Pre processing

* Remove date indications: ^\r\n\\*\\*\\*\\*\\*\\*\\*[ Aa-Zz 0-9 \\*]\*\r\n$
* Remove first line “King James Version of the Bible”

Named Entity Extraction

* Exclude lower case
* Ignore case