


MẪU BÁO CÁO CỦA MỖI HV

Họ và tên (IN HOA)	PHẠM NGỌC BẢO CƯỜNG CH2001024
Ảnh	
Số buổi vắng	1
Bonus	10
Tên đề tài (VN)	HỆ THỐNG TẠO CÂU HỎI TỪ ĐOẠN VĂN TIẾNG VIỆT
Tên đề tài (EN)	
Giới thiệu	<ul style="list-style-type: none"><i>Bài toán/vấn đề mà đề tài muốn giải quyết</i> <p><i>Việc tạo ra các câu hỏi hợp lệ về mặt cú pháp và ngữ nghĩa và có liên quan từ các đoạn văn rất hữu ích với nhiều ứng dụng. Hiện nay có đã có nhiều công trình nghiên cứu về việc sinh tự động câu hỏi (QG). Tuy nhiên, phần lớn những công trình nghiên cứu này chỉ sinh câu hỏi từ câu không xét trong ngữ cảnh đoạn văn bản và hầu như chưa có hệ thống nào phổ biến rộng rãi.</i></p>

Trong bài nghiên cứu này, tôi sẽ đề xuất giải pháp và triển khai hệ thống sinh câu hỏi từ câu trả lời trong văn bản đó, đồng thời cũng sẽ phổ biến rộng rãi hệ thống dạng dịch vụ web cho việc chạy thử nghiệm và sử dụng.

Hệ thống sẽ cung cấp các giao diện chức năng thân thiện để người dùng dễ dàng chọn lựa câu trả lời phù hợp trong quá trình tạo câu hỏi.

- *Lý do chọn đề tài*

Đặt những câu hỏi liên quan và thông minh luôn là một phần không thể thiếu trong quá trình học tập của con người, vì nó có thể giúp đánh giá mức độ hiểu biết của người học cho một phần văn bản nào đó (một phần nội dung giảng dạy, một bài báo, v.v.). Tuy nhiên, việc hình thành câu hỏi theo cách thủ công là một nhiệm vụ vất vả. Hệ thống tạo câu hỏi tự động (QG) có thể giúp giảm bớt vấn đề này bằng cách học cách tạo câu hỏi trên quy mô lớn một cách hiệu quả. Hệ thống QG có nhiều ứng dụng trong nhiều lĩnh vực như:

- *FAQs (những câu hỏi thường gặp)*
- *Hệ thống hỗ trợ học tập thông minh*
- *Tự động hóa khả năng đọc hiểu và trợ lý ảo / chatbot....*

Đối với hệ thống QG, nhiệm vụ là tạo ra các câu hỏi mạch lạc về mặt cú pháp, đúng ngữ nghĩa và một cách tự nhiên. Ngoài ra, câu hỏi được mong đợi phải liên quan đến văn bản đầu vào và câu hỏi phải xoay quanh các câu trả lời có trong văn bản.

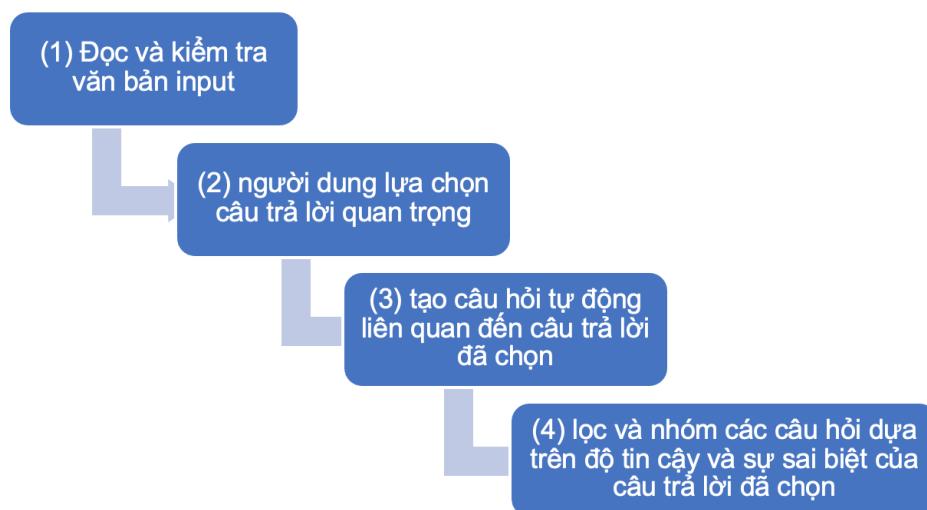
- *Mô tả input và output, nên có hình minh họa*

Ví dụ có đoạn văn bản input như sau:

	<p>Đoạn văn bản:</p> <p>OOP (viết tắt của Object Oriented Programming) – lập trình hướng đối tượng là một phương pháp lập trình dựa trên khái niệm về lớp và đối tượng.</p> <p>Mục tiêu của OOP là tối ưu việc quản lý source code, giúp tăng khả năng tái sử dụng và quan trọng hơn hết là giúp tóm gọn các thủ tục đã biết trước tính chất thông qua việc sử dụng các đối tượng.</p> <p>Câu hỏi:</p> <ul style="list-style-type: none">- Lập trình hướng đối tượng là gì? một phương pháp lập trình dựa trên khái niệm về lớp và đối tượng- Mục tiêu của lập trình hướng đối tượng là gì? tối ưu việc quản lý source code, giúp tăng khả năng tái sử dụng và quan trọng hơn hết là giúp tóm gọn các thủ tục đã biết trước tính chất thông qua việc sử dụng các đối tượng <p>Câu hỏi được sinh ra</p> <p>Dựa vào câu trả lời được người dùng sát định từ đầu</p> <p>Trong ví dụ trên, phần câu hỏi sẽ được hệ thống tự động sinh ra dựa vào đoạn văn bản input đầu vào và câu trả lời được chọn trước.</p>
Mục tiêu	<ul style="list-style-type: none">• Hoàn thành được giải pháp kỹ thuật giúp cho việc chọn lựa chọn các câu hỏi phù hợp với câu trả lời. Việc một câu trả lời có thể có nhiều câu hỏi khác nhau, trùng lặp ở những ngữ cảnh khác nhau, vì vậy hệ thống web cũng hỗ trợ cơ chế cho phép người dùng chọn lựa câu hỏi phù hợp.• Thiết kế, cài đặt và đưa vào chạy thử nghiệm được ứng dụng web, có các chức năng cho phép người dùng nhập đoạn văn bản và các chức năng cho phép chọn lựa các câu trả lời từ người dùng một cách trực quan dễ sử dụng.• Áp dụng BERT (Devlin et al., 2018) để lọc các câu hỏi không trả lời được ra khỏi văn bản.

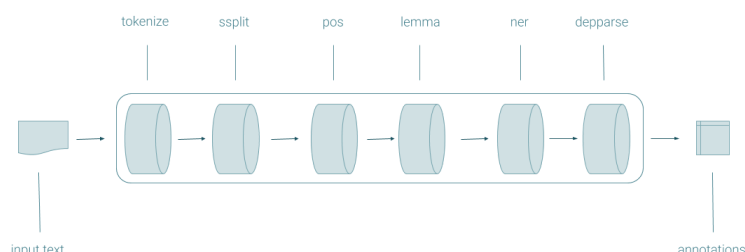
**Nội dung
và
phương
pháp thực
hiện**

- Trong bài nghiên cứu này, thống sinh câu hỏi sẽ được thực hiện qua các 4 công đoạn:



Bước (1): cho phép người dùng nhập và hệ thống tiền xử lý dữ liệu để đảm bảo dữ liệu sẵn sàng. Các ký hoặc câu lỗi sẽ được lọc bỏ, các lỗi unicode được thay thế. Bước này được thực hiện bởi người dùng, hệ thống hỗ trợ công cụ tiện ích

Bước (2): hệ thống hỗ trợ công cụ cho người dùng lựa chọn câu trả lời phù hợp. Trong giai đoạn này Stanford CoreNLP được sử dụng để xử lý cho việc trích xuất các thực thể và label.



	<p><i>Bước (3): Từ những câu trả lời được chọn ở bước 2 áp dụng hướng tiếp cận seq2seq [2] để tạo ra các câu hỏi. Module tạo câu hỏi này sẽ tạo ra các câu hỏi phù hợp nhất có thể về mặt cú pháp cũng như về mặt ngữ nghĩa.</i></p> <p><i>Việc tạo câu hỏi là việc làm rất phức tạp và khó khăn, vì vậy sau quá trình sinh ra câu hỏi (QG), các câu hỏi phải được đánh giá tính phù hợp về mặt ngữ pháp cũng như ngữ nghĩa. Vì vậy các phương pháp BLEU, GLEU và ROUGE-L được áp dụng cho việc đánh giá này.</i></p> <p><i>Bước (4): Ở bước cuối cùng này, câu hỏi không trả lời được sẽ loại bỏ bằng hướng tiếp cận BERT [3]. Kết quả cuối cùng là các câu hỏi và câu trả lời liên quan được nhóm lại với nhau.</i></p>
Kết quả dự kiến	<ul style="list-style-type: none"> • <i>Ứng dụng web giúp cho việc dễ dàng sinh câu hỏi từ văn bản đầu vào.</i> • <i>Đánh giá kết quả đạt được và so sánh giữa các phương pháp tiếp cận hiện có.</i> • <i>Đề xuất hướng cải tiến mới.</i>
Tài liệu tham khảo	<p>[1] Kiet Van Nguyen¹, Duc-Vu Nguyen, Anh Gia-Tuan Nguyen and Ngan Luu-Thuy Nguyen. 2020. A Vietnamese Dataset for Evaluating Machine Reading Comprehension</p> <p>[2] Thi-Trang Nguyen, Huu-Hoang Nguyen, Kiem-Hieu Nguyen. 2020. A Study on Seq2seq for Sentence Compression in Vietnamese</p> <p>[3] Thi-Thanh Ha, ThaiNguyen, Van-Nha Nguyen, Kiem-Hieu Nguyen, Kim-Anh Nguyen, Tien-Thanh Nguyen. 2020. Utilizing BERT for Question Retrieval in Vietnamese E-commerce Sites</p> <p>[4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.</p>

- | | |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | <p>[5] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.</p> <p>[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.</p> <p>[7] Xinya Du, Junru Shao, and Claire Cardie. 2017. Learning to ask: Neural question generation for reading comprehension. In Proceedings of the 55th ACL, pages 1342–1352. ACL.</p> <p>[8] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor OK Li. 2016. Incorporating copying mechanism in sequence-to-sequence learning. In Proceedings of the 54th ACL (Volume 1: Long Papers), volume 1, pages 1631–1640.</p> <p>[9] Lynette Hirschman, Marc Light, Eric Breck, and John D Burger. 1999. Deep read: A reading comprehension system. In ACL, pages 325–332. ACL.</p> <p>[10] Vishwajeet Kumar, Yuncheng Hua, Ganesh Ramakrishnan, Guilin Qi, Lianli Gao, and Yuan-Fang Li. 2019a. Difficulty-controllable multi-hop question generation from knowledge graphs. In ISWC.</p> |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|