(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ

(11) **1-0053719**

(51)[2022.01] **G06V 10/10**

(13) **B**

(21) 1-2023-09261

(22) 26/12/2023

(45) 25/11/2025 452

(43) 25/03/2024 432

(73) TẬP ĐOÀN CÔNG NGHIỆP - VIỄN THÔNG QUÂN ĐỘI (VN)
Lô D26 Khu đô thị mới Cầu Giấy, phường Yên Hoà, quận Cầu Giấy, thành phố Hà Nội.

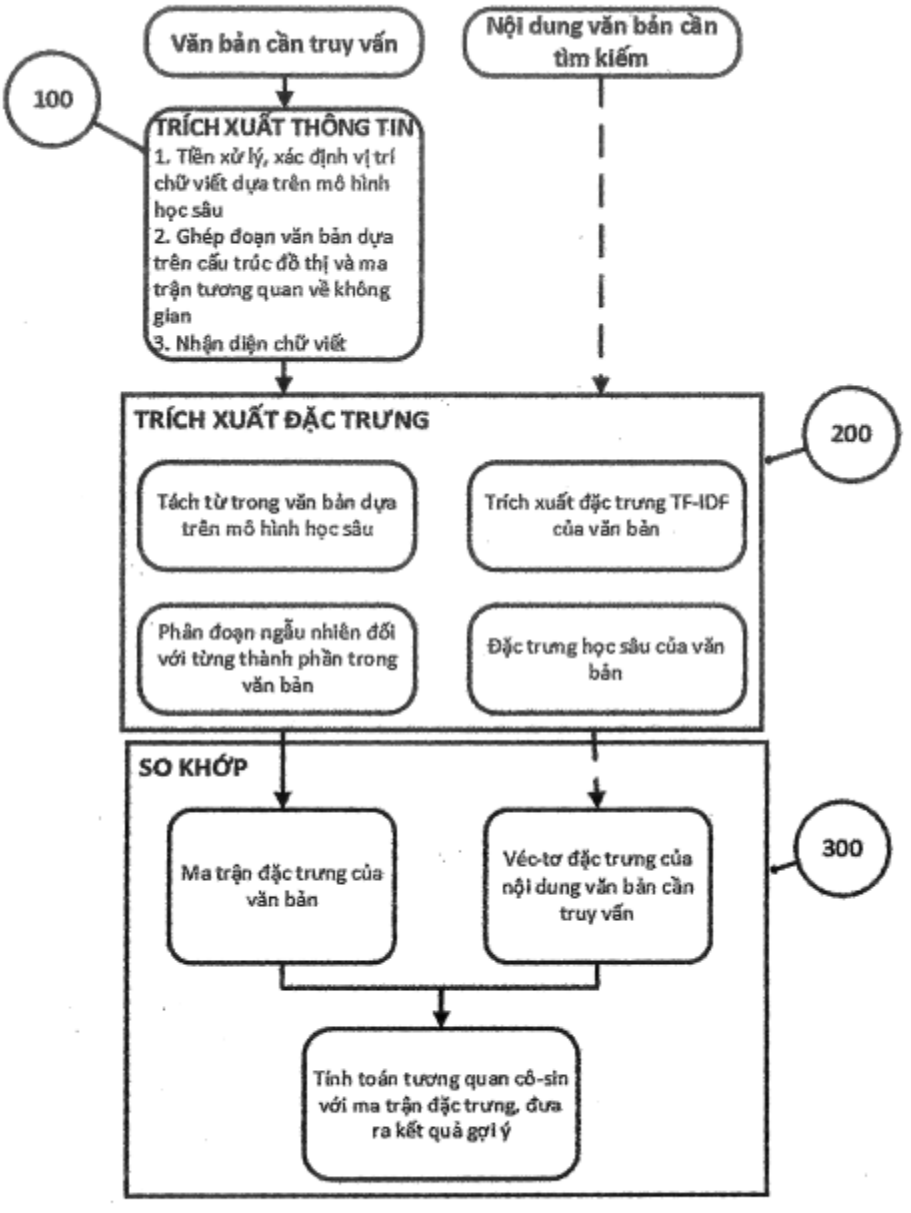
(72) Lê Văn Bằng (VN); Nguyễn Thị Hòa (VN); Đào Mai Ngọc (VN); Vũ Đức Chính (VN); Nguyễn Thị Trang (VN).

(74) Công ty TNHH NACILAW (NACILAW)

(54) PHƯƠNG PHÁP TÌM KIẾM NỘI DUNG DỰA TRÊN TRUY VẤN KHỐI VÀ ĐẶC TRƯNG VĂN BẢN

(21) 1-2023-09261

(57) Phương pháp tìm kiếm nội dung dựa trên truy vấn khối và đặc trưng văn bản bao gồm quá trình quá trình định vị, ghép đoạn cho văn bản và nhận diện chữ viết, quá trình trích xuất đặc trưng, bổ sung thông tin cho khối đặc trưng của văn bản và quá trình truy vấn nội dung văn bản. Phương pháp đề xuất đã thể hiện được khả năng tính toán chính xác cũng như hiệu năng rất cao, hiệu quả trong việc số hóa và truy vấn các văn bản tiếng việt, đặc biệt trong các trường hợp xử lý dữ liệu quy mô lớn.



Hình 1

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp tìm kiếm nội dung dựa trên truy vấn khối và đặc trưng văn bản. Cụ thể, phương pháp tìm kiếm nội dung dựa trên truy vấn khối và đặc trưng văn bản được ứng dụng trong lĩnh vực trí tuệ nhân tạo và thị giác máy tính, giúp xử lý và tìm kiếm các nội dung liên quan đến yêu cầu của người dùng.

Tình trạng kỹ thuật của sáng chế

Hiện nay, số hóa văn bản đang là một nhu cầu lớn, với các bài toán được đặt ra như lưu trữ phiên bản số các văn bản mà trước vẫn sử dụng bản giấy để lưu trữ, tổng hợp, gom nhóm quản trị văn bản, tóm tắt nội dung, tìm kiếm văn bản, cũng như nội dung trong văn bản. Với một lượng rất lớn văn bản số được sinh ra trong một ngày, để có thể phân tích và truy vấn nội dung trong đó, đặc biệt là các văn bản được quét dưới dạng ký tự quang học (OCR) phải đáp ứng được việc cấu trúc hóa và truy vấn nội dung ở tốc độ cao.

Với những công nghệ đọc hiểu và phân tích OCR, công nghệ học sâu được sử dụng rộng rãi để xác định vị trí của từng dòng văn bản, cũng như nhận diện các dòng văn bản đó. Tuy nhiên, đối với những văn bản có cấu trúc khác nhau, về sự liên quan giữa các thành phần, các đoạn văn cũng như vị trí, độ dài của các thành phần (tiêu đề, đoạn văn…) thì thông thường cần sử dụng dữ liệu mẫu để huấn luyện cho các mô hình này, quá trình xây dựng cũng như gán nhãn dữ liệu tiêu tốn rất nhiều nguồn lực và thời gian. Ngoài ra, khả năng tự tương thích với những biến đổi nhỏ trên văn bản, hoặc đối với những loại văn bản khác khá hạn chế.

Còn đối với những công nghệ truy vấn nội dung, có nhiều phương án để so khớp nội dung văn bản với đoạn văn bản cần tìm, ví dụ như đối chiếu theo mẫu, đối chiếu theo từ khóa, hoặc tìm kiếm theo đặc trưng trừu tượng được trích xuất từ các mô hình học sâu của văn bản. Đối với tìm kiếm theo từ khóa, phù hợp cho tìm kiếm những đoạn văn ngắn, hoặc là các tiêu đề không dài, còn đối với tìm kiếm theo đặc trưng học sâu thì chỉ hiệu quả khi tìm kiếm một đoạn dài, có ngữ nghĩa, ngữ cảnh. Những phương án này thường có hạn chế về việc linh hoạt tìm kiếm theo đặc trưng hoặc tìm kiếm theo từ khóa, dẫn đến hiệu quả tìm kiếm không được cao.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là khắc phục những vấn đề kỹ thuật còn tồn tại, nâng cao hiệu quả trong quá trình so khớp tìm kiếm và linh hoạt đối với nhiều loại văn bản có cấu trúc khác nhau, nội dung, chủng loại khác nhau, bằng việc đề xuất phương pháp tìm kiếm nội dung dựa trên truy vấn khối và đặc trưng văn bản. Phương pháp trong sáng chế cũng giúp cho việc xử lý linh hoạt trên các thiết bị có cấu hình khác nhau.

Trong sáng chế này, phương pháp tìm kiếm nội dung dựa trên truy vấn khối và đặc trưng văn bản được thực hiện dựa trên các bước như sau:

Bước 1: quá trình định vị, ghép đoạn cho văn bản và nhận diện chữ viết: sử dụng để xác định vị trí của từng câu văn bản, dựa trên tiền đề mảng văn bản đã được tiền xử lý, hiệu chỉnh hướng. Đồng thời nhận diện văn bản trên từng vị trí văn bản đã định vị được bởi mô hình học sâu. Dựa trên thông tin vị trí câu văn bản, những câu văn bản cùng tính chất và gần nhau được ghép lại thành các đoạn văn khác nhau.

Bước 2: quá trình trích xuất đặc trưng, bổ sung thông tin cho khối đặc trưng của văn bản: trích xuất đặc trưng của văn bản, mỗi đặc trưng này là một véc-tơ có cùng chiều dài, và được cập nhật đồng bộ với vị trí của các đoạn văn, câu văn.

Bước 3: quá trình truy vấn nội dung văn bản: tự động phân tích kết cấu của nội dung văn bản đưa vào truy vấn, sử dụng đồng thời phương án tìm kiếm theo từ khóa và tìm kiếm theo đặc trưng bằng cách tính toán tương quan cô-sin giữa đặc trưng văn bản cần so khớp và ma trận đặc trưng của văn bản đã được tính toán từ trước.

Mô tả vắn tắt các hình vẽ

Hình 1 là hình vẽ mô tả tổng thể quá trình xử lý trong phương pháp tìm kiếm nội dung dựa trên truy vấn khối và đặc trưng văn bản;

Hình 2 là hình vẽ mô tả quá trình định vị, ghép đoạn cho văn bản và nhận diện chữ viết;

Hình 3 là hình vẽ mô tả quá trình trích xuất đặc trưng, bổ sung thông tin cho khối đặc trưng của văn bản;

Hình 4 là hình vẽ mô tả quá trình truy vấn nội dung văn bản;

Hình 5 là hình vẽ mô tả ví dụ về hiệu quả của tăng cường độ tương phản của ảnh văn bản;

Hình 6 là hình vẽ mô tả cách thức hoạt động của giải thuật xác định hướng của văn bản;

Hình 7 là hình vẽ mô tả ví dụ về hiệu quả hiệu chỉnh hướng văn bản tiếng Việt;

Hình 8 là hình vẽ mô tả ví dụ về hiệu quả nhận diện chữ viết tiếng Việt kết hợp với hiệu quả gom đoạn sử dụng giải pháp đề xuất trong sáng chế;

Hình 9 là hình vẽ mô tả ví dụ về hiệu quả nhận diện văn bản tiêng Việt với độ tin cậy cao.

Mô tả chi tiết sáng chế

Để giải quyết những vấn đề tồn tại trong quá trình tìm kiếm nội dung văn bản, giải pháp được đề xuất có một số điểm đặc biệt so với những giải pháp hiện có, bao gồm:

+ Lượng hóa đặc điểm về vị trí và nội dung của từng thành phần văn bản vào trong bảng đặc trưng;

+ Sự tương quan về vị trí, cũng như về nội dung, đặc trưng học sâu của văn bản hoàn toàn được xử lý thông qua tính toán trên ma trận, bảo đảm hiệu năng tìm kiếm là cao nhất;

+ Nhiều kết quả tìm kiếm được đưa ra, được sắp xếp theo độ tin cậy từ cao xuống thấp;

Nhằm thực hiện việc đặc trưng hóa văn bản và tối ưu tìm kiếm, phương pháp tìm kiếm nội dung dựa trên truy vấn khối và đặc trưng văn bản chủ yếu bao gồm quá trình định vị, ghép đoạn cho văn bản và nhận diện chữ viết, quá trình trích xuất đặc trưng, bổ sung thông tin cho khối đặc trưng của văn bản, quá trình truy vấn nội dung văn bản. Tham chiếu Hình 1, phương pháp được thực hiện với ba mô-đun chính bao gồm:

+ Mô-đun trích xuất thông tin văn bản: chủ yếu sử dụng những phương pháp tăng cường ảnh và hiệu chỉnh hướng đối với văn bản, nhằm tăng độ tương phản và tăng hiệu quả nhận diện chữ viết. Trong đó, việc định vị và nhận diện chữ viết sử dụng mô hình học sâu để thực hiện. Ngoài ra, phương pháp đồ thị được áp dụng để đặc trưng hóa thông tin và gom nhóm, tách các thành phần trong văn bản thành các đoạn có tính phân biệt về mặt không gian.

+ Mô-đun trích xuất đặc trưng: trước hết tách những từ đơn thông thường và các ký tự đặc biệt. Tiếp theo thực hiện bóc tách những từ (có ngữ nghĩa) trong từng

câu. Các đặc trưng TF-IDF được tính toán từ xác suất xuất hiện của từ trong câu, kết hợp với xác suất không xuất hiện trong câu khác của từ đó. Bên cạnh đó, đặc trưng học sâu được trích xuất từ các mô hình ngôn ngữ lớn cũng được sử dụng để so khớp và truy vấn đối với nội dung văn bản cần truy vấn có độ dài nhất định.

+ Mô-đun so khớp tìm kiếm nội dung: nội dung văn bản cần truy vấn được đặc trưng hóa (TF-IDF và đặc trưng học sâu), sau đó tính toán tương quan cô-sin với ma trận lưu trữ đặc trưng của từng câu, từ kết quả này sẽ đưa ra gợi ý về kết quả truy vấn, với độ tin cậy từ cao xuống thấp.

Cụ thể, phương pháp tìm kiếm nội dung dựa trên truy vấn khối và đặc trưng văn bản bao gồm các bước như sau:

Bước 1: quá trình định vị, ghép đoạn cho văn bản và nhận diện chữ viết;

Tham chiếu Hình 2, văn bản cần xử lý là ảnh được quét từ những tệp cứng (số hóa văn bản). Những ảnh văn bản này có chất lượng tốt-xấu không đồng đều, ví dụ như đối với những tệp bị lệch, cong hoặc là do thời gian lưu trữ lâu bị vàng ố, dẫn đến độ tương phản thấp, chữ viết trên văn bản trở nên khó đọc hơn.

Tại bước 110, ảnh văn bản được thực hiện tiền xử lý, bao gồm:

+ Tăng cường ảnh: do đặc điểm của ảnh văn bản, đó là chữ thông thường chỉ có màu đen, màu xanh, màu đỏ hoặc các màu có phân phối màu gần với màu cơ bản (màu xanh, màu đỏ, màu xanh nước biển), và nền văn bản thường có màu trắng, hoặc các màu có phân phối gần với màu trắng (có độ xám cao), nên áp dụng phương pháp lọc và tăng cường bởi phân bố màu xám (histogram) để nâng cao chất lượng ảnh văn bản.

Trong đó, sử dụng phương pháp nhị phân hóa cục bộ dựa trên ngưỡng tự thích ứng (ngưỡng theo phân bố Gaussian) để chuyển ảnh văn bản ban đầu thành ảnh nhị phân, ảnh nhị phân M_b này được dùng để tính toán ảnh đã được tăng cường $I_{enhanced}$ từ ảnh văn bản ban đầu I_{origin}. Cụ thể:

$$I_{enhanced}(x,y) = \begin{cases} I_{origin}(x,y) \text{ if } M_b(x,y) = 1 \\ 255 \text{ others} \end{cases}$$

Công thức này là công thức chung thường được dùng để bóc tách những phần cần giữ lại trong ảnh, trong khi những vùng ảnh khác được gán giá trị tối đa (255).

+ Loại bỏ nhiễu: ảnh sau khi tăng cường được tích chập với một toán tử Gaussian để làm mượt và loại bỏ nhiễu. Trong đó, giá trị phương sai và trung bình của nhân Gaussian được tùy chỉnh ở phạm vi giá trị nhỏ, với kích thước nhân là 3x3. Trong đó, kích thước nhân 3x3 được chọn do tỉ lệ này phù hợp với việc xử lý chữ cái trong văn bản (kích thước nhỏ).

+ Hiệu chỉnh hướng văn bản: đối với nhiều loại văn bản, sai số khi đưa bản giấy vào máy quét hoặc là do bản thân bố cục có sẵn của văn bản dẫn đến từng dòng trong văn bản lệch với phương thẳng đứng một góc khác 90 độ. Điều này sẽ dẫn đến sai lệch trong việc định vị và nhận diện ký tự trên văn bản, cũng như dẫn đến khó khăn khi tiến hành gom đoạn cho văn bản.

Trong quá trình này, đầu tiên tính toán đường biên trong ảnh văn bản (sau khi đã được tăng cường và loại bỏ nhiễu) sử dụng giải thuật Canny. Phương pháp biến đổi đường thẳng Hough được sử dụng để xác định những đường thẳng có thể xuất hiện trong văn bản.

Do đặc điểm của ảnh văn bản, đó là phần lớn các dòng chữ sẽ cùng một hướng, dẫn đến các đường thẳng Hough sẽ có chỉ số góc (so với phương thẳng đứng) như nhau. Gọi tập hợp góc các đường Hough trên ảnh là *H(k)*, thì góc của văn bản được tính toán như sau:

$$\text{angle} = \text{argmax}(H(k))$$

Trong công thức trên, góc của văn bản được coi là góc có tần suất xuất hiện nhiều nhất trong văn bản, tính theo góc của các đường thẳng Hough được trích xuất từ ảnh văn bản.

Sau khi có được góc văn bản, chỉ cần xoay văn bản theo chiều ngược lại (-angle) thì sẽ được một ảnh văn bản mới, trong đó phần lớn các hàng chữ trong văn bản đều vuông góc với phương thẳng đứng.

Sau giai đoạn tiền xử lý tại bước 110, ảnh văn bản đã được cải thiện chất lượng, cũng như hiệu chỉnh lại phương hướng, từ đó bảo đảm đầu vào cho mô hình định vị chữ viết cũng như trích xuất đặc trưng.

Tại bước 120: trích xuất thông tin không gian; tại bước này các dòng chữ trong văn bản sẽ được định vị, khi sử dụng mô hình học sâu. Mô hình học sâu này có thể sử dụng các công cụ mã nguồn mở hoặc tự huấn luyện với dữ liệu (nếu đặc thù), bởi vì tính thông dụng của bài toán này rất cao, có thể sử dụng mô hình định vị cho những ngôn ngữ tương đồng (với tiếng Việt chúng ta là chữ La-tin).

Với những mô hình học sâu định vị dòng chữ viết trên văn bản thì đầu vào là ảnh văn bản (đã qua tiền xử lý), và kết quả nhận được là một chuỗi những vị trí được định vị bởi *tọa độ bốn góc đánh dấu vịt rí và độ tin cậy* $[[(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)c], \dots]$ cho mỗi dòng chữ được phát hiện trên văn bản.

Với kết quả trên, vẫn chỉ là những dòng chữ rời rạc, chưa có mối quan hệ không gian, để xác định được những dòng văn bản nào là cùng trên một đoạn văn. Quá trình xác định đoạn văn rất quan trọng, liên quan trực tiếp đến độ chính xác của đặc trưng khi tìm kiếm nội dung trong văn bản.

Trong phương pháp được đề xuất trong sáng chế này, thực hiện ghép đoạn văn theo mối quan hệ về độ cao dòng chữ, khoảng cách gần nhất giữa các dòng chữ. Hay nói cách khác, những dòng chữ cùng hướng, gần nhau có chiều cao tương đồng, sẽ được ghép lại thành một đoạn văn, được thực hiện bởi phương pháp gom nhóm đồ thị. Với khoảng cách dòng lớn hơn một ngưỡng (ngưỡng tự thích ứng) thì sẽ được coi là vị trí chuyển tiếp giữa hai đoạn văn khác nhau. Quá trình thực hiện cụ thể như sau:

+ Tách các dòng cùng hướng: từ tọa độ bốn điểm góc của dòng chữ, góc (so với phương thẳng đứng) có thể được tính toán dễ dàng theo công thức đường thẳng. Các nhóm văn bản được tách theo góc thông qua phương pháp gom nhóm phân cụm tích lũy dựa trên lan truyền quan hệ (Affinity Propagation). Đối với những nhóm dòng văn bản có góc khác so với góc phổ biến nhất là phương ngang thì việc ghép các đoạn cũng như đối với nhóm góc phương ngang, chỉ khác là thêm một thao tác xoay các nhóm đó về theo phương ngang, ở trong nội dung mô tả này chỉ mô tả phần ghép đoạn đối với các dòng phương ngang.

+ Ghép đoạn các dòng cùng hướng: chuỗi các điểm trung tâm của các dòng chữ trên văn bản L(i)=(tọa độ tâm theo trục x, tọa độ tâm theo trục y), hay $L(i) = \{\frac{x_1^i + x_2^i}{2}, \frac{y_1^i + y_2^i}{2}\}$ được gom thành từng nhóm với mức độ khoảng cách gần xa thông qua phương pháp gom nhóm dựa trên mật độ (Density Based Clustering, DBSCAN). Trong đó, giá trị mật độ nhóm *ép-si-lon* có thể điều chỉnh phù hợp sao cho phù hợp với mục tiêu văn bản cần xử lý.

Lúc này mỗi một dòng văn bản đều được gán thêm một nhãn xác định đoạn, ví dụ như dòng văn bản số 0, 1, 3 thuộc đoạn số 1, dòng văn bản số 4, 6, 8 thuộc đoạn số 0…

Tiếp theo, như mô tả tại bước 130: trích xuất thông tin nội dung, những dòng chữ này được nhận diện chữ viết bởi một mô hình học sâu (được huấn luyện bởi bộ dữ liệu với hơn 900,000 mẫu giả lập và 100,000 mẫu thực tế). Mỗi một dòng văn bản sẽ nhận diện ra một dòng chữ tương ứng, và được lưu trữ vào ma trận mô tả thông tin nội dung với cấu trúc dữ liệu được trích xuất bao gồm các trường thông tin *[vị trí dòng chữ, nhóm đoạn văn bản, nội dung chữ viết được nhận diện]*. Ma trận thông tin này được dùng để trích xuất đặc trưng văn bản, phục vụ cho quá trình truy vấn, tìm kiếm nội dung.

Bước 2: quá trình trích xuất đặc trưng;

Tham chiếu Hình 3, quá trình trích xuất đặc trưng bao gồm: phân loại ký tự đặc biệt, tách từ; và tính toán đặc trưng, ở đây có hai chủng loại đặc trưng đó là TF-IDF và đặc trưng học sâu từ mô hình mạng bộ nhớ ngắn dài (LSTM).

Tại bước 210, tiến hành phân loại ký tự đặc biệt, và tách từ ngữ. Đối với các loại ngôn ngữ nói chung, và tiếng Việt nói riêng, trước khi xử lý cần phân loại những ký tự đặc biệt ví dụ như "!", "-", hoặc những ký hiệu như "TW-QĐ"..., chủ yếu sử dụng phương pháp đối chiếu theo từ điển cố định.

Đối với lĩnh vực xử lý ngôn ngữ tự nhiên, với các ngôn ngữ trên thế giới nói chung và tiếng Việt nói riêng, việc xác định từ ngữ được tổ hợp từ các từ đơn là bước đầu tiên để phân tích ngữ nghĩa của câu văn, hay đoạn văn. Ví dụ như câu "Ông Lê Văn A hiện đang giảng dạy tại Đại học quốc gia Việt Nam", thì các từ đơn riêng lẻ như "giảng", "dạy"... không mô tả được ngữ nghĩa mà người viết muốn mô tả, thay vì đó, ghép lại với nhau sẽ xác định được thông tin cụ thể mà câu văn muốn truyền đạt, đó là "Ông" / "Lê Văn A" / "hiện" / "đang" / "giảng dạy" / "tại" / "Đại học" / "quốc gia" / "Việt Nam". Những từ ngữ được tách ra trong văn bản sẽ là căn cứ để trích xuất đặc trưng, mà có sự mô tả chính xác nhất đến nội dung được mô tả.

Mô hình ẩn mác-cốp (Hidden Markov Model, HMM) được sử dụng để tách từ ngữ trong văn bản, mô hình này được huấn luyện với 100,000 câu đã được gán nhãn, đánh dấu các loại từ loại (động từ, danh từ, trạng từ…). Sau khi huấn luyện, các câu văn bản sẽ được đưa vào HMM, và kết quả nhận được là các từ ngữ trong câu, được cấu trúc từ những từ đơn liên tiếp. Đối với văn bản cần xử lý, mô hình HMM này sẽ trích xuất ra toàn bộ số từ có trong đó, và đó là căn cứ để tính toán xác suất xuất hiện trong các phân đoạn, được tách ra từ các đoạn văn trong văn bản này tại bước 220.

Tại bước 220, tất cả những đoạn văn (số dòng tối thiểu của đoạn văn là một) được phân tách thành những phân đoạn nhỏ, những phân đoạn này có thể là một câu

hoàn chỉnh hoặc không, cũng có thể là những phân đoạn chứa nội dung của cả hai câu liên tiếp (giao giữa hai câu liên tiếp).

Khi tiến hành phân đoạn, mỗi đoạn văn được cắt theo chiều từ trái sang phải, với độ dài là *N* từ (không phải đơn từ) được tính theo những từ ngữ được phân đoạn ở bước 210 (không bao gồm dấu cách), và với độ trùng lặp giữa hai phân đoạn là *N/3*, giải thuật phân đoạn được thực hiện cụ thể như sau:

Giải thuật phân đoạn nội dung văn bản
Đầu vào: Các đoạn văn được tách ra từ văn bản Độ dài phân đoạn *N*
Đầu ra: Chuỗi phân đoạn văn bản
Thực hiện: Gọi chuỗi lưu trữ phân đoạn là *S* Trên toàn bộ các đoạn văn: Đối với mỗi đoạn văn: Lấy lần lượt N từ tính từ trái sang phải Phân đoạn sau được tính từ điểm cuối phân đoạn trước - N/3 Cập nhật phân đoạn vào *S* *Kết thúc: S chứa tất cả các phân đoạn*

Tất cả các phần tử trong *S* được trích xuất đặc trưng TF-IDF và đặc trưng học sâu, trong đó:

+ Đặc trưng TF-IDF được tính toán tại bước 230, TF-IDF (*term frequency - inverse document frequency*) là một phương thức thống kê được biết đến rộng rãi nhất để xác định độ quan trọng của một từ trong đoạn văn bản trong một tập nhiều đoạn văn bản khác nhau.

TF (Term Frequency): là tần suất xuất hiện của một từ trong một đoạn văn bản.

$$TF = \frac{\text{Số lần từ xuất hiện trong phân đoạn}}{\text{Tổng số từ trong phân đoạn}}$$

Ví dụ, với câu "Anh ấy/ có/ một/ con chó/ và/ cũng/ có/ một/ con mèo/ màu trắng", thì TF("có")=2/10.

IDF (Inverse Document Frequency): tính toán độ quan trọng của một từ, với những từ xuất hiện trong nhiều phân đoạn thì chỉ số IDF càng thấp, và ngược lại.

$$IDF(w) = \ln\left(\frac{\text{Tổng số phân đoạn}}{\text{Số phân đoạn có xuất hiện từ "w"}}\right)$$

Đặc trưng TF-IDF là một véc-tơ có chiều dài bằng tổng số từ xuất hiện trong văn bản (được coi là cơ sở dữ liệu từ ngữ), trong mỗi phân đoạn, mỗi từ sẽ được tính chỉ số TF và IDF, đặc trưng TF-IDF được xác định như sau:

$$TF - IDF(w) = TF(w) * IDF(w)$$

Những từ không xuất hiện trong phân đoạn thì đều có giá trị bằng 0.

Tại 232 đặc trưng TF-IDF của tất cả các phân đoạn được lưu trữ trên bộ nhớ để làm cơ sở so khớp, tìm kiếm nội dung văn bản.

+ Đặc trưng học sâu tại bước 231: sử dụng mô hình LSTM (được huấn luyện trên tập dữ liệu có quy mô 100,000 mẫu), khi đưa một câu văn vào mô hình, sẽ được một véc-tơ đặc trưng có chiều dài là 512. Tại 233, đặc trưng học sâu của tất cả các phân đoạn được lưu trữ trên bộ nhớ để làm cơ sở so khớp, tìm kiếm nội dung văn bản.

Bước 3: quá trình truy vấn nội dung văn bản;

Sau khi phân tích, xử lý và trích xuất đặc trưng của văn bản, nội dung cần truy vấn, tìm kiếm cũng sẽ được đưa qua mô-đun trích xuất đặc trưng tại bước 230, 231.

Khi tính đặc trưng TF-IDF, cần đồng thời cập nhật đặc trưng của cả văn bản mục tiêu lẫn đặc trưng của nội dung cần truy vấn. Coi nội dung cần truy vấn như một phân đoạn trong văn bản, và bài toán tìm kiếm nội dung sẽ chuyển thành bài toán tìm độ tương đồng của các phân đoạn. Cách tính toán đặc trưng như tại bước 230. Lúc này, độ dài của véc-tơ đặc trưng này sẽ = *(tổng số phân đoạn trong văn bản mục tiêu + 1)*.

Khi tính đặc trưng học sâu, nội dung cần truy vấn được đưa qua mô hình LSTM để trích xuất đặc trưng có số chiều là 512.

Thực hiện truy vấn (bao gồm trích xuất đặc trưng), được thể hiện như trong Hình 4. Quá trình truy vấn này được chia thành hai trường hợp:

+ Thứ nhất là khi số từ trong nội dung truy vấn nhỏ hơn N (là số từ lớn nhất trong mỗi phân đoạn). Lúc này sẽ sử dụng đặc trưng TF-IDF để tìm kiếm.

+ Thứ hai là khi số từ lớn hơn hoặc bằng N, thì sẽ sử dụng đặc trưng học sâu để tìm kiếm.

Khi so khớp tìm kiếm, độ tương quan cô-sin được sử dụng để xác định phân đoạn nào có nội dung tương đồng với nội dung cần truy vấn. Nhằm đẩy nhanh quá trình so khớp, phương pháp tính toán theo khối được áp dụng khi tính toán tương quan cô-sin (bước 310). Phương pháp này có thể thực hiện trên nhiều nền tảng phần cứng khác nhau, như CPU và GPU. Kết quả so khớp là một chuỗi các giá trị tương quan cô-sin giữa véc-tơ đặc trưng của nội dung cần truy vấn và véc-tơ đặc trưng của từng phân đoạn, các giá trị này có phạm vi từ 0 đến 1, với giá trị bằng 1 nói lên độ tương quan là cao nhất và bằng 0 là thấp nhất. Kết quả tìm kiếm là K phân đoạn có chỉ số tương quan với nội dung cần truy vấn là cao nhất, với K là hằng số được xác định từ trước. Tại 320, đưa ra gợi ý K đoạn có nội dung khớp nhất so với nội dung văn bản cần truy vấn.

Kết thúc bước 3, những đoạn văn bản (K đoạn) có nội dung tương đồng nhất với nội dung văn bản cần truy vấn, từ đó thực hiện được tính năng đề xuất trong sáng chế.

Ví dụ thực hiện sáng chế

Phương pháp được đề xuất có khả năng trích xuất thông tin cũng như có khả năng thực hiện tìm kiếm theo nội dung truy vấn nhất định, hiệu quả trong việc tìm kiếm nội dung đã được số hóa, lưu trữ.

Trong quá trình xử lý, tiền xử lý ảnh văn bản có ảnh hưởng rất lớn đến hiệu quả định vị cũng như nhận diện văn bản, trong đó quan trọng có tăng cường độ tương phản (ví dụ như trong Hình 5), và hiệu chỉnh hướng văn bản (ví dụ như trong Hình 6 và Hình 7).

Trong Hình 5, ảnh ban đầu đưa vào (A) có độ tương phản thấp, khả năng phân biệt của mô hình định vị chữ viết sẽ không tốt, còn sau khi tăng cường (B) thì phần chữ viết và nền ảnh đã có sự phân tách rất rõ ràng, là tiền đề vững chắc cho quá trình định vị dòng văn bản.

Trong Hình 6, những dòng chữ được xác định hướng (ban đầu) bởi đường thẳng Hough, sau khi xác định được hướng chính (có nhiều đường thẳng cùng hướng đó nhất)

thì sẽ xoay ảnh theo hướng ngược lại với góc lệch, sẽ hiệu chỉnh được ảnh mà các dòng chữ có phương vuông góc với phương thẳng đứng. Đây sẽ là điểm tham chiếu tin cậy cho quá trình ghép đoạn trong bước xử lý sau. Ví dụ về hiệu chỉnh hướng trên văn bản tiếng Việt như mô tả trong Hình 7.

Ví dụ về hiệu quả ghép đoạn văn được mô tả như trong Hình 8, trong đó, các đoạn đã được tách biệt về mặt không gian, mang lại ý nghĩa về mặt bóc tách các trường thông tin.

Ví dụ về hiệu quả nhận diện chữ viết tiếng Việt kết hợp với hiệu quả ghép đoạn được mô tả như trong Hình 9. Trong đó, các trường thông tin đã được gom nhóm một cách chính xác, và độ tin cậy nhận diện chữ viết cũng rất cao.

Hiệu quả có thể đạt được của sáng chế

Phương pháp được đề cập trong sáng chế đã được sử dụng trong các giải pháp trí tuệ nhân tạo và thị giác máy tính, phục vụ cho các sản phẩm số hóa và tìm kiếm nội dung văn bản của Tổng công ty Công nghiệp Công nghệ cao Viettel.

Yêu cầu bảo hộ

1. Phương pháp tìm kiếm nội dung dựa trên truy vấn khối và đặc trưng văn bản bao gồm:

bước 1: quá trình định vị, ghép đoạn cho văn bản và nhận diện chữ viết; việc định vị để xác định vị trí của từng câu văn bản, dựa trên tiền đề mảng văn bản đã được tiền xử lý, hiệu chỉnh hướng, đồng thời nhận diện văn bản trên từng vị trí văn bản đã định vị được bởi mô hình học sâu; dựa trên thông tin vị trí câu văn bản, những câu văn bản cùng tính chất và gần nhau được ghép lại thành các đoạn văn khác nhau;

bước 2: quá trình trích xuất đặc trưng, bổ sung thông tin cho khối đặc trưng của văn bản: trích xuất đặc trưng của văn bản, mỗi đặc trưng này là một véc-tơ có cùng chiều dài, và được cập nhật đồng bộ với vị trí của các đoạn văn, câu văn;

bước 3: quá trình truy vấn nội dung văn bản: tự động phân tích kết cấu của nội dung văn bản đưa vào truy vấn, sử dụng đồng thời phương án tìm kiếm theo từ khóa và tìm kiếm theo đặc trưng bằng cách tính toán tương quan cô-sin giữa đặc trưng văn bản cần so khớp và ma trận đặc trưng của văn bản đã được tính toán từ trước.

2. Phương pháp theo điểm 1, trong đó quá trình định vị, ghép đoạn cho văn bản và nhận diện chữ viết bao gồm:

quá trình tiền xử lý ảnh;

quá trình định vị và ghép đoạn các dòng trong văn bản ;

quá trình nhận diện chữ trong dòng văn bản.

3. Phương pháp theo điểm 2, trong đó quá trình tiền xử lý bao gồm:

tăng cường ảnh: áp dụng phương pháp lọc và tăng cường bởi phân bố màu xám (histogram) để nâng cao chất lượng ảnh văn bản; trong đó, sử dụng phương pháp nhị phân hóa cục bộ dựa trên ngưỡng tự thích ứng (ngưỡng theo phân bố gaussian) để chuyển ảnh văn bản ban đầu thành ảnh nhị phân, ảnh nhị phân M_b này được dùng để tính toán ảnh đã được tăng cường $I_{enhanced}$ từ ảnh văn bản ban đầu I_{origin}; cụ thể:

$$I_{enhanced}(x,y) = \begin{cases} I_{origin}(x,y) \text{ if } M_b(x,y) = 1 \\ 255 \text{ others} \end{cases}$$

loại bỏ nhiễu: ảnh sau khi tăng cường được tích chập với một toán tử gaussian để làm mượt và loại bỏ nhiễu; trong đó, giá trị phương sai và trung bình của nhân gaussian được tùy chỉnh ở phạm vi giá trị nhỏ, với kích thước nhân là 3x3;

hiệu chỉnh hướng văn bản: đầu tiên tính toán đường biên trong ảnh văn bản (sau khi tăng cường và loại bỏ nhiễu) sử dụng giải thuật canny; phương pháp biến đổi đường thẳng hough được sử dụng để xác định những đường thẳng có thể xuất hiện trong văn bản; gọi tập hợp góc các đường hough trên ảnh là *H(k)*, thì góc của văn bản được tính toán như sau:

$$\text{angle} = \text{argmax}(\text{H}(\text{k}))$$

trong công thức trên, góc của văn bản được coi là góc có tần suất xuất hiện nhiều nhất trong văn bản, tính theo góc của các đường thẳng hough được trích xuất từ ảnh văn bản;

sau khi có được góc văn bản, chỉ cần xoay văn bản theo chiều ngược lại (-angle) thì sẽ được một ảnh văn bản mới, trong đó phần lớn các hàng chữ trong văn bản đều vuông góc với phương thẳng đứng.

4. Phương pháp theo điểm 2, trong đó, tại quá trình định vị và ghép đoạn dòng văn bản:

sử dụng mô hình học sâu định vị dòng chữ viết trên văn bản, đầu vào là ảnh văn bản (đã qua tiền xử lý), và kết quả nhận được là một chuỗi những vị trí được định vị bởi *tọa độ bốn góc đánh dấu vịt rí và độ tin cậy* $[[(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)c], \dots]$ cho mỗi dòng chữ được phát hiện trên văn bản;

thực hiện ghép đoạn văn theo mối quan hệ về độ cao dòng chữ, khoảng cách gần nhất giữa các dòng chữ; hay nói cách khác, những dòng chữ cùng hướng, gần nhau có chiều cao tương đồng, sẽ được ghép lại thành một đoạn văn, được thực hiện bởi phương pháp gom nhóm đồ thị; với khoảng cách dòng lớn hơn một ngưỡng (ngưỡng tự thích ứng) thì sẽ được coi là vị trí chuyển tiếp giữa hai đoạn văn khác nhau; quá trình thực hiện cụ thể như sau:

tách các dòng cùng hướng: từ tọa độ bốn điểm góc của dòng chữ, góc (so với phương thẳng đứng) có thể được tính toán dễ dàng theo công thức đường thẳng; các nhóm văn bản được tách theo góc thông qua phương pháp gom nhóm phân cụm tích lũy dựa trên lan truyền quan hệ (affinity propagation); đối với những nhóm dòng văn bản có góc khác so với góc phổ biến nhất là phương ngang thì việc ghép các đoạn cũng như đối với nhóm góc phương ngang, chỉ khác là thêm một thao tác xoay các nhóm đó về theo phương ngang, ở trong nội dung mô tả này chỉ mô tả phần ghép đoạn đối với các dòng phương ngang;

ghép đoạn các dòng cùng hướng: chuỗi các điểm trung tâm của các dòng chữ trên văn bản $L(i) = \{\frac{x_1^i + x_2^i}{2}, \frac{y_1^i + y_2^i}{2}\}$ được gom thành từng nhóm với mức độ khoảng cách gần xa thông qua phương pháp gom nhóm dựa trên mật độ; trong đó, giá trị mật độ nhóm ép-si-lon có thể điều chỉnh phù hợp sao cho phù hợp với mục tiêu văn bản cần xử lý.

5. Phương pháp theo điểm 2, trong đó quá trình nhận diện dòng chữ bao gồm:

những dòng chữ này được nhận diện chữ viết bởi một mô hình học sâu (được huấn luyện bởi bộ dữ liệu với hơn 900,000 mẫu giả lập và 100,000 mẫu thực tế); mỗi một dòng văn bản sẽ nhận diện ra một dòng chữ tương ứng, và được lưu trữ vào ma trận mô tả thông tin nội dung với cấu trúc dữ liệu bao gồm các trường thông tin [vị trí dòng chữ, nhóm đoạn văn bản, nội dung chữ viết được nhận diện].

6. Phương pháp theo điểm 1, trong đó quá trình trích xuất đặc trưng bao gồm:

phân loại nhóm từ;

tách từ ngữ;

tính toán đặc trưng TF-IDF;

tính toán đặc trưng học sâu.

7. Phương pháp theo điểm 6, trong đó quá trình phân loại nhóm từ bao gồm:

tiến hành phân loại ký tự đặc biệt, và tách từ ngữ; đối với các loại ngôn ngữ nói chung, và tiếng Việt nói riêng, trước khi xử lý cần phân loại những ký tự đặc biệt ví dụ như "!", "-", hoặc những ký hiệu như "TW-QĐ"..., chủ yếu sửu dụng phương pháp đối chiếu theo từ điển cố định.

8. Phương pháp theo điểm 6, trong đó quá trình tách từ ngữ bao gồm:

mô hình ẩn **mác-cốp** (hidden markov model, HMM) được sử dụng để tách từ ngữ trong văn bản, mô hình này được huấn luyện với 100,000 câu đã được gán nhãn, đánh dấu các loại từ loại (động từ, danh từ, trạng từ…); sau khi huấn luyện, các câu văn bản sẽ được đưa vào HMM, và kết quả nhận được là các từ ngữ trong câu, được cấu trúc từ những từ đơn liên tiếp; đối với văn bản cần xử lý, mô hình HMM này sẽ trích xuất ra toàn bộ số từ có trong đó.

Khi tiến hành phân đoạn, mỗi đoạn văn được cắt theo chiều từ trái sang phải, với độ dài là *N* từ (không phải đơn từ) được tính theo những từ ngữ đã được phân đoạn (không bao gồm dấu cách), và với độ trùng lặp giữa hai phân đoạn là *N/3*.

9. Phương pháp theo điểm 6, trong đó trích xuất đặc trưng TF-IDF bao gồm:

TF-IDF (*term frequency - inverse document frequency*) là một phương thức thống kê được biết đến rộng rãi nhất để xác định độ quan trọng của một từ trong đoạn văn bản trong một tập nhiều đoạn văn bản khác nhau:

TF (term frequency): là tần suất xuất hiện của một từ trong một đoạn văn bản:

$$\mathrm{TF} = \frac{\text{số lần từ xuất hiện trong phân đoạn}}{\text{tổng số từ trong phân đoạn}}$$

IDF (inverse document frequency): tính toán độ quan trọng của một từ, với những từ xuất hiện trong nhiều phân đoạn thì chỉ số IDF càng thấp, và ngược lại:

$$\mathrm{IDF(w)} = \ln\left(\frac{\text{tổng số phân đoạn}}{\text{số phân đoạn có xuất hiện từ "w"}}\right)$$

đặc trưng TF-IDF là một véc-tơ có chiều dài bằng tổng số từ xuất hiện trong văn bản (được coi là cơ sở dữ liệu từ ngữ), trong mỗi phân đoạn, mỗi từ sẽ được tính chỉ số TF và IDF, đặc trưng TF-IDF được xác định như sau:

$$\mathrm{TF-IDF(w)} = \mathrm{TF(w)} * \mathrm{IDF(w)}$$

những từ không xuất hiện trong phân đoạn thì đều có giá trị bằng 0.

10. Phương pháp theo điểm 6, trong đó trích xuất đặc trưng học sâu bao gồm:

sử dụng mô hình LSTM (được huấn luyện trên tập dữ liệu có quy mô 100,000 mẫu), khi đưa một câu văn vào mô hình, sẽ được một véc-tơ đặc trưng có chiều dài là 512; đặc trưng học sâu của tất cả các phân đoạn được lưu trữ trên bộ nhớ để làm cơ sở so khớp, tìm kiếm nội dung văn bản.

11. Phương pháp theo điểm 1, trong đó quá trình truy vấn nội dung văn bản bao gồm:

xác định chủng loại đặc trưng truy vấn;

tính toán đặc trưng nội dung cần truy vấn;

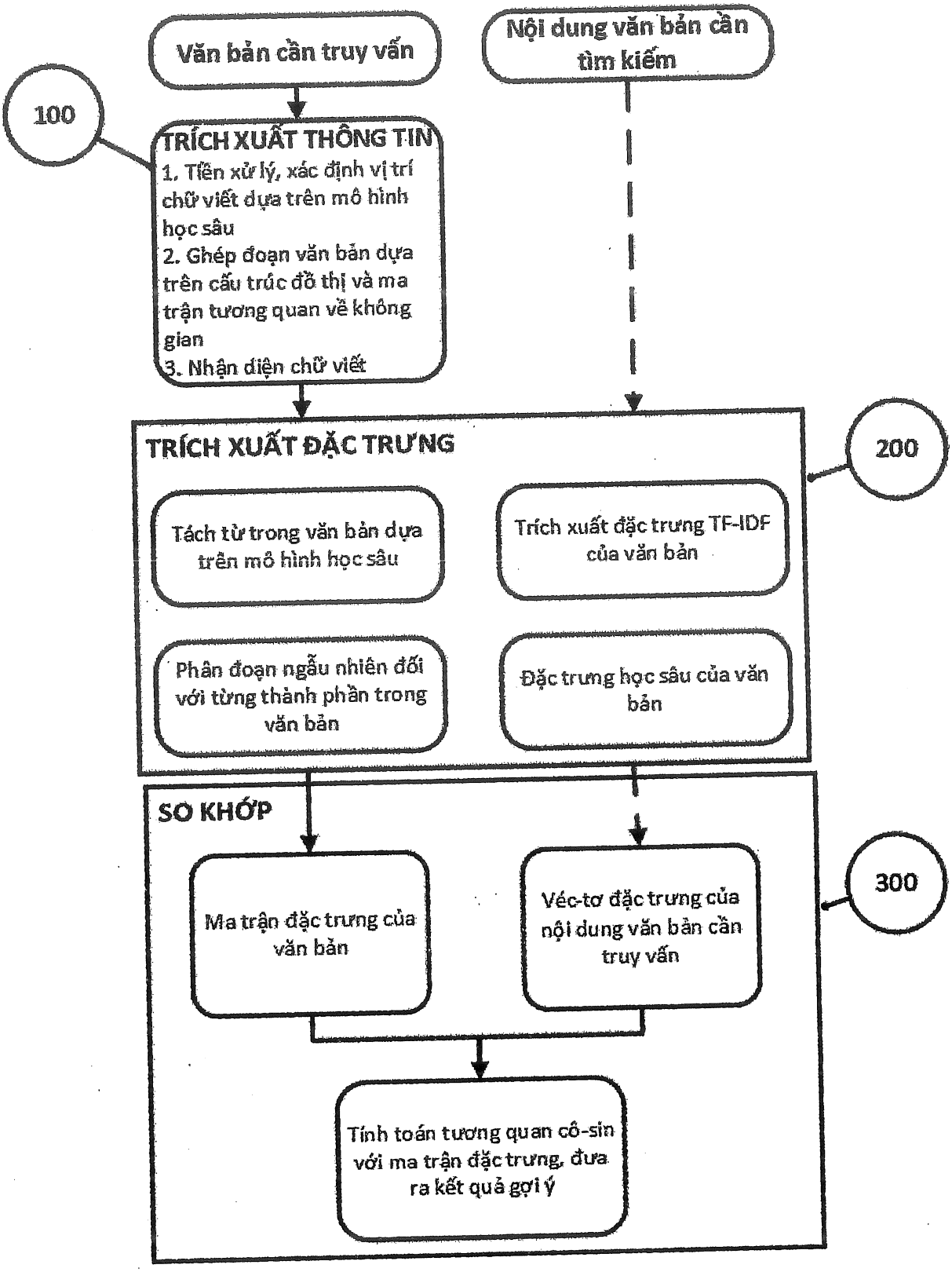
so khớp xác định những phân đoạn có nội dung tương đồng.

12. Phương pháp theo điểm 11, chủng loại đặc trưng truy vấn được xác định bởi độ dài (số từ không tính dấu cách) của nội dung cần truy vấn, nếu độ dài đó lớn hơn N (là số từ lớn nhất trong mỗi phân đoạn trong văn bản mục tiêu) thì sử dụng đặc trưng học sâu, ngược lại thì sử dụng đặc trưng TF-IDF.

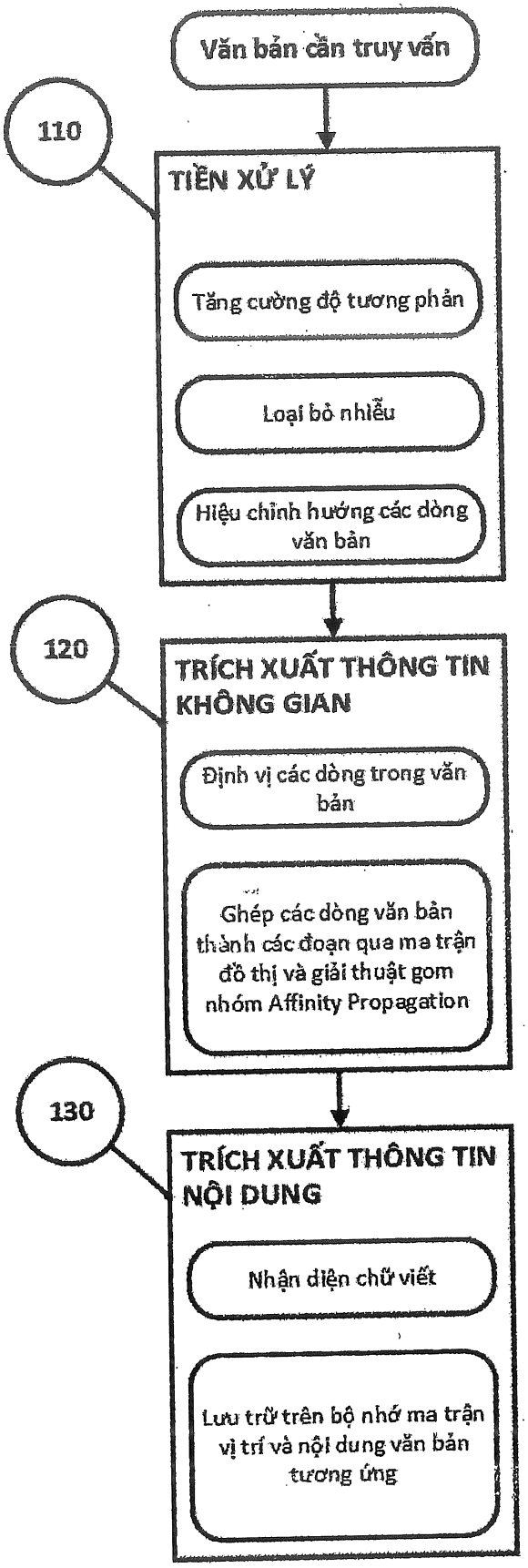
13. Phương pháp theo điểm 11, đặc trưng được tính toán như khi đặc trưng hóa văn bản mục tiêu.

14. Phương pháp theo điểm 11, trong đó tại quá trình so khớp xác định phân đoạn có nội dung tương đồng:

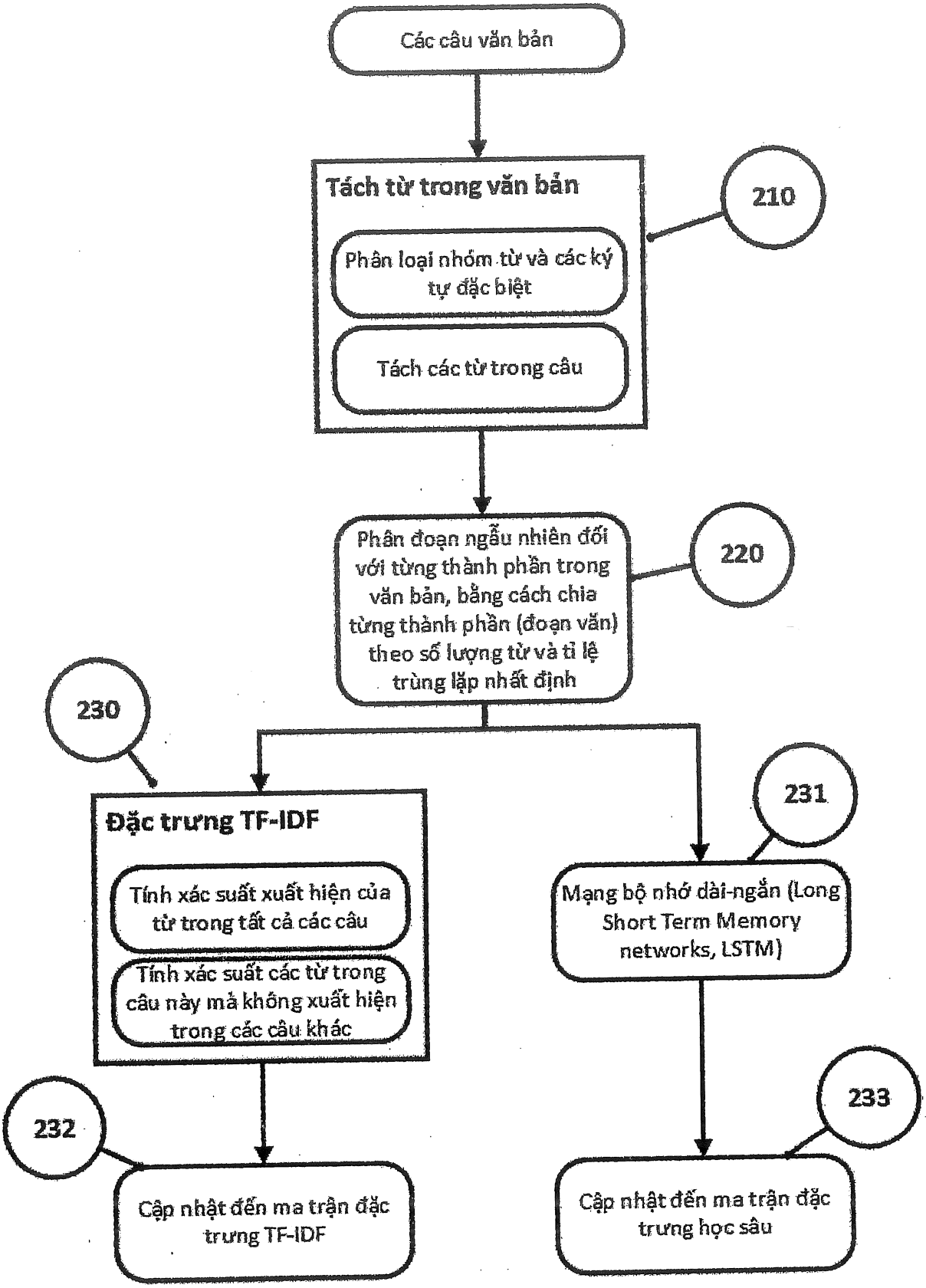
khi so khớp tìm kiếm, độ tương quan cô-sin được sử dụng để xác định phân đoạn nào có nội dung tương đồng với nội dung cần truy vấn; nhằm đẩy nhanh quá trình so khớp, phương pháp tính toán theo khối được áp dụng khi tính toán tương quan cô-sin; phương pháp này có thể thực hiện trên nhiều nền tảng phần cứng khác nhau, như CPU và GPU; kết quả so khớp là một chuỗi các giá trị tương quan cô-sin giữa véc-tơ đặc trưng của nội dung cần truy vấn và véc-tơ đặc trưng của từng phân đoạn, các giá trị này có phạm vi từ 0 đến 1, với giá trị bằng 1 nói lên độ tương quan là cao nhất và bằng 0 là thấp nhất; kết quả tìm kiếm là K phân đoạn có chỉ số tương quan với nội dung cần truy vấn là cao nhất, với K là hằng số được xác định từ trước.


Văn bản cần truy vấn
Nội dung văn bản cần tìm kiếm
100
TRÍCH XUẤT THÔNG TIN
1. Tiền xử lý, xác định vị trí chữ viết dựa trên mô hình học sâu
2. Ghép đoạn văn bản dựa trên cấu trúc đồ thị và ma trận tương quan về không gian
3. Nhận diện chữ viết
TRÍCH XUẤT ĐẶC TRƯNG
200
Tách từ trong văn bản dựa trên mô hình học sâu
Trích xuất đặc trưng TF-IDF của văn bản
Phân đoạn ngẫu nhiên đối với từng thành phần trong văn bản
Đặc trưng học sâu của văn bản
SO KHỚP
300
Ma trận đặc trưng của văn bản
Véc-tơ đặc trưng của nội dung văn bản cần truy vấn
Tính toán tương quan cô-sin với ma trận đặc trưng, đưa ra kết quả gợi ý

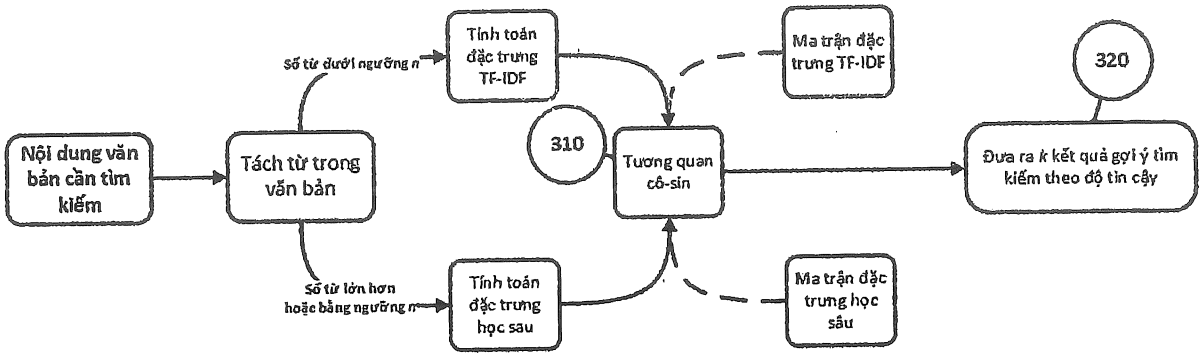
Hình 1


Văn bản cần truy vấn
110
TIỀN XỬ LÝ
Tăng cường độ tương phản
Loại bỏ nhiễu
Hiệu chỉnh hướng các dòng văn bản
120
TRÍCH XUẤT THÔNG TIN KHÔNG GIAN
Định vị các dòng trong văn bản
Ghép các dòng văn bản thành các đoạn qua ma trận đồ thị và giải thuật gom nhóm Affinity Propagation
130
TRÍCH XUẤT THÔNG TIN NỘI DUNG
Nhận diện chữ viết
Lưu trữ trên bộ nhớ ma trận vị trí và nội dung văn bản tương ứng

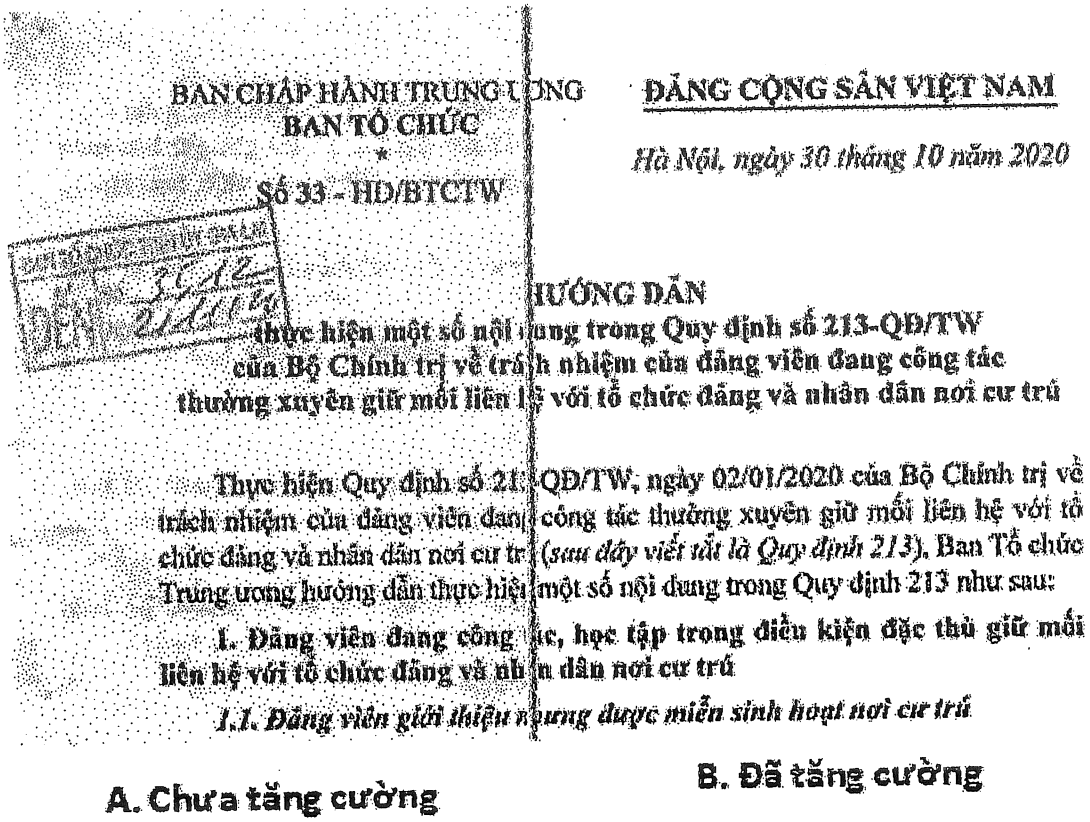
Hình 2


Các câu văn bản
Tách từ trong văn bản
210
Phân loại nhóm từ và các ký tự đặc biệt
Tách các từ trong câu
Phân đoạn ngẫu nhiên đối với từng thành phần trong văn bản, bằng cách chia từng thành phần (đoạn văn) theo số lượng từ và tỉ lệ trùng lặp nhất định
220
230
Đặc trưng TF-IDF
Tính xác suất xuất hiện của từ trong tất cả các câu
Tính xác suất các từ trong câu này mà không xuất hiện trong các câu khác
231
Mạng bộ nhớ dài-ngắn (Long Short Term Memory networks, LSTM)
232
Cập nhật đến ma trận đặc trưng TF-IDF
233
Cập nhật đến ma trận đặc trưng học sâu

Hình 3



Hình 4



BAN CHẤP HÀNH TRUNG ƯƠNG
BAN TỔ CHỨC
*
Số 33 - HD/BTCTW

ĐẢNG CỘNG SẢN VIỆT NAM

Hà Nội, ngày 30 tháng 10 năm 2020

HƯỚNG DẪN
thực hiện một số nội ung trong Quy định số 213-QĐ/TW
của Bộ Chính trị về trá h nhiệm của đảng viên đang công tác
thường xuyên giữ mối liên l với tổ chức đảng và nhân dân nơi cư trú

Thực hiện Quy định số 21 QĐ/TW, ngày 02/01/2020 của Bộ Chính trị về trách nhiệm của đảng viên đan công tác thường xuyên giữ mối liên hệ với tổ chức đảng và nhân dân nơi cư tr (*sau đây viết tắt là Quy định 213*), Ban Tổ chức Trung ương hướng dẫn thực hiệ một số nội dung trong Quy định 213 như sau:

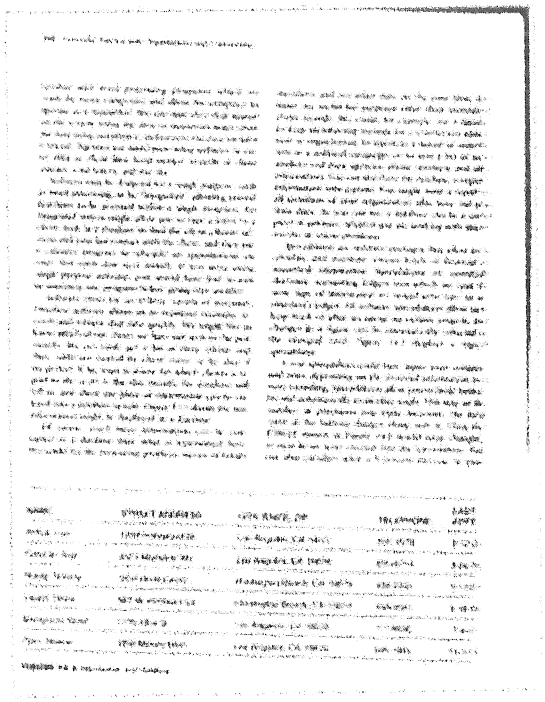
1. Đảng viên đang công c, học tập trong điều kiện đặc thù giữ mối liên hệ với tổ chức đảng và nh n dân nơi cư trú

1.1. Đảng viên giới thiệu n ung được miễn sinh hoạt nơi cư trú

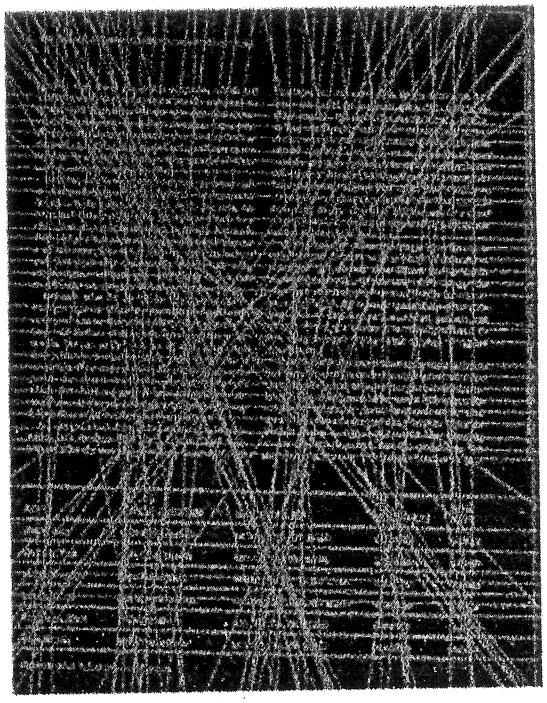
A. Chưa tăng cường

B. Đã tăng cường

Hình 5

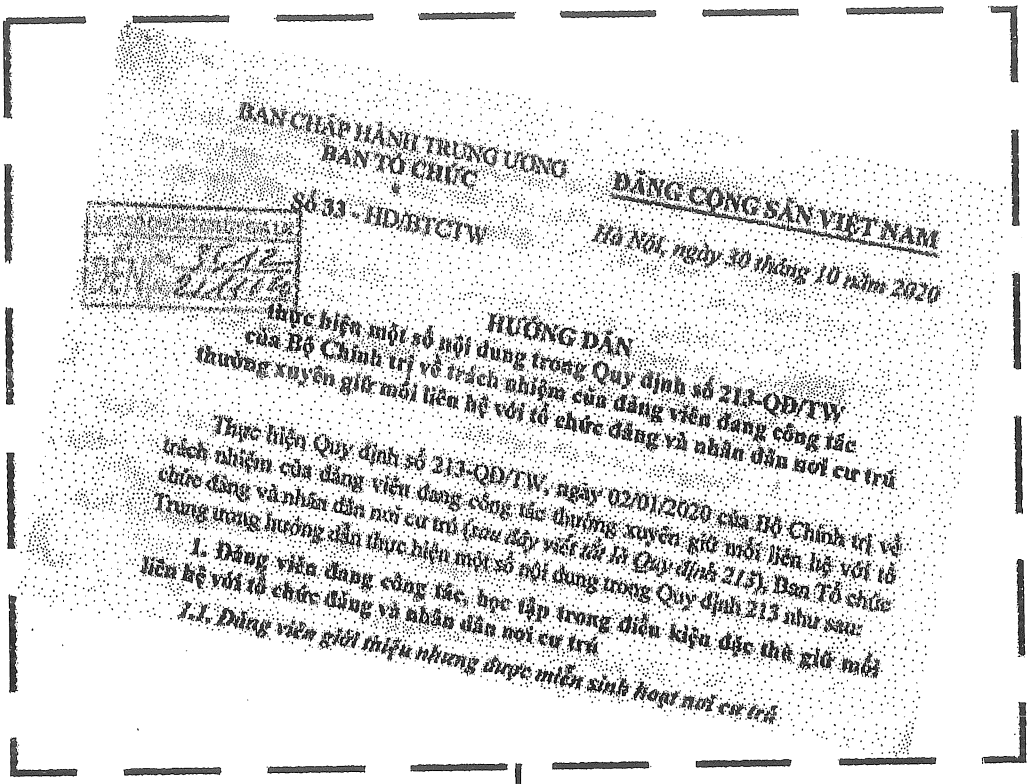


Ảnh cần xử lý



Các đường thẳng được phát hiện

Hình 6



BAN CHẤP HÀNH TRUNG ƯƠNG
BAN TỔ CHỨC
*
Số 33 - HD/BTCTW

ĐẢNG CỘNG SẢN VIỆT NAM

Hà Nội, ngày 30 tháng 10 năm 2020

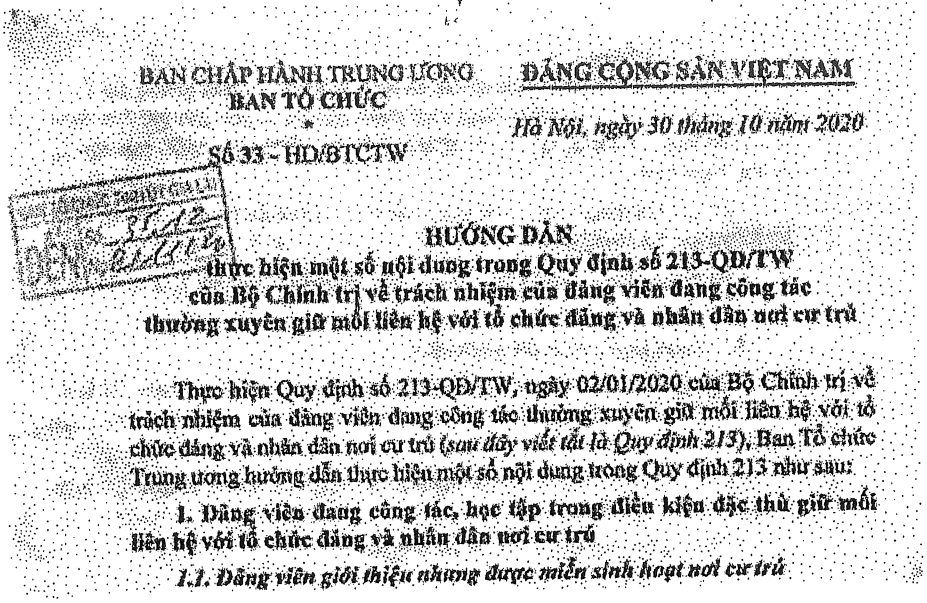
HƯỚNG DẪN
thực hiện một số nội dung trong Quy định số 213-QĐ/TW của Bộ Chính trị về trách nhiệm của đảng viên đang công tác thường xuyên giữ mối liên hệ với tổ chức đảng và nhân dân nơi cư trú

Thực hiện Quy định số 213-QĐ/TW, ngày 02/01/2020 của Bộ Chính trị về trách nhiệm của đảng viên đang công tác thường xuyên giữ mối liên hệ với tổ chức đảng và nhân dân nơi cư trú *(sau đây viết tắt là Quy định 213)*, Ban Tổ chức Trung ương hướng dẫn thực hiện một số nội dung trong Quy định 213 như sau:

1. Đảng viên đang công tác, học tập trong điều kiện đặc thù giữ mối liên hệ với tổ chức đảng và nhân dân nơi cư trú

1.1. Đảng viên giới thiệu nhưng được miễn sinh hoạt nơi cư trú

Hiệu chỉnh
hướng văn bản



BAN CHẤP HÀNH TRUNG ƯƠNG
BAN TỔ CHỨC
*
Số 33 - HD/BTCTW

ĐẢNG CỘNG SẢN VIỆT NAM

Hà Nội, ngày 30 tháng 10 năm 2020

HƯỚNG DẪN
thực hiện một số nội dung trong Quy định số 213-QĐ/TW của Bộ Chính trị về trách nhiệm của đảng viên đang công tác thường xuyên giữ mối liên hệ với tổ chức đảng và nhân dân nơi cư trú

Thực hiện Quy định số 213-QĐ/TW, ngày 02/01/2020 của Bộ Chính trị về trách nhiệm của đảng viên đang công tác thường xuyên giữ mối liên hệ với tổ chức đảng và nhân dân nơi cư trú *(sau đây viết tắt là Quy định 213)*, Ban Tổ chức Trung ương hướng dẫn thực hiện một số nội dung trong Quy định 213 như sau:

1. Đảng viên đang công tác, học tập trong điều kiện đặc thù giữ mối liên hệ với tổ chức đảng và nhân dân nơi cư trú

1.1. Đảng viên giới thiệu nhưng được miễn sinh hoạt nơi cư trú

Hình 7

BỘ NÔNG NGHIỆP VÀ PHÁT TRIỂN NÔNG THÔN

Số: 02/2011/TT-BNNPTNT

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

Hà Nội, ngày 21 tháng 01 năm 2011

THÔNG TƯ
Hướng dẫn nhiệm vụ quản lý nhà nước về chăn nuôi

Căn cứ Pháp lệnh Giống vật nuôi ngày 24 tháng 3 năm 2004;

Căn cứ Nghị định số 08/2010/NĐ-CP ngày 05 tháng 02 năm 2010 của Chính phủ về quản lý thức ăn chăn nuôi;

Căn cứ Nghị định số 01/2008/NĐ-CP ngày 03 tháng 01 năm 2008 của Chính phủ quy định chức năng, nhiệm vụ, quyền hạn và cơ cấu tổ chức của Bộ Nông nghiệp và Phát triển nông thôn; Nghị định số 75/2009/NĐ-CP ngày 10 tháng 9 năm 2009 của Chính phủ về việc sửa đổi Điều 3 Nghị định số 01/2008/NĐ-CP ngày 03 tháng 01 năm 2008 của Chính phủ quy định chức năng, nhiệm vụ, quyền hạn và cơ cấu tổ chức của Bộ Nông nghiệp và Phát triển nông thôn;

Căn cứ Thông tư liên tịch số 61/2008/TTLT-BNN-BNV ngày 15 tháng 5 năm 2008 của Bộ trưởng Bộ Nông nghiệp và Phát triển nông thôn và Bộ Nội vụ hướng dẫn chức năng, nhiệm vụ, quyền hạn và cơ cấu tổ chức của cơ quan chuyên môn thuộc Ủy ban nhân dân cấp tỉnh, cấp huyện và nhiệm vụ quản lý nhà nước của Ủy ban nhân cấp xã về nông nghiệp và phát triển nông thôn;

Bộ Nông nghiệp và Phát triển nông thôn hướng dẫn nhiệm vụ quản lý nhà nước về chăn nuôi từ trung ương đến cấp xã, phường như sau:

Hình 8

Nội dung trên cùng phía bên trái

BỘ NÔNG NGHIỆP [0,9036]

VÀ PHÁT TRIỂN NÔNG THÔN [0,8912]

Số: 02/2011/TT-BNNPTNT [0,8627]

Nội dung trên cùng phía bên phải

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM [0,9030]

Độc lập - Tự do - Hạnh phúc [0,8699]

Hà Nội, ngày 21 tháng 01 năm 2011 [0,8550]

Nội dung tiêu đề

THÔNG TƯ [0,8977]

Hướng dẫn nhiệm vụ quản lý nhà nước về chăn nuôi [0,8935]

Nội dung chính văn bản

Căn cứ Pháp lệnh Giống vật nuôi ngày 24 tháng 3 năm 2004; [0,8927]

Căn cứ Nghị định số 08/2010/NĐ-CP ngày 05 tháng 02 năm 2010 của [0,8959]

Chính phủ về quản lý thức ăn chăn nuôi; [0,8961]

Căn cứ Nghị định số 01/2008/NĐ-CP ngày 03 tháng 01 năm 2008 của [0,8955]

Chính phủ quy định chức năng, nhiệm vụ, quyền hạn và cơ cấu tổ chức của Bộ [0,9003]

Nông nghiệp và Phát triển nông thôn; Nghị định số 75/2009/NĐ-CP ngày 10 [0,8843]

Hình 9