Done

To test

To do

(1) country) non-intervention

e.g. Japan)

find the first country with bracket that appears before non-intervention.

I want the country with the half bracket closely attached to the country, there is no other token between the country and the half bracket. -> In "country)", how far before and after the keyword non-intervention should I look for the country ?

(2) Letter dated {date} from {str which contains a country} to the united nations \*.

Examples:

Letter dated 14 February 2018 from the Permanent Representative of Russian Federation to the United Nations addressed to the Secretary-General

Letter dated 14 February 2018 from Cyprus to the United Nations addressed to the United Nations addressed to the Secretary-General and the President of the Security Council

Letter dated 6 March 1995 from the Chairman of the Committee on the Elimination of Racial Discrimination to the Chargé d’affaires a.i. of the Permanent Mission of the Federal Republic of Yugoslavia to the United Nations Office at Geneva

find ten tokens after letter dated to see whether there is "country".

if there are more than 4 "letters" found in one document, eliminate this document.

(3) Cuba: draft resolution

Peru and Thailand: draft resolution

Many countries……….: draft resolution

Many countries……….: revised draft resolution

find the exact tokens " : draft resolution" or ": revised draft resolution", if there are more than 4 draft resolutions found in one document, eliminate this document. -> Re-tokenization