

Οδηγίες Εκτέλεσης Αρχείου Beta_1.4.jar

Προϋποθέσεις

1. Να είναι εγκατεστημένη η τελευταία έκδοση της java (αν όχι η τελευταία, τουλάχιστον πακέτο της 8^{ης} έκδοσης (JAVA SE 8))
2. Να υπάρχει ένα σύστημα επεξεργασίας βάσεων δεδομένων. Προτείνεται ο SQL Server 2014 Express. Οδηγίες για την εγκατάσταση θα βρείτε [εδώ](#).

ΠΡΟΣΟΧΗ: Στο 9'48" του βίντεο επιλέγουμε **ΥΠΟΧΡΕΩΤΙΚΑ** στο Authentication Mode το Mixed Mode για να έχουμε πρόσβαση στη βάση.

3. Πρέπει επίσης να έχει δημιουργηθεί μια βάση δεδομένων στην οποία μέσα πρέπει να υπάρχει πίνακας με δύο στήλες τύπου VARCHAR(2000) (αυτό προτείνεται)

Εκτέλεση

Για τις ανάγκες των εικόνων το jar αρχείο βρίσκεται στην επιφάνεια εργασίας (εσείς το βάζετε όπου θέλετε)

1. Κάνουμε εξαγωγή του αρχείου Beta.jar σε ένα φάκελο είτε στα έγγραφα μας είτε στην επιφάνεια εργασίας (ή οπουδήποτε θέλουμε)

```
C:\> Γραμμή εντολών
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. Με επιφύλαξη κάθε νόμιμου δικαιώματος.
C:\Users\Vaggelis>
```

2. Ανοίγουμε το command window και πληκτρολογούμε τα εξής:

cd (το path του φακέλου στον οποίο βρίσκεται το jar αρχείο)

```
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. Με επιφύλαξη κάθε νόμιμου δικαιώματος.
C:\Users\Vaggelis>cd Desktop
C:\Users\Vaggelis\Desktop>
```

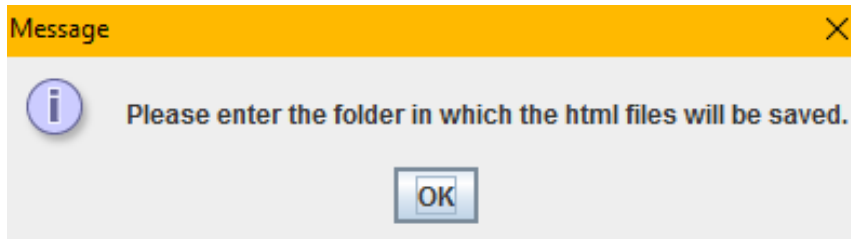
3. Για να τρέξουμε το αρχείο πληκτρολογούμε:

ο crawler αυτόματα τρέχει εκτός από το λινκ που του δίνει ο χρήστης τα εξής: edition.cnn.com, www.bloomberg.com/europe, www.businessinsider.com, www.nationalgeographic.com, www.food.com, www.imdb.com

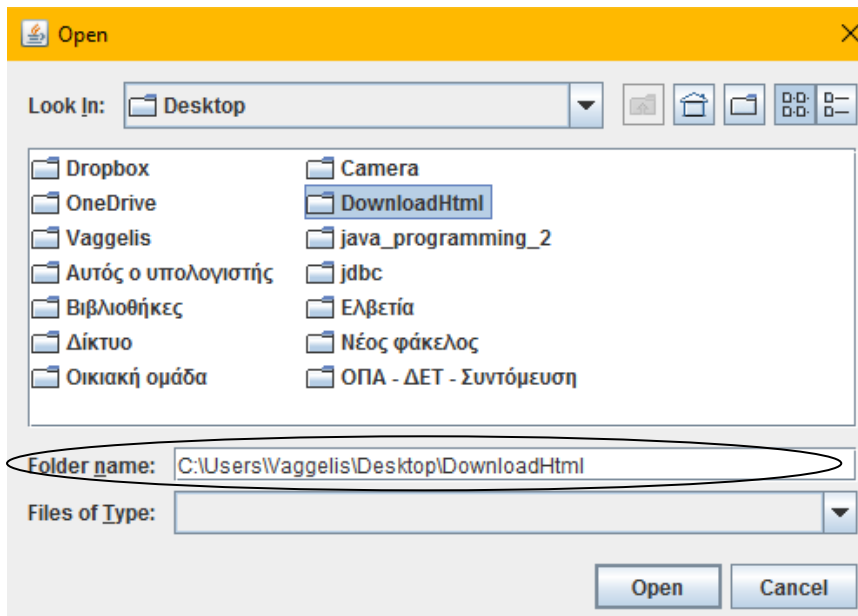
java -jar Beta_1.4.jar "(το λινκ που θέλουμε να ξεκινήσει ο crawler)"

```
C:\> Γραμμή εντολών
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. Με επιφύλαξη κάθε νόμιμου δικαιώματος.
C:\Users\Vaggelis>cd Desktop
C:\Users\Vaggelis\Desktop>java -jar Beta_1.4.jar "http://www.tovima.gr/"
```

4. Εμφανίζεται αυτό το μήνυμα

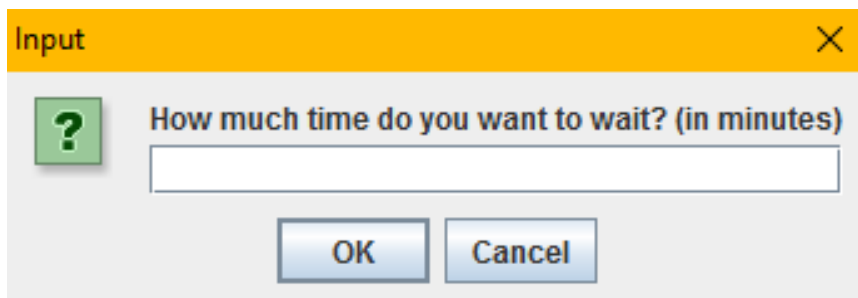


και μετά αυτό

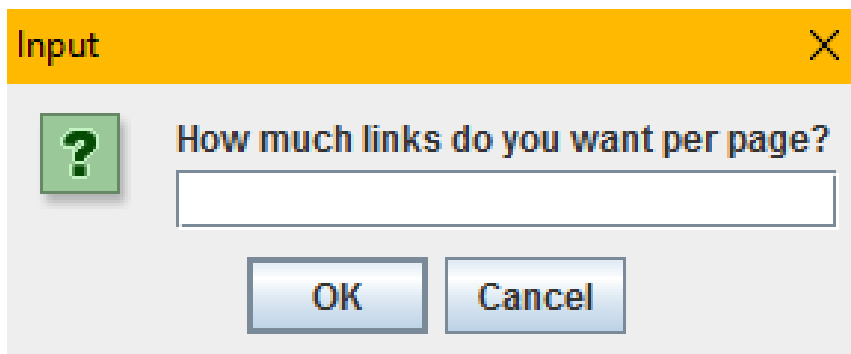


Επιλέγουμε τον φάκελο στον οποίο θα γίνει η αποθήκευση των αρχείων πατώντας Open

5. Στην ερώτηση «How much time do you want to wait? (in minutes)» απαντάμε με ακέραιο αριθμό



6. Στην ερώτηση «How much links do you want per page?» απαντάμε με ακέραιο αριθμό



7. Το πρόγραμμα τρέχει και για να δείτε ότι τρέχει σωστά πρέπει να εμφανίζει κάτι τέτοιο

Γραμμή εντολών - java -jar Beta_1.3.jar "http://www.tovima.gr/"

```
Microsoft Windows [Version 10.0.14393]
```

(c) 2016 Microsoft Corporation. Με επιφύλαξη κάθε νόμιμου δικαιώματος.

```
C:\Users\Vaggelis>cd Desktop
```

```
C:\Users\Vaggelis\Desktop>java -jar Beta_1.3.jar "http://www.tovima.gr/"
```

```
** Queue Size: 6 | Links Count: 0 | Checking 1 link: http://www.tovima.gr/
```

--- Link came up: <http://www.tovima.gr/sports/article/?aid=854587#commentForm>

```
---- Link came up:    http://www.tovima.gr/science/technology-planet/
```

```
---- Link came up:  http://www.tovima.gr/culture/people/
```

```
---- Link came up:  http://www.tovima.gr/en
```

```
---- Link came up: http://www.tovima.gr/blogs/?mode=3
```

```
---- Link came up:    http://www.tovima.gr/feed/editors/?edid=721
```

```
---- Link came up:      http://www.tovima.gr/feed/editors/?edid=963
```

```

---- Link came up:  http://www.helioskiosk.gr/defaultm.aspx?page=webproduct&id=31

```

---- Link came up: <http://www.tovima.gr/opinions/article/?aid=853884&h1=true#commentForm>

```
---- Link came up:      http://www.tovima.gr/feed/editors/?edid=7019
```

```
** Queue Size: 15 | Links Count: 10 | Checking 2 link: http://www.cnn.gr/
```

```
---- Link came up:    http://www.cnn.gr/terms
```

```
---- Link came up: https://www.google.com/+cnngreece
```

```
---- Link came up: http://www.cnn.gr/videos/viral
```

```
---- Link came up: https://www.facebook.com/cnn.greece
```

```
---- Link came up:    http://www.cnn.gr/eidhseis
```

```
---- Link came up:    http://www.mothersblog.gr/
```

```
---- Link came up: http://www.cnn.gr/insights
```

```
---- Link came up:    http://www.cnn.gr/travel/kosmo
```

```
---- Link came up:    http://www.cnn.gr/more/widgets
```

```
---- Link came up:      http://us.cnn.com/
```

```
** Queue Size: 24 | Links Count: 20 | Checking 3 link: http://www.bbc.com/
```

```

---- Link came up:    http://www.bbc.com/future/story/20161227-how-the-hoverboard-went-from-fiction-to-fact

```

```
---- Link came up:    http://www.bbc.com/autos
```

```
---- Link came up:    http://www.bbc.co.uk/cymru/
```

```
---- Link came up:    http://www.bbc.com/earth/
```

---- Link came up: <http://www.bbc.com/travel/story/20160620-the-us-one-million-acre-secret>

```
---- Link came up:    http://www.bbc.com/iwonder
```

```
---- Link came up: http://www.bbc.com/tamil,
```

---- Link came up: http://www.bbc.com/news/england/birmingham_and_black_country

```
---- Link came up:  http://www.bbc.com/persian/
```

```
---- Link came up: http://www.bbc.com/indonesia/
```

```
** Queue Size: 33 | Links Count: 30 | Checking 4 link: http://www.newsit.gr/
```

```
---- Link came up:    http://www.newsit.gr/default.asp?ID=107
```

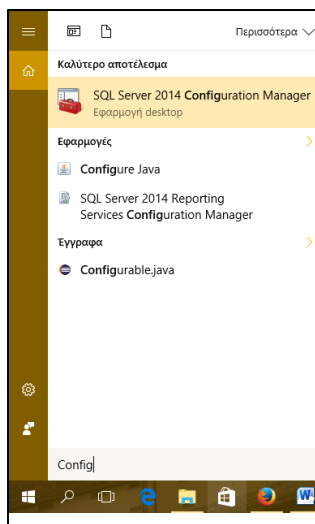
8. Όταν περάσει ο χρόνος που έχετε ορίσει ή απλά τελειώσει η λίστα με τα λινκ (αυτό θα συμβεί σπάνια) εμφανίζεται αυτή η οθόνη. Κάτω αριστερά δίπλα από την ένδειξη END φαίνεται πόσα λινκ βρέθηκαν και αποθηκεύτηκαν και πιο δίπλα βλέπετε πόσα απορρίφθηκαν γιατί υπήρχαν ήδη.

```

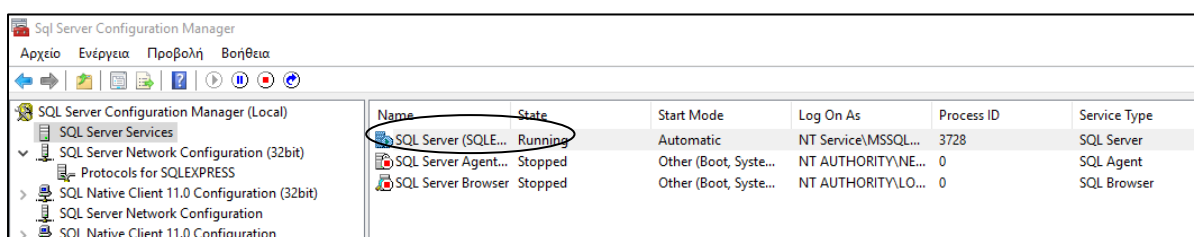
---- Link came up: http://www.newsit.gr/texnologia
---- Link came up: http://www.newsit.gr/tags/εεεεεεεεεεεε
---- Link came up: http://www.newsit.gr/rss/artrss.php
---- Link came up: http://www.newsit.gr/oikonomia/Apisteyto-Para
---- Link came up: http://www.newsit.gr/topikes-eidhseis/Den-toy
END. 79 pages were saved. Double links: 0
Driver Loaded

```

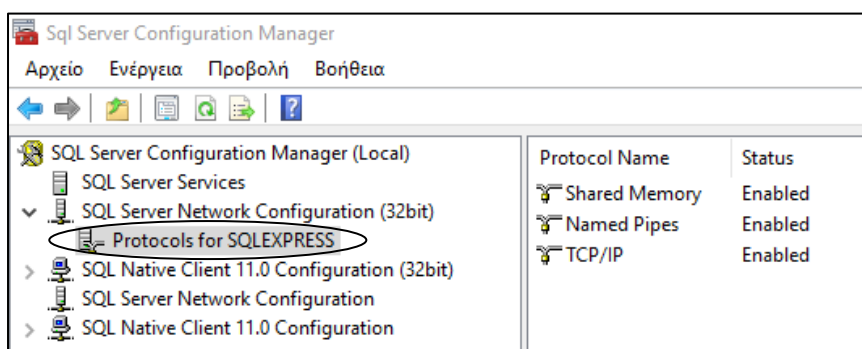
9. Κάτω από την ένδειξη END εμφανίζεται το μήνυμα πως ο driver τρέχει κανονικά. Αν ο driver δεν τρέχει κανονικά το πρόγραμμα τερματίζει και πρέπει να ακολουθήσετε τις εξής οδηγίες:
 - a. Ανοίγετε το πρόγραμμα Sql Server Configuration Manager όπως φαίνεται στην οθόνη



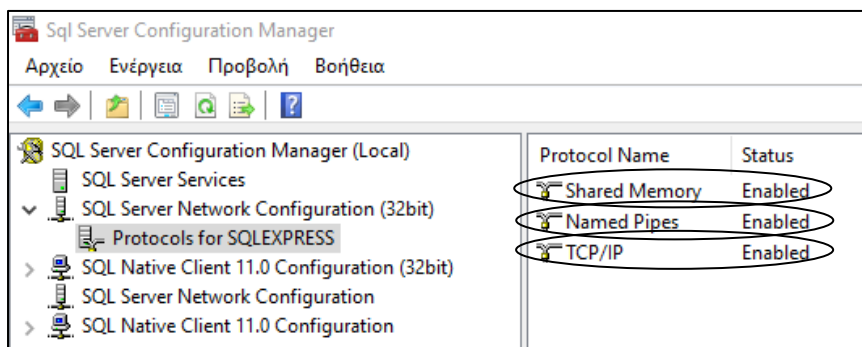
b. Βεβαιωθείτε ότι Sql Server (SQLEXPRESS) έχει την ένδειξη “Running”



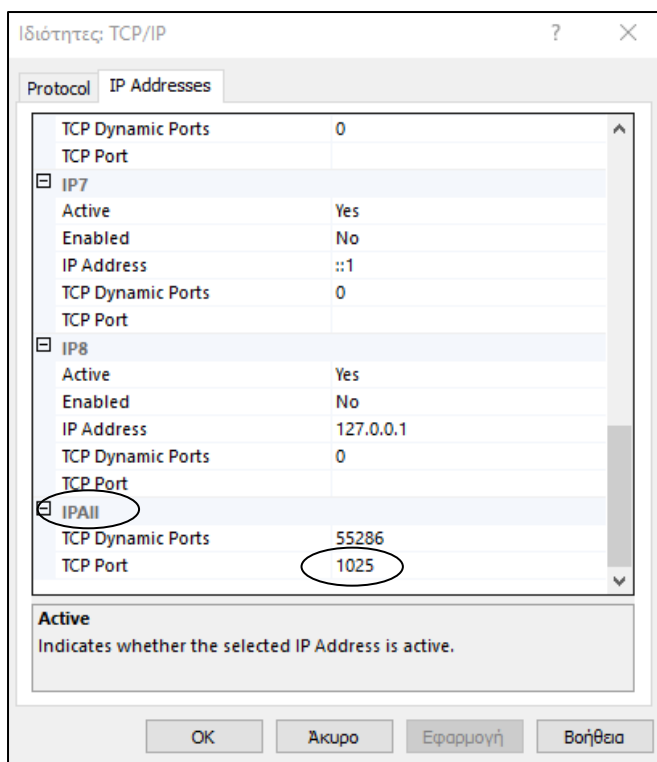
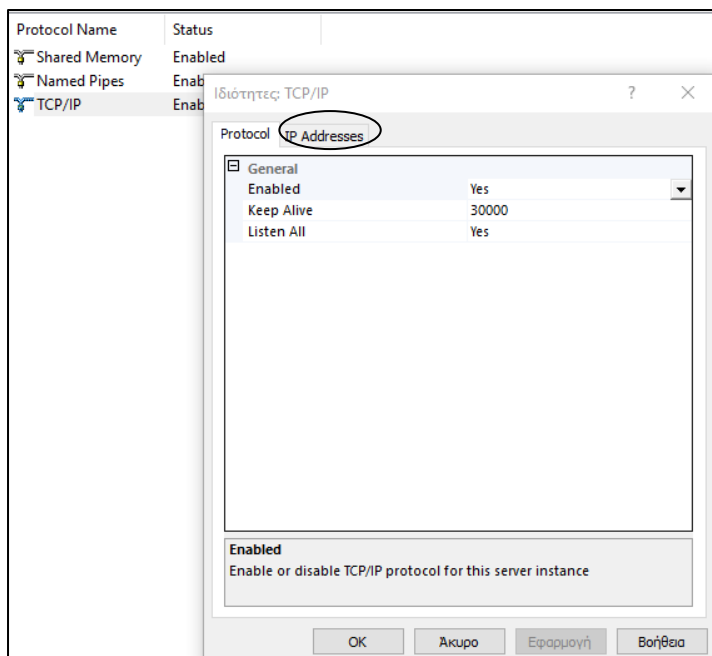
c. Αριστερά μεταβαίνουμε στη σελίδα SQL Server Network Configuration (32 bit)-> Protocols for SQLEXPRESS



d. Βεβαιωνόμαστε ότι και τα τρία πρωτόκολλα έχουν την ένδειξη «Enabled».



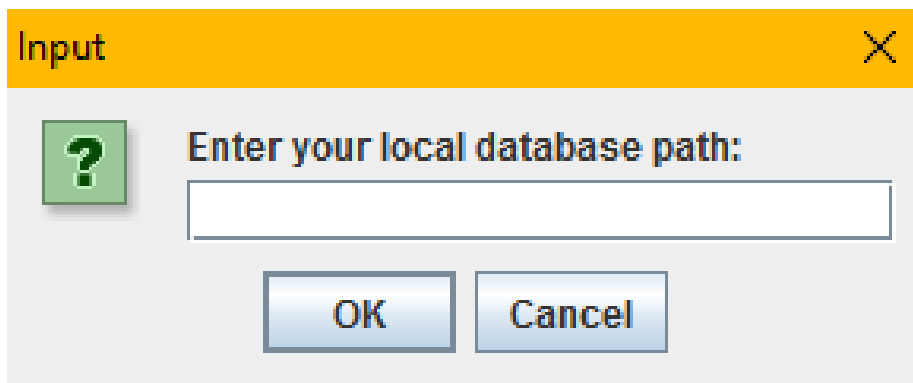
e. Δεξί κλικ στο TCP/IP -> Ιδιότητες -> καρτέλα IP Addresses -> IP All και τοποθετούμε στο TCP Port τον αριθμό 1025. Εφαρμογή -> OK



- f. Τερματίζουμε όλα τα προγράμματα, κάνουμε επανεκκίνηση και ξεκινάμε από το βήμα 10. Αν ο driver τρέχει κανονικά ("Driver Loaded") εμφανίζεται το μήνυμα «Please insert your local database path:»
Εισάγουμε το εξής:

jdbc:sqlserver://**DESKTOP-5MG7QUJ**:1025;databaseName=**MyDatabase**

κίτρινη περιοχή: βάζουμε το όνομα του server πάνω στον οποίο τρέχει η βάση δεδομένων μας
γκρι περιοχή: βάζουμε τον αριθμός της θύρας που βάλαμε στο βήμα 9e
πράσινη περιοχή: βάζουμε το όνομα της βάσης

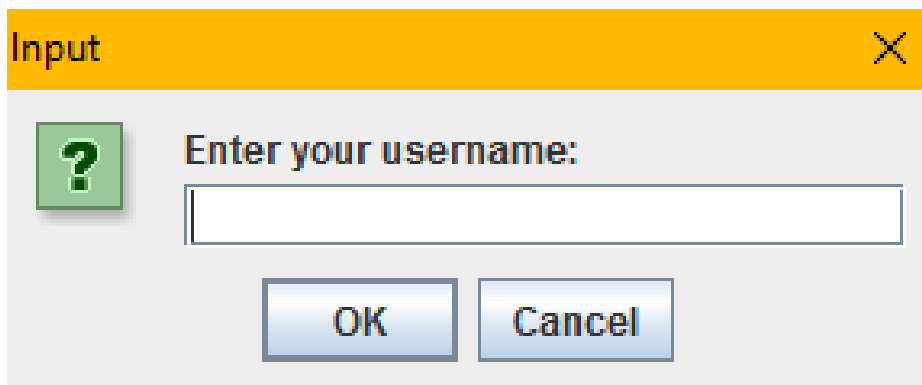


Input

Enter your local database path:

OK Cancel

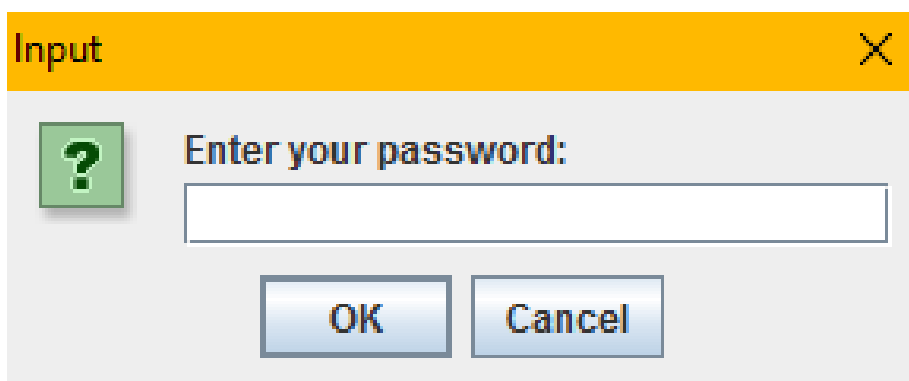
11. Στις επόμενες δύο ερωτήσεις βάζουμε το username και το password με το οποίο συνδεόμαστε στη βάση



Input

Enter your username:

OK Cancel

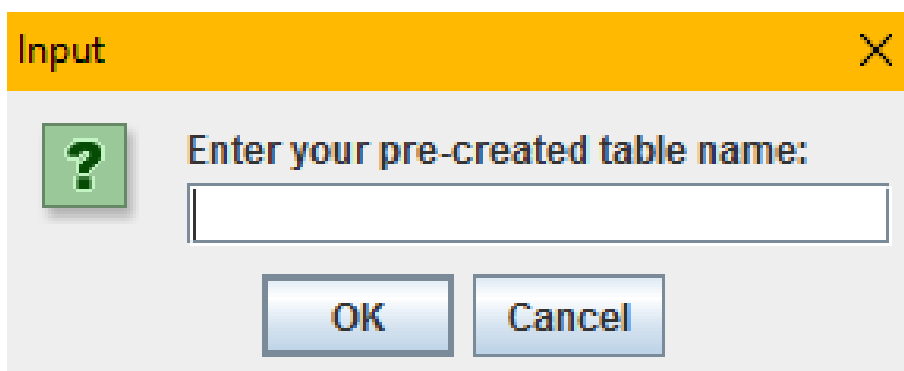


Input

Enter your password:

OK Cancel

12. Στην τελευταία ερώτηση βάζουμε το όνομα του πίνακα που θέλουμε να αποθηκευτούν τα δεδομένα



Input

Enter your pre-created table name:

OK Cancel

13. Αν όλα έχουν πάει καλά, το τελευταίο μήνυμα που εμφανίζεται είναι: "Links added to DataBase!"

```
END. 59 pages were saved. Double links: 0
Driver Loaded
Links added to DataBase!
C:\Users\Vaggelis\Desktop>
```

Βάση Δεδομένων

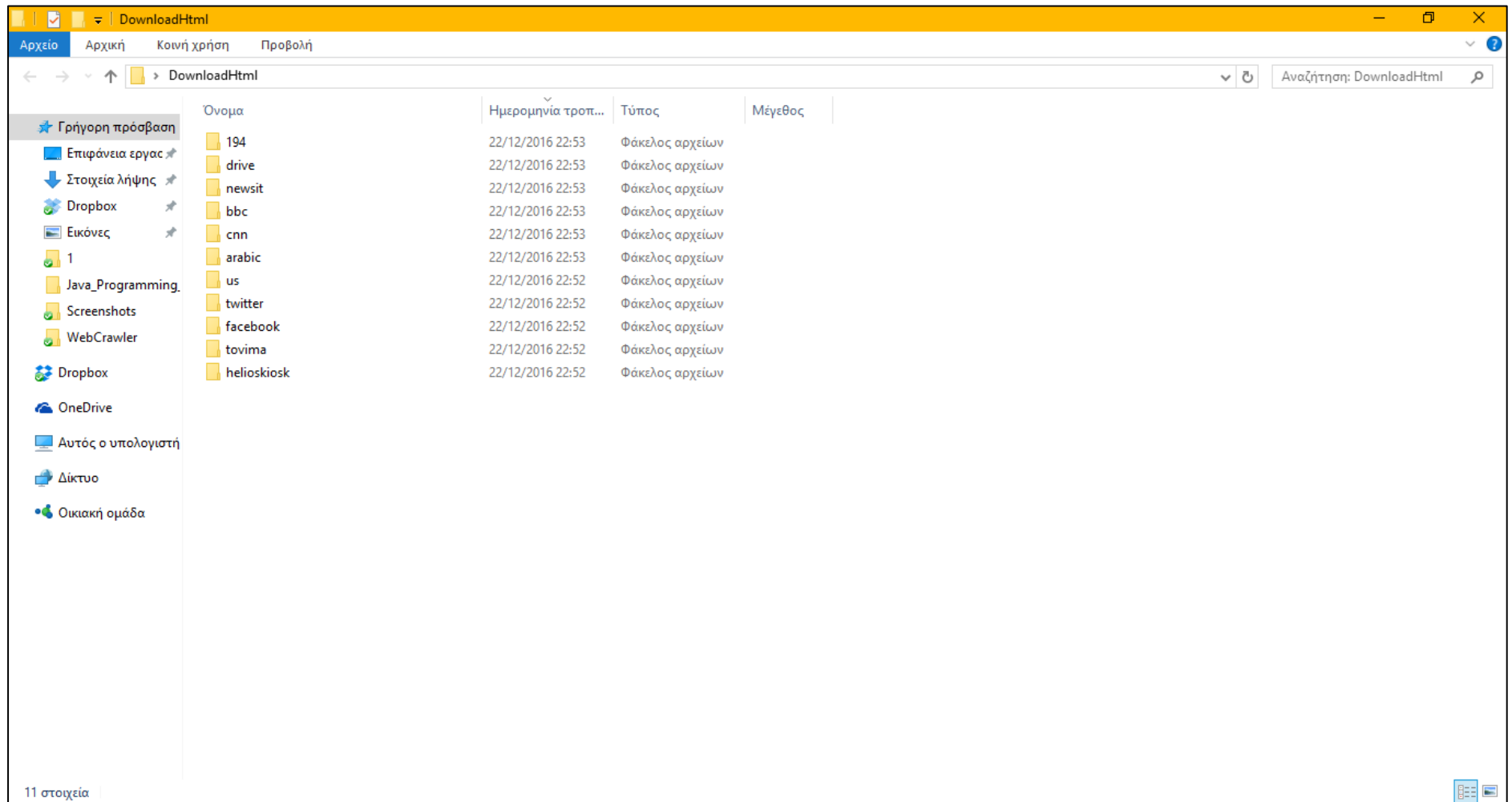
The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The left pane displays the 'Object Explorer' with the following structure:

- DESKTOP-5MG7QUA\SQLEXPRESS (+)
 - Databases
 - System Databases
 - MyDatabase
 - Database Diagrams
 - Tables
 - System Tables
 - FileTables
 - dbo.Test
 - dbo.Απλός_Πελάτης
 - dbo.Αυτοκίνητο
 - dbo.Γεωγραφικό_Διαμέρισμα
 - dbo.Ενοικίαση
 - dbo.Επιπλέον_Οδηγίες
 - dbo.Εταιρικός_Πελάτης
 - dbo.Κατηγορία
 - dbo.Πελάτης
 - dbo.Πληρωμή
 - dbo.Τηλέφωνο_Τοπίου
 - dbo.Τοποθεσία
 - Views
 - Synonyms
 - Programmability
 - Service Broker
 - Storage
 - Security
 - Security
 - Server Objects
 - Replication
 - Management

The right pane shows the 'DESKTOP-5MG7QUA\...tabase - dbo.Test' table. The table has two columns: 'link' and 'where'. The data is as follows:

link	where
http://www.newsit.gr/media/Popi-Xatzidimitrioy-I-Eleni-M...	C:\Users\Vaggelis\Desktop\DownloadHtml\newsit\6jcmqifymj.txt
http://www.cnn.gr/travel/kosmos	C:\Users\Vaggelis\Desktop\DownloadHtml\cnn\3mjjeemefh.txt
http://www.newsit.gr/tags/Ερευνα-Ερευνα-Ερευνα	C:\Users\Vaggelis\Desktop\DownloadHtml\newsit\3guxwngloj.txt
http://www.newsit.gr/topikes-eidhseis/Den-toys-ptoei-i-p...	C:\Users\Vaggelis\Desktop\DownloadHtml\newsit\3ydaqpixlwq.txt
http://www.cnn.gr/style/politismos	C:\Users\Vaggelis\Desktop\DownloadHtml\cnn\3ojvixerzn.txt
http://www.newsit.gr/rss/artss.php	C:\Users\Vaggelis\Desktop\DownloadHtml\newsit\5boinqossf.txt
http://www.tovima.gr/feed/editors/?edid=954	C:\Users\Vaggelis\Desktop\DownloadHtml\tovima\2izynadfxc.txt
http://www.newsit.gr/finder.html	C:\Users\Vaggelis\Desktop\DownloadHtml\newsit\4aaksiiczp.txt
http://www.tovima.gr/feed/editors/?edid=717	C:\Users\Vaggelis\Desktop\DownloadHtml\tovima\8scvtskvdu.txt
http://www.tovima.gr/feed/editors/?edid=959	C:\Users\Vaggelis\Desktop\DownloadHtml\tovima\4usofdbntl.txt
http://arabic.cnn.com/	C:\Users\Vaggelis\Desktop\DownloadHtml\arabic\1ghtkvadvst.txt
http://www.newsit.gr/kosmos/Syriakos-stratos-Elegxoyme...	C:\Users\Vaggelis\Desktop\DownloadHtml\newsit\3fqjveqrto.txt
http://www.cnn.gr/terms	C:\Users\Vaggelis\Desktop\DownloadHtml\cnn\6wnpajmrgf.txt
http://www.bbc.com/earth/columns/attenboroughs-story...	C:\Users\Vaggelis\Desktop\DownloadHtml\bbc\6txiaevrml.txt
http://www.newsit.gr/ygeia/Xronia-igmoritida-Aitia-sympt...	C:\Users\Vaggelis\Desktop\DownloadHtml\newsit\8bbnadoeyv.txt
http://www.tovima.gr/world/article/?aid=853059#comme...	C:\Users\Vaggelis\Desktop\DownloadHtml\tovima\5kftfzpmen.txt
http://www.bbc.com/earth/story/20161219-the-fire-salam...	C:\Users\Vaggelis\Desktop\DownloadHtml\bbc\1ozothuhpz.txt
https://twitter.com/CNNgreece	C:\Users\Vaggelis\Desktop\DownloadHtml\twitter\5aunupqvvv.txt
http://www.bbc.com/arts/	C:\Users\Vaggelis\Desktop\DownloadHtml\bbc\0vlukjovug.txt
http://www.cnn.gr/videos/viral	C:\Users\Vaggelis\Desktop\DownloadHtml\cnn\3aizbauzy.txt
http://www.bbc.com/persian/	C:\Users\Vaggelis\Desktop\DownloadHtml\bbc\0dledrclcl.txt
http://www.bbc.com/iwonder	C:\Users\Vaggelis\Desktop\DownloadHtml\bbc\2myeopktrv.txt
http://www.tovima.gr/editors/editor/?edid=4373	C:\Users\Vaggelis\Desktop\DownloadHtml\tovima\0zpllkjos.txt
http://www.cnn.gr/efimerides	C:\Users\Vaggelis\Desktop\DownloadHtml\cnn\4jchwyjkt.txt

Φάκελος



DownloadHtml

Αρχείο Αρχική Κοινή χρήση Προβολή

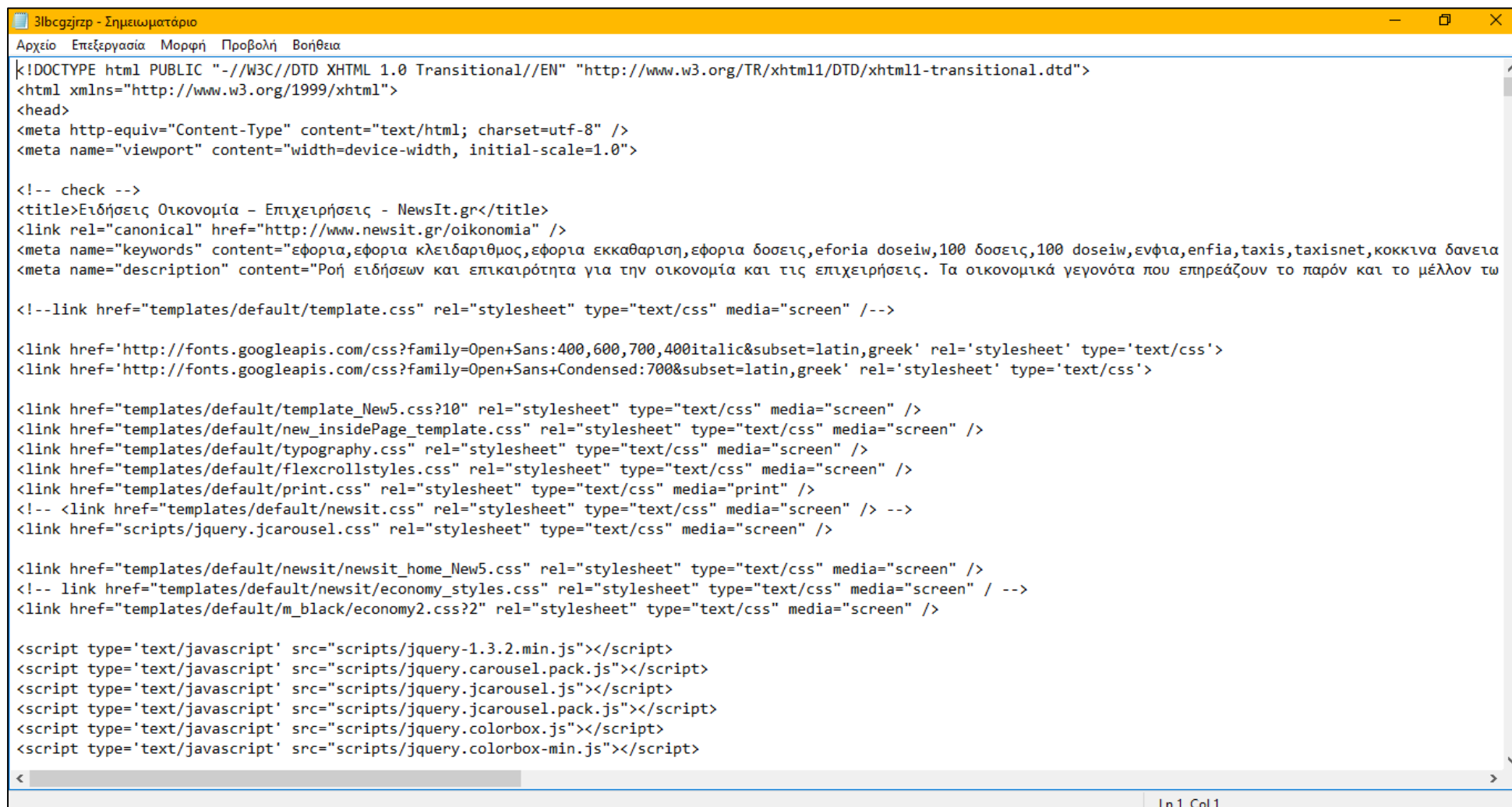
DownloadHtml

Αναζήτηση: DownloadHtml

Όνομα	Ημερομηνία τροπ...	Τύπος	Μέγεθος
194	22/12/2016 22:53	Φάκελος αρχείων	
drive	22/12/2016 22:53	Φάκελος αρχείων	
newsit	22/12/2016 22:53	Φάκελος αρχείων	
bbc	22/12/2016 22:53	Φάκελος αρχείων	
cnn	22/12/2016 22:53	Φάκελος αρχείων	
arabic	22/12/2016 22:53	Φάκελος αρχείων	
us	22/12/2016 22:52	Φάκελος αρχείων	
twitter	22/12/2016 22:52	Φάκελος αρχείων	
facebook	22/12/2016 22:52	Φάκελος αρχείων	
tovima	22/12/2016 22:52	Φάκελος αρχείων	
helioskiosk	22/12/2016 22:52	Φάκελος αρχείων	

11 στοιχεία

Αρχείο



```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<meta name="viewport" content="width=device-width, initial-scale=1.0">

<!-- check -->
<title>Ειδήσεις Οικονομία - Επιχειρήσεις - NewsIt.gr</title>
<link rel="canonical" href="http://www.newsit.gr/oikonomia" />
<meta name="keywords" content="εφορία,εφορία κλειδαριθμος,εφορία εκκαθάριση,εφορία δόσεις,eforia doseiw,100 δόσεις,100 doseiw,ενφια,enfia,taxis,taxisnet,κοκκινα δανεια
<meta name="description" content="Ποή ειδήσεων και επικαιρότητα για την οικονομία και τις επιχειρήσεις. Τα οικονομικά γεγονότα που επηρεάζουν το παρόν και το μέλλον τω

<!--link href="templates/default/template.css" rel="stylesheet" type="text/css" media="screen" /-->

<link href='http://fonts.googleapis.com/css?family=Open+Sans:400,600,700,400italic&subset=latin,greek' rel='stylesheet' type='text/css'>
<link href='http://fonts.googleapis.com/css?family=Open+Sans+Condensed:700&subset=latin,greek' rel='stylesheet' type='text/css'>

<link href="templates/default/template_New5.css?10" rel="stylesheet" type="text/css" media="screen" />
<link href="templates/default/new_insidePage_template.css" rel="stylesheet" type="text/css" media="screen" />
<link href="templates/default/typography.css" rel="stylesheet" type="text/css" media="screen" />
<link href="templates/default/flexcrollstyles.css" rel="stylesheet" type="text/css" media="screen" />
<link href="templates/default/print.css" rel="stylesheet" type="text/css" media="print" />
<!-- <link href="templates/default/newsit.css" rel="stylesheet" type="text/css" media="screen" /> -->
<link href="scripts/jquery.jcarousel.css" rel="stylesheet" type="text/css" media="screen" />

<link href="templates/default/newsit/newsit_home_New5.css" rel="stylesheet" type="text/css" media="screen" />
<!-- link href="templates/default/newsit/economy_styles.css" rel="stylesheet" type="text/css" media="screen" / -->
<link href="templates/default/m_black/economy2.css?2" rel="stylesheet" type="text/css" media="screen" />

<script type='text/javascript' src="scripts/jquery-1.3.2.min.js"></script>
<script type='text/javascript' src="scripts/jquery.carousel.pack.js"></script>
<script type='text/javascript' src="scripts/jquery.jcarousel.js"></script>
<script type='text/javascript' src="scripts/jquery.jcarousel.pack.js"></script>
<script type='text/javascript' src="scripts/jquery.colorbox.js"></script>
<script type='text/javascript' src="scripts/jquery.colorbox-min.js"></script>

<
```

Ln 1, Col 1