

# Spamszűrés

## **EZT OLVASSA EL ELŐSZÖR**

A spamszűrés, azaz a kéréstlen és fontos levelek megkülönböztetése, szűrése statisztikai alapon nyugszik, emiatt az első szabály az, hogy NAGYON JÓL be kell tanítani a szűrőt a megfelelő működéshez. Emiatt érdemes legalább 300-300 levélszemét és nem-levélszemét levéllel közvetlenül a telepítés után egy spamszűrőt betanítani, mielőtt elkezdené "élesben" használni. Második szabály, hogy minél többet használja a szűrőt, az annál jobban fog működni, mert annál több szabályt ismer fel levélszemétre és nem-levélszemétre vonatkozóan. A szűrőfeltételeknél sima szöveget és reguláris kifejezéseket is használhat. Emellett meg lehet adni fehér- és feketelistákat is, amelyekkel közvetlenül megadhatja, mi számít jó levélnek, s mi szemétnak. Továbbá általában a szűrők nyilvános DNS-feketelistákat is elérnek, azaz olyan szerverekről érkező levélszemetet is kiszűrhetnek, amely szervereket a spammerek használnak.

**FONTOS 1.:** Ne töröljön egy levelet sem a spamszűrő ítéletére alapozva. Használja a **Megjelölés levélszemétként** beállítást, s törölje kézzel a spamként megjelölt leveleket, hátha éppen egy fontos levelet szűrte ki, törölne le.

**FONTOS 2.:** A spamszűrés akkor dolgozik jól, ha tanításkor spamot ad meg spamként, s jó leveleket nem-levélszemétként, különben összezavarja a spam-felismerő mechanizmust!

## **Mi a spam?**

Kéréstlen levél. Manapság, mivel az emailért közvetlenül nem kell fizetni, "ingyen van", nagyon sok kéréstlen reklámlevelet kaphat. Ezek egyrészt eltömíthetik postaládáját, másrészt feleslegesen terhelik Önt. Habár a szolgáltatók próbálják ezeket a leveleket még az Ön postaládájába történő megérkezés előtt kiszűrni, ez egy közel lehetetlen feladat.

## **Miért kell szűrni a spamot?**

Kelleni nem kell, de ha a levelezőszoftvere, mint pl. a The Bat!, képes a kéréstlen leveleket megszüntetni, ezzel időt takaríthat meg magának, mikor letölti a leveleit.

## **Hogyan védekezhettek a spam ellen?**

Pár általános tanács:

- ne adja meg kedvenc email címét ismeretlen helyen
- ne használja kedvenc email címét elektronikus listákon
- ne nézzen meg "EGYSZER AZ ÉLETBEN VAN ILYEN AKCIÓ" stílusú levelekben lévő linkeket. Miért? Az esélye, hogy valóban nyert valamit, talán 1 az 1 millióhoz. Maximum. Viszont az emaileket gyűjtő emberek okosak. Olyan linket tesznek a levelébe, ami Önre nézve egyedí, csak Ön tudja megnézni. Azaz ha megnézi, akkor a spammer látja, hogy az Önnek szánt linket "valaki" megnézte, s tudja, hogy az Ön volt, tehát az email címe él, s már adja is el az Ön email címét más cégeknek...
- használja ki a szolgáltatója spamszűrő rendszerét, ha létezik (Freemail-en létezik pl.)
- használjon spamszűrést a levelezőprogramjában

Persze, ha egy olyan kapcsolatának lesz vírusos a gépe, akinek a címjegyzékében az Ön email címe benne van, akkor számíthat arra, hogy a vírustól függően az email címe nagy valószínűséggel kikerül az Internetre, s spamot fog kapni. Azaz tökéletes védekezés nem létezik.

## **Hogyan működnek a spamszűrők?**

Statisztikai alapon. Elemzik a letöltött leveleket, szavakat, szókapcsolatokat keresnek az egész levélben, illetve átnézik a levél fejlécét alaposan, s pontoznak bizonyos dolgokat. Ha a pontszám meghaladja azt az értéket, amit Ön beállított, akkor deklarálja a levelet szemétnak. Pár dolog elég nagy súllyal nyom a latban, amikor egy levelet spamnak minősít egy szűrő, pl.:

- nincs neve a feladónak, csak email címe
- nem Ön a közvetlen címzettje a levélnek, azaz látszik, hogy más is megkapta a levelet
- nem normális SMTP szerverrel küldték a levelet, hanem helyi gépről valami helyi SMTP szerver szoftverrel

Ennyi összefoglalva. Esetleg még egy dolog, ami érdekes lehet: vannak ún. black and white list-ek a spamszűrőknél. Ez azt jelenti, hogy ha a levél a white list-en (fehér lista) szereplő embertől jött, akkor azt a program automatikusan NEM levélszemétnek minősíti, míg ha a black list-en (fekete lista) szereplő embertől jött, akkor azt a program automatikusan levélszemétnek minősíti, függetlenül a levél szerkezetétől, tartalmától.

### Hogyan állítom be a beépített BayesIt spamszűrőt?

Ennek a beépített spamszűrőnek a telepítését **NEM ajánljuk** annak megbízhatatlansága miatt, emiatt semmilyen szinten nem támogatjuk. Javasoljuk a Bayes Filter Plugin használatát.

### Hogyan állítom be a Bayes Filter spamszűrőt?

A telepítés lépései:

- telepítse a Bayes Filter-t a letöltött telepítőprogram futtatásával, majd másolja a spamszűrő magyar nyelvi fájlját (bayesfilter.lng) a Bayes Filter telepítési könyvtárba (alapértelmezésben C:\Program Files\Bayes Filter Plugin), majd indítsa el a The Bat! Programot
- a **Beállítások/Felhasználói beállítások** menüben válassza az Védelem/Anti-spam pontot
- kattintson a **Hozzáadás...** gombra
- keresse meg a Bayes Filter könyvtárát, ez alapértelmezésben a C:\Program Files\Bayes Filter Plugin könyvtár, s az ebben lévő könyvtárból válassza a bayesit.tbp plugin fájlt. kattintson a **Beállítás** gombra, a **General** panelon válassza ki a magyar nyelvet (Hungarian) a **Language** legördülő menüben, majd nyomja meg az **OK** gombot.

A plugin beállításai:

#### Általános panel

The screenshot shows the 'Beállítás' (Settings) dialog box with the following settings:

Általános	Adatbázis	Fekete-/Fehérlista	DNS-fekete lista	RegExp szűrő	Naplófájl	Névjegy
Max. egyidejű szál:	20	Állapotpanel megjelenítési limit:	1			
Kiértékelt szavak száma:	20	Nem fontos szavak eltávolítása	<input checked="" type="checkbox"/>			
Automatikus tanulás a fekete-/fehérlistás levelekből	<input checked="" type="checkbox"/>	Tanulás besorolás korrigálásakor	<input checked="" type="checkbox"/>			
Automatikus tanítás	<input checked="" type="checkbox"/>	Nyelv	Magyar			
Tanulás						
Spam levelekből	<input checked="" type="checkbox"/>	pontszám felett:	90			
Nem-spam levelekből	<input checked="" type="checkbox"/>	pontszám alatt:	5			

Buttons: OK, Mégse

**Max. egyidejű szál:** Maximum ennyi beérkező levelet tud egyidejűleg a plugin ellenőrizni. A nagyobb szám gyorsabb ellenőrzést tesz lehetővé, 40 körüli szám megfelelő.

**Állapotpanel megjelenítési limit:** Az átnézendő levelek száma, aminél megjelenik egy ablak, hogy az ellenőrzés befejeződött. Ha pl. 10-et állít ide be, akkor 9 levélnél még nem kap visszajelzést.

**Kiértékelt szavak száma:** Ennyi szó kell egy levél spamként vagy nem-spamként való megjelöléséhez. A 20 és 70 körüli szám megfelelő érték itt.

**Tanulás besorolás korrigálásakor:** Ha egy levelet a program spamnak jelöl meg, s Ön kézzel azt nem-spamnak jelöli meg, ezt megjegyzi a program.

**Automatikus tanulás a fekete-/fehérlistás levelekből:** Ezt bejelölve a a fekete- és fehérlistás levelekből is tanul a program!

**Nyelv:** A plugin felületének nyelve.

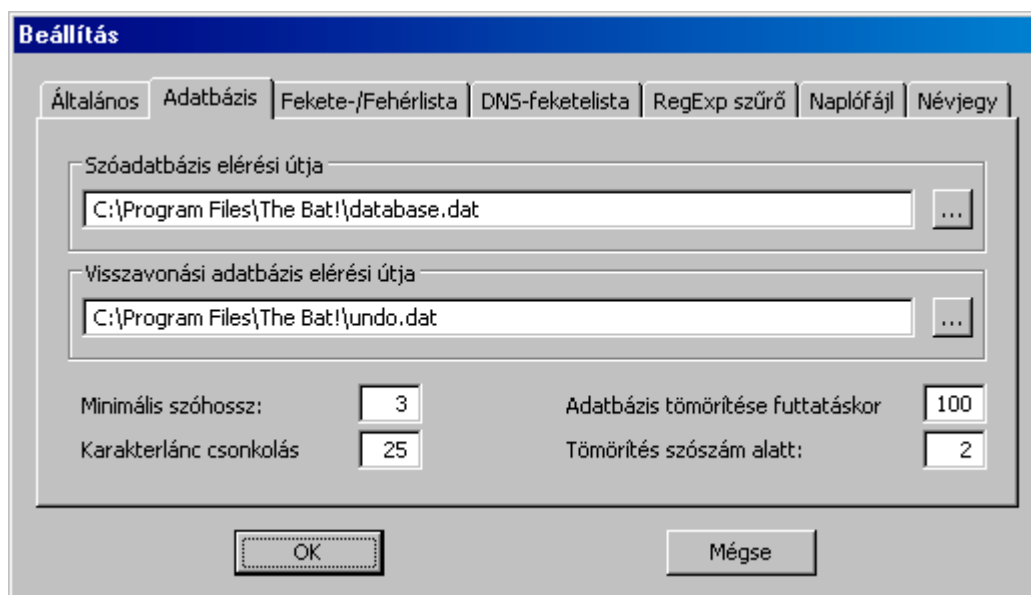
**Nem fontos szavak eltávolítása:** Bejelölve az ellenőrzés előtt a szűrő a levelek végén lévő ellenőrzés szempontjából nem fontos szavakat eltávolítja, hogy ne zavarja az ellenőrzést (spammerek néha a levél végére megzavaró szavakat tesznek).

**Automatikus tanulás:** Ezt bejelölve a program tanul a beérkező levelek általa elvégzett besorolásából.

**Tanulás - Spam levelekből:** A spamokból tanul a szűrő, ha az adott spam az adott **pontszám felett** van. Kapcsolja be, s nagy szám legyen itt!

**Tanulás - Nem-spam levelekből:** A nem-spamokból tanul a szűrő, ha az adott nem-spam az adott **pontszám alatt** van. Kis szám legyen itt!

#### Adatbázis panel



**Szóadatbázis elérési útja:** Az ellenőrzés alapját adó szavak adatbázisát tartalmazó fájl.

**Visszavonási adatbázis elérési útja:** Amikor egy levelet átminősít, annak a szavai kerülnek ide, a program a két adatbázist felhasználva dönt egy-egy levélről.

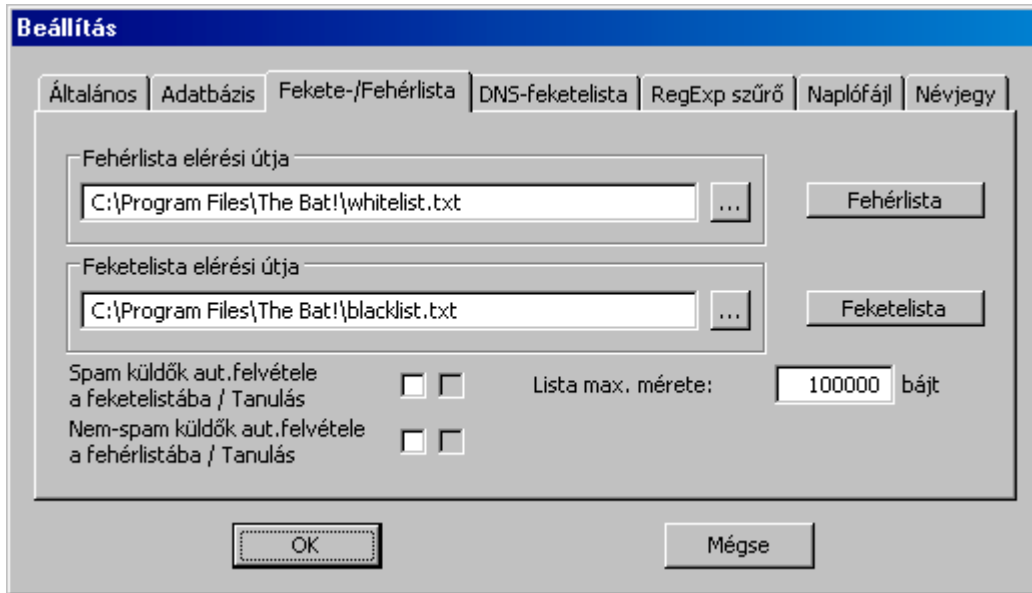
**Minimális szóhossz:** Csak a megadottnál hosszabb szavak lesznek az adatbázisban.

**Karakterlánc csonkolás:** Ha ennél nagyobbak a szavak, akkor a szót csonkolja a szűrő, s a szó, a csonkolás helyében kerül tárolásra az adatbázisban.

**Adatbázis tömörítése futtatáskor:** A szóadatbázist minden ennyi futtatás után tömöríti a plugin.

**Tömörítés szószám alatt:** A plugin tömörítéskor eltávolítja azokat a szavakat az adatbázisból, amelyek a megadott számnál kevesebbszer szerepelnek benne.

## Fekete-/Fehérlista panel



**Fehérlista elérési útja:** A mindenképpen nem-levélszemétnek minősítendő feltételeket tartalmazó fájl neve. Egyszerű tesztfájl ez, minden sora egy-egy fehérlista-tag. Használhat sima szöveget (TEXT) és reguláris kifejezést (REGEXP) szűrőnek. Példa a fájl lehetséges sorára:

TEXT: `spammer@test.com`

REGEXP: `[a-z0-9\-\_\.\,]+\@test.com`

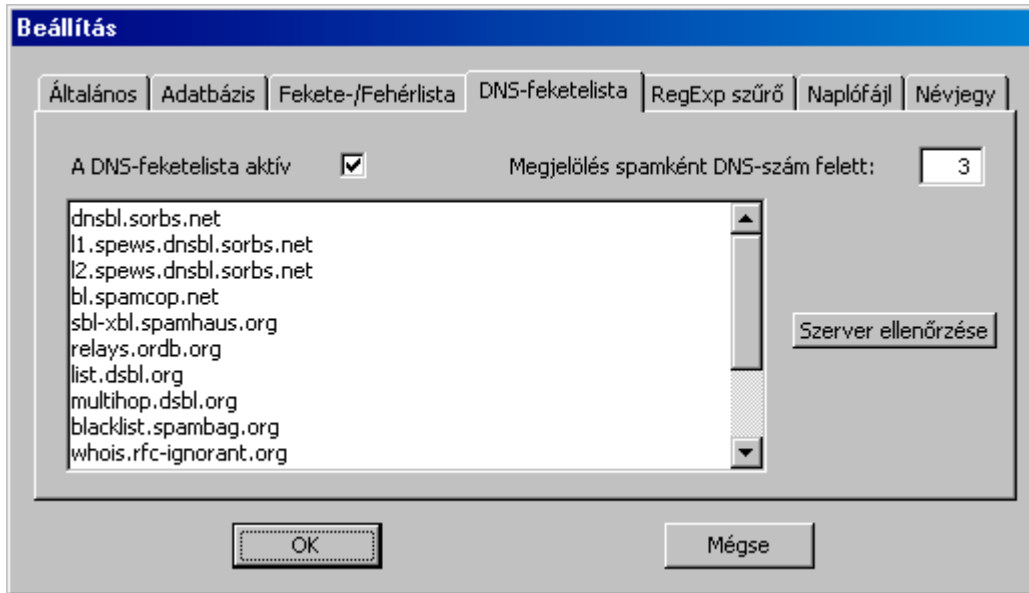
**Feketelista elérési útja:** A mindenképpen levélszemétnek minősítendő feltételeket tartalmazó fájl neve. Egyszerű tesztfájl ez, minden sora egy-egy feketelista-tag. Használhat sima szöveget (TEXT) és reguláris kifejezést (REGEXP) szűrőnek. Példákat az előző bejegyzésben láthat.

**Spam küldők aut. felvétele a feketelistába/Tanulás:** A spam küldője automatikusan a feketelistára kerül, így ezután minden levele spamnak minősül majd.

**Nem-spam küldők aut. felvétele a fehérlistába/Tanulás:** A nem-spam küldője automatikusan a fehérlistára kerül, így ezután minden levele spamnak minősül majd.

**Lista max. mérete:** Néha a fekete-/fehérlista nagyra nő, ezzel a beállítással ezen listák méretének szabhat határt.

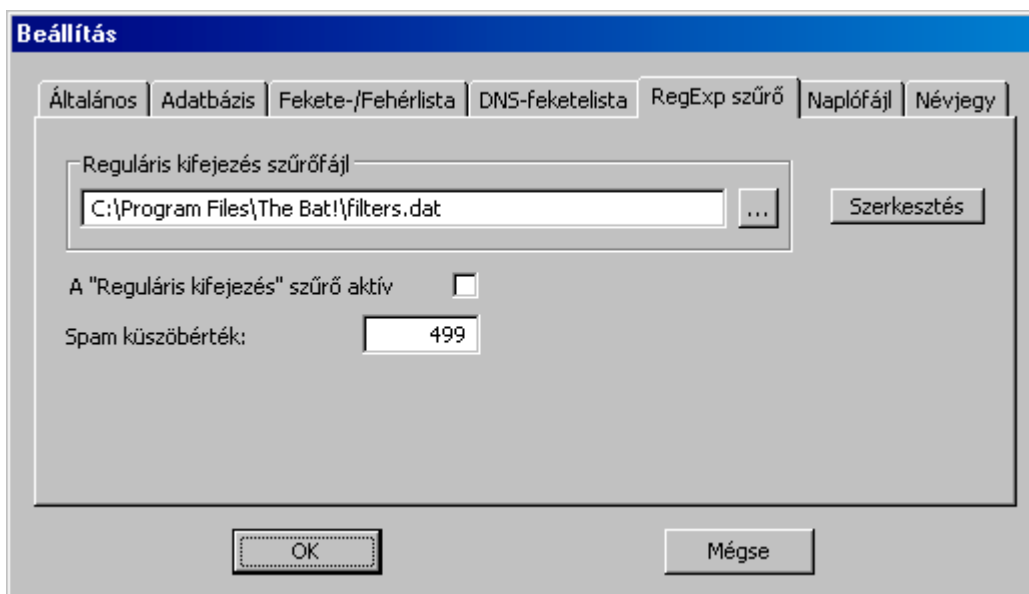
## DNS feketelista panel



**A DNS-fekete-/Fehérlista aktív:** Bejelölve a plugin megnézi, hogy a levélküldő szerepel-e nyilvános feketelista-adatbázisokban. Csak aktív internetelés esetén működik.

**Megjelölés spamként DNS-szám felett:** Ennyi feketelista adatbázisban kell szerepelni a küldőnek, hogy a levél spamnek minősüljön.

## Regexp szűrő panel

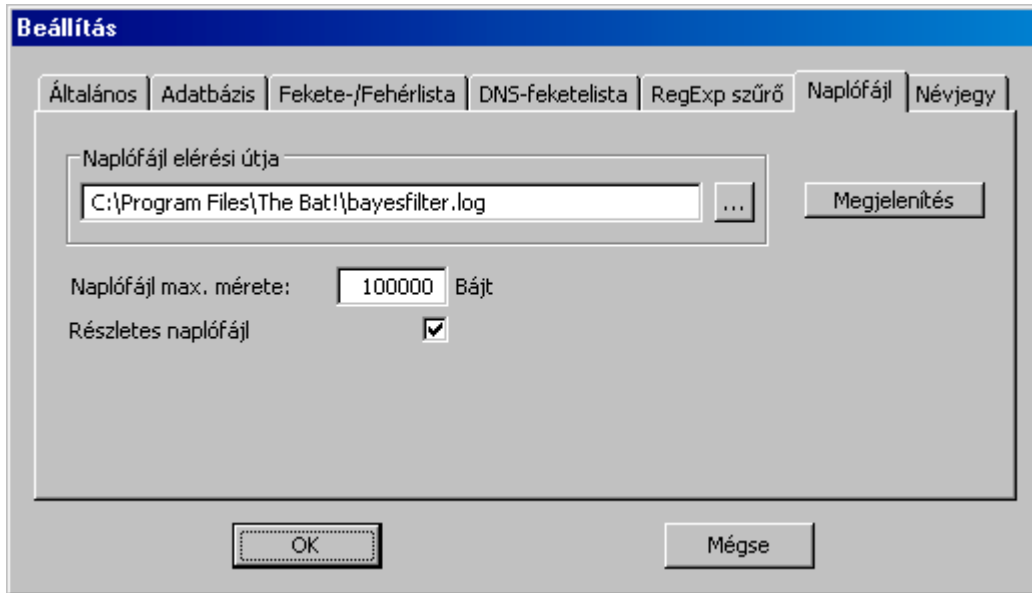


**Reguláris kifejezés szűrőfájl:** A reguláris kifejezéseket tartalmazó szűrőfájl elérhetősége.

**A "Reguláris kifejezés" szűrő aktív:** Aktiválja a reg. exp. szűrőt.

**Spam küszöbérték:** A reguláris szűrő által adott pontértéknek ennyinek kell lennie legalább a spamnak minősítéshez. Ne állítsa át az alapértelmezett értéket.

## Naplófájl panel



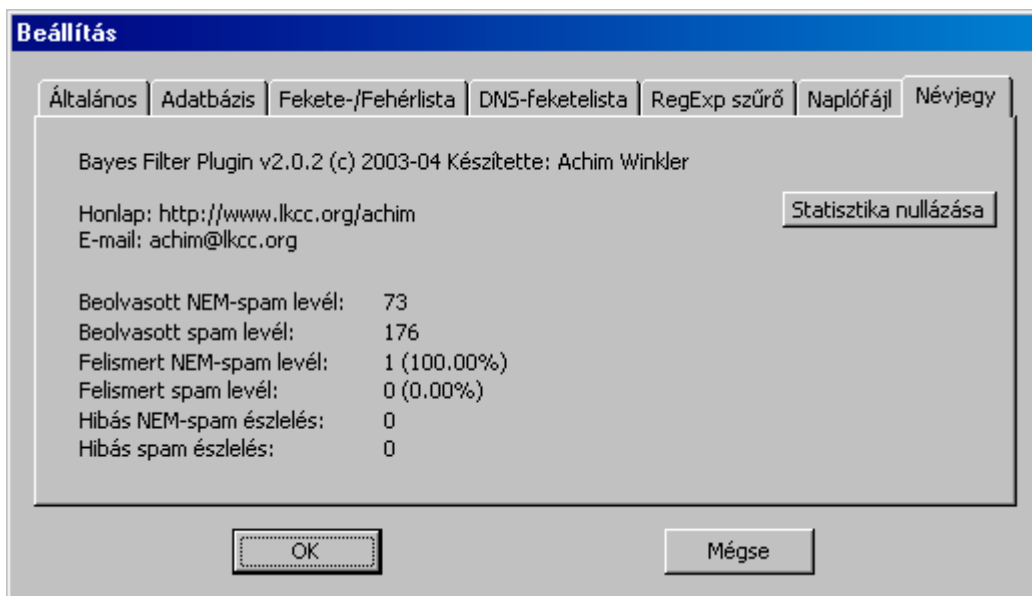
**Naplófájl elérési útja:** Ebben a fájlban naplózza a program az elvégzett műveleteket.

**Naplófájl max. mérete:** Maximális méretet szab a naplófájlnak.

**Részletes naplófájl:** Általában csak az átvizsgált levelek pontozása kerül a naplófájlba, ezzel kicsit több infót is belerak a program a naplófájlba. Példa:  
images | 1 | 153 | 3627 | 0.07313618 | 0.17873832

- 1.oszlop: a szó;
- 2.oszlop: a szó előfordulása a levélben
- 3.oszlop: a szó előfordulása minden ellenőrzött nem-spam levélben
- 4.oszlop: a szó előfordulása minden ellenőrzött spam levélben
- 5.oszlop: a számolt spamvalószínűsége a szónak
- 6.oszlop: a nem-spam/spam arány

## Névjegy panel



**Statisztika nullázása:** A nem-spam/spam statisztika nullázása, azaz a program minden tanult beállítást elfelejt.

A következőkben a The Bat! **Anti-Spam** részében eszközölhető beállítási lehetőségeket találja.

**Törölje a levelet, ha a pontszám nagyobb mint X:** ha a spamra kapott, számított pontszám ennél nagyobb, akkor törli a The Bat! a levelet (nem is rakja egyik mappába sem letöltés és spamellenőrzés után). Nem javaslom, hogy ezt bejelölje!

**Mozgassa a levelet a levélszemét mappába, ha a pontszám nagyobb mint X:** ha a spamra kapott, számított pontszám ennél nagyobb, a The Bat! letöltés után a Levélszemét mappába teszi a levelet. Bayes Filter esetén állítson ide 90-et

**Tárolt levélszemét megjelölése olvasottként:** olvasottnak, s nem újnak jelöli meg az adott spamoto a program. Én javaslom az olvasatlannak hagyást, s manuálisan átállítani, illetve letörölni a spamoto.

**Szemétként megjelölt levelek áthelyezése a levélszemét mappába:** amikor tanítja a programot, hogy melyik levél levélszemét, melyik nem, s ezt bejelöli, akkor a levélszemétként megjelöltek a Levélszemét mappába kerülnek

**Közös levélszemét-mappa használata:** minden postafiók levélszemete egy Levélszemét nevű közös mappába kerül

Ezzel a spamszűrés beállítása megtörtént. Én azt javaslom, hogy az utolsó 4 dolgot jelölje be, egyelőre NE töröltesse a leveleket. Majd amikor már kitapasztalta, hogy a szűrő jól, s megbízhatóan működik. Személyesen tapasztaltam néhány cégnél, hogy írtak nekem, megválasoltam azt, majd pár nap után reklamáltak, hogy miért nem válaszolok. Rosszul volt beállítva a spamszűrőjük, az én válaszaimat is letörölték vele. Számos apró állítási lehetőség van még, ezekkel itt nem foglalkozunk. A telepítéskor beállított értékeket nem kell változtatni, a leglényegesebb dolog a szűrő betanítása!

### **Hogyan kell betanítani a spamszűrőket?**

Pár általános tanács:

telepítés előtt érdemes egy mappában egy ideig gyűjteni a spamleveleket, hogy telepítés után legyen elegendő spamlevél a szűrő tanításához telepítés után NE töltsön le azonnal leveleket, először tanítsa a szűrőt  
És akkor a tanítás, mely nagyon egyszerű: válasszon ki NEM spam leveleket (egy mappa összes levelét a Ctrl+A paranccsal tudja kiválasztani), jobb egérgomb, **Speciális funkciók/Megjelölés NEM levélszemétként**. Ugyanígy válasszon ki spam leveleket, jobb egérgomb, **Speciális funkciók/Megjelölés levélszemétként**. Ennyi. Egy dolgot érdemes még szem előtt tartani: a spamszűrők statisztikai elven működnek, ami azt jelenti, hogy minél több levelet ad meg nekik, annál nagyobb biztonsággal találják el, hogy egy levél spam vagy sem. Személyes tapasztalat, hogy 500-500 spam és nem spam levél szükséges a megbízható működéshez.

### **A beépített spamszűrő letörölt olyan levelet, amit nem kellett volna!!!**

A gép az gép, azt csinálja, amit mondanak neki. Ha a beállítások jók, akkor nem tanította elég ideig, hogy biztosan csak azt törölje a program, amit kell. De úgy általánosságban is azt tudom mondani, hogy érdemes a szemetet egy külön mappába gyűjteni, s időnként egy perc alatt átnézni őket, majd törölni manuálisan.