Finding email with Strigi in .tar backups

Published June 3, 2007

Citation

Willighagen, E. (2007, June 3). Finding email with Strigi in .tar backups. *Chem-bla-ics*. https://doi.org/10.59350/t125s-b8827

Keywords

Kde

Abstract

Now that my CUBIC desktop machine is shutting down, I made the necessary backups, among a mail.tar for my mail correspondence of about a year. About 500MB in size for almost 8700 files. Strigi is a perfect tool to help me find messages in this archive, as it will recurse into the .tar archive, and even into email attachements.

Copyright

chem-bla-ics

Now that [my CUBIC desktop machine is shutting down](), I made the necessary backups, among a mail.tar for my mail correspondence of about a year. About 500MB in size for almost 8700 files. [Strigi]() is a perfect tool to help me find messages in this archive, as it will recurse into the .tar archive, and even into email attachements. I created an index just for the archive with:

```
strigicmd create -t clucene -d index/ mail.tar
```

It took Strigi about 30 seconds to index the whole archive. That's good performance!

Now, Strigi indexes content full text, but also uses a controlled vocabulary (among which [one specifically for chemistry]()). So I can search for email messages which have article in the subject with:

```
strigicmd query -t clucene -d index/ email.subject:article
```

However, `From:` and `To:` content was not yet extracted. That was easily patched. This allows me to find correspondence between me and, for example, Christoph:

```
strigicmd query -t clucene -d index/ email.to:Christoph AND email.from:Egon
```