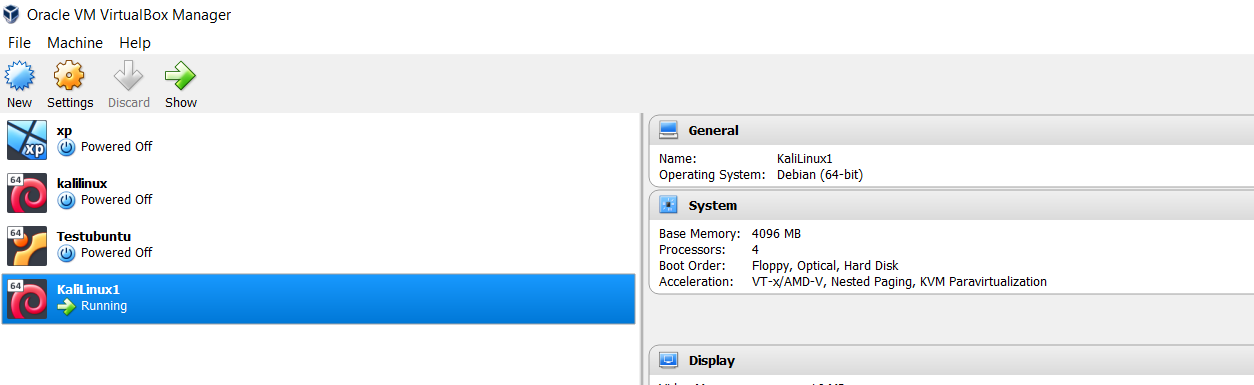


|  |  |
| --- | --- |
| **Στοιχεία Φοιτητή** | |
| Επώνυμο Φοιτητή | Κουσουννής |
| Όνομα Φοιτητή | Κωνσταντίνος |
| Αριθμός Μητρώου | Π14086 |
| Email | Kwstas654321@gmail.com |

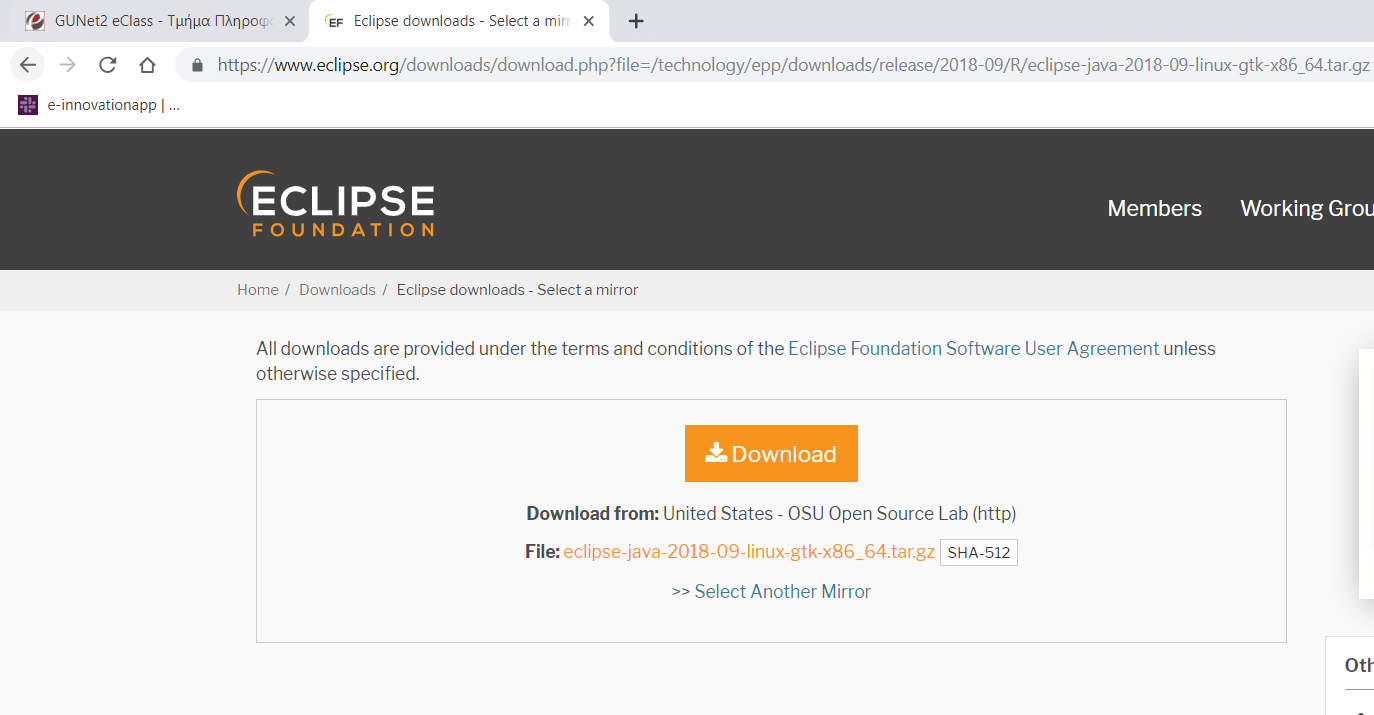
**Οδηγίες Εγκατάστασης**

Η μηχανή αναζήτησης έχει δημιουργηθεί σε λειτουργικό **Kali Linux** το οποίο δημιουργήσαμε μέσα από το **Virtual box** .



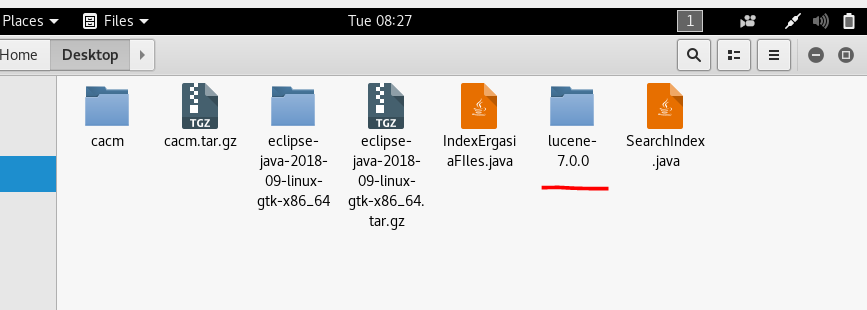
Στο λειτουργικό μας κατεβάσαμε το Eclipse IDE από την ιστοσελίδα

https://www.eclipse.org/downloads/download.php?file=/technology/epp/downloads/release/2018-09/R/eclipse-java-2018-09-linux-gtk-x86\_64.tar.gz



Στην συνέχεια κατεβάζουμε τις βιβλιοθήκες της Lucene την αντίστοιχη έκδοση από την επίσημη ιστοσελίδα της lucene.

<https://lucene.apache.org/core/downloads.html>

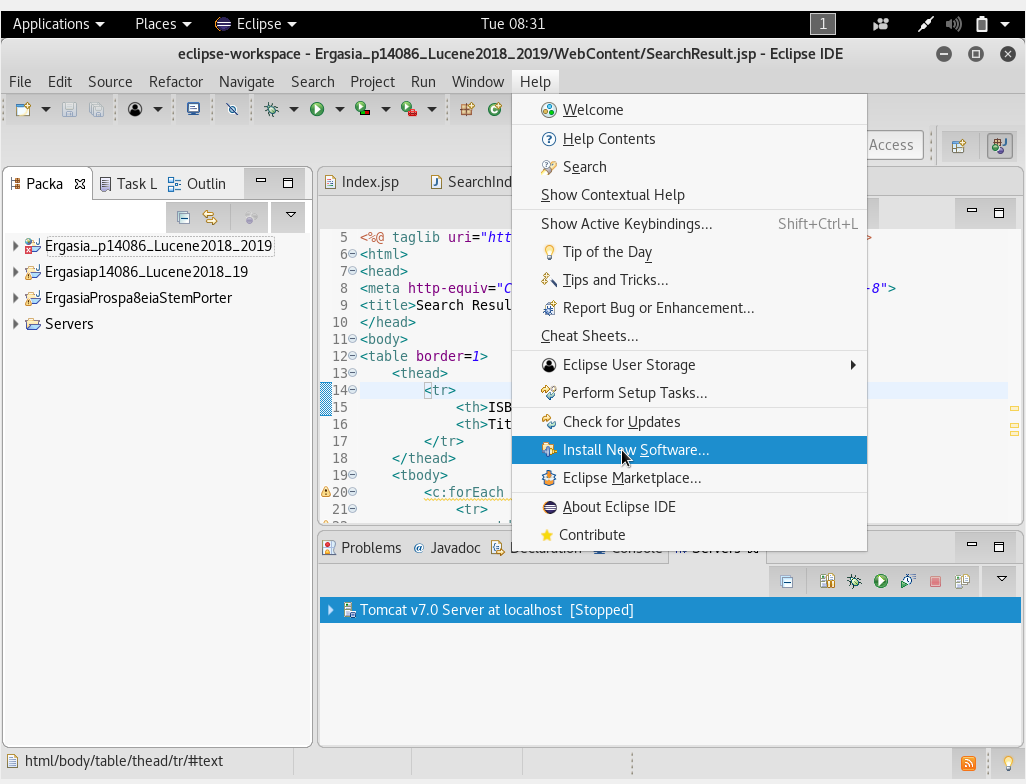


Η δικιά μου έκδοση είναι στην Lucene 7.0.0

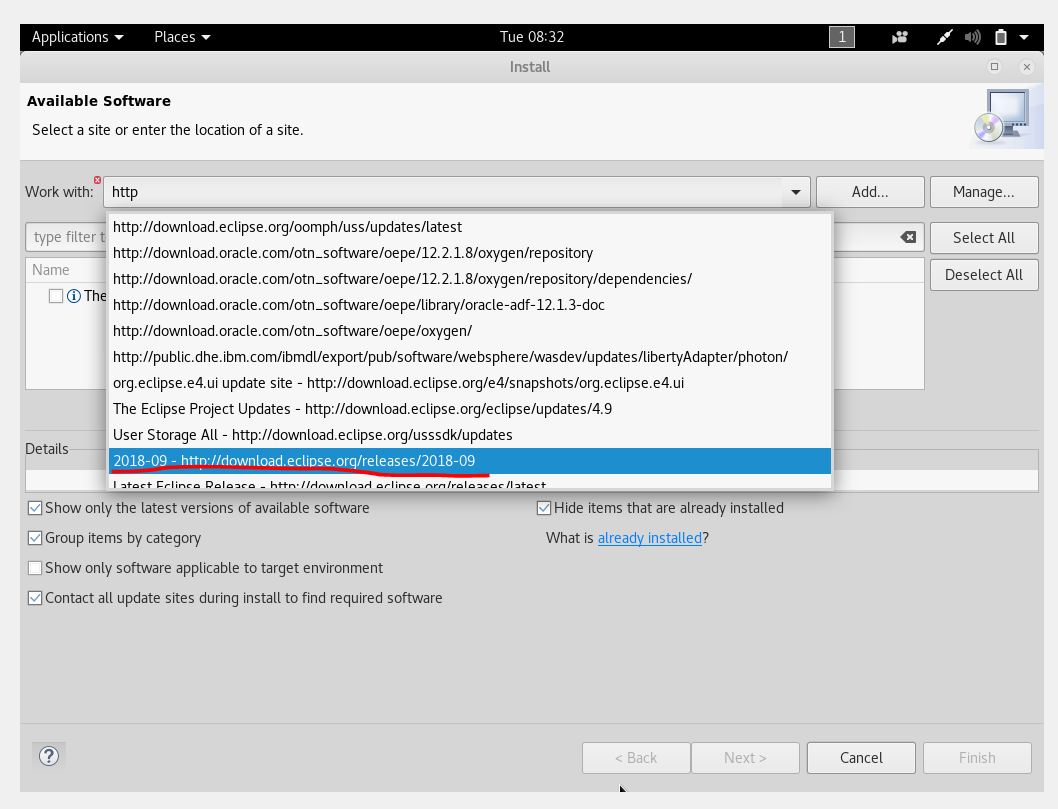
**Δημιουργία Dynamic Web Project**

Στο περιβάλλον eclipse IDE επιλέγουμε να εγκαταστήσουμε το java EE

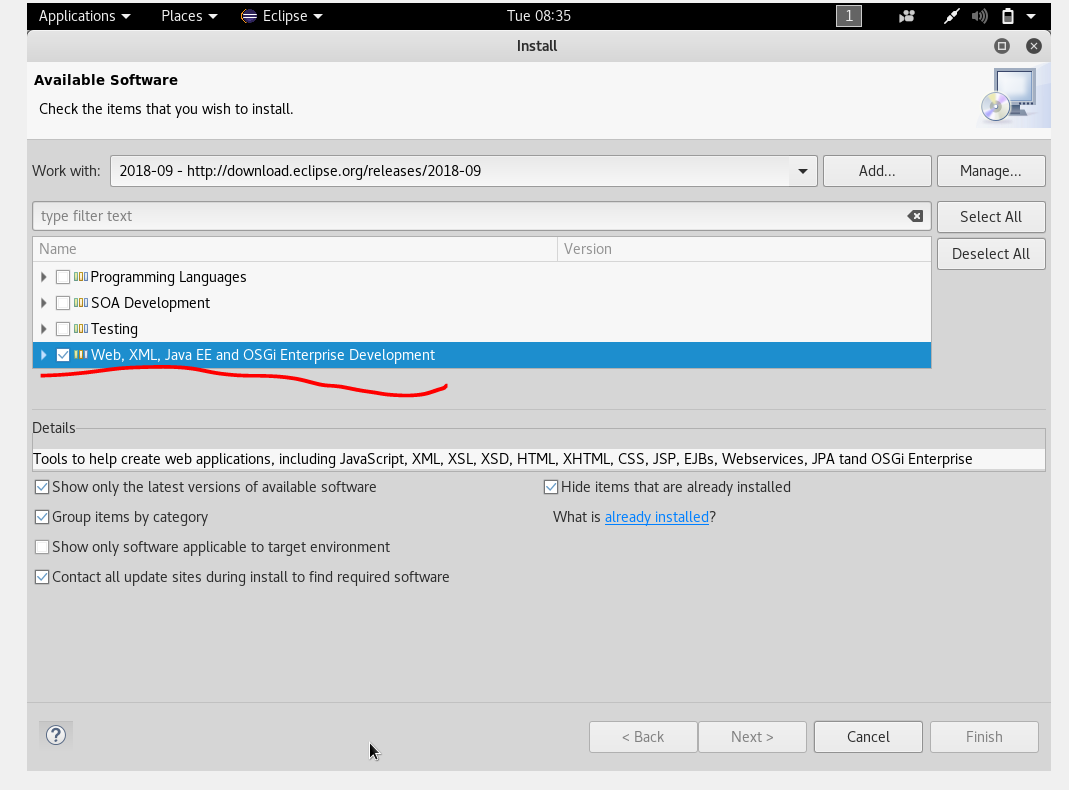
Επιλέγουμε HELP->Install New Software



Στην συνέχεια πληκτρολογούμε **http://** και επιλέγουμε το 2018-2019

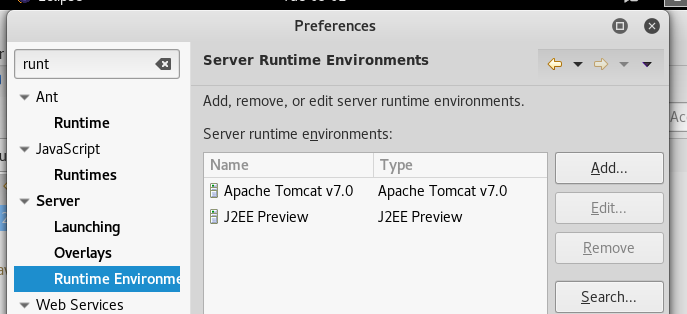


Στην συνέχεια κάνουμε εγκατάσταση τα πακέτα Web,XML,Java EE

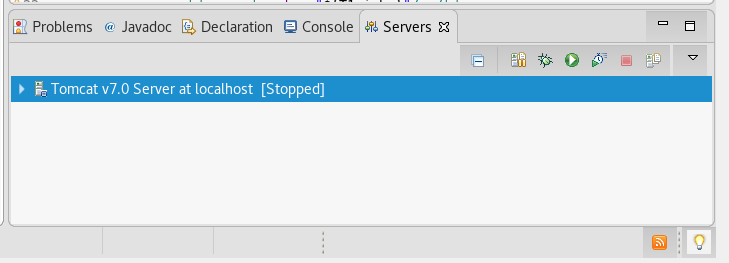


Στην συνέχεια δημιουργούμε έναν Server για να τρέξουμε το πρόγραμμα μας .

Επιλέγουμε Window->Preferences ->Runtime Environment και επιλέγουμε την αντίστοιχη έκδοση του apache που έχουμε.

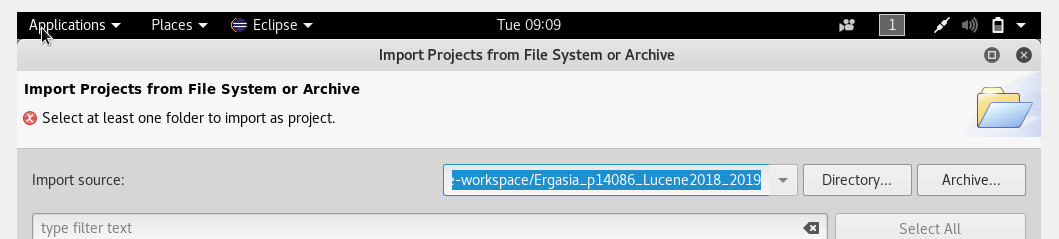


Εμείς εδώ βάλλαμε Tomcat 7.0.0



Οπότε επιλέγουμε να κάνουμε import το Project που έχουμε επιλέγουμε

**File->Open Projects From File Systems**



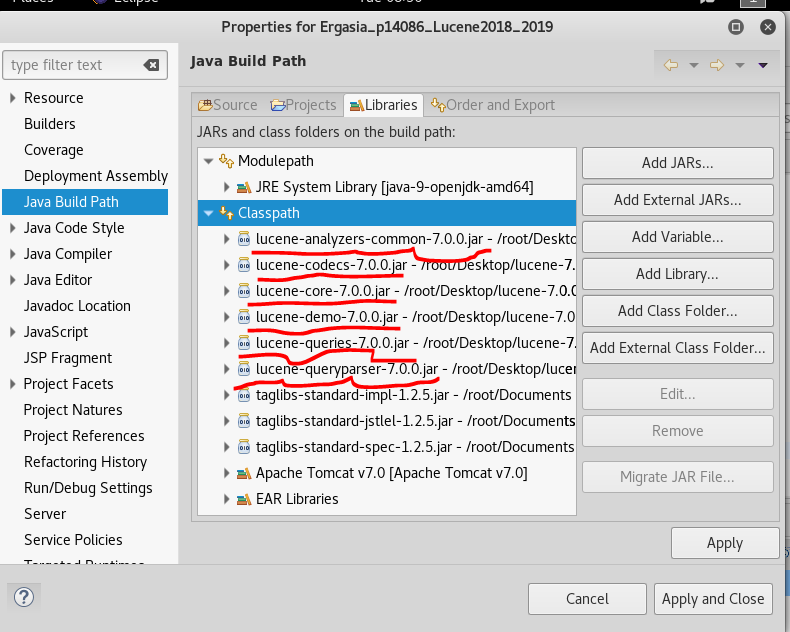
**Ergasia\_p14086\_Lucene2018\_2019**

Όταν κάνουμε import to project **Ergasia\_p14086\_Lucene2018\_2019** θα πρέπει για να τρέξει το πρόγραμμα να βάλουμε τις αντίστοιχες βιβλιοθήκες της Lucene

**Εισαγωγή Βιβλιοθήκης Lucene**

Αφού δημιουργήσουμε το project μας πατάμε δεξί κλικ **build path->configure build path**

**Add external jars** και επιλέγουμε τα αντίστοιχα jars.



Τα υπογραμμισμένα με κόκκινο είναι τα αρχεία με τα οποία τρέχουμε τον κώδικα της Lucene.

Για να τρέξει το πρόγραμμα θα πρέπει να πατήσουμε Add external Jars και να βάλουμε τις αντίστοιχα .jars αρχεία.

**Εκτέλεση Κώδικα**

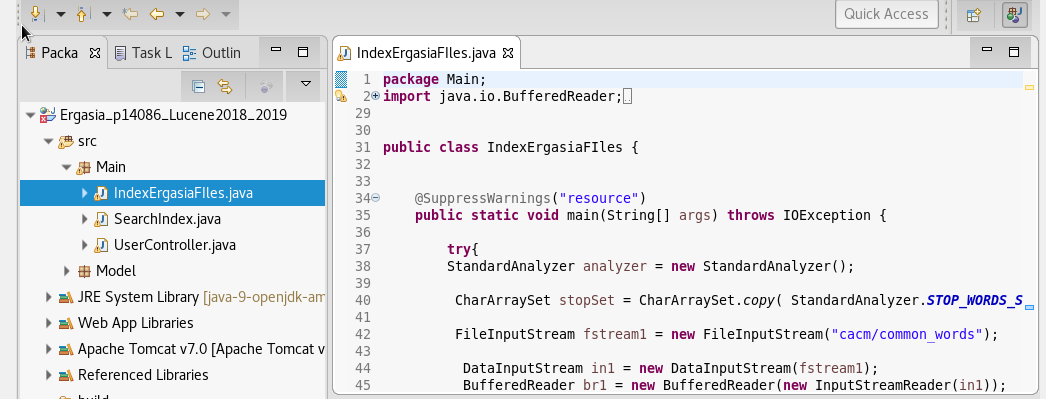
Αρχικά θέλουμε να δημιουργήσουμε τα ευρετήρια μας με την βοήθεια της βιβλιοθήκης της Lucene, τα οποία και θα ψάχνουμε για να μας επιστρέψει τα κατάλληλα κείμενα.

1ο Βήμα

**Δημιουργία Ευρετηρίων**

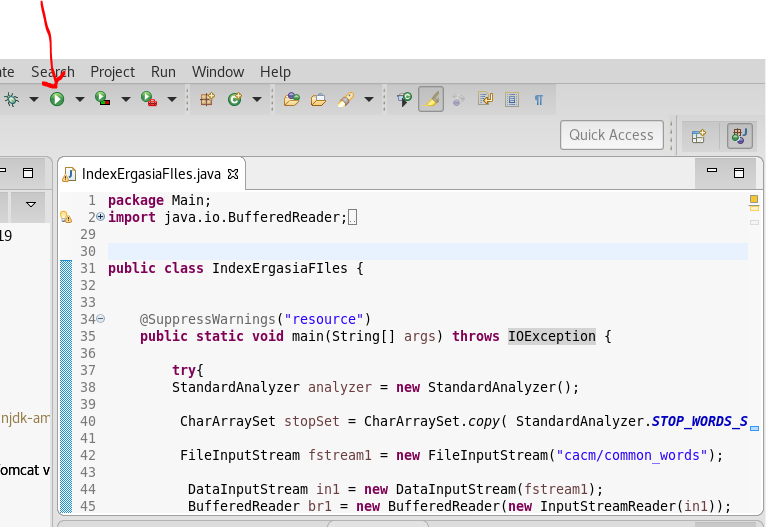
Επιλέγουμε

**Ergasia\_p14086\_Lucene2018\_2019->src->Main->IndexErgasiaFiles.java**

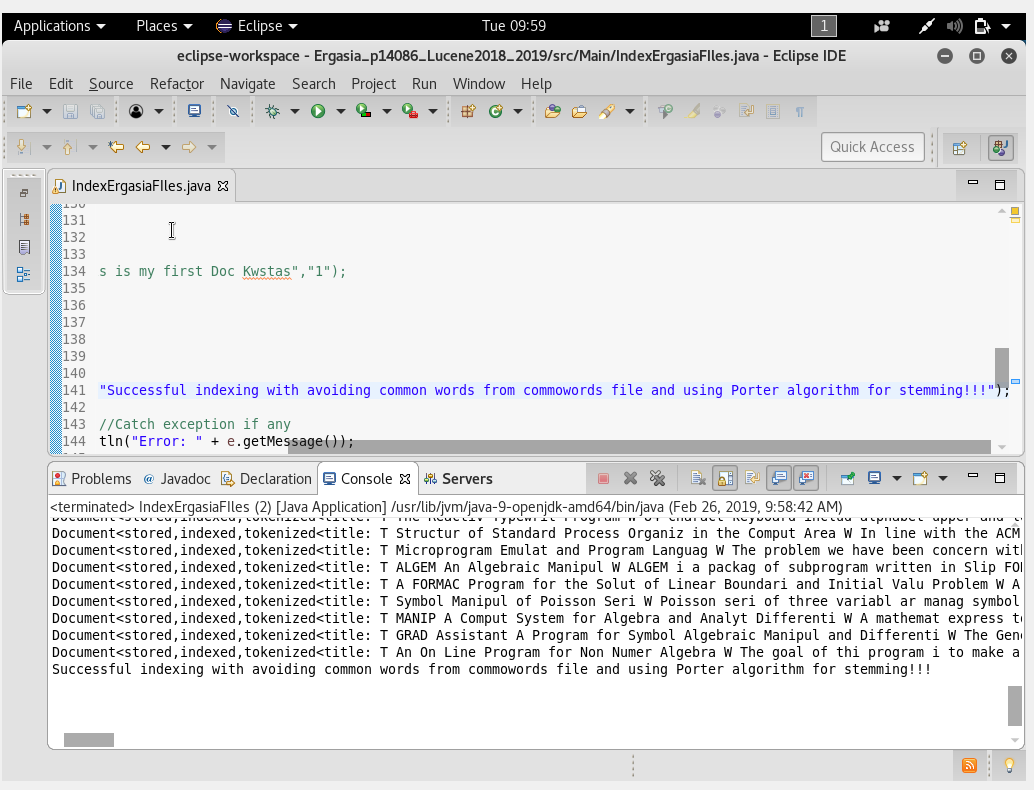


Αυτό είναι ένα java αρχείο το όποιο περιέχει μια public static void main και τρέχει από μόνο του μια φόρα.

Επιλέγουμε να τρέξουμε το αρχείο μας πατώντας play **IndexErgasiaFiles.java**

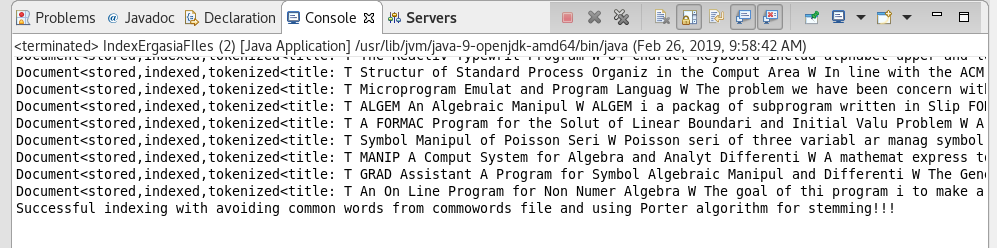


Μόλις τρέξει το αρχείο μας μπορούμε να κοιτάξουμε στην κονσόλα ότι μας έχει ενημερώσει ότι έχει διαβάσει το αρχείο cacm.all και έχει δημιουργήσει τα καταλληλά ευρετήρια αγνοώντας της λέξεις που υπάρχουν από το αρχείο common\_words και εφαρμόζοντας τον αλγόριθμο Porter για Stemming



Όπως βλέπουμε μπορούμε να παρατηρήσουμε ότι μας γράφει στα αποτελέσματα όλα τα κείμενα 3204 κείμενα τα οποία έγιναν με την βοήθεια της Lucene σε μορφή ευρετηρίου εφαρμόζοντας τον αλγόριθμο Porter για Stemming

Και αγνοώντας της λέξεις που υπήρχαν στο αρχείο common\_words.



Τέλος μας εμφανίζει ένα απλό μήνυμά επιβεβαίωσης

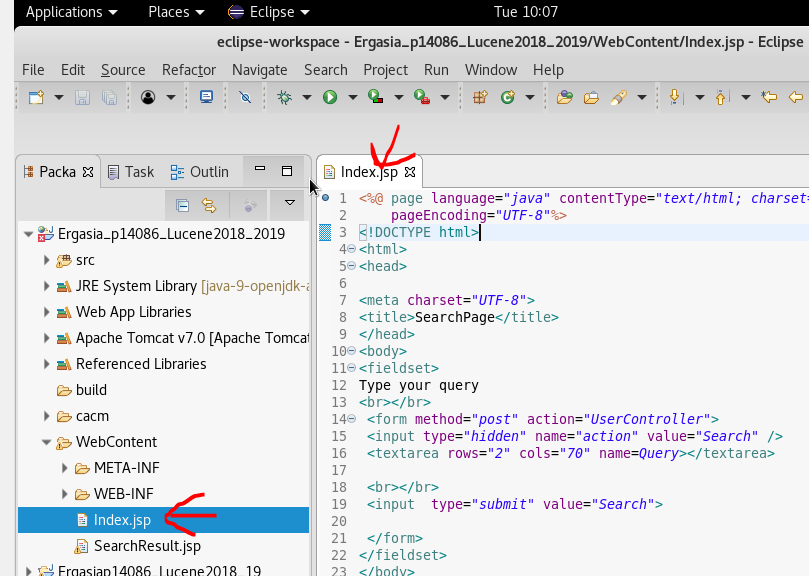
**Successful indexing with avoiding common words from commowords file and using porter algorithm for stemming.**

**2ο Βήμα**

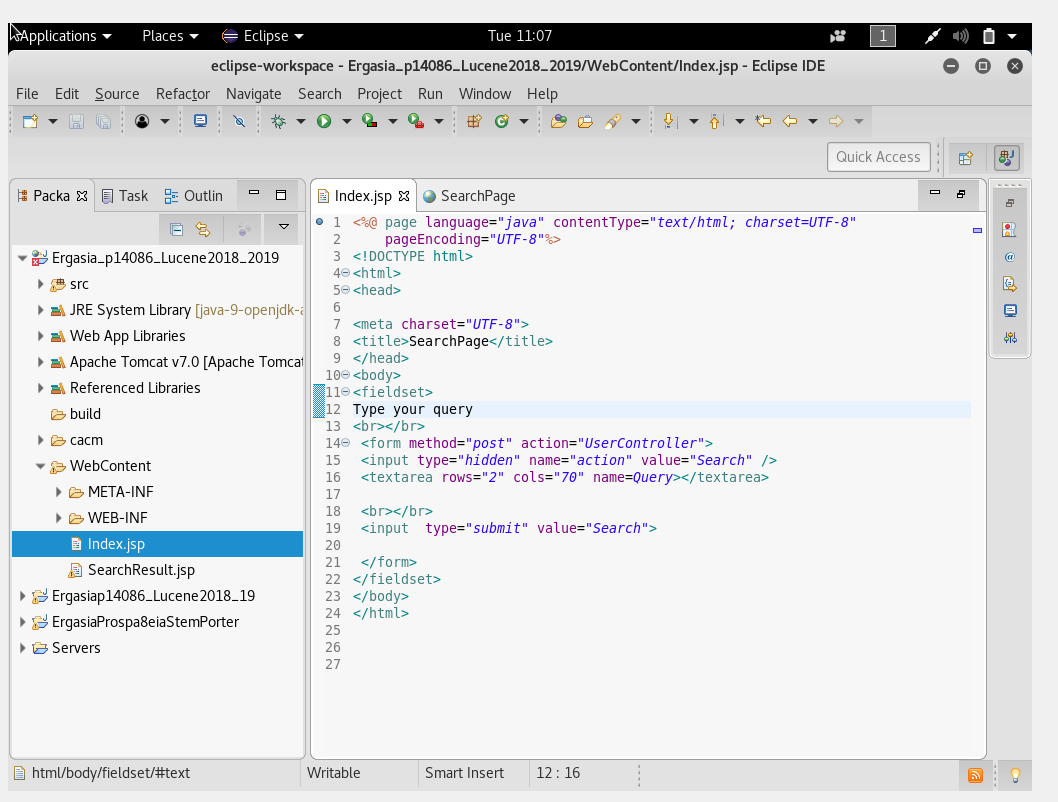
**Ερωτήματα Αναζήτηση Κειμένων**

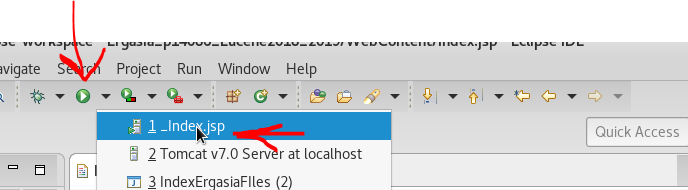
Επιλέγουμε το αρχείο

**Ergasia\_p14086\_Lucene2018\_2019->WebContent->Index.jsp**

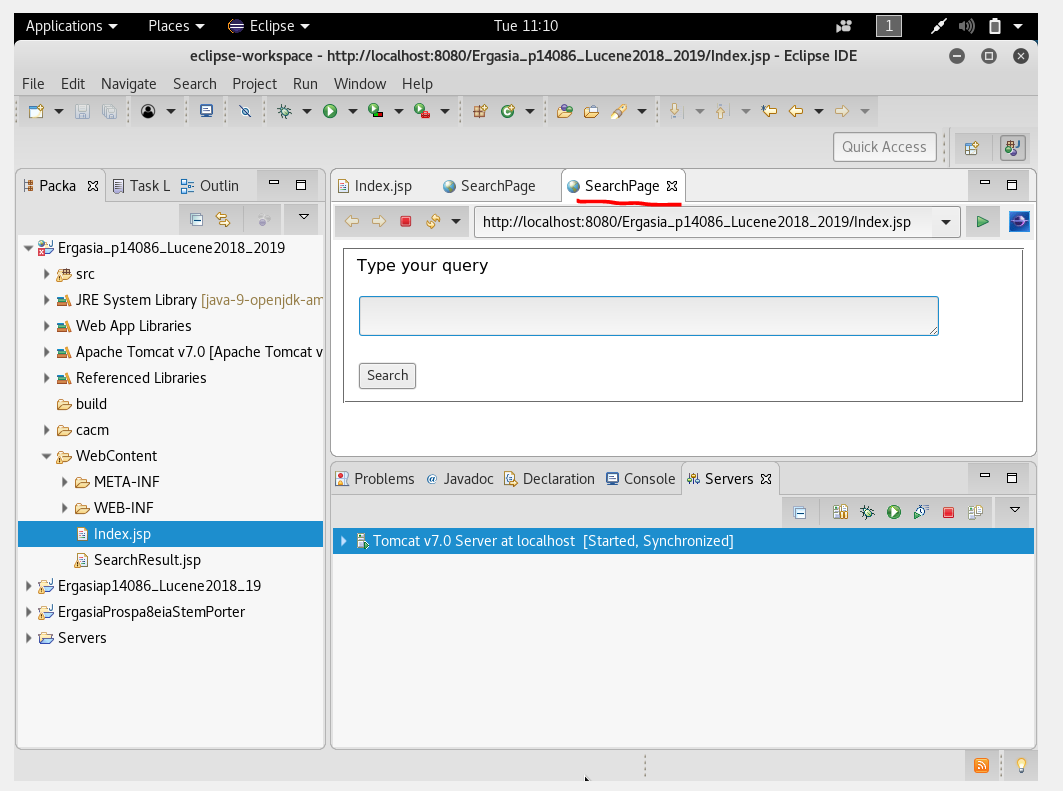


Αφού ανοίξουμε το αρχείο μας επιλέγουμε να το τρέξουμε πάνω στον server.





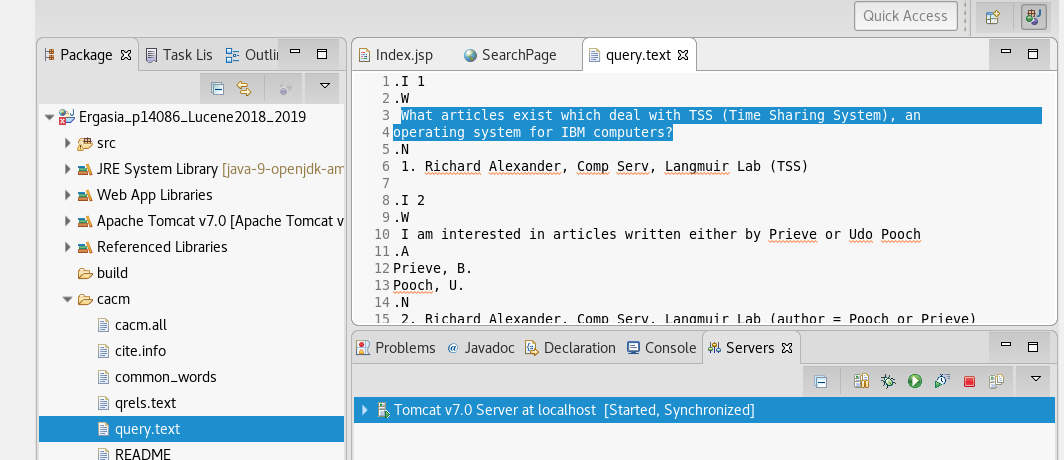
Όταν τρέξουμε το αρχείο πάνω στον server θα μας ανοίξει την αρχική ιστοσελίδα μας.



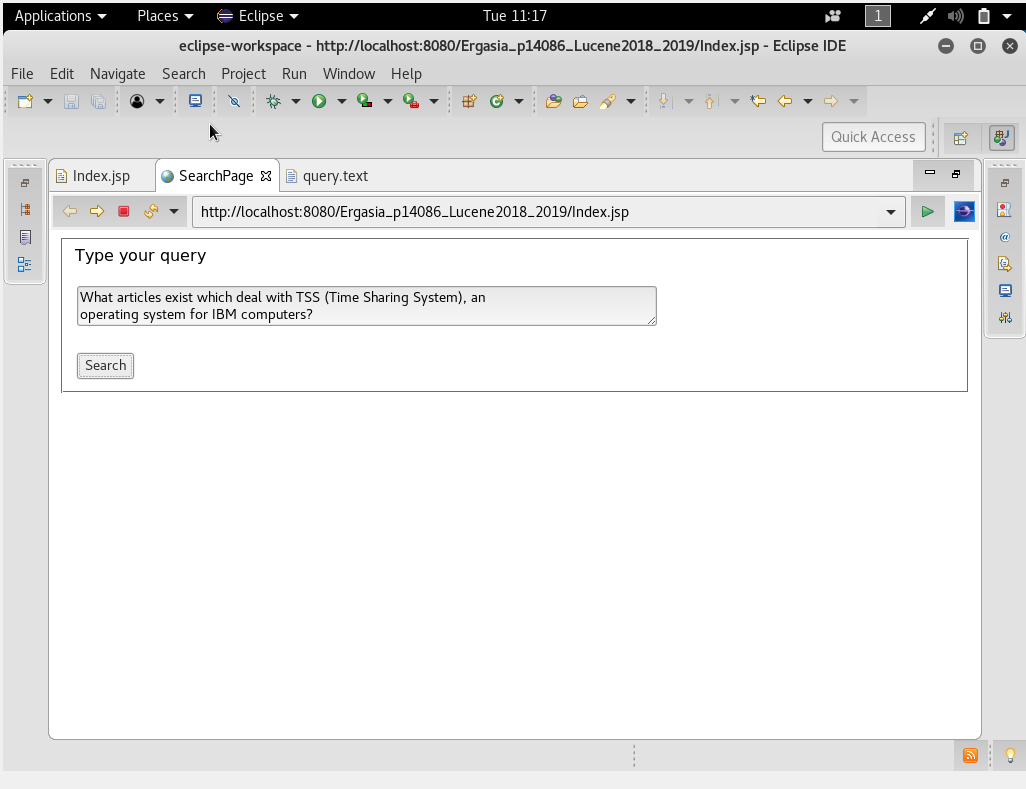
Όπως βλέπουμε έχει ανοίξει η ιστοσελίδα και μπορούμε να πληκτρολογήσουμε το αντίστοιχο ερώτημα μας στο διαθέσιμο πλαίσιο

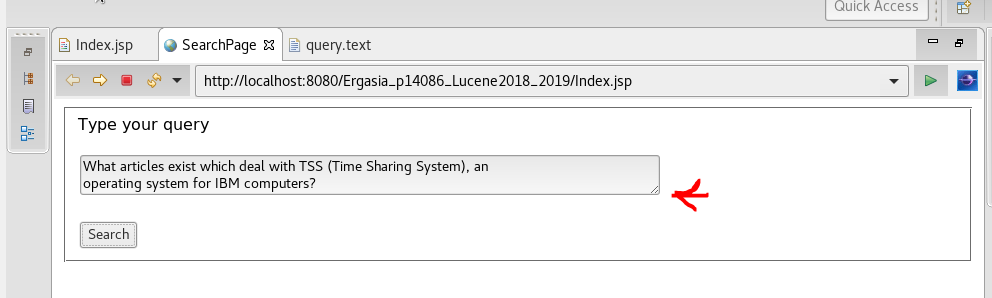
**Παράδειγμα**

Παίρνω το πρώτο ερώτημα από τον φάκελο CACM/query.txt



Πληκτρολογώ την ερώτηση μου πάνω στο πλαίσιο της ιστοσελίδας.



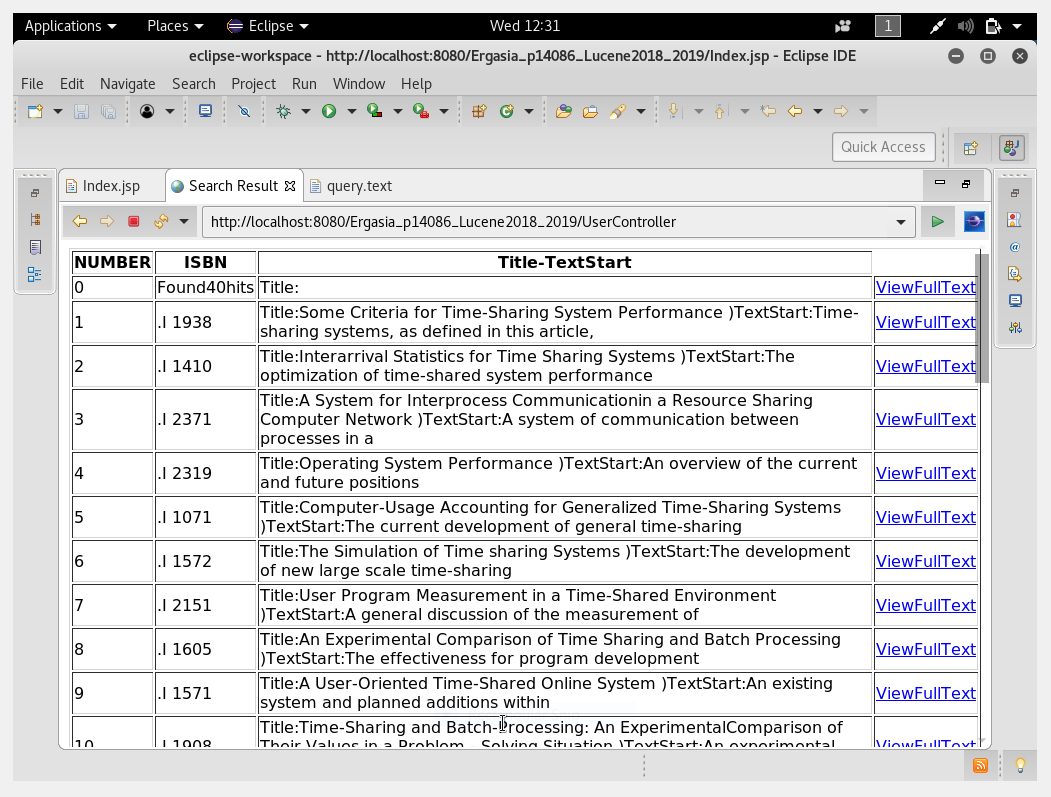


Το πλαίσιο που πληκτρολογούμε την ερώτηση μας είναι resizable δηλαδή μπορούμε να ρυθμίσουμε το πλαίσιο για να εμφανίζεται όλη η ερώτηση μας.

**Προσοχή !!!!**

Αν δεν έχουμε τρέξει στην αρχή το αρχείο για να δημιουργήσουμε τα ευρετήρια τότε θα μάς εμφανίσει ένα μήνυμα λάθους διότι δεν βρίσκει να διαβάσει κανένα ευρετήριο(**IndexErgasiaFiles.java**).

**Επιλέγουμε Search**



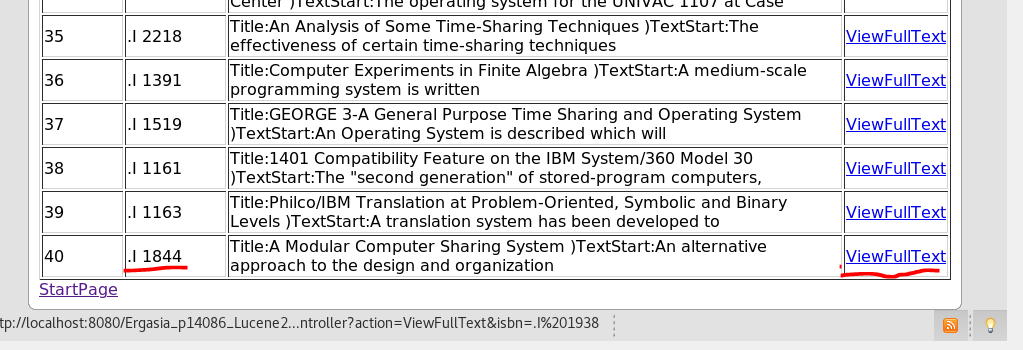
Στην πρώτη στήλη μας εμφανίζει τον αριθμό του κειμένου με σειρά προτεραιότητας (**NUMBER**)

Στην δεύτερη στήλη (**ISBN)** μας εμφανίζει τον αριθμό του κειμένου που μας επέστρεψε η βιβλιοθήκη Lucene .

Στην Τρίτη στήλη (**Title-Text Start**) μας εμφανίζει τον Τίτλο του κειμένου και ένα κομμάτι από το κανονικό κείμενο μας

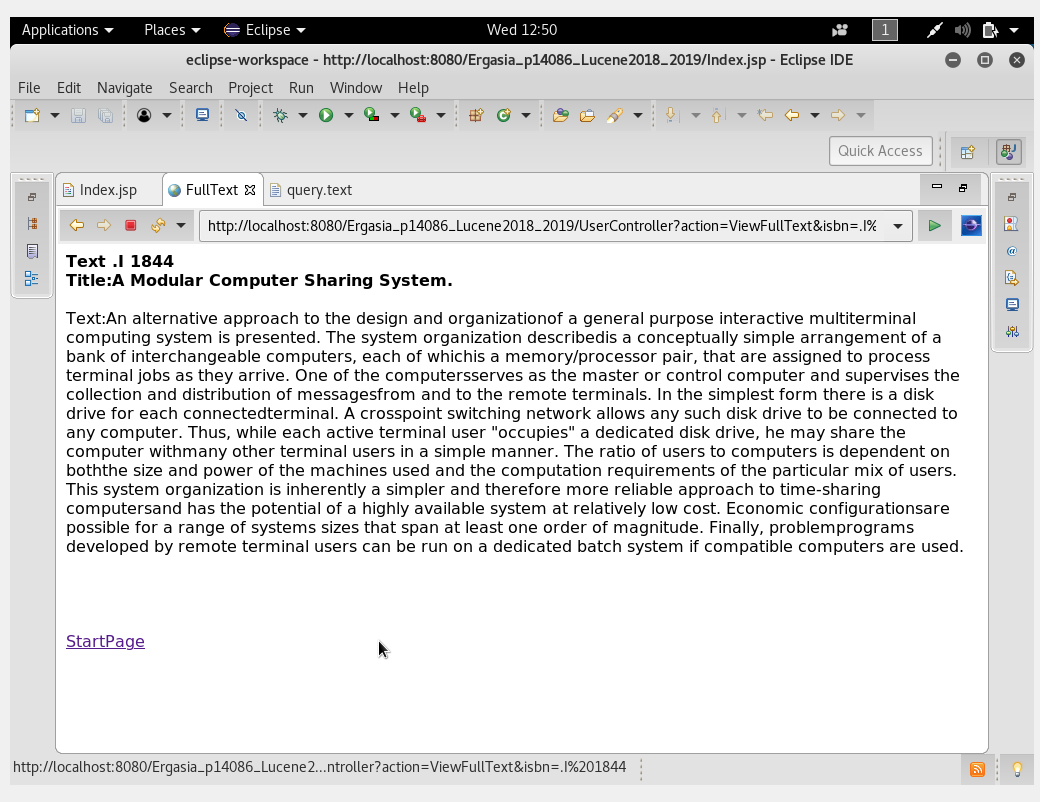
Στην Τέταρτη στήλη έχουμε την επιλογή να διαβάσουμε όλο το κείμενο για κάθε κείμενο που μας επιστράφηκε.

Έστω ότι επιλέγω να διαβάσω το τελευταίο κείμενο που μας εμφανίζει.



Αυτό είναι το .Ι 1844

Επιλέγω View Full Text



Όπως βλέπουμε μας εμφανίζει τον αριθμό του κειμένου ( **Isbn** ) .I 1844 ,τον τίτλο του κειμένου (**Title)** ,αναλυτικά το κείμενο μας (**text**) .

**Επιστροφή στην Αρχική Σελίδα**

Κάτω αριστερά σε κάθε σελίδα έχουμε την επιλογή να επιστρέψουμε στην αρχική και σελίδα και να κάνουμε ένα νέο ερώτημα **StartPage .**