
Algorithm 4 Classify Genre of Document

Input Document, Source_url

Initialize *newsSource* \leftarrow [...]

Initialize *techSource* \leftarrow [...]

/ Calculate the originality and non-article ratio of document */*

refernce_ratio \leftarrow originality_check(document)

ioc_ratio \leftarrow non_article_fmt_check(document)

/ Judge the genre cascading */*

if *Source_url* \in *newsSource* **or** *refernce_ratio* \geq 30% **then**
 determine type news

if *Source_url* \in *techSource* **or** *ioc_ratio* \geq 30% **then**
 determine type technical_report

if *multiple used_malware* in document **then**
 determine type campaign_report

else

determine type analysis_report

end if
