

Chủ đề: phân tích các bài báo có chứa cụm từ Covid

Website: vnexpress.net

Các câu hỏi có thể trả lời:

1. Độ hot của covid tại các quốc gia trên thế giới.
 - a. Các bài chứa tên các quốc gia
2. Độ hot của covid tại các tỉnh ở việt nam
 - a. Các bài chứa tên các tỉnh thành
3. Tình hình Vaccine ở các nước
 - a. Các bài có chứa cụm từ vaccine
4. Độ hot các vùng lãnh thổ về số ca nhiễm (các châu lục)
 - a. Các bài có chứa tên châu
5. Những ngôi sao có hoạt động liên quan đến covid
 - a. Các tên riêng có trong title
6. Phân tích các cụm từ đứng trước chữ “mắc”, “nhiễm”

Link:

https://timkiem.vnexpress.net/?q=Covid&media_type=all&fromdate=0&todate=0&latest=&category_code=&search_f=title.tag_list&date_format=all&page=2

GOAL:

- 0 Tạo các mảng chứa tên các nước, tình thành (2 nhóm)
- 1 detect các tên riêng có trong title
- 2 Xem các tên riêng đó nằm trong nhóm nào (quốc gia, thành phố, nước vn, celeb)
Tên riêng người ko nằm trong 2 nhóm trên
- 3 Tạo biểu đồ dựa trên số lượng các tên riêng thu được từ các nhóm