

Still Series: Recommender System βασισμένο σε Collaborative Filtering Προσέγγιση

Αποθηκεσ δεδομενων και επιχειρηματικη ευφυϊα

Επιβλέπουσα Καθηγήτρια: Μαρία Χαλκίδη

**Ιωάννα Κανδή, Κωνσταντίνος Μαυρογιώργος**

ΑΜ: ΜΕ2136, ΜΕ2144

Email: {ioannakandi,kostismvg}@gmail.com

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΙΔΙΚΕΥΣΗ: “ΠΡΟΗΓΜΕΝΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ”

# Περιεχόμενα

[Περιεχόμενα 1](#_Toc104207434)

[Κατάλογος Πινάκων 2](#_Toc104207435)

[Κατάλογος Εικόνων 3](#_Toc104207436)

[Περίληψη 4](#_Toc104207437)

[1.Θεματολογία Συστήματος 5](#_Toc104207438)

[1.1 Ορισμός του προβλήματος 5](#_Toc104207439)

[1.2 Στόχοι της εργασίας 5](#_Toc104207440)

[1.3 Δομή της εργασίας 5](#_Toc104207441)

[2. Θεωρητικό Υπόβαθρο 6](#_Toc104207442)

[Ερώτημα Α 6](#_Toc104207443)

[Ερώτημα Β 6](#_Toc104207444)

[Ερώτημα Γ 8](#_Toc104207445)

[3. Υλοποίηση Συστήματος 9](#_Toc104207446)

[3.1. Υλοποιημένες Υπηρεσίες Ιστού 9](#_Toc104207447)

[3.2. Οδηγίες Εκτέλεσης 11](#_Toc104207448)

[Απόκριση Εφαρμογής (responsiveness) 15](#_Toc104207449)

[4. Συμπεράσματα 17](#_Toc104207450)

[Βιβλιογραφία 18](#_Toc104207451)

# Κατάλογος Πινάκων

[Πίνακας 1: Πρότυπο πίνακα για υλοποιημένες υπηρεσίες. 9](#_Toc104207452)

[Πίνακας 2: Υπηρεσία δημιουργίας και ανάκτησης προτάσεων 9](#_Toc104207453)

# Κατάλογος Εικόνων

[Εικόνα 3: Η διεπαφή χρήστη του Still Series 13](#_Toc104207467)

[Εικόνα 4: Παροχή προτάσεων βάσει ομοιότητας χρηστών 14](#_Toc104207468)

[Εικόνα 5: Παροχή προτάσεων βάσει ομοιότητας αντικειμένων (σειρών) 14](#_Toc104207469)

[Εικόνα 6: Η διεπαφή του συστήματος, όπως φαίνεται από ένα iPhone X. 15](#_Toc104207470)

# Περίληψη

Η παρούσα εργασία αποτελεί την απαλλακτική εργασία για το μάθημα «Αποθήκες Δεδομένων και Επιχειρηματική Ευφυΐα» του 2ου εξαμήνου της μεταπτυχιακής ειδίκευσης «Προηγμένα Πληροφοριακά Συστήματα» του Τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιώς. Ειδικότερα, σκοπός της εργασίας είναι ο σχεδιασμός και η υλοποίηση ενός recommender system που να βασίζεται σε collaborative filtering προσέγγιση. Πιο συγκεκριμένα, η εργασία χωρίζεται σε δύο (2) μέρη. Το πρώτο μέρος περιλαμβάνει τις λύσεις σχετικά με τα τρία (3) πρώτα θεωρητικά ερωτήματα της εργασίας. Το δεύτερο μέρος αφορά την υλοποίηση του συστήματος. Πιο συγκεκριμένα, σε αυτό το μέρος αναλύονται οι υπηρεσίες που είναι διαθέσιμες ενώ παρέχεται και ένα αναλυτικό εγχειρίδιο χρήσης της συστήματος.

# 1.Θεματολογία Συστήματος

## 1.1 Ορισμός του προβλήματος

Την σημερινή εποχή παρουσιάζονται διάφορες προκλήσεις οι οποίες επηρεάζουν την καθημερινότητα και οδηγούν στη διαμόρφωση μιας νέας πραγματικότητας. Στα πλαίσια αυτής, η ανάγκη για ανάπτυξη εφαρμογών και συστημάτων που παρέχουν ολοένα και πιο εξατομικευμένες υπηρεσίες στους χρήστες. Ένα χαρακτηριστικό παράδειγμα εφαρμογών τέτοιου τύπου αποτελούν τα συστήματα προτάσεων (recommender systems). Τα συγκεκριμένα συστήματα παρέχουν προτάσεις στους χρήστες βάσει των προτιμήσεών τους και της ομοιότητας της οποίας ενδέχεται να έχουν με άλλους χρήστες. Η χρήση συστημάτων του συγκεκριμένου τύπου πραγματοποιείται από διαφορετικές εταιρείες και πλατφόρμες όπως το Netflix, η Amazon και το Yahoo, προκειμένου να παρέχουν στους χρήστες όσο το δυνατόν πιο εξατομικευμένες προτάσεις. Η συγκεκριμένη εργασία στοχεύει στην δημιουργία ενός τέτοιου συστήματος το οποίο θα παρέχει στους χρήστες του, μέσα από ένα πολύ φιλικό περιβάλλον, προτάσεις σχετικά με τηλεοπτικές εκπομπές που θα μπορούσαν να παρακολουθήσουν, βάσει των προτιμήσεών τους.

## 1.2 Στόχοι της εργασίας

Στη παρούσα εργασία, αρχικά δίνονται απαντήσεις σε μερικά θεωρητικά ερωτήματα που αφορούν διαφορετικές προσεγγίσεις σχετικά με τα recommender systems. Εν συνεχεία, υλοποιείται ένα σύστημα που ονομάζεται “Still Series” και έχει ως στόχο την παροχή προτάσεων σε χρήστες σχετικά με τηλεοπτικές εκπομπές που θα μπορούσαν να παρακολουθήσουν, βάσει των προτιμήσεών τους.

## 1.3 Δομή της εργασίας

Στο Κεφάλαιο 2 – Θεωρητικό Υπόβαθρο παρατίθενται οι λύσεις που αφορούν τα πρώτα τρία (3) ερωτήματα της εργασίας. Στο Κεφάλαιο 3 – Υλοποίηση Συστήματος περιγράφονται με κάθε λεπτομέρεια οι υπηρεσίες που παρέχονται από την εφαρμογή, όπως επίσης και παρουσιάζεται συνοπτικά η αρχιτεκτονική της, σε συνδυασμό με την ροή που ακολουθείται κατά την εκτέλεσή της. Επίσης, αναλύονται όλες οι τεχνολογίες που αξιοποιήθηκαν για την υλοποίηση. Τέλος, στο Κεφάλαιο 4 – Συμπεράσματα πραγματοποιείται μία ανασκόπηση των πεπραγμένων της εργασίας και καταγράφονται μελλοντικές βελτιώσεις που θα μπορούσαν να γίνουν στην εφαρμογή, ούτος ώστε να ανταποκρίνεται και σε ακόμα πιο σύνθετα σενάρια χρήσης.

# 2. Θεωρητικό Υπόβαθρο

Αναφορικά με τις λύσεις των πρώτων τριών (3) ασκήσεων που περιγράφονται στην εκφώνηση της εργασίας, αυτές παρατίθενται παρακάτω, ενώ αξίζει να σημειωθεί ότι αξιοποιήθηκε το λογισμικό MATLAB.

### Ερώτημα Α

Ο συγκεκριμένος πίνακας Τ είναι συμμετρικός. Τα στοιχεία της κύριας διαγωνίου αναφέρονται στο πλήθος των αντικειμένων που έχει βαθμολογήσει ο εκάστοτε χρήστης, ενώ τα υπόλοιπα στοιχεία αναφέρονται στα κοινά αντικείμενα που έχουν βαθμολογήσει οι εκάστοτε χρήστες που αντιστοιχούν στο αντίστοιχο κελί του πίνακα. Για παράδειγμα, στο ακόλουθο πίνακα Τ το τονισμένο στοιχείο υποδεικνύει ότι ο χρήστης 1 έχει αξιολογήσει 2 αντικείμενα.

Chart

Description automatically generated

Επίσης, στον ίδιο πίνακα, το δεύτερο στοιχείο στην Τρίτη γραμμή υποδεικνύει πως ο χρήστης 3 και ο χρήστης 2 έχουν αξιολογήσει ένα κοινό αντικείμενο.

Chart, box and whisker chart

Description automatically generated

### Ερώτημα Β

Έστω ο πίνακας ομοιότητας R. Υπολογίζοντας την ομοιότητα μεταξύ των στηλών (δηλαδή των αντικειμένων), προκύπτει ο ακόλουθος πίνακας S1.

Table

Description automatically generated

Text

Description automatically generated with medium confidence

Text

Description automatically generated with medium confidence

A picture containing calendar

Description automatically generated

Η προς απόδειξη σχέση είναι η ακόλουθη :

Text

Description automatically generated with low confidence

Όπου:

Chart, scatter chart

Description automatically generated

Επιλύοντας την προαναφερθείσα σχέση, προκύπτει ότι:

Text

Description automatically generated with low confidence

### Ερώτημα Γ

Item-Item collaborative filtering

Βάσει των δεδομένων της άσκησης και της δοθείσας σχέσης (Γ= RQ-1/2RTRQ-1/2) του ερωτήματος Γ, ο πίνακας προτάσεων που προκύπτει είναι ο ακόλουθος:

Text

Description automatically generated

Ειδικότερα, για τον πρώτο χρήστη θα προταθεί είτε το πρώτο είτε το τρίτο αντικείμενο. Για τον δεύτερο χρήστη, θα προταθεί το πρώτο αντικείμενο και για τον τρίτο, τα αντικείμενα 1 και 3. Τέλος, στον τέταρτο χρήστη θα προταθεί το τρίτο αντικείμενο.

User-User collaborative filtering

Στη συγκεκριμένη περίπτωση θα πρέπει να ληφθεί υπόψη και η μη-κανονικοποιημένη ομοιότητα ανάμεσα σε χρήστες (Τ=R\*RT), η οποία έχει υπολογιστεί στο ερώτημα Α. Συνεπώς, η σχέση που προκύπτει είναι η ακόλουθη :Γ= ΤRQ-1/2RTRQ-1/2. Ο αντίστοιχος πίνακας προτάσεων που προκύπτει είναι ο παρακάτω:

Table

Description automatically generated

Παρατηρούμε ότι, οι προτάσεις που προκύπτουν βάσει του παραπάνω πίνακα ταυτίζονται με αυτές της item-item προσέγγισης.

# 3. Υλοποίηση Συστήματος

## 3.1. Υλοποιημένες Υπηρεσίες Ιστού

Στη συγκεκριμένη ενότητα περιγράφεται η υλοποιημένη υπηρεσία σε μορφή πινάκων. Ο πίνακας έχει την ακόλουθη μορφή:

|  |  |
| --- | --- |
| **Όνομα Υπηρεσίας** |  |
| Περιγραφή | Περιγραφή της λειτουργίας της υπηρεσίας |
| Endpoint URL | http://localhost:5000/... |
| HTTP μέθοδος | Δυνατές τιμές: GET, POST, PUT, DELETE |
| Παράμετροι | Οι παράμετροι που χρειάζεται η υπηρεσία (αν υπάρχουν) |
| Επιστροφή | Τι επιστρέφει η υπηρεσία στον χρήστη |

Πίνακας 1: Πρότυπο πίνακα για υλοποιημένες υπηρεσίες.

|  |  |
| --- | --- |
| **getShows** |  |
| Περιγραφή | Χρησιμοποιείται για την παραγωγή προτάσεων και την ανάκτησή τους από τον εκάστοτε χρήστη |
| Endpoint URL | http://localhost:5000/ getShows/  {search\_parameter}/{search\_method} |
| HTTP μέθοδος | GET |
| Παράμετροι | search\_parameter: η παράμετρος βάσει της οποίας θα παραχθούν οι προτάσεις. Μπορεί να είναι είτε ο κωδικός ενός χρήστη, είτε το όνομα μίας σειράς  search\_method: ο τρόπος με τον οποίο θα παραχθεί η πρόταση. Αν πρόκειται για user-based, τότε λαμβάνει τη τιμή «user». Διαφορετικά, αν πρόκειται για item-based, τότε λαμβάνει τη τιμή «item» |
| Επιστροφή | Επιστρέφει δέκα (10) προτάσεις |

Πίνακας 2: Υπηρεσία δημιουργίας και ανάκτησης προτάσεων

Το σύστημα / εφαρμογή της εργασίας είναι web-based και αποτελείται από το front-end, το οποίο αποτελεί την διεπαφή του χρήστη, και το back-end που είναι όλες οι υπηρεσίες που έχει πρόσβαση ο χρήστης από την διεπαφή. Για το front-end χρησιμοποιήθηκαν οι ακόλουθες τεχνολογίες:

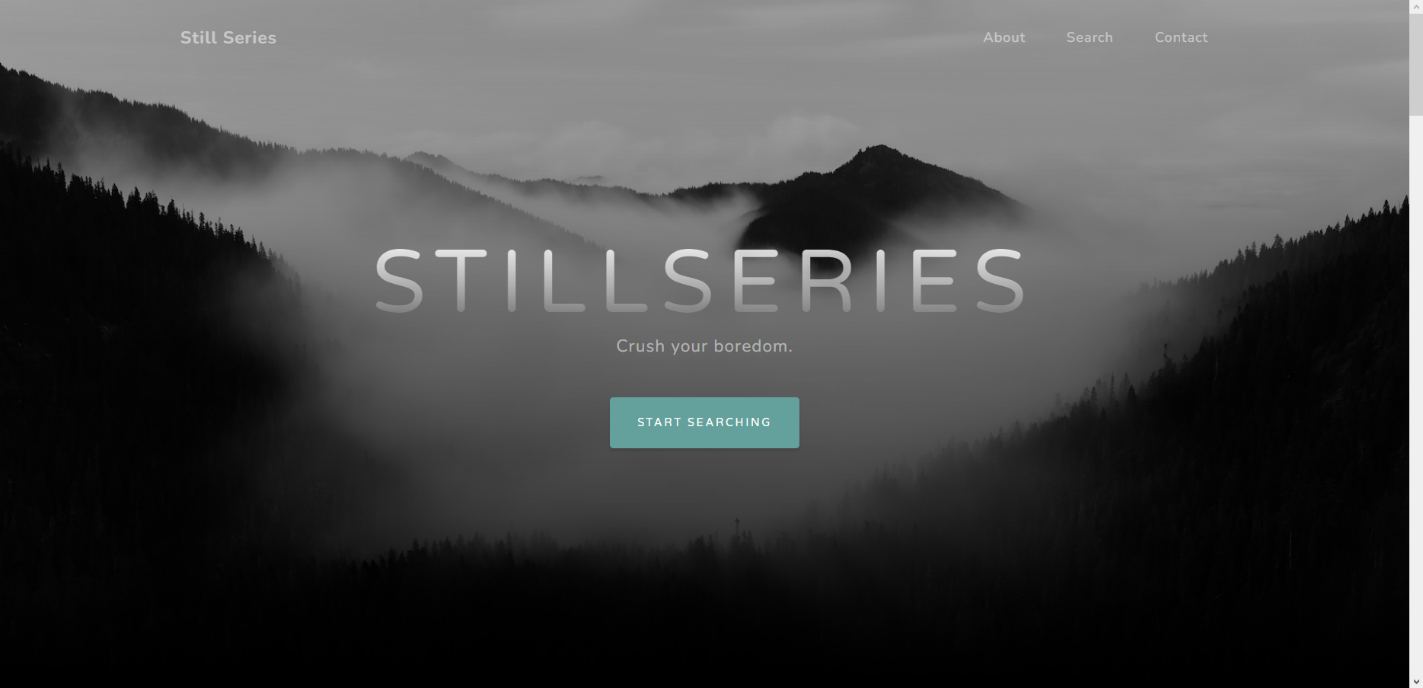
* HTML5 για την οργάνωση των επιμέρους τμημάτων από τα οποία αποτελείται η διεπαφή και την προσθήκη κανόνων στις φόρμες υποβολής του χρήστη για διασφάλιση της εγκυρότητας αυτών πριν σταλούν στο back-end.
* Bootstrap 4: για την μορφοποίηση των παραπάνω τμημάτων ούτος ώστε να είναι ευδιάκριτα και ελκυστικά προς τον χρήστη. Το Bootstrap είναι ένα framework το οποίο αποτελείται από HTML, CSS και JavaScript. Ουσιαστικά, παρέχει στους προγραμματιστές έτοιμες κλάσεις τις οποίες μπορούν να χρησιμοποιήσουν έτσι ώστε να υλοποιήσουν γρήγορα μία πολύ ελκυστική διεπαφή χρήστη.
* CSS για την μορφοποίηση των παραπάνω τμημάτων ούτος ώστε να είναι ευδιάκριτα και ελκυστικά προς τον χρήστη.
* JavaScript για την προσθήκη του στοιχείου της δυναμικότητας στο front-end, διατήρηση cookies και γενικά διαχείριση διάφορων τιμών στο front-end.
* jQuery, η οποία αποτελεί μία βιβλιοθήκη JavaScript και ουσιαστικά παίρνει πολλές κοινές εργασίες που απαιτούν πολλές γραμμές κώδικα JavaScript για να ολοκληρωθούν και τις «τυλίγει» σε μεθόδους που μπορούν να χρησιμοποιηθούν με μία γραμμή κώδικα. Μέσω της jQuery γίνονται και οι κλήσεις Ajax στα διάφορα API με τα οποία επικοινωνεί η εφαρμογή, όπως επίσης και η εφαρμογή τεχνικών form validation, αξιοποιώντας την αντίστοιχη βιβλιοθήκη (jQuery Validation Plugin).

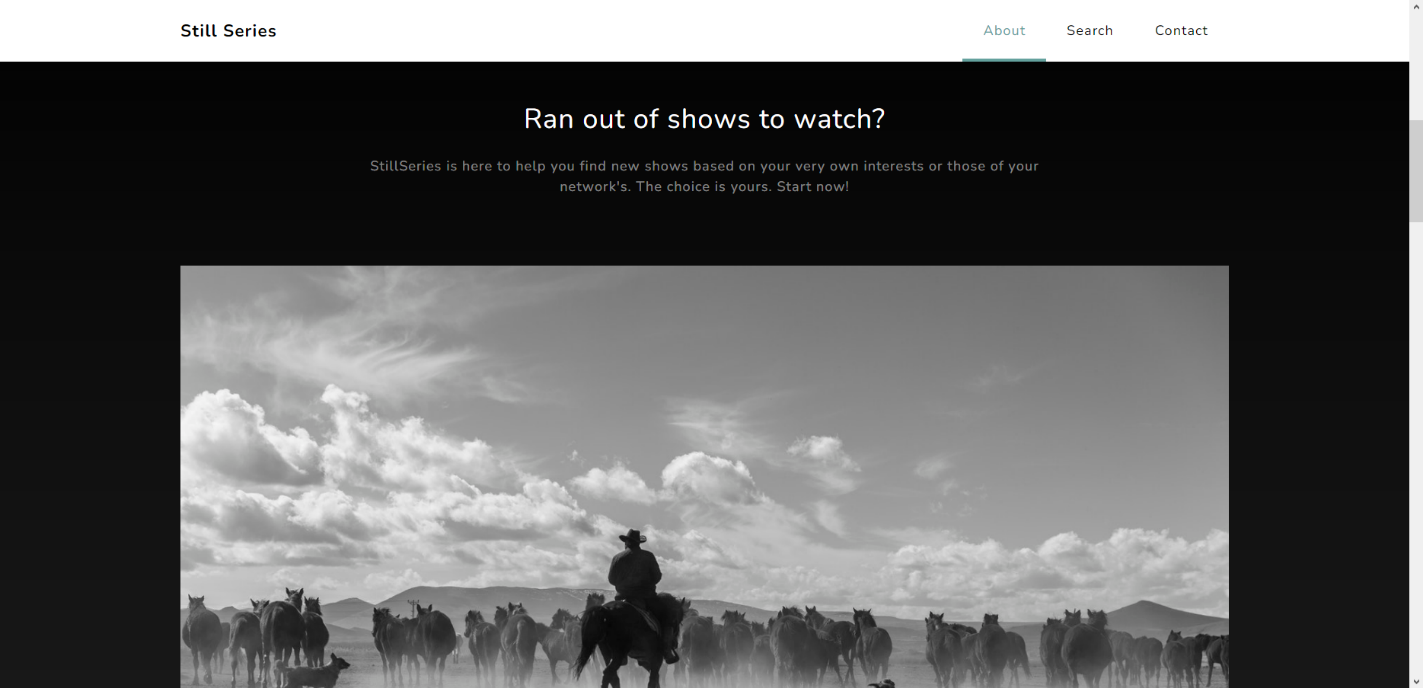
Για το back-end χρησιμοποιήθηκαν οι ακόλουθες τεχνολογίες:

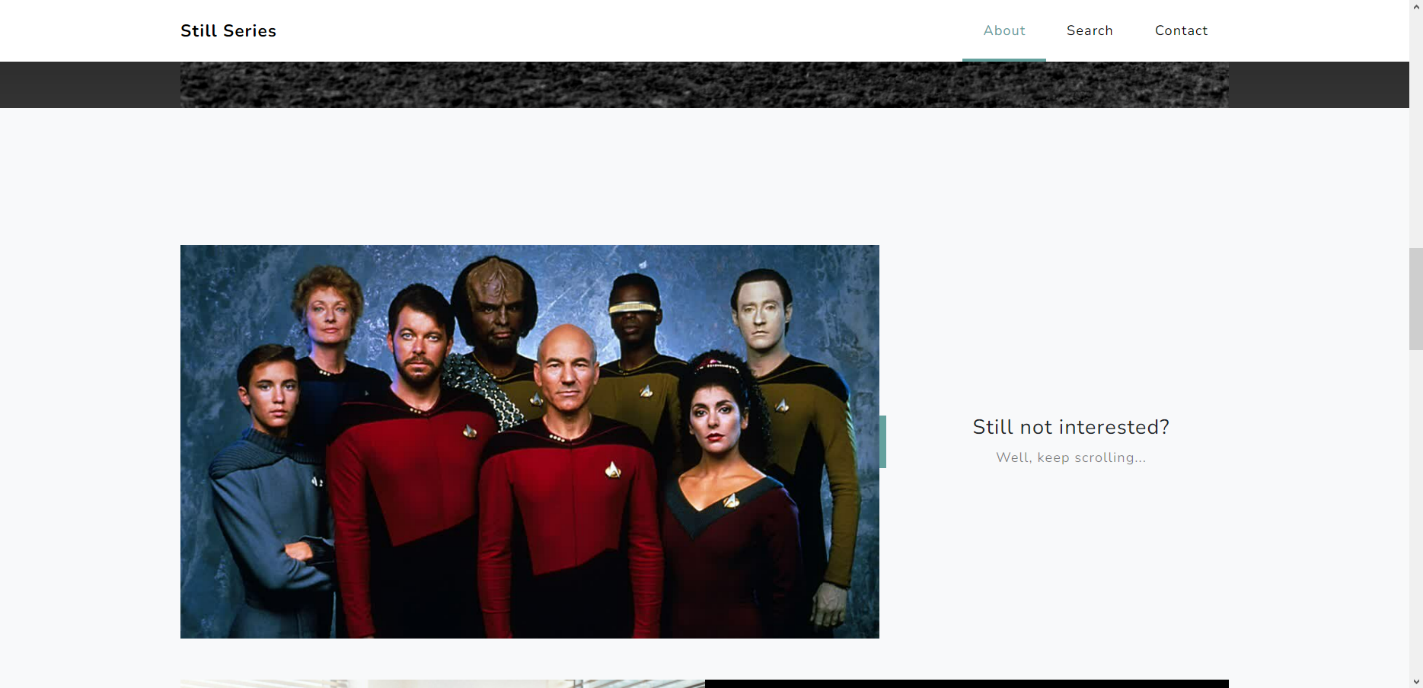
* Python για την υλοποίηση των παρεχόμενων υπηρεσιών.

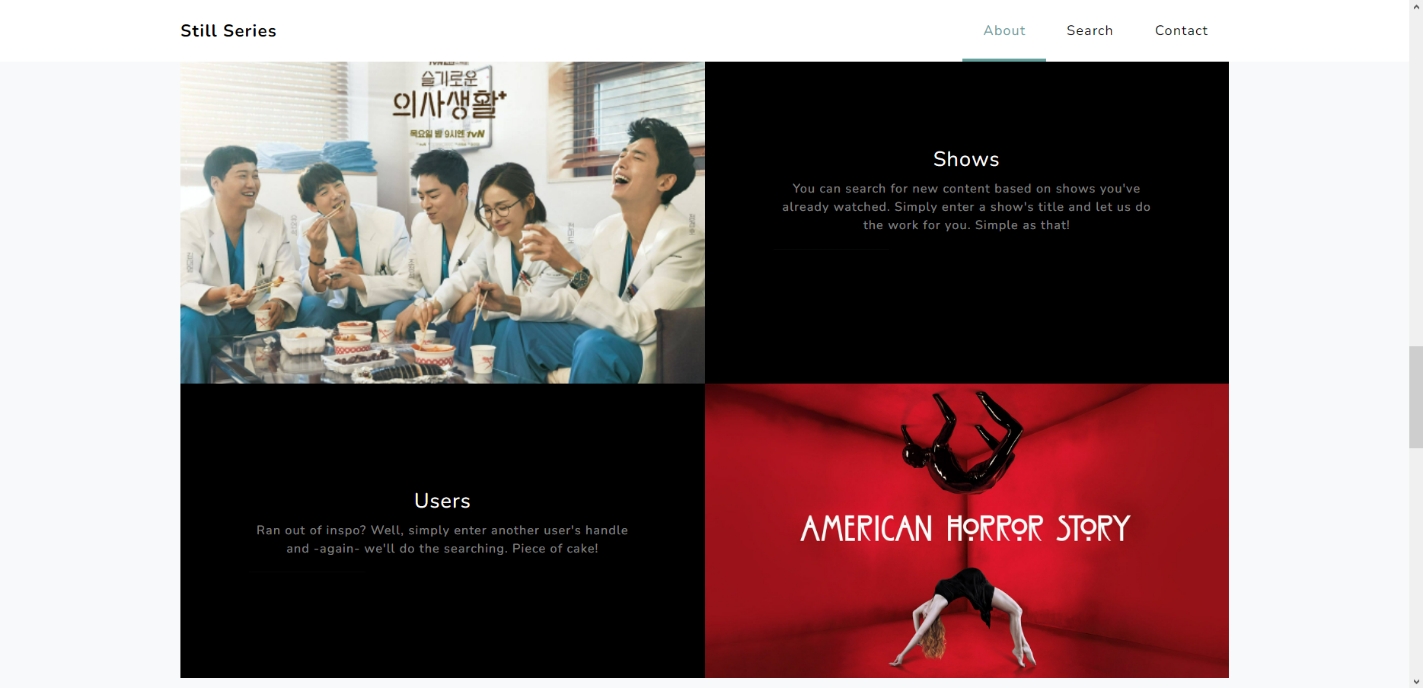
## 3.2. Οδηγίες Εκτέλεσης

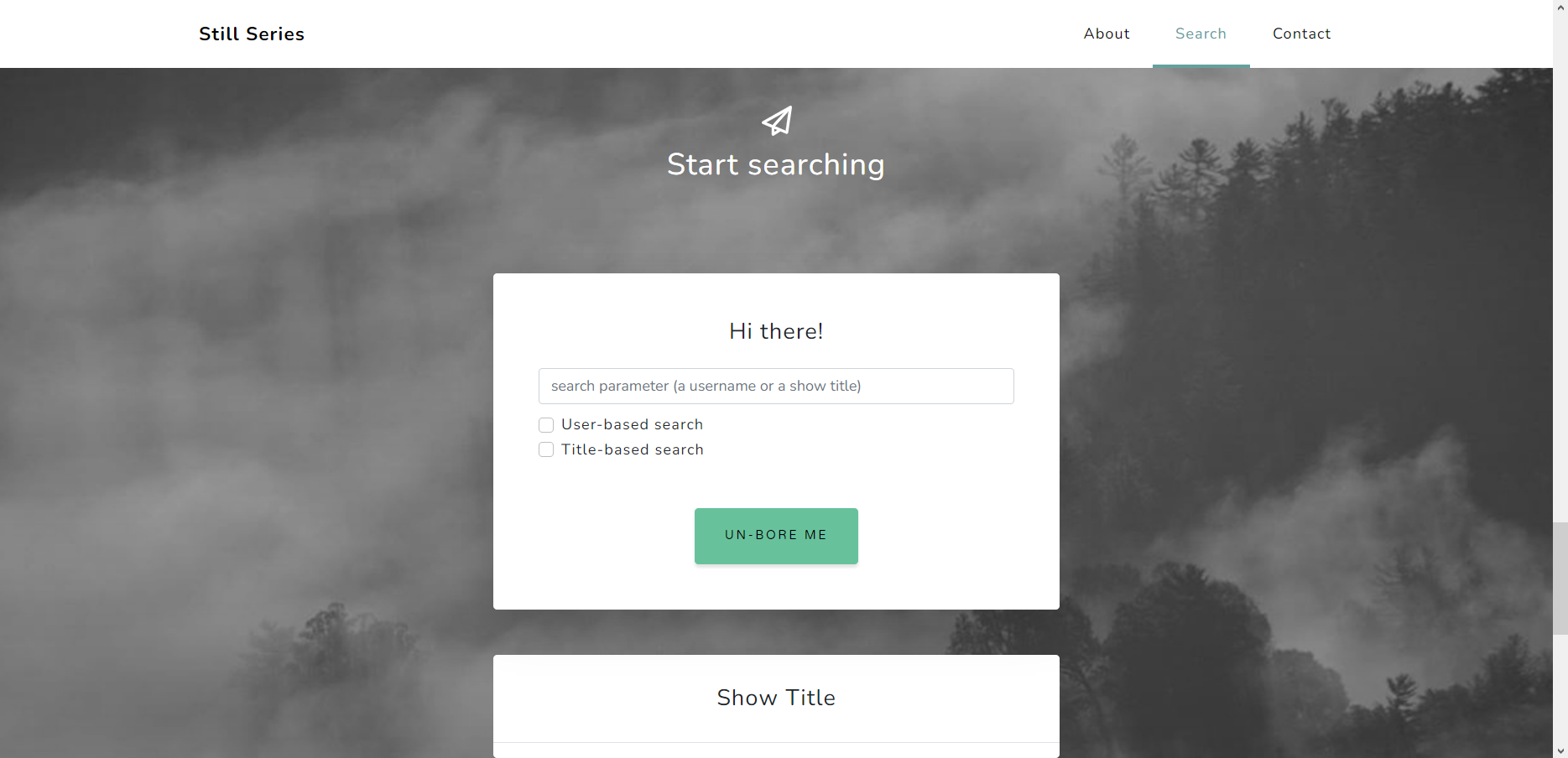
Η διεπαφή του συστήματος αποτελείται από μία ενιαία σελίδα (Εικόνα 2) και είναι διαθέσιμη στη διεύθυνση <http://localhost:5000>, δεδομένου ότι η εφαρμογή έχει εκκινήσει. Στη διεπαφή του χρήστη είναι διαθέσιμες όλες οι πληροφορίες που αφορούν τις παρεχόμενες υπηρεσίες, ενώ προς το τέλος της σελίδας παρέχεται μία φόρμα την οποία ο χρήστης μπορεί να συμπληρώσει, έτσι ώστε να ανακτήσει προτάσεις σχετικά με τηλεοπτικές σειρές που θα μπορούσε να παρακολουθήσει.

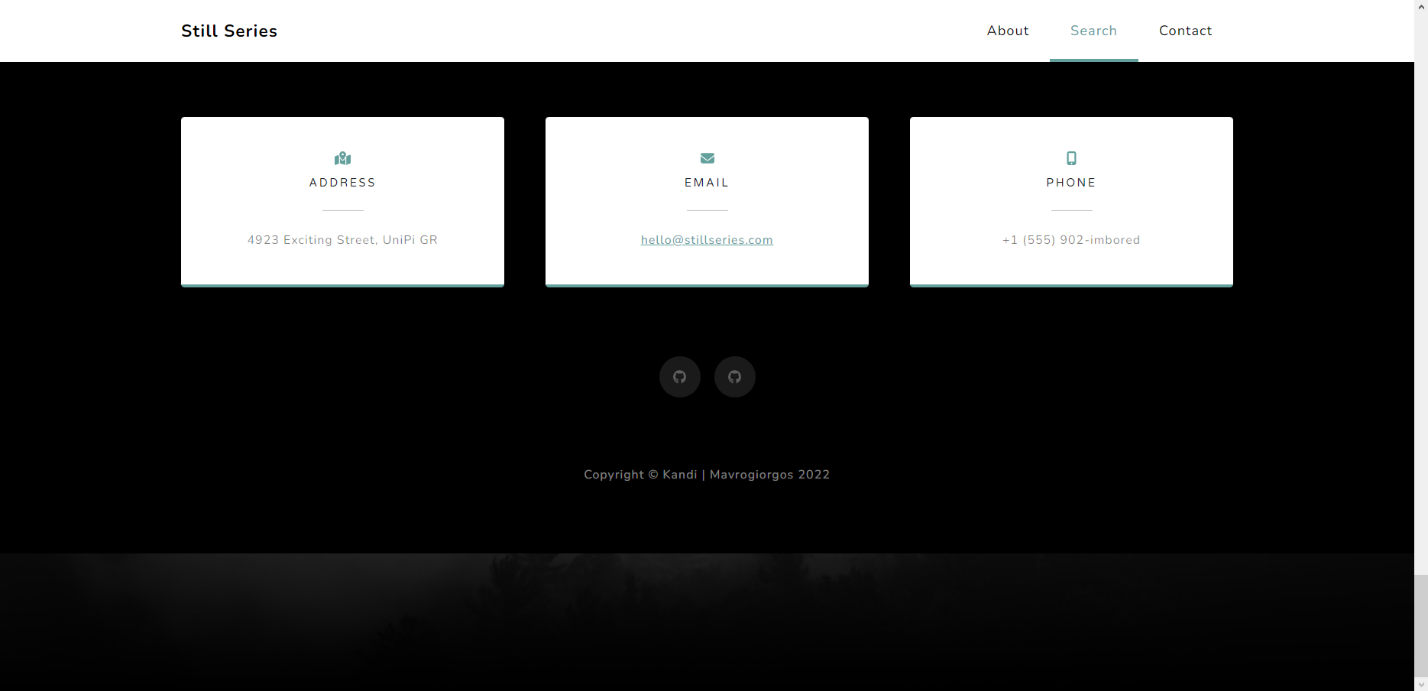












Εικόνα 3: Η διεπαφή χρήστη του Still Series

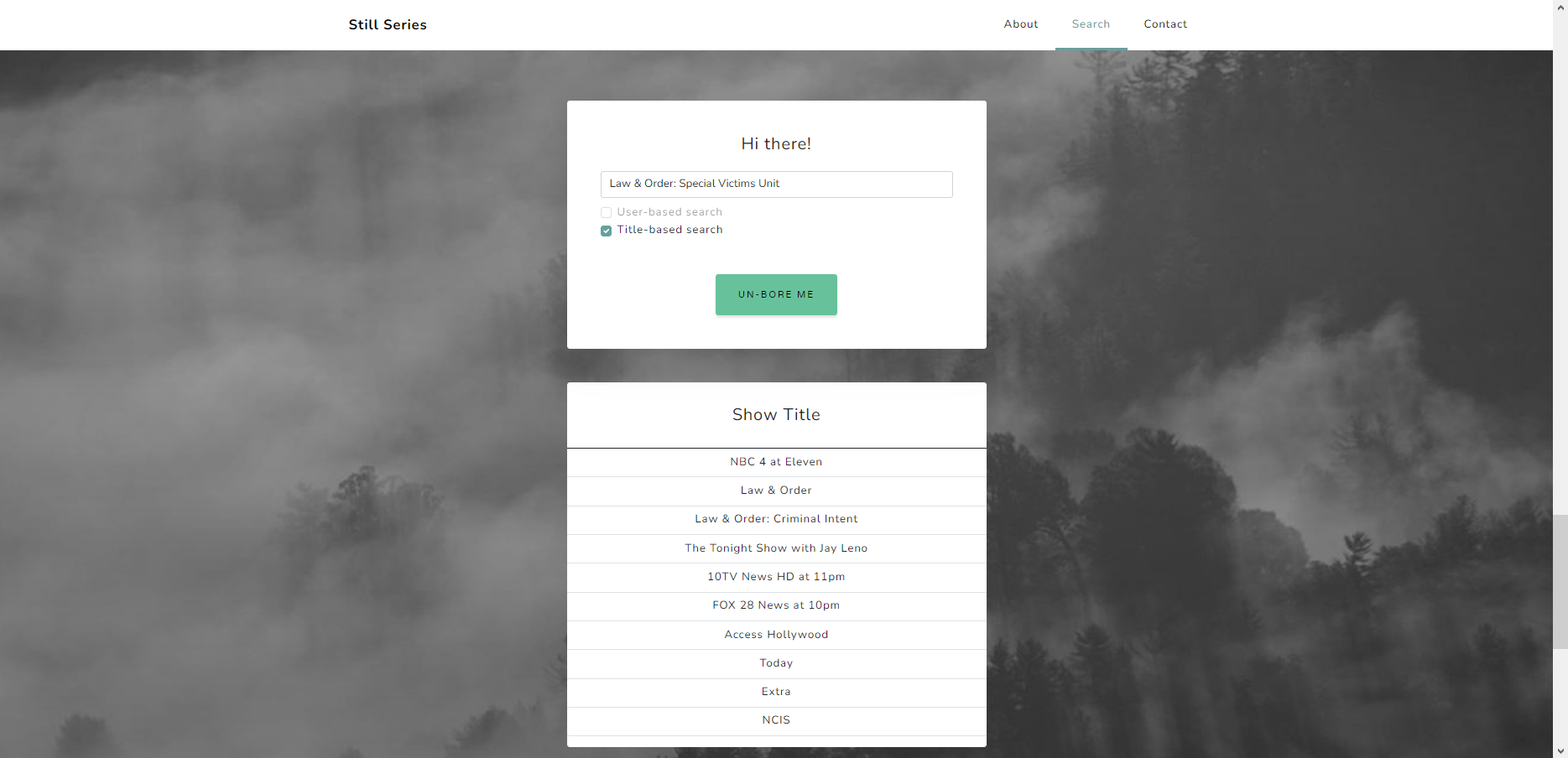
Σχετικά με την παροχή προτάσεων, αυτή μπορεί να πραγματοποιηθεί με δύο (2) τρόπους. Ο πρώτος τρόπος αφορά την παροχή προτάσεων βάσει ομοιότητας μεταξύ χρηστών. Σε αυτή τη περίπτωση, ο χρήστης επιλέγει τον κωδικό κάποιου χρήστη και του επιστρέφονται δέκα (10) προτάσεις με βάση τις σειρές που έχουν παρακολουθήσει άλλοι «παρόμοιοι» χρήστες. Ενδεικτικό παράδειγμα, όπου επιλέγεται ο χρήστης με κωδικό «1», φαίνεται στην Εικόνα 4.

A screenshot of a video game

Description automatically generated with medium confidence

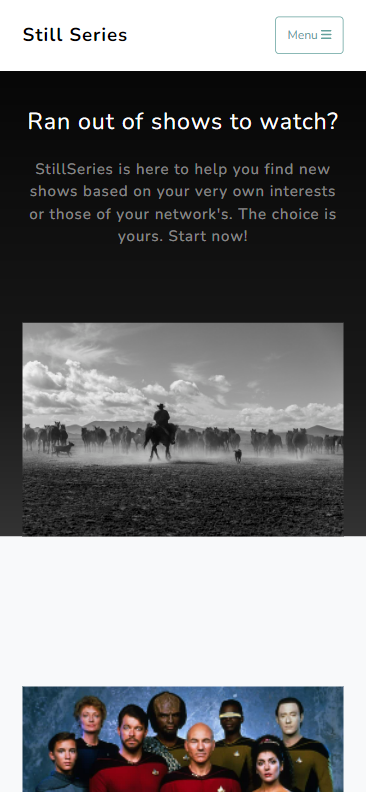
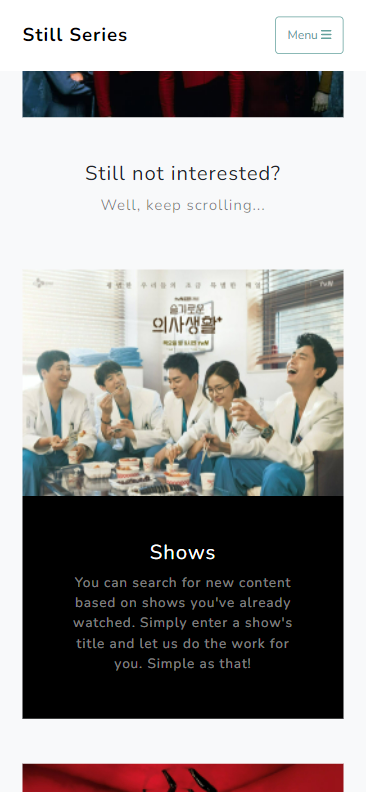
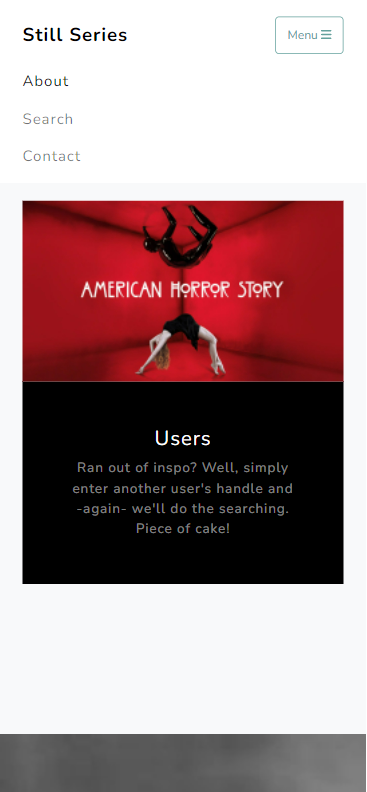
Εικόνα 4: Παροχή προτάσεων βάσει ομοιότητας χρηστών

Ο δεύτερος τρόπος αφορά την παροχή προτάσεων βάσει ομοιότητας αντικειμένων. Σε αυτή τη περίπτωση, ο χρήστης επιλέγει μία σειρά και του επιστρέφονται οι δέκα (10) πιο «παρόμοιες» σειρές σε σχέση με αυτή που επέλεξε. Ενδεικτικό παράδειγμα, όπου επιλέγεται η σειρά «Law & Order: Special Victims Unit», παρουσιάζεται στην Εικόνα 5.

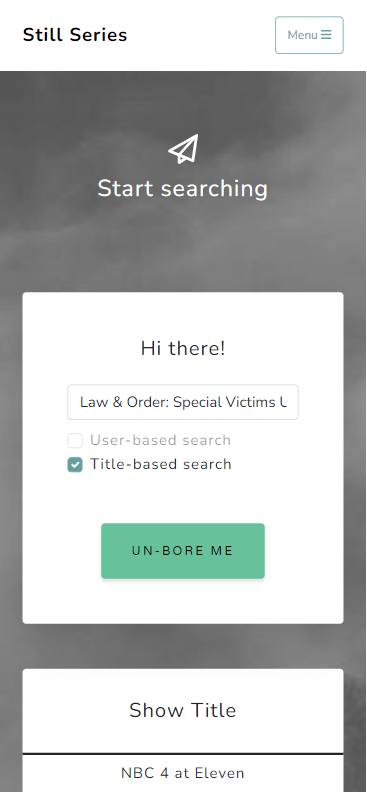


Εικόνα 5: Παροχή προτάσεων βάσει ομοιότητας αντικειμένων (σειρών)

### Απόκριση Εφαρμογής (responsiveness)

Graphical user interface, website

Description automatically generatedΑξίζει να σημειωθεί ότι το σύστημα / εφαρμογή της εργασίας έχει σχεδιαστεί με τέτοιο τρόπο, έτσι ώστε ο χρήστης να μπορεί να έχει πρόσβαση και από διαφορετικές συσκευές, εκτός του υπολογιστή. Πιο συγκεκριμένα, η μορφή της εφαρμογής προσαρμόζεται ανάλογα με την συσκευή που χρησιμοποιεί ο χρήστης. Παρακάτω παρατίθενται στιγμιότυπα οθόνης που αφορούν την χρήση της εφαρμογής σε κινητές συσκευές (Εικόνα 6).



Εικόνα 6: Η διεπαφή του συστήματος, όπως φαίνεται από ένα iPhone X.

Επίσης, αξίζει να σημειωθεί ότι ο κώδικας της εφαρμογής είναι ανεβασμένος και σε ιδιωτικό repository στο GitHub, στη διεύθυνση <https://github.com/ioannakandi/recSys>. Για να έχει κανείς πρόσβαση θα πρέπει να επικοινωνήσει σε ένα από τα emails που αναγράφονται στο εξώφυλλο της εργασίας, έτσι ώστε να προστεθεί ως contributor.

# 4. Συμπεράσματα

Η συγκεκριμένη εργασία είχε ως στόχο την ανάπτυξη μιας web εφαρμογής για την ψηφιοποίηση μιας διαδικασίας στον τομέα της Υγείας. Πιο συγκεκριμένα, αυτή η διαδικασία αφορά τον χρονοπρογραμματισμό ακτινολογικών πράξεων. Πλέον υπάρχει τεράστια ανάγκη για δημιουργία εφαρμογών οι οποίες θα μετατρέπουν τις σχετικές διαδικασίες, όπως για παράδειγμα η παραγγελία εξετάσεων, σε διαδικασίες κατανοητές και φιλικές προς τους χρήστες, ενώ συγχρόνως τις εκτελούν πολύ πιο αποδοτικά.

Σε αυτό το πλαίσιο, η συγκεκριμένη εργασία έχει πετύχει τον στόχο της. Έχει αναπτυχθεί μία web εφαρμογή που αυτοματοποιεί τις διαδικασίες παραγγελίας και εκτέλεσης ακτινολογικών πράξεων, με τρόπο κατανοητό και φιλικό προς τον χρήστη και μάλιστα, αξιοποιώντας ένα ευρέως χρησιμοποιούμενο είδος υπηρεσιών ιστού όπως είναι οι REST. Φυσικά, υπάρχουν και περιθώρια βελτίωσης. Για να μπορέσει να χρησιμοποιηθεί αυτή η εφαρμογή σε σύνθετα σενάρια χρήσης της πραγματικότητας θα πρέπει να πραγματοποιηθούν διάφορες βελτιώσεις όπως: ο αυτόματος χρονοπρογραμματισμός ακτινολογικών πράξεων και η διαμόρφωση πολιτικής ασφάλειας για τα προσωπικά δεδομένα των χρηστών όπως είναι το όνομα χρήστη και ο κωδικός πρόσβασης.

# Βιβλιογραφία

1. The State of Healthcare Industry – Statistics for 2021, https://policyadvice .net /insurance/insights/healthcare-statistics/
2. IBM - The Basics of Business Process Modeling and Notation (BPMN), https://www.ibm.com/cloud/blog/bpmn