

## A.BAŁABAŃSKA PJN

Szanowny Panie Doktorze,

jako, że nie mam planu swojej pracy edytowałam artykuł na temat NLP. Mam nadzieję, że nie będzie to stanowić problemu.

Dziękuję

### Analyzing the structure and effectiveness of news headlines using NLP

- Posted by [Mike Waldron](<https://www.datasciencecentral.com/profile/MikeWaldron>) on June 9, 2016 at 1:30am

\_We wanted to gather and analyze news content in order to look for similarities and differences in the way two journalists write headlines for their respective news articles and blog posts. The two reporters we selected operate in, and write about, two very different industries/topics and have two very different writing styles:\_

- **Finance:** Akin Oyedele of Business Insider, who covers market updates.
  - **Celebrity:** Carly Ledbetter of the Huffington Post, who mainly writes about celebrities.
- ![Akin Oyedele and Carly Ledbetter]([https://66.media.tumblr.com/978efd3b038dee7d0364105ffde292ea/tumblr\\_inline\\_o6elyj52xr1u37g00\\_500.png](https://66.media.tumblr.com/978efd3b038dee7d0364105ffde292ea/tumblr_inline_o6elyj52xr1u37g00_500.png))

### The Approach

- \* Collect news headlines from both of our journalists
- \* Create parse trees from collected headlines (we explain parse trees below!)
- \* Extract information from each parse tree that is indicative of the overall headline structure
- \* Define a simple sequence similarity metric to quantitatively compare any pair of headlines
- \* Apply the same metric to all headlines collected for each author to find similarity
- \* Use K-Means and tSNE to produce a visual map of all the headlines so we can clearly see the differences between our two journalists

### Data

In total we gathered about 700 article headlines for both journalists using the [AYLIEN News API] (<https://t.umblr.com/redirect?z=https%3A%2F%2Fnewsapi.aylien.com%2Fdemo&t=YzYzNmZiYzZjOTlmNTZlZjc1MWM1ZDZhMWRLYmVjMDIzYzdiY2NjMSxKdXNPMVRLVg%3D%3D>) which we then analyzed using Python. If you'd like to give it a go yourself, you can grab the Pickled data files directly from the GitHub repository (link), or by using the data collection notebook we prepared for this project.

>First we loaded all the headlines for Akin Oyedele, then we created parse trees for all **700** of them, and finally we stored them together with some basic information about the headline in the same Python object.

Then using a sequence similarity metric, we compared all of these headlines two by two, to build a similarity matrix.

## MARKDOWN TUTORIAL

`_ text _` - kursywa

`**text**` - pogrubienie

`**_ text _**` - i kursywa i pogrubienie

6 wielkości nagłówków – dajemy tyle znaków # przed nagłówkiem ile ma mieć wielkości

# - największy ##### - najmniejszy

### Header – będzie w stopniu 3, czyli taka średnia wielkość

`[Visit GitHub!](www.github.com)` – sposób dodawania linków

to co jest w brakecie widzi użytkownik

to co jest w nawiasie jest właściwym adresem

`[**Visit GitHub!**]` - można w ten sposób pogrubić

#####The Latest News from the [BBC](www.bbc.com/news) – to będzie w rozmiarze nagłówek 4, widoczne dla użytkownika będzie BBC a jak kliknie to wejdzie na ten adres

! - wykrzyknik pozwala na wstawianie zdjęcia

przykład:

`![Benjamin Bannekat](https://octodex.github.com/images/bannekat.png)`

1. wykrzyknik na początku

2. w brakecie to co widzi użytkownik

3. w nawiasie faktyczny link do obrazka

BLOCKQUOTE – to taki cytat specjalnie umieszczony na stronie żeby był widoczny, zazwyczaj duża czcionka, jakby wyrwany z jakiejś książki. i żeby to zrobić wystarczy wstawić znak > na początku i tyle. ( w każdej linijce). można cały Quote oczywiście dowolnie modyfikować.

LISTY

żeby zrobić listę z ładnymi kropkami trzeba w każdej linijce dać znak \* i koniecznie spację przed słowem UNORDERED LIST

ORDERED LIST znaczy że ma liczby nie kropki i normalnie numerki kropka spacja i tekst

Jak chcemy listę stopniowaną, to po prostu dodajemy jedną, dwie lub więcej spacji

---

W Markdownie jeśli dodajemy każdą linijkę jak w normalnym edytorze to zapisze wszystko w jednej linii, więc żeby zrobić normalne przejście do następnej linii trzeba dać dwie spacje na końcu , co jest niestety niewidoczne. Przykład

```
Do I contradict myself?..
```

```
Very well then I contradict myself,...
```

```
(I am large, I contain multitudes.)
```

