MailEntityProcessor

<!> Solr1.4

Where it is used

This is used for indexing mails from a mail box. Currently IMAP protocol is supported. Since Java Mail API is used, it should be able to support other protocols as well in future.

Fields emitted by MailEntityProcessor

Each mail gets indexed as one document. The MailEntityProcessor emits the following fields for each mail. The consumer is free to consume fields of interest, transform etc.

single valued fields :

- messageId
- subject
- from
- sentDate
- xMailer

multi valued fields

- allTo
- flags : possible flags are 'answered', 'deleted', 'draft', 'flagged' , 'recent', 'seen'
- content
- attachment
- attachmentNames;

How to configure

The data-config.xml should have the below configuration at a minimum.

```
<document>
   <entity processor="MailEntityProcessor"
           user="somebody@gmail.com"
           password="something"
           host="imap.gmail.com"
           protocol="imaps"
           folders = "x,y,z"/>
</document>
```

Additional configuration

The below attributes help fine tune the indexing. These are all optional.

folders to index

- **recurse** : true|false. Index the sub folders recursively.
- **include** : comma separated list of regex patterns to include
- **exclude** : comma separated list of regex patterns to exclude

Timeouts

- **connectTimeout** : maximum time to wait in milliseconds while connecting. Default is 30 seconds.
- **readTimeout** : maximum time to wait in milliseconds while fetching mails from mail server. Default is 60 seconds.

handling attachments

- **processAttachement** : true|false. If true, the attachments are also indexed. They can be retrieved or searched using "attachment" and ""attachmentNames" multi valued fields in the indexed documents.

The MailEntityProcessor uses Apache Tika.

fetching mails since a certain date

- **fetchMailsSince** : This should be in the format "yyyy-MM-dd HH:mm:ss"

custom filter for fetching mails

- **customFilter** : Should implement the below interface, where SearchTerm and Folder are from Java Mail API.

```
public static interface MailEntityProcessor.CustomFilter {
  public SearchTerm getCustomSearch(javax.mail.Folder folder);
}
```

Other attributes for tuning mail fetching from mail server

- **batchSize** : How many mails to fetch at a time. Default is 20.
- **fetchSize** : Sets the "mail.imap.fetchsize" property. Default is 32KB.