



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

<E-Commerce Scraper>

Έγγραφο Περιγραφής Απαιτήσεων Λογισμικού

Ιστορικό

Ημερομηνία	Έκδοση	Περιγραφή	Συγγραφέας
<15/04/24>	<1.1>	<προσθήκες/τροποποιήσεις/διαγραφές >	<Ομάδα Ανάπτυξης>

Σανίδα Μαρίνα
Καζάς Μηνάς
Αγαθαγγελίδου Ανθή
Αϊσόπουλος Δημήτριος

Απρίλιος, 2024
Κοζάνη

Πίνακας Περιεχομένων

Ιστορικό.....	1
Πίνακας Περιεχομένων.....	2
1. Εισαγωγή.....	4
1.1 Σκοπός.....	4
1.2 Γενική Άποψη.....	4
1.3 Ορισμοί, Ακρωνύμια και Συντομογραφίες.....	4
1.4 Αναφορές.....	5
1.5 Επισκόπηση.....	5
2. Γενική Περιγραφή.....	6
2.1 Προοπτική του Προϊόντος.....	6
2.2 Λειτουργίες του Προϊόντος.....	6
2.3 Χαρακτηριστικά Χρηστών.....	6
2.4 Περιορισμοί.....	6
3. Ειδικές Απαιτήσεις.....	8
3.1 Λειτουργικές Απαιτήσεις.....	8
3.1.1. Εξαγωγή Πληροφοριών.....	8
3.1.2. Παρακολούθηση Τιμών.....	8
3.1.3. Σύγκριση Τιμών.....	9
3.1.4.α. Αναζήτηση Με Φίλτρα Τοποθεσίας.....	10
3.1.4.β. Αναζήτηση Με Φίλτρα Τιμών.....	10
3.1.5. Σύγκριση Τιμών.....	11
3.1.6. Εμφάνιση Μέσης Τιμής.....	11
3.1.7. Εμφάνιση Υψηλότερης - Χαμηλότερης Τιμής.....	12
3.1.8. Εμφάνιση Κοντινότερης Τοποθεσίας και Αριθμού Προϊόντων.....	12
3.2 Μη Λειτουργικές Απαιτήσεις.....	13
4. Παραρτήματα.....	14

1. Εισαγωγή

1.1 Σκοπός

Σκοπός του εγγράφου είναι η καταγραφή των απαιτήσεων του scraper αγοραπωλησίας, το οποίο θα εξάγει πληροφορίες από διαδικτυακές ιστοσελίδες για να παρουσιάσει στον χρήστη τα δεδομένα που αναζητά με έναν οργανωμένο και συγκεντρωμένο τρόπο.

Το συγκεκριμένο λογισμικό απευθύνεται τόσο στον χρήστη που ενδιαφέρεται για την αγοραπωλησία προϊόντων, τόσο και στους διευθυντές εταιρειών που ενδιαφέρονται για τα πρόσφατα trends στον χώρο της αγοράς.

1.2 Γενική Άποψη

Μερικά από τα πλεονεκτήματα και οι λειτουργίες τα οποία θα προσφέρει ο E-CS είναι η σύγκριση τιμών, παρουσιάζοντας την καλύτερη τιμή προϊόντων από διάφορες ιστοσελίδες, για διάφορες κατηγορίες προϊόντων. Η αναζήτηση θα μπορεί να γίνει είτε βάσει τιμής είτε βάσει τοποθεσίας. Στην πρώτη περίπτωση ο χρήστης θα έχει τη δυνατότητα να αναζητήσει προϊόντα με βάση ενός εύρους τιμών, με ανώτατο ή κατώτατο όριο. Ο χρήστης θα μπορεί επίσης να αναζητήσει προϊόντα με βάση την τοποθεσία του, εμφανίζοντας προϊόντα τα οποία είναι διαθέσιμα στην περιοχή του.

Ένα τέτοιο λογισμικό θα είναι χρήσιμο για online αγορές, επιτρέποντας στους χρήστες να βρίσκουν τις καλύτερες δυνατές προσφορές για προϊόντα που τους ενδιαφέρουν, γλιτώνοντας έτσι χρόνο. Επιπλέον, η δυνατότητα αναζήτησης βάσει τοποθεσίας ή εντός ενός χιλιομετρικού εύρους, μπορεί να προσφέρει μια έξτρα ευκολία στους χρήστες, επιτρέποντάς τους να εντοπίζουν τοπικές προσφορές ή ακόμα και να εξοικονομούν επιπλέον χρήματα από τα έξοδα αποστολής.

1.3 Ορισμοί, Ακρωνύμια και Συντομογραφίες

- *E-CS*: E-Commerce Scraper
- *E-Commerce*: Αγοραπωλησία
- *Scraper*: Λογισμικό για απόσπαση περιεχομένου και δεδομένων από έναν ή πολλούς ιστότοπους.
- *URL*: Uniform Resource Locator - Ενιαίος Εντοπιστής Πόρων.
- *HTML*: HyperText Markup Language - Καθιερωμένη γλώσσα σήμανσης για ιστοσελίδες.

1.4 Αναφορές

1. McKeever, G. (2023, December 20). *What is scraping | About price & web scraping tools | Imperva*. Learning Center.

1.5 Επισκόπηση

Η σύνθεση του εγγράφου πραγματοποιήθηκε με την βοήθεια του Google Docs. Για την σχεδίαση των διαγραμμάτων έγινε χρήση του προγράμματος Visual Paradigm.

- Στην ενότητα **2. Γενική Περιγραφή** περιέχει πληροφορίες σχετικά με το λογισμικό και τους χρήστες του.
 - Η υποενότητα **2.1 Προοπτική του Προϊόντος** εξηγεί την αυτονομία του προϊόντος, και αν εξαρτάται από άλλα έργα.
 - Στην υποενότητα **2.2 Λειτουργίες του Προϊόντος** γίνεται αναφορά στις κυριότερες λειτουργίες του scraper.
 - Στην **2.3 Χαρακτηριστικά Χρηστών** απαριθμούνται οι διάφοροι χρήστες που μπορούν να χρησιμοποιήσουν το scraper, καθώς και οι προαπαιτούμενες γνώσεις που πρέπει να έχει κάθε ομάδα χρηστών.
 - Η υποενότητα **2.4 Περιορισμοί** εξηγεί τους περιορισμούς του υλικού, όπως και τις απαιτήσεις ασφάλειας.
 - Η υποενότητα **2.5 Παραδοχές** περιέχει τις παραδοχές από τις οποίες επηρεάζονται οι απαιτήσεις.
- Η ενότητα **3. Ειδικές Απαιτήσεις** περιέχει την λεπτομερή εξήγηση του λογισμικού και τις απαιτήσεις του.
 - Στην υποενότητα **3.1 Λειτουργικές Απαιτήσεις** περιγράφονται οι λειτουργικές απαιτήσεις υπό μορφή Περιπτώσεων Χρήσης.
 - Στις **3.2 Μη Λειτουργικές Απαιτήσεις** γίνεται αναφορά στις μη-λειτουργικές απαιτήσεις, όπως απαιτήσεις επίδοσης, ασφάλειας, και άλλα.

2. Γενική Περιγραφή

2.1 Προοπτική του Προϊόντος

Ο scraper που θα υλοποιήσουμε θα είναι ένα αυτόνομο σύστημα και δεν θα είναι μέρος ενός μεγαλύτερου συστήματος. Ειδικότερα ο Scraper είναι ανεξάρτητος από άλλα συστήματα και μπορεί να εκτελεστεί χωρίς κάποια εξωτερική εξάρτηση (μπορεί να τρέξει σε κάποιο υπολογιστή ή διακομιστή). Επιπλέον δεν θα κάνει χρήση άλλων υποκαταστημάτων για να λειτουργήσει καθώς θα συλλεγεί πληροφορίες από ιστοσελίδες αυτόνομα. Τέλος ο scraper μας λειτουργεί για συγκεκριμένες ιστοσελίδες (αγοραπωλησιών) και δεν μπορεί να χρησιμοποιηθεί σε κάποιο άλλο είδος site.

2.2 Λειτουργίες του Προϊόντος

Το E-CS θα επιτρέπει στον χρήστη να αναζητήσει προϊόντα αγοραπωλησίας από τρεις ιστότοπους (eBay, Skroutz, Car.gr).

Συγκεκριμένα θα έχει την δυνατότητα:

- Αναζήτησης συγκεκριμένου προϊόντος (με την χρήση λέξεων-κλειδιών)
- Φιλτραρίσματος για πιο εξειδικευμένες αναζητήσεις (συγκεκριμένη τοποθεσία, εύρος τιμών)
- Εμφάνισης στατιστικών και άλλων πληροφοριών (μέσος όρος τιμών, κοντινότερη τοποθεσία προϊόντος κ.α.)

2.3 Χαρακτηριστικά Χρηστών

Είτε είναι ενδιαφερόμενος αγοραστής, είτε μέλος εταιρείας αγοραπωλησίας, ο χρήστης απαιτείται να έχει μόνο βασικές γνώσεις πληροφορικής, και όχι απαραίτητα γνώσεις προγραμματισμού, καθώς η αλληλεπίδραση με το λογισμικό θα γίνεται απλά και με κατανοητό για όλους τρόπο. Θα πρέπει, βέβαια, να ξέρει κάποια βασικά Αγγλικά, καθώς ενδέχεται να κληθεί να αναζητήσει προϊόντα ή να διαβάσει περιγραφές που έχουν γραφτεί στα Αγγλικά.

2.4 Περιορισμοί

Το λογισμικό δεν έχει συγκεκριμένες απαιτήσεις υλικού, όπως ο επεξεργαστής, η διαθέσιμη μνήμη, ή το λειτουργικό σύστημα. Θα είναι, όμως, απαραίτητη η σύνδεση του χρήστη στο Internet.

Ο χρήστης δεν τίθεται να κάνει λογαριασμό με username και κωδικό σε κανένα στάδιο της εξαγωγής πληροφοριών από τον E-CS.

2.5 Παραδοχές

Κάποιες παραδοχές που μπορεί να έχουν ως αποτέλεσμα την αλλαγή των απαιτήσεων του scraper είναι οι ακόλουθες:

- Σε περίπτωση αλλαγής της δομής μιας ιστοσελίδας ο scraper θα πρέπει να είναι ευέλικτος και να μπορεί να προσαρμοστεί στις αλλαγές αυτές με ελάχιστες τροποποιήσεις.
- Ο scraper θα πρέπει να είναι σε θέση να προσαρμοστεί σε ενδεχόμενους περιορισμούς ή αποκλίσεις που μπορεί να συνάρτηση κατά την διάρκεια λειτουργίας του.

3. Ειδικές Απαιτήσεις

Παρακάτω αναφέρονται οι λειτουργικές και μη-λειτουργικές απαιτήσεις του E-CS.

3.1 Λειτουργικές Απαιτήσεις

Ακολουθούν οι λειτουργικές απαιτήσεις του λογισμικού:

3.1.1. Εξαγωγή Πληροφοριών

A. Σύντομη Περιγραφή

Γίνεται ανάκτηση των πληροφοριών για τα προϊόντα για τα οποία ενδιαφέρεται ο χρήστης από τις τρεις ιστοσελίδες αγοραπωλησίας.

B. Κύρια Ροή Γεγονότων

Εισαγωγή του URL της σελίδας.

Έλεγχος του HTML κώδικα για τις πληροφορίες των προϊόντων

Ανάκτηση ονομάτων, τιμών και περιγραφών προϊόντων.

Αποθήκευση δεδομένων σε βάση δεδομένων.

C. Εναλλακτικές Ροές Γεγονότων

Η σελίδα δεν ανταποκρίνεται → Επιστροφή σφάλματος.

3.1.2. Παρακολούθηση Τιμών

A. Σύντομη Περιγραφή

Ο χρήστης θέλει να μένει ενήμερος για τυχόν αλλαγές στις τιμές των προϊόντων.

B. Κύρια Ροή Γεγονότων

Καθορισμός προϊόντων και ιστοσελίδων.

Καθορισμός συχνότητας ελέγχου τιμών.

Επαναλαμβανόμενη ανάκτηση τιμών προϊόντων.

Σύγκριση με προηγούμενες τιμές και ενημέρωση για την αλλαγή.

C. Εναλλακτικές Ροές Γεγονότων

Κάποιο προϊόν δεν διαθέτει τιμή → Αγνόηση προϊόντος από τον έλεγχο.

3.1.3. Σύγκριση Τιμών

A. Σύντομη Περιγραφή

Ο χρήστης επιθυμεί να συγκρίνει τιμές για το ίδιο προϊόν από διαφορετικές ιστοσελίδες.

B. Κύρια Ροή Γεγονότων

Εισαγωγή τιμών προϊόντων για σύγκριση.

Ανάκτηση των τιμών από την κάθε ιστοσελίδα.

Παρουσίαση αποτελεσμάτων.

C. Εναλλακτικές Ροές Γεγονότων

Μια σελίδα δεν ανταποκρίνεται → Αγνόηση της συγκεκριμένης ιστοσελίδας και συνέχεια με την σύγκριση των άλλων.

Οι τιμές είναι ίδιες → Εμφάνιση απουσίας σύγκρισης.

3.1.4.α. Αναζήτηση Με Φίλτρα Τοποθεσίας

A. Σύντομη Περιγραφή

Ο χρήστης επιθυμεί να κάνει μια πιο εξειδικευμένη αναζήτηση με φίλτρα τοποθεσίας ή εύρους τιμών.

B. Κύρια Ροή Γεγονότων

Επιλογή φίλτρων.

Καθορισμός ιστοσελίδων που θα συμμετέχουν στην αναζήτηση.

Ανάκτηση προϊόντων που ταιριάζουν στα επιλεγόμενα φίλτρα.

Παρουσίαση προϊόντων.

C. Εναλλακτικές Ροές Γεγονότων

Απουσία προϊόντων που αντιστοιχούν στα συγκεκριμένα φίλτρα → Εμφάνιση κατάλληλου μηνύματος.

3.1.4.β. Αναζήτηση Με Φίλτρα Τιμών

A. Σύντομη Περιγραφή

Ο χρήστης επιθυμεί να κάνει μια πιο εξειδικευμένη αναζήτηση με φίλτρα εύρους τιμών.

B. Κύρια Ροή Γεγονότων

Επιλογή φίλτρων.

Καθορισμός ιστοσελίδων που θα συμμετέχουν στην αναζήτηση.

Ανάκτηση προϊόντων που ταιριάζουν στα επιλεγόμενα φίλτρα.

Παρουσίαση προϊόντων.

C. Εναλλακτικές Ροές Γεγονότων

Απουσία προϊόντων που αντιστοιχούν στα συγκεκριμένα φίλτρα → Εμφάνιση κατάλληλου μηνύματος.

3.1.5. Σύγκριση Τιμών

A. Σύντομη Περιγραφή

Ο χρήστης επιθυμεί να συγκρίνει τιμές για το ίδιο προϊόν από διαφορετικές ιστοσελίδες.

B. Κύρια Ροή Γεγονότων

Εισαγωγή τιμών προϊόντων για σύγκριση.

Ανάκτηση των τιμών από την κάθε ιστοσελίδα.

Παρουσίαση αποτελεσμάτων.

C. Εναλλακτικές Ροές Γεγονότων

Μια σελίδα δεν ανταποκρίνεται → Αγνόηση της συγκεκριμένης ιστοσελίδας και συνέχεια με την σύγκριση των άλλων.

Οι τιμές είναι ίδιες → Εμφάνιση απουσίας σύγκρισης.

3.1.6. Εμφάνιση Μέσης Τιμής

A. Σύντομη Περιγραφή

Ο χρήστης επιθυμεί να δει τον μέσο όρο τιμών ενός συγκεκριμένου προϊόντος μετά από σύγκριση.

B. Κύρια Ροή Γεγονότων

Συγκέντρωση τιμών για το επιλεγμένο προϊόν.

Υπολογισμός του μέσου όρου τιμών.

Εμφάνιση του μέσου όρου τιμών.

C. Εναλλακτικές Ροές Γεγονότων

Δεν υπάρχουν διαθέσιμες τιμές για κάποια προϊόντα → Αγνόηση των συγκεκριμένων προϊόντων και συνέχεια με τον υπολογισμό.

Δεν υπάρχουν διαθέσιμες τιμές για κανένα προϊόν → Εμφάνιση κατάλληλου μηνύματος.

3.1.7. Εμφάνιση Υψηλότερης - Χαμηλότερης Τιμής

A. Σύντομη Περιγραφή

Ο χρήστης ενδιαφέρεται για τα δύο άκρα στο εύρος τιμών του διαθέσιμου προϊόντος.

B. Κύρια Ροή Γεγονότων

Εντοπισμός της υψηλότερης και χαμηλότερης τιμής από τα προϊόντα.

Παρουσίαση αποτελεσμάτων.

C. Εναλλακτικές Ροές Γεγονότων

Δεν υπάρχουν διαθέσιμες τιμές για κάποια προϊόντα → Αγνόηση των συγκεκριμένων προϊόντων και συνέχεια με τον υπολογισμό.

Δεν υπάρχουν διαθέσιμες τιμές για κανένα προϊόν → Εμφάνιση κατάλληλου μηνύματος.

3.1.8. Εμφάνιση Κοντινότερης Τοποθεσίας και Αριθμού Προϊόντων

A. Σύντομη Περιγραφή

Ο χρήστης θέλει να εντοπίσει την πιο κοντινή τοποθεσία όπου είναι διαθέσιμο ένα συγκεκριμένο προϊόν και τον αριθμό των προϊόντων που υπάρχουν εκεί.

B. Κύρια Ροή Γεγονότων

Εντοπισμός της πιο κοντινής τοποθεσίας στον χρήστη.

Ανάκτηση του αριθμού προϊόντων στην συγκεκριμένη τοποθεσία.

Εμφάνιση αποτελεσμάτων.

C. Εναλλακτικές Ροές Γεγονότων

Δεν υπάρχουν διαθέσιμα προϊόντα στην ζητούμενη τοποθεσία → Εμφάνιση κατάλληλου μηνύματος.

3.2 Μη Λειτουργικές Απαιτήσεις

- Απαιτήσεις Απόδοσης

Το σύστημα πρέπει να ανταποκρίνεται σε αιτήματα ανάκτησης δεδομένων σε αποδεκτό χρόνο, όπως και τα δεδομένα να βρίσκονται σε βάση δεδομένων που επιτρέπει την αποδοτική αποθήκευση και ανάκτηση τους.

- Απαιτήσεις Ασφάλειας

Το σύστημα πρέπει να είναι σε θέση να ανιχνεύσει και να αντιμετωπίσει κακόβουλες κινήσεις από τις ιστοσελίδες που θέλει να εξάγει πληροφορίες (scraping blocking attempts).

- Απαιτήσεις Προστασίας

Είναι σημαντική η τήρηση των πολιτικών ασφαλείας και ο περιορισμός πρόσβασης σε ευαίσθητες πληροφορίες από μη-εξουσιοδοτημένους χρήστες.

- Απαιτήσεις Ποιότητας Λογισμικού

Θα πρέπει να γίνεται συνεχής έλεγχος από τους testers, για την εξασφάλιση της ορθής λειτουργίας του λογισμικού. Στην περίπτωση που υπάρχει πρόβλημα με κάποια λειτουργία του συστήματος, πρέπει να αναφερθεί άμεσα για την γρηγορότερη αντιμετώπισή του. Το σύστημα πρέπει να λειτουργεί σωστά ακόμα και κάτω από συνθήκες μεγάλου φόρτου.

- Απαιτήσεις Συντηρησιμότητας Λογισμικού

Κρίνεται απαραίτητη η συνεχή παρακολούθηση των ιστοσελίδων από τις οποίες έχει γίνει εξαγωγή των πληροφοριών, σε περίπτωση που έχει αλλάξει το interface ή η δομή της ιστοσελίδας. Ακόμη, θα πρέπει ο κώδικας του συστήματος να είναι ευανάγνωστο και εύκολα τροποποιήσιμο για την καλύτερη δυνατή ενημέρωση που τυχόν χρειαστεί.

4. Παραρτήματα

Τελικό Διάγραμμα των Περιπτώσεων Χρήσης:

