Text File Format Identification

Santhilata K.V.

02/04/2019

Agenda

- Introduction
- Research question
- Time line
- Work done so far
  - Papers reviewed / to review
  - Software available
- Targets for the next meeting

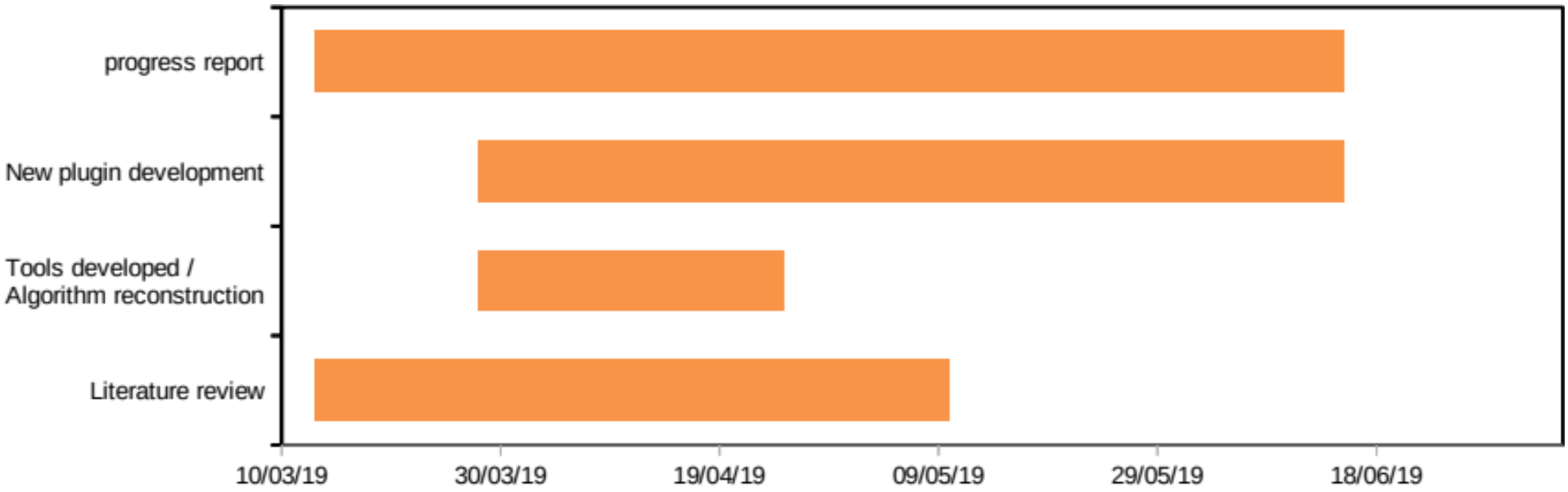
Introduction

Source:
https://scientificwritingtips.wordpress.com/the-cartoons/writingtips-08/

Research Question

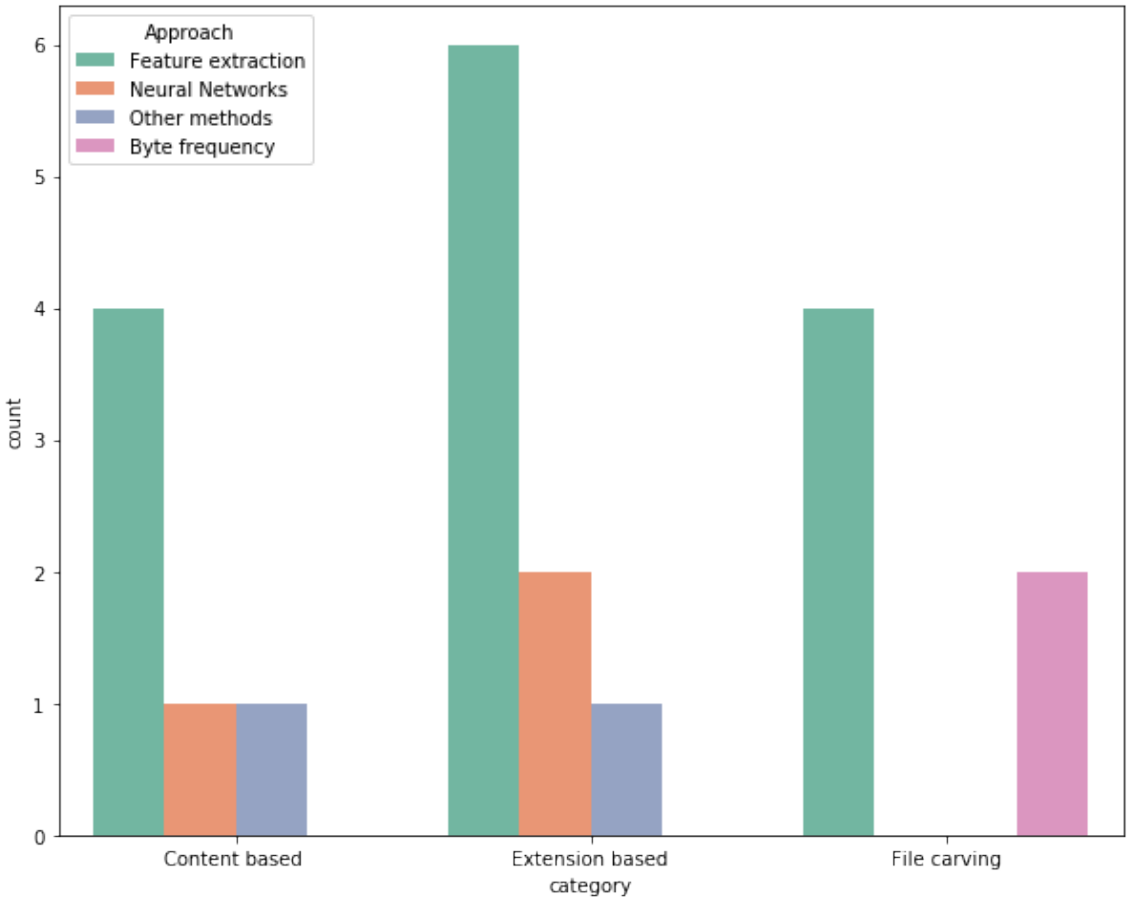
- How to correctly identify the file type from text files?
  - A normal text file is a flat file structured as a sequence of lines electronic text.
  - Text files often contain code segments, technical scripts or unidentified data objects. These files often pass through ASCII/ Unicode filters.

Timeline

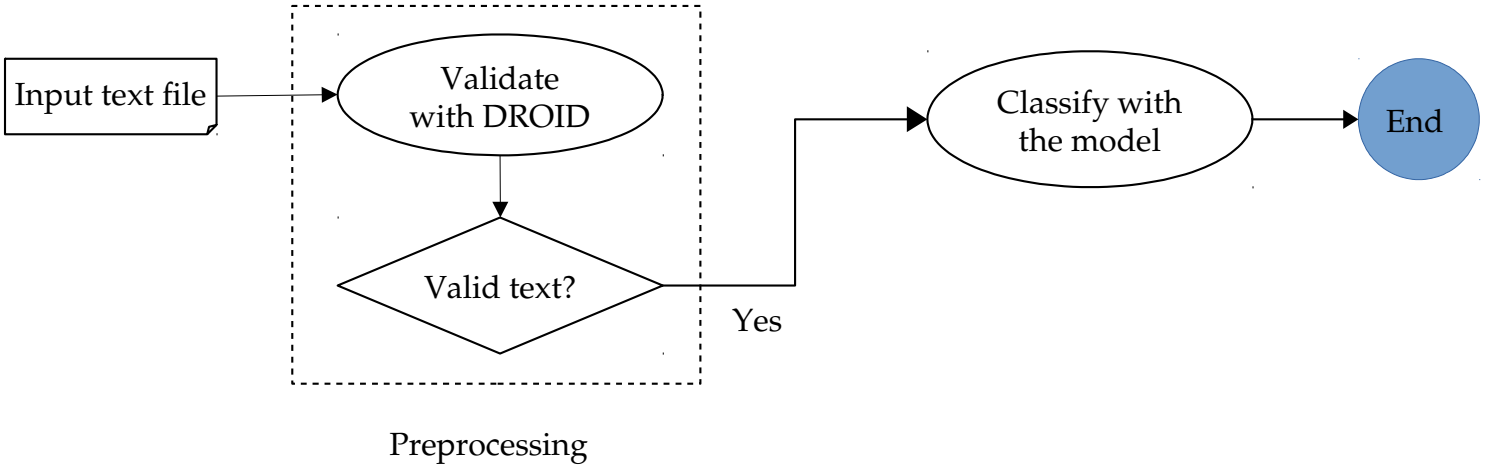

progress report
New plugin development
Tools developed /
Algorithm reconstruction
Literature review
10/03/19
30/03/19
19/04/19
09/05/19
29/05/19
18/06/19

Work done so far

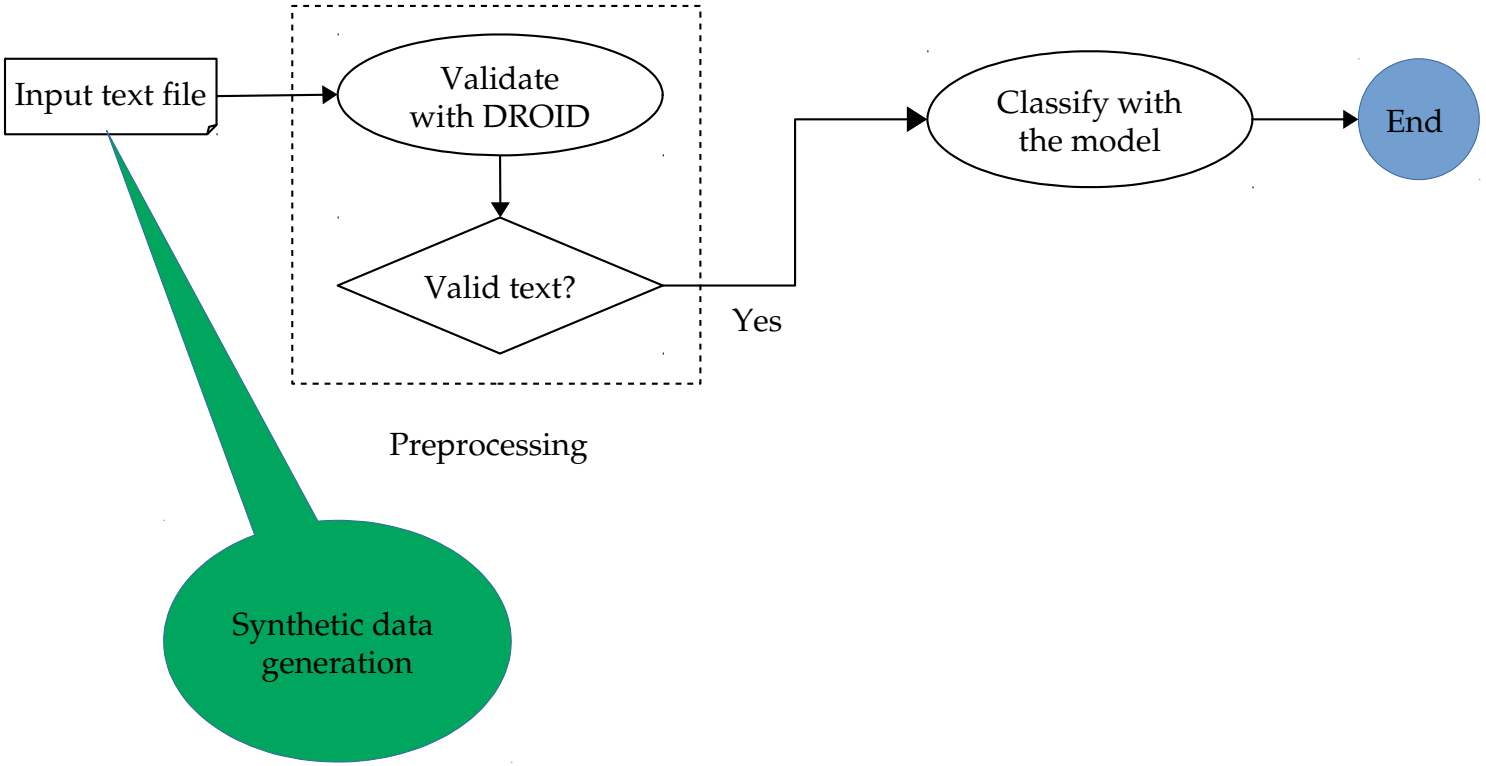
- Papers Reviewed / To read



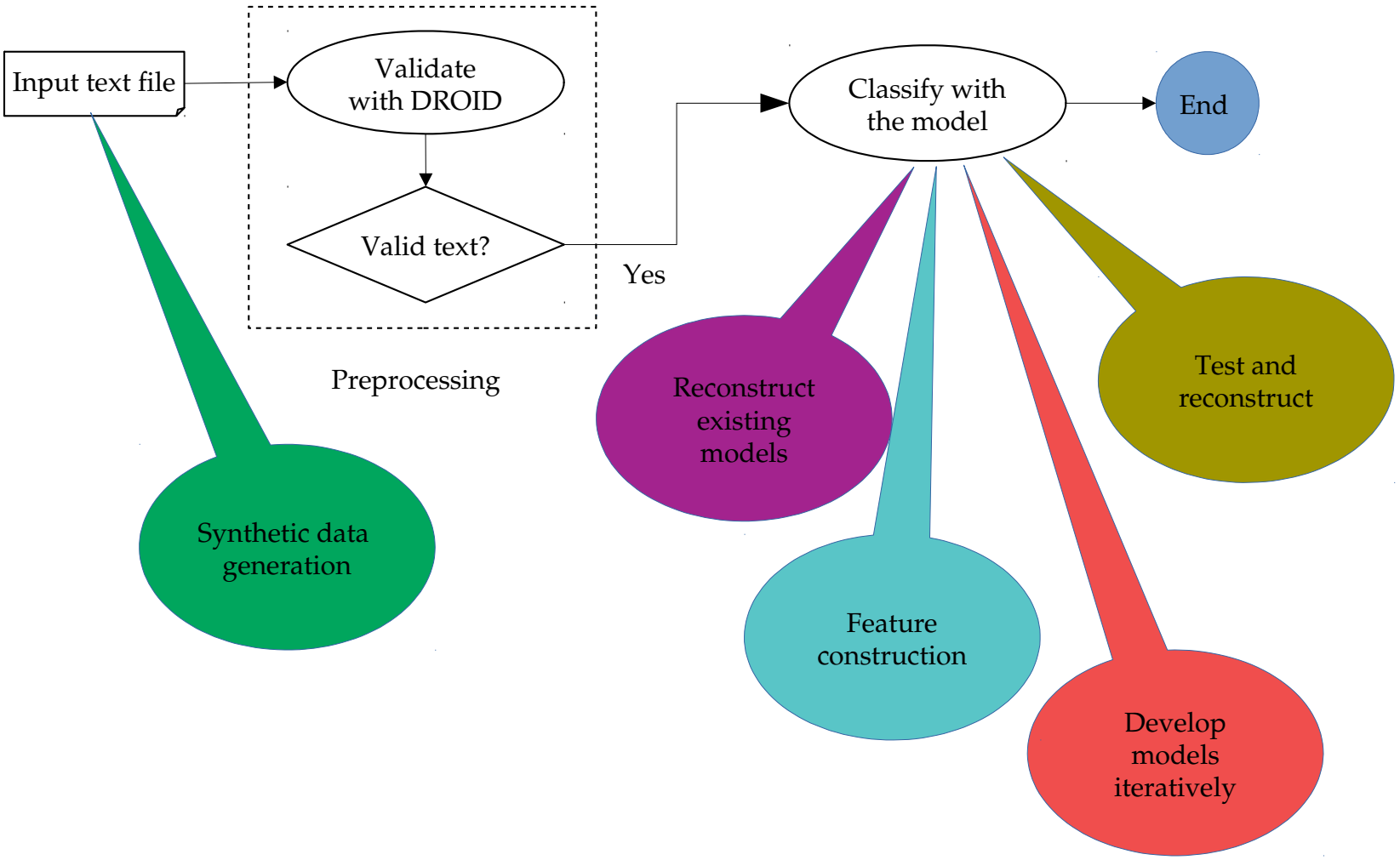
Workflow (top view)



Workflow (top view)



Workflow (top view)



Hickups

- Majority of the papers are published by IEEE
- Number of papers only discuss methods / approach but neither tool nor sample datasets. A challenge to decide whether a particular approach is useful!
- Algorithm reconstruction is going to take some time.