

# PHÂN TÍCH YÊU CẦU CHỨC NĂNG HỆ THỐNG

Input: Bài báo khoa học định dạng: .pdf, đoạn text. Bài báo khoa học gồm có text, biểu đồ, hình ảnh.

Output: Phần trăm mức độ sao chép.

Hilight những đoạn bị xem là có sao chép

Link dẫn tới những tài liệu nguồn.

Trích xuất báo cáo.

Cách thức nhập input: copy đoạn text dán vào khung tìm kiếm

upload file từ thiết bị.

URL tới bài báo.

Kho liệu dùng để so sánh là những bài báo khoa học trên internet hoặc được lưu trữ cục bộ.

Chức năng chính:

So sánh input với kho dữ liệu cục bộ.

So sánh input với kho dữ liệu trên internet.

Chi tiết:

Tìm kiếm từ kho dữ liệu trên Internet	
<b>Pre-Condition:</b>	<ul style="list-style-type: none"><li>• Actor: Người dùng</li><li>• Điều kiện: Người dùng đang ở 1 trang bất kì của ứng dụng. Ứng dụng đang rồi.</li></ul>
<b>Trigger:</b>	<ul style="list-style-type: none"><li>• Người dùng nhấp vào biểu tượng [Check online] trên màn ứng dụng</li></ul>
<b>Basic flow:</b>	<ul style="list-style-type: none"><li>• Hệ thống hiển thị các phương thức nhập input cho người dùng lựa chọn</li><li>• Người dùng nhập input</li><li>• Người dùng nhấn nút [Check]</li><li>• Hệ thống tiến hành xử lí, sau đó, trả về kết quả.</li></ul>
<b>Alternative flow:</b>	<ul style="list-style-type: none"><li>• Hủy tiến trình đang chạy:<ul style="list-style-type: none"><li>▪ Nếu hệ thống xử lí quá lâu (do input quá dài hay kết nối internet kém), người dùng nhấn nút [Cancel] bên cạnh nút</li></ul></li></ul>

<p>[Check], hệ thống sẽ hủy toàn bộ tiến trình đang xử lí, các thông tin input đều bị xóa.</p> <ul style="list-style-type: none"> <li>▪ Hệ thống đưa người dùng về màn hình nhập input ban đầu.</li> </ul> <ul style="list-style-type: none"> <li>• Input rỗng: <ul style="list-style-type: none"> <li>▪ Người dùng chưa nhập input, nút [Check] sẽ bị “Disable”.</li> </ul> </li> </ul>
<p><b>Business rule:</b></p> <ul style="list-style-type: none"> <li>• Input nhập vào không quá 2000 từ.</li> </ul>
<p>Appendix:</p> <div style="text-align: center;"> <p><b>Mockup Screen</b></p> </div>

Tìm kiếm từ kho dữ liệu cục bộ	
<b>Pre-Condition:</b>	<ul style="list-style-type: none"> <li>• Actor: Người dùng</li> <li>• Điều kiện: Người dùng đang ở 1 trang bất kì của ứng dụng. Ứng dụng đang rồi.</li> </ul>
<b>Trigger:</b>	<ul style="list-style-type: none"> <li>• Người dùng nhấp vào biểu tượng [Check offline] trên màn ứng dụng</li> </ul>
<b>Basic flow:</b>	<ul style="list-style-type: none"> <li>• Hệ thống hiển thị các phương thức nhập input cho người dùng lựa chọn</li> <li>• Người dùng nhập input</li> <li>• Người dùng chỉ đường dẫn đến kho dữ liệu cục bộ trên máy.</li> </ul>

- Người dùng nhấn nút [Check]
- Hệ thống tiến hành xử lý, sau đó, trả về kết quả.

#### Alternative flow:

- Hủy tiến trình đang chạy:
  - Nếu hệ thống xử lý quá lâu (do input quá dài hay kết nối internet kém), người dùng nhấn nút [Cancel] bên cạnh nút [Check], hệ thống sẽ hủy toàn bộ tiến trình đang xử lý, các thông tin input đều bị xóa.
  - Hệ thống đưa người dùng về màn hình nhập input ban đầu.
- Input rỗng:
  - Người dùng chưa nhập input, nút [Check] sẽ bị “Disable”.

#### Business rule:

- Input nhập vào không quá 2000 từ.

#### Appendix:

##### Mockup Screen

#### Các trường hợp xử lý:

Không trích dẫn nguồn:

**Copy toàn văn:** Nghĩa là trong input có chứa những đoạn văn bản giống 100% với một hoặc nhiều tài liệu trong kho dữ liệu

Với trường hợp này, hệ thống cần chỉ ra đúng đoạn text được coi là sao chép (highlight). Đồng thời, phải chỉ nguồn bị sao chép dưới dạng liên kết đến tài liệu nguồn.

**Copy toàn văn nhưng có thay đổi 1 số từ:** giống như sao chép toàn văn nhưng người viết đã thay đổi 1 vài từ trong câu như chủ ngữ, tân ngữ, mạo từ... để câu văn khác đi 1 chút nhưng không thay đổi về ý nghĩa.

Với trường hợp này, hệ thống cần tiến hành so sánh từ câu. Nếu câu nào có 70% trở lên số từ trong câu giống với 1 các từ trong 1 câu nào đó trong kho dữ liệu thì đánh dấu câu đó là đạo văn.

**Viết lại ý của người khác:** dựa vào ý của người khác nhưng diễn đạt lại theo cách khác, bằng từ ngữ khác.

Với trường hợp này, hệ thống cần phân rã các câu trong input, mỗi câu lấy ra những từ khóa của câu. Tìm những từ khóa đó trong kho dữ liệu (kiểm tra luôn các từ đồng nghĩa)

**Trích dẫn lại bài viết của chính tác giả:** trường hợp tác giả trích dẫn lại chính bài viết của mình, thì vẫn xử lý như đạo văn của người khác.

### **Kết quả trả về gồm:**

Đánh giá mức độ sao chép: hệ thống cần tính toán tỉ lệ giữa độ dài phần bị coi là sao chép với độ dài của input (độ chia nhỏ nhất là từ), từ đó đưa ra tỉ lệ % mức độ sao chép của input và hiển thị kết quả cho người dùng.

Highlight kết quả: đối với những trường hợp đạo văn khác nhau sẽ qui định highlight những màu khác nhau để tiện trong việc đánh giá kết quả. Ví dụ, đối với những đoạn bị cho là copy toàn văn thì highlight màu đỏ, đạo ý thì highlight màu xanh lá... và trong hiển thị kết quả phân tích, t cũng chú thích màu gì có ý nghĩa gì để người dùng tiện kiểm tra.

Link dẫn đến tài liệu nguồn: là liên kết đến tài liệu nguồn trên mạng hoặc đường dẫn tuyệt đối đến tài liệu nguồn nếu tài liệu nguồn được lưu trữ local.

Xuất báo cáo kết quả: ứng dụng cho phép người dùng chọn định dạng báo cáo tải về là .pdf . Người dùng nhấn nút [Export] để tải báo cáo về máy.

Database.

**Result:**

42%

**1.**

"Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla quam velit, vulputate eu pharetra nec, mattis ac neque. Duis vulputate commodo lectus, ac blandit elit tincidunt id. Sed rhoncus, tortor sed eleifend tristique, tortor mauris molestie elit, et lacinia ipsum quam nec dui. Quisque nec mauris sit amet elit iaculis pretium sit amet quis magna. Aenean velit odio, elementum in tempus ut, vehicula eu diam."

**Source:** [cpmputertch.uk/mlp.pdf](http://cpmputertch.uk/mlp.pdf)**Rank:**  12%**Detail:****2.**

"Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla quam velit, vulputate eu pharetra nec, mattis ac neque. Duis vulputate commodo lectus, ac blandit elit tincidunt id. Sed rhoncus, tortor sed eleifend tristique, tortor mauris molestie elit, et lacinia ipsum quam nec dui. Quisque nec mauris sit amet elit iaculis pretium sit amet quis magna. Aenean velit odio, elementum in tempus ut, vehicula eu diam."

**Source:** [cpmputertch.uk/mlp.pdf](http://cpmputertch.uk/mlp.pdf)**Rank:**  30%**Detail:**