Named Entity Recognition For Information Extraction From A Text

07.02.2021

Aslantas Serdal
Data Science (411)
University of Bucharest
Bucharest,Romania

Overview

Named entity recognition is an important concept in NLP. It is used to categorize the names in a sentence as the names of places, organizations, persons, numerical, money, date, etc. NER is widely used in NLP to solve so many real-world problems like:

Where is the location of a specific organization mentioned in the news?

Who are the people in the article about Trump's contribution to the conflict between Saudi Arabia and Yemen?

Which countries are mentioned in the list of most corrupt countries?

Goals

In this project we are going to use SpaCy and NLTK to extract information from an article.

These libraries are two of the most popular tools in NLP for named entity recognition.

Information Extraction

I took a sentence from The New York Times, “Democrats pressed past Republicans’ objections to remove the Georgia freshman from her two committee posts in a vote without precedent in the modern Congress.”

Then we apply word tokenization and part-of-speech tagging to the sentence.

Let’s see what we get:

[('Democrats', 'NNPS'), ('pressed', 'VBD'), ('past', 'JJ'), ('Republicans', 'NNPS'), ('’', 'VBP'), ('objections', 'NNS'), ('to', 'TO'), ('remove', 'VB'), ('the', 'DT'), ('Georgia', 'NNP'), ('freshman', 'NN'),

('from', 'IN'), ('her', 'PRP$'), ('two', 'CD'), ('committee', 'NN'), ('posts', 'NNS'), ('in', 'IN'), ('a', 'DT'), ('vote', 'NN'), ('without', 'IN'), ('precedent', 'NN'), ('in', 'IN'), ('the', 'DT'), ('modern', 'JJ'), ('Congress', 'NNP')]

Noun Phrase Chunk

We get a list of tuples containing the individual words in the sentence and their associated part-of-speech.

Now we’ll implement noun phrase chunking to identify named entities using a regular expression consisting of rules that indicate how sentences should be chunked.

Our chunk pattern consists of one rule, that a noun phrase, NP, should be formed whenever the chunker finds an optional determiner, DT, followed by any number of adjectives, JJ, and then a noun, NN.

np = 'NP: {<DT>?<JJ>*<NN>}'

Using this pattern, we create a chunk parser and test it on our sentence.

(S
Democrats/NNPS pressed/VBD past/JJ Republicans/NNPS ’/VBP objections/NNS to/TO
remove/VB the/DT Georgia/NNP (NP freshman/NN) from/IN her/PRP$ two/CD
(NP committee/NN) posts/NNS in/IN (NP a/DT vote/NN)without/IN (NP precedent/NN) in/IN
the/DT modern/JJ Congress/NNP)

IOB tags have become the standard way to represent chunk structures in files, and we will also be using this format.

[('Democrats', 'NNPS', 'O'), ('pressed', 'VBD', 'O'), ('past', 'JJ', 'O'), ('Republicans', 'NNPS', 'O'),

('’', 'VBP', 'O'), ('objections', 'NNS', 'O'), ('to', 'TO', 'O'), ('remove', 'VB', 'O'), ('the', 'DT', 'O'),

('Georgia', 'NNP', 'O'), ('freshman', 'NN', 'B-NP'), ('from', 'IN', 'O'), ('her', 'PRP$', 'O'), ('two', 'CD', 'O'),

('committee', 'NN', 'B-NP'), ('posts', 'NNS', 'O'), ('in', 'IN', 'O'), ('a', 'DT', 'B-NP'), ('vote', 'NN', 'I-NP'), ('without', 'IN', 'O'), ('precedent', 'NN', 'B-NP'), ('in', 'IN', 'O'), ('the', 'DT', 'O'), ('modern', 'JJ', 'O'), ('Congress', 'NNP', 'O')]

In this representation, there is one token per line, each with its part-of-speech tag and its named entity tag. Based on this training corpus, we can construct a tagger that can be used to label new sentences.; and use the nltk.chunk.conlltags2tree() function to convert the tag sequences into a chunk tree.

(S Democrats/NNPS pressed/VBD past/JJ Republicans/NNPS'/VBP objections/NNS to/TO

remove/VB the/DT Georgia/NNP (NP freshman/NN) from/IN her/PRP$ two/CD

(NP committee/NN) posts/NNS in/IN (NP a/DT vote/NN) without/IN (NP precedent/NN) in/IN

the/DT modern/JJ Congress/NNP)

With the function nltk.ne_chunk(), we can recognize named entities using a classifier, the classifier adds category labels such as PERSON, ORGANIZATION, and GPE.

(S Democrats/NNPS pressed/VBD past/JJ Republicans/NNPS '/VBP objections/NNS to/TO

remove/VB the/DT **(GPE Georgia/NNP)** freshman/NN from/IN her/PRP$ two/CD committee/NN posts/NNS in/IN a/DT vote/NN without/IN precedent/NN in/IN the/DT modern/JJ **(ORGANIZATION Congress/NNP))**

Spacy As A Tool For Named Entity Recognition

We are using the same sentence, "Democrats pressed past Republicans' objections to remove the Georgia freshman from her two committee posts in a vote without precedent in the modern Congress." One of the nice things about Spacy is that we only need to apply nlp once, the entire background pipeline will return the objects.

Sentence = nlp(sent)

This is the label of the each name in the sentence.

[('Democrats', 'NORP'),

('Republicans', 'NORP'),

('Georgia', 'GPE'),

('two', 'CARDINAL'),
('Congress', 'ORG')]

Democrats and Republicans are NORP (nationalities or religious or political groups), Georgia is a state, two is a numeric value and Congress is an organization. They are all correct.

This is the pos tag of each word in the sentence.

[Democrats PROPN pressed VERB past ADP Republicans PROPN' PART objections NOUN to PART remove VERB the DET Georgia PROPN freshman NOUN from ADP her DET two NUM committee NOUN posts NOUN in ADP a DET vote NOUN without ADP precedent NOUN in ADP the DET modern ADJ Congress PROPN]

Token

During the above example, we were working on entity level, in the following example, we are demonstrating token-level entity annotation using the BILUO tagging scheme to describe the entity boundaries.

TAG	DESCRIPTION
B EGIN	The first token of a multi-token entity.
I N	An inner token of a multi-token entity.
L AST	The final token of a multi-token entity.
U NIT	A single-token entity.
O UT	A non-entity token.

[(Democrats, 'B', 'NORP'), (pressed, 'O', ''), (past, 'O', ''), (Republicans, 'B', 'NORP'), (', 'O', ''), (objections, 'O', ''), (to, 'O', ''), (remove, 'O', ''), (the, 'O', ''), (Georgia, 'B', 'GPE'), (freshman, 'O', ''), (from, 'O', ''), (her, 'O', ''), (two, 'B', 'CARDINAL'), (committee, 'O', ''), (posts, 'O', ''), (in, 'O', ''), (a, 'O', ''), (vote, 'O', ''), (without, 'O', ''), (precedent, 'O', ''), (in, 'O', ''), (the, 'O', ''), (modern, 'O', ''), (Congress, 'B', 'ORG')]

Extracting NER from an article

Now let's get serious with SpaCy and extracting named entities from a New York Times article, '***Biden Signals Break With Trump Foreign Policy in a Wide-Ranging State Dept. Speech.'***

There are 272 entities in the article and they are represented as 12 unique labels:

{'DATE': 34, 'PRODUCT': 2, 'ORG': 35, 'GPE': 75, 'PERSON': 54,

'CARDINAL': 11, 'NORP': 49, 'MONEY': 2, 'ORDINAL': 6, 'LOC': 2, 'TIME': 1, 'FAC': 1}

The following are three most frequent tokens.
[('American', 26), ('Biden', 25), ('Yemen', 15)]

Let's randomly select one sentence to learn more.

'Soon after Iran-allied Houthi forces took over Yemen's capital, Sana, in the fall of 2014, the Saudis and their gulf allies began airstrikes and then bought billions of dollars in American weaponry, with the goal of ousting the Houthi rebels from northern Yemen.'

Let's run displacy.render to generate the raw markup.

Next, we verbatim, extract part-of-speech and lemmatize this sentence.

[('Soon', 'ADV', 'soon'),
('Iran', 'PROPN', 'Iran'),
('allied', 'PROPN', 'allied'),
('Houthi', 'PROPN', 'Houthi'),
('forces', 'NOUN', 'force'),
('took', 'VERB', 'take'),
('Yemen', 'PROPN', 'Yemen'),
('capital', 'NOUN', 'capital'),
('Sana', 'PROPN', 'Sana'),
('fall', 'NOUN', 'fall'),
('2014', 'NUM', '2014'),
('Saudis', 'PROPN', 'Saudis'),
('gulf', 'PROPN', 'gulf'),
('allies', 'NOUN', 'ally'),
('began', 'VERB', 'begin'),

('airstrikes', 'NOUN', 'airstrike'),

('bought', 'VERB', 'buy'),

('billions', 'NOUN', 'billion'),

('dollars', 'NOUN', 'dollar'),

('American', 'ADJ', 'american'),

('weaponry', 'NOUN', 'weaponry'),

('goal', 'NOUN', 'goal'),

('ousting', 'VERB', 'oust'),

('Houthi', 'PROPN', 'Houthi'),

('rebels', 'NOUN', 'rebel'),

('northern', 'ADJ', 'northern'),
('Yemen', 'PROPN', 'Yemen')]

Named entity extraction are correct except “Sana”.

{'Iran': 'GPE',

'Houthi': 'ORG',

'Yemen': 'GPE',

'Sana': 'PERSON',

'the fall of 2014': 'DATE',

'Saudis': 'NORP',

'billions of dollars': 'MONEY',
'American': 'NORP'}
Finally, we visualize the entity of the entire article.

[Biden Signals Break With Trump Foreign Policy in a Wide-Ranging State Dept., Speech - The New York Times , SectionsSEARCHSkip **ORG** , to contentSkip to site indexPoliticsLog, inToday’s PaperPolitics|Biden Signals Break With Trump Foreign Policy in a Wide-Ranging State Dept., SpeechAdvertisementContinue reading the main storySupported byContinue **PRODUCT** reading the main storyBiden Signals Break With Trump Foreign Policy in a Wide-Ranging State Dept., SpeechThe **ORG** president said that he would end support for Saudi Arabia **GPE** in its intervention in Yemen **GPE** and that the U.S. **GPE** would no longer be “rolling over in the face of Russia **GPE** ’s aggressive actions., ”Aid workers distributing rations last month **DATE** in Sana **GPE** , Yemen **GPE** ., Credit..., Yahya Arhab/EPA **ORG** , via ShutterstockBy David E. Sanger **PERSON** and Eric SchmittFeb **PERSON** . 4 **CARDINAL** , 2021WASHINGTON **CARDINAL** —, President Biden **PERSON** on Thursday **DATE** ordered an end to arms sales and other support to Saudi Arabia **GPE** for a war in Yemen **GPE**

that he called a “humanitarian and strategic catastrophe” and declared that the United States GPE would no longer be “rolling over in the face of Russia GPE ’s aggressive actions., ”The announcement was the clearest signal Mr. Biden PERSON has given of his intention to reverse the way President Donald J. Trump PERSON dealt with two CARDINAL of the hardest issues in American NORPforeign policy., Mr., Trump regularly rejected calls to rein in the Saudis NORP for the indiscriminate bombing they carried out in their intervention in the civil war in Yemen GPE as well as for the killing of a dissident journalist, Jamal Khashoggi PERSON , on the grounds that American NORP sales of arms to Riyadh GPE “creates hundreds of thousands CARDINAL of jobs” in the United States GPE ., And he repeatedly dismissed evidence of interference by President Vladimir V. Putin PERSON of Russia GPE in American NORP elections and Russia GPE ’s role in a highly sophisticated hacking of the United States GPE government., Saudi NORP leaders knew that the move was coming., Mr. Biden PERSON had promised to stop selling arms to them during the presidential campaign, and it follows the new administration’s announcement last month DATE that it was pausing the sale of $478 million MONEY in precision-guided munitions to Saudi Arabia GPE , a transfer the State Department ORG approved in DecemberDATE over strong objections in Congress ORG ., The administration has also announced a review of major AmericanNORP arms sales to the United Arab Emirates GPE ., But Mr. Biden PERSON ’s order on Thursday DATE went further, appearing to also end providing the Saudis NORP targeting data and logistical support., It was not only a rejection of Trump administration policy but a reversal of American NORP support for the Saudi NORP effort that dated to the Obama PERSON administration — and that Mr. Biden PERSON and his newly appointed secretary of state ORG , Antony J. Blinken PERSON , helped formulate., Soon after Iran GPE -allied Houthi ORG forces took over Yemen GPE ’s capital, Sana PERSON , in the fall of 2014 DATE , the Saudis NORP and their gulf allies began airstrikes and then bought billions of dollars MONEY in American NORP weaponry, with the goal of ousting the Houthi ORG rebels from northern Yemen GPE ., President Barack Obama PERSON gave the war his qualified approval, in part to assuage Saudi NORP anger over the Iran GPE nuclear deal in 2015 DATE ., Two years later DATE , Mr. Trump PERSON doubled down, embracing the Saudi NORP crown prince, Mohammed bin Salman PERSON , despite mounting evidence that American NORP fingerprints — and American NORP -made munitions — were all over civilian deaths in the brutal civil war, which helped create the world’s greatest humanitarian crisis and a famine that is engulfing the country., Now Mr. Biden PERSON is no longer making the case that American NORPsupport was helping bring the war to a conclusion that would stop the civilian deaths., His goal is to force the SaudisNORP into a diplomatic solution, and he appointed a longtime career diplomat, Timothy Lenderking PERSON , to act as special envoy to negotiate a settlement., “This war has to end,” Mr. Biden PERSON said Thursday DATE at the State Department ORG , in his first ORDINAL major foreign policy speech since taking office., He said the speech was intended to “send a clear message to the world: America GPE is back., ”But, Mr. Biden PERSON also made clear that while he was seeking to force the Saudis NORP to face up to the huge human toll of their intervention in YemenGPE , he was not leaving them alone to deal with a hostile Iran GPE ., He said he would continue sales of defensive weapons to Saudi Arabia GPE that were designed to protect against missiles, drones and cyberattacks from TehranGPE ., “We’re going to continue to support and help Saudi Arabia GPE defend its sovereignty and its territorial integrity and its people,” the president said., He said nothing about the possibilities of imposing sanctions on the crown prince for his involvement in the Khashoggi LOC killing, though Mr. Biden PERSON ’s director of national

intelligence, Avril D. Haines PERSON , has said she plans to declassify intelligence about the killing., In another reversal of Trump-era ORG policy, Mr. Biden PERSON also announced he was “stopping any planned troop withdrawals from Germany GPE ,” halting Mr. Trump PERSON ’s, order to redeploy 12,000 CARDINAL troops stationed in GermanyGPE ., National security experts from both parties had called that order shortsighted, saying it was rooted in Mr. TrumpPERSON ’s dislike of Chancellor Angela Merkel PERSON and his determination to force NATO ORG nations to pay more for their own defenses, no matter what the strategic costs to the United States GPE ., The New WashingtonLive UpdatesUpdated Feb. 4, 2021 DATE , 5:23 p.m. TIME , ETThe Senate Intelligence Committee ORG will examine anti-government extremists., Schumer PERSON , seeking to pressure Biden LOC , pushes a $ 50,000 MONEYstudent debt forgiveness plan., Romney PERSON proposes monthly DATE payments to parents to fight child poverty., But strategically, it is Mr. Biden PERSON ’s warning to Moscow GPE that may, over the long run, say more about the redirection of American NORP foreign policy than the decision to limit Saudi Arabia GPE , ’s ability to prosecute a regional war., He is the first ORDINAL president since the fall of the Soviet Union GPE who has decided against trying a “reset” with Russia GPE , instead announcing what amounts to a new strategy of deterrence, if not containment., Mr., Biden LOC hardened his vow to respond to Russian NORP efforts to disrupt American NORPdemocracy and to the SolarWinds hacking, a vast intrusion into American NORP government and private networks whose dimensions are still a mystery., He said that in a call with Mr. Putin PERSON last week DATE , he told the Russian NORP leader “in a manner very different from my predecessor, that the days of the United States GPE rolling over in the face of Russia GPE ’s aggressive actions — interfering with our elections, cyberattacks, poisoning its citizens — are over.”Mr., Biden PERSON called on Moscow GPE to release the imprisoned dissident Aleksei A. Navalny ORG , adding, “We will not hesitate to raise the cost on Russia GPE .”, But he did not specify how he would accomplish that, and his options may be limited., While the president hinted at a response “in kind” to the cyberattack, that could set off a round of escalation that has many American NORP officials concerned., Mr., Biden LOC ’s announcement came a day DATE after the United States GPE and Russia GPE formally approved the five-yearDATE extension of New START GPE , the one CARDINAL remaining nuclear arms treaty between the twoCARDINAL countries., Mr. Trump PERSON had insisted on amendments, but Mr. Biden PERSON concluded that it was wiser to get a prospective nuclear arms race off the table at a time of heightened competition in other arenas., He said that strong alliances were key to deterring Moscow GPE , along with the “growing ambitions of China GPE to rival the United States GPE .”, And Mr. Biden PERSON ’s aides concede that it is a powerful, rising, technologically sophisticated China GPE , not a declining and disruptive Russia GPE , that poses a deeper long-term threat., But the president spent little time on China GPE in his speech, a recognition that his administration will be spending months DATE trying to reformulate its approach to Beijing GPE ., VideotranscriptBackbars0:00/1:56-0:00transcript‘America Is Back’:, Biden Outlines Vision of Global LeadershipPresident Biden ORG said the United States GPE would repair alliances after years of “abuse.”, He also announced a freeze on troop withdrawals from Germany GPE and a diplomatic focus on ending the war in Yemen GPE ., America GPE is back, America GPEis back., Diplomacy is back at the center of our foreign policy., As I said in my inaugural address, we will repair our alliances and engage with the world once again, not to meet yesterday DATE ’s challenges, but, today DATE ’s and tomorrow’s., American NORP leadership must meet this new moment of advancing authoritarianism, including the growing ambitions of China GPE to rival the

United States GPE , and the determination of Russia GPE to damage and disrupt our democracy., We must meet the new moment of accelerating global, accelerating global challenges from the pandemic to the climate crisis to nuclear proliferation, challenging the will only to be solved by nations working together., Rebuilding the muscle of Democratic NORP alliances that have atrophied over the past few years DATE of neglect, and I would argue, abuse., American NORP alliances are our greatest asset, and leading with diplomacy means standing shoulder to shoulder with our allies and key partners, once again., Today DATE , I'm announcing additional steps to course correct our foreign policy and better uniting our democratic values with our diplomatic leadership., To begin, Defense ORG Secretary Austin PERSON will be leading a global posture review of our forces so that our military footprint is appropriately aligned with our foreign policy and national security priorities., And while this review is taking place, we'll be stopping any planned troop withdrawals from Germany GPE ., We're also stepping up our diplomacy to end the war in Yemen GPE , a war which has created humanitarian and strategic catastrophe., President Biden PERSON said the United States GPE would repair alliances after years of "abuse.", He also announced a freeze on troop withdrawals from Germany GPE and a diplomatic focus on ending the war in Yemen GPE ., CreditCreditORG ..., Stefani Reynolds PERSON for The New York TimesMr., Biden LOC started his visit to the State Department ORG with a talk to an incoming class of 165 CARDINAL young diplomats — an annual DATE inflow of talent that was suspended in Mr. Trump PERSON 's first year DATE — promising to rebuild "the muscle of democratic alliances that have atrophied over the past few years DATE of neglect, and, I would argue, abuse., "His choice of venue was deliberate:, His predecessor preferred visiting the Pentagon ORG and the C.I.A. GPE , symbols of American NORP hard power, in Mr. Trump PERSON 's view., Mr. Biden PERSON , who spent decades DATEin the Senate ORG on the Foreign Relations Committee ORG , made the headquarters of American NORPdiplomacy his first ORDINAL stop, telling its 70,000 CARDINAL employees that "I'm going to have your back., "Those diplomats have found the American NORP support for the Saudis NORP increasingly hard to defend., When Saudi F-15 ORG warplanes took off from an air base in southern Saudi Arabia GPE for a bombing run over Yemen GPE , it was not just a plane and bombs that were American NORP ., American NORP mechanics serviced the jet and carried out repairs on the ground., American NORP technicians upgraded the targeting software and other classified technology, which Saudis NORP were not allowed to touch., The pilot was likely to have been trained by the United States Air Force ORG ., At a flight operations room in the capital, Riyadh GPE , Saudi NORP commanders sat near American NORP military officials who provided intelligence and tactical advice, mainly aimed at stopping the SaudisNORP from killing Yemeni NORP civilians., While the Pentagon ORG and State Department ORG have denied knowing whether American NORP bombs were used in the war's most notorious airstrikes — which have struck weddings, mosques FAC and funerals — a former senior State Department ORG official told The New York TimesORG in 2018 DATE that the United States GPE had access to records of every airstrike over Yemen GPE since the early days DATE of the war., At the same time, the Saudis NORP whitewashed an American NORP -sponsored initiative to investigate errant airstrikes and often ignored a voluminous no-strike list., ImageA Saudi NORP airstrike in July DATE in Sana. GPE , Mr. Biden PERSON ordered an end to the sale of weapons and other support to Saudi Arabia GPE for a war in Yemen GPE that he called a "humanitarian and strategic catastrophe."Credit, ..., Khaled Abdullah/ReutersMr PERSON ., Biden PERSON 's actions drew praise from human rights groups and their advocates in Congress ORG ., "The shift from a failed war strategy toward a comprehensive

diplomatic approach cannot come a moment too soon," David Miliband PERSON , a former senior British NORP diplomat who is now the president and chief executive of the International Rescue Committee ORG , said in a statement., Senator Christopher S. MurphyPERSON , Democrat NORP of Connecticut GPE , who is one CARDINAL of the staunchest congressional critics of the Saudi NORP -led campaign in Yemen GPE , echoed that sentiment: "For the last six years DATE , the war in Yemen GPE has led to a horrific humanitarian crisis that has hurt U.S. GPE security and moral credibility., TodayDATE 's actions by President Biden PERSON are a decisive first ORDINAL step to bring this nightmare to an end., "Mr., Biden PERSON 's decision to make the State Department ORG his first ORDINAL stop in an executive branch department was calculated, part of his broader effort to argue that Mr. Trump PERSON 's four years DATE in office were an aberration — and an affront to American NORP values., "Though many of these values have come under intense pressure in recent years DATE , even pushed to the brink in the last few weeks DATE , the American NORPpeople are going to emerge from this moment stronger, more determined and better equipped to unite the world in fighting to defend democracy," he said, "because we have fought for it ourselves., "Still, , the speech had the feel of a president speaking a bit into the diplomatic void., The State Department ORG is largely empty because of the pandemic, and Mr. Blinken PERSON , who was Mr. Biden PERSON 's top foreign policy adviser for the past two decades DATE , works on a depopulated seventh ORDINAL floor., The nominees for the department's other top posts have not yet had Senate ORG confirmation hearings., Michael Crowley PERSON contributed reporting., AdvertisementContinue ORGreading the main storySite IndexSite Information Navigation© ORG 2021 The New York Times CompanyNYTCoContact UsWork ORG with usAdvertiseT Brand StudioYour Ad ChoicesPrivacy PolicyTerms of ServiceTerms of SaleSite NORP MapCanadaInternationalHelpSubscriptions]