Assignment 1
L715/B659
Due Monday, September 8

Your first assignment is going to be to browse the literature on a topic and find papers that you’re interested in reading. I have provided a rough spreadsheet of possible readings on oncourse\(^1\), and here are some other pointers for tracking down readings:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Some starting points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authorship attribution (Stylometry)</td>
<td><a href="http://pan.webis.de">http://pan.webis.de</a></td>
</tr>
<tr>
<td>Deception &amp; plagiarism detection</td>
<td><a href="http://pan.webis.de">http://pan.webis.de</a></td>
</tr>
<tr>
<td>Author profiling / social media</td>
<td><a href="http://pan.webis.de">http://pan.webis.de</a>, <a href="https://sites.google.com/site/wcprst/home/wcpr14">https://sites.google.com/site/wcprst/home/wcpr14</a></td>
</tr>
<tr>
<td>Dialect identification</td>
<td><a href="http://corporavm.uni-koeln.de/vardial/sharedtask.html">http://corporavm.uni-koeln.de/vardial/sharedtask.html</a></td>
</tr>
<tr>
<td>Native language identification</td>
<td><a href="https://sites.google.com/site/nlisharedtask2013/home">https://sites.google.com/site/nlisharedtask2013/home</a></td>
</tr>
<tr>
<td>Language proficiency identification</td>
<td>[no single site I know of]</td>
</tr>
</tbody>
</table>

For example, you might look at teams competing in the PAN challenges and track back to their previous work, in addition to following up on some of their references. For any shared task, I highly recommend starting with the overview paper and working backwards, as they tend to provide the context of the topic as a whole, as well as recent methods being used.

From different perspectives, the following appear to be generally useful (as well as trusty internet searches):

- The ACL Anthology: http://anthology.aclweb.org
- The ACM Digital Library: http://dl.acm.org

More specific notes:

- I encourage you to chat with me and with each other about topics, papers, etc. over email, before/after class, in my office, and so forth.
- Note that, although I’d like to get some spread over topics, multiple people can cover the same general issue, seeing as how there are some very broad topics out there, with some readings focusing on more foundational topics and others describing into cutting-edge work. Working together on leading a topic discussion is also allowed.
- You won’t be exhaustively researching a topic (and you won’t understand everything you skim right now), so think about what it is that you find interesting: particular subtasks? incorporating linguistic features? gathering data? combining different classifiers? using unsupervised learning? etc.

This assignment is not about committing to a particular research topic, but it doesn’t hurt to be thinking about the types of projects you’ll eventually be interested in working on.\(^2\)

- On the 8th, I would like you to turn in an approximately two-page report on the topic of your choice, what you hope to discover in that topic, a list of (5–10?) readings that seem to be most relevant to that hope, and any other notes you deem worth reporting.
- On the 8th, we will then go about figuring out the exact order of discussions and scheduling meetings with me to talk about the topics.

---

\(^1\)The spreadsheet was compiled by Fan Huang, who deserves much thanks!

\(^2\)More on projects in probably a month or so.