

automated_post ver. 0.1

**Μια εφαρμογή γραμμένη σε python 3.5 και php που επιτρέπει την
αυτόματη δημοσίευση αρχείων στο Πανελλήνιο Σχολικό Δίκτυο**

**Κωνσταντίνος Χερτούρας, Διπλωματούχος ΗΜΜΥ Πολυτεχνείου Κρήτης , Εκπαιδευτικός
ΠΕ19 , ΕΠΑΛ Ροδόπολης Σερρών
chertour@sch.gr**

**Ο κώδικας έχει άδεια GPL 3.0 και είναι πλήρως ελεύθερος
Δεν δίνεται καμία εγγύηση και δεν φέρεται καμία ευθύνη από την μεριά μου για ότι
προκύψει από τη χρήση της εφαρμογής.**

Κεφάλαιο 1: Γενικές Πληροφορίες

Η συγκεκριμένη εφαρμογή δημιουργήθηκε ώστε να γίνει απλούστερος ο τρόπος δημοσίευσης αρχείων που αφορούν την σχολική ζωή, στο Πανελλήνιο σχολικό δίκτυο (ΠΣΔ). Με τη χρήση της ο κάθε ενδιαφερόμενος μπορεί απλά αποθηκεύοντας ένα αρχείο σε ένα συγκεκριμένο φάκελο του υπολογιστή του, να επιτύχει την αυτοματοποιημένη αποστολή του έναν συγκεκριμένο φάκελο στην περιοχή του στο ΠΣΔ και την ταυτόχρονη προβολή στο WWW του αρχείου που μεταφορτώθηκε, μέσω μιας διασυνδεδεμένης ιστοσελίδας.

Θα μπορούσαμε να αναπαραστήσουμε σε βήματα τον τρόπο λειτουργίας της εφαρμογής, ως εξής

1. Θέλω να βγάλω μια ανακοίνωση ή να δημοσιεύσω μια φωτογραφία.
2. Αφού δημιουργήσω το αρχείο, το μεταφέρω σε έναν φάκελο που έχω ορίσει εξ αρχής στις παραμέτρους του προγράμματος.
3. Ο φάκελος ελέγχεται από ένα script python σε τακτά χρονικά διαστήματα.
4. Το script εντοπίζει τα αρχεία που υπάρχουν και επικοινωνεί με της ρουτίνες του ΠΣΔ, ώστε να διαπιστώσει πια αρχεία είναι ήδη δημοσιευμένα και ποια όχι. Επίσης εντοπίζει ποια αρχεία έχουν διαγραφεί από το χρήστη του προγράμματος (από τον φάκελο του τοπικού δίσκου) ώστε αυτά με τη σειρά τους να διαγραφούν από το ΠΣΔ, και να σταματήσει η δημοσίευσή τους.
5. Τα αρχεία ανεβαίνουν με την χρήση του HTTP POST ή όσα πρέπει διαγράφονται με την κλήση της αντίστοιχης ρουτίνας. Όλες οι ιστοσελίδες είναι password protected αλλά το σύνολο της ανταλλαγής των κωδικών γίνεται με τη μορφή plain text (χωρίς encryption) , για λόγους απλότητας.
6. Κάθε ένας που θέλει να δει τα διαθέσιμα αρχεία , μπορεί να επισκέπτεται την συγκεκριμένη ιστοσελίδα προβολής, όπου δημιουργείται δυναμικά και περιλαμβάνει το σύνδεσμο προβολής του αρχείου, το όνομα του αρχείου και την ημερομηνία δημιουργίας του.

Κεφάλαιο 2: Λεπτομέρειες εγκατάστασης και χρήσης (client)

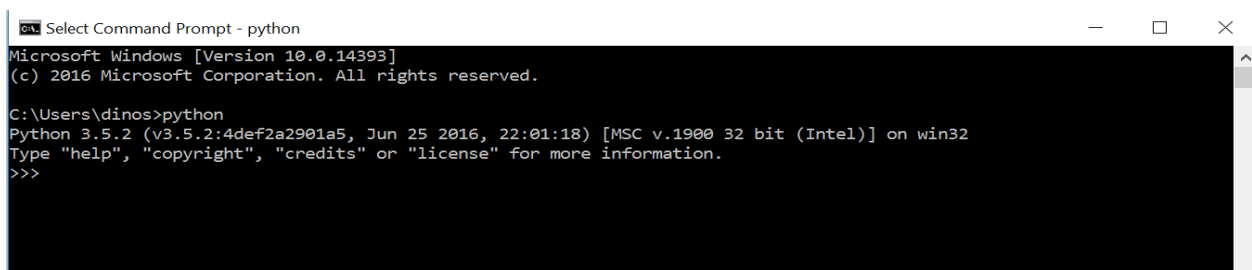
Η εφαρμογή περιλαμβάνει δυο μέρη:

- α) Το μέρος που εκτελείτε στον προσωπικό υπολογιστή και
- β) Το μέρος που εκτελείτε στο ΠΣΔ.

α. Στον προσωπικό υπολογιστή εκτελείται ένα script γραμμένο σε Python 3.5

Για να εκτελεστεί θα πρέπει :

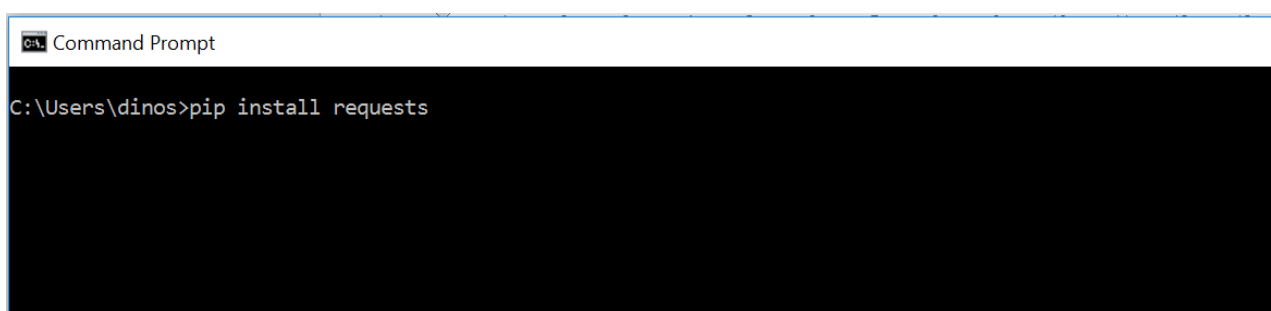
1. Να είναι εγκατεστημένη η έκδοση της python 3.4 και μεταγενέστερη. Πιθανότατα μπορεί να τρέξει και σε παλαιότερες εκδόσεις της έκδοσης 3.4 (πιθανά όχι όμως και της 2.x) αλλά δεν έχουν γίνει δοκιμές για να πιστοποιηθεί κάτι τέτοιο.
2. Να υπάρχει η python στο path των windows ώστε π.χ. από το cmd να μπορεί να κληθεί ο python interpreter από οποιοδήποτε φάκελο :



```
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. All rights reserved.

C:\Users\dinos>python
Python 3.5.2 (v3.5.2:4def2a2901a5, Jun 25 2016, 22:01:18) [MSC v.1900 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
```

3. Να έχει εγκατασταθεί το **module requests** με την εντολή από το cmd: **pip install requests** (το πρόγραμμα pip είναι ήδη εγκατεστημένο στις σύγχρονες διανομές python). Σε αντίθετη περίπτωση θα πρέπει να εγκατασταθεί το πρόγραμμα pip (λεπτομέρειες στο <https://pip.pypa.io/en/stable/installing/>)



```
C:\Users\dinos>pip install requests
```

Στη συνέχεια από τον φάκελο στον οποίο βρίσκεται το script (**στον ίδιο φάκελο μαζί με το script πρέπει να υπάρχει και το συνοδευτικό αρχείο παραμέτρων parameters_php.py - που αναφέρεται παρακάτω**), μπορείτε να εκτελέσετε το σενάριο ως εξής:

python automated_post.py

Ταυτόχρονα, σε αυτό τον φάκελο όπου βρίσκονται τα δύο αρχεία της εφαρμογής, δημιουργείται και ένα αρχείο κειμένου με το όνομα text.log ώστε να βοηθηθεί και το debugging της εφαρμογής (Ο κώδικας είναι εύκολα εντοπίσιμος στο script και μπορεί να αφαιρεθεί).

Όπως ήδη αναφέρθηκε, μαζί με το κύριο script της εφαρμογής, υπάρχει και ένα βοηθητικό script με το όνομα **parameters_php.py** το οποίο περιλαμβάνει τον εξής κώδικα (σημειώνονται μόνον οι κρίσιμες παράμετροι):

```
-----  
path = "documents_for_www"  
drive = "c:\\"
```

Το όνομα του φακέλου στον οποίο βρίσκονται τα αρχεία που θέλουμε να ανεβάσουμε στο Internet. Στην πιο απλή περίπτωση είναι ο φάκελος c:\ documents_for_www όπου αν δεν είναι δημιουργημένος, δημιουργείται από το πρόγραμμα κατά την πρώτη εκτέλεση, και προφανώς είναι άδειος (Παρατήρηση: εάν για κάποιο λόγο εμφανιστεί το πρόβλημα της αδυναμίας δημιουργίας του φακέλου από το πρόγραμμα, προτείνεται η δημιουργία του από το χρήστη). Προφανώς απαιτούνται τροποποιήσεις στις δύο παραμέτρους στην περίπτωση χρήσης Linux).

```
#username and password needed to delete and update files
username="xxxxxxx"
password="xxxxxxx"
```

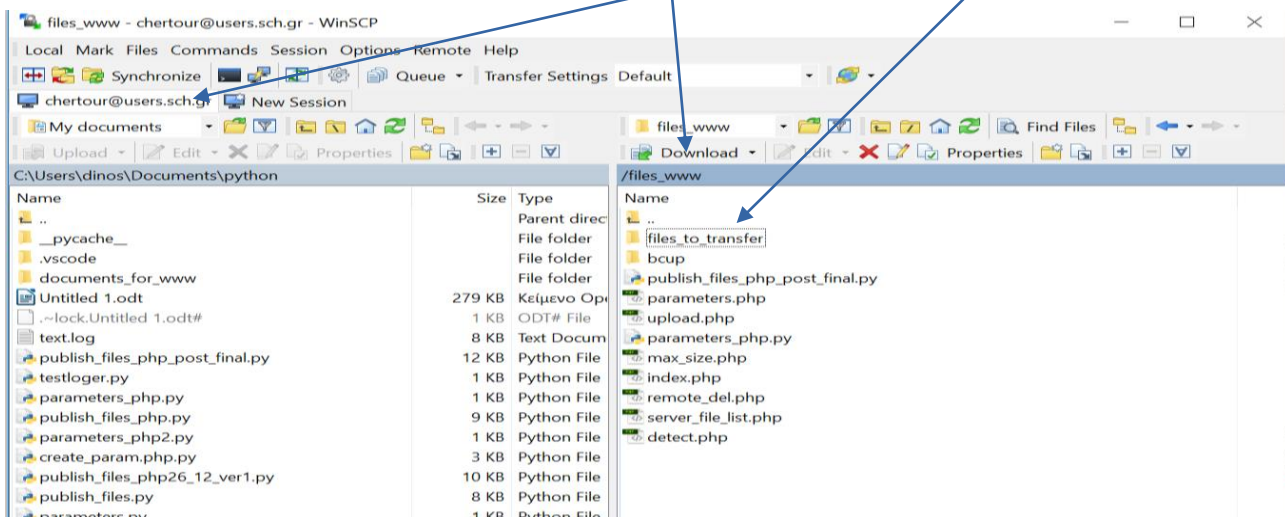
Τα στοιχεία με βάση τα οποία επιτρέπεται ή όχι η πρόσβαση στην δυνατότητα διαγραφής ή δημοσίευσης αρχείων στο ΠΣΔ. Στο αρχείο parameters.php που υπάρχει στο ΠΣΔ (φάκελος server στο αρχείο automated_post.rar της εφαρμογής) πρέπει να υπάρχουν οι ίδιες τιμές για τους κωδικούς. Δεν είναι αναγκαίο να ταυτίζονται με τους κωδικούς πρόσβασης του σχολείου στο Internet.

ΣΗΜΑΝΤΙΚΟ : Προτείνεται η χρήση διαφορετικών κωδικών από τους κωδικούς του σχολείου στο ΠΣΔ για λόγους ασφαλείας.

```
#where to upload the files
url_to_receive = 'http://users.sch.gr/username/files_www/upload.php'
```

Το url στο οποίο βρίσκονται τα αρχεία php του server. Όπως θα αναφερθεί και στη συνέχεια, θα πρέπει στο ΠΣΔ να δημιουργηθεί ένας φάκελος στην κεντρική περιοχή του χρήστη με όνομα **files_www** όπου θα τοποθετηθούν τα php αρχεία που αποτελούν το server side κομμάτι της εφαρμογής. Τα αρχεία που θα μεταφορτώνονται από το σενάριο python για δημοσίευση, θα τοποθετούνται στον φάκελο **files_to_transfer** (βρίσκεται μέσα στο φάκελο **files_www**).

Για παράδειγμα, στη δική μου περίπτωση ο χρήστης είναι ο **chertour** και δημιουργήθηκαν οι δύο φάκελοι όπως δείχνει η παρακάτω εικόνα:



```
filenames_excluded = ['httpserver.log', 'server_files.json']
```

Πιθανά ονόματα αρχείων που κάποιος χρήστης δεν θέλει να περιλαμβάνονται σε αυτά που γίνονται upload, παρά το γεγονός ότι αυτά βρίσκονται στο φάκελο documents_for_www

```
#server side script residing url  
server_name = 'users.sch.gr/xxxxx/' ή για σχολική μονάδα 'xxxxx.thess.sch.gr'
```

Το URL του χρήστη στο πανελλήνιο σχολικό δίκτυο.

```
seconds_to_pause = 15
```

Ο χρόνος που ορίζει κάθε πόσα ΔΕΥΤΕΡΟΛΕΠΤΑ θα ελέγχεται ο φάκελος για καινούργια αρχεία.

Ειδική μνεία για τα ονόματα αρχείων με ελληνικούς χαρακτήρες

Σημαντικά προβλήματα εμφανίστηκαν σε πολλά στάδια της ανάπτυξης της εφαρμογής, που αφορούσαν την σωστή αναπαράσταση των ελληνικών χαρακτήρων στα ονόματα αρχείων. Τα αρχεία μεταφέρονται και αποθηκεύονται με κωδικοποίηση χαρακτήρων ISO - 8859-7, καθώς αποδείχτηκε στην πράξη και για την περίπτωση αυτή, η καταλληλότερη στο να διαβάσει και να προβάλει σωστά τα ονόματα αρχεία που δημιουργήθηκαν στα ελληνικά windows έκδοσης 10.

Η έκδοση Winscp 5.9.3 επιτρέπει και την σωστή αναπαράσταση των ονομάτων αρχείων και μέσω ftp, αν κάποιος επιθυμεί για λόγους test, να βλέπει να αρχεία που ανέβηκαν στο φάκελο username/files_www/files_to_tranfer του ΠΣΔ.

Κεφάλαιο 3: ΣΗΜΑΝΤΙΚΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ:

1. Όπως ήδη αναφέρθηκε υπάρχει ένα ζεύγος τιμών username / password που αποφασίζουν αν θα επιτρέψουν την δημιουργία ή την διαγραφή αρχείων στο ΠΣΔ. Λόγω του ότι δεν γίνεται κρυπτογράφηση, **προτείνεται η χρήση διαφορετικών κωδικών από τους κωδικούς του σχολείου.**

2. Για τον ευκολότερο χειρισμό των ονομάτων αρχείων, κάθε αρχείο που αποθηκεύεται στο φάκελο που υπάρχουν τα αρχεία προς δημοσίευση, μετονομάζεται αυτόματα με όνομα το ίδιο που το αρχείο είχε πριν αλλά με αφαίρεση των κενών (αν υπάρχουν) ανάμεσα στις λέξεις και αντικατάστασή τους με την κάτω παύλα (_). Για παράδειγμα το αρχείο:

Ανακοίνωση για εκδρομή Α-Β-Γ δημοτικού.doc θα μετατραπεί αυτόματα σε
Ανακοίνωση_για_εκδρομή_Α-Β-Γ_δημοτικού.doc

3. Η εφαρμογή ανταποκρίνεται καλά στην αποστολή αρχείων τα οποία είναι ήδη ανοικτά. Για την αποφυγή προβλημάτων όμως που μπορεί να προκύψουν από την περίπτωση του να υπάρχει ένα αρχείο ανοιχτό την στιγμή που η εφαρμογή προσπαθεί να το μετονομάσει, προτείνεται στην αρχή και μέχρι να εξοικειωθεί ο χρήστης με τον τρόπο λειτουργίας της, να δημιουργεί / τροποποιεί τα αρχεία έξω από το τον φάκελο των αρχείων προς δημοσίευση. Στην συνέχεια μπορεί να τα μεταφέρει στο εν λόγω φάκελο, ώστε με την σειρά τους να δημοσιευτούν. **ΠΡΟΣΟΧΗ: ΑΝ ΚΑΤΑ ΤΗΝ ΠΡΟΣΠΑΘΕΙΑ ΜΕΤΟΝΟΜΑΣΙΑΣ ΕΝΟΣ ΑΡΧΕΙΟΥ ΒΡΕΘΕΙ ΙΔΙΟ ΑΡΧΕΙΟ ΣΤΟΝ ΦΑΚΕΛΛΟ, ΤΟ ΠΡΟΓΡΑΜΜΑ ΘΑ ΚΑΝΕΙ EXIT ΠΕΡΙΜΕΝΟΝΤΑΣ ΑΠΟ ΤΟΝ ΧΡΗΣΤΗ ΝΑ ΔΙΑΛΕΞΕΙ ΠΟΙΟ ΑΠΟ ΤΑ ΔΥΟ ΑΡΧΕΙΑ ΜΕ ΤΟ ΙΔΙΟ ΟΝΟΜΑ ΕΠΙΘΥΜΕΙ. ΠΡΟΤΕΙΝΕΤΑΙ Η ΔΙΑΓΡΑΦΗ ΤΩΝ ΠΑΛΑΙΩΝ ΑΡΧΕΙΩΝ ΟΤΑΝ ΘΕΛΟΥΜΕ ΝΑ ΔΗΜΟΣΙΕΥΤΟΥΝ ΝΕΩΤΕΡΕΣ ΕΚΔΟΣΕΙΣ ΤΟΥΣ ΜΕ ΤΟ ΙΔΙΟ ΟΝΟΜΑ.**

4. Όπως αναφέρθηκε, δημιουργείτε και ένα log (στο φάκελο που η εφαρμογή είναι εγκατεστημένη) ώστε να μπορέσουν να βρεθούν ευκολότερα τα σφάλματα που θα προκύψουν.

5. Η εφαρμογή δεν επιτρέπει την αποστολή αρχείων μεγαλύτερων από το μέγεθος που ορίζει ο χρήστης μέσω μια μεταβλητής `php` στον `server`. Το ίδιο το ΠΣΔ δεν επιτρέπει την άνοδο αρχείων μεγαλύτερων από 100MB, όπως προκύπτει από το `phpinfo()`;

Κεφάλαιο 4: Λεπτομέρειες εγκατάστασης (server)

Για την εγκατάσταση και λειτουργία της εφαρμογής στο ΠΣΔ απαιτούνται οι 5 συνολικά σελίδες `php` που αποτελούν το `server side` μέρος της εφαρμογής καθώς και δύο φάκελοι. Πιο συγκεκριμένα για την εγκατάσταση της εφαρμογής απαιτείται :

A. Η δημιουργία ενός φακέλου στην κεντρική περιοχή με το όνομα **files_www**

B. Η δημιουργία ενός ακόμα φακέλου (μέσα στο **files_www**) με το όνομα **files_to_transfer**

Στον φάκελο `files_www` θα τοποθετηθούν οι 5 σελίδες `php`, των οποίων η λειτουργικότητα αναλύεται παρακάτω, ενώ στον φάκελο `files_to_transfer` θα μεταφορτώνονται τα αρχεία από το πρόγραμμα της `python`.

Τα αρχεία `php` είναι τα εξής:

1. **Upload.php**: Το αρχείο `upload.php` υλοποιεί την μεταφόρτωση των αρχείων μέσω του HTTP Post, από το φάκελο του προσωπικού υπολογιστή στο ΠΣΔ. Ο τρόπος λειτουργίας του συνοψίζεται στα εξής: Αρχικά γίνεται ο έλεγχος του `username / password` του οποίου το αποτέλεσμα θα κρίνει σε πρώτο επίπεδο, το αν θα επιτραπεί η μεταφόρτωση αρχείων. Στην συνέχεια και εφόσον οι κωδικοί είναι σωστοί γίνεται έλεγχος για το αν το μέγεθος του αρχείου είναι μικρότερο από το όριο που έχει τεθεί από τον χρήστη στο αρχείο `parameters.php` (θα αναφερθούμε σε αυτό παρακάτω). Αν ναι τότε εξετάζεται το αν η κατάληξη του αρχείου ανήκει στις επιτρεπτές προς μεταφόρτωση καταλήξεις καθώς επίσης εξετάζεται το αν υπάρχει διπλή κατάληξη ή δεν υπάρχει καθόλου κατάληξη στο αρχείο. Εάν όλοι οι έλεγχοι είναι επιτυχείς τότε επιτρέπεται η μεταφόρτωση του αρχείου στον `server`. Να σημειωθεί ότι η ημερομηνία του κάθε αρχείου στέλνεται σαν παράμετρος και «τοποθετείται » στο αρχείο εκ των υστέρων μια και κατά την μεταφόρτωση ενός αρχείου στον `server` , ως ημερομηνία δημιουργίας του θεωρείτε η ημερομηνία μεταφόρτωσης.

2. **remote_del.php:** Το αρχείο **remote_del.php** υλοποιεί τη διαγραφή των αρχείων που έχουν διαγραφεί από τον φάκελο του τοπικού υπολογιστή στο server. Από το python script καλείτε η ιστοσελίδα και δίνεται ως παράμετρος (πέρα από τα Username / password τα οποία χάριν συντομίας να υποθέσουμε ότι είναι σωστά) το md5 sum του αρχείου (αρχείο + όνομα αρχείου) που θέλουμε να διαγραφεί. Η ιστοσελίδα αφού ανοίξει ένα json αρχείο αποθηκευμένο που περιέχει τις αντιστοιχίες hashes – ονόματα αρχείων (αναφέρεται παρακάτω), εκτελεί τη διαγραφή του αρχείου με βάση το όνομά του.
3. **server_file_list.php:** Το αρχείο **server_file_list.php** αναλαμβάνει: **α)** να δημιουργήσει την λίστα των αρχείων που υπάρχουν στο server **β)** να βρει το άθροισμα md5 για το καθένα από αυτά **γ)** να αποστείλει την λίστα στο python script που την ζήτησε και **δ)** να τα αποθηκεύσει σε ένα αρχείο με την μορφή json ώστε στη συνέχεια να μπορέσει το αρχείο αυτό να διαβαστεί από την σελίδα **remote_del.php** Τα περιεχόμενα που επιστρέφει η συγκεκριμένη ιστοσελίδα στον client είναι αυτά με βάση τα οποία θα ληφθεί η απόφαση για το ποια αρχεία λείπουν από τον server και ποια αρχεία είναι ήδη σε αυτόν.
4. **index.php:** Η ιστοσελίδα προβολής της λίστας των αρχείων που δημοσιεύσαμε στο Internet. Η ιστοσελίδα ακολουθεί την εξής λογική στην υλοποίησή της: α. Εξετάζει ποια αρχεία υπάρχουν στο φάκελο **/files_www/files_to_transfer/** του server και τα τοποθετεί σε μια λίστα (προγραμματιστική δομή). Η λίστα περιέχει διανύσματα που έχουν στοιχεία α) το όνομα του αρχείου, β) το όνομα του αρχείου σε urlencoded μορφή (απαραίτητο για το download λόγω της ύπαρξης ελληνικών χαρακτήρων) και γ) την ημερομηνία τροποποίησης (modification time) του κάθε αρχείου. Στη συνέχεια η λίστα ταξινομείται με βάση τις ημερομηνίες τροποποίησης και έτσι προκύπτει ένας ταξινομημένος κατά ημερομηνία πίνακας διανυσμάτων. Ο πίνακας αυτό τροφοδοτεί μια «δομή» Javascript η οποία τον προβάλλει αλλά και επιτρέπει τόσο την σελιδοποίηση του όσο και την περαιτέρω ταξινόμηση του. Στην πρώτη στήλη τοποθετείται ο σύνδεσμος με την χρήση του οποίου ο κάθε ενδιαφερόμενος μπορεί να επιλέγει το να κατεβάσει κάποιο αρχείο. Στις επόμενες στήλες υπάρχουν τα υπόλοιπα στοιχεία.

5. Parameters.php : Η σελίδα αυτή περιέχει βασικές παραμέτρους που επιτρέπουν την ευκολότερη παραμετροποίηση της εφαρμογής σε σχέση με τις ανάγκες του χρήστη. Υπάρχουν οι εξής παράμετροι (βασικές):

- για το username / password (πρέπει να ταυτίζονται με αυτά στο client/parameters_php.py)
- για το μέγιστο μέγεθος σε bytes που μπορεί να έχει το προς μεταφόρτωση αρχείο
- για τη λίστα με τα ονόματα των αρχείων που δεν θέλουμε να εμφανίζονται στην ιστοσελίδα προβολή (**ΠΡΟΣΟΧΗ – Τα αρχεία εξακολουθούν να είναι προσβάσιμα σε κάποιον που γνωρίζει ακριβώς το όνομα τους με την χρήση του αντίστοιχου URL**)
- για τις καταλήξεις αρχείων που επιτρέπεται να γίνουν uploaded (filetype whitelist)
- για τον φάκελο που φιλοξενεί τα αρχεία

Κεφάλαιο 5: Ζητήματα ασφάλειας.

5.1 Γενικά

Σε κάθε εφαρμογή που επιτρέπεται η μεταφόρτωση αρχείων από τους χρήστες στον server υπάρχουν πάντα οι περιπτώσεις όπου κακόβουλοι χρήστες θα μπορούσαν να κατορθώσουν τη μεταφόρτωση ενός αρχείου στο οποίο θα περιέχεται κώδικας που αν εκτελεστεί θα μπορούσε να έχει άγνωστα αποτελέσματα για τα ήδη υπάρχοντα αρχεία στην περιοχή του χρήστη. Καμία εφαρμογή τέτοιου τύπου και χωρίς βλάβη της χρηστικότητας της δεν μπορεί να είναι απολύτως ασφαλής.

5.2 Μέτρα ασφαλείας της εφαρμογής

Η συγκεκριμένη εφαρμογή λαμβάνει την εξής μέριμνα για την όσο το δυνατό μεγαλύτερη προστασία του χρήστη:

α. Πρόσβαση στις ιστοσελίδες με την χρήση username / password (δεν χρησιμοποιείται encryption για λόγους απλότητας της εφαρμογής)

β. Έλεγχος του μεγέθους του αρχείου και απόρριψη του σε ανάλογη περίπτωση.

γ. Έλεγχος της κατάληξης του αρχείου προς μεταφόρτωση και έλεγχος για την ύπαρξη ή όχι κατάληξης ή και ύπαρξης διπλής κατάληξης. Σκοπός είναι η προστασία του χρήστη από μεταφόρτωση αρχείων που θα μπορούν να εκτελεστούν από τον server και όχι να προταθούν για λήψη.

Το σύνολο της προσπάθειας αντιμετώπισης των όποιων προβλημάτων μπορεί να προκύψουν αφήνεται στον server και όχι στον client.