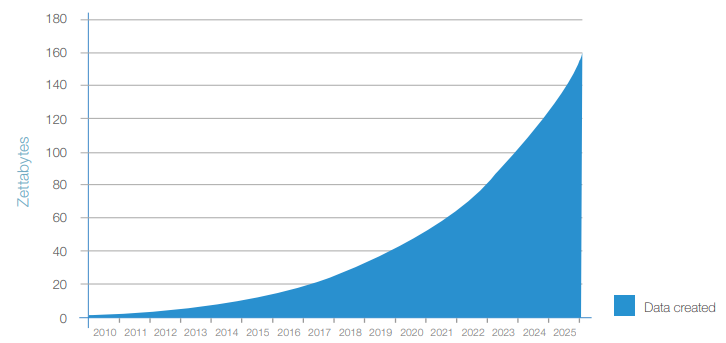
**CHƯƠNG 1: ĐẶC TẢ YÊU CẦU**

1. **Giới thiệu**
   1. **Dữ liệu**

Mỗi ngày trôi qua sẽ có một lượng thông tin khổng lồ được tạo ra khiến cho con người chúng ta không thể ghi nhớ hết tất cả. Vì thế, chúng ta cần phải lưu trữ lại thành những dữ liệu để có thể trích xuất khi cần thiết giúp cho việc xử lý lượng thông tin lớn một cách dễ dàng. Dữ liệu bao gồm hình ảnh, bản ghi âm, video, bảng thống kê,… có nội dung có thể chuyển sang dạng văn bản. Những dữ liệu này có là dạng phi ngôn ngữ thường được lưu dưới dạng một cấu trúc cố định như tệp cở sở dữ liệu quan hệ hoặc bảng tính.

Hình . Thống kê dữ liệu được tạo ra từng năm

* 1. **Chuyển dữ liệu thành văn bản**

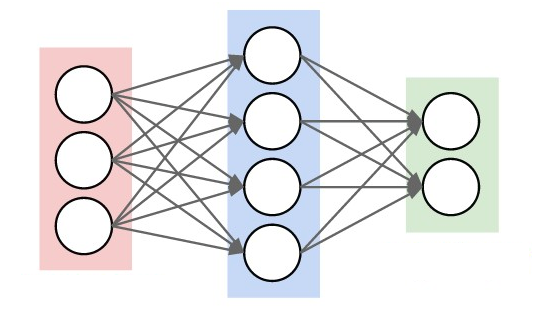
Trong sự phát triển ngày càng lớn về công nghệ trí tuệ nhân tạo dẫn đến nhu cầu tạo dữ liệu thành văn bản với tập dữ liệu quy mô lớn và mô hình mạng nơ-ron được đào tạo từ đầu đến cuối ngày càng cao. Do đó, đã có nhiều nhà phát triển đã triển khai nhiều công trình nghiên cứu vấn đề data-to-text nhầm đáp ứng nhu cầu trên. Nhưng những nghiên cứu đó lại không có bất kỳ lựa chọn hoặc lập kế hoạch nội dung nào hoặc lựa chọn nội dung và lập kế hoạch. Vì thế mục đích của bài báo là trình bày kiến trúc mạng nơ-ron kết hợp lựa chọn nội dung có kế hoạch.

Để có thế thực thi được phương pháp kết hợp mạng nơ-ron và việc chọn nội dung dùng cho việc lập kế hoạch, nhóm tác giả đã chia thành hai nhiệm vụ chính: Lập kế hoạch với nội dung quan trọng được chọn ra từ dữ liệu đồng thời sắp xếp thứ tự dựa vào nội dung kế hoạch; và tạo tài liệu văn bản dựa trên kế hoạch ban đầu. Từ đó, đó tạo ra những lợi thế cho việc tạo ra tài liệu nhiều câu:

* Đại diện cho một nhóm dữ liệu cấp cao của cấu trúc tài liệu cho phép bộ giải mã tập trung vào các nhiệm vụ. Từ đó làm cho việc lập kế hoạch và hiện thực bề mặt hóa dễ dàng hơn.
* Làm cho quá trình tạo dữ liệu thành tài liệu văn bản dễ hiểu hơn bằng cách tạo ra một đại diện trung gian.
* Giảm dư thừa trong đầu ra vì có ít khả năng nội dung kế hoạch chứa cùng một thông tin dữ liệu ở nhiều nơi.

1. **Mô hình sử dụng**
   1. **Mạng nơ-ron**

Mạng nơ-ron nhân tạo (Artificial Neural Networks - ANN) được xây dựng, mô phỏng và mô hình hóa dựa trên sự hoạt động mạng lưới thần kinh não người bằng  một chuỗi các [thuật toán](https://vietnambiz.vn/giao-dich-thuat-toan-algorithmic-trading-la-gi-nhung-dac-diem-can-luu-y-20191205025413985.htm)được đưa ra để nỗ lực tìm kiếm các [mối quan hệ](https://vietnambiz.vn/tuong-quan-dong-bien-positive-correlation-la-gi-nhung-dac-diem-can-luu-y-20191227113029107.htm) cơ bản trong một tập hợp dữ liệu. Một mạng nơ-ron nhân tạo là tập hợp của nhiều nơ-ron nhân tạo được nối kết với nhau, và xử lý thông tin bằng cách truyền theo các kết nối và tính giá trị mới tại các nút.



Hình .Mô hình mạng nơ-ron nhân tạo (ANN)

Mạng nơ-ron là sự kết hợp của của các tầng perceptron hay còn được gọi là perceptron đa tầng (multilayer perceptron) là mạng truyền thẳng, tất cả các nơron chỉ có thể được kết nối với nhau theo một hướng tới một hay nhiều các nơron khác trong lớp kế tiếp (trừ các nơ-ron ở output layer). Multilayer perceptron gồm có 3 tầng: tầng vào (input layer), tầng ẩn (hidden layer), tầng ra (output layer). Mỗi nút trong mạng là một sigmoid nơ-ron nhưng hàm kích hoạt của chúng có thể khác nhau thường cùng dạng với nhau.

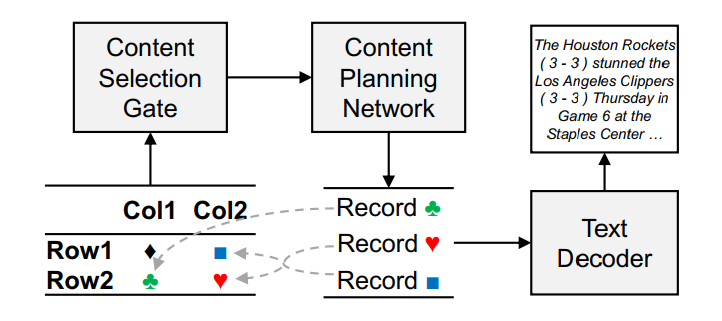
Hình. Mô hình multilayer perceptron

Mạng lưới thần kinh nhân tạo đang mở đường cho các ứng dụng thay đổi cuộc sống được phát triển để sử dụng trong tất cả các lĩnh vực của nền kinh tế. Các nền tảng trí tuệ nhân tạo được xây dựng trên mạng ANNs đang phá vỡ những cách làm truyền thống. Từ việc dịch các trang web sang các ngôn ngữ khác đến việc có một cửa hàng tạp hóa trợ lý ảo trực tuyến để trò chuyện với chatbot để giải quyết vấn đề, các nền tảng AI đang đơn giản hóa các giao dịch và làm cho các dịch vụ có thể truy cập được với chi phí không đáng kể.

* 1. **Chọn lọc nội dung và tạo kế hoạch (Content Selection and Planning)**

Kiến trúc tổng thể của mô hình lập lựa chọn và lập kế hoạch nội dung được chia thành hai giai đoạn chính:

* Giai đoạn 1: Lựa chọn nội dung và lập kế hoạch hoạt động dựa trên các bản ghi đầu vào của cơ sở dữ liệu, từ đó tạo ra một kế hoạch nội dung xác định những bản ghi nào sẽ được diễn đạt bằng lời trong tài liệu và theo thứ tự nào.
* Giai đoạn 2: Tạo văn bản tạo ra văn bản đầu ra với kế hoạch nội dung làm đầu vào; ở mỗi bước giải mã, mỗi thành phần trong mô hình chú ý đến các vector được biểu diễn ở từng bản ghi trong kế hoạch nội dung.



Hình. Mô hình hoạt động của Content Selection and Planning

Content Selection and Planning là một hệ thống xử lý dữ liệu vô cùng phức tạp, dữ liệu được xử lý trải qua nhiều quá trình tính toán khác nhau bao gồm:

* Bộ mã hóa bản ghi: tại đây dữ liệu sẽ được chuyển hóa thành các dạng véc-tơ, sau đó sẽ được các multilayer perceptron trong mô hình sử dụng.
* Cổng lựa chọn nội dung: Từ các véc-tơ thu được ở bộ mã hóa bản ghi, các nhà nghiên cứu thực hiện các phép tính để thu được những thông tin tối ưu, tránh gây trùng lập.
* Lập nội dung kế hoạch: Lập bản đồ văn bản trong phần tóm tắt dựa vào các thực thể ở đầu vào (mối quan hệ giữa các thực thể). Mỗi kế hoạch là một chuỗi các con trỏ với mỗi mục nhập trỏ đến các véc-tơ và sử dụng bộ giải mã LSTM để tạo mã thông báo.
* Tạo văn bản: Mã hóa kế hoạch nội dung thành các đơn vị LSTM, dựa vào đó sử dụng mạng nơ-ron tuần hoàn sử dụng các đơn vị LSTM và LSTM sao chép để tăng hiệu năng xử lý.
* Đào tạo và suy luận: Được đào tạo để tối đa hóa khả năng ghi nhật kí của mô hình. Đối với mỗi giai đoạn, họ sử dụng tìm kiếm chùm tia để đạt được kết quả tốt nhất.