Assignment

* Add a database component to your website

  * Just focus on the website part, not the JSON API.

* Due March 29th.

* Send me an archive with all code and instructions on what to install and how to run.

* Or host the site somewhere accessible via HTTP:

  * For example on https://www.pythonanywhere.com/

    * Python hosting with support for Flask websites

    * Free for a limited account with one website.

Tool accepts plain text ✔

When Sebastian Thrun started working on self-driving cars at Google in 2007, few people outside of the company took him seriously. "I can tell you very senior CEOs of major American car companies would shake my hand and turn away because I wasn't worth talking to," said Thrun, in an interview with Recode earlier this week.

Tool generates markup ✔

<markup>When <entity class="PERSON">Sebastian Thrun</entity> started working on self-driving cars at Google in <entity class="DATE">2007</entity>, few people outside of the company took him seriously. "I can tell you very senior CEOs of major <entity class="NORP">American</entity> car companies would shake my hand and turn away because I wasn't worth talking to," said <entity class="PERSON">Thrun</entity>, in an interview with <entity class="PERSON">Recode</entity> <entity class="DATE">earlier this week</entity>.</markup>

Tool generates markup ✔

```
>>> import ner
>>> text = open('input.txt').read()
>>> text
'When Sebastian Thrun started working on self-driving cars at Google in 2007, few
people outside of the company took him seriously. "I can tell you very senior CEOs
of major American car companies would shake my hand and turn away because I wasn't
worth talking to," said Thrun, in an interview with Recode earlier this week.\n'
>>> ner.entity_markup(text)
'<markup>When <entity class="PERSON">Sebastian Thrun</entity> started working on
self-driving cars at Google in <entity class="DATE">2007</entity>, few people
outside of the company took him seriously. "I can tell you very senior CEOs of
major <entity class="NORP">American</entity> car companies would shake my hand and
turn away because I wasn't worth talking to," said <entity class="PERSON">Thrun</
entity>, in an interview with <entity class="PERSON">Recode</entity> <entity
class="DATE">earlier this week</entity>.\n</markup>'
>>>
```

Example Solution ✔

```python
def entity_markup(text):
    entities = nlp(text).ents
    starts = {e.start_char: e.label_ for e in entities}
    ends = set([e.end_char for e in entities])
    buffer = io.StringIO()
    for p, char in enumerate(text):
        if p in ends:
            buffer.write('</entity>')
        if p in starts:
            buffer.write('<entity class="%s">' % starts[p])
        buffer.write(char)
    markup = buffer.getvalue()
    return '<markup>%s</markup>' % markup
```

RESTful service ✔

GET request returns some descriptive JSON

```
$> curl http://127.0.0.1:5000/api
{
    "description": "Interface to the spaCy entity extractor",
    "usage": "curl -X POST -d@input.txt http://127.0.0.1:5000/"
}
```

RESTful service ✔

POST request returns JSON with markup

```
$> curl -X POST -d@input.txt http://127.0.0.1:5000/api
{
    "input": "When Sebastian Thrun started working on self-driving cars at Google
in 2007, few people outside of the company took him seriously. \u201cI can tell
you very senior CEOs of major American car companies would shake my hand and turn
away because I wasn\u2019t worth talking to,\u201d said Thrun, in an interview
with Recode earlier this week.",
    "output": "<markup>When <entity class=\"PERSON\">Sebastian Thrun</entity>
started working on self-driving cars at Google in <entity class=\"DATE\">2007</
entity>, few people outside of the company took him seriously. \u201cI can tell
you very senior CEOs of major <entity class=\"NORP\">American</entity> car
companies would shake my hand and turn away because I wasn\u2019t worth talking
to,\u201d said <entity class=\"PERSON\">Thrun</entity>, in an interview with
<entity class=\"PERSON\">Recode</entity> <entity class=\"DATE\">earlier this
week</entity>.</markup>"
}
```
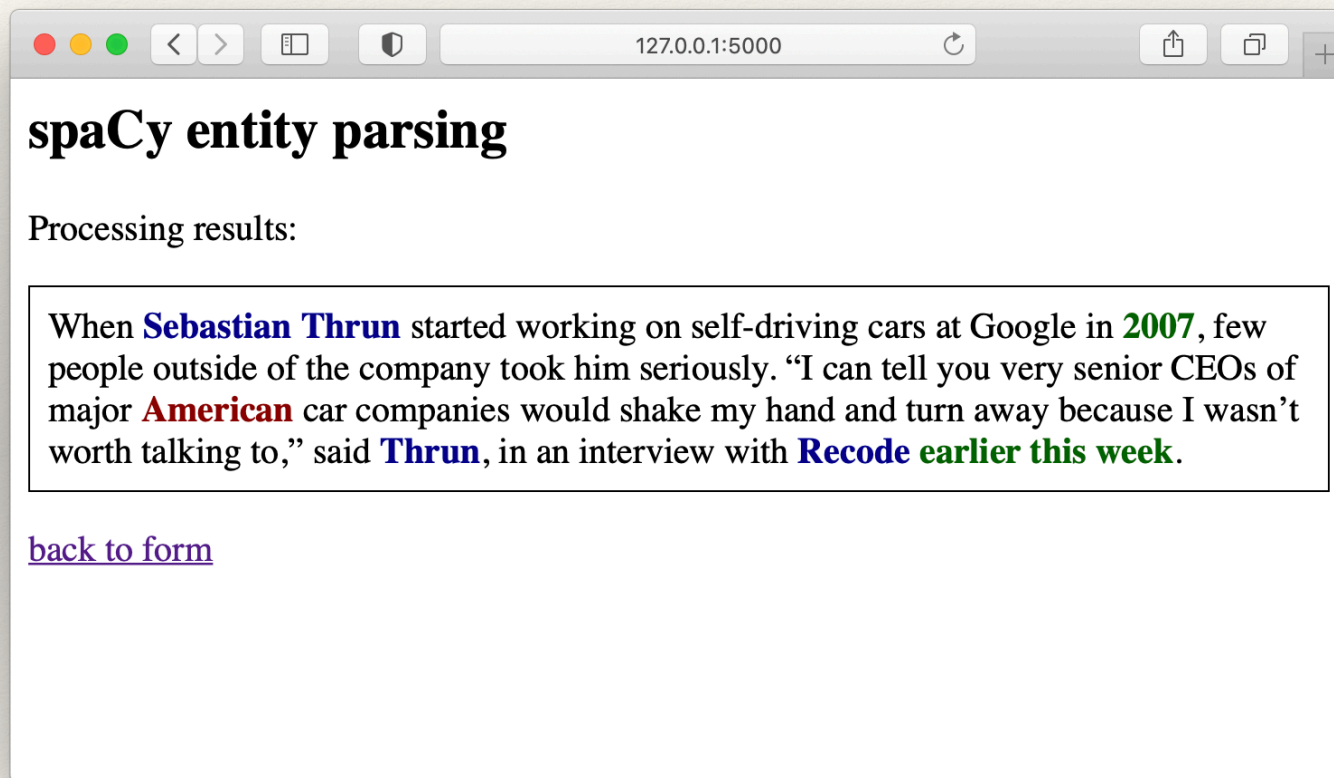
Web Interface ✔

spaCy entity parsing

Replace the text below (or use it) and click submit.

When Sebastian Thrun started working on self-driving cars at Google in 2007, few people outside of the company took him seriously. "I can tell you very senior CEOs of major American car companies would shake my hand and turn away because I wasn't worth talking to," said Thrun, in an interview with Recode earlier this week.

Submit

Web Interface

✔

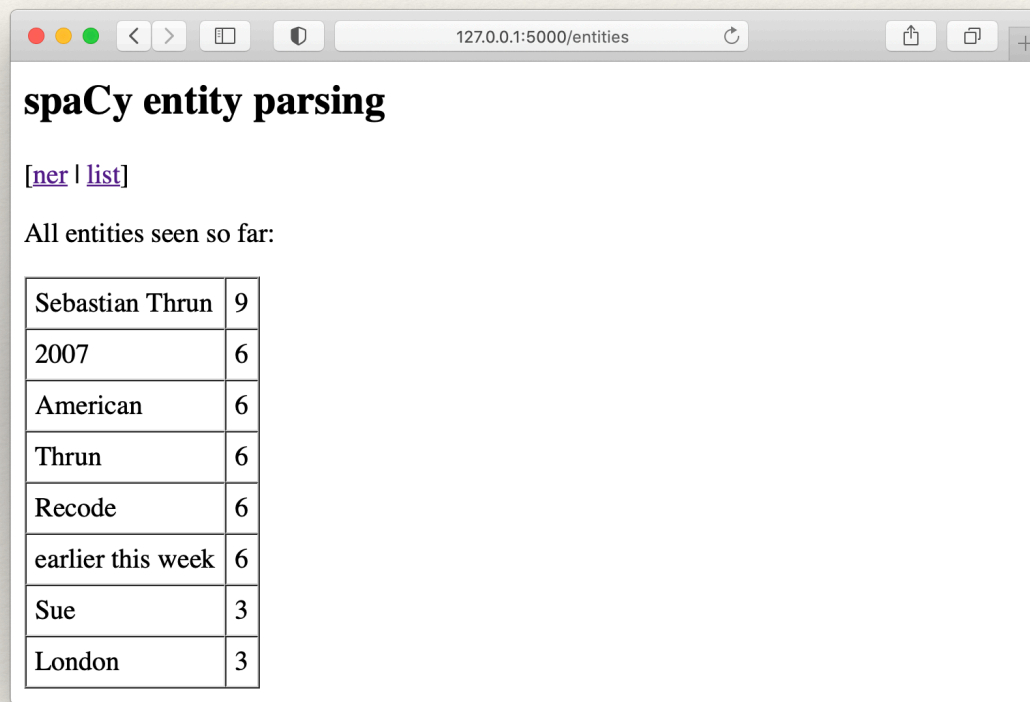**spaCy entity parsing**

Processing results:

When **Sebastian Thrun** started working on self-driving cars at Google in **2007**, few people outside of the company took him seriously. "I can tell you very senior CEOs of major **American** car companies would shake my hand and turn away because I wasn't worth talking to," said **Thrun**, in an interview with **Recode** **earlier this week**.

back to form

Web Interface

spaCy entity parsing

[ner | list]

All entities seen so far:

Sebastian Thrun	9
2007	6
American	6
Thrun	6
Recode	6
earlier this week	6
Sue	3
London	3