Assignment Demo:

For starters take any document and apply your processing techniques until tune everything to your liking, this will help you apply it to the other documents without any issues.

You need to preprocess the text as described in the assignment document. After that you need to divide each document into sentences as every sentence is a string o words that ends with a period (.). let’s call each sentence a Mini-Doc, for Mini-Doc you need to create an id consisting of Doc number and Mini-Doc number you can create the number by the order of sentences and documents that are read.

For example:

We have the following documents:

Doc1:

“Nor should the world find much solace in Clinton’s counter-jihad plans, which she outlined in a 2014 interview with The Atlantic’s Jeffrey Goldberg. Likening the **struggle** against Islamic terrorists to the struggle against the Soviet Union, Clinton acknowledged that, during the cold war, “we supported really **nasty** guys, we did some things that we are not particularly proud of, from Latin America to Southeast Asia”.

In the end, however, communism collapsed: “We achieved it.” Nevermind all the foreign bloodshed – or that whole link between the anti-Soviet war and the present jihad.

Regarding certain objectively “**nasty** guys” on the contemporary scene, meanwhile, Clinton had nothing but positive things to say. Questioned by Goldberg about the most recent Israeli assault on the Gaza Strip – in which well over 2,000 Palestinians were killed, the vast majority of them civilians – the presidential hopeful opined: “I think Israel did what it had to do.”

She then went as far as to propose possible anti-Semitic motives for opposing Israeli actions.

And this is not the only instance in which we’d be forgiven for thinking Clinton was actually running for office in Israel; during her previous campaign in 2008, she warned that “we would be able to totally obliterate” Iran in the event of an attack on the Jewish state.”

Doc2:

“In a November essay for Harper’s, titled “Stop Hillary!“, Doug Henwood marvels at the “widespread liberal fantasy of [Clinton] as a progressive paragon” – an image not easily reconciled with her tendency to out-hawk the hawks. “She backed an escalation of the Afghanistan war,” Henwood writes, “lobbied on behalf of a continuing military presence in Iraq, urged Obama to bomb Syria, and supported the intervention in Libya”.

He tacks on an observation by TIME Magazine’s Michael Crowley: “On at least three crucial issues -Afghanistan, Libya, and the bin Laden raid – Clinton took a more aggressive line than [Defense Secretary Robert] Gates, a Bush-appointed Republican.”

So much for that hallucination of an impassable chasm between the Democratic and Republican camps, by means of which partisan fanatics can blame everything on the other side rather than on the system itself.

The subheadline of Henwood’s article contains the following appeal: “Vote no to a Clinton dynasty.” He’s got my vote – but while we’re at it, let’s also vote “no” to a putrid duopoly that feigns irreparable division while remaining fully united in cultivating and exporting calamity.

Belen Fernandez is the author of The Imperial Messenger: Thomas Friedman at Work, published by Verso. She is a contributing editor at Jacobin Magazine.

The views expressed in this article are the author’s own and do not necessarily reflect Al Jazeera’s editorial policy.”

We break the documents into sentences:

D1S1: Nor should the world find much solace in Clinton’s counter-jihad plans, which she outlined in a 2014 interview with the Atlantic’s Jeffrey Goldberg.

D1S2: Likening the struggle against Islamic terrorists to the struggle against the Soviet Union, Clinton acknowledged that, during the cold war, “we supported really nasty guys, we did some things that we are not particularly proud of, from Latin America to Southeast Asia”.

…..

If we read “Nor should the world find much solace in Clinton’s counter-jihad plans, which she outlined in a 2014 interview with The Atlantic’s Jeffrey Goldberg” from Doc1 first then it’s id would be d1s2 or something similar.

Your terms index should be built as following:

Term Postings List

Struggle 🡪 [d1s2, ….] (this may be a “trust” term

Nasty 🡪 [ d1s2, d1s5 …] (emotional term, probably belongs to “Hate”)

….

This will allow you to lookup where each word is located and how many Mini-Docs it appears in.

In order to know which term belongs to an emotion according to the given lexicon you can either create a lookup table where each emotional term maps to its corresponding emotion. After that you can use the created index to lookup which trust terms and neighbors’ emotional terms are in the same sentence and check if they appear in the same sentence.

Example for emotions lookup table:

Term Emotion

Hit 🡪 angry

Cry 🡪 sad

Nasty 🡪 hate

So if for example, the term “Struggle” appears in id d1s2 and term “Nasty” appears in the same id, that means they appear in the same sentence so you need to count their cooccurrence.

After that you can follow the rest of the assignment requirements.