

This homework explores coreference resolution for the task of timeline generation: for a given biography on Wikipedia, can you extract all of the events associated with the people mentioned and create one timeline for each person? For this homework, be sure you are using the `anlp_spacy2` anaconda environment that we set up in `CorefSetup.ipynb` (that has Spacy 2 installed, along with the `neuralcoref` package).

Within this environment, install the wikipedia package:

```
pip install wikipedia
```

```
In [ ]: import wikipedia
import spacy
import neuralcoref
import re
import operator
```

```
In [ ]: nlp = spacy.load('en_core_web_sm')
coref = neuralcoref.NeuralCoref(nlp.vocab)
nlp.add_pipe(coref, name='neuralcoref')
```

```
In [ ]: biography = wikipedia.page("Harrison Ford", auto_suggest=False)
```

```
In [ ]: doc=nlp(biography.content)
```

Q1. Create a method `extract_timeline` to construct a timeline for the subject of the Wikipedia biography; the timeline should be comprised of events (i.e., verbs) in the biography that the subject participates in and that can be grounded to a year. Your criteria:

- Only include events in sentences that contain a year.
- Only include events that the target of the Wikipedia biography is involved in.
- Only include events where that target person is the subject or direct object of the verb.

Your function should return a dict mapping a sentence where those criteria are satisfied to the date mentioned in that sentence. For example, assume the following to be a biography of the target *Bach*:

Bach is a composer. He was born in 1685. He composed 6 suites for solo cello in 1717. Anna Magdalena Wilcke met him in 1721. Handel was a contemporary and was also born in 1685. Bach's dog was born in 1723.

Your function should return the following dict:

```
{"He was born in 1685": 1685, "He composed 6 suites for solo cello in 1717": 1717, "Anna Magdalena Wilcke met him in 1721": 1721}
```

We exclude the other sentences because:

- "Bach is a composer" does not mention a date.
- Bach is not mentioned in "Handel was a contemporary and was also born in 1685."
- Bach is not the subject or direct object of "Bach's dog was born in 1723"

```
In [ ]: def extract_timeline(doc):

    events={}
    subject="Harrison Ford"

    for chain in doc._.coref_clusters:

        Ford_bool = False
        for mention in chain.mentions:
            if mention.text in (subject):
                Ford_bool = True

        if Ford_bool:
            # is year in the sentence?
            year = re.search("[1-3][0-9]{3}", mention.sent.__str__())
            # is ford dobj or nsubj?
            if year is not None and mention.root.dep_in ('nsubj', 'dobj'):
                events[mention.sent.__str__()] = int(year[0])

    return events
```

Let's print that extracted timeline from oldest to newest events.

```
In [ ]: events=extract_timeline(doc)
for k,v in sorted(events.items(), key=operator.itemgetter(1)):
    print("%s\t%s" % (v, k))
```

```
1960    Ford graduated in 1960 from Maine East High School in Park Ridge, Illi
nois.
1960    Ford soon dropped the "J" and worked for Universal Studios, playing mi
nor roles in many television series throughout the late 1960s and early 1970s,
including Gunsmoke, Ironside, The Virginian, The F.B.I., Love, American Style
and Kung Fu.
1960    Ford began flight training in the 1960s at Wild Rose Idlewild Airport
in Wild Rose, Wisconsin, flying in a Piper PA-22 Tri-Pacer, but at $15 an hour
(equivalent to $128 in 2020), he could not afford to continue the training.
1964    In 1964, after a season of summer stock with the Belfry Players in Wis
consin, Ford traveled to Los Angeles to apply for a job in radio voice-overs.
1964    He was first married to Mary Marquardt from 1964 until their divorce i
n 1979.
1968    He appeared in the western Journey to Shiloh (1968) and had an uncredi
ted, non-speaking role in Michelangelo Antonioni's 1970 film Zabriskie Point a
s an arrested student protester.
1969    French filmmaker Jacques Demy chose Ford for the lead role of his firs
t American film, Model Shop (1969), but the head of Columbia Pictures thought
Ford had "no future" in the film business and told Demy to hire a more experie
nced actor.
1970    ===
Ford began to receive bigger roles in films throughout the late 1970s, includi
ng Heroes (1977), Force 10 from Navarone (1978) and Hanover Street (1979).
1973    In addition to heroic roles, he played morally ambiguous and darker ch
```

characters in films such as the coming-of-age comedy drama *American Graffiti* (1973), the conspiracy thriller *The Conversation* (1974), the survival drama *The Mosquito Coast* (1986), the legal drama *Presumed Innocent* (1990), and the supernatural suspense thriller

1973 Casting director and fledgling producer Fred Roos championed the young Ford and secured him an audition with George Lucas for the role of Bob Falfa, which Ford went on to play in *American Graffiti* (1973).

1974 *The Conversation* (1974) and *Apocalypse Now* (1979); in the latter film, Ford played an army officer named "G. Lucas".

1977 Following the initial phase of his career in bit parts and supporting roles, Ford gained worldwide fame for his starring role as Han Solo in the epic space opera film *Star Wars* (1977), reprising the role in four sequels over the course of the next 42 years.

1980 Ford also proved his versatility throughout the 1980s with dramatic parts in films such as *Witness* (1985), *The Mosquito Coast* (1986), and *Frantic* (1988) as well as the romantic male lead opposite Melanie Griffith and Sigourney Weaver in the comedy-drama *Working Girl* (1988).

1982 Following his leading-man success as Indiana Jones, he played Rick Deckard in Ridley Scott's dystopian science-fiction film *Blade Runner* (1982).

1983 During the June 1983 filming of *Temple of Doom* in London, Ford herniated a disc in his back.

1984 Ford went on to reprise the role throughout the rest of the decade in the prequel *Indiana Jones and the Temple of Doom* (1984), which co-starred Spielberg's future-wife Kate Capshaw, and the sequel *Indiana Jones and the Last Crusade* (1989), which co-starred Sean Connery as Indy's father, Henry Jones Sr.

1985 Ford has portrayed heroic characters in other films such as the thrillers *Witness* (1985), for which he received his only Academy Award nomination, *The Fugitive* (1993), and *Air Force One* (1997), as well as the historical sports drama *42* (2013).

1990 In the 1990s, Ford became the second of five actors to portray Jack Ryan in two films of the film series based on the literary character created by Tom Clancy: *Patriot Games* (1992) and *Clear and Present Danger* (1994), both co-starring Anne Archer and James Earl Jones.

1990 Ford took over the role from Alec Baldwin who had previously played the character in *The Hunt for Red October* (1990).

1990 He also played more straight dramatic roles in *Presumed Innocent* (1990) and *Regarding Henry* (1991) as well as another romantic lead in *Sabrina* (1995), a remake of a classic 1954 film with the same name.

1990 In the late 1990s, Ford started appearing in several critically derided and/or commercially disappointing films that failed to match his earlier successes, including *Six Days, Seven Nights* (1998), *Random Hearts* (1999),

1990 In the mid-1990s, he bought a used Gulfstream II and asked one of his pilots, Terry Bender, to give him flying lessons.

1993 Ford also played leading roles in other action-based thrillers throughout the decade such as the critically acclaimed *The Fugitive* (1993), *The Devil's Own* (1997), and *Air Force One* (1997).

1995 On September 7, 1995, Ford testified before the U.S. Senate Foreign Relations Committee in support of the Dalai Lama and an independent Tibet.

2003 On May 30, 2003, Ford received a star on the Hollywood Walk of Fame.

2004 In 2004, Ford declined a chance to star in the thriller *Syriana*, later commenting that "I didn't feel strongly enough about the truth of the material"

2004 In March 2004, Ford officially became chairman of the Young Eagles program of the Experimental Aircraft Association (EAA).

2005 In July 2005, at the gathering in Oshkosh, Ford agreed to accept the position for another two years.

2008 In 2008, Ford enjoyed success with the release of *Indiana Jones and the Kingdom of the Crystal Skull*, the first *Indiana Jones* film in nineteen years and another collaboration with Lucas and Spielberg.

2008 Ford later said he would like to star in another sequel, "...if it didn't take another 20 years to digest." Other 2008 work included *Crossing Over*, directed by Wayne Kramer.

2009 Ford filmed the medical drama *Extraordinary Measures* in 2009 in Portland, Oregon.

2009 Ford stepped down as program chairman in 2009 and was replaced by Captain Chesley Sullenberger and First Officer Jeff Skiles.

2009 In 2009, Ford signed a petition calling for the release of film director Roman Polanski, who had been arrested in Switzerland in relation to his 1977 charge for drugging and raping a 13-year-old girl.

2011 In July 2011, Ford starred alongside Daniel Craig and Olivia Wilde in the science-fiction/western hybrid film *Cowboys & Aliens*.

2011 Also in 2011, Ford starred in Japanese commercials advertising the video game

2014 During filming on June 11, 2014, Ford suffered what was said to be a fractured ankle when a hydraulic door fell on him.

2015 In 2015, Ford received the Albert R. Broccoli Britannia Award for Worldwide Contribution to Entertainment from BAFTA Los Angeles.

2016 On July 28, 2016, Ford flew the two millionth Young Eagle at the EAA AirVenture convention.

2017 On February 13, 2017, Ford landed an Aviat Husky at John Wayne Airport in Orange County, California, on the taxiway left of runway 20L.

2019 In 2019, Ford had his first voice role in an animated film, as a dog named Rooster in *The Secret Life of Pets 2*.

2019 In 2019, on behalf of Conservation International, Ford gave an impassioned speech during the United Nations' Climate Action Summit in New York on the destruction of the Amazon rainforest and its effect on climate change for the rest of the world.

2020 On April 24, 2020, at the Los Angeles Hawthorne Airport, Ford was piloting a plane when he crossed a runway where another aircraft was landing.

2020 For the first time ever endorsing a political candidate for office, Ford endorsed Joe Biden's 2020 presidential campaign against Donald Trump.

2020 Along with Mark Hamill, Ford worked with the anti-Trump Republican group The Lincoln Project to produce and narrate a 2020 election ad attacking Trump's disparaging of Anthony Fauci.