

*«Ανάπτυξη στοιχείων λογισμικού (plug-in) για την ενσωμάτωσή τους σε υπηρεσίες ηλεκτρονικού ταχυδρομείου τύπου Thunderbird και Outlook για το σύστημα «Arachnoid»»*



**ΒΡΥΝΙΟΣ ΠΑΝΑΓΙΩΤΗΣ**

04/2654

**ΕΠΙΒΛΕΠΩΝ: ΗΛΙΟΥΔΗΣ ΧΡΗΣΤΟΣ**

ΘΕΣΣΑΛΟΝΙΚΗ, 2014



Τμήμα Μηχανικών  
Πληροφορικής ΑΤΕΙΘ

# Εισαγωγικά Στοιχεία



- **Αντικείμενο εργασίας**

Σε αυτή την εργασία παρουσιάζεται το *Spam*, οι τρόποι με τους οποίους διαδίδεται στο διαδίκτυο, καθώς και οι μηχανισμοί που έχουν αναπτυχθεί για να καταπολεμηθεί. Μελετούμε το *Thunderbird* και αναπτύσσουμε μία επέκταση φίλτρου.

- **Στόχος εργασίας**

Η επέκταση «*Arachnoid*» έχει σαν στόχο να ανιχνεύει και να απομακρύνει όλα τα μηνύματα *Spam* ενός χρήστη καθώς αυτά φθάνουν στον *Thunderbird*.

# Τι είναι το Spam



**RE: bogdan.botezatu@gmail.com, Buy One - Get Four Free!** Spam | X

★ from   
to 

[hide details](#) Nov 14 (4 days ago) [Reply](#) ▼

date Fri, Nov 14, 2008 at 1:33 PM

subject RE: bogdan.botezatu@gmail.com, Buy One - Get Four Free!

**Warning: This message may not be from whom it claims to be. Beware of following any links in it or of providing the sender with any personal information. [Learn more](#)**

If you have images disabled or have trouble viewing this message, please [click here](#).

[Weekly Special SALE- Order Now While Quantities Last!](#)

## Save up to 50% off Retail

[DSW Shoes](#)

**\*Easy Monthly Payments \*No Risk 30-Day Money Back Guarantee**

**Please feel free to forward this special announcement  
to your friends so they can enjoy this special offer.**

[VIEW PRIVACY POLICY](#)

This email was sent to: @gmail.com

You have told us you would like to receive exciting e-mail offers from

[CLICK HERE TO UNSUBSCRIBE](#)

Please allow 24 - 48 hours for processing.

Products limited and may sell out at any time. Prices are subject to change.

# Χαρακτηριστικά *Sprint*



- Έχει την μορφή ενημερωτικών ή διαφημιστικών μηνυμάτων για προϊόντα ή υπηρεσίες
- Χαρακτηρίζεται από χαμηλό κόστος αποστολής
- Είναι απρόκλητο αφού δεν υπάρχει κάποια σχέση μεταξύ παραλήπτη και αποστολέα
- Είναι εμπορικό με σκοπό την προβολή και την διαφήμιση προϊόντων και υπηρεσιών για την προσέλκυση πελατών και την πραγματοποίηση πωλήσεων
- Είναι μαζικό από τον αποστολέα σε ένα πλήθος παραληπτών

# Τρόποι Διάδοσης



- Πρωταρχική τακτική είναι η συλλογή πραγματικών διευθύνσεων μέσω:
  - *mailing list archives*
  - ροές *usenet*
  - δικτυακούς τόπους
  - εξαπλώνοντας κακόβουλα προγράμματα
  - αγοράς λιστών διευθύνσεων από παράνομες πηγές
- Αποστολή των μηνυμάτων αγοράζοντας *bandwidth* ή κάνοντας χρήση *botnet*

# Επιπτώσεις



- Είναι φαινόμενο δυσάρεστο, ενοχλητικό και απαράδεκτο από τους παραλήπτες
- Οδηγεί σε κατάχρηση πόρων του Διαδικτύου
- Θέτει σε κίνδυνο την ασφάλεια και την αξιοπιστία του διαδικτύου

# Τρόποι Αποφυγής και Αντιμετώπισης



- Δημιουργία διεύθυνσης ηλεκτρονικού ταχυδρομείου αποκλειστικά για συναλλαγές στον Ιστό
- Χρήση διαδικτυακής υπηρεσίας δωρεάν ηλεκτρονικού ταχυδρομείου
- Η διεύθυνση ηλεκτρονικού ταχυδρομείου να παραμένει προσωπική
- Χρήση «καμουφλαρισμένων» διευθύνσεων
- Έλεγχος στις πολιτικές απορρήτου των διαδικτυακών τοποθεσιών

Υπάρχουν δύο τρόποι προστασίας από *spam*:

- Μέσω εφαρμογών *spam blockers*
- Με τη χρήση *spam* φίλτρων

# Μηχανισμοί Αντιμετώπισης 1/2



- **Μέθοδος Ανάλυσης Περιεχομένου**

Στηρίζονται στην ανάλυση του περιεχομένου των μηνυμάτων, με σκοπό την ομαδοποίησή τους βάσει κοινών χαρακτηριστικών

- **Το *Checksum-based* φιλτράρισμα**

Ελέγχει τις λέξεις σε κάθε εισερχόμενο μήνυμα και εάν βρει λέξεις οι οποίες έχουν θεωρηθεί ως *spam*, τις μαρκάρει. Δημιουργεί ένα μοναδικό αριθμό κλειδί με βάση τις λέξεις που βρήκε και το αποθηκεύει για μελλοντική χρήση

- **Οι *DNS Blacklists***

Όταν το σύστημα δέχεται πολλά *spam* από μια διεύθυνση IP, στέλνει την διεύθυνση αυτή σε ένα DNSBLs και την επόμενη φορά τα εισερχόμενα από αυτήν *email* θεωρούνται αυτόματα ως *spam*.



# Μηχανισμοί Αντιμετώπισης 2/2



- **Οι Μαύρες Λίστες**

Χρησιμοποιεί μια γνωστή λίστα διευθύνσεων των οποίων τα μηνύματα ηλεκτρονικής αλληλογραφίας απορρίπτονται άμεσα ως *Spam* από το εκάστοτε φίλτρο.

- **Η Μέθοδος Φήμης**

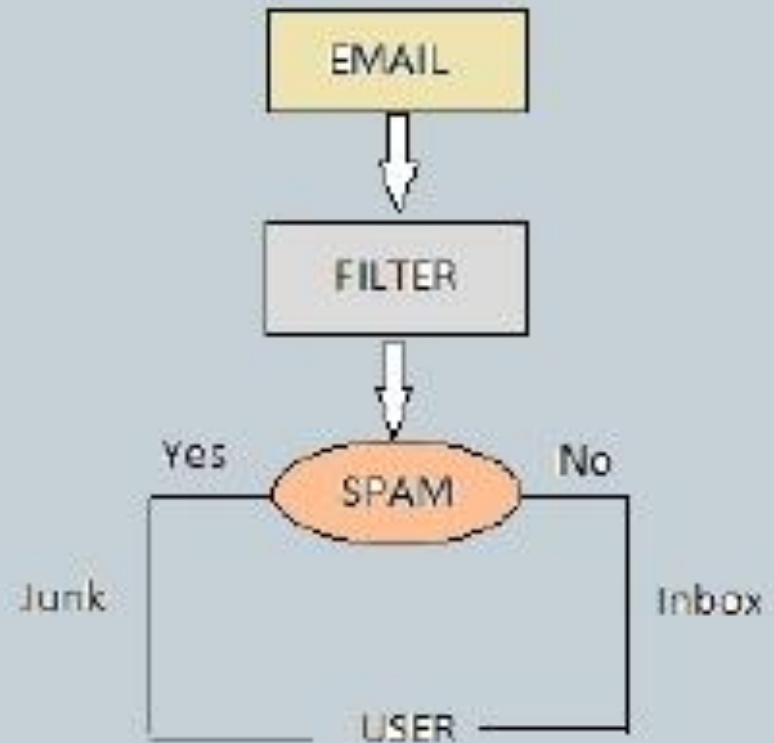
Στηρίζεται στην απόδοση ενός βαθμού αξιοπιστίας σε κάθε αποστολέα ηλεκτρονικής αλληλογραφίας. Όταν ο βαθμός αυτός είναι υψηλότερος από ένα όριο τα μηνύματα γίνονται αποδεκτά ενώ αν είναι χαμηλότερος απορρίπτονται

# Σχεδιασμός της Επέκτασης «Arachnoid»



Για την αντιμετώπιση των ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου σχεδιάστηκε φίλτρο με το όνομα «Arachnoid» υπό την μορφή επέκτασης. Στην συνέχεια ενσωματώθηκε στο πρόγραμμα ανάγνωσης μηνυμάτων ηλεκτρονικού ταχυδρομείου της Mozilla, το *Thunderbird*.

Το «Arachnoid» φίλτρο που αναπτύχθηκε επεκτείνει την απόδοση του προϋπάρχοντος φίλτρου του *Thunderbird* χρησιμοποιώντας μια ανεξάρτητη βιβλιοθήκη λέξεων κλειδιών ορισμένων από το χρήστη και μέσω κανόνων χαρακτηρίζει τα μηνύματα ως ανεπιθύμητα.



# Λειτουργικές απαιτήσεις



- Ανίχνευση μηνυμάτων *Spam*
- Ανίχνευση μηνυμάτων *Hoax*
- Σάρωση μηνυμάτων
- Σάρωση φακέλων
- Μαύρη Λίστα Αποστολέων
- Προσθήκη στην Μαύρη Λίστα
- Λευκή Λίστα Αποστολέων
- Προσθήκη στην Λευκή Λίστα
- Προσθήκη στην Λίστα Λέξεων Κλειδιών
- Δείκτης επιπέδου *Spam*
- Δείκτης χαρακτηρισμού *Spam*

# Μη Λειτουργικές απαιτήσεις



- **Χρόνος Απόκρισης**

Πρέπει πάντοτε να ανταποκρίνεται άμεσα με αντίστοιχα μηνύματα

- **Απαιτήσεις συστήματος**

Πρέπει να λειτουργεί σε όλες τις διαθέσιμες εκδόσεις του *Thunderbird* και να αναβαθμίζεται εύκολα

# Τεχνολογικό Περιβάλλον Ανάπτυξης

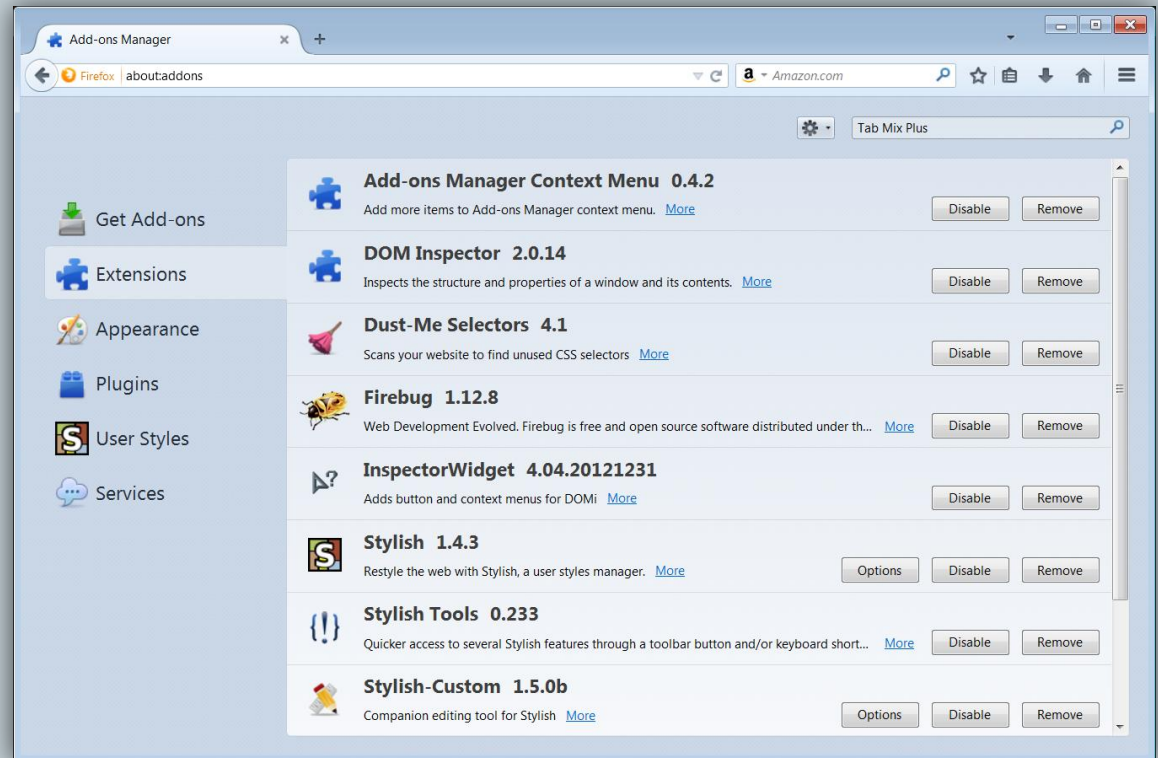


- Η ανάπτυξη της εφαρμογής έγινε με την χρήση των λογισμικών *Notepad++* και *NetBeans* δ
- Οι γλώσσες που χρησιμοποιήθηκαν κατά την ανάπτυξη της εφαρμογής είναι η *Javascript* και η *XUL* ή αλλιώς *XML User Interface Language*
- Απαιτούμενη προϋπόθεση για την χρήση των προγραμμάτων ήταν η εγκατάσταση των εργαλείων της *Java*, *Java Development Kit (JDK)*
- Για την αποσφαλμάτωση χρησιμοποιήθηκε ο ενσωματωμένος *Debugger* του *Mozilla Firefox* και το *Error Console* που βρίσκεται ενσωματωμένο στο *Thunderbird*.

# Τι είναι το Extension



Τα extensions είναι μικρά προγράμματα που μπορούν να εγκατασταθούν από τους χρήστες για να προστεθούν χαρακτηριστικά ή να βελτιστοποιήσουν τον *Thunderbird*



# Η δομή ενός Extension



Η δομή του περιεχομένου του ΧΡΙ αρχείου έχει ως εξής:

*extension.xpi:*

- */install.rdf*
- */defaults/*
- */defaults/preferences/\*.js*
- */chrome.manifest*
- */chrome/*
- */chrome/content/*
- */chrome/skin*
- */chrome/locale*
- */chrome/modules*

# Η Γλώσσα Προγραμματισμού *XUL*



Η *XUL* (*XML-based User Interface Language*), με τύπο αρχείων *\*.xul*,

- είναι μια γλώσσα διεπαφής χρήστη βασισμένη στην *XML*
- επιτρέπει στους προγραμματιστές να σχεδιάσουν *cross-platform* εφαρμογές που μπορούν να τρέχουν *online* αλλά και *offline*

Η *XUL* ορίζει διάφορα είδη στοιχείων, όπως:

- στοιχεία ανώτερων επιπέδων, π.χ. *window*, *page*, *dialog* κλπ.
- *widgets*, π.χ. *label*, *button*, *textbox*, *listbox*, *menulist*, *menupopup*, *checkbox*, *checkbox*, *menu*, *toolbar*, *groupbox*, *tabbox*, *colorpicker*, *spacer*, *splitter* κλπ.
- *events* και *scripts*
- πηγές δεδομένων
- και άλλα όπως *overlay*, *iframe* κ.λ.π.



# Η Γλώσσα Προγραμματισμού *JavaScript*



Η *JavaScript*, με τύπο αρχείων *\*.js*:

- Είναι μια αντικειμενοστραφής scripting γλώσσα, ανεξάρτητη πλατφόρμας και χρησιμοποιείται την για ανάπτυξη *client-side* εφαρμογών
- Το επίσημο όνομά της είναι *ECMAScript* και αναπτύχθηκε από τον *Brendan Eich* στη *Netscape*
- Είχε το όνομα *Mocha* στη συνέχεια μετονομάστηκε σε *LiveScript* και τελικά σε *JavaScript*
- Η *JavaScript* υποστηρίζει τη σύνταξη δομημένου προγραμματισμού της *C*
- Τα αντικείμενα της δε βασίζονται σε κλάσεις αλλά σε «πρωτότυπα»
- Δε γίνεται διαχωρισμός μεταξύ των συναρτήσεων και των μεθόδων παρά μόνο κατά την κλήση μιας συνάρτησης

# Η Γλώσσα Προγραμματισμού *JavaScript*



Η *JavaScript*, με τύπο αρχείων *\*.js*:

- Υποστηρίζει κανονικές εκφράσεις
- Ανταποκρίνεται πολύ γρήγορα στις ενέργειες του χρήστη
- Μπορεί να ανιχνεύσει τις ενέργειες του χρήστη

Χρησιμοποιείται στις επεκτάσεις της *Mozilla* για να παρέχει λειτουργικότητα κατά το «χρόνο εκτέλεσης» και αλληλεπίδρασης με το χρήστη, την αποθήκευση των προτιμήσεων του χρήστη και για τη δημιουργία στοιχείων *XPCOM*.

# Υλοποίηση Εφαρμογής



Η προέκταση *Arachnoid Spam Filter* απαιτεί εγκατάσταση από τον χρήστη κάνοντας *drag-n-drop* το αρχείο *arachnoidspamfilter.xpi* στο μενού *Tools* και επιλογή *Add-ons*.



## Arachnoid Spam Filter 1.0

By [Vrynios Panagiotis](#) for ATEI Thessalonikis

Plug-in Για την Εφαρμογή Mozilla Thunderbird στο πλαίσιο του προγράμματος Arachnoid.

Automatic Updates

Default  On  Off

Last Updated

Τρίτη, 7 Οκτωβρίου 2014

Homepage

[www.it.teithe.gr](http://www.it.teithe.gr)

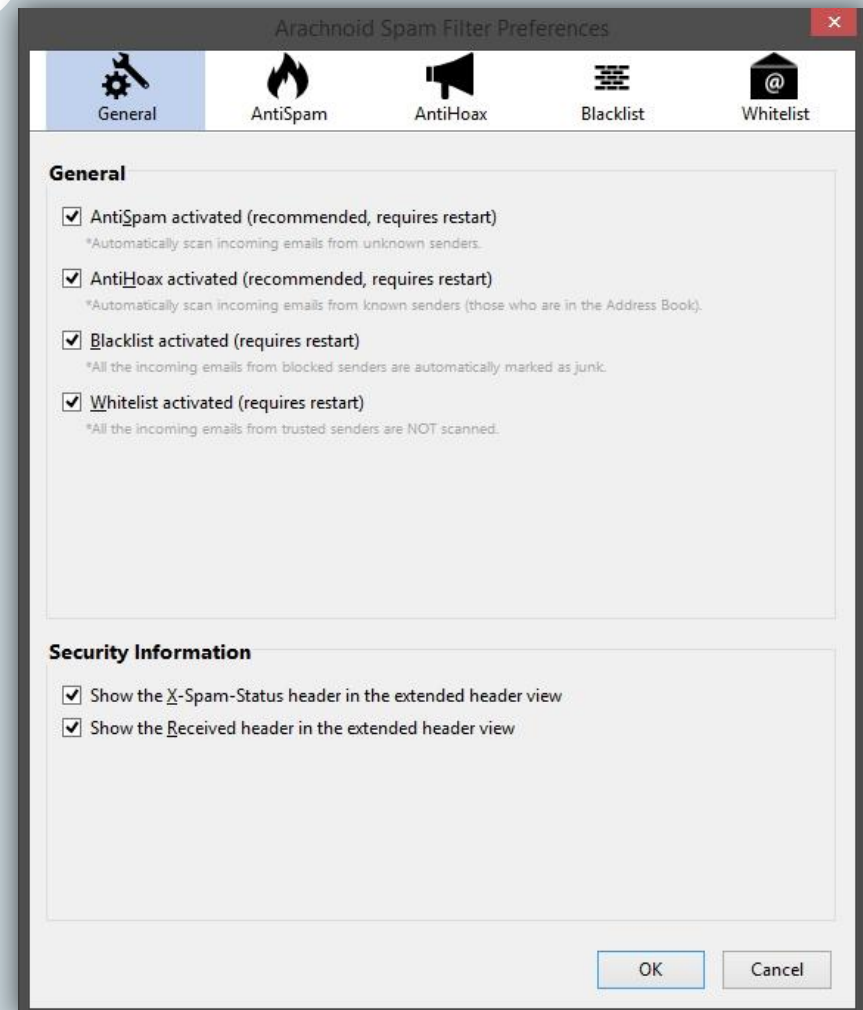
[Options](#)

# Η Καρτέλα επιλογών *General*

Ανοίγοντας το παράθυρο επιλογών παρουσιάζεται στον χρήστη η καρτέλα *General* με τις γενικές επιλογές του φίλτρου.

Εδώ ο χρήστης μπορεί να

- Ενεργοποιήσει την προστασία *AntiSpam* και *AntiHoax*
- Επιλέξει αν θέλει να κάνει χρήση της Μαύρης ή και της Άσπρης Λίστας αποστολέων.

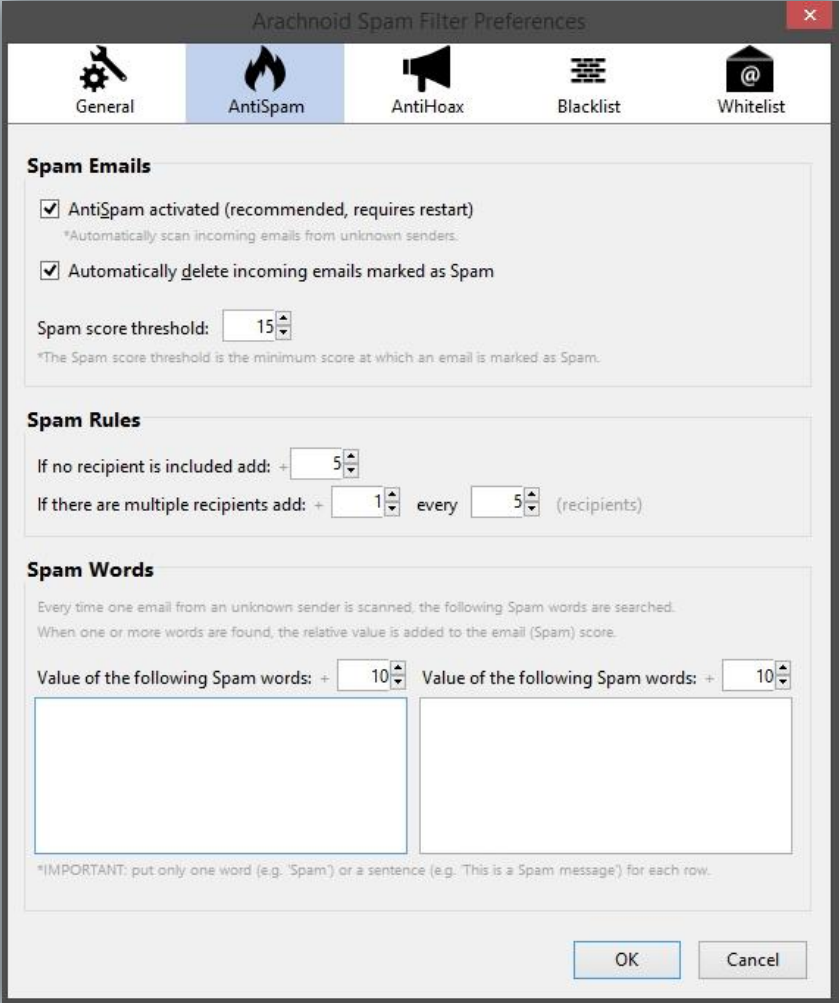


# Η Καρτέλα επιλογών *AntiSpam*

Στην καρτέλα *AntiSpam* ο χρήστης μπορεί να επιλέξει τις ρυθμίσεις οι οποίες διέπουν την ανίχνευση των μηνυμάτων *Spam*.

Δύνονται οι επιλογές:

- Ενεργοποίησης Ανίχνευσης
- Αυτόματης διαγραφής
- Κατώφλι πόντων ανίχνευσης
- Ορισμός πόντων σε διάφορες συνθήκες
- Λίστες με λέξεις κλειδιά με δύο ξεχωριστές βαρύτητες



The screenshot shows the 'Arachnoid Spam Filter Preferences' dialog box with the 'AntiSpam' tab selected. The 'Spam Emails' section has 'AntiSpam activated' and 'Automatically delete incoming emails marked as Spam' checked, with a 'Spam score threshold' of 15. The 'Spam Rules' section has 'If no recipient is included add: + 5' and 'If there are multiple recipients add: + 1 every 5 (recipients)'. The 'Spam Words' section has two input fields, each with a value of 10. The dialog has 'OK' and 'Cancel' buttons at the bottom right.

Arachnoid Spam Filter Preferences

General AntiSpam AntiHoax Blacklist Whitelist

**Spam Emails**

AntiSpam activated (recommended, requires restart)  
\*Automatically scan incoming emails from unknown senders.

Automatically ddelete incoming emails marked as Spam

Spam score threshold:   
\*The Spam score threshold is the minimum score at which an email is marked as Spam.

**Spam Rules**

If no recipient is included add: +

If there are multiple recipients add: +  every  (recipients)

**Spam Words**

Every time one email from an unknown sender is scanned, the following Spam words are searched.  
When one or more words are found, the relative value is added to the email (Spam) score.

Value of the following Spam words: +  Value of the following Spam words: +

\*IMPORTANT: put only one word (e.g. 'Spam') or a sentence (e.g. 'This is a Spam message') for each row.

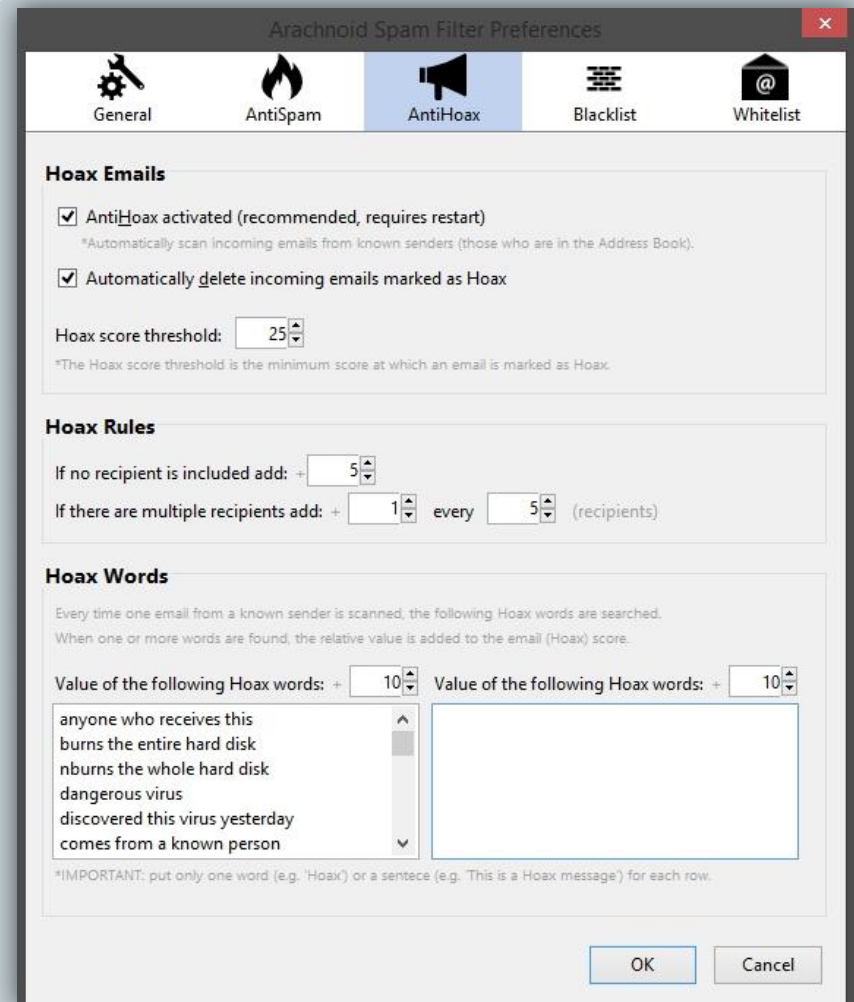
OK Cancel

# Η Καρτέλα επιλογών *AntiHoax*

Στην καρτέλα *AntiHoax* ο χρήστης μπορεί να επιλέξει τις ρυθμίσεις οι οποίες διέπουν την ανίχνευση των μηνυμάτων *Hoax*.

Δύνονται οι επιλογές:

- Ενεργοποίησης Ανίχνευσης
- Αυτόματης διαγραφής
- Κατώφλι πόντων ανίχνευσης
- Ορισμός πόντων σε διάφορες συνθήκες
- Λίστες με λέξεις κλειδιά με δύο ξεχωριστές βαρύτητες



The screenshot shows the 'Arachnoid Spam Filter Preferences' dialog box with the 'AntiHoax' tab selected. The dialog has a title bar with a close button and a navigation bar with icons for General, AntiSpam, AntiHoax, Blacklist, and Whitelist. The 'AntiHoax' section is active and contains the following settings:

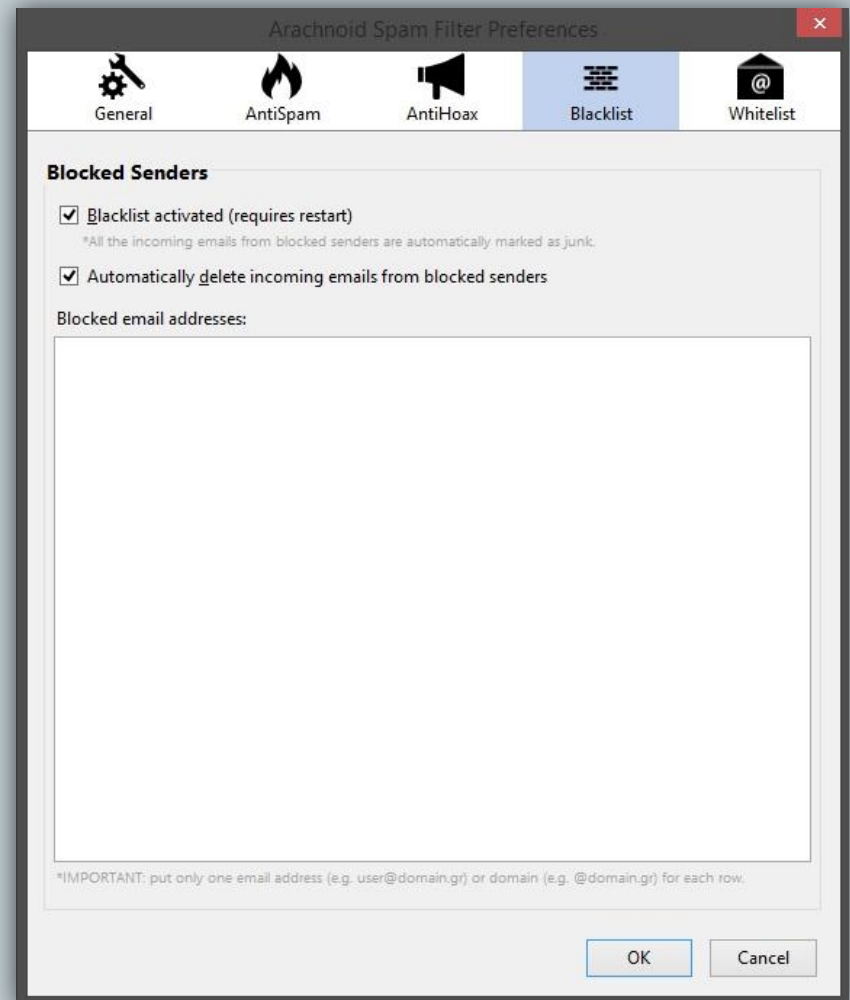
- Hoax Emails**
  - AntiHoax activated (recommended, requires restart)  
\*Automatically scan incoming emails from known senders (those who are in the Address Book).
  - Automatically delete incoming emails marked as Hoax
  - Hoax score threshold:   
\*The Hoax score threshold is the minimum score at which an email is marked as Hoax.
- Hoax Rules**
  - If no recipient is included add: +
  - If there are multiple recipients add: +  every  (recipients)
- Hoax Words**
  - Every time one email from a known sender is scanned, the following Hoax words are searched. When one or more words are found, the relative value is added to the email (Hoax) score.
  - Value of the following Hoax words: +  Value of the following Hoax words: +
  - A list of words is shown in a scrollable area:
    - anyone who receives this
    - burns the entire hard disk
    - nburns the whole hard disk
    - dangerous virus
    - discovered this virus yesterday
    - comes from a known person
  - \*IMPORTANT: put only one word (e.g. 'Hoax') or a sentence (e.g. 'This is a Hoax message') for each row.

At the bottom right, there are 'OK' and 'Cancel' buttons.

# Η Καρτέλα επιλογών *Blacklist*

Στην καρτέλα *Blacklist* δύνονται οι επιλογές:

- Ενεργοποίησης Μαύρης Λίστας
- Αυτόματης διαγραφής
- Προσθήκης/Διαγραφής διευθύνσεων και *Domain* από την Λίστα



# Η Καρτέλα επιλογών *Whitelist*

Στην καρτέλα *Whitelist* δύνονται οι επιλογές:

- Ενεργοποίησης Λευκής Λίστας
- Προσθήκης/Διαγραφής διευθύνσεων και *Domain* από την Λίστα



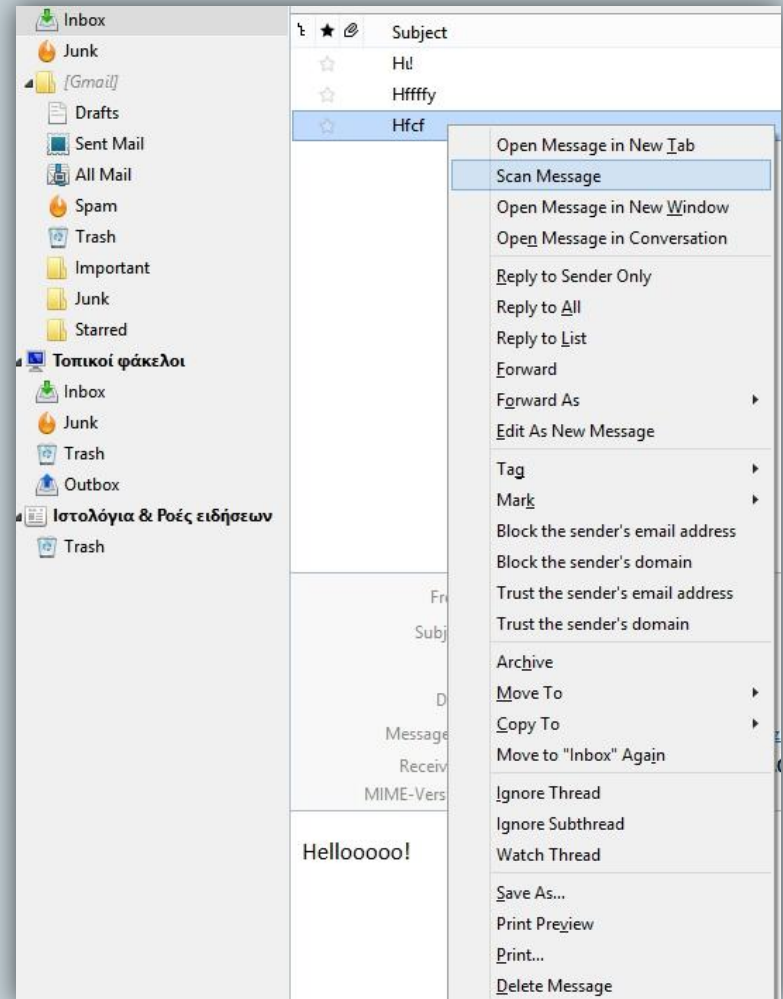


# Το μενού Επιλογών



Με την επιλογή ενός ή πολλαπλών μηνυμάτων ή φακέλων ο χρήστης μπορεί να:

- Εκκινήσει διαδικασία σάρωσης για ανίχνευση μηνυμάτων *Spam*.
- Προσθέσει την διεύθυνση ηλεκτρονικού ταχυδρομείου ή το *domain* του αποστολέα στην Μαύρη Λίστα ή στην Λευκή Λίστα



# Η Επιλογή *Scan*



Η σάρωση ενός μηνύματος ολοκληρώνεται με την εμφάνιση σχετικού παράθυρου διεπαφής ενημέρωσης του χρήστη

**Arachnoid Spam Filter**

**Scan**  
Status: ---

Start time: Τρίτη, 7 Οκτώβριος 2014 18:14:58  
End time: Τρίτη, 7 Οκτώβριος 2014 18:14:58

**Scan Progress**

100%

All Mail:CAMqRC4kxF\_ttPWEs5GzPrfksb+fzXABnDpvocO+Q4-KQ\_OMxbw@mail.gmail.com

**Security Information**

Sender: @gmail.com [unknown]  
Received: from mail-wi0-x243.google.com (10.07.08.10) by mx.google.com (10.216.188.197)

Spam Rate: > **100%**  
X-Spam-Status: ---

OK

# Η Επιλογή *Scan* Φακέλου



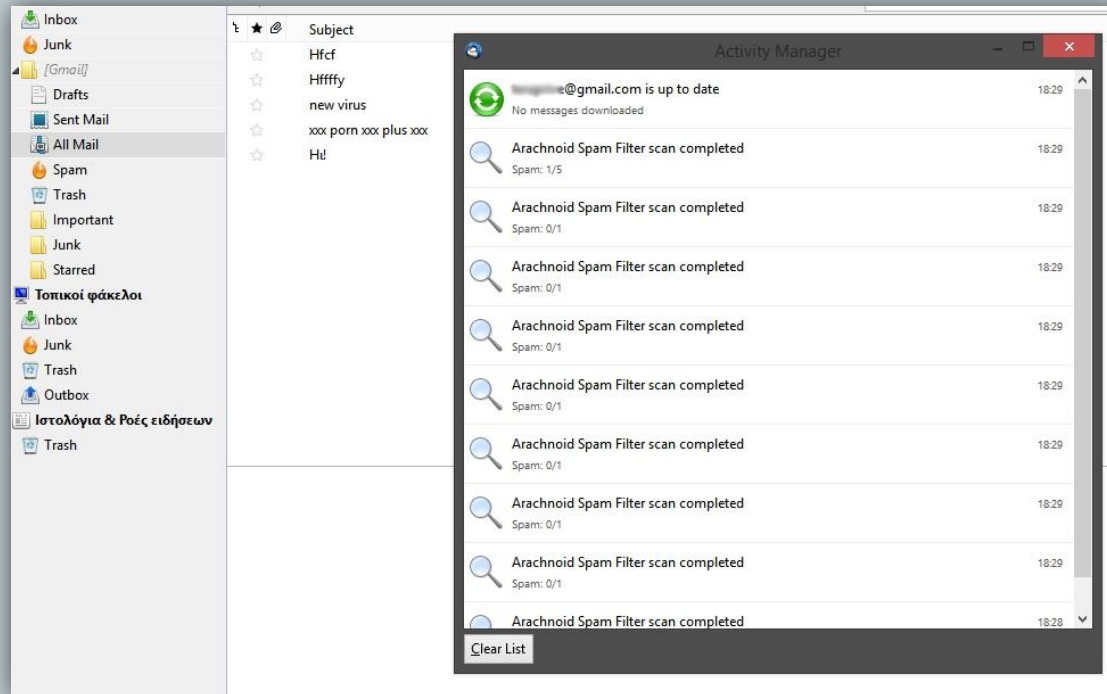
Η σάρωση ενός φακέλου ολοκληρώνεται με την εμφάνιση σχετικού παράθυρου διεπαφής ενημέρωσης του χρήστη

The screenshot displays an email client interface with a sidebar on the left containing folders like 'Inbox', 'Junk', '[Gmail]', 'Drafts', 'Sent Mail', 'All Mail', 'Spam', 'Trash', 'Important', 'Junk', 'Starred', and 'Τοπικοί φάκελοι'. The main pane shows an email with subject 'Subject' and headers 'Hd', 'Hffffy', and 'Hfcf'. A dialog box titled 'Arachnoid Spam Filter' is overlaid, showing a 'Scan' progress bar that is nearly full. The scan details include: Start time: Τρίτη, 7 Οκτώβριος 2014 18:05:16; End time: Τρίτη, 7 Οκτώβριος 2014 18:05:16; Status: ---. The scan progress bar is labeled 'Scan Progress' and shows a green bar at 100% completion for the folder 'Inbox:CAFswvs2U6qugXE\_DdLh9MTgwTou5zhVYmUgaLayh6LLXyePyMw@mail.gmail.com'. Below the progress bar is a 'Security Information' section with the following details: Sender: [redacted]@gmail.com [unknown]; Received: from mail-ie0-x243.google.com (10.07.08.02) by mx.google.com (10.64.14.2); Spam Rate: ~ 0.00%; X-Spam-Status: ---. An 'OK' button is located at the bottom right of the dialog box. The email content below the dialog box shows 'MIME-Version 1.0' and 'Helloooooo!'.

# Ο Διαχειριστής Ενεργειών



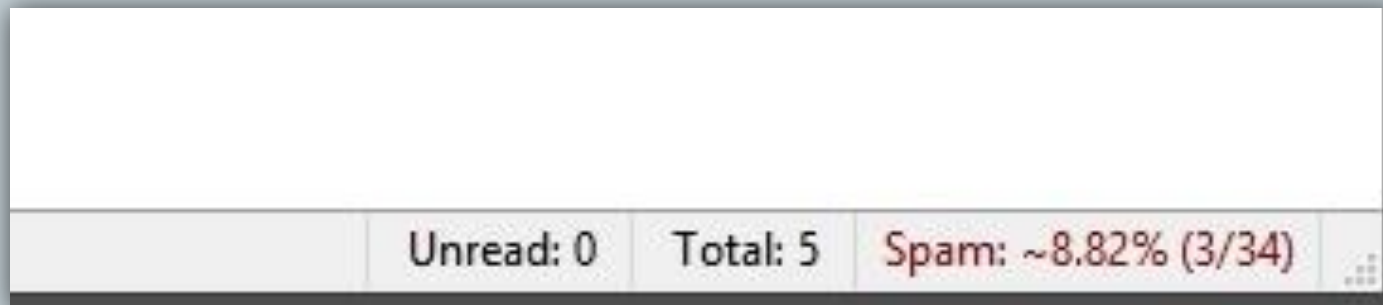
Μετά την ολοκλήρωση κάθε ενέργειας, η εφαρμογή προσθέτει ανάλογο μήνυμα στον ενσωματωμένο *Activity Manager* όπου ο χρήστης μπορεί να ενημερωθεί για τις τελευταίες ενέργειες του φίλτρου καθώς και την ώρα που ολοκληρώθηκαν.



# Ο Μετρητής *Spam*



Με την άφιξη κάθε μηνύματος η εφαρμογή αυξάνει τον μετρητή ληφθέντων μηνυμάτων στην γραμμή κατάστασης του *Mozilla Thunderbird*. Σε αναλογία με τα μηνύματα *Ham* ο μετρητής μετρά και των αριθμών των ληφθέντων μηνυμάτων τα οποία χαρακτηρίστηκαν ως *Spam* και παρουσιάζει ένα ποσοστό λήψης *Spam* στον χρήστη



# Αξιολόγηση Φίλτρων *Sram*



- Η αξιολόγηση ενός συστήματος ταξινομητή είναι ένα κρίσιμο σημείο και απαιτούνται τυποποιημένες μετρήσεις, συλλογές και διαδικασίες, οι οποίες να επιτρέπουν την σύγκριση διαφορετικών ερευνητικών έργων
- Τα κύρια θέματα στην αξιολόγηση των φίλτρα *sram* είναι οι συλλογές δοκιμών, η λειτουργική διαδικασία, και οι μετρήσεις αξιολόγησης

# Οι Συλλογές Δοκιμών



- Μια συλλογή δοκιμών είναι ένα σύνολο ήδη διαβαθμισμένων μηνυμάτων που αποστέλλονται σε ένα ταξινομητή προκειμένου να μετρηθεί η αποτελεσματικότητά του, όσον αφορά τις επιτυχίες και τα λάθη.
- Μερικές ενδεικτικές συλλογές είναι οι εξής:
  - SpamBase23, που περιέχει 4601 μηνύματα, από τα οποία τα 1813 (39%) έχουν χαρακτηριστεί ως *spam*.
  - PU1, που περιέχει 1099 μηνύματα, από τα οποία τα 481 (43%) έχουν χαρακτηριστεί ως *spam*.
  - Lingspam, που περιέχει 2893 μηνύματα, από τα οποία τα 481 (16%) έχουν χαρακτηριστεί ως *spam*.

# Η Διαδικασία Δοκιμής



- Η πιο συχνή διαδικασία αξιολόγησης για τα φίλτρα *spam* είναι η αξιολόγηση παρτίδας όπου το φίλτρο εκπαιδεύεται σε ένα σύνολο μηνυμάτων και εφαρμόζεται σε ένα διαφορετικό σύνολο μηνυμάτων για δοκιμή.
- Τα δοκιμαστικά μηνύματα επισημαίνονται ως *Ham* ή *Spam*, και είναι δυνατόν να συγκριθεί η απόφαση του φίλτρου υπολογίζοντας τις επιτυχίες και τα λάθη
- Είναι σημαντικό, η συλλογή δοκιμής να είναι παρόμοια, αλλά και διαφορετική από αυτήν που εκπαίδευσε το φίλτρο ώστε να εξομοιώνει όσο πιο πιστά γίνεται τις πραγματικές συνθήκες



# Οι Μετρικές Αξιολόγησης



Χωρίζονται σε 3 ομάδες:

- η βασική μετρική
  - που χρησιμοποιείται στα αρχικά έργα
- η μέθοδος ROCCH
  - ως μία προχωρημένη μέθοδος για πρωταρχικά λάθη
- και οι μετρικές TREC
  - ως το ισχύον πρότυπο

# Πειραματική Αξιολόγηση Φίλτρου *Arachnoid*



- Οι δοκιμές έγιναν με διαφορετικές ρυθμίσεις ευαισθησίας του φίλτρου και σε συνδυασμό με το ενσωματωμένο φίλτρο του *Thunderbird* ή χωρίς.
- Χρησιμοποιήθηκε το σύνολο μηνυμάτων *Enron-Spam* που περιέχει επιθυμητά και ανεπιθύμητα μηνύματα έξι ψευδό-χρηστών. Στα μηνύματα της παραπάνω συλλογής αξιολογήσαμε μόνο το σώμα και το θέμα του κάθε μηνύματος, αγνοώντας συνημμένα αρχεία, ετικέτες *HTML* και κεφαλίδες.
- Για την αναπαράσταση των αποτελεσμάτων χρησιμοποιήθηκαν πίνακες συσχέτισης. Αυτό το είδος αναπαράστασης αποτελεσμάτων εστιάζει στην προγνωστική ικανότητα του μοντέλου.
- Σε κάθε πείραμα χρησιμοποιήθηκαν 300 μηνύματα, εκ των οποίων 200 ήταν *Spam* και 100 ήταν *Ham*

# Μετρικές Πειράματος



Οι εξισώσεις που προκύπτουν μέσα από τις μετρικές  $TP$ ,  $FP$ ,  $FN$ ,  $TN$  των δυαδικών προβλημάτων ταξινόμησης, ονομάζονται:

- Ευαισθησία (*Sensitivity* -  $TPR$ )
  - Αντιπροσωπεύει το ποσοστό ανάκλησης ανεπιθύμητων μηνυμάτων (*Spam Recall*,  $SR$ ) και  $SR = TP / (TP + FN)$
- Σαφήνεια (*Specificity* -  $TNR$ )
  - Αντιπροσωπεύει το ποσοστό ανάκλησης επιθυμητών μηνυμάτων (*Ham Recall*,  $HR$ ) και  $HR = TN / (TN + FP)$

Όπου:

- $TP$  : **True Positives**, δηλαδή πόσα ανεπιθύμητα μηνύματα κατετάγησαν σωστά
- $TN$  : **True Negatives**, δηλαδή πόσα επιθυμητά μηνύματα κατετάγησαν σωστά
- $FP$ : **False Positive**, αριθμός ανεπιθύμητων μηνυμάτων που κατετάγησαν **λανθασμένα**
- $FN$ : **False Negative**, αριθμός επιθυμητών μηνύματα που κατετάγησαν **λανθασμένα**

# Αποτελέσματα Πειράματος 1



Στο πρώτο πείραμα τα μηνύματα σαρώθηκαν κατά την άφιξη τους από το προεπιλεγμένο φίλτρο του *Thunderbird*. Από τα αποτελέσματα προέκυψε:

- $Sensitivity = TP / (TP + FN) = 163 / (163 + 37) = 0.815 = 81,5\%$
- $Specificity = TN / (TN + FP) = 74 / (74 + 26) = 0.74 = 74\%$

# Αποτελέσματα Πειράματος 2



Στο δεύτερο πείραμα τα μηνύματα σαρώθηκαν κατά την άφιξη τους από την επέκταση *Arachnoid* ενώ το φίλτρο της *Mozilla* είχε απενεργοποιηθεί χειροκίνητα από τις ρυθμίσεις. Στις επιλογές της επέκτασης στην καρτέλα «*AntiSpam*» χρησιμοποιήθηκε «*Spam Score Threshold*» ίσο με 25, «*High Spam Word Value*» ίσο με 7 «*Low Spam Word Value*» ίσο με 1. Από τα αποτελέσματα προέκυψε:

- $Sensitivity = TP / (TP + FN) = 159 / (159 + 41) = 0.815 = 79,5\%$
- $Specificity = TN / (TN + FP) = 72 / (72 + 28) = 0.72 = 72\%$

# Αποτελέσματα Πειράματος 3



Στο τρίτο πείραμα στις επιλογές της επέκτασης στην καρτέλα «*AntiSpam*» χρησιμοποιήθηκε «*Spam Score Threshold*» ίσο με 100, «*High Spam Word Value*» ίσο με 91 και «*Low Spam Word Value*» ίσο με 6. Από τα αποτελέσματα προέκυψε:

- $Sensitivity = TP / (TP + FN) = 167 / (167 + 33) = 0.815 = 83,5\%$
- $Specificity = TN / (TN + FP) = 69 / (69 + 31) = 0.69 = 69\%$

# Αποτελέσματα Πειράματος 4



Στο τέταρτο πείραμα τα μηνύματα σαρώθηκαν κατά την άφιξη τους από τον συνδυασμό των δύο φίλτρων. Η βάση δεδομένων λέξεων κλειδιών της επέκτασης *Arachnoid* βοήθησε το προεπιλεγμένο φίλτρο να περάσει πιο γρήγορα την φάση της εκπαίδευσης και να αρχίσει να αποδίδει. Από τα αποτελέσματα προέκυψε:

- $Sensitivity = TP / (TP + FN) = 169 / (169 + 31) = 0.815 = 84,5\%$
- $Specificity = TN / (TN + FP) = 78 / (78 + 22) = 0.74 = 78\%$

# Συμπεράσματα Πειραμάτων



Από τα πειράματα τα οποία εκτελέστηκαν κατά την διάρκεια της συγγραφής της εργασίας παρατηρήθηκε ότι:

- Με βάση τις τιμές των μετρικών και ειδικά με βάση την υψηλή τιμή της σαφήνειας έναντι της ευαισθησίας, η θετική πρόβλεψη του φίλτρου είναι πιο επιθετική.
- Ενώ η απόδοση των φίλτρων ξεχωριστά δεν παρουσιάζει μεγάλες διαφορές, σε συνδυασμό πετυχαίνουν αρκετά καλύτερα αποτελέσματα και συνολική ευστοχία.



Ευχαριστώ πολύ για την προσοχή σας



Thank You!