UploadedCorporaIndependentMassCheck

Using uploaded corpora with an independent mass-check

The NewUploadedCorporaUser page describes setting up a ruleQA user so that an uploaded corpus will be mass-checked using the mass-check client /server setup. However, a bug means that spamassassin2.zones.apache.org doesn't support C/S mode, for some reason, so instead to use that resource, some of the uploaded corpora are scanned separately in traditional single-machine non-distributed mode. Here are the commands used to set up a new uid on that machine, for PMC members.

First, log into spamassassin2.zones.apache.org. (You'll probably need to have an account created for you first.)

set some variables:

```
BBUSERNAME=bb-jm
```

create a uid:

```
sudo useradd -c "Nightly mass-check jm" $BBUSERNAME
sudo passwd $BBUSERNAME
[give the new account a random password. this is needed for cron to work!]
sudo mkdir -p /export/home/$BBUSERNAME
sudo chown $BBUSERNAME /export/home/$BBUSERNAME
sudo -H -u $BBUSERNAME bash
```

you are now running as the new uid. Follow instructions similar to http://wiki.apache.org/spamassassin/NightlyMassCheck :

```
cd $HOME
mkdir tmp
echo 'sa-nightlymc-user@jmason.org' > .forward
svn co http://svn.apache.org/repos/asf/spamassassin/trunk svn
```

Accept (p)ermanently when asked.

```
cp trunk/masses/rule-qa/corpus.example ~/.corpus
vi ~/.corpus
```

use something like this:

```
opts_weekly="--net -j 8 --reuse --cache --cachedir=/tmp/aicache_nightly --restart=500 ham:detect:/export/h
ome/bbmass/uploadedcorpora/jm/ham/* --after="-15552000" --tail=40000 --scanprob=0.3 spam:detect:/export/ho
me/bbmass/uploadedcorpora/jm/spam/*"
opts_nightly="--reuse --cache --cachedir=/tmp/aicache_nightly --restart=500 ham:detect:/export/home/bbmass
/uploadedcorpora/jm/ham/* --after="-15552000" --tail=40000 --scanprob=0.3 spam:detect:/export/home/bbmass/
uploadedcorpora/jm/spam/*"
tmp=$HOME/tmp
tree=$HOME/svn
prefs_weekly=$HOME/user_prefs.weekly
prefs_nightly=$HOME/user_prefs.nightly
username=__BBUSERNAME__
password=__RSYNC_PASSWORD__
```

Replace BBUSERNAME with the value of $BBUSERNAME, and RSYNC_PASSWORD with the correct pwd for that rsync user.

Then, run the mass-check just to see if it works (feel free to CTRL-C once you're happy):

```
bash $HOME/svn/masses/rule-qa/corpus-nightly
```

Then set up the cron using 'EDITOR=vi crontab -e':

```
0 9 * * * bash svn/masses/rule-qa/corpus-nightly
```

Hopefully that should do it 😉