

# RescoreDetails310

## Rescore Mass-Check Instructions

(These are the instructions for the now completed re-run of 3.1.0 mass-checks; see [RescoreMassCheck](#) for the overview of the general process in toto. This page left as-is for the next time we have to do it!)

Here's the procedure you'll need to follow, if you wish to submit data for the rescoring run for 3.1.0 using [MassCheck](#):

Clean up the corpus of mail you intend to [MassCheck](#) (see [CorpusCleaning](#)), and get an rsync account (see [RsyncAccounts](#)). The latter can be done while mass-check is running, btw, it's not needed until the end; and the 'checking for false positives and false negatives' stage of corpus cleaning can be done afterwards as well.

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
- 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.

Note that it's essential that you mass-check *both* ham and spam for this run, as otherwise the Bayes rules will be affected.

Then run these commands:

```
wget http://people.apache.org/~jm/devel/Mail-SpamAssassin-3.1.0-pre4.tar.gz
tar xvfz Mail-SpamAssassin-3.1.0-pre4.tar.gz
cd Mail-SpamAssassin-3.1.0
perl Makefile.PL < /dev/null
make

cd masses
mkdir spamassassin
rm -f spamassassin/*
echo "bayes_auto_learn 0" > spamassassin/user_prefs
echo "lock_method flock" >> spamassassin/user_prefs
echo "bayes_store_module Mail::SpamAssassin::BayesStore::SDBM" >> spamassassin/user_prefs
echo "use_auto_whitelist 0" >> spamassassin/user_prefs

nohup ./mass-check --progress --bayes --net -j 4 --restart=400 --learn=35 --reuse \
--after=1072933200 <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.

Do not use --reuse if you have scanned with SA, but have configured that scanner to run with -L, or you have disabled common network tests or SPF. This is because it relies on the presence of the X-Spam-Status line to pick up hits on those rules, and currently cannot detect those conditions.

This takes \*ages\* to run. -j 4 controls the number of processes to use; 4 should be OK for a single-processor machine, since most of the time they'll be waiting for network results to arrive. If you have adequate RAM and don't mind the load, you can use -j 6 or -j 8. There's not much benefit in going higher than -j 8.

The --after=1072933200 option tells mass-check to ignore messages older than 18 months ago (in this case January 1 2004). 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. A good way to tell is if you see no SPF\_PASS results – SPF will not be used if the message passes through one or more trusted relays.

Once it finishes, check that the results are sane. See [CorpusCleaning](#) to remove any result lines that deal with misclassified or corrupt messages.

Then submit your results!

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

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

*(note: previously, we used -C on those rsync commands. it should be removed as the current host seems to be running a version of rsync that cannot handle that, giving this error: 'filter rules are too modern for remote rsync. rsync error: syntax or usage error (code 1) at exclude.c(1119)'.)*

That's it!

The results for this run will need to be in by Friday July 22nd (tentatively). If you're still running then, submit what you have so far and beg for more time. We may be pushing it out a little further anyway depending on how things go 🙄