|  |  |
| --- | --- |
|  |  |
| **TRƯỜNG ĐẠI HỌC XÂY DỰNG HÀ NỘI**  **---------------------------------------** | |
| Đại học Xây dựng Hà Nội | |
|  | |
| **NGUYỄN VĂN A** | |
|  | |
|  | |
|  | |
| **ĐỀ CƯƠNG**  **ĐỒ ÁN TỐT NGHIỆP KỸ SƯ** | |
|  | |
|  | |
| HÀ NỘI - 2020 | |
|  |  |
| **TRƯỜNG ĐẠI HỌC XÂY DỰNG HÀ NỘI**  **---------------------------------------** | |
| Đại học Xây dựng Hà Nội | |
|  | |
| **NGUYỄN VĂN A** | |
|  | |
|  | |
|  | |
| **XÂY DỰNG ABC XYZ** | |
| CHUYÊN NGÀNH : KHOA HỌC MÁY TÍNH | |
| MÃ SỐ:  **ĐỀ CƯƠNG ĐỒ ÁN TỐT NGHIỆP KỸ SƯ**  **NGƯỜI HƯỚNG DẪN KHOA HỌC**  **Ths. Nguyễn Văn B**  HÀ NỘI – 2020 | |
|  | |

1. **MỞ ĐẦU**
2. **Lý do chọn đề tài**

Trình bày chi tiết lý do chọn đề tài.

1. **Mục đích đề tài**

Sinh viên trình bày mục đích thực hiện đề tài.

1. **Đối tượng và phạm vi nghiên cứu/đề tài**

* Đối tượng:
* Phạm vi nghiên cứu:

1. **Phương pháp nghiên cứu/thực hiện**

* Trình bày phương pháp mà sinh viên thực hiện đề tài.

1. **NỘI DUNG**

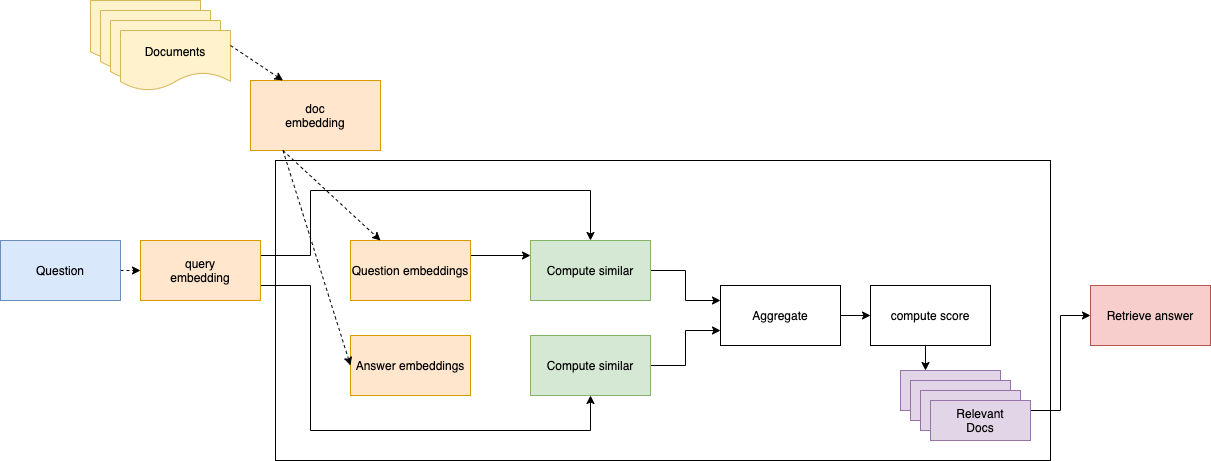
**Chương 1: TỔNG QUAN**

Giới thiệu chương: Chương này tập trung vào trình bày khái quát

Nội dung chương 1 sẽ bố cục theo các mục sau:

* 1. .
  2. .
  3. .

Ví dụ: Mục này sẽ đưa ra kiến trúc sẽ áp dụng trong việc xây dựng hệ thống hỏi đáp, kiến trúc chung được mô tả bằng hình vẽ sau:



Trong kiến trúc này câu hỏi của người dùng hay các tài liệu liên quan sẽ được encode thành các embedding vectors sau đó tôi tiến hành tính khoảng cách (độ giống nhau) từ câu hỏi của người dùng đến câu hỏi và đến câu trả lời trong tập dữ liệu huấn luyện và cuối cùng là kết hợp lại để tính ra điểm để đánh giá mức độ phù hợp giữa câu hỏi và câu trả lời nhận được.

* 1. .
  2. .
     1. .
        1. .
        2. .
  3. Kết luận chương

**Chương 2: PHƯƠNG PHÁP THỰC HIỆN**

Giới thiệu chương: Trình bày ngắn gọn về CÁCH THỨC THỰC HIỆN ĐỀ TÀI.

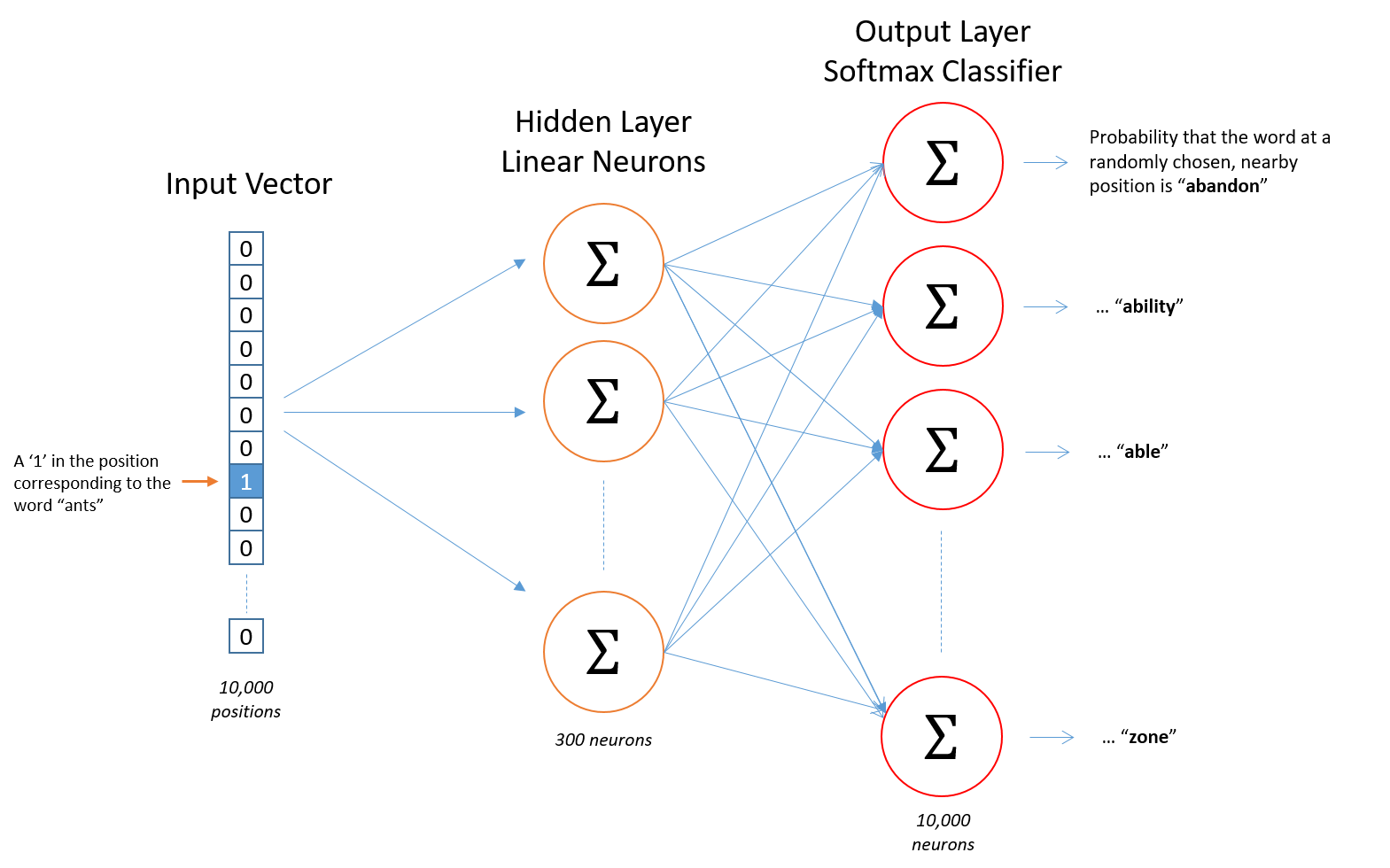
Nội dung chương 2 sẽ bố cục theo các mục sau:

* 1. Công nghệ/thuật toán sử dụng.
     1. Ngôn ngữ lập trình C#
     2. Framework Unity..

Trong mục này, sinh viên trình bày về các công nghệ/thuật toán trong đề tài của mình

* + 1. Mô hình xxxx.

Ví dụ: Để biểu diễn từ tôi sử dụng bộ ngữ liệu gồm tất cả các bài viết tiếng Việt trên wikipedia, sau đó tiến hành tách từ, tách câu để huấn luyện cho mô hình skip-gram dùng để biểu diễn từ dưới dạng vector.



* + 1. Xây dựng mô hình truy xuất thông tin và trích xuất câu trả lời.

Sau khi đã có biểu diễn vector các từ của câu hỏi của người dùng, câu hỏi và câu trả lời trong bộ dữ liệu huấn luyện, tôi sử dụng phương pháp sau để xếp hạng các câu trả lời phù hợp:

Với hai từ wt và wu, độ giống nhau giữa 2 từ được định nghĩa là khoảng các cosine giữa 2 vector biểu diễn 2 từ đó:

Biểu thức này cho phép tính độ giống giữa 2 từ, vậy để tính toán độ giống nhau giữa 2 túi từ (bags-of-words – coi mỗi câu hỏi, câu trả lời là các túi từ). Cần đưa ra phương pháp tính độ giống giữa 1 từ với 1 túi từ. Cụ thể, để tính độ giống giữa từ w và túi từ T, ta lấy độ giống lớn nhất giữa từng cặp từ w với một từ bất kỳ w’ thuộc vào T:

Sau khi có hàm xếp hạng f ta có thể xếp hạng được các câu hỏi trong tập huấn luyện tương tự với câu hỏi mà người dùng nhập vào, sau đó lấy câu trả lời để đưa ra làm câu trả lời cho hệ thống. Tuy nhiên ta cần xác định được tham số xác định ngưỡng (threshold) để biết được câu hỏi đó có đúng và phù hợp với câu hỏi trong có sở dữ liệu hay không. Ngưỡng này sẽ được chọn dựa trên đánh giá độ chính xác của mô hình.

* 1. Cài đặt hệ thống.
  2. Kết luận chương

**Chương 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ**

Giới thiệu chương: Chương này tập trung vào các phương pháp thử nghiệm và đánh giá.

Nội dụng chương 3 sẽ bố cục theo các mục sau (ví dụ):

* 1. Tập dữ liệu thử nghiệm.
  2. Các phép đánh giá.
  3. So sánh các phương pháp đánh giá.
  4. Độ phức tạp và thời gian thực hiện phương pháp.
  5. Triển khai thử nghiệm và kết quả thử nghiệm
  6. Kết luận chương

1. **KẾT LUẬN**

Áp dụng mô hình hỏi đáp vào việc hỗ trợ sinh viên trường Đại học Xây dựng mang lại hiệu quả cao, giúp sinh viên dễ dàng tiếp cận thông tin từ phía nhà trường đồng thời làm giảm khối lượng công việc tiếp nhận và giải quyết thắc mắc, nhu cầu thông tin từ phía sinh viên cho các phòng ban trong trường. Tuy nhiên mô hình đưa ra vẫn còn nhiều hạn chế cần cải thiện trong tương lai và đề xuất các hướng đi tiếp theo nhằm nâng cao hơn nữa hiệu quả của việc hỏi đáp.

1. **DANH MỤC CÁC TÀI LIỆU THAM KHẢO**
2. Ai, Q., L. Yang, J. Guo, and W. B. Croft. 2016a. “Analysis of the paragraph vector model for information retrieval”. In: Proc. ICTIR. ACM. 133–142..
3. Arora, S., Y. Li, Y. Liang, T. Ma, and A. Risteski. 2015. “Rand-walk: A latent variable model approach to word embeddings”. arXiv preprint

arXiv:1502.03520..

1. Bailey, P., N. Craswell, and D. Hawking. 2003. “Engineering a multipurpose test collection for web retrieval experiments”. Information Processing & Management. 39(6): 853–871.
2. J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: experiments on recommending content from information streams. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1185-1194. ACM, 2010.
3. Bullinaria, J. A. and J. P. Levy. 2012. “Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and SVD”. Behavior research methods..
4. Ganguly, D., D. Roy, M. Mitra, and G. J. Jones. 2015. “Word Embedding based Generalized Language Model for Information Retrieval”. In: Proc. SIGIR. ACM..
5. 2016. From word embeddings to document similarities for improved information retrieval in software engineering. ICSE '16: Proceedings of the 38th International Conference on Software Engineering May 2016 404–415
6. **DỰ KIẾN KẾ HOẠCH THỰC HIỆN**

|  |  |  |
| --- | --- | --- |
| **TT** | **Nội dung** | **Thời gian** |
| 1 |  | 01/11/2020 – 30/11/2020 |
| 2 |  | 01/12/2020 – 15/12/2020 |
| 3 |  | 16/12/2020 – 30/02/2021 |
| 4 |  | 01/03/2021 – 30/03/2021 |
| 5 |  | 01/04/2021 – 30/04/2021 |

|  |  |
| --- | --- |
| **Ý KIẾN CỦA GIÁO VIÊN HƯỚNG DẪN** | **NGƯỜI LẬP ĐỀ CƯƠNG** |