



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ  
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

# ΠΛΑΝΟ ΕΡΓΟΥ: E-Commerce Web Scraper

Σανίδα Μαρίνα  
Καζάς Μηνάς  
Αγαθαγγελίδου Ανθή  
Αϊσόπουλος Δημήτριος

Μάρτιος, 2024  
Κοζάνη

---

## ΠΕΡΙΕΧΟΜΕΝΑ:

<b>1. ΕΙΣΑΓΩΓΗ.....</b>	<b>3</b>
1.1. Τρέχουσα κατάσταση.....	3
1.2. Προηγούμενες ενέργειες.....	3
1.3. Στόχοι του συστήματος.....	4
1.4. Κατηγορίες χρηστών.....	4
1.5. Επισκόπηση του συστήματος.....	4
1.6. Ανάλυση SWOT.....	5
<b>2. ΟΡΓΑΝΩΣΗ ΤΟΥ ΕΡΓΟΥ.....</b>	<b>6</b>
2.1. Μέθοδος ανάπτυξης.....	6
2.2. Προσωπικό: Ρόλοι και υπευθυνότητες.....	7
<b>3. ΔΙΑΧΕΙΡΙΣΗ ΤΟΥ ΕΡΓΟΥ.....</b>	<b>8</b>
3.1. Προτεραιότητες του έργου.....	8
3.2. Βασικά ορόσημα.....	8
3.3. Παρακολούθηση και έλεγχος του έργου.....	8
3.4. Διαχείριση κινδύνων.....	9
3.5. Εκτιμήσεις έργου.....	10
<b>4. ΤΕΧΝΟΛΟΓΙΚΑ ΖΗΤΗΜΑΤΑ.....</b>	<b>11</b>
4.1. Γλώσσες προγραμματισμού.....	11
4.2. Πλατφόρμα ανάπτυξης.....	11
4.3. Διαχείριση ανάπτυξης.....	11
4.4. Εξασφάλιση ποιότητας.....	12
4.5. Τεκμηρίωση συστήματος.....	12
<b>Βιβλιογραφία.....</b>	<b>13</b>

---

# 1. ΕΙΣΑΓΩΓΗ

## 1.1. Τρέχουσα κατάσταση

Ο χώρος του ηλεκτρονικού λιανεμπορίου αλλάζει με ταχύτατους ρυθμούς. Όλο και περισσότεροι άνθρωποι στρέφονται στην αγορά και πώληση προϊόντων μέσω Διαδικτύου, λόγω της ευκολίας των συναλλαγών, καθώς και την ευχέρεια στην αναζήτηση προϊόντων από όλο τον κόσμο. Η διαδικτυακή δραστηριότητα των επιχειρήσεων αυξάνεται ραγδαία, καθώς οι επιχειρηματίες συνειδητοποιούν ότι το μέλλον των αγοραπωλησιών βρίσκεται στον ψηφιακό κόσμο. Λόγω αυτού, οι επιχειρήσεις παγκοσμίως ωθούνται να εξελίξουν την διαδικτυακή τους παρουσία με εργαλεία και λειτουργίες πιο φιλικές προς τον χρήστη.

Έτσι, έχουν αναπτυχθεί ορισμένες ιστοσελίδες που αποσκοπούν στην αγοραπωλησία τεράστιας ποικιλίας προϊόντων μαζεμένα σε έναν ιστότοπο. Παρόλη την αλματώδη ανάπτυξη στον τομέα αυτό, εξακολουθούν να υπάρχουν αδυναμίες στην εξυπηρέτηση μεγάλου μέρους πελατών. Κύρια αιτία καθίσταται η μερικώς χαοτική και διασκορπισμένη προβολή προϊόντων από κάθε πλευρά του κόσμου, ιδιαίτερα όταν κάποιος πελάτης αναζητά προϊόντα κοντά στην κατοικία του.

## 1.2. Προηγούμενες ενέργειες

Η συνεχής διαδικτυακή παρακολούθηση για τον τομέα του ηλεκτρονικού εμπορίου δεν είναι κάτι καινούργιο ή πρωτοπόρο. Παρόλα αυτά, παρουσιάζεται αμείωτο ενδιαφέρον από ένα ευρύ κοινό ανά τον κόσμο, το οποίο πια αναφέρεται σε βελτιώσεις για την καλύτερη πλατφόρμα σύγκρισης τιμών, παρακολούθησης των αλλαγών τους, την ανάλυση των τάσεων της αγοράς και πολλές ακόμα παροχές. Έως τώρα, υπάρχουν διάφοροι τρόποι με τους οποίους γίνεται το scrapping σε ιστοσελίδες ηλεκτρονικού εμπορίου, ανάλογα με την προσέγγιση που επιλέγεται, η οποία εξαρτάται από τις απαιτήσεις και την πολυπλοκότητα του ιστότοπου, αλλά και από τις ανάγκες του χρήστη. Επί παραδείγματι, μια πολύ διαδεδομένη τεχνική είναι αυτή της χρήσης API, τα οποία παρέχονται δημόσια από τις εκάστοτε ιστοσελίδες. Η χρήση των API προτιμάται γενικότερα, αφού θεωρείται ως πιο ηθική και προσφέρει πιο σταθερά δεδομένα.

Αμέσως μετά, έχουμε την χρήση βιβλιοθηκών scrapping, όπως η BeautifulSoup της Python, με την οποία μπορούμε να εξάγουμε δεδομένα από το HTML των ιστοσελίδων. Η αναφερόμενη τεχνική είναι και αυτή που επιλέχθηκε ως μέθοδος scrapping από την ομάδα.

Η σύσταση και η διαχείριση της οποίας για την ανάπτυξη ενός τέτοιου έργου κρίνεται σημαντική. Όσον αφορά την συγκρότησή της, λήφθηκε υπ όψιν η παρελθοντική συνεργασία με κάποια από τα μέλη που την απαρτίζουν, αλλά και η διάθεση από νέες προσθήκες. Παλαιότερες εργασίες κύλησαν με ομαλό ρυθμό και παρατηρήθηκε αποτελεσματική συνεργασία, γεγονός που

επίσης συντέλεσε στην κατοχύρωση όλων των μελών υπό μία, ενιαία καθοδήγηση και πορεία, με σκοπό την ανάπτυξη του παρόντος έργου.

### 1.3. Στόχοι του συστήματος

Στόχος του συστήματος που θα υλοποιηθεί είναι η ανάπτυξη ενός αποτελεσματικού και λειτουργικού τρόπου συλλογής προϊόντων, που ενδιαφέρουν έναν πιθανό πελάτη, από τις αρμόδιες ιστοσελίδες (eBay, Amazon, Skrutz, Vendra...), και της διαχώρισής τους ανάλογα με την τοποθεσία που θα έχει επιλέξει ο χρήστης. Αυτό θα λύσει ένα μεγάλο πρόβλημα που επικρατεί στα σημερινά συγκεντρωτικά site: η αναζήτηση ενός προϊόντος πρέπει να γίνει ξεχωριστά σε κάθε ιστοσελίδα επιλέγοντας πάντα την επιθυμητή τοποθεσία του φυσικού καταστήματος παραλαβής, καθώς δεν είναι δυνατή η συλλογική εμφάνιση όλων των καταστημάτων από όλα τα site σε μια ενιαία σελίδα.

Το σύστημα θα έχει ως αποτέλεσμα:

- Την εξοικονόμηση χρόνου στο στάδιο της αναζήτησης.
- Την διευκόλυνση του πελάτη στην διερεύνηση προϊόντων.

### 1.4. Κατηγορίες χρηστών

Το συγκεκριμένο έργο απευθύνεται στον χρήστη που επιθυμεί μια πιο εξειδικευμένη αναζήτηση για κάποια πιθανή αγορά. Θα του δίνεται η δυνατότητα να αναζητήσει ένα συγκεκριμένο προϊόν από πολλαπλές ιστοσελίδες αγοραπωλησίας στην τοποθεσία που τον ενδιαφέρει. Έχοντας αντλήσει προϊόντα από πολλαπλές ιστοσελίδες, το σύστημα είναι ιδανικό για χρήστες που ενδιαφέρονται να αγοράσουν οποιοδήποτε είδος προϊόντων που παρέχεται από τα συγκεκριμένα site.

Το σύστημα μπορεί επίσης να χρησιμοποιηθεί από επιχειρηματίες και εταιρείες για την ανάλυση διαφόρων δεδομένων και στατιστικών, σχετικά με τις τιμές, τάσεις της αγοράς και στρατηγικές πιθανών ανταγωνιστών, καθώς και να προσφέρουν στοχευμένες προσφορές στους πελάτες τους.

### 1.5. Επισκόπηση του συστήματος.

Το σύστημα του E-Commerce Scraper αποσκοπεί στην εξαγωγή πληροφοριών προϊόντων από ιστοσελίδες αγοραπωλησίας και παρουσίασής τους σε ένα συγκεντρωμένο περιβάλλον. Κάποιες από τις πληροφορίες περιλαμβάνουν περιγραφές και τιμές προϊόντων, διαθεσιμότητα, αξιολογήσεις πελατών, και τοποθεσία πώλησης.

Ο χρήστης θα έχει την δυνατότητα να:

- Αναζητήσει κάποιο προϊόν που ενδιαφέρεται να αγοράσει με λέξεις-κλειδιά.
- Εφαρμόσει τα κατάλληλα φίλτρα τοποθεσίας και πλαισίου τιμής.
- Επισκεφτεί την ιστοσελίδα πώλησης του προϊόντος που επέλεξε για την ολοκλήρωση της διαδικασίας της αγοράς.

## 1.6. Ανάλυση SWOT

Ακολουθεί το διάγραμμα Ανάλυσης SWOT, όπου αναγράφονται τα χαρακτηριστικά που θα διευκολύνουν και θα δυσκολέψουν την διαδικασία υλοποίησης, είτε λόγω χαρακτηριστικών της οργάνωσης, είτε του περιβάλλοντος:

<p><b>Strengths:</b></p> <ul style="list-style-type: none"> <li>• <u>Αυτοματοποίηση</u>: Με την συγκέντρωση των ιστοσελίδων σε ένα scraper μειώνεται σημαντικά η αναζήτηση σε πολλαπλές ιστοσελίδες.</li> <li>• <u>Προσαρμογή</u>: Ο χρήστης έχει τη δυνατότητα επιλογής εμφάνισης συγκεκριμένων αποτελεσμάτων κατά την αναζήτηση.</li> </ul>	<p><b>Weaknesses:</b></p> <ul style="list-style-type: none"> <li>• <u>Εξάρτηση από Δομή</u>: Κάθε αλλαγή στην δομή των ιστοσελίδων μπορεί να προκαλέσει αναστατώσεις στην διαδικασία του scraping, οδηγώντας σε λανθασμένη εξαγωγή πληροφοριών.</li> <li>• <u>Ποιότητα Δεδομένων</u>: Τυχόν δυναμικό περιεχόμενο στις ιστοσελίδες ή ελλιπείς πληροφορίες μπορεί να οδηγήσει στην ημιτελή παρουσίαση των προϊόντων.</li> </ul>
<p><b>Opportunities:</b></p> <ul style="list-style-type: none"> <li>• <u>Εξατομίκευση</u>: Αξιοποιώντας τα συλλεγόμενα δεδομένα, οι επιχειρήσεις μπορούν να βελτιώσουν τις εμπειρίες των πελατών, εξατομικεύοντας τις συστάσεις προϊόντων, τις στρατηγικές τιμολόγησης και τις προσφορές προώθησης.</li> <li>• <u>Συνεργασίες</u>: Τα scrapers αγοραπωλησίας μπορούν να ανταλλάσσουν δεδομένα με άλλες επιχειρήσεις ή πλατφόρμες, δημιουργώντας νέες ροές εσόδων και βελτιώνοντας υπάρχουσες.</li> </ul>	<p><b>Threats:</b></p> <ul style="list-style-type: none"> <li>• <u>Νομική Δράση</u>: Συγκεκριμένες ιστοσελίδες αγοραπωλησίας δεν επιτρέπουν το scraping, και η εξαγωγή πληροφοριών από αυτές μπορεί να οδηγήσει σε δικαστικές διαμάχες και πιθανά πρόστιμα.</li> <li>• <u>Ασφάλεια Δεδομένων</u>: Η αποθήκευση και η διαχείριση των scraped δεδομένων μπορεί να ενέχει κινδύνους ασφαλείας (παράβαση δεδομένων, μη εξουσιοδοτημένη πρόσβασης, κλπ.), θέτοντας σε κίνδυνο την ιδιωτικότητα των χρηστών.</li> </ul>

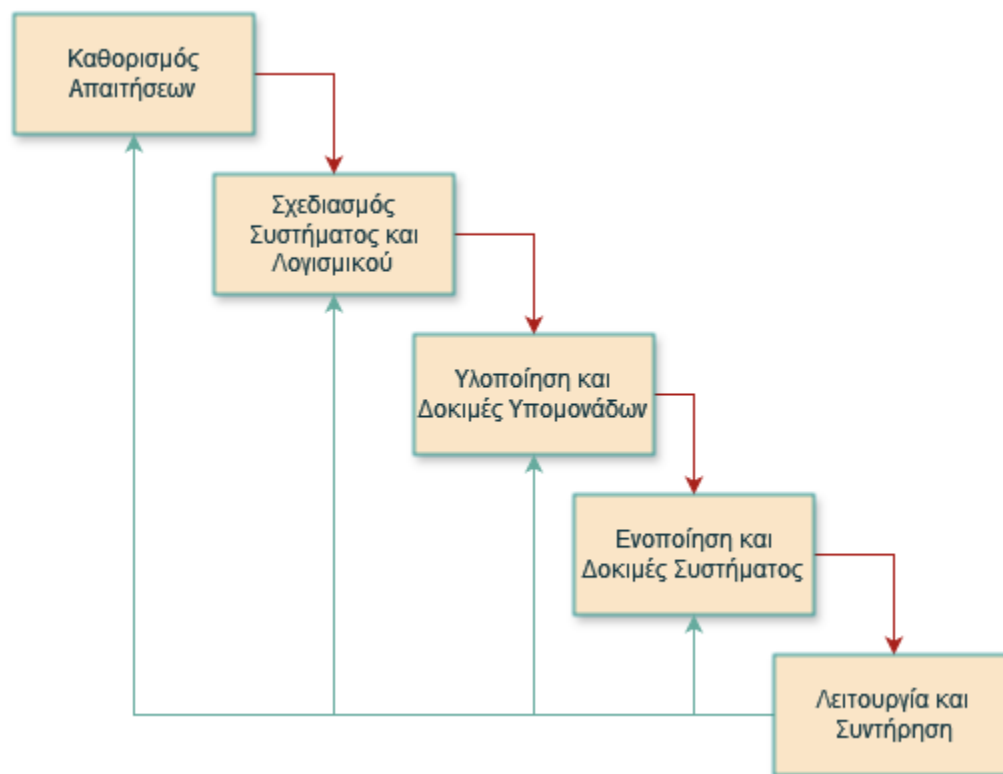
## 2. ΟΡΓΑΝΩΣΗ ΤΟΥ ΕΡΓΟΥ

### 2.1. Μέθοδος ανάπτυξης

Η επιλογή που επιλέχθηκε για την υλοποίηση του έργου είναι το Μοντέλο Καταρράκτη.<sup>[1]</sup> Πρόκειται για ένα γραμμικό μοντέλο που αξιοποιεί άκαμπτες φάσεις. Δηλαδή, όταν μία φάση τελειώνει, τότε αρχίζει η επόμενη. Με την πραγματοποίηση ενός βήματος, δεν είναι εφικτή η επιστροφή σε προηγούμενο βήμα. Τα βήματα ρέουν από την πηγή προς τα κάτω, όπως το νερό στον καταρράκτη.

Σε πραγματικές καταστάσεις, οι φάσεις ενός έργου αλληλεπικαλύπτονται και γίνονται συνεχώς επιστροφές σε παρελθοντικά στάδια. Γι' αυτόν τον λόγο, έχει αναπτυχθεί μία παραλλαγή του μοντέλου, η οποία καθιστά την διαδικασία πιο ευέλικτη. Συγκεκριμένα, επιτρέπει την οπισθοδρόμηση σε προηγούμενες φάσεις για τυχόν διορθώσεις και βελτιώσεις συγκεκριμένων στελεχών του συστήματος.

Το παρακάτω σχεδιάγραμμα απεικονίζει τις φάσεις του Μοντέλου Καταρράκτη:



**Εικόνα 1:** Μοντέλο Καταρράκτη. Τα κόκκινα βέλη δείχνουν την βασική ροή των φάσεων, ενώ με πράσινο χρώμα φαίνονται τα βέλη της οπισθοδρόμησης στην παραλλαγή του μοντέλου.

## 2.2. Προσωπικό: Ρόλοι και υπευθυνότητες

Τα μέλη της ομάδας είναι σε θέση να ασχοληθούν με πολλαπλούς ρόλους κατά την ανάπτυξη του έργου, λαμβάνοντας υπόψη τα δυνατά σημεία του καθενός, για μία πιο αποτελεσματική υλοποίηση.

Ρόλος	Υπεύθυνοι
Συγγραφή Κειμένων	Αγαθαγγελίδου, Καζάς, Σανίδα, Αϊσόπουλος
Μορφοποίηση Αρχείων / Δημιουργία Διαγραμμάτων	Αγαθαγγελίδου
Προγραμματισμός	Καζάς, Σανίδα, Αγαθαγγελίδου
Έλεγχος Ποιότητας	Αϊσόπουλος

### 3. ΔΙΑΧΕΙΡΙΣΗ ΤΟΥ ΕΡΓΟΥ

#### 3.1. Προτεραιότητες του έργου

Για το συγκεκριμένο έργο, είναι ζωτικής σημασίας η διασφάλιση της ιδιωτικότητας των χρηστών. Πρέπει να ληφθούν, λοιπόν, ορισμένα μέτρα για την προστασία των προσωπικών τους δεδομένων. Κάποια από αυτά αποτελούν:

- **Περιορισμός Πρόσβασης:** Τα προσωπικά δεδομένα θα είναι προσβάσιμα μόνο στους διαχειριστές του συστήματος.
- **Εφαρμογή Κανόνων Απορρίπτου:** Το σύστημα δεν θα συλλέγει παραπάνω πληροφορίες από αυτές που απαιτούνται για την ορθή λειτουργία του.
- **Ανίχνευση Παραβάσεων:** Θα υπάρχουν κατάλληλοι μηχανισμοί, οι οποίοι θα μπορούν να αντιμετωπίζουν πιθανές απόπειρες ανεπιθύμητης πρόσβασης ή διαρροής δεδομένων.

#### 3.2. Βασικά ορόσημα

Τα βασικά ορόσημα του πρότζεκτ αναγράφονται στον παρακάτω πίνακα. Το χρονοδιάγραμμα παρέχει έναν οδηγό για την αποτελεσματική οργάνωση της υλοποίησης και την ανάθεση απαραίτητων αρμοδιοτήτων στα μέλη της ομάδας, για κάθε χρονικό διάστημα.

Ημερομηνία Προθεσμίας	Φάση Ανάπτυξης
20/3/2024	Πλάνο Έργου
20/4/2024	Απαιτήσεις Έργου
15/5/2024	Σχέδιο Λογισμικού
15/6/2024	Κώδικας Λογισμικού, Έλεγχος Λογισμικού

#### 3.3. Παρακολούθηση και έλεγχος του έργου

Η παρακολούθηση του έργου θα γίνει μέσω δια-ζώσης και διαδικτυακών συναντήσεων, συγκεκριμένα χρησιμοποιώντας την πλατφόρμα του Discord. Ο έλεγχος γίνεται με βάση το



χρονοδιάγραμμα που υλοποιήθηκε στην αρχή του προτζεκτ. Συγκεκριμένα, θα γίνονται συχνά rain-checks για την διασφάλιση της ομαλής λειτουργίας και εκπλήρωσης όλων των ρόλων των μελών. Θα υπάρχουν συνεχείς έλεγχοι από όλα τα μέλη της ομάδας για προβλήματα που μπορεί να προκύψουν, τα οποία θα μοιράζονται σε συγκεκριμένο κανάλι στην πλατφόρμα επικοινωνίας. Με αυτόν τον τρόπο, τα προβλήματα θα είναι συγκεντρωμένα και έτοιμα για επίλυση ανά πάσα στιγμή από τα μέλη της ομάδας.

### 3.4. Διαχείριση κινδύνων

Στην υλοποίηση μιας εργασίας τέτοιου είδους είναι φυσικό να υπάρξουν προβλήματα, κάποια από τα οποία είναι αναμενόμενα και κάποια απρόβλεπτα. Πιθανά προβλήματα που ενδέχεται να προκύψουν είναι τα εξής:

1. Κάποιες ιστοσελίδες απαγορεύουν ρητά το scraping οπότε ίσως αποκλείσουμε κάποιες κυρίες ιστοσελίδες λόγω των πολιτικών τους.

*Αντιμετώπιση:* Έλεγχος πολιτικών σελίδας πριν την εκκίνηση του scrapping.

2. Αν μια ιστοσελίδα αλλάξει την δομή δεδομένων της, υπάρχει μεγάλη πιθανότητα το scraper που θα έχουμε υλοποιήσει να μην μπορεί να προσαρμοστεί στα νέα δεδομένα και να μην δουλέψει.

*Αντιμετώπιση:* Παρακολουθούμε τακτικά τις αλλαγές στη δομή των ιστοσελίδων και προσαρμόζουμε τον κώδικα μας κατάλληλα.

3. Αν κάποια ιστοσελίδα ανιχνεύσει έντονη δραστηριότητα από μια συγκεκριμένη IP Address, υπάρχει κίνδυνος αποκλεισμού από την ιστοσελίδα αυτή.

*Αντιμετώπιση:* Προσαρμογή του scraping για να μην υπερφορτώνει τις ιστοσελίδες και χρήση προσωρινών διακομιστών (proxies) για αποφυγή blocking.

4. Η συλλογή μεγάλου όγκου πληροφοριών μπορεί να υπερφόρτωση την σελίδα μας με αποτέλεσμα να μην λειτουργεί.

*Αντιμετώπιση:* Χρησιμοποίηση καθυστερήσεων μεταξύ των αιτήσεων (frame limiting) και διαχείριση την ποσότητα των δεδομένων που συλλέγουμε.

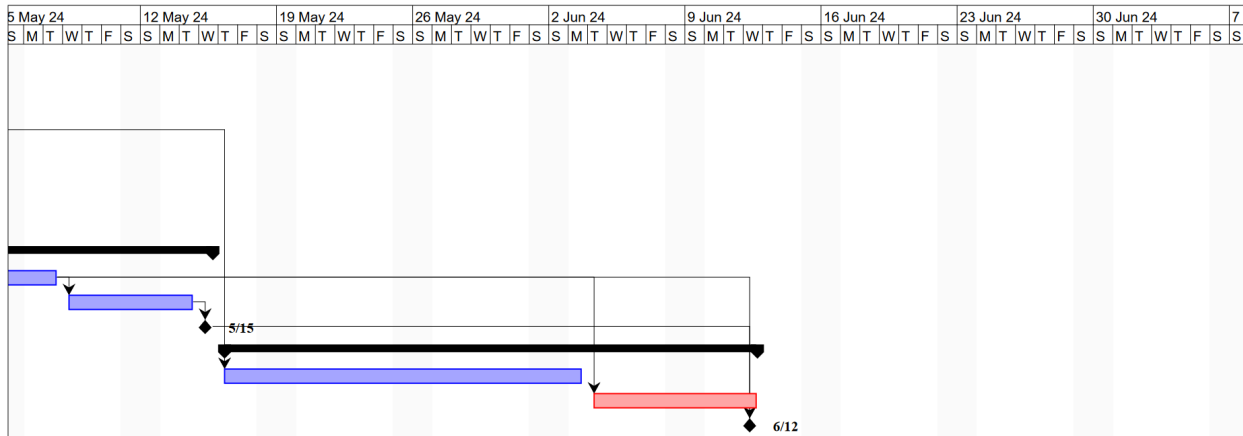
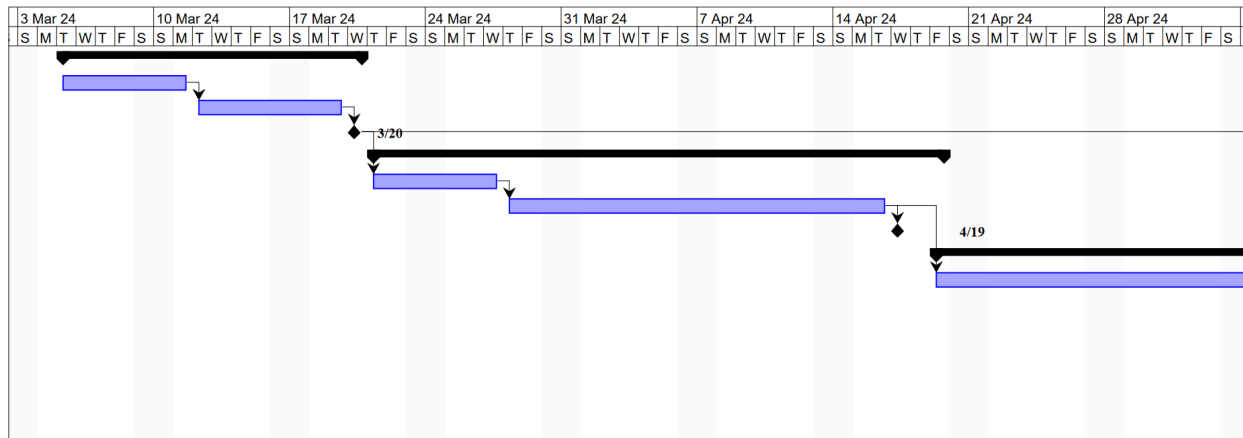
5. Ορισμένες πληροφορίες μπορεί να είναι προστατευμένες και να μην επιτρέπεται η εξόρυξη τους από μη εξουσιοδοτημένα άτομα.

Δεν υπάρχει αντιμετώπιση. Απαιτείται η επιλογή άλλης ιστοσελίδας ή κλάδου πληροφοριών.

### 3.5. Εκτιμήσεις έργου

Ακολουθεί το Διάγραμμα Gantt για την εκτίμηση του χρονοδιαγράμματος του έργου. Οι διάρκειες κάθε φάσης έχουν τεθεί με βάση τα ορόσημα που αναφέρθηκαν στο 3.2.

Name	Duration	Start	Finish	Predecessors
☐ Ορισμός απαιτήσεων	12 days	3/5/24 8:00 AM	3/20/24 5:00 PM	
Καθορισμός Στόχου	5 days	3/5/24 8:00 AM	3/11/24 5:00 PM	
Καθορισμός Περιπτώσεων >	6 days	3/12/24 8:00 AM	3/19/24 5:00 PM	2
Έγκριση πελάτη	1 day	3/20/24 8:00 AM	3/20/24 5:00 PM	3
☐ Σχεδίαση Κώδικα	22 days	3/21/24 8:00 AM	4/19/24 5:00 PM	
Επιλογή Αρχιτεκτονικής	5 days	3/21/24 8:00 AM	3/27/24 5:00 PM	4
Σχεδίαση κλάσεων	14 days	3/28/24 8:00 AM	4/16/24 5:00 PM	6
Έγγραφο σχεδίου	3 days	4/17/24 8:00 AM	4/19/24 5:00 PM	7
☐ Υλοποίηση	19 days	4/19/24 8:00 AM	5/15/24 5:00 PM	
Υλοποίηση κλάσεων	13 days	4/19/24 8:00 AM	5/7/24 5:00 PM	7
Ενοποίηση κώδικα	5 days	5/8/24 8:00 AM	5/14/24 5:00 PM	10
Παράδοση κώδικα	1 day	5/15/24 8:00 AM	5/15/24 5:00 PM	11
☐ Έλεγχος κώδικα	20 days	5/16/24 8:00 AM	6/12/24 5:00 PM	
Συγγραφή Περιπτώσεων ελ	13 days	5/16/24 8:00 AM	6/3/24 5:00 PM	4
Δοκιμές	7 days	6/4/24 8:00 AM	6/12/24 5:00 PM	10
Παράδοση συστήματος	1 day	6/12/24 8:00 AM	6/12/24 5:00 PM	10;12



## 4. ΤΕΧΝΟΛΟΓΙΚΑ ΖΗΤΗΜΑΤΑ

### 4.1. Γλώσσες προγραμματισμού

Η γλώσσα που θα χρησιμοποιηθεί στο έργο είναι η Python. Η Python είναι μία αντικειμενοστραφής γλώσσα υψηλού επιπέδου, η οποία βρίσκει χρήση σε πολλούς τομείς της μηχανικής υπολογιστών, ειδικότερα στην ανάπτυξη ιστοσελίδων και ψηφιακών παιχνιδιών, στην μηχανική μάθηση, και φυσικά στο web scrapping.<sup>[2]</sup> Η ευελιξία της γλώσσας επιτρέπει την υλοποίηση του καλύτερου δυνατού συστήματος.

### 4.2. Πλατφόρμα ανάπτυξης

Το προγραμματιστικό περιβάλλον στο οποίο θα αναπτυχθεί το παρόν project είναι το PyCharm, το οποίο ενδείκνυται για την χρήση της γλώσσας Python, στην οποία θα αναπτυχθεί ο κώδικας, καθώς επίσης προτείνεται και για την ανάπτυξη project από αρχάριους.

Θα χρησιμοποιηθούν οι εξής βιβλιοθήκες για την υλοποίηση του E-Commerce Scraper:

- **Beautiful Soup:** αναλύει έγγραφα HTML & XML και εξάγει δεδομένα από αυτά.
- **Requests:** για αποστολή αιτημάτων HTTP, κάτι που είναι απαραίτητο για την αλληλεπίδραση με ιστότοπους ηλεκτρονικού εμπορίου και την άντληση δεδομένων από αυτά.
- **Scrapy:** για τον ορισμό της δομής των ιστοσελίδων για scrapping, τον χειρισμό των αιτημάτων και των απαντήσεων και την επεξεργασία των δεδομένων.
- **Pandas:** για ανάλυση δεδομένων.
- **Splash:** σχεδιασμένο ειδικά για το scrapping ιστοσελίδων, ιδιαίτερα αυτές που είναι *JavaScript-heavy*. Μπορεί να ενσωματωθεί με το Scrapy ή να χρησιμοποιηθεί ανεξάρτητα για το scrapping δυναμικού περιεχομένου από ιστοσελίδες e-commerce.

Για την δημιουργία των διαγραμμάτων και άλλων σχεδιαγραμμάτων χρησιμοποιείται το ProjectLibre, όπως και το Visual Paradigm.

### 4.3. Διαχείριση ανάπτυξης

Η χρήση του εργαλείου Git είναι σαφώς αναγκαία, αφού πρόκειται για ένα DevOps εργαλείο το οποίο χρησιμοποιείται για την διαχείριση πηγαίου κώδικα. Είναι ένα δωρεάν και ανοικτού κώδικα σύστημα διαχείρισης, το οποίο χρησιμοποιείται τόσο σε μικρά όσο και σε μεγαλύτερα έργα. Επιτρέπει σε πολλούς προγραμματιστές να εργάζονται μαζί, κρατώντας

παράλληλα το ιστορικό των αλλαγών που γίνονται από αυτούς, στον κώδικα. Το γεγονός ότι καθένας από αυτούς μπορεί να έχει ένα local αντίγραφο του κώδικα και ανά πάσα στιγμή πρόσβαση στο πλήρες ιστορικό αλλαγών, η μεγάλη άνεση που προσφέρει στην ομάδα όσον αφορά την διαχείριση αλλά και η υποστήριξη μη γραμμικής ανάπτυξης έργου (που επιτρέπει ουσιαστικά τους προγραμματιστές να πειραματίζονται με νέες ιδέες και να τον τροποποιούν χωρίς να επηρεάζουν τον κύριο κώδικα) καθιστούν το Git ένα πολύτιμο εργαλείο, χωρίς το οποίο η ανάπτυξη του project στο οποίο αναφερόμαστε θα ήταν δυσλειτουργική.

#### 4.4. Εξασφάλιση ποιότητας

Για την εξασφάλιση της αξιοπιστίας και της λειτουργικότητας του τελικού προϊόντος, θα γίνουν πολλαπλές δοκιμές σε κάθε στάδιο του έργου. Συγκεκριμένα:

- Κάνοντας **δοκιμές μονάδων** (unit testing) εξασφαλίζεται η αποτελεσματικότητα κάθε μέρους του κώδικα.
- Με την γενική **δοκιμή λειτουργικότητας** επιβεβαιώνεται η ορθή λειτουργία του προγράμματος.
- Με την **δοκιμή απόδοσης** ελέγχεται η ταχύτητα και η αποδοτικότητά του κάτω από διάφορες συνθήκες χρήσης.
- Σημαντικό θα είναι να γίνουν και **δοκιμές ασφάλειας**, κατοχυρώνοντας την προστασία των χρηστών και των ιδιωτικών τους δεδομένων.

#### 4.5. Τεκμηρίωση συστήματος

Το σύστημα θα συνοδεύεται από μία εκτενή περιγραφή στο Github, όπου θα γίνεται επισκόπηση των λειτουργιών του scraper και αναφορά στα συγκεκριμένα φίλτρα που μπορούν να εφαρμοστούν.

## Βιβλιογραφία

1. Tekinerdogan, B., Alí, N., Grundy, J., Mistrík, I., & Soley, R. M. (2016). Quality concerns in large-scale and complex software-intensive systems. In Elsevier eBooks (pp. 1–17). <https://doi.org/10.1016/b978-0-12-802301-3.00001-6> / <https://www.sciencedirect.com/topics/computer-science/waterfall-model>
2. GfG. (2024, February 27). What is Python? its Uses and Applications. GeeksforGeeks. <https://www.geeksforgeeks.org/what-is-python/>