



QUÁ TRÌNH THU THẬP VÀ XỬ LÝ DỮ LIỆU

Thu thập dữ
liệu web

Văn bản
gốc

Loại bỏ
stopword

Loại bỏ các dấu
cách, các đoạn
mã, kí tự dư thừa

Tách câu, tách từ

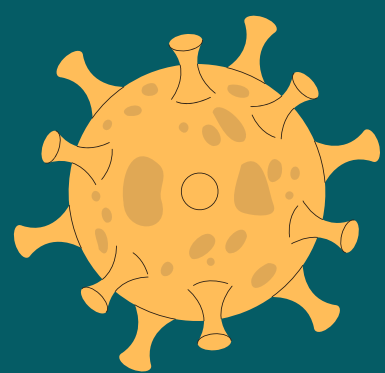
Đồ thị hóa văn bản,
tính trọng số cạnh

Áp dụng TextRank,
sắp xếp từ khóa
theo trọng số

Trích rút
từ khóa

Văn bản
tóm tắt

XÂY DỰNG BẢNG TIN RÚT GỌN VỀ DỊCH BỆNH COVID-19



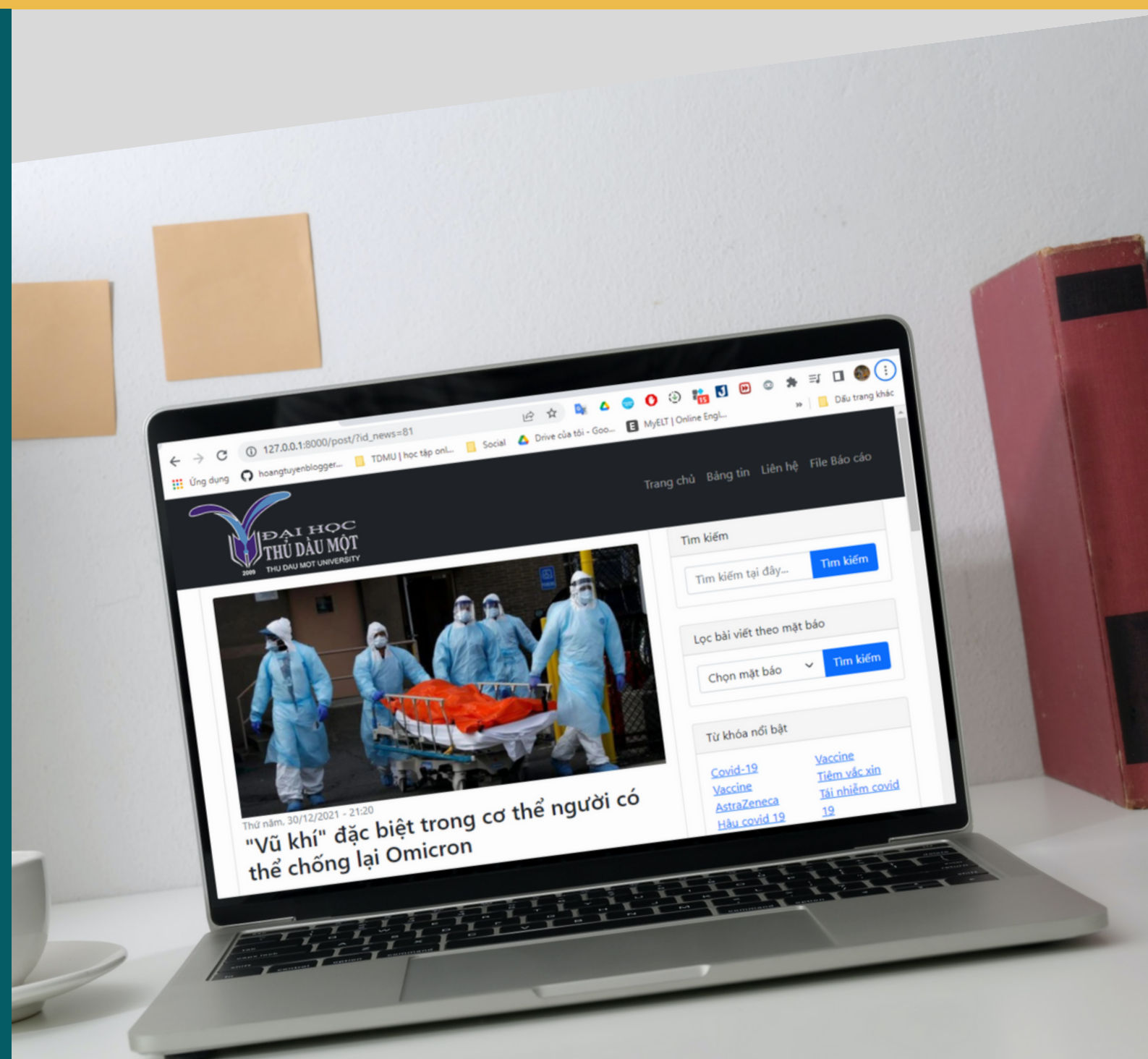
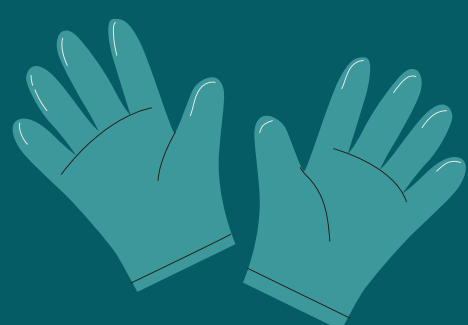
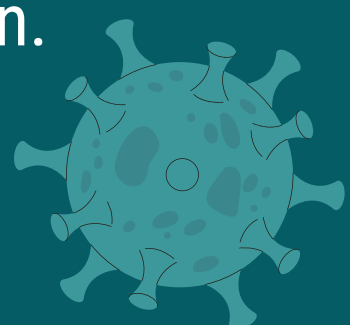
BẢNG KỸ THUẬT TEXT MINING

Hoàng Kim Tuyến - D18HT01

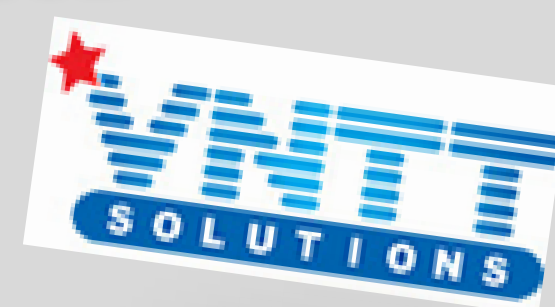
Ngành Hệ Thống Thông Tin

Khóa 2018-2022

Đề tài đã nghiên cứu và thực hiện các phương pháp thu thập dữ liệu văn bản web, trích rút từ khóa và tóm tắt văn bản, áp dụng vào các bài báo về dịch bệnh Covid-19, xây dựng một bản tin Covid-19 với nội dung ngắn gọn, giúp người đọc tiết kiệm thời gian và nắm được nội dung chính của bài báo, giúp cho việc tìm hiểu và tổng hợp thông tin được dễ dàng và thuận tiện hơn.



SPONSORED BY



<https://et.tdmu.edu.vn>



<https://www.facebook.com/VienKTCN.TDMU>