1. ~1k articles for which the file hasn’t been generated.
2. ~1k articles (mostly NBC) for which the xpath is unable to get anything.

CNN: id:55 Cuomo. In some cases, they are writing in “div” class instead of “p”. Maybe try with div or p for the same class?

This is an example of “h\_” class. But, despite this, it is actually selecting multiple articles.

Remedies: 1. Use the same Xpath that for edition.cnn.com.

2 Use “//p[@class=] or //div[]” as xpath.