# Structure of file

All data had been anonymized (recipient name, domain, terms-words of subject – stopwords were removed)

## Accounts used for downloading emails

Company employees, one of them use more mail accounts (addresses)

* **Senior software developer, software** **architect (man, 35+):** name00001, name00002, name00003, name00004, name00005
* **Software developer (woman, 35+):** name00006
* **Software developer (man, 35+):** name00007
* **Team leader, software architect (man, 50+):** name00008
* **CEO, business person (man, 50+):** name00009
* **Software developer, technology specialist (man, 30+):** name00010

<Accounts>

<Account smtpAddress="name00001@domain0001.com" />

<Account smtpAddress="name00002@domain0002.cz" />

<Account smtpAddress="name00003@domain0003.cz" />

<Account smtpAddress="name00004@domain0004.cz" />

<Account smtpAddress="name00005@domain0005.cz" />

<Account smtpAddress="name00006@domain0004.cz" />

<Account smtpAddress="name00007@domain0004.cz" />

<Account smtpAddress="name00008@domain0004.cz" />

<Account smtpAddress="name00009@domain0004.cz" />

<Account smtpAddress="name00010@domain0004.cz" />

</Accounts>

## Email

* **Sender**: email address (anonymized)
* **Sent**: data and time
* **MessageId**: unique ID (anonymized); can be missing
* **InReplyTo** (optional): MessageId of previous mail (replayed or forwarded); can be missing in the dataset (previous email could be deleted)
* **Subject** (optional): Set of terms (anonymized)
* **Recipient**: email addresses of email recipients (anonymized); type “To” or “Cc”

</Message><Message Sender="name00008@domain0004.cz" Sent="2010-01-13T17:00:48.000Z" MessageId="9fce8b1b-2db3-447a-b422-aba74d8e4264" InReplyTo="1f29f3b6-f77e-42a5-b272-f37adcf6b50d" Subject="term00053 term00054 term00055 ">

<Recipient Type="To">name00045@domain0028.cz</Recipient>

<Recipient Type="Cc">name00056@domain0002.cz</Recipient>