

1. Giới thiệu chung

Hội thi Giải pháp ứng dụng trí tuệ nhân tạo năm 2022 Chủ đề cuộc thi AI-Challenge lần thứ III năm 2022 là " *Truy vấn sự kiện từ dữ liệu thị giác (Event Retrieval from Visual Data)*).

Cuộc thi được tổ chức với sự phối hợp của **Sở Thông tin và Truyền thông Tp. Hồ Chí Minh** và **Đại học Quốc gia Tp. Hồ Chí Minh, Sở Khoa học và Công nghệ, Hội Tin học, Thành Đoàn TP. Hồ Chí Minh, Hiệp hội Phần mềm và Dịch vụ Công nghệ thông tin Việt Nam (VINASA)**.

Trong cuộc thi này, mục tiêu đặt ra là tìm kiếm đoạn video sự kiện với thông tin biết trước. Cụ thể, ứng với mỗi truy vấn cho biết trước thông tin một trích đoạn ngắn hoặc một hình ảnh, mục tiêu là tìm đoạn video chứa thông tin đó trong kho dữ liệu video tin tức của các báo và đài truyền hình tại Việt Nam.

Vấn đề được đặt ra giúp giải quyết bài toán tìm kiếm nội dung video trong bối cảnh dữ liệu video ngày càng lớn, được cập nhật thêm qua từng ngày và còn nhiều thách thức so với tìm kiếm nội dung văn bản.

Việc tìm kiếm sự kiện từ video là bài toán khó, có nhiều thử thách. Yêu cầu đặt ra trong cuộc thi không phải là một giải pháp tự động hoàn toàn để tìm kiếm sự kiện theo các mô tả cho trước, mà mỗi đội cần phát triển một ứng dụng phần mềm cho phép hỗ trợ người dùng trong quá trình tìm kiếm đoạn video đáp ứng đúng yêu cầu truy vấn một cách tiện dụng, nhanh chóng và chính xác.

2. Thể thức cuộc thi:

Thể thức của cuộc thi được kế thừa từ hai cuộc thi quốc tế là Video Browser Showdown – VBS (<https://videobrowsershowdown.org/>) và Lifelog Search Challenge – LSC (<http://lsc.dcu.ie/>).

Trong ban tổ chức và ban giám khảo của cuộc thi AI Challenge 2022 có các chuyên gia quốc tế là người sáng lập cũng như thành viên lâu năm của hai cuộc thi quốc tế này.

Video Browser Showdown (VBS) là cuộc thi thường niên về tìm kiếm video, trong đó các nhà nghiên cứu trên thế giới trình diễn trực tiếp và đánh giá tính hiệu quả của các công cụ truy vấn video trên một tập dữ liệu chung đã được công bố trước.

Lifelog Search Challenge (LSC) được xây dựng dựa trên hình thức của cuộc thi Video Browser Showdown, tập trung vào việc truy vấn hình ảnh (tĩnh) từ kho dữ liệu hình ảnh nhật ký sinh hoạt hằng ngày.

Dưới đây là một số hình ảnh giới thiệu buổi thi trực tiếp của cuộc thi Video Browser Showdown năm 2022, được tổ chức trong khuôn khổ hội nghị MMM 2022 tại Việt Nam vào đầu tháng 6 năm 2022.



3. Nội dung cuộc thi

Yêu cầu trong cuộc thi AI Challenge lần thứ III năm 2022 là phát triển một ứng dụng phần mềm hỗ trợ người dùng tìm kiếm một sự kiện trong kho dữ liệu video tin tức theo yêu cầu truy vấn cụ thể.

Yêu cầu truy vấn thuộc dạng Known-Item Search (KIS, tạm dịch là “*tìm kiếm đối tượng được mô tả trước*”), được thể hiện theo một trong hai dạng cụ thể sau:

- **Yêu cầu truy vấn dạng văn bản (Textual KIS):** Ban giám khảo cung cấp mô tả bằng ngôn ngữ tự nhiên của một sự kiện. Các đội dự thi cần tìm ra chính xác đoạn video của sự kiện này. Đoạn mô tả có thể gồm nhiều ý, nhiều câu văn.
 - Ở **vòng sơ tuyển**: nội dung đoạn mô tả được *cung cấp sẵn, trọn vẹn* để các đội dự thi có thể tìm kiếm video theo mô tả,
 - Ở **vòng chung kết**: nội dung đoạn mô tả sẽ được *cung cấp dần dần* trong thời gian dành cho câu truy vấn. Nếu đội dự thi tự tin vào kết quả tìm kiếm video từ những gợi ý ban đầu của đoạn mô tả, đội dự thi có thể nộp kết quả ngay để có thể được điểm rất cao cho câu truy vấn này (nếu kết quả đúng). Tuy nhiên, đội dự thi có thể thận trọng để chờ thêm các thông tin mô tả về sự kiện để kiểm chứng kết quả tìm được.
- **Yêu cầu truy vấn dạng hình ảnh hay video (Video KIS):** thay vì mô tả bằng ngôn ngữ tự nhiên, các đội dự thi sẽ xem một đoạn video ngắn (không quá 20 giây) được trích ra từ một sự kiện nào đó trong kho dữ liệu video đã cho. Các đội không được phép chụp ảnh, ghi hình đoạn video này để đưa vào công cụ của mình mà phải tìm cách phù hợp để diễn tả yêu cầu tìm kiếm, ví dụ như mô tả bằng ngôn ngữ tự nhiên nội dung video hay vẽ lại bối cảnh mình nhìn thấy... Dạng truy vấn này chỉ được áp dụng ở **vòng chung kết** của cuộc thi.

4. Dữ liệu được cấp cho các đội dự thi:

Ban tổ chức cung cấp cho các đội dự thi:

- Video tin tức (ở định dạng MP4)

- Các keyframe trong mỗi video tin tức (ở định dạng JPG).
- Metadata của các keyframe, bao gồm:
 - Vector đặc trưng của mỗi keyframe được rút trích bằng CLIP (mô hình ViT-B/16)
 - Danh sách các khái niệm (concept) được cung cấp bằng kết quả của mô hình phát hiện vật thể với kiến trúc Faster RCNN được huấn luyện trên tập Open Images V4. Số lượng vật thể tối đa mỗi bức ảnh là 100 thuộc 600 loại vật thể của tập dữ liệu Open Images. Chi tiết mô hình cụ thể tại :
https://tfhub.dev/google/faster_rcnn/openimages_v4/inception_resnet_v2/1)
 - Các mô tả (description, nếu có) của video.

Ban tổ chức **dự kiến** cung cấp cho các đội dự thi dữ liệu theo các mốc thời gian và khối lượng dữ liệu như sau:

- 01/10/2022: video và metadata của khoảng 100 giờ video tin tức để các đội làm quen với bài toán.
- 15/10/2022: thêm video và metadata của khoảng 100 giờ video tin tức. Dữ liệu của 200 giờ video đã cung cấp được dùng cho vòng sơ tuyển của cuộc thi
- Đầu tháng 11/2022: thêm video và metadata của khoảng 100 giờ video tin tức. Toàn bộ dữ liệu của 300 giờ video đã cung cấp được dùng cho vòng chung kết của cuộc thi.

5. Một số gợi ý cho các đội khi xây dựng giải pháp cho cuộc thi

- Cuộc thi không yêu cầu một giải pháp tự động tìm kiếm video theo mô tả mà chỉ cần phát triển công cụ hỗ trợ người dùng tìm kiếm video theo mô tả. Do đó, các đội dự thi nên tập trung vào việc xây dựng giao diện tương tác hiệu quả, tự nhiên, dễ sử dụng cho việc truy vấn.
- Người dùng hệ thống sẽ quan sát các kết quả mà công cụ tìm thấy, từ đó có thể chọn và quyết định kết quả nào phù hợp nhất với yêu cầu của câu truy vấn đề nộp.

- Do khối lượng dữ liệu hình ảnh/video khá lớn, các đội nên lưu ý vấn đề về quản lý nội dung, thao tác đọc dữ liệu và hiển thị kết quả truy vấn trên công cụ phần mềm của mình.
- Các đội dự thi có thể sử dụng metadata (gồm các nội dung mô tả - nếu có, các vector đặc trưng CLIP hay các khái niệm/đối tượng được rút trích sẵn từ các keyframe) mà ban tổ chức đã cung cấp. Tuy nhiên, các đội **không bắt buộc** phải sử dụng metadata này mà có thể tự xử lý theo cách riêng của mình. Ngoài ra, các đội cũng có thể **xử lý thêm** các frame khác trong video (nếu cần thiết).

6. Quy trình triển khai

6.1 Công bố hướng dẫn về chủ đề cuộc thi

Ban tổ chức công bố hướng dẫn về chủ đề cuộc thi qua email của đại diện đội dự thi và trên website cuộc thi: <https://aichallenge.hochiminhcity.gov.vn/>

Bên cạnh nội dung hướng dẫn chung về cuộc thi, Ban tổ chức sẽ cung cấp dữ liệu của khoảng 100 giờ video (kèm theo metadata) để các đội **làm quen với bài toán và dữ liệu**.

Ngoài ra, Ban tổ chức còn cung cấp **website thử nghiệm** để các đội làm quen cách mô tả của các câu truy vấn (dạng Textual KIS) và đánh giá thử nghiệm kết quả.

6.2 Vòng sơ tuyển

Ban tổ chức và Ban giám khảo sẽ công bố các câu truy vấn cho vòng sơ tuyển. Các câu truy vấn ở vòng sơ tuyển theo dạng Textual KIS và chỉ tập trung trong khoảng 200 giờ video mà Ban tổ chức cung cấp cho các đội dự thi.

Mỗi đội sẽ xử lý câu truy vấn bằng chính công cụ mà mình phát triển và nộp kết quả cho Ban tổ chức. Bên cạnh kết quả nộp cho mỗi câu truy vấn, mỗi đội cần có phần mô tả tóm tắt cách mà đội mình đã làm để tìm ra được kết quả cho mỗi câu truy vấn để Ban giám khảo có thể tham khảo, đánh giá khi cần thiết.

Số lượng câu truy vấn, quy định về độ đo đánh giá, định dạng kết quả khi nộp trực tuyến, số lần nộp tối đa, hình thức nộp trực tuyến và số lượng đội tối đa (dự kiến) được chọn vào vòng chung kết sẽ được thông báo sau.

Ban giám khảo có thể yêu cầu thẩm định giải pháp của một số đội để đảm bảo kết quả của đội đã nộp được thực hiện từ chính công cụ của đội.

Các đội có kết quả tốt trong vòng sơ tuyển sẽ được chọn vào vòng chung kết.

6.3 Vòng chung kết

Các đội vào vòng chung kết sẽ sử dụng công cụ phần mềm của nhóm mình để tìm kiếm các video phù hợp cho các câu truy vấn trong vòng chung kết.

Các đội sẽ dự thi trong một hội trường rộng. Trong một số trường hợp đặc biệt, Ban tổ chức có thể cho phép một số đội từ nước ngoài tham gia dự thi trực tuyến. Ban giám khảo sẽ lần lượt đưa ra mô tả/yêu cầu cho mỗi câu truy vấn. Các câu truy vấn ở vòng chung kết gồm cả dạng Textual KIS và Video KIS, khai thác nội dung của toàn bộ dữ liệu (khoảng 300 giờ video và metadata) đã cung cấp cho các đội vào vòng chung kết.

Với mỗi câu truy vấn, mỗi đội sẽ sử dụng công cụ phần mềm mà mình đã phát triển để tìm ra chính xác đoạn video theo yêu cầu mà Ban giám khảo đưa ra trong thời gian cho phép và nộp kết quả trực tuyến lên server của Ban giám khảo. Nếu kết quả là chính xác, đội sẽ được ghi nhận điểm cho câu truy vấn này. Nếu kết quả chưa chính xác, đội sẽ được phép nộp kết quả mới cho đến khi có được kết quả đúng hoặc hết thời gian dành cho câu truy vấn này.

Số điểm tối đa dành cho mỗi câu truy vấn là bằng nhau (ví dụ như 100 điểm/câu truy vấn). Điểm của đội cho mỗi câu truy vấn tùy thuộc vào thời gian mà đội cần dùng để tìm ra đoạn video chính xác, cũng như số lần nộp kết quả của đội. Nói cách khác, nếu đội nộp kết quả càng trễ hay nộp kết quả sai nhiều lần thì điểm số mà đội nhận được sẽ bị giảm.

Quy tắc tính điểm cũng như cách giao tiếp với hệ thống chấm thi trực tuyến sẽ được thông báo riêng đến các đội vào vòng chung kết.

7. Các mốc thời gian quan trọng

- Thời gian đăng ký tham gia: từ ngày phát động đến ngày 15/10/2022.
- Thời gian tổ chức tập huấn cho thí sinh tham gia hội thi: tháng 9, 10/2022.
- Thời gian công bố chủ đề, yêu cầu của cuộc thi: ngày 01/10/2022.

- Thời gian công bố dữ liệu (khoảng 100 giờ video và metadata) để các đội làm quen cuộc thi: ngày 01/10/2022.
- Thời gian công bố website thử nghiệm để các đội làm quen cuộc thi: ngày 05/10/2022.
- Thời gian diễn ra vòng sơ tuyển: từ ngày 10/10/2022 đến ngày 25/10/2022.
- Thời gian công bố dữ liệu bổ sung cho vòng sơ tuyển (thêm khoảng 100 giờ video và metadata) cùng quy định chi tiết của vòng sơ tuyển: ngày 10/10/2022.
- Thời gian công bố kết quả vòng sơ tuyển: ngày 31/10/2022.
- Thời gian diễn ra vòng chung kết: từ 01/11/2022 đến 25/11/2022

Khung thời gian này có thể có thay đổi tùy theo tình hình thực tế tổ chức cuộc thi, Ban tổ chức sẽ có thông báo cho đội thi nếu có sự thay đổi về thời gian.