

# RescoreSet23Details

## These Rescoring Runs Have Finished

This is an old page, left for reference only.

## Rescore Mass-checks for Set 2 and Set 3

The bayes+nonet and bayes+net mass-check runs for 3.0.0 have started! Here's the procedure you'll need to follow, if you wish to submit logs for the rescoring run:

First, send mail to <submit.at.spamassassin.org>, and ask for a log-submission account if you haven't already got one.

It's helpful, but not required, to have some or all of the helper applications installed:

- the Mail::SPF::Query module
- the Net::DNS module
- Razor
- DCC
- Pyzor

If you're running nightly mass-checks, please feel free to disable them when running the rescore mass-check runs. Also, please note that the nightly submission accounts will work for rescore submissions as well.

Then run these commands:

```
wget http://old.SpamAssassin.org/released/Mail-SpamAssassin-3.0.0-pre3.tar.gz
tar xvfz Mail-SpamAssassin-3.0.0-pre3.tar.gz
cd Mail-SpamAssassin-3.0.0
perl Makefile.PL < /dev/null; make

cd masses
rm -rf spamassassin; mkdir spamassassin
echo "use_bayes 1" > spamassassin/user_prefs
echo "use_auto_whitelist 0" >> spamassassin/user_prefs
rm ham.log spam.log

./mass-check --bayes --net -j 4 --restart=400 --after=1041397200 --all <targets>
```

<targets> is the list of directories, mboxs, etc., like  
spam:dir:~/Mail/spam. See the comments at the top of "mass-check" for details.

This takes a long time to run. Due to Bayes DB lock contention, you will not want to create too many processes running concurrently. -j 2 controls the number of processes to use; 2 should be OK for a single-processor machine, since most of the time there will be one processing while the other is writing to the DB. -j 4 may be good depending on network response speed. Also, if your Bayes DB isn't on an NFS filesystem, you will want to add lock\_method flock to the user\_prefs file so [SpamAssassin](#) can use the more efficient flock locking method.

The --after=1041397200 option tells mass-check to ignore messages older than 18 months ago (in this case January 1 2003). This is useful if your corpus has older messages intermingled with your newer messages.

If you have an unusual network layout, you may need to specify trusted\_networks and/or internal\_networks in the spamassassin/user\_prefs file. But SA should be able to infer it in most cases. If you get less than a 10% or 15% spam hit rate for RCVD\_IN\_XBL, then you might need to use these configuration parameters.

Once it finishes:

```
USER="[whatever your username is]"
RSYNC_PASSWORD="[whatever your password is]"
export RSYNC_PASSWORD

rsync -CPcvuzb ham.log $USER@rsync.spamassassin.org::submit/ham-bayes-net-$USER.log
rsync -CPcvuzb spam.log $USER@rsync.spamassassin.org::submit/spam-bayes-net-$USER.log
```

Next, redo *without* --net:

```
cd masses
rm -rf spamassassin; mkdir spamassassin
echo "use_bayes 1" > spamassassin/user_prefs
echo "use_auto_whitelist 0" >> spamassassin/user_prefs
rm ham.log spam.log

./mass-check --bayes -j 2 --restart=400 --after=1041397200 --all <targets>
```

See the above notes for other options that may be useful.

Once it finishes:

```
USER="[whatever your username is]"
RSYNC_PASSWORD="[whatever your password is]"
export RSYNC_PASSWORD

rsync -CPcvzb ham.log $USER@rsync.spamassassin.org::submit/ham-bayes-nonet-$USER.log
rsync -CPcvzb spam.log $USER@rsync.spamassassin.org::submit/spam-bayes-nonet-$USER.log
```

That's it!

The results for these two runs will need to be in by Wednesday July 28th, 2004.