DATeCH 2017 – PoCoTo Workshop – Introduction

Florian Fink & Uwe Springmann

Centrum für Informations- und Sprachverarbeitung (CIS)
Ludwig-Maximilians-Universität München (LMU)

May 31, 2017

- 13:00 - 13:15: Introduction
- 13:15 - 14:00: PoCoTo
- 14:00 - 15:00: Hands on PoCoTo
- 15:00 - 15:30: Profiler
- 15:30 - 16:00: Coffee break
- 16:00 - 17:00: Hands on Profiler
- There will be time between the sessions for short coffee breaks!

This Workshop covers two main topics:

1. The PoCoTo PostCorrectionTool: A tool to do interactive postcorrection of OCR'ed (historical) documents.

2. The language profiler: A tool that enables error detection and error correction for historical documents.

- PoCoTo is a tool for the interactive postcorrection of OCR'ed documents.
- The language profiler is a error detection and correction system for historical documents.
- PoCoTo and the language profiler are independent from each other.
- PoCoTo can work together with the profiler to detect errors and error pattern and to use correction suggestions.

- The USB stick, that you received at the registration contains a data file `datech-2017-workshop-pocoto.zip` (DATeCH 2017 - Goettingen/Preconference-workshops/PoCoTo)
- There is a github repository, that contains a lot more data related to historical OCR and our developed tools (including slides of this and other workshops and the sources of the profiler and PoCoTo): https://github.com/cisocrgroup

The data file contains data to be used in this workshop:

- The PoCoTo-manual
- The profiler-manual
- An OCR-project of *Die Grenzboten, 1844* (ABBYY)
- An OCR-project of *Heiligenleben, 1488* (Ocropus)
- A binary version of PoCoTo

Thanks for your attention!