Consulting for digital humanists

The cultural shock developing tools and pedagogy
Centre for Digital Humanities
Uppsala
Department of ALM
Uppsala University
Marie Dubremetz

email: marie.dubremetz@abm.uu.se
mastodon: @mardub@piaille.fr
www.mariedubremetz.com

Engineer in Digital Humanities

Located here:

And on spared time:

LILLE
UPPSALA
UNIVERSITET

Raoull
Past digital humanists at FOSDEM

Past digital Humanists at FOSDEM

From pipe dreams and waste to functional accretion: building a capable infrastructure for the Digital Humanities. Thank you FOSDEM! Ian Johnson

Executable papers in the Humanities, or how did we land to the Journal of Digital History Daniele Guido

Using Advene to accompany research in AudioVisual Digital Humanities Olivier Aubert

Stylo : a user friendly text editor for humanities scholars Antoine Fauchié
What is an engineer in Digital Humanities?

someone technical that will have expertise in one or several domain of computer science and whose mission is to help humanists in their projects (i.e. philosophy, literature, history, sociology, archeology…)
Examples
Helping theologists showing the colors of the bible.
Building online databasis for art historians
<table>
<thead>
<tr>
<th>Main applicant</th>
<th>Department</th>
<th>Seniority</th>
<th>Gender</th>
<th>Title</th>
<th>Technical need</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agnieszka Backman</td>
<td>Scandinavian Languages</td>
<td>Postdoc</td>
<td>Female</td>
<td>Networked Accessibility and Discoverability of Digital Cultural Heritage Data</td>
<td>Web scraping</td>
</tr>
<tr>
<td>Tuva Haglund</td>
<td>Literature</td>
<td>Postdoc</td>
<td>Female</td>
<td>Among the Wattpad readers</td>
<td>Web scraping, text mining</td>
</tr>
<tr>
<td>Harald Hammarström</td>
<td>Linguistics and Philology</td>
<td>Associate professor</td>
<td>Male</td>
<td>Hosting Gramfinder</td>
<td>Virtual machine</td>
</tr>
<tr>
<td>Ashleigh Harris</td>
<td>English</td>
<td>Professor</td>
<td>Female</td>
<td>The Contemporary African Ephemeral Literatures Archive (CAELA)</td>
<td>Online portal with crowdsourcing capacity</td>
</tr>
<tr>
<td>Isto Huvila</td>
<td>ALM</td>
<td>Professor</td>
<td>Male</td>
<td>Automatic identification of paradata</td>
<td>NLP, object recognition</td>
</tr>
<tr>
<td>Beata Megyési</td>
<td>Linguistics and Philology</td>
<td>Professor</td>
<td>Female</td>
<td>Automatic decryption of historical manuscripts</td>
<td>Computational space and resources, GUI</td>
</tr>
<tr>
<td>Matti La Mela</td>
<td>FEK/ALM</td>
<td>Postdoc</td>
<td>Male</td>
<td>Text mining commodification</td>
<td>Classification of OCRd text</td>
</tr>
<tr>
<td>Michaela Vance</td>
<td>English</td>
<td>Postdoc</td>
<td>Female</td>
<td>Attribution study on France’s Brooke’s 18th Century Libretto Marian</td>
<td>OCR, stylometry</td>
</tr>
<tr>
<td>Main applicant</td>
<td>Department</td>
<td>Seniority</td>
<td>Gender</td>
<td>Title</td>
<td>Technical need</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------</td>
<td>-----------</td>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Agnieszka Backman</td>
<td>Scandinavian Languages</td>
<td>Postdoc</td>
<td>Female</td>
<td>Networked Accessibility and Discoverability of Digital Cultural Heritage Data</td>
<td>Web scraping</td>
</tr>
<tr>
<td>Tuva Haglund</td>
<td>Literature</td>
<td>Postdoc</td>
<td>Female</td>
<td>Among the Wattpad readers</td>
<td>Web scraping, text mining</td>
</tr>
<tr>
<td>Harald Hammarström</td>
<td>Linguistics and Philology</td>
<td>Associate professor</td>
<td>Male</td>
<td>Hosting Gramfinder</td>
<td>Virtual machine</td>
</tr>
<tr>
<td>Ashleigh Harnis</td>
<td>English</td>
<td>Professor</td>
<td>Female</td>
<td>The Contemporary African Ephemeral Literatures Archive (CAELA)</td>
<td>Online portal with crowdsourcing capacity</td>
</tr>
<tr>
<td>Isto Huvila</td>
<td>ALM</td>
<td>Professor</td>
<td>Male</td>
<td>Automatic identification of paradata</td>
<td>NLP, object recognition</td>
</tr>
<tr>
<td>Beata Megyési</td>
<td>Linguistics and Philology</td>
<td>Professor</td>
<td>Female</td>
<td>Automatic decryption of historical manuscripts</td>
<td>Computational space and resources, GUI</td>
</tr>
<tr>
<td>Matti La Mela</td>
<td>FEK/ALM</td>
<td>Postdoc</td>
<td>Male</td>
<td>Text mining commodification</td>
<td>Classification of OCRd text</td>
</tr>
<tr>
<td>Michaela Vance</td>
<td>English</td>
<td>Postdoc</td>
<td>Female</td>
<td>Attribution study on France’s Brooke’s 18th Century Libretto Marian</td>
<td>OCR, stylometry</td>
</tr>
<tr>
<td>Main applicant</td>
<td>Department</td>
<td>Seniority</td>
<td>Gender</td>
<td>Title</td>
<td>Technical need</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------</td>
<td>-----------</td>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>Agnieszka Backman</td>
<td>Scandinavian Languages</td>
<td>Postdoc</td>
<td>Female</td>
<td>Networked Accessibility and Discoverability of Digital Cultural Heritage Data</td>
<td>Web scraping</td>
</tr>
<tr>
<td>Tuva Haglund</td>
<td>Literature</td>
<td>Postdoc</td>
<td>Female</td>
<td>Among the Wattpad readers</td>
<td>Web scraping, text mining</td>
</tr>
<tr>
<td>Harald Hammarström</td>
<td>Linguistics and Philology</td>
<td>Associate professor</td>
<td>Male</td>
<td>Hosting Gramfinder</td>
<td>Virtual machine</td>
</tr>
<tr>
<td>Ashleigh Harris</td>
<td>English</td>
<td>Professor</td>
<td>Female</td>
<td>The Contemporary African Ephemeral Literatures Archive (CAELA)</td>
<td>Online portal with crowdsourcing capacity</td>
</tr>
<tr>
<td>Isto Huvila</td>
<td>ALM</td>
<td>Professor</td>
<td>Male</td>
<td>Automatic identification of paradata</td>
<td>NLP, object recognition</td>
</tr>
<tr>
<td>Beata Megyési</td>
<td>Linguistics and Philology</td>
<td>Professor</td>
<td>Female</td>
<td>Automatic decryption of historical manuscripts</td>
<td>Computational space and resources, GUI</td>
</tr>
<tr>
<td>Matti La Mela</td>
<td>FEK/ALM</td>
<td>Postdoc</td>
<td>Male</td>
<td>Text mining commodification</td>
<td>Classification of OCRd text</td>
</tr>
<tr>
<td>Michaela Vance</td>
<td>English</td>
<td>Postdoc</td>
<td>Female</td>
<td>Attribution study on France’s Brooke’s 18th Century Libretto Marian</td>
<td>OCR, stylometry</td>
</tr>
<tr>
<td>Main applicant</td>
<td>Department</td>
<td>Seniority</td>
<td>Gender</td>
<td>Title</td>
<td>Technical need</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------</td>
<td>------------</td>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Agnieszka Backman</td>
<td>Scandinavian Languages</td>
<td>Postdoc</td>
<td>Female</td>
<td>Networked Accessibility and Discoverability of Digital Cultural Heritage Data</td>
<td>Web scraping</td>
</tr>
<tr>
<td>Tuva Haglund</td>
<td>Literature</td>
<td>Postdoc</td>
<td>Female</td>
<td>Among the Wattpad readers</td>
<td>Web scraping, text mining</td>
</tr>
<tr>
<td>Harald Hammarström</td>
<td>Linguistics and Philology</td>
<td>Associate professor</td>
<td>Male</td>
<td>Hosting Gramfinder</td>
<td>Virtual machine</td>
</tr>
<tr>
<td>Ashleigh Harris</td>
<td>English</td>
<td>Professor</td>
<td>Female</td>
<td>The Contemporary African Ephemeral Literatures Archive (CAELA)</td>
<td>Online portal with crowdsourcing capacity</td>
</tr>
<tr>
<td>Isto Huvila</td>
<td>ALM</td>
<td>Professor</td>
<td>Male</td>
<td>Automatic identification of paradata</td>
<td>NLP, object recognition</td>
</tr>
<tr>
<td>Beata Megyési</td>
<td>Linguistics and Philology</td>
<td>Professor</td>
<td>Female</td>
<td>Automatic decryption of historical manuscripts</td>
<td>Computational space and resources, GUI</td>
</tr>
<tr>
<td>Matti La Mela</td>
<td>FEK/ALM</td>
<td>Postdoc</td>
<td>Male</td>
<td>Text mining commodification</td>
<td>Classification of OCRd text</td>
</tr>
<tr>
<td>Michaela Vance</td>
<td>English</td>
<td>Postdoc</td>
<td>Female</td>
<td>Attribution study on France’s Brooke’s 18th Century Libretto Marian</td>
<td>OCR, stylometry</td>
</tr>
</tbody>
</table>
CS, Engineers  VS  Humanists

Male audience
Programming is the norm (Linux/Mac, latex, terminal app…)
Industry influence (salary, practices…)
Broad, commercial topics…
Metrics oriented (quantified performance, speed)

Female/mixed audience
Point and click and often proprietary is the norm (Windows, Mac, pptx, word…)
Academic job culture
Scarce data, niche topics.
Question oriented
How can we help each others?
Digital humanists and free software

- Lack of visibility
- Lack of cohesion
- Lack of common references (but not of commun values!)
- Some Free softwares…. but also some Freewares
Digital humanists and free software

- Receptive audience: Enlightenment was studied in the humanities before Stallman.
- Lot of activism
- Nascent varied motivated community
- Financial means (grants)
- Diverse different from our usual dev crowd
- Know communication, science of education, ethics, think critical.
Questions?

- Ideas to reach this community?
- How to make them feel welcome?
THANK YOU!

The slides after this are kept for illustration purpose for the question session.
References

CDHU Website: https://www.abm.uu.se/cdhu-eng

Marie Dubreemetz’s website www.mariedubreemetz.com

Philosophy:

Ethique du Libre: une lecture philosophique par Véronique Bonnet

https://videos.ubuntu-paris.org/videos/watch/1c850141-7958-4e84-a49a-f665401f67f6

Pedagogy

Introduction to Cultural Analytics & Python — Introduction to Cultural Analytics & Python (melaniewalsh.github.io)

https://mardub.gitlab.io/song/
Marie Dubremetz

Located here:

And on spared time:

email: marie.dubremetz@abm.uu.se
mastodon: mardub@piaille.fr
Asking the right question as engineer

Context: what is your domain? What does this domain focus on?
What is the problem you try to solve?
(eventually) What did you already try to solve your problem? What tools did you use?
What are the data you have (or hope to collect?)
Can you present me your data? Size, format, how has it been collected?
Dante Running from the Three Beasts

Person on red dress protecting the angel. At Commedia...
Digital humanists and free software

- Lack of visibility
- Lack of cohesion
- Lack of common references (but not of commun values!)
- Some Free softwares…. but also some Freewares

- Receptive audience: Enlightenment existed in the humanities before Stallman.
- Lot of activism
- Nascent varied motivated community
- Financial means (grants)
- Know communication, science of education, ethics…
Examples of projects

- Webscraping of old Swedish manuscripts in national libraries
- Researcher studying 10 ebooks.
Examples of projects

- Webscraping of old manuscripts
- Researcher studying 10 ebooks.
Examples of projects

- Webscraping of old manuscripts in national libraries
- Researcher studying 10 ebooks.

Both freeware, neither free software!
Did you say *Engineer in Humanities*!?
Diverse source of data and formats
Tools

some OS and Free softwares….  

but also some Freewares

spaCy

Python Beautiful Soup

GitLab

GitHub