

```
In [1]: import re
        from spacy import displacy
```

```
In [2]: [m.start() for m in re.finditer('test', 'test test test test')]
        #[0, 5, 10, 15]
```

```
Out[2]: [0, 5, 10, 15]
```

```
In [3]: txt = "On February 23, 2019, Eminem released a re-issue of The Slim Shady LP. On
            "The New York Times Magazine listed Eminem among hundreds of artists whose
            "reportedly destroyed in the 2008 Universal Studios fire."
```

```
In [4]: GT = [
        {'start': 22, 'end': 28, 'mention_txt': 'Eminem', 'entity_txt': 'Eminem'},
        {'start': 52, 'end': 69, 'mention_txt': 'The Slim Shady LP', 'entity_txt': 'The Sli
        {'start': 89, 'end': 116, 'mention_txt': 'The New York Times Magazine', 'entity_
        {'start': 124, 'end': 130, 'mention_txt': 'Eminem', 'entity_txt': 'Eminem'},
        {'start': 149, 'end': 156, 'mention_txt': 'artists', 'entity_txt': 'Artist'},
        {'start': 204, 'end': 231, 'mention_txt': '2008 Universal Studios fire', 'entity
        ]
```

```
In [5]: model = [
        {'start': 52, 'end': 69, 'mention_txt': 'The Slim Shady LP', 'entity_txt': 'The Sli
        {'start': 93, 'end': 116, 'mention_txt': 'New York Times Magazine', 'entity_txt'
        {'start': 149, 'end': 156, 'mention_txt': 'artists', 'entity_txt': 'Artist (film
        {'start': 209, 'end': 226, 'mention_txt': 'Universal Studios', 'entity_txt': 'Un
        ]
```

```
In [6]: def process(anno, title=''):
        ents = []
        for anno_ele in anno:
            start = anno_ele['start']
            end = anno_ele['end']
            mention = anno_ele['mention_txt']
            entity = anno_ele['entity_txt']

            assert txt[start:end] == mention
            ent = {
                'start': start,
                'end': end,
                'label': entity,
            }
            ents.append(ent)

        instance = {
            "text": txt,
            "ents": ents,
            'title': title,
        }
        print(repr(instance))
        return instance
```

```
In [7]: process(model)
```

```
{'text': 'On February 23, 2019, Eminem released a re-issue of The Slim Shady LP.
On June 25, 2019, The New York Times Magazine listed Eminem among hundreds of ar
tists whose material was reportedly destroyed in the 2008 Universal Studios fir
```

```
e.', 'ents': [{ 'start': 52, 'end': 69, 'label': 'The Slim Shady LP'}, { 'start': 93, 'end': 116, 'label': 'The New York Times Magazine'}, { 'start': 149, 'end': 156, 'label': 'Artist (film)'}, { 'start': 209, 'end': 226, 'label': 'Universal Pictures'}], 'title': ''}
```

```
Out[7]: {'text': 'On February 23, 2019, Eminem released a re-issue of The Slim Shady LP. On June 25, 2019, The New York Times Magazine listed Eminem among hundreds of artists whose material was reportedly destroyed in the 2008 Universal Studios fire.', 'ents': [{ 'start': 52, 'end': 69, 'label': 'The Slim Shady LP'}, { 'start': 93, 'end': 116, 'label': 'The New York Times Magazine'}, { 'start': 149, 'end': 156, 'label': 'Artist (film)'}, { 'start': 209, 'end': 226, 'label': 'Universal Pictures'}], 'title': ''}
```

```
In [8]: GT_pre = {'text': 'On February 23, 2019, Eminem released a re-issue of The Slim Shady LP. On June 25, 2019, The New York Times Magazine listed Eminem among hundreds of artists whose material was reportedly destroyed in the 2008 Universal Studios fire.', 'ents': [{ 'start': 22, 'end': 28, 'label': 'Eminem'}, { 'start': 52, 'end': 69, 'label': 'The Slim Shady LP'}, { 'start': 89, 'end': 116, 'label': 'The New York Times Magazine'}, { 'start': 124, 'end': 130, 'label': 'Eminem'}, { 'start': 149, 'end': 156, 'label': 'Artist'}, { 'start': 204, 'end': 231, 'label': '2008 Universal Studios fire'}], 'title': ''}
```

```
In [9]: model_pre = {'text': 'On February 23, 2019, Eminem released a re-issue of The Slim Shady LP. On June 25, 2019, The New York Times Magazine listed Eminem among hundreds of artists whose material was reportedly destroyed in the 2008 Universal Studios fire.', 'ents': [{ 'start': 52, 'end': 69, 'label': 'The Slim Shady LP'}, { 'start': 93, 'end': 116, 'label': 'The New York Times Magazine'}, { 'start': 149, 'end': 156, 'label': 'Artist (film)'}, { 'start': 209, 'end': 226, 'label': 'Universal Pictures'}], 'title': ''}
```

```
In [10]: colors = {
    'Eminem': "#ff8197",
    'The Slim Shady LP': "#bfeeb7",
    'The New York Times Magazine': "#e4e7d2",
    '2008 Universal Studios fire': "#ff9561",
    'Universal Pictures': 'yellow',
    'Artist': "#bfeeb7",
    'Artist (film)': "#ffeb80",
}
options = {'colors': colors}
```

```
In [11]: displacy.render(GT_pre, style='ent', manual=True, options=options)
```

On February 23, 2019, **Eminem** **Eminem** released a re-issue of **The Slim Shady LP** **The Slim Shady LP** . On June 25, 2019, **The New York Times Magazine** **The New York Times Magazine** listed **Eminem** **Eminem** among hundreds of **artists** **Artist** whose material was reportedly destroyed in the **2008 Universal Studios fire** **2008 Universal Studios fire** .

```
In [12]: displacy.render(model_pre, style='ent', manual=True, options=options)
```

On February 23, 2019, Eminem released a re-issue of **The Slim Shady LP** **The Slim Shady LP** .
On June 25, 2019, The **New York Times Magazine** **The New York Times Magazine** listed

Eminem among hundreds of **artists** **Artist (film)** whose material was reportedly destroyed in the 2008 **Universal Studios** **Universal Pictures** fire.

In []: