα! Το ταξίδι του ξεκινά από το όρος Μαυροβούνι κοντά στο Μέτσοβο, στη διαδρομή του διέρχεται ορμητικός από φαράγγια άγριας ομορφιάς, έχοντας τα Τζουμέρκα από τη μία πλευρά και το Ξηροβούνι από την άλλη, μέχρι την τεχνητή λίμνη Πουρναρίου. Από εκεί τα νερά του κυλούν πιο ήρεμα πια μέχρι να καταλήξει στον Αμβρακικό κόλπο, δημιουργώντας στο διάβα του μοναδικές λιμνοθάλασσες (Λογαρού, Ροδιά και Τσουκαλιό) οι οποίες αποτελούν τμήμα της ιδιαίτερα σημαντικής προστατευόμενης περιοχής Υγροτόπων Αμβρακικού. Στα βόρεια του ξεκινά το Εθνικό Πάρκο Βόρειας Πίνδου, ενώ εντάσσεται μέσα στο Εθνικό Πάρκο Τζουμέρκων, Περιστερίου και χαράδρας Άραχθου. Στην αρχαιότητα, ο Άραχθος, ήταν γνωστός με το όνομα Ίναχος, ενώ από το 1881 έως το 1912, αποτελούσε το φυσικό σύνορο μεταξύ Ελλάδας και Τουρκίας. Αυτό που τον κάνει να ξεχωρίζει είναι ότι κατά μήκος του συναντάμε 55 πέτρινα γεφύρια, με πιο γνωστά το Γεφύρι της Πλάκας, το μεγαλύτερο μονότοξο γεφύρι των Βαλκανίων και το ξακουστό Γεφύρι της Άρτας. Το συνολικό μήκος του ποταμού φτάνει τα 150 χλμ., με το μήκος του κυρίως ποταμού να φτάνει 110 χλμ. Από τους σημαντικότερους παραπόταμους του Αράχθου είναι ο Καλλαρύτικος, ο Μετσοβίτικος, ο Μελισσουργιώτικος κι ο Βάρδας. Από τα ομορφότερα τοπία που συναντά κανείς στο στενό φαράγγι του Αράχθου με τις κατακόρυφες ορθοπλαγιές ύψους 700μ., είναι ο εντυπωσιακός καταρράκτης της Κλίφκης, τα νερά του οποίου σχηματίζουν μια πανέμορφη λίμνη, ενώ δεκάδες πηγές αναβλύζουν από τους βράχους με πεντακάθαρο νερό.

Στο μεγαλύτερο τμήμα του, ο Άραχθος καλύπτεται από μεγάλα παρόχθια δάση με ιτιές, λεύκες, πλατάνια, σκλήθρα, σφενδάμια και πικροκαστανιές, ενώ λίγο πιο ψηλά στις πηγές του, συναντάμε ελατοδάση.

Ο Άραχθος, αποτελεί ιδανική επιλογή για καταβάσεις με rafting και canoe kayak.

Εικόνα 2: Λεπτομέρειες σημείου ενδιαφέροντος με χωρικές πληροφορίες σε ιστοσελίδα epirusforallseasons.gr

3.2 Σύνολα δεδομένων από Πολιτιστικό Ίδρυμα Ομίλου Πειραιώς (ΠΙΟΠ)

Το Πολιτιστικό Ίδρυμα Ομίλου Πειραιώς (ΠΙΟΠ) υποστηρίζει τη διάσωση και ανάδειξη της πολιτιστικής κληρονομιάς της χώρας με έμφαση στη βιοτεχνική και βιομηχανική τεχνολογία, και προωθεί τη σύνδεση του Πολιτισμού με το Περιβάλλον. Το ΠΙΟΠ συμμετέχει στο έργο του e-χνηλάτη, έχει ένα σύστημα θεματικών μουσείων με δυο από αυτά να βρίσκονται στη περιοχή μελέτης του έργου και να αποτελούν τον χώρο διεξαγωγής των πιλοτικών δοκιμών του έργου. Τα μουσεία αναφοράς είναι το μουσείο Μετάξης στο Σουφλί και το μουσείο αργυροτεχνίας στα Ιωάννινα και αποτελέσαν σημεία αναφοράς, με πολύ εκτενή πληροφορία για όλα τα πεδία της βάσης δεδομένων.

Τα συγκεκριμένα σημεία προτέθηκαν με αναλυτικές εκτενείς πληροφορίες για να αποτελέσουν πρότυπο εισαγωγής για όλα τα σημεία ενδιαφέροντος της βάσης δεδομένων που αφορούν το έργο του e-ΧΝΗΛΑΤΗ.

The screenshot shows the website of the Cultural Foundation of the Piraeus Group (ΠΙΟΠ). The main content is a map of Greece with several museum locations marked with icons and labels: Σουφλί, Ιωάννινα, Βόλος, Λέσβος, Χίος, Στυμφαλία, Τήνος, Δημητσάνα, and Σπάρτη. On the right side, there is a detailed information box for the 'Μουσείο Αργυροτεχνίας' (Silverware Museum) in Ioannina, including its address (Acropolis of Itea, Kalé, Káastro, Ioannína), contact numbers (Tel: 26510 64065, Fax: 26510 32012), and opening hours (March 1st - October 15th, daily 10:00-18:00; October 16th - February 28th, daily 10:00-17:00). The ticket prices are listed as 4 euros for regular and 2 euros for reduced. A 'Μάθε Περισσότερα' (Learn More) button is also visible.

Εικόνα 3: Θεματικά μουσεία ΠΙΟΠ

3.3 Σύνολα δεδομένων από OpenStreetMaps

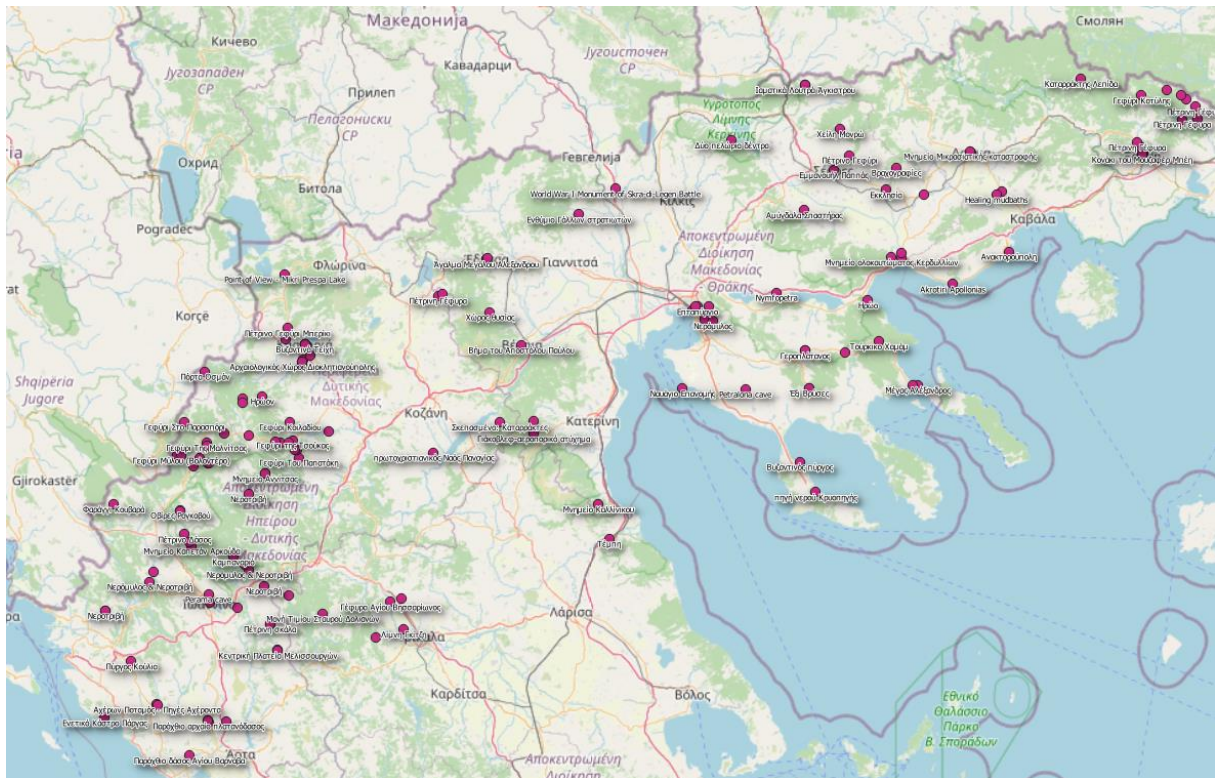
Το OpenStreetMap (OSM)¹ είναι ένας χάρτης με ελεύθερη άδεια ο οποίος αναπτύσσεται από μια κοινότητα εθελοντών που συνεισφέρουν και διατηρούν δεδομένα σχετικά με δρόμους και διάφορα σημεία ενδιαφέροντος σε όλο τον κόσμο. Το OpenStreetMap έχει ελεύθερα δεδομένα που διατίθενται με άδεια Open Data Commons Open Database License (ODbL) Τα δεδομένα διατίθενται με την άδεια Creative Commons Attribution-ShareAlike 2.0 license (CC

¹ <https://download.geofabrik.de/europe/greece.html>

BY-SA). Το OpenStreetMap περιέχει πληθώρα δεδομένων που μπορούν να αξιοποιηθούν από το σύστημα του e-χνηλατη. Για την περίπτωση του έργου κρατήσαμε τα σημεία ενδιαφέροντος από της περιοχής μελέτης και φιλτράραμε με βάση τις θεματικές ενότητες τα σημεία που σχετίζονται με το έργο: α) φυσικού περιβάλλοντος, β) πολιτιστικού ενδιαφέροντος γ) οικιστικά σύνολα. Μετά τη συλλογή τους έγινε η κατάλληλη μετατροπή στη μορφή της βάσης δεδομένων για να ενταχθούν στο σύστημα.

```
{
  "type": "Feature",
  "properties": {
    "wc": null,
    "food": null,
    "name": "Επιταύργιο",
    "dimos": null,
    "email": null,
    "is_ar": null,
    "awards": null,
    "refuge": null,
    "routes": null,
    "unesco": null,
    "address": null,
    "parking": null,
    "s_descr": null,
    "website": null,
    "category": null,
    "cultural": null,
    "name_eng": null,
    "shopping": null,
    "periferia": null,
    "protected": null,
    "published": null,
    "telephone": null,
    "exhibition": null,
    "full_descr": null,
    "restaurant": null,
    "source_cat": "OpenStreetMap",
    "time_spent": null,
    "amea_access": null,
    "per_enotita": null,
    "s_descr_eng": null,
    "subcategory": null,
    "ticket_cost": null,
    "interchanges": null,
    "accommodation": null,
    "certification": null,
    "info_material": null,
    "observatories": null,
    "difficulty_lvl": null,
    "full_descr_eng": null,
    "interchange_id": null,
    "kallikratis_id": null,
    "historic_period": null,
    "outdoor_company": null,
    "nearest_neighbour": null,
    "outdoor_activities": null,
    "recreational_areas": null,
    "work_hours": null,
    "image": null
  },
  "geometry": {
    "type": "Point",
    "coordinates": [22.9620398, 40.644387]
  }
}
```

Παράδειγμα εγγραφής από OpenStreetMap



Εικόνα 4: Συλλεγμένα σημεία ενδιαφέροντος από OpenStreetMaps

4. ΣΥΜΠΕΡΑΣΜΑΤΑ & ΕΠΟΜΕΝΑ ΒΗΜΑΤΑ

Σε αυτό το παραδοτέο, παρουσιάστηκε η τελική έκδοση των μεθόδων που αναπτύχθηκαν για τους crawlers του e-ΧΝΗΛΑΤΗ, ένα γενικό πλαίσιο που ενσωματώνει και συνδέει καταληκτικά ένα σύνολο διαφορετικών πόρων βασισμένο στο μόντουλο του scrapping περιεχομένου. Η πρώτη έκδοσή που παρουσιάστηκε στο παραδοτέο Π2.1.1 είχε πολύ καλά αποτελέσματα και σύλλεξε δεδομένα για να χρησιμοποιηθούν από τα υπόλοιπα εργαλεία του συστήματος. Σε αυτή την έκδοσή προτέθηκαν παραπάνω πηγές για άντληση περισσότερων δεδομένων στη περιοχή μελέτης του έργου που θα αποθηκευτούν για να χρησιμοποιηθούν από το σύστημα του e-ΧΝΗΛΑΤΗ.

Ο στόχος για τον οποίο αναπτύχθηκε το συγκεκριμένο πλαίσιο για το έργο είναι η συλλογή χρήσιμου πολυμεσικού περιεχομένου και μεταδεδομένων που μπορούν εν συνεχεία να αποθηκευτούν και να συμπεριληφθούν σε αλγορίθμους δρομολόγησης, προσωποποίησης και συλλογιστικής από τις υποενότητες ΥΕ2.3 και ΥΕ2.4. Πρόκειται για το πρώτο βήμα μιας ροής διεργασιών προ-επεξεργασίας που δημιουργεί περιεχόμενο που θα χρησιμοποιηθεί από την πλατφόρμα του έργου.

Ακόμη έγινε λόγος για τη δόμηση της μοντελοποίησης των δεδομένων και τη σημασία της, καθώς κάθε ξεχωριστή διαδικτυακή πηγή είχε εξίσου ξεχωριστά δομημένη τη ζητούμενη πληροφορία προς ανάκτηση. Εν συνεχεία, παρουσιάστηκαν τα τωρινά σύνολα δεδομένων που σχηματίστηκαν από τις διαδικασίες που περιγράφηκαν παραπάνω.

Το λογισμικό των crawlers του e-ΧΝΗΛΑΤΗ είναι διαθέσιμο στο παρακάτω σύνδεσμο: <https://github.com/tassosblackg/easIE/releases/tag/1.0.2> . Τα αρχεία ρυθμίσεων για το χειρισμό των crawler βρίσκονται στο παρακάτω σύνδεσμο: https://github.com/AKokkalas/etracer_crawlers.

ΑΝΑΦΟΡΕΣ

- Battle, R., & Kolas, D. (2011). Geosparql: enabling a geospatial semantic web. *Semantic Web Journal*, 3(4), 355-370.
- Gkatziaki, V., Papadopoulos, S., Mills, R., Diplaris, S., Tsampoulatidis, I., & Kompatsiaris, I. (2018). easIE: Easy-to-use information extraction for constructing CSR databases from the web. *ACM Transactions on Internet Technology (TOIT)*, 18(4), 1-21.
- Hobbs, J. R., & Pan, F. (2006). Time ontology in OWL. W3C working draft, 27, 133.
- Knublauch, H., & Kontokostas, D. (2017). Shapes Constraint Language (SHACL), W3C Recommendation. World Wide Web Consortium.
- Maria, G., Akrivi, K., Costas, V., George, L., & Constantin, H. (2007). Creating an ontology for the user profile: Method and applications. In *Proceedings AI* AI Workshop RCIS*.
- Motik, B., Grau, B. C., Horrocks, I., Wu, Z., Fokoue, A., & Lutz, C. (2009). OWL 2 web ontology language profiles. W3C recommendation, 27, 61.
- Olston, C., & Najork, M. (2010). Web crawling. *Foundations and Trends in Information Retrieval*, 4(3), 175-246.
- Paliokas I, Patenidis AT, Mitsopoulou EE, Tsita C, Pehlivanides G, Karyati E, Tsafaras S, Stathopoulos EA, Kokkalas A, Diplaris S, Meditskos G, Vrochidis S, Tasiopoulou E, Riggas C, Votis K, Kompatsiaris I, Tzouvaras D. A Gamified Augmented Reality Application for Digital Heritage and Tourism. *Applied Sciences*. 2020; 10(21):7868. <https://doi.org/10.3390/app10217868>
- Shaw, R., Troncy, R., & Hardman, L. (2009, December). Lode: Linking open descriptions of events. In *Asian semantic web conference* (pp. 153-167). Springer, Berlin, Heidelberg.
- Stathopoulos, E. A., Kokkalas, A., Mitsopoulou, E. E., Patenidis, A. T., Meditskos, G., Diplaris, S., ... & Kompatsiaris, I. (2020, June). Knowledge-Based Management and Reasoning on Cultural and Natural Touristic Routes. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 355-367). Springer, Cham.
- Stathopoulos, E. A., Paliokas, I., Meditskos, G., Diplaris, S., Tsafaras, S., Valkouma, E., ... & Tzouvaras, D. (2019, October). Smart Discovery of Cultural and Natural Tourist Routes. In *IEEE/WIC/ACM International Conference on Web Intelligence-Companion Volume* (pp. 208-214).

ΠΑΡΑΡΤΗΜΑ

Παράρτημα Α: Παράδειγμα Αρχείου διαμόρφωσης για την ιστοσελίδα της Ιεράς Αρχιεπισκοπής Αθηνών:

```
{
  "url": {
    "base_url": "http://www.religiousgreece.gr"
  },
  "group_of_urls": ["/web/guest/west-central-macedonia", "/web/guest/east-macedonia-thrace"],
  "source_name": "religiousgreece",
  "table_selector": "#dv-top-attr-container div.dv-carousel div.owl-item",
  "company_info": [
    {
      "label": "name",
      "value": {
        "selector": "span.dv-title",
        "type": "text"
      }
    }
  ],
  "metrics": [
    {
      "label": "crawl_to",
      "value": {
        "selector": "div.dv-content > a:nth-child(1)",
        "type": "link"
      }
    }
  ],
  "crawl": {
    "metrics": [{
      "label": "full_descr",
      "value": {
        "selector": ".dv-content-area",
        "type": "text"
      }
    }
  ],
  {
    "label": "lat",
    "value": {
      "selector": "a.dv-widget-link",
      "type": "href",
      "replace": {
        "regex": [".*@", " ,[^,]*z", ".*,", ""],
        "with": ["", "", ""]
      }
    }
  }
  ,
  {
    "label": "lng",
    "value": {
      "selector": "a.dv-widget-link",
      "type": "href",
      "replace": {
        "regex": [".*@", " ,[^,]*z", ".*,", ""],
        "with": ["", "", ""]
      }
    }
  }
  ,
  {
    "label": "Info",
    "value": {
      "selector": ".dv-widget-content:eq(1) li",
      "type": "text"
    }
  }
}
```



```

    }
  ]
},
"store": {
  "format": "json",
  "hd_path": "D:/Documents/CERTH/Crawler easie/Religious
Greece/results_group_url_religious.json"
},
"dynamic_page": true
}

```

Πηγαίος κώδικας StaticWrapperGenerator:

```

public class StaticWrapperGenerator extends WrapperGenerator {
public StaticWrapperGenerator(Configuration configuration) {
    super(configuration);
}

@Override
public void execute() throws InterruptedException, Exception {
    if (configuration.url.relative_url != null) {
        StaticHTMLExtractor wrapper = new StaticHTMLExtractor(
            configuration.url.base_url,
            configuration.url.relative_url
        );

        if (configuration.next_page_selector != null) {
            PaginationIterator pagination = new PaginationIterator(
                configuration.url.base_url,
                configuration.url.relative_url,
                configuration.next_page_selector
            );
            extraction_handler.execute(pagination, configuration);
        } else {
            extraction_handler.execute(wrapper, configuration);
        }
    } else if (configuration.group_of_urls != null && configuration.url.base_url != null &&
configuration.url.relative_url == null) {
        if (configuration.next_page_selector != null) {
            throw new Exception("Pagination can not be conducted in Group Wrapper mode");
        }
        GroupHTMLExtractor wrapper = new GroupHTMLExtractor(configuration.group_of_urls,
configuration.url.base_url);
        extraction_handler.execute(wrapper, configuration);
    }
}
}
}

```

Πηγαίος κώδικας DynamicWrapperGenerator:

```

public class DynamicWrapperGenerator extends WrapperGenerator {
private String ChromeDriverPath;
public DynamicWrapperGenerator(Configuration configuration, String ChromeDriverPath) {
    super(configuration);
    this.ChromeDriverPath = ChromeDriverPath;
}

@Override
public void execute() throws InterruptedException, Exception {
    if (configuration.url.relative_url != null) {
        DynamicHTMLExtractor wrapper = new DynamicHTMLExtractor(
            configuration.url.base_url,
            configuration.url.relative_url,
            ChromeDriverPath
        );
    }
}
}

```

```

        if (configuration.events != null) {
            this.execute_events(wrapper);
        }else{
            extraction_handler.execute(wrapper, configuration);
        }

        wrapper.browser_emulator.close();
    } else if (configuration.group_of_urls != null && configuration.url.base_url != null &&
configuration.url.relative_url == null) {
        HashSet<String> group_of_urls = configuration.group_of_urls;

        Iterator<String> it = group_of_urls.iterator();
        while (it.hasNext()) {
            String current_url = it.next().replace(configuration.url.base_url, "");

            DynamicHTMLExtractor wrapper = new DynamicHTMLExtractor(
                configuration.url.base_url,
                current_url,
                ChromeDriverPath
            );

            if (configuration.events != null) {
                this.execute_events(wrapper);
            }else{
                extraction_handler.execute(wrapper, configuration);
            }

            wrapper.browser_emulator.close();
        }
    }

    } else {
        throw new Exception("relative_url is not defined");
    }
}

private void execute_events(DynamicHTMLExtractor wrapper) throws InterruptedException,
Exception {
    if (configuration.events instanceof ArrayList) {
        ArrayList<Configuration.Event> list_of_events = (ArrayList<Configuration.Event>)
configuration.events;
        for (int i = 0; i < list_of_events.size(); i++) {
            if (list_of_events.get(i).type.equals("CLICK")) {
                int times_to_repeat = list_of_events.get(i).times_to_repeat;
                for (int j = 0; j < times_to_repeat; j++) {
                    wrapper.browser_emulator.clickEvent(list_of_events.get(i).selector);
                }
                extraction_handler.execute(wrapper, configuration);
            } else if (list_of_events.get(i).type.equals("SCROLL_DOWN")) {
                int times_to_repeat = list_of_events.get(i).times_to_repeat;
                for (int j = 0; j < times_to_repeat; j++) {
                    wrapper.browser_emulator.scrollDownEvent();
                }
                extraction_handler.execute(wrapper, configuration);
            }
        }
    }
    } else {
        Configuration.Event event = (Configuration.Event) configuration.events;
        System.out.println(event.extraction_type);
        if (event.type.equals("CLICK")) {
            if (event.extraction_type.equals("AFTER_ALL_EVENTS")) {
                for (int i = 0; i < event.times_to_repeat; i++) {
                    wrapper.browser_emulator.clickEvent(event.selector);
                }
                extraction_handler.execute(wrapper, configuration);
            } else {
                extraction_handler.execute(wrapper, configuration);
                for (int i = 0; i < event.times_to_repeat; i++) {
                    wrapper.browser_emulator.clickEvent(event.selector);
                    extraction_handler.execute(wrapper, configuration);
                }
            }
        }
    }
} else {

```

```
        if (event.times_to_repeat != null) {
            wrapper.browser_emulator.scrollDownEvent(event.times_to_repeat);
        } else {
            wrapper.browser_emulator.scrollDownEvent();
        }
        extraction_handler.execute(wrapper, configuration);
    }
}
}
```

Παράδειγμα απόκρισης Ticketmaster Event Search API για ένα γεγονός

```
{
  "_embedded": {
    "events": [
      {
        "name": "Harry Styles | Platinum",
        "type": "event",
        "id": "Z698xZu0Za675",
        "test": false,
        "url": "https://www.ticketmaster.cz/event/harry-styles-%7C-platinum-
tickets/28939?language=en-us",
        "locale": "en-us",
        "images": [
          {
            "ratio": "3_2",
            "url": "https://s1.ticketm.net/dam/a/fc3/2d35621d-71b3-46c6-87af-
e5163f3c9fc3_1206431_TABLET_LANDSCAPE_3_2.jpg",
            "width": 1024,
            "height": 683,
            "fallback": false
          }
        ],
        { },
        { },
        { },
        { },
        { },
        { },
        { },
        { }
      ],
      "sales": {
        "public": {
          "startDateTime": "2019-11-22T10:44:00Z",
          "startTBD": false,
          "endDateTime": "2021-02-23T22:00:00Z"
        }
      },
      "dates": {
        "access": {
          "startDateTime": "2021-02-23T18:30:00Z",
          "startApproximate": false,
          "endApproximate": false
        }
      },
      "start": {
        "localDate": "2021-02-23",
        "localTime": "20:00:00",
        "dateTime": "2021-02-23T19:00:00Z",
        "dateTBD": false,
        "dateTBA": false,
        "timeTBA": false,
        "noSpecificTime": false
      }
    }
  }
}
```

```
},
  "timezone": "Europe/Prague",
  "status": {
    "code": "rescheduled"
  },
  "spanMultipleDays": false
},
  "classifications": [
    {
      "primary": true,
      "segment": {
        "id": "KZFzniwnSyZfZ7v7nJ",
        "name": "Music"
      }
    }
  ],
  "genre": {
    "id": "KnvZfZ7vAeA",
    "name": "Rock"
  },
  "subGenre": {
    "id": "KZazBEonSMnZfZ7v6F1",
    "name": "Pop"
  },
  "family": false
}
},
  "promoter": {
    "id": "923",
    "name": "LIVE NATION CZECH REPUBLIC s.r.o."
  },
  "promoters": [
    {
      "id": "923",
      "name": "LIVE NATION CZECH REPUBLIC s.r.o."
    }
  ]
},
  "_links": {
    "self": {
      "href": "/discovery/v2/events/Z698xZu0Za675?locale=en-us"
    }
  },
  "attractions": [
    {
      "href": "/discovery/v2/attractions/K8vZ9174XZ0?locale=en-us"
    }
  ]
},
  "venues": [
    {
      "href": "/discovery/v2/venues/Z598xZu0Zdv61?locale=en-us"
    }
  ]
},
  "_embedded": {
    "venues": [
      {
        "name": "O2 Arena",
        "type": "venue",
        "id": "Z598xZu0Zdv61",

```

```
        "test": false,
        "url": "https://www.ticketmaster.cz/venue/o2-arena-praha-9-
vstupenky/o2arena/105",
        "locale": "en-us",
        "images": [
            {
                "ratio": "16_9",
                "url": "https://media.ticketmaster.eu/czechrepublic/b18a88b1506370ac889a0bd8d26da3bb.jpg",
                "width": 205,
                "height": 115,
                "fallback": false
            }
        ],
        "postalCode": "19000",
        "timezone": "Europe/Prague",
        "city": {
            "name": "Praha 9"
        },
        "country": {
            "name": "Czech Republic",
            "countryCode": "CZ"
        },
        "address": {
            "line1": "Českomoravská 2345/17"
        },
        "location": {
            "longitude": "14.49377",
            "latitude": "50.10479"
        },
        "upcomingEvents": {
            "_total": 32,
            "mfx-cz": 32
        },
        "_links": {
            "self": {
                "href": "/discovery/v2/venues/Z598xZu0Zdv61?locale=en-us"
            }
        }
    },
    "attractions": [
        {
            "name": "Harry Styles",
            "type": "attraction",
            "id": "K8vZ9174XZ0",
            "test": false,
            "url": "http://www.ticketmaster.cz/artist/984352",
            "locale": "en-us",
            "externalLinks": {
                "youtube": [
                    {
                        "url": "https://www.youtube.com/channel/UCbOCbp5gXL8jigIBZLqMPrw"
                    }
                ]
            }
        }
    ],
    "twitter": [
        {
            "url": "https://twitter.com/Harry_Styles"
        }
    ]
}
```

```
],
    "itunes": [
        {
            "url": "https://itunes.apple.com/us/artist/id471260289"
        }
    ],
    "lastfm": [
        {
            "url": "https://www.last.fm/music/Harry%20Styles"
        }
    ],
    "wiki": [
        {
            "url": "https://en.wikipedia.org/wiki/Harry_Styles"
        }
    ],
    "facebook": [
        {
            "url": "https://www.facebook.com/harrystyles"
        }
    ],
    "instagram": [
        {
            "url": "https://instagram.com/harrystyles/"
        }
    ],
    "musicbrainz": [
        {
            "id": "7eb1ce54-a355-41f9-8d68-e018b096d427"
        }
    ],
    "homepage": [
        {
            "url": "http://hstyles.co.uk/"
        }
    ],
    ],
    "images": [
        {
            "ratio": "3_2",
            "url": "https://s1.ticketm.net/dam/a/fc3/2d35621d-71b3-46c6-87af-e5163f3c9fc3_1206431_TABLET_LANDSCAPE_3_2.jpg",
            "width": 1024,
            "height": 683,
            "fallback": false
        },
        { },
        { },
        { },
        { },
        { },
        { },
        { },
        { },
        { },
        { }
    ],
    "classifications": [
```

```
{
  "primary": true,
  "segment": {
    "id": "KZFzniwnSyZfZ7v7nJ",
    "name": "Music"
  },
  "genre": {
    "id": "KnyZfZ7vAev",
    "name": "Pop"
  },
  "subGenre": {
    "id": "KZazBEonSMnZfZ7vkE1",
    "name": "Pop Rock"
  },
  "type": {
    "id": "KZAyXgnZfZ7v7lt",
    "name": "Event Style"
  },
  "subType": {
    "id": "KZFzBErXgnZfZ7vA6J",
    "name": "Concert"
  },
  "family": false
},
"upcomingEvents": {
  "_total": 101,
  "tmr": 8,
  "mfx-de": 17,
  "mfx-es": 3,
  "mfx-at": 3,
  "mfx-dk": 4,
  "mfx-nl": 3,
  "ticketmaster": 57,
  "mfx-cz": 3,
  "mfx-pl": 3
},
"_links": {
  "self": {
    "href": "/discovery/v2/attractions/K8vZ9174XZ0?locale=en-us"
  }
}
},
{
  "name": "King Princess",
  "type": "attraction",
  "id": "K8vZ9179uW0",
  "test": false,
  "url": "http://www.ticketmaster.cz/artist/999960",
  "locale": "en-us",
  "externalLinks": {
    "musicbrainz": [
      {
        "id": "158778fe-55c1-4dd5-8a3a-9d1b968a5e52"
      }
    ]
  }
},
"images": [
  {
    "ratio": "3_2",
```

```
        "url": "https://s1.ticketm.net/dam/a/a42/56a913a9-ba02-4f7d-b8b1-
d68fae76da42_1223031_ARTIST_PAGE_3_2.jpg",
        "width": 305,
        "height": 203,
        "fallback": false
    },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { }
],
    "classifications": [
        {
            "primary": true,
            "segment": {
                "id": "KZFzniwnSyZfZ7v7nJ",
                "name": "Music"
            }
        },
        {
            "genre": {
                "id": "KnvZfZ7vAev",
                "name": "Pop"
            }
        },
        {
            "subGenre": {
                "id": "KZazBEonSMnZfZ7vkl1t",
                "name": "Pop"
            }
        },
        {
            "type": {
                "id": "KZAyXgnZfZ7v71a",
                "name": "Individual"
            }
        },
        {
            "subType": {
                "id": "KZFzBErXgnZfZ7vAd7",
                "name": "Musician"
            }
        },
        {
            "family": false
        }
    ]
},
    "upcomingEvents": {
        "_total": 6,
        "ticketmaster": 3,
        "mfx-cz": 3
    },
    "_links": {
        "self": {
            "href": "/discovery/v2/attractions/K8vZ9179uW0?locale=en-us"
        }
    }
}
}
}
]
}
},
    { },
    { },
```



```
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { },
    { }
]
},
  "_links":{
    "first":{
      "href":"/discovery/v2/events.json?page=0&size=20"
    },
    "self":{
      "href":"/discovery/v2/events.json"
    },
    "next":{
      "href":"/discovery/v2/events.json?page=1&size=20"
    },
    "last":{
      "href":"/discovery/v2/events.json?page=5918&size=20"
    }
  },
  "page":{
    "size":20,
    "totalElements":118375,
    "totalPages":5919,
    "number":0
  }
}
```

Παράδειγμα απόκρισης Open Weather One Call API για ένα σημείο

```
{
  "lat": 33.44,
  "lon": -94.04,
  "timezone": "America/Chicago",
  "timezone_offset": -18000,
  "current": {
    "dt": 1588935779,
    "sunrise": 1588936856,
    "sunset": 1588986260,
    "temp": 16.75,
    "feels_like": 16.07,
    "pressure": 1009,
    "humidity": 93,
    "dew_point": 15.61,
    "uvi": 8.97,
    "clouds": 90,
    "visibility": 12874,
    "wind_speed": 3.6,
    "wind_deg": 280,
```

```
"weather": [
  {
    "id": 501,
    "main": "Rain",
    "description": "moderate rain",
    "icon": "10n"
  },
  {
    "id": 200,
    "main": "Thunderstorm",
    "description": "thunderstorm with light rain",
    "icon": "11n"
  }
],
"rain": {
  "1h": 2.79
},
"minutely": [
  {
    "dt": 1588935780,
    "precipitation": 2.789
  },
  ...
],
"hourly": [
  {
    "dt": 1588935600,
    "temp": 16.75,
    "feels_like": 13.93,
    "pressure": 1009,
    "humidity": 93,
    "dew_point": 15.61,
    "clouds": 90,
    "wind_speed": 6.66,
    "wind_deg": 203,
    "weather": [
      {
        "id": 501,
        "main": "Rain",
        "description": "moderate rain",
        "icon": "10n"
      }
    ],
    "rain": {
      "1h": 2.92
    }
  },
  ...
]
"daily": [
  {
    "dt": 1588960800,
    "sunrise": 1588936856,
    "sunset": 1588986260,
    "temp": {
      "day": 22.49,
      "min": 10.96,
      "max": 22.49,
      "night": 10.96,
      "eve": 18.45,
      "morn": 18.14
    },
    "feels_like": {
      "day": 18.72,
      "night": 6.53,
      "eve": 16.34,
      "morn": 16.82
    },
    "pressure": 1014,
    "humidity": 60,
    "dew_point": 14.35,
    "wind_speed": 7.36,
    "wind_deg": 342,
```

```
"weather": [  
  {  
    "id": 502,  
    "main": "Rain",  
    "description": "heavy intensity rain",  
    "icon": "10d"  
  }  
],  
"clouds": 68,  
"rain": 15.38,  
"uvi": 8.97  
},  
...  
}
```