



BA/SA/Group/Lab:

## Authorship Attribution Task

Imagine that you are a detective searching for evidence that will allow you to identify a dreaded criminal. Your only clue at this point is a short machine-typed piece of text left by the perpetrator. Can you find out who it was?

Indeed, in forensics the following problem is relevant: Given a set of *suspects*, each of which is associated with a (rather limited) set of known *original* texts, and an *unknown* text document, is it possible to attribute the true authorship of the unknown text to one of the suspects?

There has been a lot of research on this topic. We may start by looking at some of the known strategies and their attribution capabilities; however, original approaches are very welcome. To evaluate our implementation, we may ultimately participate in the PAN shared tasks (<https://pan.webis.de/tasks.html>) and compete against international scientists from this specific domain.



**Requirements:** Programming experience (especially in Python) might be an advantage. The topic is broad and can be tailored to your previous knowledge and interests. There will be weekly meetings with your supervisor to discuss progress and open questions. For a group project, there might be the option to do an intense hackathon weekend to start off.

**Interested? Please contact us for more details!**

### Contacts

- Roland Schmid: [roschmi@ethz.ch](mailto:roschmi@ethz.ch), ETZ G94