



NLP

Natural Language Processing



Where do we see NLP?

- Email spam filters
- Smart assistants
- Search results
- Predictive text
- Language translations
- Spell checking
- Sentiment Analysis
- Digital Phone Calls
- Social Media Monitoring
- Banning systems

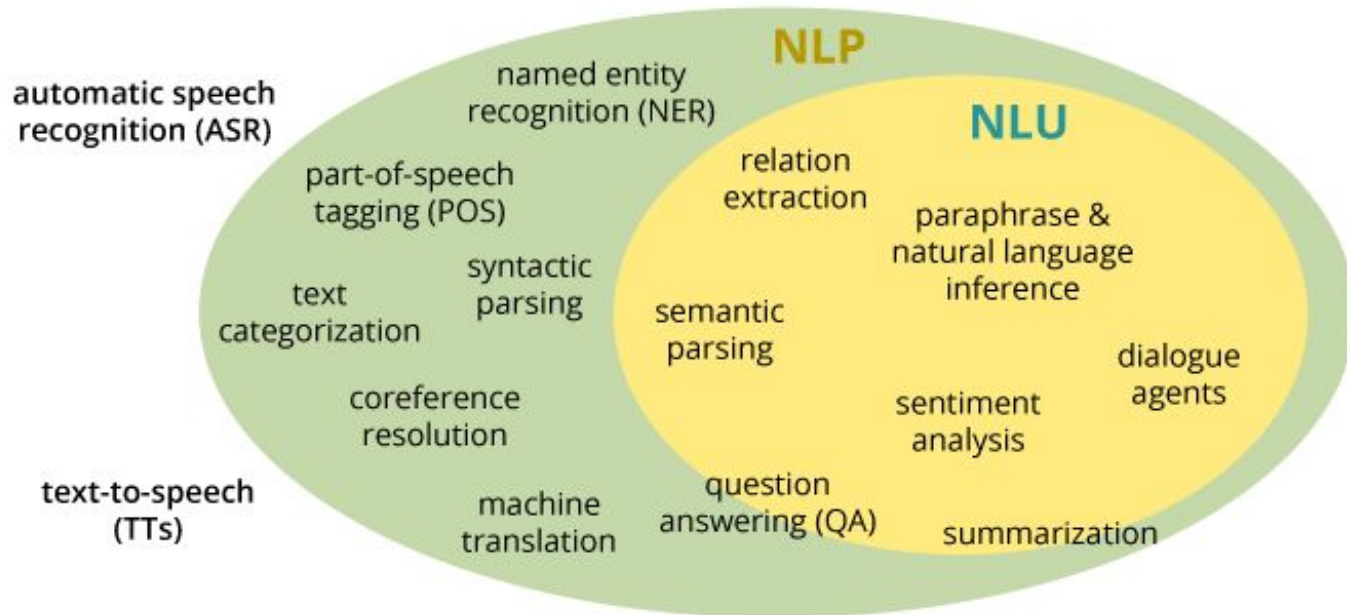
What is Natural Language Processing?

Makes the computer understand human languages.

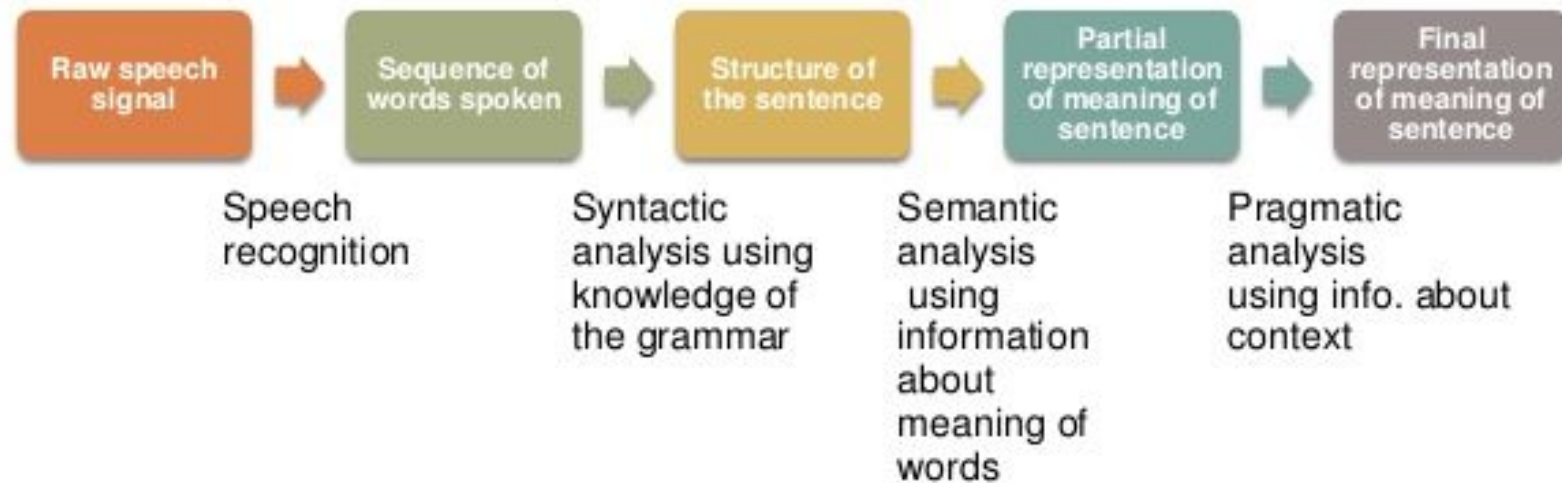
Translating a text to something that a machine can understand.



Terminology: NLU vs. NLP vs. ASR



Natural Language Understanding

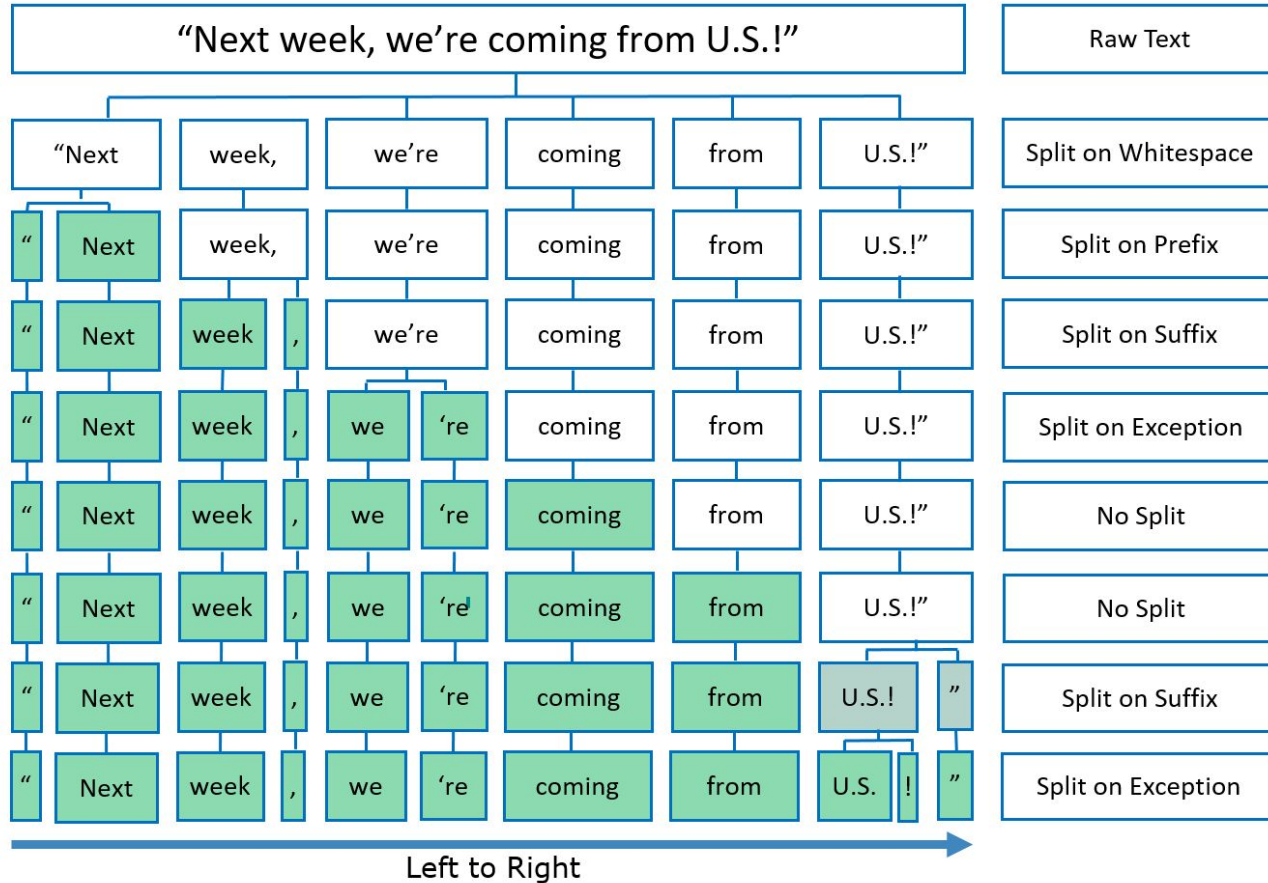


Tokenization

"It gets every word of the text into a new element of a list, not only the unique ones, so the list will have duplicates."

```
['It', 'gets', 'every', 'word', 'of', 'the', 'text',  
'into', 'a', 'new', 'element', 'of', 'a', 'list', ',',  
'not', 'only', 'the', 'unique', 'ones', ',', 'so',  
'the', 'list', 'will', 'have', 'duplicates', '.']
```

Tokenization



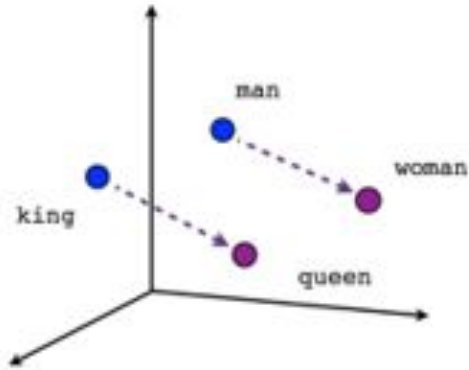
<https://ashutoshtripathi.com/>

Named Entity recognition

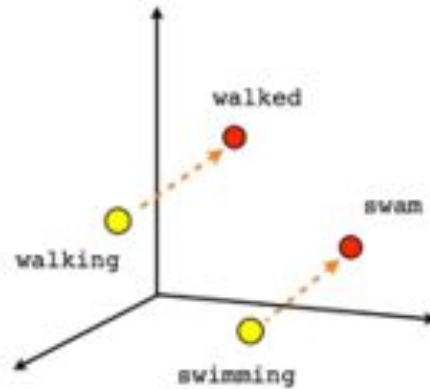
In fact, the Chinese NORP market has the three CARDINAL most influential names of the retail and tech space – Alibaba GPE , Baidu ORG , and Tencent PERSON (collectively touted as BAT ORG), and is betting big in the global AI GPE in retail industry space . The three CARDINAL giants which are claimed to have a cut-throat competition with the U.S. GPE (in terms of resources and capital) are positioning themselves to become the ‘future AI PERSON platforms’. The trio is also expanding in other Asian NORP countries and investing heavily in the U.S. GPE based AI GPE startups to leverage the power of AI GPE . Backed by such powerful initiatives and presence of these conglomerates, the market in APAC AI is forecast to be the fastest-growing one CARDINAL , with an anticipated CAGR PERSON of 45% PERCENT over 2018 - 2024 DATE .

To further elaborate on the geographical trends, North America LOC has procured more than 50% PERCENT of the global share in 2017 DATE and has been leading the regional landscape of AI GPE in the retail market. The U.S. GPE has a significant credit in the regional trends with over 65% PERCENT of investments (including M&As, private equity, and venture capital) in artificial intelligence technology. Additionally, the region is a huge hub for startups in tandem with the presence of tech titans, such as Google ORG , IBM ORG , and Microsoft ORG .

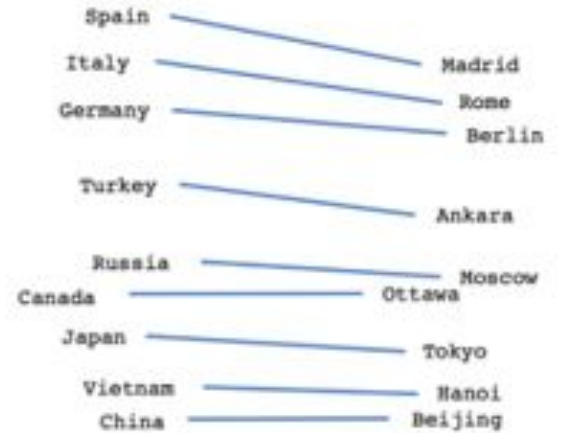
Word vectorisation



Male-Female

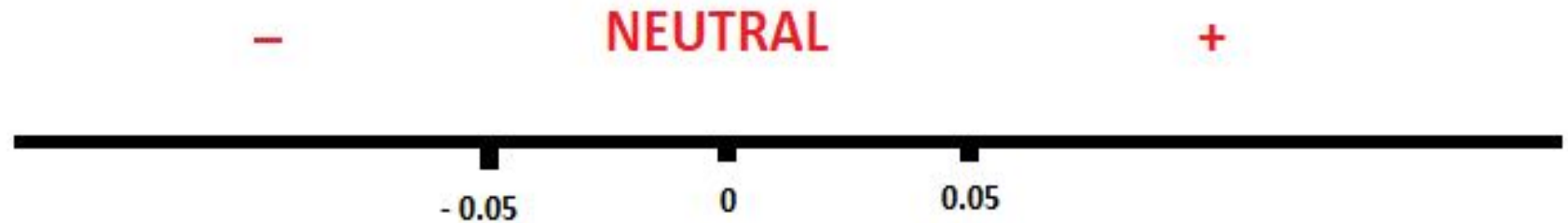


Verb tense



Country-Capital

Sentiment Analysis



positive sentiment: compound score ≥ 0.05

neutral sentiment: (compound score > -0.05) and (compound score < 0.05)

negative sentiment: compound score ≤ -0.05

Lots of sentiment toolkits available (free and paid for)



Natural Language
Analyses with NLTK





Process

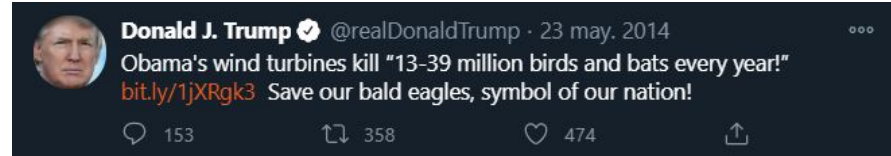
cleaning and summarising text
to get to insight



Example : tweet

First, clean the data

- Special characters
- Numbers
- Stopwords



Stopwords



We have declassified a picture of the wonderful dog (name not declassified) that did such a GREAT JOB in capturing and killing the Leader of ISIS, Abu Bakr al-Baghdadi!

We want to keep only the meaningful words

We have declassified a picture of the wonderful dog (name not declassified) that did such a GREAT JOB in capturing and killing the Leader of ISIS, Abu Bakr al-Baghdadi!

Stopwords

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', 'you're', 'you've', 'you'll', 'you'd', 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', 'that'll', 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', ...]

Now we have a cleaner list

```
['We', 'declassified', 'picture', 'wonderful', 'dog', 'name',  
'declassified', 'GREAT', 'JOB', 'capturing', 'killing', 'Leader', 'ISIS',  
'Abu', 'Bakr', 'al-Baghdadi']
```

we should have changed everything
to lower case first....



Ok ok... now we do have it

```
['declassified', 'picture', 'wonderful', 'dog', 'name',  
'declassified', 'great', 'job', 'capturing', 'killing', 'leader',  
 'isis', 'abu', 'bahr', 'al-baghdadi']
```

But, if we had “dogs” and “dog”, or “captured” and “capturing”, or “kill” and “killing”, we will have different elements even though they mean the same

Solving redundancy

- **Stemming:** process of reducing inflected (or sometimes derived) words to their word stem, base or root form

['declassifi', 'pictur', 'wonder', 'dog', 'name', 'declassifi',
'great', 'job', 'captur', 'kill', 'leader', 'isi', 'abu', 'bakr',
'al-baghdadi']

Solving redundancy

- **Lemmatization:** process of grouping together the inflected forms of a word so they can be analysed as a single item, identified by the word's lemma, or dictionary form

| Word | Root (stem) | Lemma |
|----------|-------------|-------|
| was | wa | be |
| studying | study | study |
| studies | studi | study |

And what else?

Sometimes the number of words is still too big, so we can thin it out a bit, but which words do we keep?

- Most frequent words
- Those who belong to a theme
- ...
- Other ideas?

What information can we look for?

- **Morphological:** about the structure of the text:
 - How long is it?
 - How many phrases does it have?
 - How many words? How long they are? Any typos?
 - How many caps? Percentage of caps in the message?
 - Does it use punctuation? Which kind?

What information can we look for?

- **Syntactic and Grammatical:** about the kind of words used
 - Does it use adjectives? Maybe it's describing something
 - In which verbal tense is narrating?
 - Which pronouns is it using? Is the author talking about itself or someone else?

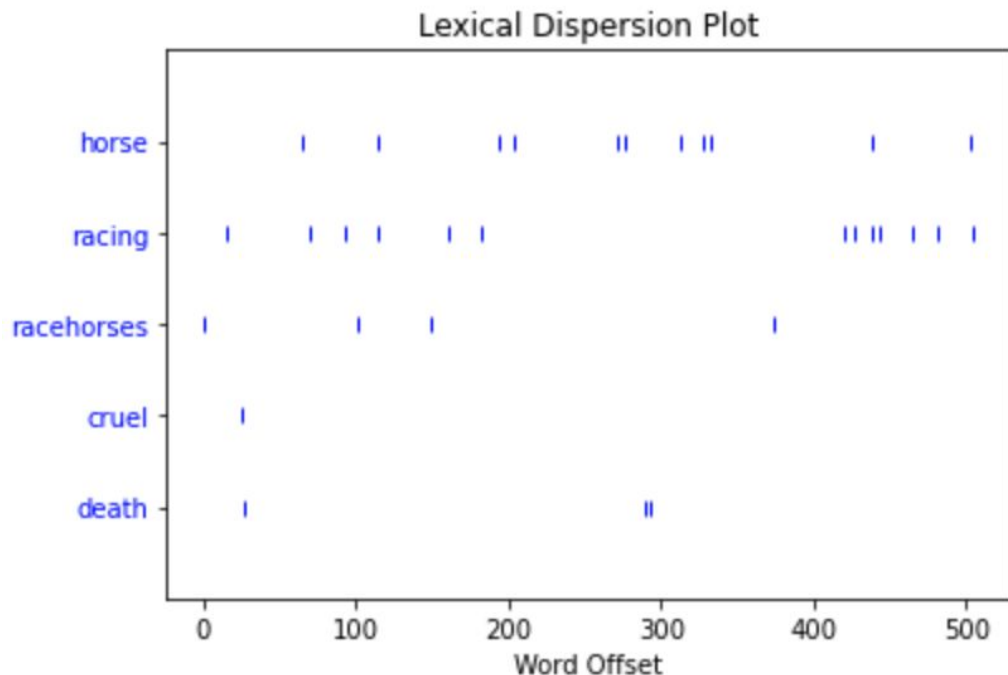
Part of Speech Taggers gives us this analysis

What information can we look for?

- **Buzzwords:** words related to something we are studying
 - Look for insults, criticism
 - Look for dates
 - Look for names - how often are they mentioned?

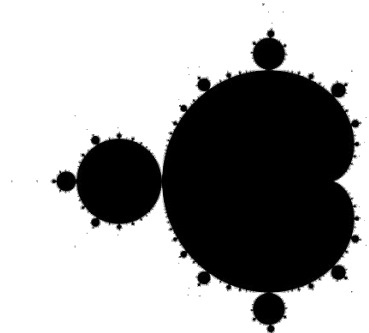
Activity 1

News article text summarisation



Activity 2

Sentiment on product reviews



TextBlob



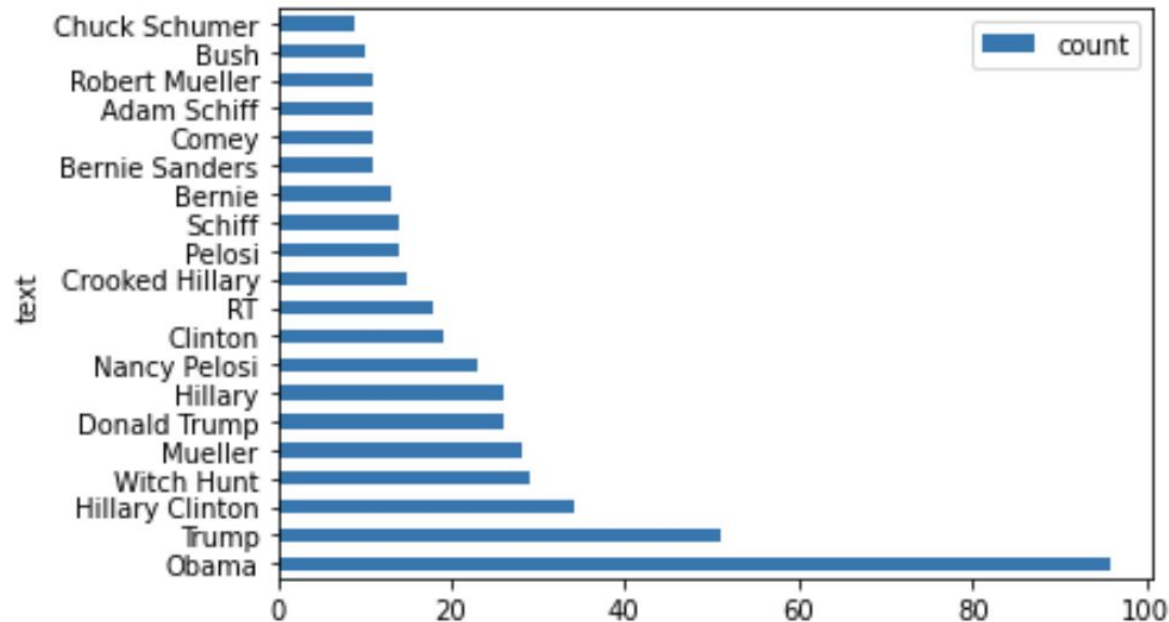
TextBlob is also available as a [conda](#) package. To install with conda, run

```
$ conda install -c conda-forge textblob  
$ python -m textblob.download_corpora
```



Activity 3 - demonstration

Trump tweets - most common named persons





Not for profit

Qualitative analysis of impact



Civil society Monitoring and Evaluation (M&E)

- Interviews
- Focus group discussions
- Social media mentions
- Observations
- Surveys/ questionnaires
- Case studies
- Photography
- Video

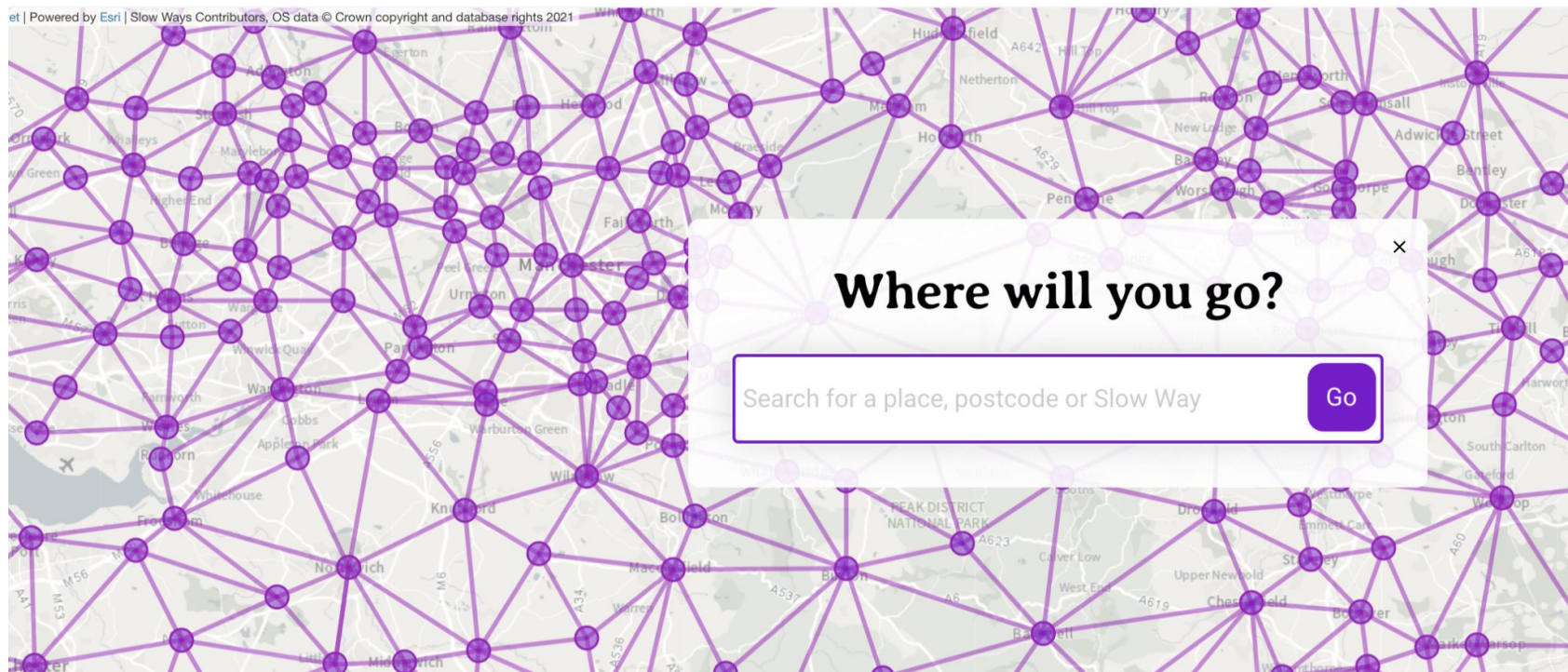




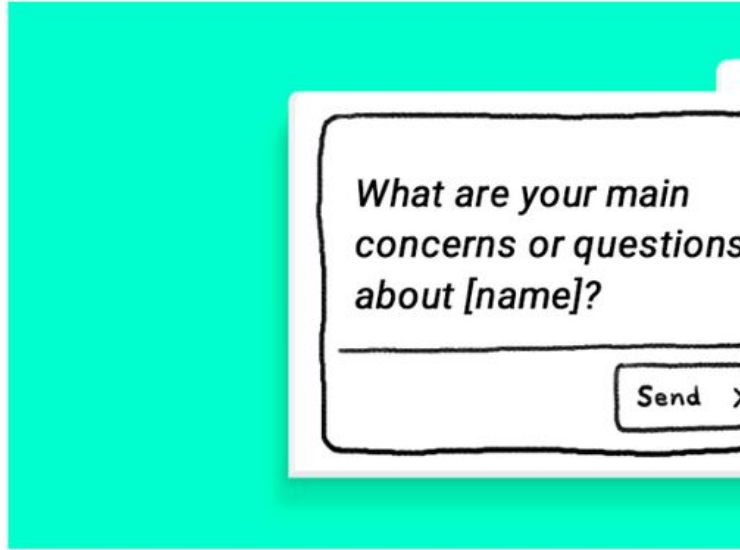
Help create a national walking network

[Home](#)[About](#)[Get Involved](#)[Journey](#)

et | Powered by Esri | Slow Ways Contributors, OS data © Crown copyright and database rights 2021

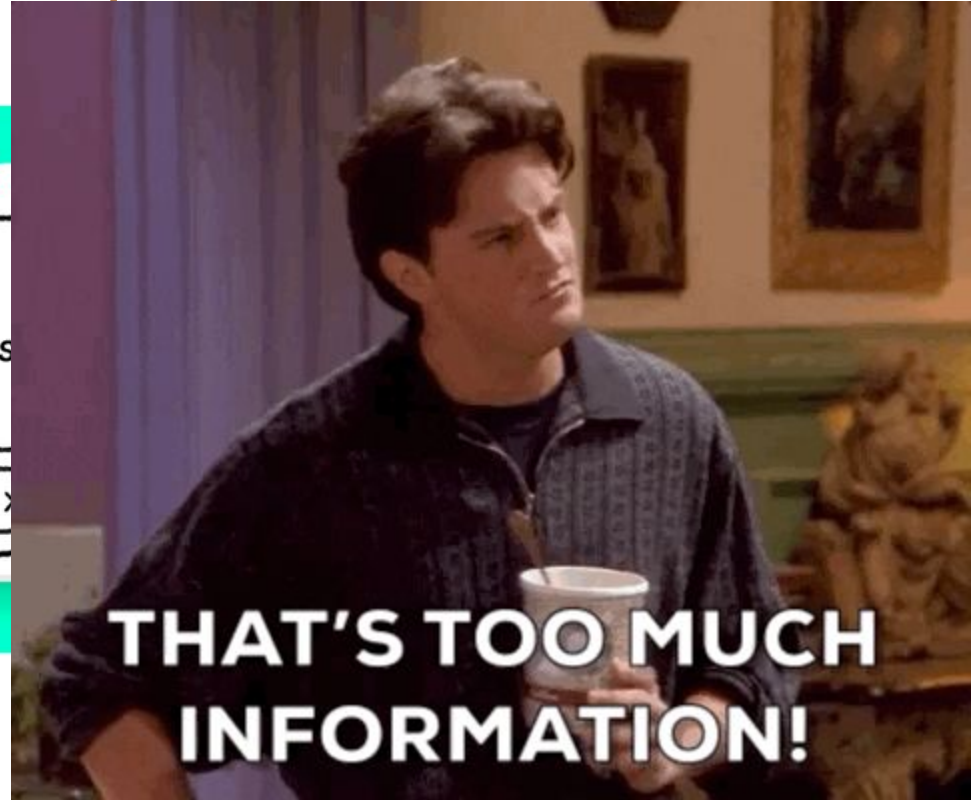


Perceptual Feedback- open ended



What are your main concerns or questions about [name]?

Send >



Perceptual Feedback- open ended

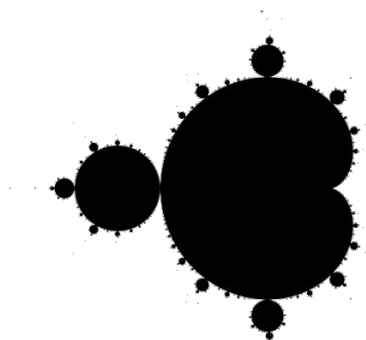
I assume the "seasonal nesting birds" question will be covered in training - pretty much any route will have some kind of seasonal nesting birds somewhere.

Why say livestock when it's only really cows and bulls that are an issue?

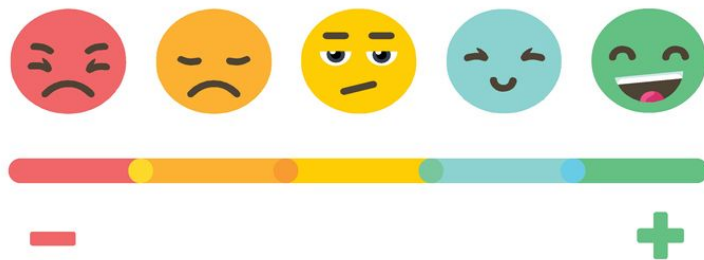
The facilities question is still unclear. It implies not to include any facilities that are in settlements, but why not?

I'm a bit confused by the use of Yes-Maybe-No-Don't Know. If you've walked a route surely maybe and don't know should be redundant?

Sentiment on survey responses - useful?



TextBlob



TextBlob is also available as a [conda](#) package. To install with conda, run

```
$ conda install -c conda-forge textblob  
$ python -m textblob.download_corpora
```



OR social media mentions- trending



377

mentions



11.2K

interactions



1.9M

reach



9K

shares



2.2K

likes

Context of Mentions

Context No. of uses






















| | |
|--------------------|---|
| 96. teemillstore h | 7 |
| 97. time httpstco | 7 |
| 98. colleagues asl | 7 |
| 99. nationalparkc | 7 |
| 100. epping newa | 7 |
| 101. routes time | 7 |
| 102. newaddingto | 7 |
| 103. weekend epp | 7 |
| 104. aslnews weel | 7 |

waysthank comments wareham swanage
walking route
lovely walk
town city walk rumi
rt leeinroyston walk town
adding slow rt debrimfeng road walk
travel fund lovely path jo riley
city villages
slow travel
rt kyukimasa
start newccar
london nationalparkcity
slow waysthank leeinroyston road
proceeds invested journeysdesigned urbangooodcic
walk walk
explore beautiful
mountfield jo
afternoon spent
survey lovely
improve slow
expenses people
mines looked
mines












Sentiment

social media mentions

Latest Mentions

| Post | Site | Performance | Date | Mood |
|---|-------------|---|----------|---|
|  RT @DanRavenEllison: Superb 60km hike from Epping to New Addington... | twitter.com |  100 | 11:18 PM |  |
|  RT @SlowWaysUK: We're slowly adding Slow Ways county t-shirts.... | twitter.com |  100 | 09:49 PM |  |
|  RT @DanRavenEllison: Superb 60km hike from Epping to New Addington... | twitter.com |  100 | 09:30 PM |  |
|  Superb 60km hike from Epping to New Addington over 2.5 days, including two... | twitter.com |  100 | 09:28 PM |  |
|  RT @RunningMM: A lovely day exploring a # slowways route from Warwick-... | twitter.com |  100 | 09:11 PM |  |
|  A lovely [redacted] a # slowways route from Warwick-Kineton. I'll review... | twitter.com |  78 | 08:33 PM |  |
|  @SedgemoorDC folks here could check out the slow ways initiative web [redacted] ... | twitter.com |  100 | 04:12 PM |  |

Top Influencers

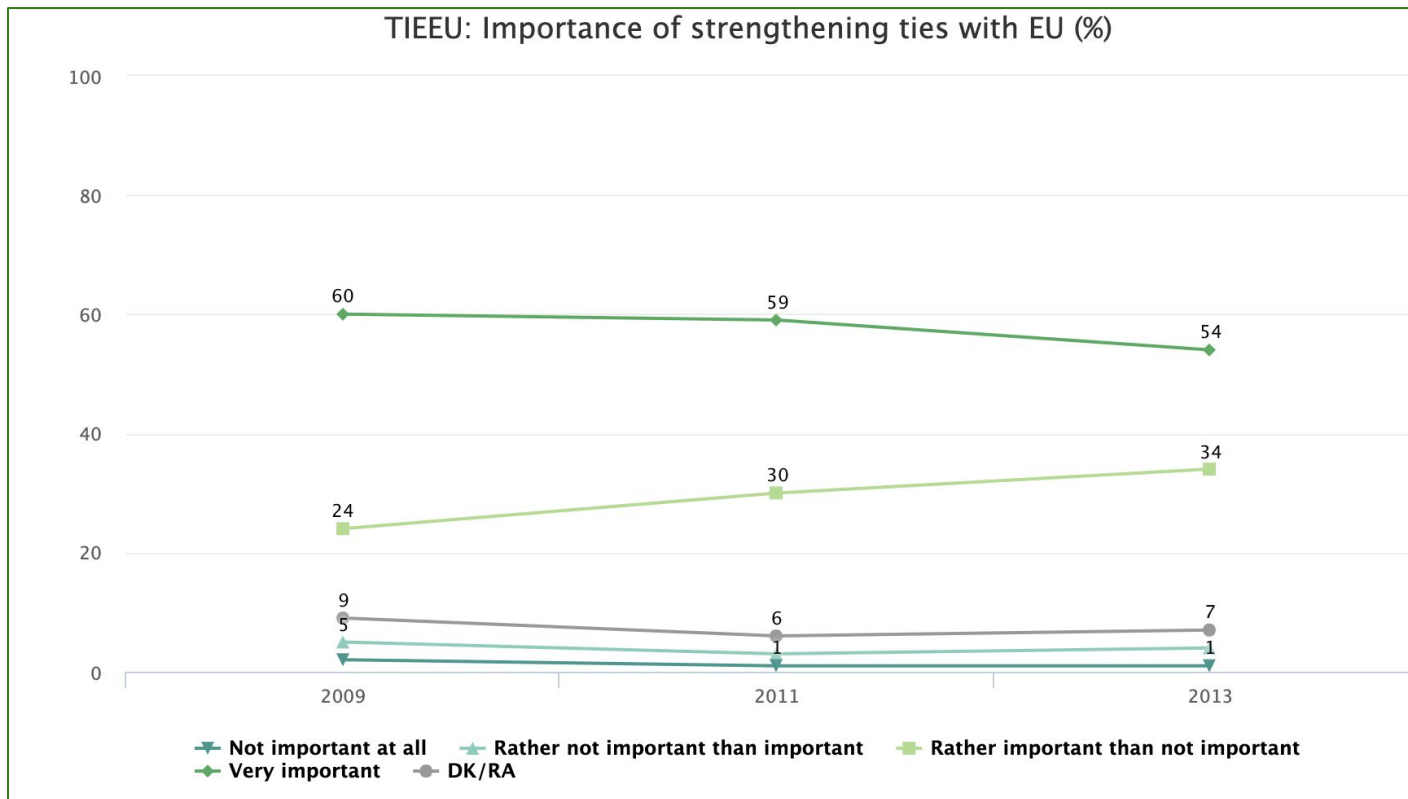
| # of followers | Mentions |
|--|--|
|  Slow Ways @SlowWaysUK 24.7K followers  | 44  |
|  Daniel Raven-Ellison @DanRavenEllison 18.1K followers | 10  |
|  Chris Lee @leeinroyston 1.1K followers  | 6  |
|  Rita Mather @pilgrim_49 110 followers  | 4  |

Related Hashtags

| Hashtag | No. of uses |
|-------------------------|-------------|
| 1. newcastleemlyn | 29 |
| 2. bnw | 29 |
| 3. bnwphotography | 29 |
| 4. blackandwhitephotogr | 29 |
| 5. monochrome | 29 |
| 6. carmarthenshire | 29 |

What other ways can a not for profit organisation track its engagement with their target population?

Perceptual Feedback- heavily structured



Temperature reading- unstructured summary

Submission for the Universal Periodic Review on Georgia

Focused on rights of national minorities in Georgia

Prepared by: Public Movement “Multinational_Georgia



8. The feeling of sometimes being abandoned by the state results in particular from ethnic minorities' under-involvement in public and political life. It is generally admitted, even in the absence of reliable figures, that the ethnic minorities are under-represented in administration, political institutions such as parliament and the government, and also local institutions, even in those regions where they are the majority community. The situation is of course closely linked to the language issue, but that is not the sole factor: some members of ethnic minorities have stated that, quite apart from the question of command of the language, it can be very difficult for a member of an ethnic minority to gain promotion in an administrative service.

9. In general there is an urgent need to strengthen trust between the ethnic minorities on the one hand and the state and majority population on the other.