CliDEID 1.0
Installation and Usage Guide

CliDEID is a command line tool for the de-identification of clinical narratives. It was developed as part of the CEGS N-GRID 2016 text mining challenge and therefore has been trained and optimised on the challenge data.

For example, for input:

Mr. Smith was admitted to the ABCD hospital.
He is 50 years old and works as a mechanic.
Visit with Dr. Ramirez on 10/21/2077.

CliDEID will produce the following output (stand-off):

```
<?xml version="1.0" encoding="UTF-8" ?>
<NGRID_deId>
<TEXT><![CDATA[Mr. Smith was admitted to the ABCD hospital.
He is 50 years old and works as a mechanic.
Visit with Dr. Ramirez on 10/21/2077 .
]]></TEXT>
<TAGS>
<AGE id="P0" start="52" end="54" text="50" TYPE="AGE" comment="" />
<DATE id="P1" start="117" end="127" text="10/21/2077" TYPE="DATE" comment="" />
<NAME id="P2" start="106" end="113" text="Ramirez" TYPE="DOCTOR" comment="" />
<LOCATION id="P3" start="30" end="34" text="ABCD" TYPE="HOSPITAL" comment="" />
<PROFESSION id="P4" start="80" end="88" text="mechanic" TYPE="PROFESSION" comment="" />
<NAME id="P5" start="4" end="9" text="Smith" TYPE="PATIENT" comment="" />
</TAGS>
</NGRID_deId>
```

More details on the architecture and the performance of the tool can be found in the paper below. Please cite this publication if you use CliDEID:

> Dehghan A., Kovačević A., Karystianis G, Keane JA., & Nenadic, G. (2017). Learning to identify protected health information by integrating knowledge- and data-driven algorithms: a case study on psychiatric evaluation notes. *Journal of Biomedical Informatics*. Submitted.

Contact Aleksandar Kovacevic (http://informatika.ftn.uns.ac.rs/AleksandarKovacevic/, kocha78@uns.ac.rs) with any questions, bugs and/or suggestions.

Installation:

1. Unzip the CliDEID archive to the folder of your choice.
2. Download the CliDEID_resources.7z from https://drive.google.com/open?id=0BwLyTML0dZsqb1ozSndjbl9rUGM and unzip it in the same folder as CliDEID.
3. If you are using an operating system other than Windows you need to install CRF++ following the instructions from https://taku910.github.io/crfpp/.

Running CliDEID:

Prerequisites: You need to have Java JRE installed and added to the path.

You can run the tool using the following line:
<u>*Windows:*</u>
      runCliDEID.bat *input_folder output_folder*
<u>*Linux:*</u>
      runCliDEID.sh *input_folder output_folder*

Where:
--*input_folder* is the folder with plain text files
--*output_folder* is the folder where the resulting files will be stored

Example input and output are provided in the /doc folder.

CliDEID can theoretically run on any platform that supports Java, but it has been tested on Windows 7 Professional (X86-64).