



APACHECON

ASIA 2022

New features in
SpamAssassin 4.0

Giovanni Bechis <gbechis@apache.org>

2022.07.30

SpamAssassin 4.0 dev story

- 4.0 branch started on 2015 as a copy of 3.4.1 version
- 4.0 branch has started primarily to improve utf-8 parser
- While developing on 4.0 branch 5 minor versions have been released
- Most recent SpamAssassin released version is 3.4.6 released on 04/12/2021

UTF8 support



- With UTF8 and UTF16 support
SpamAssassin can easily block spam in different languages

UTF8 support

Match a " " on the email Subject

```
header EMOJI_CAL Subject =~ /\xf0\x9f\x93\x86/i
```

Bayes improvements

Naive Bayes

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

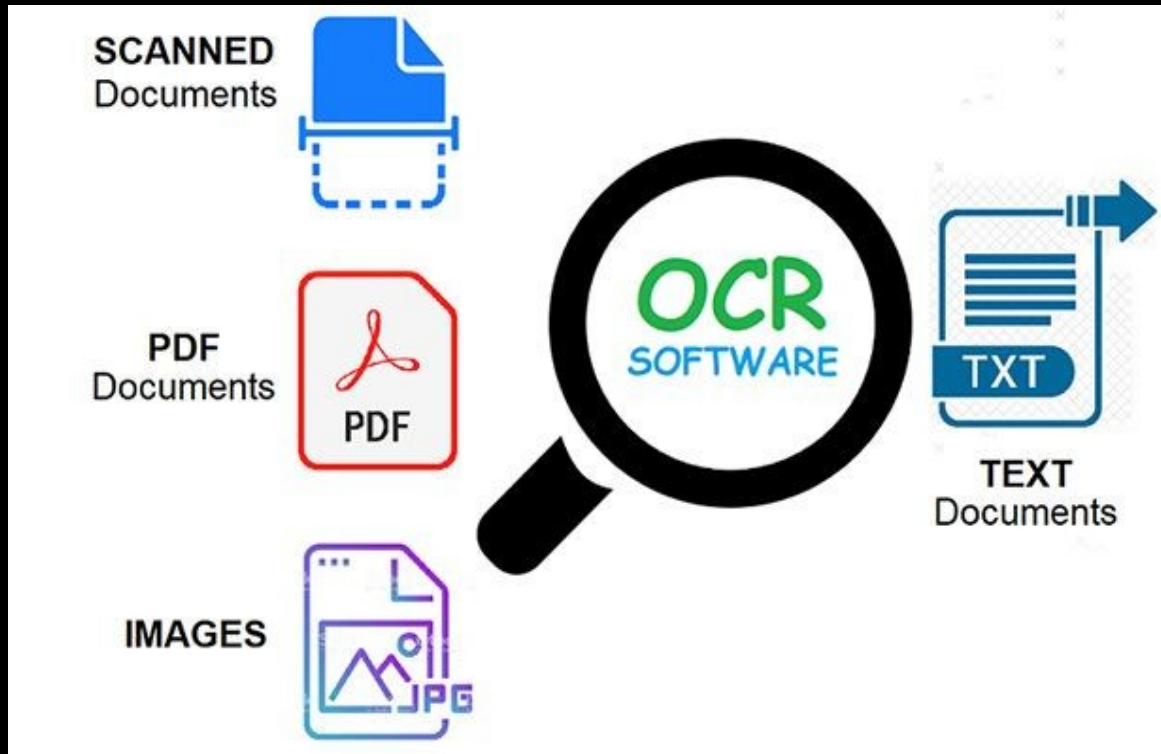
- It's now possible to skip common words used in 15 different languages to improve Bayes algorithm on non-english emails

Bayes improvements

Danish stopwords

```
bayes_stopword_dk (?^:(?:a(?:lle|nden)|bl(?:ev|  
iver?))|d(?:e(?:nne|res|tte)|isse)|e(?:fter|ller)|  
h(?:a(?:ns|v(?:de|e))|endes?|v(?:ad|is|or))|ikke|  
kunne|m(?:ange|eget|ine)|nog(?:et|le)|o(?:gs\xc3\x  
a5|ver)|s(?:elv|ine|k(?:al|ulle)|\xc3\xa5dan)|  
under|v(?:ille|\xc3\xa6ret?)))
```

ExtractText plugin



- ExtractText.pm is a plugin that extract text from attachments using external softwares

ExtractText plugin

```
extracttext_external tesseract {OMP_THREAD_LIMIT=1}  
/usr/local/bin/tesseract -c page_separator= {} -  
extracttext_use tesseract .bmp .jpg .png .tif
```

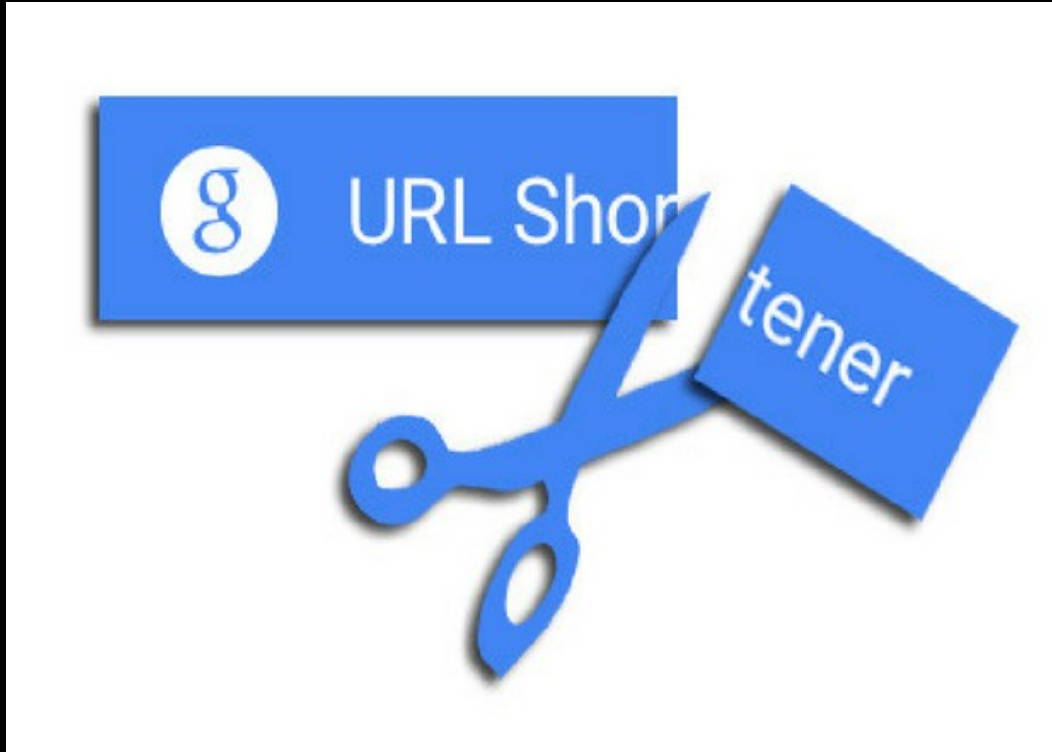
```
extracttext_external xls2txt  
/usr/local/scripts/xls2txt.pl -u -i {}  
extracttext_use xls2txt .xls  
application/(? :vnd\.?)?ms-excel.*
```

OLEVBMacro plugin



- OLEVBMacro plugin can now extract uris from Office documents metadata

Short urls



- `DecodeshortURLs` is a plugin that checks the url shorteners target

Short urls

```
Ifplugin Mail::SpamAssassin::Plugin::DecodeShortURLs
```

```
  url_shortener .short.gy
```

```
  url_shortener bit.ly
```

```
  body HAS_SHORT_URL
```

```
  eval:short_url()
```

```
  score HAS_SHORT_URL
```

```
  0.01
```

```
endif
```

Captured tags

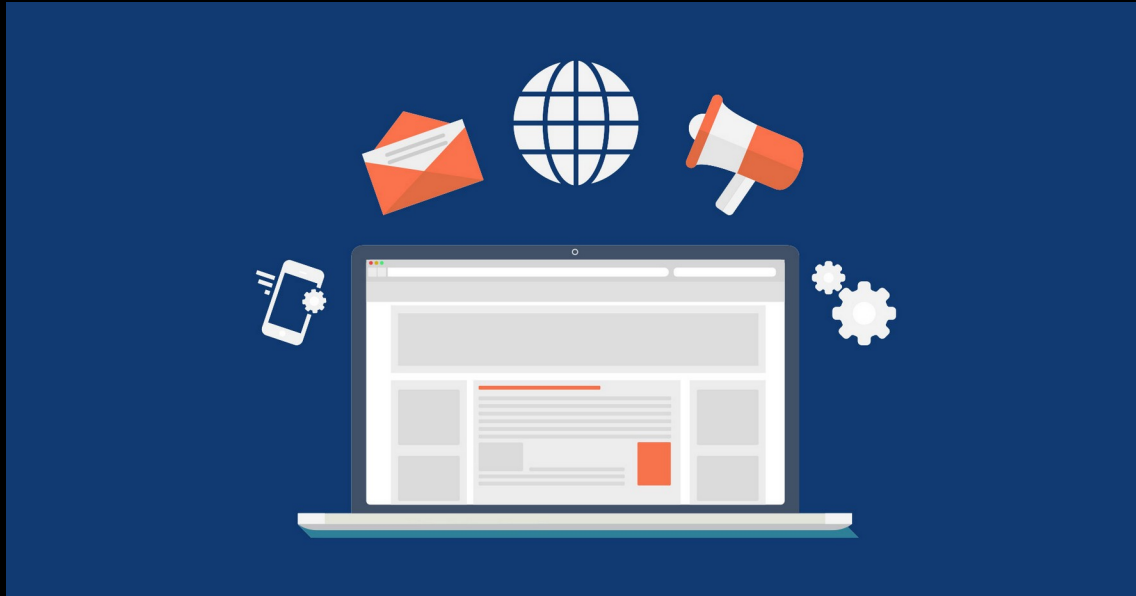


- Captured tags can be used in other rules as a matching string

Captured tags

```
if can(Mail::SpamAssassin::Conf::feature_capture_rules)
  header __FROM_NAME From:name =~ /( ?<FROM_NAME>\w+)/
  header FROM_NAME_SUBJ Subject =~ /\b%{FROM_NAME}\b/i
endif
```

sa-update(1) improvements

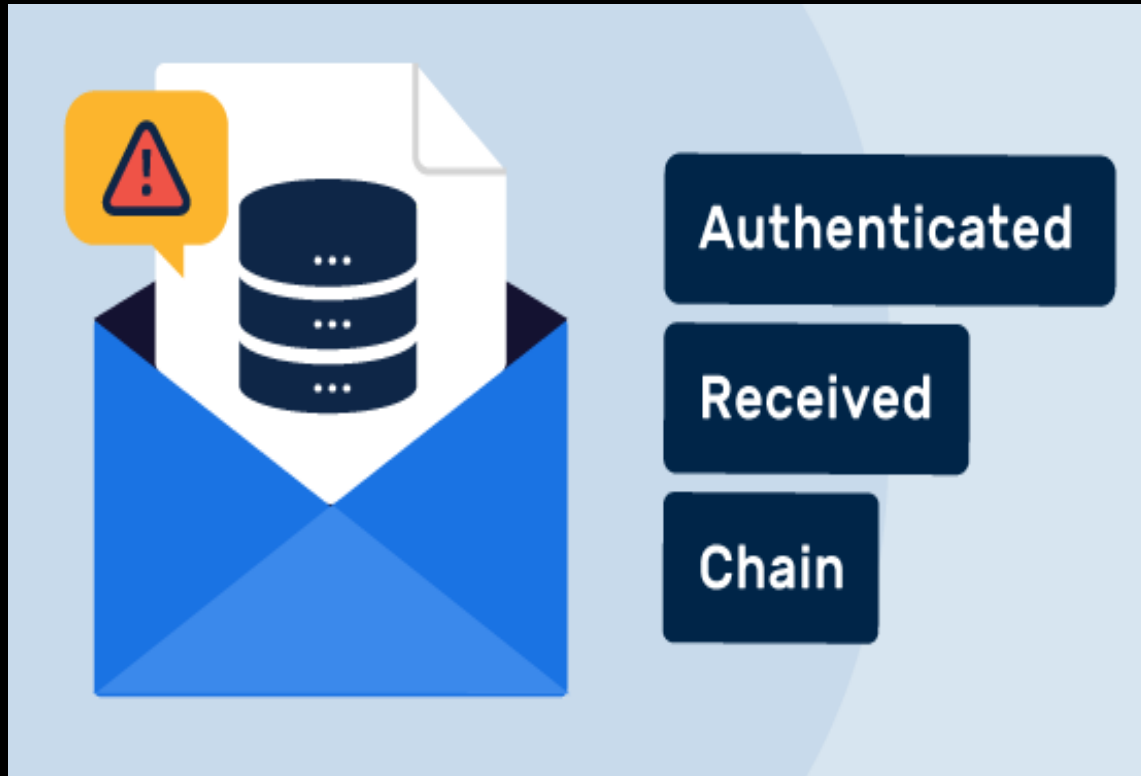


- sa-update(1) has been improved with new options:
 - forcemirror
 - score-multiplier
 - score-limit

sa-update(1) improvements

- `--forcemirror`
forces sa-update to use a specific mirror server
- `--score-multiplier`
adjust all scores from update channel, multiply with given value
- `--score-limit`
adjust all scores from update channel, limit to given value

ARC Signatures



- ARC signatures are used to validate an email when the email's SPF and DKIM records are rendered invalid by an intermediate server

DMARC policies



- DMARC policies give email domain owners the ability to protect their domain from unauthorized use

DMARC policies

```
ifplugin Mail::SpamAssassin::Plugin::Dmarc
# dmarc_save_reports 1
header DMARC_PASS eval:check_dmarc_pass()
describe DMARC_PASS DMARC pass policy
priority DMARC_PASS 500
tflags DMARC_PASS net nice
reuse DMARC_PASS
endif
```

Inclusion changes

Some keywords have been changed by using more inclusive words.

- whitelist -> welcomelist
- blacklist -> blocklist

Old keywords will continue to work until SpamAssassin 4.1 will be released



Thanks

 gbechis@apache.org

 [giovannibechis](#)

 [@_gbechis](#)

APACHECON ASIA 2022