

In [1]:

```
import spacy
from spacy import displacy
from collections import Counter
import en_core_web_sm
from bs4 import BeautifulSoup
import requests
import re

# Load English language model
nlp = en_core_web_sm.load()

# Process a text document
doc = nlp("""Hello! My name is Aurin and I live in Dhaka.
        I study "Robotics and Mechatronics
        Engineering" at University of Dhaka
        which was established
        on 15th May, 2015. Currently, I work as a Teacher and earn about $8000/month"""))
```

In [2]:

```
# Print entities with their IOB (Inside, Outside, Beginning) and entity type
print([(X, X.ent_iob_, X.ent_type_) for X in doc])

[(Hello, 'O', ''), (!, 'O', ''), (My, 'O', ''), (name, 'O', ''), (is, 'O', ''), (Aurin, 'B', 'PERSON'), (and, 'O', ''), (I, 'O', ''), (live, 'O', ''), (in, 'O', ''), (Dhaka, 'B', 'GPE'), (., 'O', ''), (
    , 'O', ''), (I, 'O', ''), (study, 'O', ''), (", 'O', ''), (Robotics, 'B', 'WORK_OF_ART'), (and, 'I', 'WORK_OF_ART'), (Mechatronics, 'I', 'WORK_OF_ART'), (
    , 'I', 'WORK_OF_ART'), (Engineering, 'I', 'WORK_OF_ART'), (", 'O', ''), (at, 'O', ''), (University, 'B', 'ORG'), (of, 'I', 'ORG'), (Dhaka, 'I', 'ORG'), (
    , 'O', ''), (which, 'O', ''), (was, 'O', ''), (established, 'O', ''), (
    , 'O', ''), (on, 'O', ''), (15th, 'B', 'DATE'), (May, 'I', 'DATE'), (, 'I', 'DATE'), (2015, 'I', 'DATE'), (., 'O', ''), (Currently, 'O', ''), (, 'O', ''), (I, 'O', ''), (work, 'O', ''), (as, 'O', ''), (a, 'O', ''), (Teacher, 'O', ''), (and, 'O', ''), (earn, 'O', ''), (about, 'B', 'MONEY'), ($, 'I', 'MONEY'), (8000, 'I', 'MONEY'), (/ , 'I', 'MONEY'), (month, 'I', 'MONEY')]
```

In [3]:

```
# Visualize named entities in the document
displacy.render(nlp(str(doc)), jupyter=True, style='ent')
```

Hello! My name is **Aurin PERSON** and I live in **Dhaka GPE** .

I study " **Robotics and Mechatronics Engineering WORK\_OF\_ART** " at **University of Dhaka ORG**

which was established

on **15th May, 2015 DATE** . Currently, I work as a Teacher and earn **about \$8000/month MONEY**

In [4]:

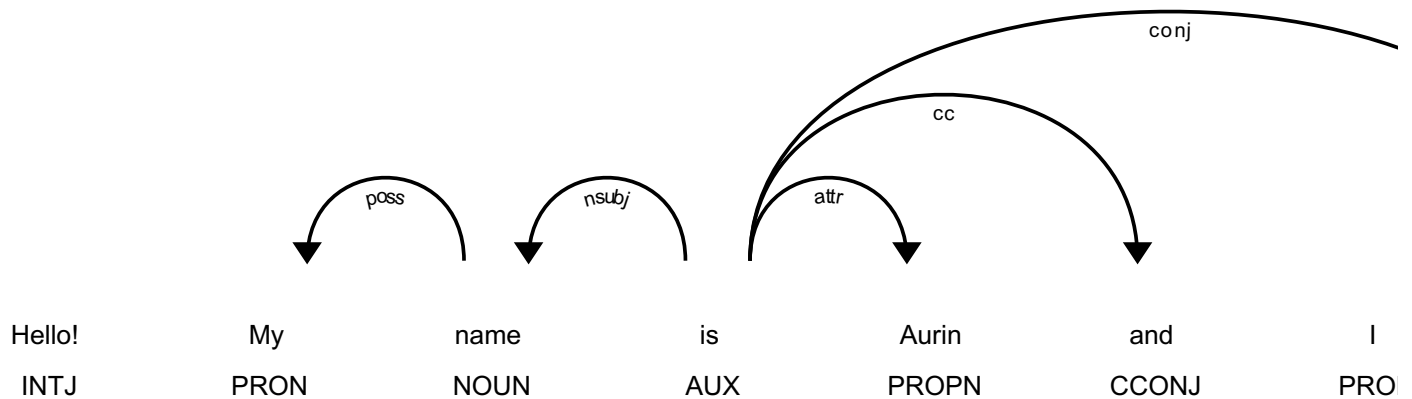
```
# Visualize syntactic dependencies in the document
displacy.render(nlp(str(doc)), jupyter=True, style='dep', options={'distance': 120})

# Sentences written in a certain way make a Huge Impact

# Two example sentences for NER analysis
sentence1 = "Alice Walker has booked a flight for December 10th, 2023, to travel from Dhaka to Cox's Bazar and experience the stunning coastal beauty."
sentence2 = "In her travel itinerary, Alice Walker is set to journey from Dhaka to Cox's Bazar on December 10th, 2023, immersing herself in the scenic wonders of the coastline."

# Process the sentences with NLP pipeline
```

```
doc1 = nlp(sentence1)
doc2 = nlp(sentence2)
```



In [5]:

```
# Visualize named entities in the sentences
displacy.render(doc1, style="ent", jupyter=True)
```

Alice Walker **PERSON** has booked a flight for December 10th, 2023 **DATE**, to travel from Dhaka **ORG** to Cox's Bazar **ORG** and experience the stunning coastal beauty.

In [6]:

```
displacy.render(doc2, style="ent", jupyter=True)
```

In her travel itinerary, Alice Walker **PERSON** is set to journey from Dhaka **ORG** to Cox's Bazar **ORG** on December 10th, 2023 **DATE**, immersing herself in the scenic wonders of the coastline.

In [7]:

```
# Extract text from a URL and process it
def url_to_string(url):
    res = requests.get(url)
    html = res.text
    soup = BeautifulSoup(html, 'html.parser')
    for script in soup(["script", 'style', "aside"]):
        script.extract()
    return " ".join(re.split(r'[\n\t]+', soup.get_text()))

# Convert the content of a URL to a SpaCy document
convert = url_to_string('https://dailycampus.com/2024/01/11/mens-basketball-huskies-escape-cintas-with-close-80-75-win/')
article = nlp(convert)

# Print the length of the processed document
len(article)
```

Out[7]:

1588

In [8]:

Out[8]:

Men's Basketball: Huskies escape Cintas with close 80-75 win | The Daily Campus Spotify Sign in News News Home Life Life Home Campus Events Movie & TV Reviews Music Reviews Opinion Opinion Home Letters to the Editor + Op-Eds Culture Shock Sports Sports Home Sports Staff Photo Photo Home Video Comics About About Us Our Staff Board of Directors Executive History Print Editions Special Editions Join Us Tip Line Sign in Welcome! Log into your account your username your password Forgot your password? Password recovery Recover your password your email Search Trending Now Men's Basketball: No. 4 Huskies rout Hoyas 80-67 in scrappy Big East showdown Resignation of Claudine Gay fuels firestorm over plagiarism, politics on campus An Analysis of UConn's 10-Year Plan. Part 1: The endowment Search 29 F Storrs The Daily CampusThe Daily Campus Tuesday, January 16, 2024 News News Home Life Life Home Campus Events Movie & TV Reviews Music Reviews Opinion Opinion Home Letters to the Editor + Op-Eds Culture Shock Sports Sports Home Sports Staff Photo Photo Home Video Comics About About Us Our Staff Board of Directors Executive History Print Editions Special Editions Join Us Tip Line Home Sports Men's Basketball: Huskies escape Cintas with close 80-75 win Sports Men's Basketball: Huskies escape Cintas with close 80-75 win By Evan Rodriguez - January 11, 2024 0 267 FacebookTwitter View this post on Instagram A post shared by UConn Huskies (@uconnhuskies) While 80% of the top 5 seeds in college basketball weren't able to grab a win this week, one team successfully grabbed a victory. That team happened to be the UConn Huskies, who beat the Xavier Musketeers at the Cintas Center on Wednesday night, 80-75. Coming into this game, UConn averaged 8.5 three-point shots per game, but the way that the program approached Wednesday evening's game would make you think that they shot even more. The Huskies' first three shots were all from beyond the arc and following a triple from Cam Spencer, it was a 9-4 start for Connecticut. But, with the history that both of these programs have in this building, the reality was that it would not be this easy throughout the entire matchup. Despite Spencer nearly making it a double-digit deficit, the Musketeers fought back with vengeance and really turned up their effort. With the duo of Dayvion McKnight and Quincy Olivari leading the way, the lead narrowed to just a 14-11 game. Just like that, the pressure was on to respond and UConn's answer came in the form of Alex Karaban. In under two minutes, the sophomore forward hit two massive jumpers from beyond the arc to give his squad a seven-point lead with 12:37 left in the first half. The Massachusetts native became the catalyst for a massive UConn run to make it a 26-17 game in favor of the reigning national champions. While the Huskies would push their lead to as much as 10 with 7:35 to go in the half, the Musketeers would get back in this one in a big way. Once again, it was McKnight and Olivari who led the charge to get the game to a two-point deficit. As for the Rice transfer specifically, he'd end the first half with 13 points, 4 rebounds and an assist. That would be critical in not only keeping this game close for his squad, but also setting up a Desmond Claude score inside to tie the game at 35 with 2:58 left. With momentum up for grabs heading into the half, both teams fought valiantly and in the end, it would be UConn who would win the first-half battle with a 42-38 advantage. From the first second-half whistle, it was Xavier who quickly got the game tied. Suddenly, it was a brand-new basketball game and the momentum had shifted toward a team with one of the most deadly home advantages in not just the Big East conference, but all of college basketball. It certainly wasn't the most ideal situation to be in, but the Huskies' confidence never shifted. While the most impressive first-half performance belonged to Karaban, the second half belonged to Tristen Newton. The East Carolina transfer made plays as a facilitator in the first half, but the second half really showed off his ability to put the ball in the basket. His shotmaking would give Connecticut a 5 point lead and a massive boost in offensive firepower. When coach Dan Hurley decided to put in Hassan Diarra, that's when his squad shifted to another level. The senior guard got UConn up to a 9-point lead with 11:53 to go in the 2nd half on a 3-pointer. View this post on Instagram A post shared by UConn Men's Basketball (@uconnmbb) But Xavier wasn't ready just yet to allow the Huskies to run away with this one. Following a Gytis Nemeiksa three-point shot and Abou Ousmane inside score, UConn's lead dropped to five. Once again, the Musketeers had given the Huskies another challenge to overcome, and just as before, the No. 4 team in the nation pushed past another obstacle. With the help of Diarra, Connecticut extended their lead to 10 with the help of the Queens native's 5 points in under a minute of play. Thanks to a jumper from Spencer and a tough layup from Castle, it was a massive 14 point UConn lead. As the game clock showed 5:39, the momentum was all in UConn's hands and all signs pointed to a huge win for the visitors. But instead, the Musketeers had other plans and after they went on a stunning 9-0 run, it quickly became a five point game. Even with 1:06 left, this game wasn't over and coach Sean Miller's squad certainly showed a relentless effort on the defensive end. Some grueling full-court pressure would help get Xavier a three-point game, but ultimately, it would not be enough. While the home squad did everything they could to push Connecticut to their limits, some key defensive plays from Spencer would be the deciding factor in the 80-75 win. While the Rutgers transfer's 19 points was certainly a highlight of Wednesday's victory, it was impossible to avoid the complete team showcase of the Huskies. Whether it was 16 points and 11 assists from Newton or 11 po

ints from Diarra off the bench, UConn showed why they're one of the most feared teams in the nation in an environment that they've notably struggled with in the past. This is certainly an incredible way for Dan Hurley and Co. to elevate their confidence as they prepare to potentially make a huge leap in the AP poll. Coming off this massive victory, Connecticut heads back home to the XL Center for a matchup with Georgetown at noon on Sunday. TAGSbasketballrecapUConn men's basketball FacebookTwitter Evan Rodriguez RELATED ARTICLES MORE FROM AUTHOR Big East Baller Update No. 5: Karaban leads Huskies to first AP No. 1 ranking since 2009 Men's Basketball: No. 4 Huskies rout Hoyas 80-67 in scrappy Big East showdown Women's Basketball: Stifling defense dictates No. 13 Huskies' victory over Providence Women's Basketball: Huskies beat Hoyas, powered by well-rounded effort Men's Basketball: No. 4 Huskies keep the momentum rolling with 88-81 comeback win over Butler Leave a ReplyCancel reply Sign in Welcome! Log into your account your username your password Forgot your password? Get help Password recovery Recover your password your email A password will be e-mailed to you. © Copyright - The Daily Campus THE DAILY CAMPUS1266 STORRS ROADSTORRS, CT 06269, USA The Daily Campus is UConn's independent, student-run newspaper. We strive to provide the UConn community with fair, accurate, relevant and editorially independent content, to take on student interests as our own interests, and to provide the best possible environment for students to learn, experience and develop skills related to news media. ADVERTISE WITH US InstagramTwitterFacebookSpotifyYoutube Newsletter sign-up Sign in / Join Never miss a beat! Interested in getting blog post updates? Simply click the button below to stay in the loop! Subscribe

In [9]:

```
# Count the occurrences of different entity labels in the document
labels = [x.label_ for x in article.ents]
Counter(labels)
```

Out[9]:

```
Counter({'CARDINAL': 51,
        'ORG': 30,
        'GPE': 29,
        'PERSON': 19,
        'DATE': 17,
        'LOC': 8,
        'TIME': 8,
        'NORP': 4,
        'ORDINAL': 3,
        'WORK_OF_ART': 2,
        'EVENT': 2,
        'PERCENT': 1,
        'FAC': 1,
        'PRODUCT': 1})
```

In [10]:

```
# Get the three most common entity items and their occurrences
items = [x.text for x in article.ents]
Counter(items).most_common(3)
```

Out[10]:

```
[('UConn', 12), ('Huskies', 6), ('80-75', 5)]
```

In [11]:

```
# Display the 30th sentence in the document
sentences = [x for x in article.sents]
print(sentences[30])
```

View this post on Instagram

A post shared by UConn Men's Basketball (@uconnmbb)

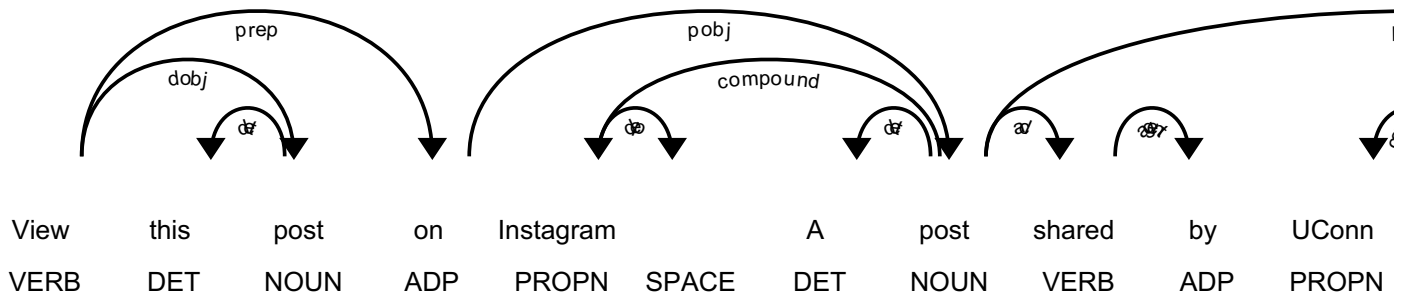
In [12]:

```
# Visualize syntactic dependencies in the 30th sentence
displacy.render(nlp(str(sentences[30])), jupyter=True, style='dep', options={'distance': 70})

# Extract relevant information from the 30th sentence, excluding stop words and punctuation
```

```
[(x, x.pos_, x.lemma_, x.ent_type_) for x in [y
    for y
    in nlp(str(sentences[30]))
    if not y.is_stop and y.pos_ != 'PUNCT']]

# Print entities and their labels in the 25th sentence
dict([(str(x), x.label_) for x in nlp(str(sentences[25])).ents])
```



Out[12]:

```
{'first-half': 'DATE',
 'Karaban': 'ORG',
 'the second half': 'DATE',
 'Tristen Newton': 'ORG'}
```