



(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)



1-0053541

(51)^{2022.01} G06F 16/36; G06F 16/383

(13) B

(21) 1-2022-08462

(22) 25/05/2021

(86) PCT/JP2021/019885 25/05/2021

(87) WO 2021/241602 02/12/2021

(30) 2020-093683 28/05/2020 JP

(45) 25/11/2025 452

(43) 25/05/2023 422A

(73) JFE STEEL CORPORATION (JP)

2-3, Uchisaiwai-cho 2-chome, Chiyoda-ku, Tokyo 1000011, Japan

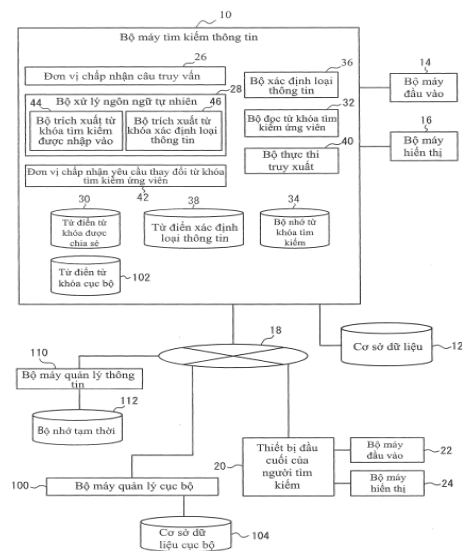
(72) TAKEMURA Yusuke (JP); KARIYA Akira (JP); KUWABARA Satoshi (JP); ETO Ayaka (JP); MOROOKA Nobuyuki (JP); SHIGEOKA Daiki (JP); TAKENAKA Shudo (JP); ITO Daigo (JP).

(74) Công ty Cổ phần Sở hữu công nghiệp INVESTIP (INVESTIP)

(54) HỆ THỐNG TÌM KIẾM THÔNG TIN

(21) 1-2022-08462

(57) Sáng chế đề cập đến hệ thống tìm kiếm thông tin mà bằng nó tìm kiếm tốc độ cao được sử dụng thường xuyên khả thi ở nhiều quận, hệ thống bao gồm: cơ sở dữ liệu (12) mà lưu trữ nhiều mảnh thông tin mà tìm kiếm được bằng văn bản; đơn vị chấp nhận câu truy vấn (26) mà chấp nhận câu truy vấn; bộ trích xuất từ khóa tìm kiếm được nhập vào (44) mà trích xuất từ khóa tìm kiếm được nhập vào từ câu truy vấn; bộ thực thi truy xuất (40) mà thực thi quá trình xử lý truy xuất từ cơ sở dữ liệu sử dụng từ khóa tìm kiếm được nhập vào; bộ máy quản lý cục bộ (100) mà lưu trữ tư liệu quận trong cơ sở dữ liệu cục bộ (104); và bộ máy quản lý thông tin (110) mà thực thi quá trình xử lý trích xuất kí tự đối với tư liệu được lưu trữ trong cơ sở dữ liệu cục bộ và chuyển đổi định dạng tệp tin của tư liệu theo kích cỡ của nó, lưu trữ tư liệu trong bộ nhớ tạm thời làm tư liệu được lưu trữ, và xuất ra tư liệu được lưu trữ đến cơ sở dữ liệu.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến hệ thống tìm kiếm thông tin.

Tình trạng kỹ thuật của sáng chế

Dữ liệu được viết trên giấy (dữ liệu giấy) đã được sử dụng làm tư liệu cho các hoạt động bảo trì trong việc sản xuất và các lĩnh vực công nghiệp khác. Điều quan trọng là số hóa dữ liệu giấy như vậy và lưu trữ nó cùng với các cơ sở dữ liệu trong các hệ thống tìm kiếm thông tin cho các hoạt động bảo trì. Các lợi thế trong công nghệ truy xuất đã cải thiện sự chính xác của các kết quả tìm kiếm, và dữ liệu giấy được số hóa và được tổng hợp thành các cơ sở dữ liệu.

Ví dụ, trong các ngành công nghiệp sản xuất, các hỏng hóc và lỗi sai của các bộ máy hoặc tương tự như vậy, các nguyên nhân và các giải pháp của chúng, và v.v. được tổng hợp thành các cơ sở dữ liệu mỗi khi chúng diễn ra, và các cơ sở dữ liệu được tham chiếu đến làm các trường hợp quá khứ trong sự kiện của các hỏng hóc hoặc lỗi sai tương tự, để ngăn chặn các hỏng hóc và rút ngắn các lần hồi phục. Các cơ sở dữ liệu cũng chứa thông tin, chẳng hạn như các tiêu chuẩn làm việc hoặc các cảm nang. Lượng thông tin được lưu trữ trong các cơ sở dữ liệu vì thế là khổng lồ.

Các ví dụ về các phương pháp thông thường của việc thu về thông tin cần thiết từ lượng dữ liệu khổng lồ như vậy bao gồm xử lý tìm kiếm văn bản mà trong đó chuỗi cho trước được sử dụng làm từ khóa và dữ liệu bất kỳ khớp với từ khóa được xuất ra làm các kết quả tìm kiếm. Ví dụ, tài liệu sáng chế 1 mô tả hệ thống tìm kiếm thông tin mà cho phép những người có kiến thức và kinh nghiệm hạn chế mà có khó khăn trong việc nhập từ khóa phù hợp để đạt được các kết quả tìm kiếm tốt.

Danh mục tài liệu trích dẫn

Tài liệu sáng chế

Tài liệu sáng chế 1: JP 2019-121392 A

Bản chất kỹ thuật của sáng chế

Vấn đề kỹ thuật

Ở đây, trong trường hợp mà các nhà máy, văn phòng, hoặc tương tự như vậy được

đặt tại nhiều quận sử dụng hệ thống tìm kiếm thông tin chung, tìm kiếm tốc độ cao có thể không khả thi nếu tư liệu được thu thập đồng nhất từ nhiều quận, do kích cỡ được tăng lên của cơ sở dữ liệu.

Điều hữu ích là cung cấp hệ thống tìm kiếm thông tin mà tìm kiếm tốc độ cao sẽ khả thi kể cả trong trường hợp mà nó được sử dụng thường xuyên ở nhiều quận.

Giải pháp cho vấn đề

Hệ thống tìm kiếm thông tin theo sáng chế bao gồm:

cơ sở dữ liệu được tạo cấu hình để lưu trữ nhiều mảnh thông tin tìm kiếm được bằng văn bản;

đơn vị chấp nhận câu truy vấn được tạo cấu hình để chấp nhận câu truy vấn theo định dạng ngôn ngữ tự nhiên;

bộ trích xuất từ khóa tìm kiếm được nhập vào được tạo cấu hình để trích xuất từ khóa tìm kiếm được nhập vào từ câu truy vấn;

bộ thực thi truy xuất được tạo cấu hình để thực thi quá trình xử lý truy xuất từ cơ sở dữ liệu bằng việc sử dụng từ khóa tìm kiếm được nhập vào, cùng với từ khóa liên quan đến từ khóa tìm kiếm được nhập vào;

bộ máy quản lý cục bộ được tạo cấu hình để lưu trữ tư liệu quận trong cơ sở dữ liệu cục bộ; và

bộ máy quản lý thông tin được tạo cấu hình để thực thi quá trình xử lý trích xuất kí tự đối với tư liệu được lưu trữ trong cơ sở dữ liệu cục bộ và chuyển đổi định dạng tệp tin của tư liệu theo kích cỡ của nó, lưu trữ tư liệu trong bộ nhớ tạm thời làm tư liệu được lưu trữ, và xuất ra tư liệu được lưu trữ đến cơ sở dữ liệu.

Hiệu quả đạt được của sáng chế

Theo sáng chế, hệ thống tìm kiếm thông tin mà tìm kiếm tốc độ cao là khả thi kể cả trong trường hợp mà nó được sử dụng thường xuyên ở nhiều quận có thể được cung cấp.

Mô tả vắn tắt các hình vẽ

Trong các hình vẽ đi kèm:

Fig.1 là sơ đồ khối minh họa cấu hình dạng giản đồ của hệ thống tìm kiếm thông tin theo phương án của sáng chế;

Fig.2 minh họa các ví dụ của các từ khóa được lưu trữ trong từ điển từ khóa được

chia sẻ;

Fig.3 minh họa các ví dụ của các từ khóa xác định loại thông tin mà được lưu trữ trong từ điển xác định loại thông tin;

Fig.4 là lưu đồ minh họa các hoạt động của bộ máy quản lý thông tin được bao gồm trong hệ thống tìm kiếm thông tin theo phương án;

Fig.5 minh họa kết cấu phân cấp của cơ sở dữ liệu;

Fig.6 minh họa các kết quả tìm kiếm mà được hiển thị;

Fig.7 là lưu đồ minh họa các hoạt động tìm kiếm thông tin của hệ thống tìm kiếm thông tin theo phương án;

Fig.8 là sơ đồ khối chức năng minh họa máy tính; và

Fig.9 minh họa cơ sở dữ liệu xử lý truy xuất ngôn ngữ.

Mô tả chi tiết sáng chế

Trong phương án sáng chế, hệ thống tìm kiếm thông tin mà tổng hợp cơ sở dữ liệu thông tin tìm kiếm được bằng văn bản từ thông tin công việc bảo trì thiết bị trong các lĩnh vực công nghiệp sản xuất sẽ được mô tả bằng cách ví dụ. Hệ thống tìm kiếm thông tin cho phép người tìm kiếm truy xuất thông tin về các trường hợp quá khứ hữu ích từ cơ sở dữ liệu khi hỏng hóc hoặc lỗi sai diễn ra, bằng việc nhập tình huống như là câu. Người tìm kiếm có thể thực hiện công việc phục hồi dựa trên thông tin được truy xuất. Ở đây, thông tin công việc bảo trì thiết bị bao gồm, ví dụ, thông tin, chẳng hạn các tiêu chuẩn hoặc cảm nang công việc mà mô tả các hỏng hóc và lỗi sai của các bộ máy hoặc tương tự như vậy, các nguyên nhân và giải pháp của chúng, và v.v. Sau đây, các trường hợp liên quan đến các hỏng hóc hoặc lỗi sai có thể được gọi là “các trường hợp hỏng hóc.”

Việc sử dụng có chủ đích của truy xuất bằng việc sử dụng hệ thống tìm kiếm thông tin theo sáng chế không bị giới hạn ở trên. Hệ thống tìm kiếm thông tin theo sáng chế có thể được sử dụng cho các cách sử dụng có chủ đích hơn là để hỗ trợ công việc bảo trì thiết bị, chẳng hạn như công việc phục hồi thiết bị hỏng. Hệ thống tìm kiếm thông tin theo sáng chế có thể được áp dụng, ví dụ, trong lĩnh vực bán hàng, để truy xuất các trường hợp quá khứ và cung cấp các phản hồi tối ưu cho các thắc mắc của khách hàng. Hệ thống tìm kiếm thông tin theo sáng chế có thể được áp dụng, ví dụ, trong lĩnh vực thiết kế sản phẩm, để tìm kiếm các tài liệu hoặc hồ sơ sản xuất tương tự, chẳng hạn như các giấy tờ kỹ thuật, và đề

xuất các thiết kế sản phẩm tối ưu, phản hồi các yêu cầu từ khách hàng.

Fig.1 là sơ đồ khối minh họa cấu hình dạng giản đồ của hệ thống xử lý thông tin theo phương án của sáng chế. Hệ thống tìm kiếm thông tin được tạo cấu hình để bao gồm bộ máy tìm kiếm thông tin 10 để thực thi quá trình xử lý tìm kiếm, cơ sở dữ liệu 12 mà lưu trữ nhiều mảnh thông tin mà tìm kiếm được bằng văn bản, bộ máy đầu vào 14, chẳng hạn như bàn phím hoặc chuột, mà được kết nối với bộ máy tìm kiếm thông tin 10, và bộ máy hiển thị 16, chẳng hạn như màn hình hiển thị, mà được kết nối với bộ máy tìm kiếm thông tin 10. Hệ thống tìm kiếm thông tin cũng được tạo cấu hình để bao gồm thiết bị đầu cuối 20 của người tìm kiếm, bộ máy quản lý cục bộ 100, và bộ máy quản lý thông tin 110 mà được kết nối với bộ máy tìm kiếm thông tin 10 qua mạng 18.

Mạng 18 là mạng truyền thông, chẳng hạn như mạng cục bộ (Local Area Network, LAN). Mạng 18, tuy nhiên, không bị giới hạn ở đó. Mạng 18 có thể là mạng truyền thông công cộng, chẳng hạn như Internet, hoặc mạng truyền thông, chẳng hạn như mạng diện rộng (Wide Area Network, WAN) hoặc mạng riêng ảo (Virtual Private Network, VPN), mà mở rộng qua mạng công cộng thường. Bộ máy tìm kiếm thông tin 10, thiết bị đầu cuối 20 của người tìm kiếm, bộ máy quản lý cục bộ 100, và mỗi bộ máy quản lý thông tin 110 được thể hiện bởi máy tính và chương trình. Các ví dụ của thiết bị đầu cuối 20 của người tìm kiếm có thể bao gồm máy tính cá nhân (Personal Computer, PC) và máy tính bảng, và bộ máy đầu vào 22 và bộ máy hiển thị 24 có thể được kết nối đến hoặc được tích hợp với thiết bị đầu cuối 20 của người tìm kiếm. Các ví dụ của bộ máy quản lý cục bộ 100 bao gồm máy tính cá nhân với cơ sở dữ liệu cục bộ 104. Bộ máy quản lý cục bộ 100 có thể bao gồm bộ máy đầu vào và bộ máy hiển thị. Các ví dụ của bộ máy quản lý thông tin 110 bao gồm máy tính cá nhân với bộ nhớ tạm thời 112. Bộ máy quản lý thông tin 110 có thể bao gồm bộ máy đầu vào và bộ máy hiển thị.

Ở đây, cấu hình của hệ thống tìm kiếm thông tin của Fig.1 là ví dụ, và vài thành phần có thể được bỏ qua. Hệ thống tìm kiếm thông tin cũng có thể bao gồm các thành phần khác. Ví dụ, hệ thống tìm kiếm thông tin có thể được tạo cấu hình mà không cần bộ máy đầu vào 14 và bộ máy hiển thị 16, và thông tin có thể được nhập vào hoặc xuất ra bằng việc sử dụng thiết bị đầu cuối 20 của người tìm kiếm. Hệ thống tìm kiếm thông tin cũng có thể bao gồm nhiều bộ máy quản lý cục bộ 100 được đặt trong nhiều quận. Trong phương án sáng chế,

hệ thống tìm kiếm thông tin bao gồm nhiều bộ máy quản lý cục bộ 100, mặc dù chỉ một bộ máy quản lý cục bộ 100 được minh họa trên Fig.1 vì các cấu hình của chúng giống nhau. Mỗi bộ máy trong số nhiều bộ máy quản lý cục bộ 100 được sắp xếp trong sự tương ứng một trong nhiều quận mà thường xuyên sử dụng hệ thống tìm kiếm thông tin.

Cơ sở dữ liệu 12 bao gồm khu vực lưu trữ tệp tin mà trong đó các tài liệu, các hình vẽ, các hình ảnh, hoặc tương tự như vậy được lưu trữ. Các câu không bị giới hạn ở văn bản được nhập vào ban đầu, và chúng có thể bao gồm văn bản được chuyển đổi từ âm thanh bằng việc sử dụng công cụ nhận diện giọng nói, hoặc văn bản được chuyển đổi từ các câu được viết trên phương tiện giấy bằng việc sử dụng nhận dạng kí tự quang học (Optical Character Recognition, OCR). Trong trường hợp mà các hình vẽ hoặc ảnh được lưu trữ, quá trình xử lý tìm kiếm có thể được thực hiện đối với các kí tự được đánh trong các hình vẽ hoặc các tên tệp tin. Thông tin được lưu trữ trong cơ sở dữ liệu 12 bao gồm thông tin, chẳng hạn như các tiêu chuẩn công việc, các cảm nang, hoặc tương tự như vậy, về các hỏng hóc và lỗi sai của các bộ máy hoặc tương tự như vậy, các nguyên nhân và giải pháp của chúng, và v.v. Trong phương án sáng chế, cơ sở dữ liệu 12 chứa nhiều trường hợp hỏng hóc. Tốt hơn là việc quản lý tệp tin được thực hiện bằng việc sử dụng kết cấu phân cấp trong cả hai bộ nhớ tạm thời 112 và khu vực lưu trữ tệp tin.

Hơn nữa, cơ sở dữ liệu 12 được tạo cấu hình để bao gồm cơ sở dữ liệu xử lý truy xuất ngôn ngữ, mà được liên kết với tư liệu được lưu trữ trong khu vực lưu trữ tệp tin và được sử dụng bằng bộ thực thi truy xuất 40 để thực hiện xử lý truy xuất ngôn ngữ. Như sẽ được mô tả sau, cơ sở dữ liệu xử lý truy xuất ngôn ngữ chứa thông tin kí tự đạt được từ quá trình xử lý được thực hiện bằng bộ máy quản lý thông tin 110 đối với tư liệu được lưu trữ trong cơ sở dữ liệu cục bộ 104, thông tin phân cấp cho tư liệu trong cơ sở dữ liệu cục bộ 104, và liên kết thông tin đến tư liệu được nén được lưu trữ trong cơ sở dữ liệu 12. Vì cơ sở dữ liệu 12 bao gồm cơ sở dữ liệu xử lý truy xuất ngôn ngữ, bộ thực thi truy xuất 40 không cần tìm kiếm trực tiếp tư liệu được lưu trữ trong cơ sở dữ liệu 12 nhưng chỉ cần tìm kiếm cơ sở dữ liệu xử lý truy xuất ngôn ngữ để truy xuất tư liệu theo cách hiệu quả.

Ở đây, như được minh họa trên Fig.9, ví dụ, cơ sở dữ liệu xử lý truy xuất ngôn ngữ là cơ sở dữ liệu quan hệ (Relational Database, RDB) mà từ đó dữ liệu dưới các phần đầu cột cho các quận mà trong đó các trường hợp quá khứ tồn tại, các phân loại, các loại chẳng

hạn như các trường hợp quá khứ (danh mục 1), các loại tài liệu (danh mục 2), các đường liên kết đến các tệp tin, thông tin kí tự được trích xuất từ các tệp tin, hoặc tương tự như vậy có thể được trích xuất. Cơ sở dữ liệu xử lý truy xuất ngôn ngữ có thể được tạo cấu hình để lưu trữ những danh mục này cùng với nhau như là hồ sơ duy nhất. Cơ sở dữ liệu xử lý truy xuất ngôn ngữ có thể được tạo cấu hình để lưu trữ những danh mục này riêng biệt trong các bảng khác nhau, và các bảng khác nhau có thể được nối liền bằng việc sử dụng thông tin định danh (chẳng hạn như số ID) làm chìa khóa.

Lượng thông tin truy xuất được từ cơ sở dữ liệu 12 tăng lên, bằng việc thường xuyên bổ sung tư liệu được lưu trữ trong các cơ sở dữ liệu cục bộ 104 được đặt trong nhiều quận. Như sẽ được mô tả chi tiết sau, tư liệu được lưu trữ trong cơ sở dữ liệu cục bộ 104 cho mỗi quận được chuyển đổi thành định dạng tệp tin phù hợp cho sự lưu trữ trong cơ sở dữ liệu 12 bằng bộ máy quản lý thông tin 110, và tư liệu được chuyển đổi được lưu trữ trong bộ nhớ tạm thời 112 làm tư liệu được lưu trữ. Bộ máy quản lý thông tin 110 thường xuyên xuất ra tư liệu được lưu trữ trong bộ nhớ tạm thời 112 cùng lúc với lượng thông tin trong cơ sở dữ liệu 12 tăng lên. Nghĩa là, tư liệu được lưu trữ trong các cơ sở dữ liệu cục bộ 104 cho các quận tương ứng được kết hợp thành cơ sở dữ liệu 12, sau khi các định dạng tệp tin được chuyển đổi bằng bộ máy quản lý thông tin 110. Tại cùng thời điểm, các hồ sơ dữ liệu cần thiết được thêm vào cơ sở dữ liệu xử lý truy xuất ngôn ngữ.

Bộ máy tìm kiếm thông tin 10 được tạo cấu hình để bao gồm đơn vị chấp nhận câu truy vấn 26, bộ xử lý ngôn ngữ tự nhiên 28, từ điển từ khóa được chia sẻ 30, từ điển từ khóa cục bộ 102, bộ đọc từ khóa tìm kiếm ứng viên 32, bộ nhớ từ khóa tìm kiếm 34, bộ xác định loại thông tin 36, từ điển xác định loại thông tin 38, bộ thực thi truy xuất 40, và đơn vị chấp nhận yêu cầu thay đổi từ khóa tìm kiếm ứng viên 42. Bộ xử lý ngôn ngữ tự nhiên 28 bao gồm bộ trích xuất từ khóa tìm kiếm được nhập vào 44 và bộ trích xuất từ khóa xác định loại thông tin 46. Ở đây, cấu hình của bộ máy tìm kiếm thông tin 10 trên Fig.1 là ví dụ, và vài thành phần có thể được bỏ qua. Bộ máy tìm kiếm thông tin 10 cũng có thể bao gồm các thành phần khác. Ví dụ, bộ máy tìm kiếm thông tin 10 có thể được tạo cấu hình mà không cần từ điển từ khóa được chia sẻ 30, từ điển từ khóa cục bộ 102, và từ điển xác định loại thông tin 38, và những từ điển này có thể được đặt ngoài bộ máy tìm kiếm thông tin 10 và được truy cập qua mạng 18.

Nói chung, các khối chức năng này có các chức năng sau. Đơn vị chấp nhận câu truy vấn 26 chấp nhận tuần tự, qua mạng 18, câu truy vấn trong định dạng ngôn ngữ tự nhiên mà đã được nhập bằng người tìm kiếm qua bộ máy đầu vào 22 của thiết bị đầu cuối 20 của người tìm kiếm. Ví dụ, các câu truy vấn, chẳng hạn như “Hồng học của bộ máy $\times\times$ đã diễn ra trong thiết bị oo. Có bất kì ví dụ tương tự nào trong quá khứ không?”, “Các hồng học với loại hồng học $\Delta\Delta$ là gì?”, “Quy trình cho việc thay thế bộ phận $\square\square$ là gì?”, hoặc “Nguyên nhân của ngắt biến tần là gì?”, có thể được nhập vào tuần tự thiết bị đầu cuối 20 của người tìm kiếm trong định dạng ngôn ngữ tự nhiên.

Bộ xử lý ngôn ngữ tự nhiên 28 thực hiện sự xử lý ngôn ngữ tự nhiên, chẳng hạn như phân tích hình thái học, đối với câu truy vấn được nhận bởi đơn vị chấp nhận câu truy vấn 26, để trích xuất các từ. Bộ trích xuất từ khóa tìm kiếm được nhập vào 44 của bộ xử lý ngôn ngữ tự nhiên 28 trích xuất, từ câu truy vấn được nhập vào, từ khóa tìm kiếm được nhập vào được sử dụng cho việc truy xuất. Với sự trích xuất này, bộ trích xuất từ khóa tìm kiếm được nhập vào 44 có thể tham chiếu, ví dụ, đến từ điển từ khóa được chia sẻ 30 và từ điển từ khóa cục bộ 102, mà sẽ được giải thích chi tiết sau, để trích xuất từ được ghi vào trong từ điển từ khóa được chia sẻ 30 hoặc từ điển từ khóa cục bộ 102 làm từ khóa tìm kiếm được nhập vào. Ví dụ, trong trường hợp mà các câu truy vấn “Hồng học của bộ máy $\times\times$ đã diễn ra trong thiết bị oo. Có ví dụ bất kỳ tương tự trong quá khứ không?” được nhập vào, bộ trích xuất từ khóa tìm kiếm được nhập vào 44 trích xuất các từ “thiết bị oo” và “bộ máy $\times\times$ ” làm các từ khóa tìm kiếm được nhập vào. Ví dụ, trong trường hợp mà trong đó câu truy vấn “Các hồng học với loại hồng học $\Delta\Delta$ là gì?” được nhập vào, bộ trích xuất từ khóa tìm kiếm được nhập vào 44 trích xuất từ “hồng học $\Delta\Delta$ ” là từ khóa tìm kiếm được nhập vào. Ví dụ, trong trường hợp mà trong đó câu truy vấn “Quy trình cho việc thay thế bộ phận $\square\square$ là gì?” được nhập vào, bộ trích xuất từ khóa tìm kiếm được nhập vào 44 trích xuất từ “bộ phận $\square\square$ ” làm từ khóa tìm kiếm được nhập vào. Ví dụ, trong trường hợp mà câu truy vấn “Nguyên nhân của ngắt biến tần là gì?” được nhập vào, bộ trích xuất từ khóa tìm kiếm được nhập vào 44 trích xuất các từ “biến tần” và “ngắt” làm các từ khóa tìm kiếm được nhập vào.

Bộ trích xuất từ khóa xác định loại thông tin 46 của bộ xử lý ngôn ngữ tự nhiên 28 trích xuất, từ câu truy vấn được nhập vào, từ khóa xác định loại thông tin mà được sử dụng để xác định loại thông tin tương ứng với mục đích của tìm kiếm. Với sự trích xuất này, bộ

trích xuất từ khóa xác định loại thông tin 46 có thể tham chiếu đến từ điển xác định loại thông tin 38, mà sẽ được mô tả chi tiết sau, để trích xuất từ được ghi vào trong từ điển xác định loại thông tin 38 làm từ khóa xác định loại thông tin. Ví dụ, trong trường hợp mà các câu truy vấn “Hồng hóc của bộ máy $\times\times$ đã diễn ra trong thiết bị oo. Có ví dụ bất kỳ tương tự trong quá khứ không?” được nhập vào, bộ trích xuất từ khóa xác định loại thông tin 46 trích xuất các từ “hồng hóc” và “ví dụ tương tự” làm các từ khóa xác định loại thông tin. Ví dụ, trong trường hợp mà trong đó câu truy vấn “Các hồng hóc với loại hồng hóc $\Delta\Delta$ là gì?” được nhập vào, bộ trích xuất từ khóa xác định loại thông tin 46 trích xuất từ “loại hồng hóc” làm từ khóa xác định loại thông tin. Ví dụ, trong trường hợp mà trong đó câu truy vấn “Quy trình cho việc thay thế bộ phận $\square\square$ là gì?” được nhập vào, bộ trích xuất từ khóa xác định loại thông tin 46 trích xuất từ “quy trình thay thế” làm từ khóa xác định loại thông tin. Ví dụ, trong trường hợp mà câu truy vấn “Nguyên nhân của ngắt biến tần là gì?” được nhập vào, bộ trích xuất từ khóa xác định loại thông tin 46 trích xuất từ “nguyên nhân” làm từ khóa xác định loại thông tin.

Trong từ điển từ khóa được chia sẻ 30, các từ khóa liên quan được ghi vào và lưu trữ trước trong sự liên kết với nhau. Trong phương án sáng chế, các từ khóa mà đồng nghĩa với nhau được ghi vào trong sự liên kết với nhau trong từ điển từ khóa được chia sẻ 30. Các từ khóa mà có các nghĩa tương tự với nhau cũng được ghi vào trong sự liên kết với nhau trong từ điển từ khóa được chia sẻ 30. Cần được lưu ý rằng, tuy nhiên, chỉ các từ khóa mà đồng nghĩa với nhau, hoặc chỉ các từ khóa mà có các nghĩa tương tự với nhau có thể được ghi vào trong từ điển từ khóa được chia sẻ 30. Theo cách khác, hoặc ngoài ra, các từ khóa mà được cân nhắc là liên quan cao với nhau có thể được ghi vào trước trong sự liên kết với nhau.

Fig.2 minh họa các ví dụ của các từ khóa được ghi vào trước trong từ điển từ khóa được chia sẻ 30. Các từ khóa thuộc về cùng hàng liên quan với nhau và được liên kết với nhau.

Bộ đọc từ khóa tìm kiếm ứng viên 32 tham chiếu đến từ điển từ khóa được chia sẻ 30 mỗi lần bộ trích xuất từ khóa tìm kiếm được nhập vào 44 trích xuất từ khóa tìm kiếm được nhập vào, để đọc ra các từ khóa mà liên quan đến từ khóa tìm kiếm được nhập vào làm các từ khóa tìm kiếm ứng viên. Ví dụ, trong trường hợp mà từ khóa tìm kiếm được

nhập vào là “thiết bị oo”, bộ đọc từ khóa tìm kiếm ứng viên 32 đọc ra, từ từ điển từ khóa được chia sẻ 30, “thiết bị AA”, “thiết bị A’A’”, và “A”A’” làm các từ đồng nghĩa, và “aaa” và “a’a’a’” làm các từ mà có các nghĩa tương tự. Nghĩa là, bộ đọc từ khóa tìm kiếm ứng viên 32 đọc ra “thiết bị AA”, “thiết bị A’A’”, “A”A’”, “aaa”, và “a’a’a’” làm các từ khóa tìm kiếm ứng viên. Tại thời điểm này, bộ đọc từ khóa tìm kiếm ứng viên 32 có thể không đọc ra tất cả, mà chỉ vài từ đồng nghĩa và các từ mà có các nghĩa tương tự từ từ điển từ khóa được chia sẻ 30 và sử dụng chúng làm các từ khóa tìm kiếm ứng viên. Trong một ví dụ, bộ đọc từ khóa tìm kiếm ứng viên 32 có thể đọc ra chỉ các từ đồng nghĩa.

Từ điển từ khóa cục bộ 102 ghi vào các từ khóa quận mà là các từ khóa được sử dụng trong các quận cụ thể. Các từ khóa quận có thể được ghi vào trong từ điển từ khóa cục bộ 102 bởi người quản lý. Đối với trường hợp với từ điển từ khóa được chia sẻ 30, các từ khóa thuộc về cùng hàng liên quan với nhau và được liên kết với nhau trong từ điển từ khóa cục bộ 102. Đối với trường hợp với từ điển từ khóa được chia sẻ 30, các từ khóa mà đồng nghĩa với nhau được ghi vào trong sự liên kết với nhau trong từ điển từ khóa cục bộ 102. Đối với trường hợp với từ điển từ khóa được chia sẻ 30, các từ khóa mà các nghĩa tương tự với nhau cũng được ghi vào trong sự liên kết với nhau trong từ điển từ khóa cục bộ 102. Vì từ điển từ khóa cục bộ 102 ghi vào các từ khóa quận và tốt hơn là được tham chiếu đến bằng bộ đọc từ khóa tìm kiếm ứng viên 32, điều không cần thiết là quản lý tất cả các thuật ngữ trong từ điển từ khóa được chia sẻ 30. Điều này ngăn chặn sự giảm tốc độ tìm kiếm do sự tăng kích cỡ của từ điển từ khóa được chia sẻ 30. Từ điển từ khóa cục bộ 102 có thể được cung cấp nhiều, và nhiều từ điển từ khóa cục bộ 102 có thể ghi vào các từ khóa quận cho các quận khác nhau. Theo cách khác, từ điển từ khóa cục bộ 102 có thể là một từ điển duy nhất được cung cấp, và từ điển từ khóa cục bộ 102 được tạo cấu hình để xác định trong quận nào mỗi từ khóa quận đã được ghi vào được sử dụng.

Trong phương án sáng chế, trong trường hợp mà trong đó từ khóa tìm kiếm được nhập vào không được ghi vào trong từ điển từ khóa cục bộ 102, bộ đọc từ khóa tìm kiếm ứng viên 32 tham chiếu đến từ điển từ khóa được chia sẻ 30, để đọc ra các từ khóa mà liên quan đến từ khóa tìm kiếm được nhập vào làm các từ khóa tìm kiếm ứng viên. Như được nhắc đến ở trên, từ khóa quận là từ khóa được sử dụng trong quận cụ thể. Từ khóa quận có thể, ví dụ, là từ mà được sử dụng chỉ trong các nhà máy hoặc các văn phòng trong quận bị

giới hạn, hoặc nó có thể là từ mà có các từ vựng được chia sẻ tương tự nhưng thường được sử dụng trong quận giới hạn. Ví dụ, trong trường hợp mà trong đó có nhiều từ điển từ khóa cục bộ 102, từ khóa quận cho quận X được ghi vào trong từ điển từ khóa cục bộ 102 cho quận X. Ví dụ, trong trường hợp mà trong đó có một từ điển từ khóa 102, từ khóa quận cho quận X được liên kết với quận X và được ghi vào trong từ điển từ khóa cục bộ 102. Trong trường hợp mà trong đó từ khóa tìm kiếm được nhập vào được ghi vào trong từ điển từ khóa cục bộ 102, bộ đọc từ khóa tìm kiếm ứng viên 32 đọc ra các từ khóa liên quan từ từ điển từ khóa cục bộ 102 làm các từ khóa tìm kiếm ứng viên. Trong trường hợp mà trong đó từ khóa tìm kiếm được nhập vào không được ghi vào trong từ điển từ khóa cục bộ 102, bộ đọc từ khóa tìm kiếm ứng viên 32 tham chiếu đến từ điển từ khóa được chia sẻ 30. Ví dụ, trong trường hợp mà trong đó từ khóa tìm kiếm được nhập vào là “thiết bị XX”, mà là từ khóa quận, bộ đọc từ khóa tìm kiếm ứng viên 32 đọc ra, từ từ điển từ khóa cục bộ 102, “thiết bị EE” và từ đồng nghĩa của nó, cùng với “thiết bị X’X” làm từ mà có nghĩa tương tự, làm các từ khóa tìm kiếm ứng viên.

Ở đây, trong từ điển từ khóa cục bộ 102, các từ khóa quận có thể được liên kết với các từ khóa được ghi vào trong từ điển từ khóa được chia sẻ 30. Trong trường hợp này, trong trường hợp mà trong đó từ khóa tìm kiếm được nhập vào là từ khóa quận, bộ đọc từ khóa tìm kiếm ứng viên 32 có thể tham chiếu đến từ điển từ khóa được chia sẻ 30 và từ điển từ khóa cục bộ 102, để đọc ra các từ khóa liên quan đến từ khóa tìm kiếm được nhập vào. Ví dụ, trong từ điển từ khóa cục bộ 102, “thiết bị XX”, mà là từ khóa quận, có thể được liên kết với “thiết bị oo” mà được ghi vào trong từ điển từ khóa được chia sẻ 30 làm từ mà có nghĩa tương tự. Ví dụ, trong trường hợp từ khóa tìm kiếm được nhập vào là “thiết bị XX”, bộ đọc từ khóa tìm kiếm ứng viên 32 tham chiếu đến từ điển từ khóa cục bộ 102, để đọc ra thiết bị oo”, và cũng tham chiếu đến từ điển từ khóa được chia sẻ 30, để đọc ra “thiết bị AA”, “thiết bị A’A’”, “A’A’”, “aaa”, và “a’a’a’”, làm các từ khóa tìm kiếm ứng viên. Bằng việc tham chiếu đến cả hai từ điển từ khóa được chia sẻ 30 và từ điển từ khóa cục bộ 102, bộ đọc từ khóa tìm kiếm ứng viên 32 có thể đặt các từ khóa tìm kiếm ứng viên mà không có các từ khóa bị bỏ sót.

Bộ nhớ từ khóa tìm kiếm 34 lưu trữ từ khóa tìm kiếm được nhập vào và các từ khóa tìm kiếm ứng viên mà được thu về như được mô tả ở trên, làm các từ khóa tìm kiếm. Khi,

ví dụ, câu truy vấn được nhập vào bộ sung, bộ nhớ từ khóa tìm kiếm 34 có thể lưu trữ theo cách tích lũy từ khóa tìm kiếm được nhập vào bộ sung, cùng với các từ khóa tìm kiếm ứng viên bộ sung, làm các từ khóa tìm kiếm.

Bộ xác định loại thông tin 36 xác định mục đích của tìm kiếm, dựa trên từ khóa xác định loại thông tin mà đã được trích xuất từ câu truy vấn. Bộ xác định loại thông tin 36 tham chiếu đến từ điển xác định loại thông tin 38 và xác định loại thông tin cần được truy xuất, dựa trên từ khóa xác định loại thông tin. Fig.3 minh họa các ví dụ của các từ khóa xác định loại thông tin được lưu trữ trong từ điển xác định loại thông tin 38. Trong từ điển xác định loại thông tin 38, các từ khóa xác định loại thông tin, mà có thể được trích xuất từ các câu truy vấn, và các loại thông tin mà tương ứng với các từ khóa xác định loại thông tin được ghi vào trong sự liên kết với nhau. Ví dụ, “trường hợp hồng học” được ghi vào làm loại thông tin của từ khóa xác định loại thông tin “hồng học” và “ví dụ tương tự.” Ví dụ, “cảm nang” được ghi vào làm loại thông tin của các từ khóa xác định loại thông tin “loại hồng học” và “quy trình thay thế.” Mặt khác, thông tin được lưu trữ trong cơ sở dữ liệu 12 được phân loại thành nhiều nhóm bằng loại thông tin cho các mục đích quản lý. Ví dụ, trong cơ sở dữ liệu 12, “trường hợp hồng học” bất kỳ có thể được lưu trữ dưới một thư mục được đặt tên “các trường hợp quá khứ”, và “cảm nang” bất kỳ có thể được lưu trữ dưới thư mục khác.

Bộ thực thi truy xuất 40 thực hiện xử lý truy xuất dựa vào văn bản về thông tin được lưu trữ trong cơ sở dữ liệu 12 bằng việc sử dụng các từ khóa tìm kiếm được lưu trữ trong bộ nhớ từ khóa tìm kiếm 34. Nghĩa là, bộ thực thi truy xuất 40 thực thi quá trình xử lý truy xuất từ cơ sở dữ liệu 12 bằng việc sử dụng từ khóa tìm kiếm được nhập vào, cùng với các từ khóa liên quan đến từ khóa tìm kiếm được nhập vào. Bộ thực thi truy xuất 40 thực hiện quá trình xử lý truy xuất đối với thông tin mà thuộc về nhóm tương ứng với ít nhất loại thông tin được xác định bởi bộ xác định loại thông tin 36 trong nhiều nhóm trong cơ sở dữ liệu 12.

Đơn vị chấp nhận yêu cầu thay đổi từ khóa tìm kiếm ứng viên 42 chấp nhận, từ người tìm kiếm, sự bổ sung và xóa đi đến và từ các từ được ghi vào trong từ điển từ khóa được chia sẻ 30 và từ điển từ khóa cục bộ 102. Ví dụ, người tìm kiếm có thể bổ sung các từ mới mà có các nghĩa tương tự, để bộ thực thi truy xuất 40 có thể thực thi quá trình xử lý truy

xuất mà sự truy xuất bị bỏ lỡ được ngăn chặn hơn nữa.

Mỗi bộ máy quản lý cục bộ 100 được đặt trong một trong nhiều quận tương ứng mà sử dụng thường xuyên hệ thống tìm kiếm thông tin, để quản lý cơ sở dữ liệu cục bộ 104 cho quận. Bộ máy quản lý cục bộ 100 có thể bao gồm bộ máy đầu vào, ví dụ, và nó có thể được vận hành bởi người quản lý của quận.

Mỗi cơ sở dữ liệu cục bộ 104 lưu trữ tư liệu được tạo ra tại các nhà máy hoặc các văn phòng trong quận. Tư liệu được tạo ra có thể được lưu trữ tự động trong cơ sở dữ liệu cục bộ 104 bằng hệ thống của quận, bao gồm bộ máy quản lý cục bộ 100, hoặc nó có thể được lưu trữ trong cơ sở dữ liệu cục bộ 104 bởi người quản lý. Các ví dụ của tư liệu được lưu trữ trong cơ sở dữ liệu cục bộ 104 bao gồm tư liệu trong các trường hợp hỏng hóc của các bộ máy hoặc tương tự như vậy, các cảm nang, và v.v. Tư liệu được lưu trữ trong cơ sở dữ liệu cục bộ 104 được chuyển đổi thành định dạng tệp tin phù hợp cho cơ sở dữ liệu 12 bằng bộ máy quản lý thông tin 110 như được mô tả ở trên, và tư liệu được chuyển đổi thường được lưu trữ trong cơ sở dữ liệu 12.

Bộ máy quản lý thông tin 110 chuyển đổi các định dạng tệp tin của tư liệu được lưu trữ trong các cơ sở dữ liệu cục bộ 104 được đặt trong nhiều quận, và lưu trữ nó trong bộ nhớ tạm thời 112 làm tư liệu được lưu trữ. Bộ máy quản lý thông tin 110 thông thường đưa ra tư liệu được lưu trữ tại cùng thời điểm với lượng thông tin trong cơ sở dữ liệu 12 tăng lên.

Fig.4 là lưu đồ minh họa các hoạt động của bộ máy quản lý thông tin 110. Bộ máy quản lý thông tin 110 thu về tư liệu mới mà đã được lưu trữ trong các cơ sở dữ liệu cục bộ 104 bất kỳ được đặt trong nhiều quận, ví dụ, thường xuyên. Trong ví dụ khác, bộ máy quản lý thông tin 110 có thể thu về tư liệu mới ngay sau khi nó đã được lưu trữ trong cơ sở dữ liệu cục bộ 104.

Trong trường hợp mà kích cỡ của tư liệu được thu về nhỏ hơn hoặc bằng ngưỡng (Có trong bước S1), bộ máy quản lý thông tin 110 tiến đến quá trình xử lý bước S2. Trong một ví dụ, ngưỡng là 1 MB.

Bộ máy quản lý thông tin 110 thực thi quá trình xử lý trích xuất kí tự để trích xuất chỉ thông tin kí tự từ tư liệu được thu về (Bước S2). Thông tin kí tự được lưu trữ trong cơ sở dữ liệu 12, để được sử dụng để tăng tốc quá trình xử lý truy xuất được dựa vào văn bản

được thực hiện bởi bộ thực thi truy xuất 40.

Bộ máy quản lý thông tin 110 chuyển đổi tư liệu có kích cỡ nhỏ hơn hoặc bằng ngưỡng thành định dạng tệp tin thứ nhất. Định dạng tệp tin thứ nhất là định dạng tệp tin mà cho phép toàn bộ tư liệu có thể được nén, để nó có thể được tải xuống trực tiếp bởi người tìm kiếm từ màn hình tìm kiếm (tham chiếu đến Fig.6). Định dạng tệp tin thứ nhất có thể là, ví dụ, định dạng tài liệu di động (Portable Document Format, PDF). Bộ máy quản lý thông tin 110 sau đó lưu trữ thông tin kí tự được trích xuất, cùng với tư liệu được chuyển đổi thành định dạng tệp tin thứ nhất, trong bộ nhớ tạm thời 112 làm tư liệu được lưu trữ (Bước S3).

Trong trường hợp mà kích cỡ của tư liệu được thu về không nhỏ hơn hoặc bằng ngưỡng (Không trong bước S1), bộ máy quản lý thông tin 110 tiến đến sự xử lý của bước S4.

Bộ máy quản lý thông tin 110 thực thi quá trình xử lý trích xuất kí tự để trích xuất chỉ thông tin kí tự từ tư liệu được thu về (Bước S4). Quá trình xử lý của bước S4 giống với của bước S2.

Bộ máy quản lý thông tin 110 chuyển đổi tư liệu có kích cỡ lớn hơn ngưỡng thành định dạng tệp tin thứ hai. Định dạng tệp tin thứ hai là định dạng tệp tin mà bao gồm thông tin đường liên kết cho toàn bộ tư liệu hoặc cho bất kì bộ phận của tư liệu ngoài thông tin kí tự. Đường liên kết là nơi tư liệu được thu về bằng bộ máy quản lý thông tin 110 đã được lưu trữ ban đầu, nghĩa là, cơ sở dữ liệu cục bộ 104 cho quận cụ thể. Người tìm kiếm có thể truy cập tư liệu trong cơ sở dữ liệu cục bộ được liên kết 104 từ màn hình tìm kiếm (tham chiếu đến Fig.6). Định dạng tệp tin thứ hai, ví dụ, định dạng ngôn ngữ đánh dấu siêu văn bản (HyperText Markup Language, HTML). Bộ máy quản lý thông tin 110 sau đó lưu trữ thông tin kí tự được trích xuất, cùng với tư liệu được chuyển đổi thành định dạng tệp tin thứ hai, trong bộ nhớ tạm thời 112 làm tư liệu được lưu trữ (Bước S5).

Bộ máy quản lý thông tin 110 xuất ra tư liệu được lưu trữ trong bộ nhớ tạm thời 112 đến cơ sở dữ liệu 12, ví dụ, thường xuyên (Bước S6).

Bây giờ, nếu lượng lớn tư liệu được tạo ra trong nhiều quận được lưu trữ trong cơ sở dữ liệu 12 hiện tại, tìm kiếm tốc độ cao có thể không khả thi do kích cỡ được tăng lên của cơ sở dữ liệu 12. Bên cạnh đó, dung lượng của cơ sở dữ liệu 12 và dung lượng của bộ

nhớ tạm thời 112 cần được tăng lên. Trong phương án sáng chế, bộ máy quản lý thông tin 110 thực hiện quá trình xử lý ở trên, và thông tin kí tự, cùng với tư liệu được nén hoặc thông tin đường liên kết, được lưu trữ trong cơ sở dữ liệu 12 và trong bộ nhớ tạm thời 112. Nghĩa là, việc tăng kích cỡ của cơ sở dữ liệu 12 có thể được ngăn chặn, và do đó, tìm kiếm tốc độ cao có thể khả thi bằng hệ thống tìm kiếm thông tin theo phương án sáng chế. Hơn nữa, vì việc tăng dung lượng của cơ sở dữ liệu 12 và việc tăng dung lượng của bộ nhớ tạm thời 112 có thể được ngăn chặn, hệ thống nhỏ gọn có thể được tạo cấu hình. Trong một ví dụ, bộ máy quản lý thông tin 110 có thể nén tư liệu mà chứa hình ảnh lớn hơn hoặc bằng 3 MB đến xấp xỉ 5 kB, bằng cách chuyển đổi nó thành HTML làm định dạng tệp tin thứ hai.

Ở đây, mỗi cơ sở dữ liệu cục bộ 104 được đặt trong nhiều quận tốt hơn là có kết cấu phân cấp chung với cơ sở dữ liệu 12. Fig.5 minh họa kết cấu phân cấp của cơ sở dữ liệu. Ví dụ, kết cấu phân cấp bao gồm, từ mức trên cùng, quận, sự phân loại, danh mục 1, danh mục 2, hoặc tương tự như vậy. Cơ sở dữ liệu cục bộ 104 cho quận X có kết cấu phân cấp của các thư mục như được minh họa trên Fig.5 trong ví dụ cụ thể. Kết cấu phân cấp này giống với cơ sở dữ liệu 12, và tốt hơn là phổ biến với quận Y và quận Z. Trong cách này, cơ sở dữ liệu cục bộ 104 có cùng kết cấu phân cấp như cơ sở dữ liệu 12. Do đó, người tìm kiếm có thể dễ dàng nhận ra vị trí lưu trữ của tư liệu mà được hiển thị với đường liên kết trong các kết quả tìm kiếm, nghĩa là, tư liệu trong đó phần thân tư liệu được lưu trữ chỉ trong cơ sở dữ liệu cục bộ 104.

Hơn nữa, vì cơ sở dữ liệu 12 và mỗi cơ sở dữ liệu cục bộ 104 có kết cấu phân cấp của các thư mục như được minh họa trên Fig.5, tư liệu, chẳng hạn như các hình vẽ và ảnh, mà không chứa thông tin kí tự có thể được truy xuất chính xác. Như được mô tả ở trên, trong trường hợp mà tư liệu bao gồm các hình vẽ và ảnh, quá trình xử lý truy xuất có thể được thực hiện không chỉ đối với các kí tự được gắn mà còn đối với các tên tệp tin. Bằng các tên tệp tin bao gồm sự phân cấp của các thư mục như được minh họa trên Fig.5, nội dung của tư liệu có thể được xác định chính xác hơn trong quá trình xử lý truy xuất.

Ví dụ, giả sử rằng hai ảnh với tên tệp tin “bộ máy BB” mà không chứa thông tin kí tự được lưu trữ dưới thư mục được đặt tên “các trường hợp quá khứ” và thư mục được đặt tên “các cảm nang” trong cơ sở dữ liệu cục bộ 104 cho quận X. Trong quá trình xử lý trích xuất kí tự (bước S2 hoặc bước S4 trên Fig.4), bộ máy quản lý thông tin 110 đặt thông tin kí

tự cho “bộ máy BB” được lưu trữ dưới thư mục được đặt tên “các trường hợp quá khứ” đến, ví dụ “quận X ¥ cán nguội ¥ các trường hợp quá khứ ¥ bộ máy BB.” Theo cách tương tự, trong quá trình xử lý trích xuất kí tự, bộ máy quản lý thông tin 110 đặt thông tin kí tự cho “bộ máy BB” được lưu trữ dưới thư mục được đặt tên “các cảm nang” là, ví dụ, “quận X ¥ cán nguội ¥ các cảm nang ¥ bộ máy BB.” Nghĩa là, trong trường hợp mà tư liệu không chứa thông tin kí tự, bộ máy quản lý thông tin 110 thực thi quá trình xử lý trích xuất kí tự để trích xuất tên tệp tin và đường dẫn thư mục của tư liệu làm thông tin kí tự. Do đó, quá trình xử lý trích xuất kí tự được thực thi bởi bộ máy quản lý thông tin 110 cho phép tư liệu mà không chứa thông tin kí tự mà không phân biệt được bằng tên tệp tin của nó để được lưu trữ trong cơ sở dữ liệu 12 theo cách mà nó có thể phân biệt được. Hơn nữa, vì thông tin kí tự cho tư liệu như vậy chứa đường dẫn thư mục, nội dung của tư liệu có thể được suy ra. Ví dụ, khi truy xuất “ví dụ thực của sự hỏng hóc của bộ máy BB”, bộ thực thi truy xuất 40 có thể trích xuất tư liệu với thông tin kí tự “quận X ¥ cán nguội ¥ các trường hợp quá khứ ¥ bộ máy BB” làm thông tin được truy xuất, trong khi loại trừ tư liệu với thông tin kí tự ‘quận X ¥ cán nguội ¥ các cảm nang ¥ bộ máy BB.” Theo cách tương tự, khi truy xuất “phương pháp vận hành bộ máy BB”, bộ thực thi truy xuất 40 có thể trích xuất tư liệu với thông tin kí tự “quận X ¥ cán nguội ¥ các cảm nang ¥ bộ máy BB” làm thông tin được truy xuất, trong khi loại trừ tư liệu với thông tin kí tự “quận X ¥ cán nguội ¥ các trường hợp quá khứ ¥ bộ máy BB.” Hơn nữa, khi truy xuất “bề ngoài của bộ máy BB trong quận Y”, bộ thực thi truy xuất 40 có thể ngăn chặn tư liệu ở trên khỏi bị bao gồm theo cách không đúng trong thông tin được truy xuất. Do đó, tư liệu, chẳng hạn như các hình vẽ hoặc các ảnh, mà không chứa thông tin kí tự có thể được truy xuất chính xác. Mặc dù trường hợp mà sự phân cấp các thư mục được bao gồm trong các tên tệp tin đã được mô tả, trong ví dụ khác, bộ máy quản lý thông tin 110 có thể trích xuất tên tệp tin và đường dẫn thư mục theo cách riêng biệt và sử dụng mỗi loại trong số chúng làm mục tiêu truy xuất. Trong ví dụ cụ thể, bộ máy quản lý thông tin 110 có thể trích xuất “quận X ¥ cán nguội ¥ các trường hợp quá khứ” và “bộ máy BB” theo cách riêng biệt, và sử dụng chúng làm thông tin kí tự.

Ở đây, các hình vẽ và các ảnh thường lớn hơn về kích cỡ so với các tài liệu. Do đó, bằng việc đặt ngưỡng (Bước S1 trên Fig.4) đến giá trị phù hợp, bộ máy quản lý thông tin 110 có thể phân biệt tài liệu, chẳng hạn như các hình vẽ và các ảnh, từ tư liệu tài liệu. Trong

trường hợp này, bộ máy quản lý thông tin 110 thực thi quá trình xử lý trích xuất kí tự của bước S4 của Fig.4 đối với tư liệu, chẳng hạn như các hình vẽ và ảnh, để chuyển đổi nó thành định dạng tệp tin thứ hai mà bao gồm thông tin đường liên kết. Nghĩa là, các ảnh mà không chứa thông tin kí tự có thể được truy xuất trong khi chúng được lưu trữ trong cơ sở dữ liệu cục bộ 104, mà không bị sao chép vào cơ sở dữ liệu 12. Bằng việc thực thi quá trình xử lý trích xuất kí tự để trích xuất tên tệp tin và đường dẫn thư mục của tư liệu làm thông tin kí tự, điều khả thi là ngăn chặn các ảnh hoặc tương tự như vậy mà được lưu trữ trong cơ sở dữ liệu cục bộ 104 và không chứa thông tin kí tự khỏi việc bị bỏ lỡ trong khi truy xuất, mà không cần tăng dung lượng của cơ sở dữ liệu 12.

Ở đây, xử lý trích xuất kí tự và xử lý nén tư liệu là xử lý tải cao cho bộ máy quản lý thông tin 110, và nó cần phải có các thông số kĩ thuật thông lượng cao cho việc tính toán để xử lý tải lớn. Trong phương án sáng chế, chức năng lên lịch được sử dụng cho bộ máy quản lý thông tin 110 để thực hiện quá trình xử lý trên cơ sở quận, và tải xử lý có thể vì thế được giảm nhẹ. Nghĩa là, tải lớn có thể được xử lý mà không cần yêu cầu bộ máy quản lý thông tin 110 phải có thông lượng tính toán cao.

Fig.6 minh họa các kết quả tìm kiếm mà được hiển thị, ví dụ, trên bộ máy hiển thị 24 sau khi xử lý truy xuất. Trong ví dụ của Fig.6, bộ thực thi truy xuất 40 thực hiện quá trình xử lý truy xuất từ cơ sở dữ liệu 12 đáp lại câu truy vấn “Nguyên nhân của ngắt biến tần là gì?”, và thu về nhiều mảnh thông tin được truy xuất. Sau đây, nhiều mảnh thông tin được truy xuất có thể được tham chiếu đến như là “thông tin được phát hiện.” Thông tin được phát hiện bao gồm sự kết hợp của thông tin trong nhiều định dạng tệp tin mà trong đó tư liệu ban đầu đã được chuyển đổi theo kích cỡ bởi bộ máy quản lý thông tin 110. Trong ví dụ của Fig.6, thông tin chính với tệp tin tải xuống được đính kèm, và thông tin phụ với đường liên kết với tư liệu được lưu trữ trong cơ sở dữ liệu cục bộ 104 được minh họa cho ví dụ. Bộ thực thi truy xuất 40 truy xuất thông tin chính và thông tin phụ mà được lưu trữ trong cơ sở dữ liệu 12 làm văn bản do sự xử lý trích xuất kí tự được thực thi bởi bộ máy quản lý thông tin 110. Bộ thực thi truy xuất 40 hiển thị các ngày đăng ký, các số tham chiếu, các quận đăng ký, và các ý chính dưới dạng văn bản. Bộ thực thi truy xuất 40 cũng hiển thị các tệp tin tải xuống được hoặc các đường liên kết đến người tìm kiếm làm các kết quả tìm kiếm. Bất kể tư liệu được chuyển đổi thành định dạng tệp tin nào bởi bộ máy quản lý thông

tin 110, người tìm kiếm có thể biết nội dung đầy đủ của tư liệu được bao gồm trong các kết quả tìm kiếm.

Hơn nữa, trong phương án sáng chế, bộ thực thi truy xuất 40 có thể lọc các kết quả tìm kiếm theo kết cấu phân cấp được nhắc đến trước đó của cơ sở dữ liệu 12, và cơ sở dữ liệu cục bộ 104, trước khi chúng được hiển thị. Điều này khiến các kết quả tìm kiếm dễ nhìn hơn.

Fig.7 là lưu đồ mà minh họa các hoạt động xử lý tìm kiếm của hệ thống tìm kiếm thông tin.

Thứ nhất, đơn vị chấp nhận câu truy vấn 26 của bộ máy tìm kiếm thông tin 10 chấp nhận câu truy vấn mà người tìm kiếm đã nhập ở thiết bị đầu cuối 20 của người tìm kiếm qua bộ máy đầu vào 22 (bước S101).

Thứ hai, bộ xử lý ngôn ngữ tự nhiên 28 thực hiện quá trình xử lý ngôn ngữ tự nhiên, chẳng hạn như phân tích hình thái học, đối với câu truy vấn được nhận bởi đơn vị chấp nhận câu truy vấn 26, để trích xuất các từ (các từ độc lập) (Bước S102).

Thứ ba, bộ trích xuất từ khóa tìm kiếm được nhập vào 44 trích xuất (lựa chọn) từ khóa tìm kiếm được nhập vào từ các từ được trích xuất bởi quá trình xử lý ngôn ngữ tự nhiên (Bước S103).

Sau đó, bộ đọc từ khóa tìm kiếm ứng viên 32 tham chiếu đến các từ điển từ khóa (từ điển từ khóa được chia sẻ 30 và từ điển từ khóa cục bộ 102), để đọc ra các từ khóa liên quan đến từ khóa tìm kiếm được nhập vào, làm các từ khóa tìm kiếm ứng viên (Bước S104). Tại thời điểm này, trong trường hợp mà từ khóa tìm kiếm được nhập vào không được ghi vào trong từ điển từ khóa cục bộ 102, bộ đọc từ khóa tìm kiếm ứng viên 32 tham chiếu đến từ điển từ khóa được chia sẻ 30.

Bộ nhớ từ khóa tìm kiếm 34 lưu trữ từ khóa tìm kiếm được nhập vào, cùng với các từ khóa tìm kiếm ứng viên được đọc ra bởi bộ đọc từ khóa tìm kiếm ứng viên 32 (Bước S105).

Bộ máy tìm kiếm thông tin 10 khiến bộ máy hiển thị 24 hiển thị các từ khóa tìm kiếm được lưu trữ trong bộ nhớ từ khóa tìm kiếm 34, và, nếu cần thiết, cho phép người tìm kiếm loại bỏ từ khóa tìm kiếm không được họ mong muốn từ nhiều từ khóa tìm kiếm được hiển thị (Bước S106).

Bộ xác định loại thông tin 36 tham chiếu đến từ điển xác định loại thông tin 38, để xác định loại thông tin (mục đích của tìm kiếm), dựa trên từ khóa xác định loại thông tin được trích xuất bởi bộ trích xuất từ khóa xác định loại thông tin 46 (Bước S107).

Sử dụng các từ khóa tìm kiếm, bộ thực thi truy xuất 40 thực hiện quá trình xử lý truy xuất dựa trên văn bản cho ví dụ về thông tin được lưu trữ trong nhóm trong cơ sở dữ liệu 12 mà tương ứng với loại thông tin được xác định bởi bộ xác định loại thông tin 36 hoặc loại thông tin được lựa chọn thủ công bởi người tìm kiếm (Bước S108).

Bộ máy tìm kiếm thông tin 10 hiển thị các kết quả tìm kiếm về bộ máy hiển thị 24 của người tìm kiếm (Bước S109).

Ở đây, bộ máy tìm kiếm thông tin 10, thiết bị đầu cuối 20 của người tìm kiếm, bộ máy quản lý cục bộ 100, và mỗi bộ máy quản lý thông tin 110 có thể là máy tính như được minh họa trên Fig.8. Trong máy tính, bộ nhớ (bộ lưu trữ), CPU (bộ xử lý), ổ đĩa cứng (Hard Disk Drive, HDD), bộ máy hiển thị 16 hoặc bộ kiểm soát hiển thị mà kiểm soát sự hiển thị của bộ máy hiển thị 24, bộ kiểm soát truyền thông cho việc kết nối mạng 18, hoặc tương tự như vậy được kết nối qua kênh. Hệ thống vận hành (Operating System, OS) và các chương trình ứng dụng cho việc triển khai sự xử lý trong phương án ở trên có thể được lưu trữ trong HDD, và khi được thực thi bởi CPU, chúng đọc ra từ HDD đến bộ nhớ. Nếu cần thiết, CPU kiểm soát bộ kiểm soát hiển thị, bộ kiểm soát truyền thông, hoặc tương tự như vậy để thực hiện các hoạt động cần thiết. Dữ liệu mà được xử lý được lưu trữ trong bộ nhớ và, nếu cần thiết, được lưu trữ trong HDD. Nhiều chức năng khác nhau được nhắc đến ở trên được nhận ra bởi sự hợp tác hệ thống giữa phần cứng, chẳng hạn như CPU hoặc bộ nhớ, OS, và các chương trình ứng dụng cần thiết.

Như đã được mô tả, theo hệ thống tìm kiếm thông tin của phương án sáng chế với cấu hình ở trên, tìm kiếm tốc độ cao khả thi kể cả trong trường hợp mà nó được sử dụng thường xuyên ở nhiều quận.

Mặc dù sáng chế đã được mô tả dựa trên các ví dụ được minh họa, nó không bị giới hạn ở phương án ở trên, và nhiều sự cải tiến và sửa đổi khác nhau có thể được tạo ra mà không rời khỏi ý chính của nó.

Khả năng ứng dụng công nghiệp

Theo sáng chế, hệ thống tìm kiếm thông tin mà tìm kiếm tốc độ cao khả thi kể cả

trong trường hợp mà nó được sử dụng thường xuyên ở nhiều quận có thể được cung cấp.

Danh sách số chỉ dẫn

- 10 Bộ máy tìm kiếm thông tin
- 12 Cơ sở dữ liệu
- 14 Bộ máy đầu vào
- 16 Bộ máy hiển thị
- 18 Mạng
- 20 Thiết bị đầu cuối của người tìm kiếm
- 22 Bộ máy đầu vào
- 24 Bộ máy hiển thị
- 26 Đơn vị chấp nhận câu truy vấn
- 28 Bộ xử lý ngôn ngữ tự nhiên
- 30 Từ điển từ khóa được chia sẻ
- 32 Bộ đọc từ khóa tìm kiếm ứng viên
- 34 Bộ nhớ từ khóa tìm kiếm
- 36 Bộ xác định loại thông tin
- 38 Từ điển xác định loại thông tin
- 40 Bộ thực thi truy xuất
- 42 Đơn vị chấp nhận yêu cầu thay đổi từ khóa tìm kiếm ứng viên
- 44 Bộ trích xuất từ khóa tìm kiếm được nhập vào
- 46 Bộ trích xuất từ khóa xác định loại thông tin
- 100 Bộ máy quản lý cục bộ
- 102 Từ điển từ khóa cục bộ
- 104 Cơ sở dữ liệu cục bộ
- 110 Bộ máy quản lý thông tin
- 112 Bộ nhớ tạm thời

YÊU CẦU BẢO HỘ

1. Hệ thống tìm kiếm thông tin, bao gồm:

cơ sở dữ liệu được tạo cấu hình để lưu trữ nhiều mảnh thông tin mà tìm kiếm được bằng văn bản;

đơn vị chấp nhận câu truy vấn được tạo cấu hình để chấp nhận câu truy vấn theo định dạng ngôn ngữ tự nhiên;

bộ trích xuất từ khóa tìm kiếm được nhập vào được tạo cấu hình để trích xuất từ khóa tìm kiếm được nhập vào từ câu truy vấn;

bộ thực thi truy xuất được tạo cấu hình để thực thi xử lý truy xuất từ cơ sở dữ liệu sử dụng từ khóa tìm kiếm được nhập vào, cùng với từ khóa liên quan đến từ khóa tìm kiếm được nhập vào;

bộ máy quản lý cục bộ được tạo cấu hình để lưu trữ tư liệu quận trong cơ sở dữ liệu cục bộ; và

bộ máy quản lý thông tin được tạo cấu hình để thực thi quá trình xử lý trích xuất kí tự đối với tư liệu được lưu trữ trong cơ sở dữ liệu cục bộ và chuyển đổi định dạng tệp tin của tư liệu theo kích cỡ của nó, lưu trữ tư liệu trong bộ nhớ tạm thời làm tư liệu được lưu trữ, và xuất ra tư liệu được lưu trữ đến cơ sở dữ liệu.

2. Hệ thống tìm kiếm thông tin theo điểm 1, trong đó, trong trường hợp mà kích cỡ của tư liệu nhỏ hơn hoặc bằng ngưỡng, bộ máy quản lý thông tin chuyển đổi tư liệu thành định dạng tệp tin thứ nhất, và trong trường hợp mà kích cỡ của tư liệu không nhỏ hơn hoặc bằng ngưỡng, bộ máy quản lý thông tin chuyển đổi tư liệu thành định dạng tệp tin thứ hai mà bao gồm thông tin đường liên kết đến cơ sở dữ liệu cục bộ.

3. Hệ thống tìm kiếm thông tin theo điểm 1 hoặc điểm 2, trong đó cơ sở dữ liệu cục bộ có kết cấu phân cấp mà giống với cơ sở dữ liệu.

4. Hệ thống tìm kiếm thông tin theo điểm bất kì trong số các điểm từ 1 đến 3, trong đó hệ thống tìm kiếm thông tin bao gồm nhiều bộ máy quản lý cục bộ mà được đặt trong nhiều

quận và được kết nối với bộ máy quản lý thông tin qua mạng.

5. Hệ thống tìm kiếm thông tin theo điểm bất kì trong số các điểm từ 1 đến 4, trong đó, trong trường hợp mà tư liệu không chứa thông tin kí tự, bộ máy quản lý thông tin thực thi quá trình xử lý trích xuất kí tự để trích xuất tên tệp tin và đường dẫn thư mục của tư liệu làm thông tin kí tự.

6. Hệ thống tìm kiếm thông tin theo điểm bất kì trong số các điểm từ 1 đến 5, trong đó cơ sở dữ liệu được tạo cấu hình để bao gồm cơ sở dữ liệu xử lý truy xuất ngôn ngữ và khu vực lưu trữ tệp tin mà tư liệu được thu về từ bộ nhớ tạm thời được lưu trữ.

7. Hệ thống tìm kiếm thông tin theo điểm 6, trong đó việc quản lý tệp tin được thực hiện sử dụng cùng kết cấu phân cấp trong cả hai bộ nhớ tạm thời và khu vực lưu trữ tệp tin được bao gồm trong cơ sở dữ liệu.

1/9

FIG. 1

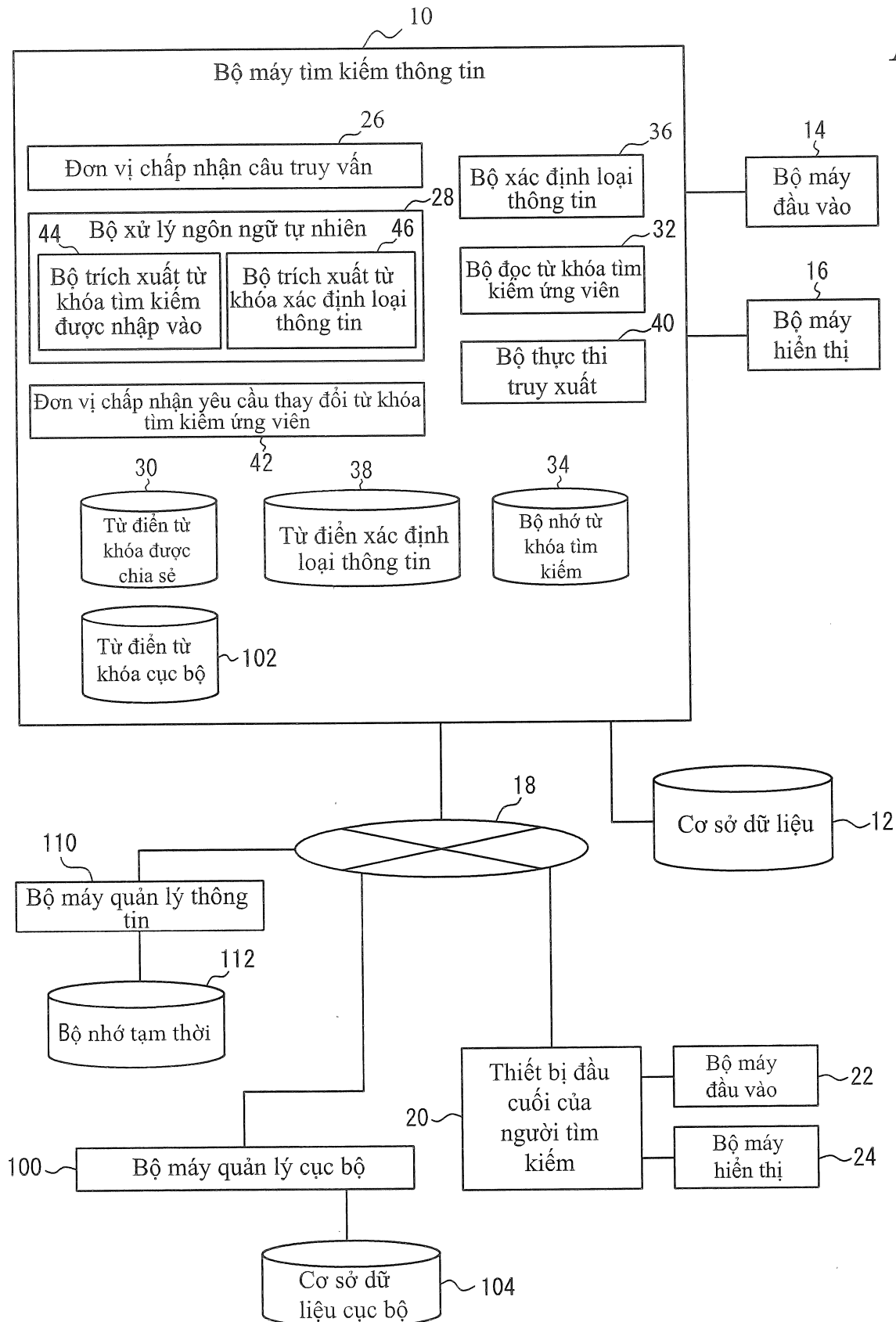


FIG.2

30

	Các từ đồng nghĩa					Các từ có các nghĩa tương tự		
	Thiết bị AA x x	Thiết bị OO	Thiết bị A' A'	A'' A''	...	aaa	a' a' a'	...
1	Bộ máy	Bộ máy BB	Máy B' B'	B'' B''	...	bbb	b' b' b'	...
2	Hồng học CC	Lỗi sai C' C'	C'' C''	Hồng học ΔΔ	...	ccc	c' c' c'	...
3	Bộ phận DD	Mẫu D' D'	Bộ phận □□	D'' D''	...	ddd	d' d' d'	...
4	...	...	...	...	...	...	...	...

3 / 9

FIG.3

38

Loại		Các từ khóa xác định loại thông tin			
A	Trường hợp hỏng hóc	Hỏng hóc	Ví dụ tương tự	...	...
B	Cẩn nang	Loại hỏng hóc	Quy trình thay thế	...	...
C	...	...	...	...	...
D	...	...	...	...	...
..	...	...	...	...	...

4 / 9

FIG.4

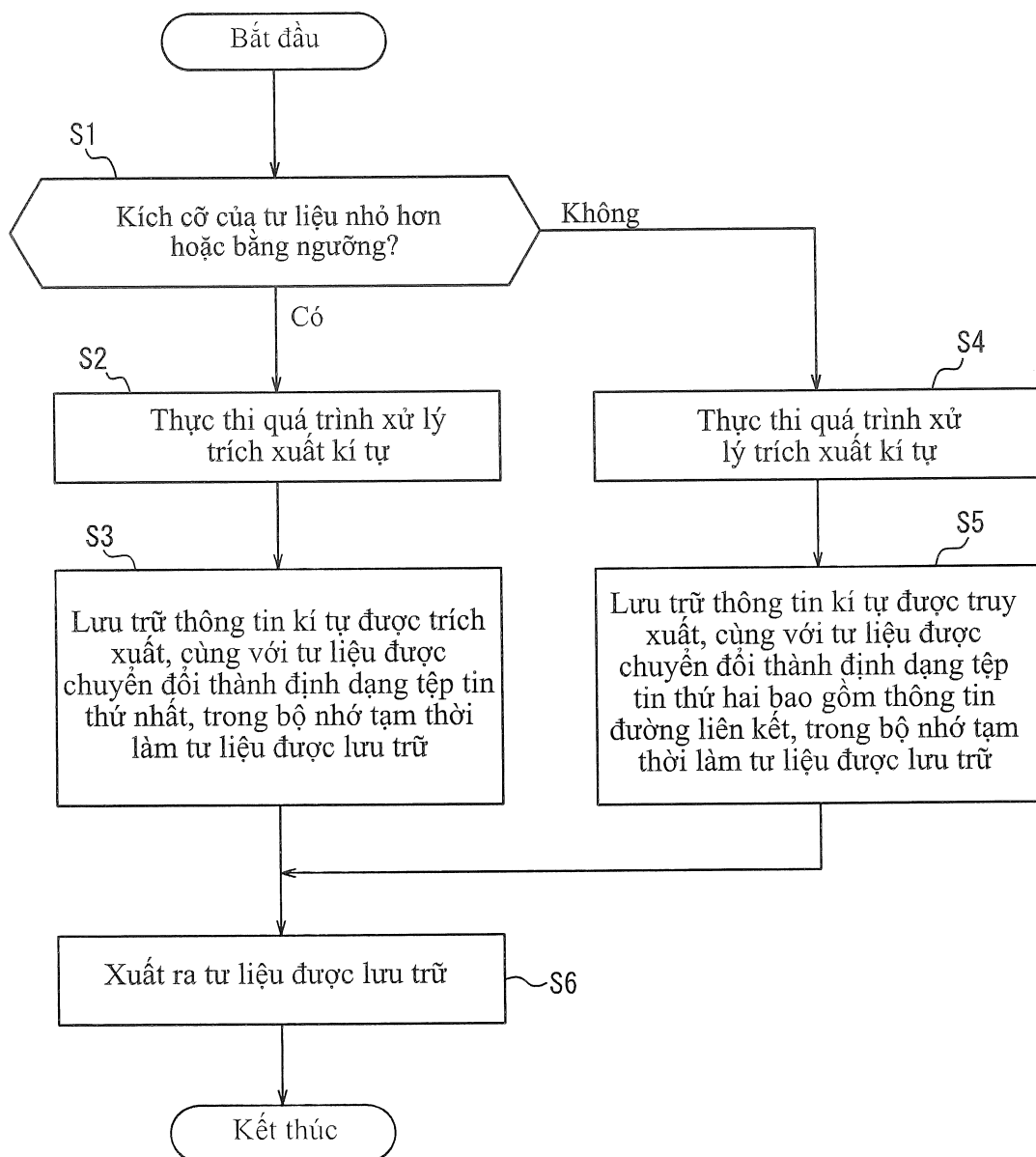


FIG.5

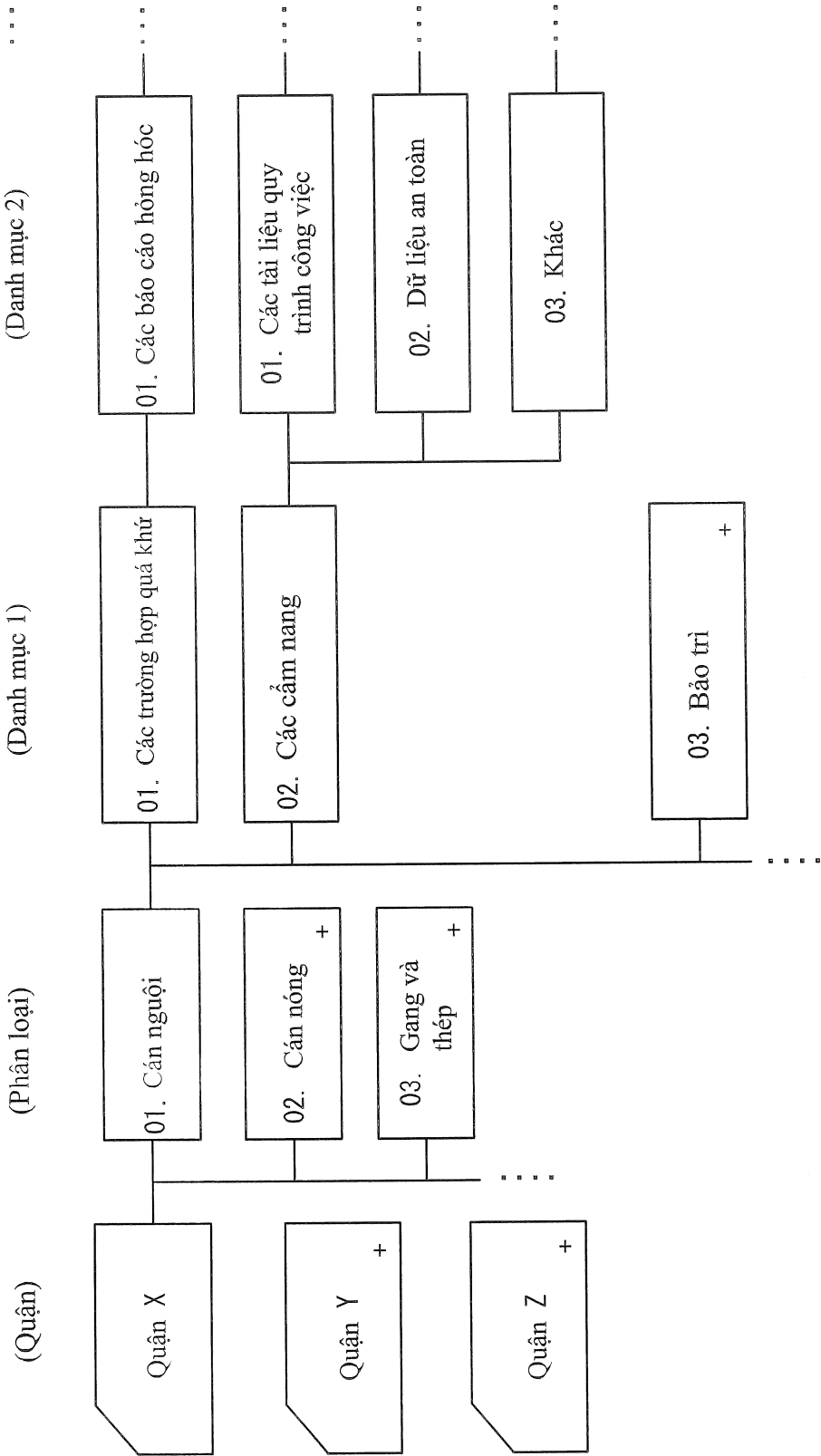


FIG.6

24

Nguyên nhân của ngất biến tần là gì?

Tìm kiếm

(Các kết quả tìm kiếm)

Các trường hợp quá khứ

Các cảm nang

Bảo trì

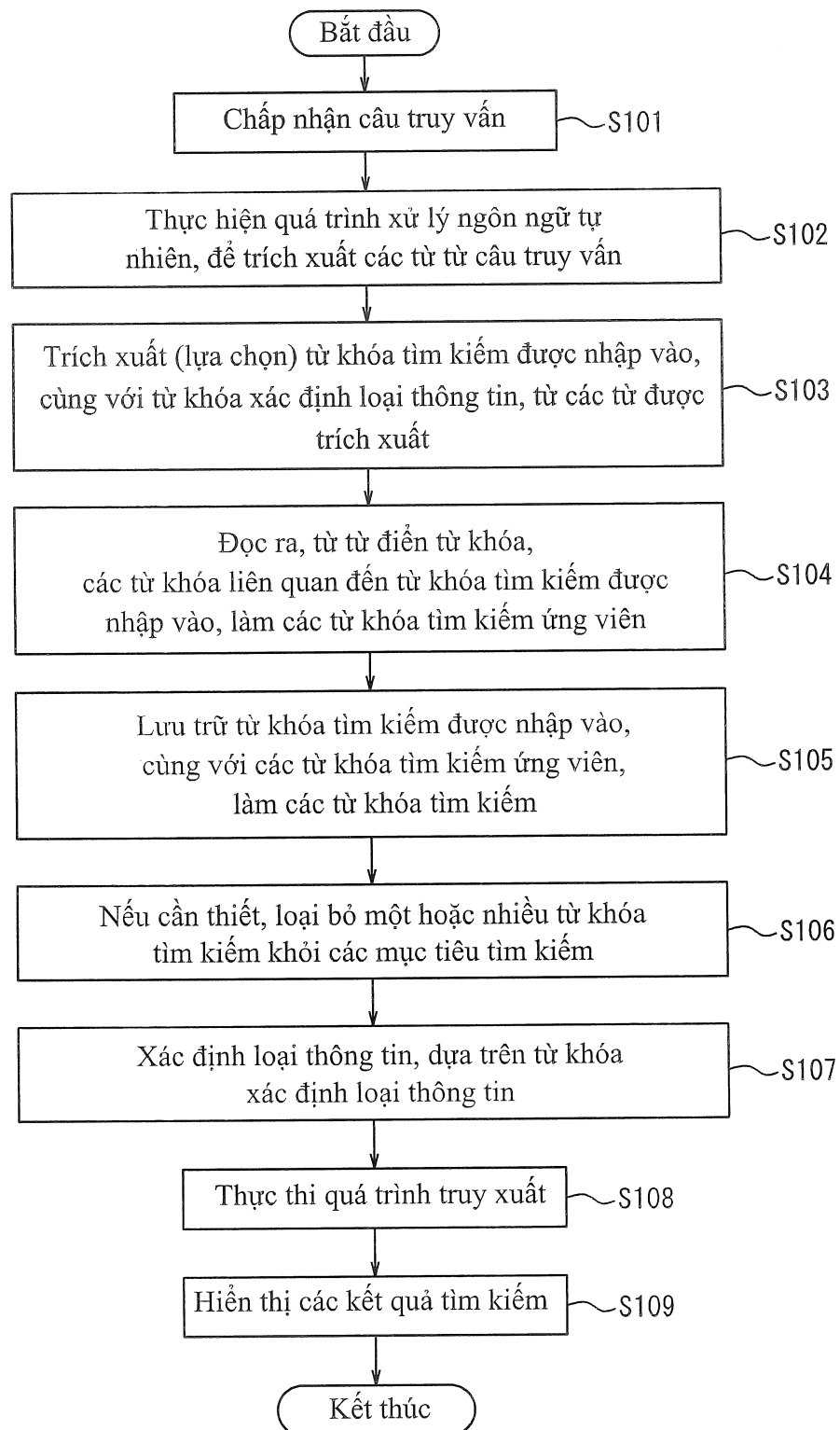
Ngày đăng ký: ○○
Số tham chiếu: △△
Quận đăng ký: X
Ý chính: biến tần
(Tên tin)

Ngày đăng ký: ◆◆
Số tham chiếu: □□
Quận đăng ký: Y
Ý chính: nguyên nhân ngất
(Đường liên kết)

Ngày đăng ký: ☆☆

FIG. A

7/9

FIG. 7

8/9

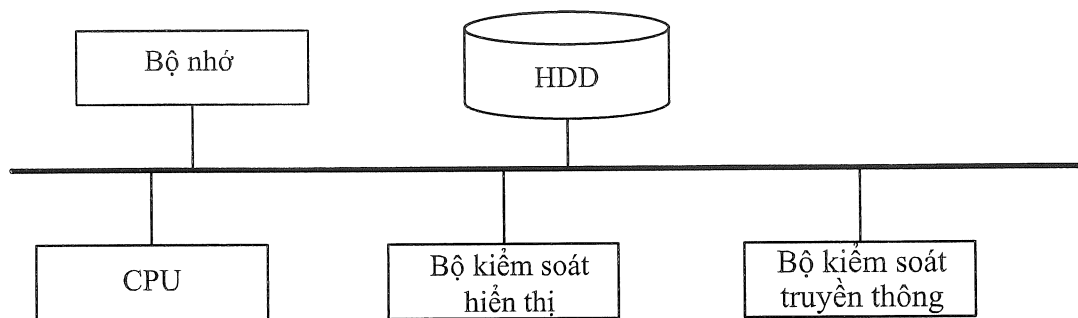
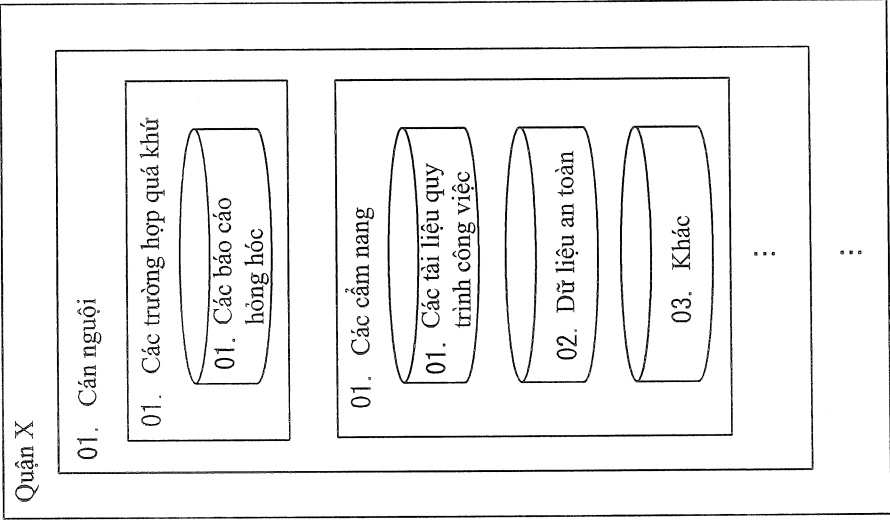
FIG.8

FIG.9



Quận	Phân loại	Danh mục 1	Danh mục 2	...	Thông tin đường liên kết tệp tin	Thông tin kí tự được trích xuất
Quận X	01. Cán ngườ	01. Các trường hợp quá khứ	01. Các báo cáo hông học	...	***	***
:	:	:	:	:	:	:

