



(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)



1-0033247

(51)⁷ G06F 17/30

(13) B

(21) 1-2019-03226

(22) 17/11/2017

(86) PCT/CN2017/111604 17/11/2017

(87) WO 2018/095281 31/05/2018

(30) 201611055619.8 (CN) 25/11/2016 CN

(45) 25/09/2022 414

(43) 26/08/2019 377A

(73) Advanced New Technologies Co., Ltd. (KY)

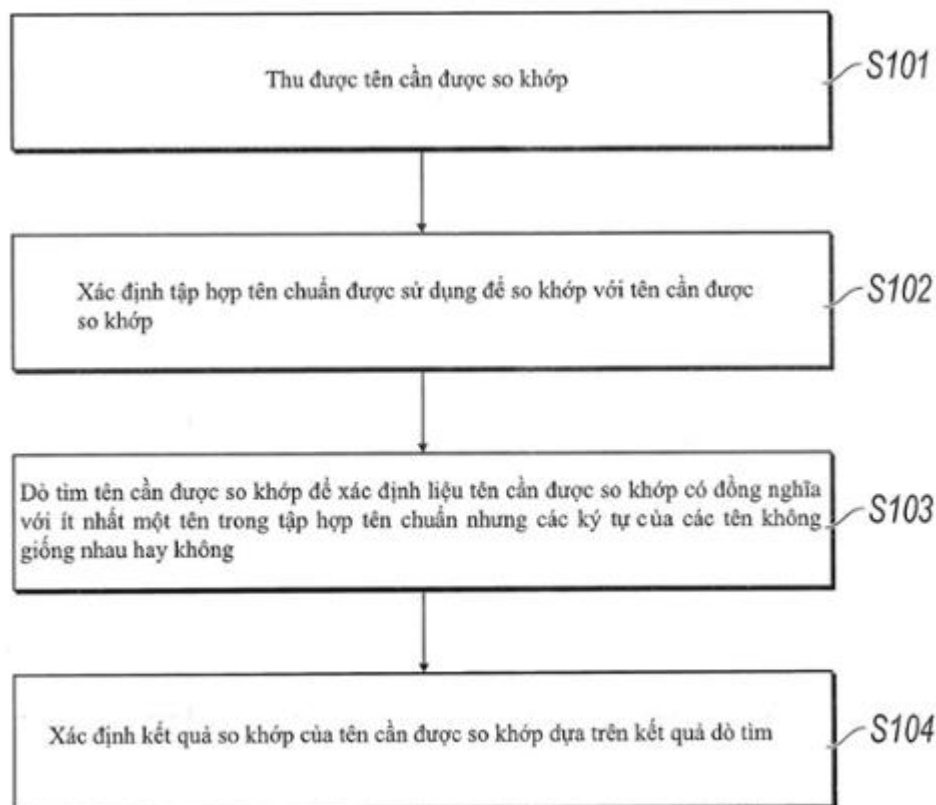
Cayman Corporate Centre, 27 Hospital Road, George Town, Grand Cayman KY1-9008, Cayman Islands

(72) SUN, Qingqing (CN).

(74) Công ty TNHH Tầm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) PHƯƠNG PHÁP VÀ THIẾT BỊ SO KHỚP TÊN

(57) Các phương án thực hiện theo sáng chế bộc lộ phương pháp và thiết bị so khớp tên. Phương pháp bao gồm các bước sau đây: thu được tên cần được so khớp; xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp; dò tìm tên cần được so khớp để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không; và xác định kết quả so khớp của tên cần được so khớp dựa trên kết quả dò tìm. Độ chính xác của việc so khớp có thể được cải thiện và tỷ lệ cảnh báo sai của hệ thống kiểm soát rủi ro có thể được giảm bớt bằng cách sử dụng các phương án thực hiện theo sáng chế.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực công nghệ phần mềm máy tính, và cụ thể, đến phương pháp và thiết bị để so khớp tên.

Tình trạng kỹ thuật của sáng chế

So khớp tên người là công nghệ rất quan trọng trong lĩnh vực kiểm soát rủi ro. Ví dụ, hệ thống kiểm soát rủi ro ghi lại tên người của những người sử dụng được xác định là trái phép trong danh sách đen. Sau đó, khi thực hiện hoạt động kiểm soát rủi ro, thì đối với mỗi người sử dụng hiện đang thực hiện một dịch vụ, tên người của mỗi người sử dụng được so khớp với mỗi tên người trong danh sách đen thông qua việc quét. Nếu việc so khớp thành công, thì người sử dụng này có thể được coi là người sử dụng trái phép, và dịch vụ của người sử dụng này sẽ bị từ chối, để ngăn ngừa các rủi ro nhất định.

So khớp tên người có thể được phân loại thành so khớp tên người chính xác và so khớp mờ tên người. Khi so sánh, so khớp mờ tên người khó khăn hơn về mặt công nghệ bởi vì khó kiểm soát độ mờ phù hợp.

Theo công nghệ đã có, thuật toán so khớp chuỗi thường được sử dụng để thực hiện so khớp mờ tên người, và mức độ trùng khớp chuỗi xác định độ mờ. Tuy nhiên, mức độ trùng khớp chuỗi được thiết lập hoàn toàn theo kinh nghiệm. Để giảm sự thiếu sót, mức độ trùng khớp chuỗi thường được thiết lập ở giá trị tương đối thấp. Do đó, độ chính xác của việc so khớp là tương đối thấp, và tỷ lệ cảnh báo sai của hệ thống kiểm soát rủi ro là tương đối cao.

Bản chất kỹ thuật của sáng chế

Các phương án thực hiện theo sáng chế đề xuất phương pháp và thiết bị để so khớp tên, để giảm vấn đề kỹ thuật sau đây: Theo công nghệ đã có, độ chính xác của việc so khớp tương đối thấp và tỷ lệ cảnh báo sai của hệ thống tương đối cao xảy ra

khi so khớp mờ tên người được thực hiện bằng cách sử dụng thuật toán so khớp chuỗi.

Để giảm bớt vấn đề kỹ thuật trên đây, các phương án thực hiện theo sáng chế được thực hiện như sau:

Phương án thực hiện của sáng chế đề xuất phương pháp so khớp tên, bao gồm các bước sau đây: thu được tên cần được so khớp; xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp; dò tìm tên cần được so khớp để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không; và xác định kết quả so khớp của tên cần được so khớp dựa trên kết quả dò tìm.

Phương án thực hiện của sáng chế đề xuất thiết bị so khớp tên, bao gồm các môđun sau đây: môđun thu thập, được tạo cấu hình để thu được tên cần được so khớp; môđun xác định, được tạo cấu hình để xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp; môđun dò tìm, được tạo cấu hình để dò tìm tên cần được so khớp để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không; và môđun so khớp, được tạo cấu hình để xác định kết quả so khớp của tên cần được so khớp dựa trên kết quả dò tìm.

Ít nhất một giải pháp kỹ thuật được sử dụng trong các phương án thực hiện theo sáng chế có thể đạt được các hiệu quả có ích sau: Tên có thể bao gồm tên người. Trong thực tế, tên người cần được so khớp có thể khác với tên người thực tế do tính hợp thời và không chắc chắn trong quá trình nhập dữ liệu và tính biến đổi của tên người. Đây cũng là lý do để thực hiện so khớp mờ. Theo giải pháp của sáng chế, đối với lý do này, tên người cần được so khớp được dò tìm để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không, và kết quả so khớp tên người được xác định dựa trên kết quả dò tìm. So sánh với công nghệ đã có mà độ mờ chỉ được kiểm soát bằng cách sử dụng mức độ trùng khớp được thiết lập theo kinh nghiệm, thì sáng chế có nhiều lợi ích hơn để cải thiện độ chính xác của việc kiểm soát độ mờ. Như vậy, độ chính xác của việc so khớp có thể được cải thiện, và tỷ lệ cảnh báo sai của hệ thống kiểm soát rủi ro có thể được giảm bớt. Do đó, một số hoặc toàn bộ vấn đề trong các công nghệ đã có có thể

được giảm bớt.

Mô tả vắn tắt các hình vẽ

Để mô tả các giải pháp kỹ thuật theo các phương án thực hiện theo sáng chế hoặc theo công nghệ đã có một cách rõ ràng hơn, phần sau đây sẽ mô tả ngắn gọn các hình vẽ kèm theo cần thiết để mô tả các phương án thực hiện hoặc công nghệ đã có. Rõ ràng, các hình vẽ kèm theo trong phần mô tả sau đây chỉ thể hiện một số phương án thực hiện của sáng chế và người có hiểu biết trung bình trong lĩnh vực kỹ thuật vẫn có thể suy ra các hình vẽ khác từ các hình vẽ kèm theo này mà không cần sự nỗ lực sáng tạo.

FIG.1 là hình vẽ lưu đồ giản lược minh họa phương pháp so khớp tên, theo phương án thực hiện của sáng chế;

FIG.2 là hình vẽ lưu đồ giản lược minh họa phương án thực hiện của sàng lọc sơ bộ trong phương pháp so khớp tên trong kịch bản ứng dụng thực tế, theo phương án thực hiện của sáng chế;

FIG.3 là hình vẽ lưu đồ giản lược minh họa phương án thực hiện của thực hiện so khớp mờ bằng cách sử dụng thuật toán được tích hợp trong phương pháp so khớp tên trong kịch bản ứng dụng thực tế, theo phương án thực hiện của sáng chế;

FIG.4 là hình vẽ lưu đồ giản lược minh họa phương án thực hiện của phương pháp so khớp tên trong kịch bản ứng dụng thực tế, theo phương án thực hiện của sáng chế; và

FIG.5 là hình vẽ sơ đồ cấu trúc minh họa thiết bị so khớp tên và tương ứng với FIG.1, theo phương án thực hiện của sáng chế.

Mô tả chi tiết sáng chế

Các phương án thực hiện theo sáng chế đề xuất phương pháp và thiết bị so khớp tên.

Để làm cho người có hiểu biết trung bình trong lĩnh vực kỹ thuật hiểu các giải pháp kỹ thuật trong sáng chế tốt hơn, phần sau đây sẽ mô tả rõ ràng và toàn diện các

giải pháp kỹ thuật theo các phương án thực hiện theo sáng chế dựa vào các hình vẽ kèm theo trong các phương án thực hiện theo sáng chế. Rõ ràng, các phương án được mô tả chỉ là một số mà không phải toàn bộ các phương án thực hiện theo sáng chế. Toàn bộ các phương án thực hiện khác thu được bởi người có hiểu biết trung bình trong lĩnh vực kỹ thuật dựa trên các phương án thực hiện theo sáng chế mà không cần sự nỗ lực sáng tạo nên nằm trong phạm vi bảo hộ của sáng chế.

Như được mô tả ở trên, trong thực tế, tên người cần được so khớp có thể khác với tên người thực tế do tính hợp thời và không chắc chắn trong quá trình nhập dữ liệu và tính biến đổi của tên người (mà chủ yếu chỉ ra sự thay đổi trong "hình thức (cụ thể là, ký tự)" của tên người này). Để dễ hiểu thì tên người bằng tiếng Anh được sử dụng làm ví dụ. Các kiểu biến thể và ví dụ phổ biến của tên người bằng tiếng Anh được thể hiện trong Bảng 1.

Bảng 1

Kiểu biến thể phổ biến của tên người bằng tiếng Anh	Ví dụ của kiểu biến thể tên người
Viết sai chính tả	Jaxson, Jakson, và Jaxon ban đầu nhằm thể hiện cùng một tên người. SMYTH và SMITH ban đầu nhằm thể hiện cùng một tên người. Viết sai chính tả.
Sự khác biệt chính tả ngữ âm	Michel, Michal, và Miguel ban đầu nhằm thể hiện cùng một tên người. Viết sai chính tả gây ra bởi các cách phát âm tương tự.
Biệt danh	Mike, Mick, và Mikey thể hiện biệt danh của cùng một người.
Thuật ngữ xưng hô	Dr., Mr., v.v. là các thuật ngữ chỉ người, và không ảnh hưởng đến nghĩa của tên người khi chúng xuất hiện cùng với tên người này.
Từ đồng nghĩa	Mohamed, Mohammed, và Mohammad là đồng nghĩa khi xuất hiện trong tên người.

Từ viết tắt	Tên người ROB là từ viết tắt của ROBBIN, và ROB và ROBBIN là đồng nghĩa. Tên người BOB là từ viết tắt của BOBBY, và ROB và BOBBY là đồng nghĩa.
Sự khác biệt của ngôn ngữ đa quốc gia	Tiếng Nga BORISOV có nghĩa khác với BORISOVNA, và tiếng Tây Ban Nha JUANLTA đồng nghĩa với JUAN.

Nếu chỉ phương pháp so khớp chuỗi được sử dụng để so khớp tên người, thì việc so khớp có thể được thử đối với kiểu biến thể là "viết sai chính tả" một chữ cái nhất định (thực ra, việc so khớp thành công một cách tình cờ, và không có căn cứ vững chắc). Tuy nhiên, độ chính xác của việc so khớp là rất thấp đối với các kiểu biến thể khác.

Trong các giải pháp của sáng chế, sự dò tìm cụ thể chẳng hạn như dò tìm từ viết tắt, dò tìm thuật ngữ xung hô, dò tìm đa ngôn ngữ, hoặc dò tìm bí danh có thể được thực hiện đối với các kiểu biến thể khác, sao cho trường hợp mà tên người là đồng nghĩa nhưng các ký tự của tên người không giống nhau (nói các khác, tên này là đồng nghĩa nhưng khác nhau về các hình thức) có thể được xem xét một cách toàn diện, và ngoài ra độ chính xác của việc so khớp có thể được cải thiện. Đáng lưu ý là "các hình thức khác nhau" có thể là các hình thức khác nhau mà "sai" do viết sai chính tả; nhưng theo các phương án thực hiện sau đây, các hình thức khác nhau này có thể chủ yếu là các hình thức khác nhau một cách "hợp lý và đúng" do các kiểu biến thể khác.

Các giải pháp của sáng chế không chỉ áp dụng được cho so khớp tên người, mà còn áp dụng được để so khớp tên khác với tên người, ví dụ, tên địa điểm hoặc tên đối tượng.

Các giải pháp của sáng chế được mô tả chi tiết sau đây.

FIG.1 là hình vẽ lưu đồ giản lược minh họa phương pháp so khớp tên, theo phương án thực hiện của sáng chế. Thiết bị để chương trình thực thi quy trình có thể được cài đặt trên bao gồm nhưng không bị giới hạn ở máy tính cá nhân, máy tính lớn hoặc trung bình, cụm máy tính, điện thoại di động, máy tính dạng bảng, thiết bị thông minh đeo được, động cơ phương tiện, v.v.. Quá trình này có thể thường được sử dụng

trong lĩnh vực kiểm soát rủi ro, và được thực thi bởi hệ thống kiểm soát rủi ro hoặc hệ thống liên quan.

Quy trình được thể hiện trên FIG.1 có thể bao gồm các bước sau đây:

S101. Thu được tên cần được so khớp.

Theo phương án thực hiện này của sáng chế, ngôn ngữ mà tên cần được so khớp thuộc về không bị giới hạn, và có thể là tiếng Anh, tiếng Nga, tiếng Tây Ban Nha, hoặc tiếng Trung Quốc. Để đơn giản cho việc mô tả, phương án thực hiện sau đây được mô tả chủ yếu bằng cách sử dụng ví dụ là ngôn ngữ mà tên cần được so khớp thuộc về là tiếng Anh.

S102. Xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp.

Theo phương án thực hiện này của sáng chế, tập hợp tên chuẩn có thể là tập hợp con được sàng lọc từ tập hợp tên lớn, hoặc có thể là chính tập hợp tên lớn. Để đơn giản cho việc mô tả, ở đây việc sàng lọc có thể được gọi là "sàng lọc sơ bộ". Ví dụ, trong kịch bản trong tình trạng kỹ thuật, tập hợp tên lớn có thể là danh sách đen được giữ bởi hệ thống kiểm soát rủi ro.

Trong tình huống trên đây, tập hợp con có thể là tập hợp mà chỉ bao gồm tên tương tự với tên cần được so khớp. Ở đây "Tương tự" có thể không quá nghiêm ngặt bởi vì có một chuỗi các hoạt động tiếp theo để xác định thêm sự tương tự. Có nghĩa là "sàng lọc mịn" được thực hiện về sau. Khoảng so khớp có thể được giảm nhanh chóng thông qua sàng lọc sơ bộ, sao cho khối lượng công việc của sàng lọc mịn tiếp theo có thể được giảm bớt, và sự thích hợp của sàng lọc mịn có thể được cải thiện, điều này có lợi cho việc cải thiện hiệu quả của các giải pháp của sáng chế.

S103. Dò tìm tên cần được so khớp để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Theo phương án thực hiện này của sáng chế, "tên là đồng nghĩa nhưng các ký tự của tên không giống nhau" mà cần được dò tìm và được xác định được gây ra chủ yếu

bởi một hoặc nhiều kiểu biến thể trong Bảng 1, và sự dò tìm này có thể bao gồm ít nhất một trong số các sự dò tìm chẳng hạn như dò tìm từ viết tắt, dò tìm thuật ngữ xung hô, dò tìm đa ngôn ngữ, hoặc dò tìm bí danh, mà được mô tả chi tiết sau đây.

Theo phương án thực hiện này của sáng chế, khi có các sự dò tìm, thì các sự dò tìm này có thể được thực hiện một cách liên tục theo thứ tự. Các sự dò tìm còn lại có thể không được thực hiện nếu kết quả so khớp của tên cần được so khớp có thể được xác định trong quá trình dò tìm. Chắc chắn, để cải thiện hiệu quả thực thi, thì các sự dò tìm có thể được thực hiện song song và sau đó các kết quả dò tìm có thể được tổng kết lại.

S104. Xác định kết quả so khớp của tên cần được so khớp dựa trên kết quả dò tìm.

Theo phương án thực hiện này của sáng chế, bằng cách thực hiện bước S103, kết quả so khớp của tên cần được so khớp có thể được xác định trực tiếp nếu xác định được rằng tên cần được so khớp đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau. Trong tình huống này, quá trình dò tìm trong bước S103 thực sự là toàn bộ quá trình so khớp của tên cần được so khớp.

Nếu xác định được rằng tên cần được so khớp không thỏa mãn điều kiện sau đây: tên cần được so khớp đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau, việc so khớp có thể còn được thực hiện đối với tên cần được so khớp bằng cách sử dụng phương pháp so khớp khác để xác định kết quả so khớp của tên cần được so khớp.

Thông qua phương pháp được thể hiện trên FIG.1, tên có thể bao gồm tên người. Trong thực tế, tên người cần được so khớp có thể khác với tên người thực tế do tính hợp thời và không chắc chắn trong quá trình nhập dữ liệu và tính biến đổi của tên người. Đây cũng là lý do để thực hiện so khớp mờ. Theo giải pháp của sáng chế, đối với lý do này, tên người cần được so khớp được dò tìm để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không, và kết quả so khớp tên người được xác định dựa trên kết quả dò tìm. So sánh với công nghệ đã có mà độ mờ chỉ được kiểm soát bằng cách sử dụng mức độ trùng khớp được thiết lập theo kinh nghiệm, thì sáng chế có nhiều lợi

ích hơn để cải thiện độ chính xác của việc kiểm soát độ mờ. Như vậy, độ chính xác của việc so khớp có thể được cải thiện, và tỷ lệ cảnh báo sai của hệ thống kiểm soát rủi ro có thể được giảm bớt. Do đó, một số hoặc toàn bộ vấn đề trong các công nghệ đã có có thể được giảm bớt.

Dựa vào phương pháp trên FIG.1, phương án thực hiện này của sáng chế còn đề xuất một số phương án thực hiện của phương pháp và giải pháp mở rộng, mà được mô tả dưới đây.

Theo phương án thực hiện này của sáng chế, tính phức tạp của tên khác nhau cần được so khớp có thể là khác nhau, và thông tin được bao gồm cũng có thể khác nhau. Đối với một số tên cần được so khớp mà có khối lượng thông tin nhỏ hoặc dấu hiệu thông tin đơn giản, sẽ rất khó để giá trị của kết quả so khớp thu được đạt kỳ vọng ngay cả khi việc so khớp được thực hiện đối với tên cần được so khớp. Ví dụ, nếu tên quá đơn giản và chung chung, chẳng hạn như tên người bằng tiếng Anh "Jim", "Jimmy", "David", "John", hoặc "Mike", thì rất khó để xác định một người nhất định ngay cả khi việc so khớp thành công.

Để giảm sự lãng phí của các tài nguyên xử lý được sử dụng để so khớp tên do tình huống này, thì tên cần được so khớp có thể được lọc trước sau khi thu được, để xác định liệu có tiếp tục việc so khớp hay không. Ngoài ra, nếu tên cần được so khớp nằm trong danh sách trắng, thì sẽ tồn tại vấn đề tương tự, và tên cần được so khớp cũng có thể được xử lý bằng cách sử dụng phương pháp này.

Đối với bước S101 và bước S102, sau khi tên cần được so khớp thu được, và trước khi tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp được xác định, các bước sau đây có thể còn được thực hiện: thu được tập hợp tên định trước mà không cần được so khớp; xác định liệu tên cần được so khớp có nằm trong tập hợp tên mà không cần được so khớp này hay không; và nếu có, tiếp tục thực hiện các bước tiếp theo. Nếu không, việc so khớp không thể được thực hiện đối với tên cần được so khớp.

Theo phương án thực hiện này của sáng chế, phương án thực hiện của bước S102 (mà tương ứng với tình huống trên đây về "sàng lọc sơ bộ") được sử dụng làm ví dụ để mô tả. Đối với bước S102, việc xác định tập hợp tên chuẩn được sử dụng để so

khớp với tên cần được so khớp có thể bao gồm sau đây: xác định tập hợp tên thứ nhất mà có thể được sử dụng để so khớp với tên cần được so khớp; và xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp bằng cách thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi từ nằm trong tên trong tập hợp tên thứ nhất.

Cũng có các phương án thực hiện về cách để thực hiện so khớp tương tự. So khớp phân đoạn từ có thể được thực hiện đối với tên cần được so khớp, hoặc so khớp toàn bộ văn bản có thể được thực hiện đối với tên cần được so khớp, v.v..

Theo ví dụ của so khớp phân đoạn từ, việc xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp bằng cách thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi từ nằm trong tên trong tập hợp tên thứ nhất có thể bao gồm các bước sau đây: thu được chỉ mục của mỗi tên nằm trong tập hợp tên thứ nhất, trong đó chỉ mục của tên là từ bất kỳ nằm trong tên này; phân đoạn tên cần được so khớp để thu được mỗi từ nằm trong tên cần được so khớp; thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi chỉ mục để thu được tập hợp con của tập hợp tên thứ nhất. Tập hợp con thu được bao gồm tên được lập chỉ mục bởi mỗi chỉ mục được so khớp thành công; và xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp dựa trên mỗi tập hợp con.

Chỉ mục trong ví dụ trên đây được thiết lập trước, và ưu điểm của việc thực hiện so khớp phân đoạn từ dựa trên chỉ mục là tốc độ của việc thu được tên cần thiết trong bộ theo quá trình so khớp có thể được tăng tốc một cách hiệu quả. Nếu so khớp phân đoạn từ không được thực hiện dựa trên chỉ mục, thì so khớp phân đoạn từ vẫn có thể được thực hiện (ví dụ, bảng dữ liệu lưu trữ bộ được truy vấn trực tiếp cho tên cần thiết bằng cách sử dụng câu lệnh Chọn để so khớp phân đoạn từ), ngoại trừ hiệu quả có thể bị ảnh hưởng.

Ngoài ra, việc thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi chỉ mục có thể bao gồm bước sau đây: thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi chỉ mục bằng cách sử dụng thuật toán so khớp chuỗi, trong đó thuật toán so khớp chuỗi có thể bao gồm một hoặc nhiều thuật toán, ví dụ, thuật toán so khớp cây tiền tố, thuật toán so khớp cây từ điển, thuật

toán so khớp tương tự chuỗi, và thuật toán so khớp tương tự cách phát âm. Ở đây, sử dụng thuật toán so khớp chuỗi chỉ là phương pháp tốt hơn, hoặc thuật toán khác, chẳng hạn như thuật toán so khớp văn bản, mà có thể được sử dụng để thực hiện so khớp tương tự có thể được sử dụng.

Có thể có các phương án thực hiện của việc xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp dựa trên mỗi tập hợp con. Ví dụ, nếu N thuật toán so khớp chuỗi được sử dụng, đối với mỗi trong số M chữ thu được thông qua phân đoạn từ, mỗi thuật toán so khớp chuỗi được sử dụng để so khớp từ với mỗi chỉ mục thu được N tập hợp con tương ứng. Sau đó bộ kết hợp của N tập hợp con được lấy ra. Toàn bộ M bộ kết hợp thu được, và mỗi tên mà tổng số lần xuất hiện trong M bộ kết hợp vượt quá ngưỡng chỉ định được chọn và được xác định là tập hợp tên chuẩn. Đối với ví dụ khác, N tập hợp con thu được, và sau đó bộ giao của N tập hợp con được lấy ra và được xác định là tập hợp tên chuẩn.

Quá trình "sàng lọc sơ bộ" được mô tả ở trên. Để dễ hiểu phương án thực hiện của sáng chế còn đề xuất lưu đồ giản lược minh họa phương án thực hiện của sàng lọc sơ bộ trong phương pháp so khớp tên trong kịch bản ứng dụng thực tế, như được thể hiện trên FIG.2.

Trong kịch bản ứng dụng thực tế, giả sử rằng tập hợp tên thứ nhất là danh sách tên người bằng tiếng Anh, chỉ mục của mỗi tên trong danh sách này được định trước (thông tin liên quan khác với tên tương ứng có thể được lập chỉ mục bằng cách sử dụng chỉ mục), chỉ mục này là từ trong tên tương ứng với chỉ mục, mỗi chỉ mục nằm trong bảng chỉ mục được thiết lập bằng cách sử dụng từ tương ứng với chỉ mục làm khóa chính, tên cần được so khớp là tên người bằng tiếng Anh "Kit Wai Jackson Wong", và khoảng trống được sử dụng làm dấu phân cách để phân đoạn tên người, và kết quả phân đoạn từ được thể hiện trong Bảng 2.

Bảng 2

Tên người cần được so khớp	Kit Wai Jackson Wong			
Sau khi phân đoạn từ	Từ 1: Kit	Từ 2: Wai	Từ 3: Jackson	Từ 4: Wong

Trên FIG.2, $M=4$, và kết quả phân đoạn từ là {Kit, Wai, Jackson, Wong}. Mỗi từ thu được thông qua phân đoạn từ được so khớp với mỗi chỉ mục bằng cách sử dụng thuật toán so khớp cây tiền tố, thuật toán so khớp cây từ điển, thuật toán so khớp tương tự chuỗi (chẳng hạn như thuật toán Simstring), và thuật toán so khớp tương tự cách phát âm (chẳng hạn như thuật toán metaphone), để kết xuất tập hợp tên tương ứng với các chỉ mục của từ đơn thu được bằng cách sử dụng bốn thuật toán so khớp: kết quả so khớp 1, 2, 3, và 4.

Bộ kết hợp của các kết quả so khớp của mỗi từ được lấy ra để thu được kết quả so khớp toàn diện của mỗi từ.

Tên nằm trong ít nhất hai kết quả so khớp toàn diện được chọn để tạo thành bộ, và bộ này được sử dụng làm kết quả sàng lọc sơ bộ được tạo ra.

Theo phương án thực hiện này của sáng chế, sau khi tập hợp tên chuẩn được xác định, tên cần được so khớp có thể được dò tìm. Tuy nhiên, một hoặc nhiều loại của tiền xử lý có thể còn được thực hiện trước khi dò tìm. Tiền xử lý có lợi cho việc cải thiện độ tin cậy của kết quả dò tìm tiếp theo. Tiền xử lý có thể bao gồm xử lý căn chỉnh, xử lý hợp nhất chữ hoa và chữ thường, xử lý truyền thống được đơn giản hóa, v.v..

Xử lý căn chỉnh được sử dụng làm ví dụ. Đối với bước S103, trước việc dò tìm tên cần được so khớp, việc xử lý sau đây có thể còn được thực hiện: căn chỉnh mỗi từ nằm trong tên trong tập hợp tên chuẩn với mỗi từ nằm trong tên cần được so khớp dựa trên tính tương tự giữa mỗi từ nằm trong tên trong tập hợp tên chuẩn và mỗi từ nằm trong tên cần được so khớp.

Trong thực tế, nguyên lý của việc xử lý căn chỉnh dựa trên tính tương tự có thể là nguyên lý tối đa hóa dựa trên độ tương tự. Để cụ thể, vị trí của tên bao gồm các từ với độ tương tự tối đa được căn chỉnh.

Ví dụ, giả sử rằng tên cần được so khớp là "Kate Lee Smith", và kết quả căn chỉnh của tên chẳng hạn như "Smith Catherine Lee" trong tập hợp tên chuẩn được thể hiện trong Bảng 3.

Bảng 3

Tên cần được so khớp	Kate	Lee	Smith
Tên thu được sau khi xử lý căn chỉnh trong tập hợp tên chuẩn	Catherine	Lee	Smith

Ngoài ra, khi xử lý căn chỉnh được thực hiện, đối với bước S103, việc dò tìm tên cần được so khớp có thể bao gồm bước sau đây: dò tìm tên cần được so khớp dựa trên tập hợp tên chuẩn được căn chỉnh để xác định liệu tên cần được so khớp đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Theo phương án thực hiện này của sáng chế, việc dò tìm trong bước S103 là phần quan trọng để cải thiện độ chính xác của việc so khớp của so khớp mờ tên người. Đối với bước S103, việc dò tìm tên cần được so khớp bao gồm bước sau đây: thực hiện ít nhất một trong số dò tìm từ viết tắt, dò tìm thuật ngữ xưng hô, dò tìm đa ngôn ngữ, hoặc dò tìm bí danh đối với tên cần được so khớp để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không. Một vài sự dò tìm được mô tả một cách riêng biệt dưới đây.

Theo phương án thực hiện này của sáng chế, sự dò tìm từ viết tắt bằng chữ cái đầu là phổ biến nhất trong dò tìm từ viết tắt. Ngoài sự dò tìm từ viết tắt bằng chữ cái đầu, thì có dò tìm từ viết tắt bằng sự thiếu sót của từ định lượng. Theo phương án thực hiện, việc thực hiện dò tìm từ viết tắt đối với tên cần được so khớp có thể bao gồm các bước sau đây: thu được dữ liệu kết hợp đối chiếu từ viết tắt định trước, trong đó mỗi sự kết hợp đối chiếu từ viết tắt phản ánh mối quan hệ ánh xạ từ viết tắt giữa ít nhất một từ và từ viết tắt của từ này; dò tìm liệu từ nằm trong tên cần được so khớp có mối quan hệ ánh xạ từ viết tắt với từ nằm trong tên trong tập hợp tên chuẩn dựa trên dữ liệu kết hợp đối chiếu từ viết tắt hay không; và xác định, dựa trên kết quả dò tìm, liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Ví dụ, nếu được định trước là tên người bằng tiếng Anh "Ben Williams" có thể

được viết tắt là "B. Williams", "Ben Williams" và "B. Williams" có thể được sử dụng làm sự kết hợp đối chiếu từ viết tắt. Nếu được dò tìm rằng tên cần được so khớp và tên trong tập hợp tên chuẩn là sự kết hợp đối chiếu từ viết tắt, thì có thể xác định được rằng có tình huống tên là đồng nghĩa nhưng các ký tự của tên không giống nhau.

Theo phương án thực hiện này của sáng chế, việc thực hiện dò tìm thuật ngữ xung hô đối với tên cần được so khớp có thể bao gồm các bước sau đây: thu được dữ liệu thuật ngữ xung hô định trước; dò tìm liệu tên cần được so khớp có bao gồm thuật ngữ xung hô hay không, trong đó được xem là thuật ngữ xung hô không ảnh hưởng đến nghĩa của tên chứa thuật ngữ xung hô; và xác định, dựa trên kết quả dò tìm, liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Trong các thuật ngữ của tên người, thuật ngữ xung hô thường là từ được gắn vào ít nhất một phần của tên người ban đầu, chẳng hạn như kính ngữ (chẳng hạn như Mr. hoặc Miss.) hoặc chức vụ (chẳng hạn như Dr. hoặc Prof.). Trong môi trường so khớp tên người, thuật ngữ xung hô không ảnh hưởng đến nghĩa của ít nhất một phần của tên người ban đầu mà tương ứng với thuật ngữ xung hô này. Do đó, nếu xác định được rằng tên cần được so khớp bao gồm thuật ngữ xung hô, và các phần khác khác với thuật ngữ xung hô có thể trùng khớp với tên trong tập hợp tên chuẩn, thì có thể xác định được rằng tên cần được so khớp đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau.

Theo phương án thực hiện này của sáng chế, việc thực hiện dò tìm bí danh đối với tên cần được so khớp có thể bao gồm các bước sau đây: thu được dữ liệu kết hợp đối chiếu bí danh định trước, trong đó mỗi dữ liệu kết hợp đối chiếu bí danh phản ánh mối quan hệ ánh xạ bí danh giữa ít nhất một từ và bí danh của từ này; dò tìm liệu từ nằm trong tên cần được so khớp có mối quan hệ ánh xạ bí danh với từ nằm trong tên trong tập hợp tên chuẩn hay không dựa trên dữ liệu kết hợp đối chiếu bí danh; và xác định, dựa trên kết quả dò tìm, liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Trong thực tế, bí danh có thể bao gồm biệt danh (ví dụ, Mick trong Bảng 1 là biệt danh của Mikey) của tên tương ứng với bí danh hoặc tên đồng nghĩa của tên

tương ứng với bí danh trong các lĩnh vực khác nhau. Đối với sau này, các lĩnh vực khác nhau có thể là các khu vực khác nhau (ví dụ, các nước khác nhau hoặc các tỉnh khác nhau), các ngôn ngữ khác nhau (ví dụ, các ngôn ngữ của các nước khác nhau hoặc các ngôn ngữ của các dân tộc khác nhau), các ngành công nghiệp khác nhau, v.v..

Do đó, ít nhất một trong số các việc sau đây được thực hiện: biệt danh của tên cần được so khớp được dò tìm, hoặc tên đồng nghĩa của tên cần được so khớp trong các lĩnh vực khác nhau được dò tìm.

Bí danh cũng không ảnh hưởng đến nghĩa của tên tương ứng với bí danh này. Do đó, nếu xác định được rằng tên cần được so khớp là bí danh của tên trong tập hợp tên chuẩn, có thể xác định được rằng tên cần được so khớp đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau.

Đáng lưu ý là các dạng lưu trữ của dữ liệu kết hợp đối chiếu và dữ liệu thuật ngữ xung hô được mô tả ở trên không bị giới hạn theo sáng chế. Phương pháp phổ biến để lưu trữ dữ liệu kết hợp đối chiếu và dữ liệu thuật ngữ xung hô trong bảng dữ liệu tương ứng và đọc dữ liệu từ cơ sở dữ liệu khi dữ liệu này cần được sử dụng.

Theo phương án thực hiện này của sáng chế, việc thực hiện dò tìm đa ngôn ngữ đối với tên cần được so khớp có thể bao gồm các bước sau đây: xác định ngôn ngữ tương ứng với tên cần được so khớp; thu được ít nhất một trong số quy tắc đồng nghĩa biến dạng chính tả hoặc quy tắc đồng âm biến dạng chính tả của ít nhất một trong số ngôn ngữ này hoặc ngôn ngữ khác; và dò tìm tên cần được so khớp theo ít nhất một trong số quy tắc đồng nghĩa biến dạng chính tả hoặc quy tắc đồng âm biến dạng chính tả để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Dò tìm đa ngôn ngữ chủ yếu nhằm vào tình huống sau đây, ví dụ, tiếng Anh "Pooh" được đánh vần là "puh" trong tiếng Đức, và hai từ này là đồng nghĩa khi xuất hiện trong tên người.

Đáng lưu ý là thuật toán được sử dụng để tính toán độ trùng khớp chuỗi có thể được sử dụng trong các sự dò tìm được thực hiện trên đây.

Theo phương án thực hiện này của sáng chế, việc xác định kết quả so khớp của

tên cần được so khớp dựa trên kết quả dò tìm có thể bao gồm các bước sau đây: xác định ít nhất một tên làm kết quả so khớp của tên cần được so khớp nếu xác định được rằng tên cần được so khớp đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau; hoặc nếu không (để cụ thể, khi việc so khớp thất bại thông qua lần dò tìm trước), xác định kết quả so khớp của tên cần được so khớp bằng cách so khớp với tên cần được so khớp với tên trong tập hợp tên chuẩn bằng cách sử dụng một hoặc nhiều thuật toán tương tự.

Theo phương án thực hiện này của sáng chế, các thuật toán tương tự có thể là dựa trên các chiều khác nhau, và do đó có thể là có lợi cho việc cải thiện độ tin cậy của kết quả so khớp. Theo ý tưởng như vậy, thuật toán tương tự có thể là thuật toán (chẳng hạn như thuật toán n-gram) được sử dụng để tính toán độ trùng khớp văn bản, thuật toán (chẳng hạn như thuật toán Phonex) được sử dụng để tính toán độ trùng khớp ngữ âm, thuật toán (chẳng hạn như thuật toán Jaro-Winkler) được sử dụng để tính toán độ trùng khớp chuỗi, v.v..

Khi các thuật toán tương tự này được sử dụng, thì các kết quả so khớp mà tương ứng với thuật toán tương tự có thể được đo một cách toàn diện để thu được kết quả so khớp toàn diện. Phương pháp đo không bị giới hạn theo sáng chế, và phương pháp phổ biến là tính tổng trọng số.

Ví dụ, khi thuật toán n-gram được sử dụng, đầu vào thuật toán là mỗi từ nằm trong tên cần được so khớp và mỗi từ tại vị trí được căn chỉnh của từ này, và đầu ra thuật toán là độ trùng khớp văn bản của mỗi cặp từ được căn chỉnh, và được ký hiệu là F1.

Khi thuật ngữ Jaro-Winkler được sử dụng, đầu vào thuật toán là mỗi từ nằm trong tên cần được so khớp và mỗi từ tại vị trí được căn chỉnh của từ này, và đầu ra thuật toán là độ trùng khớp chuỗi của mỗi cặp từ được căn chỉnh, và được ký hiệu là F2.

Khi thuật toán Phonex được sử dụng, đầu vào thuật toán là mỗi từ nằm trong tên cần được so khớp và mỗi từ tại vị trí được căn chỉnh của từ này, và đầu ra thuật toán là độ trùng khớp ngữ âm của mỗi cặp từ được căn chỉnh, và được ký hiệu là F3.

Độ trùng khớp toàn diện F của mỗi cặp từ được căn chỉnh thu được bằng cách thực hiện tính tổng trọng số đối với độ trùng khớp văn bản, độ trùng khớp chuỗi, và độ trùng khớp ngữ âm, như được thể hiện trong phương trình sau đây:

$$F = w_1 * F_1 + w_2 * F_2 + w_3 * F_3, \text{ trong đó } w_1 + w_2 + w_3 = 1.$$

Dựa trên độ trùng khớp toàn diện F của mỗi cặp từ, kết quả của việc so khớp giữa tên cần được so khớp và tên trong tập hợp tên chuẩn thu được bằng cách tính giá trị trung bình. Ví dụ, đối với cặp tên trong Bảng 3, kết quả so khớp thu được được thể hiện trong Bảng 4.

Bảng 4

Tên được dò tìm	Kate	Lee	Smith
Tên trong danh sách	Catherine	Lee	Smith
Độ trùng khớp toàn diện	0.872	1	1
Kết quả so khớp	$(0.872 + 1 + 1) / 3 = 0.957$		

Quá trình dò tìm và so khớp sau khi sàng lọc sơ bộ được mô tả ở trên. Trong quá trình này, các thuật toán liên quan đến so khớp có thể được sử dụng. Trong phương án thực hiện của các giải pháp của sáng chế, các thuật toán có thể được sử dụng có thể được tích hợp, và quá trình này là quá trình của việc thực hiện so khớp mờ bằng cách sử dụng thuật toán được tích hợp.

Theo phương án thực hiện này của sáng chế, sau khi so khớp mờ được thực hiện bằng cách sử dụng thuật toán tích hợp, một số lọc bài dựa trên các quy tắc nhất định có thể còn được thực hiện, ví dụ, độ trùng khớp trong kết quả so khớp được ánh xạ tới thông tin mô tả văn bản, hoặc độ trùng khớp được bù hoặc được giảm một cách hợp lý dựa trên kịch bản nhất định.

Dựa trên phần mô tả trên đây, trực quan hơn, phương án thực hiện của sáng chế còn đề xuất lưu đồ giản lược minh họa phương án thực hiện của thực hiện so khớp mờ bằng cách sử dụng thuật toán được tích hợp trong phương pháp so khớp tên trong kịch bản ứng dụng thực tế, như được thể hiện trên FIG.3. Trên FIG.3, trình tự của việc thực

hiện các dò tìm khác nhau và việc tính toán các độ trùng khớp theo các phương pháp khác nhau chỉ là ví dụ, và không để giới hạn sáng chế.

Trên FIG.3, nếu việc so khớp của tên cần được so khớp thành công thông qua bất kỳ một trong số các dò tìm trên đây, kết quả so khớp có thể được kết xuất trực tiếp; nếu không, kết quả so khớp của tên cần được so khớp có thể được xác định và được kết xuất bằng cách sử dụng một hoặc nhiều phương pháp để tính toán độ trùng khớp (ví dụ, tính toán độ trùng khớp văn bản, tính toán độ trùng khớp ngữ âm, và tính toán độ trùng khớp chuỗi).

Chắc chắn, các dò tìm khác nhau và tính toán độ trùng khớp theo các phương pháp khác nhau cũng có thể được thực hiện một cách hoàn toàn, và sau đó các kết quả dò tìm và các kết quả tính toán độ trùng khớp khác nhau theo các phương pháp khác nhau có thể được xem xét một cách toàn diện, để xác định kết quả so khớp của chuỗi cần được so khớp.

Ngoài ra, phương án thực hiện của sáng chế còn đề xuất lưu đồ giản lược minh họa phương án thực hiện của phương pháp so khớp tên trong kịch bản ứng dụng thực tế, như được thể hiện trên FIG.4. Trên FIG.4, việc định trước dựa trên các quy tắc nhất định có thể bao gồm quy tắc sau đây: xác định liệu tên cần được so khớp có không nằm trong tập hợp tên mà không cần được so khớp hay không, trong đó chỉ mục danh sách quét là chỉ mục của tên trong tập hợp tên chuẩn.

Các bước trên FIG.3 và FIG.4 được mô tả chi tiết ở trên, và các chi tiết được lược bỏ ở đây để đơn giản hóa.

Phương pháp so khớp tên được đề xuất theo phương án thực hiện này của sáng chế được mô tả ở trên. Như được thể hiện trên FIG.5, dựa trên ý tưởng sáng chế giống nhau, phương án thực hiện của sáng chế còn đề xuất thiết bị tương ứng.

FIG.5 là hình vẽ sơ đồ cấu trúc minh họa thiết bị so khớp tên và tương ứng với FIG.1, theo phương án thực hiện của sáng chế. Đường đứt đoạn thể hiện mô đun tùy chọn, và thiết bị này bao gồm các mô đun sau đây: mô đun thu thập 501, được tạo cấu hình để thu được tên cần được so khớp; mô đun xác định 502, được tạo cấu hình để xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp; mô đun dò

tìm 503, được tạo cấu hình để dò tên cần được so khớp để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không; và môđun so khớp 504, được tạo cấu hình để xác định kết quả so khớp của tên cần được so khớp dựa trên kết quả dò tìm.

Tùy chọn, trước khi xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp, môđun xác định 502 được tạo cấu hình để thu được tập hợp tên định trước mà không cần được so khớp, và xác định rằng tên cần được so khớp không nằm trong tập hợp tên mà không cần được so khớp.

Tùy chọn, môđun xác định 502 được tạo cấu hình để xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp bao gồm các bước sau đây:

Môđun xác định 502 được tạo cấu hình để thu được tập hợp tên thứ nhất có thể được sử dụng để so khớp với tên cần được so khớp, và xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp bằng cách thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi từ nằm trong tên trong tập hợp tên thứ nhất.

Tùy chọn, môđun xác định 502 được tạo cấu hình để xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp bằng cách thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi từ nằm trong tên trong tập hợp tên thứ nhất bao gồm các bước sau đây:

Môđun xác định 502 được tạo cấu hình để thu được chỉ mục của mỗi tên nằm trong tập hợp tên thứ nhất, trong đó chỉ mục của tên là từ bất kỳ nằm trong tên này; phân đoạn tên cần được so khớp để thu được mỗi từ nằm trong tên cần được so khớp; thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi chỉ mục để thu được tập hợp con của tập hợp tên thứ nhất. Tập hợp con thu được bao gồm tên được lập chỉ mục bởi mỗi chỉ mục được so khớp thành công; và xác định tập hợp tên chuẩn được sử dụng để so khớp với tên cần được so khớp dựa trên tập hợp con này.

Tùy chọn, môđun xác định 502 được tạo cấu hình để thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi chỉ mục bao gồm bước sau đây:

Môđun xác định 502 được tạo cấu hình để thực hiện so khớp tương tự đối với

mỗi từ nằm trong tên cần được so khớp và mỗi chỉ mục bằng cách sử dụng thuật toán so khớp chuỗi, trong đó thuật toán so khớp chuỗi bao gồm ít nhất một trong số thuật toán sau đây: thuật toán so khớp cây tiền tố, thuật toán so khớp cây từ điển, thuật toán so khớp tương tự chuỗi, hoặc thuật toán so khớp tương tự cách phát âm.

Tùy chọn, thiết bị này còn bao gồm các môđun sau đây: môđun căn chỉnh 505, được tạo cấu hình để căn chỉnh mỗi từ nằm trong tên trong tập hợp tên chuẩn với mỗi từ nằm trong tên cần được so khớp dựa trên tính tương tự giữa mỗi từ nằm trong tên trong tập hợp tên chuẩn và mỗi từ nằm trong tên cần được so khớp trước khi môđun dò tìm 503 dò tìm tên cần được so khớp.

Môđun dò tìm 503 được tạo cấu hình để dò tên cần được so khớp bao gồm bước sau đây:

Môđun dò tìm 503 được tạo cấu hình để dò tên cần được so khớp dựa trên tập hợp tên chuẩn được căn chỉnh để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Tùy chọn, môđun dò tìm 503 được tạo cấu hình để dò tên cần được so khớp bao gồm bước sau đây: môđun dò tìm 503 được tạo cấu hình để thực hiện ít nhất một trong số dò tìm từ viết tắt, dò tìm thuật ngữ xung hô, dò tìm đa ngôn ngữ, hoặc dò tìm bí danh đối với tên cần được so khớp để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Tùy chọn, môđun dò tìm 503 được tạo cấu hình để thực hiện dò tìm từ viết tắt đối với tên cần được so khớp bao gồm bước sau đây:

Môđun dò tìm được tạo cấu hình để thu được dữ liệu kết hợp đối chiếu từ viết tắt định trước, trong đó mỗi sự kết hợp đối chiếu từ viết tắt phản ánh mối quan hệ ánh xạ từ viết tắt giữa ít nhất một từ và từ viết tắt của từ này; dò tìm liệu từ nằm trong tên cần được so khớp có mối quan hệ ánh xạ từ viết tắt với từ nằm trong tên trong tập hợp tên chuẩn dựa trên dữ liệu kết hợp đối chiếu từ viết tắt hay không; và xác định, dựa trên kết quả dò tìm, liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập

hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Tùy chọn, môđun dò tìm 503 được tạo cấu hình để thực hiện dò tìm thuật ngữ xung hô đối với tên cần được so khớp bao gồm các bước sau đây:

Môđun dò tìm 503 được tạo cấu hình để thu được dữ liệu thuật ngữ xung hô định trước; dò tìm liệu tên cần được so khớp có bao gồm thuật ngữ xung hô hay không, trong đó được xem là thuật ngữ xung hô không ảnh hưởng đến nghĩa của tên chứa thuật ngữ xung hô này; và xác định, dựa trên kết quả dò tìm, liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Tùy chọn, bí danh bao gồm biệt danh của tên tương ứng với bí danh này, hoặc tên đồng nghĩa của tên tương ứng với bí danh trong các lĩnh vực khác nhau. Môđun dò tìm 503 được tạo cấu hình để thực hiện dò tìm bí danh đối với tên cần được so khớp bao gồm các bước sau đây:

Môđun dò tìm 503 được tạo cấu hình để thực hiện ít nhất một trong số sau đây: dò tìm biệt danh của tên cần được so khớp, hoặc dò tìm tên đồng nghĩa của tên cần được so khớp trong các lĩnh vực khác nhau.

Tùy chọn, môđun dò tìm 503 được tạo cấu hình để thực hiện dò tìm đa ngôn ngữ đối với tên cần được so khớp bao gồm các bước sau đây: môđun dò tìm 503 được tạo cấu hình để xác định ngôn ngữ tương ứng với tên cần được so khớp; thu được ít nhất một trong số quy tắc đồng nghĩa biến dạng chính tả hoặc quy tắc đồng âm biến dạng chính tả của ngôn ngữ này; và dò tìm tên cần được so khớp theo ít nhất một trong số quy tắc đồng nghĩa biến dạng chính tả hoặc quy tắc đồng âm biến dạng chính tả, để xác định liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Tùy chọn, môđun dò tìm 503 được tạo cấu hình để thực hiện dò tìm bí danh đối với tên cần được so khớp bao gồm các bước sau đây:

Môđun dò tìm 503 được tạo cấu hình để thu được dữ liệu kết hợp đối chiếu bí danh định trước, trong đó mỗi dữ liệu kết hợp đối chiếu bí danh phản ánh mối quan hệ ánh xạ bí danh giữa ít nhất một từ và bí danh của từ này; dò tìm liệu từ nằm trong tên

cần được so khớp có mối quan hệ ánh xạ bí danh với từ nằm trong tên trong tập hợp tên chuẩn dựa trên dữ liệu kết hợp đối chiếu bí danh hay không; và xác định, dựa trên kết quả dò tìm, liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau hay không.

Tùy chọn, môđun so khớp 504 được tạo cấu hình để xác định kết quả so khớp của tên cần được so khớp dựa trên kết quả dò tìm bao gồm các bước sau đây:

Môđun so khớp 504 được tạo cấu hình để xác định ít nhất một tên làm kết quả so khớp của tên cần được so khớp nếu môđun dò tìm 503 xác định rằng tên cần được so khớp đồng nghĩa với ít nhất một tên trong tập hợp tên chuẩn nhưng các ký tự của tên không giống nhau; hoặc nếu không, xác định kết quả so khớp của tên cần được so khớp bằng cách so khớp với tên cần được so khớp với tên trong tập hợp tên chuẩn bằng cách sử dụng một hoặc nhiều thuật toán tương tự.

Tùy chọn, thuật toán tương tự bao gồm ít nhất một trong số các thuật toán sau đây: thuật toán được sử dụng để tính toán độ trùng khớp văn bản, thuật toán được sử dụng để tính toán độ trùng khớp ngữ âm, hoặc thuật toán được sử dụng để tính toán độ trùng khớp chuỗi.

Tùy chọn, tên là tên người.

Các thiết bị được đề xuất theo các phương án thực hiện theo sáng chế là tương ứng một-một với các phương pháp. Do đó, các thiết bị và các phương pháp có các hiệu quả kỹ thuật có lợi tương tự. Các hiệu quả kỹ thuật có lợi tương tự của các phương pháp đã được mô tả chi tiết ở trên, và do đó các hiệu quả kỹ thuật có lợi tương tự của các thiết bị được lược bỏ ở đây để đơn giản hóa.

Ngoài ra, các kịch bản ứng dụng của các thiết bị và các phương pháp trên đây không bị giới hạn theo sáng chế. Ngoài lĩnh vực kiểm soát rủi ro (ví dụ, các lĩnh vực chẳng hạn như chống rửa tiền và xác thực người sử dụng) được đề cập trong tình trạng kỹ thuật, các giải pháp của sáng chế áp dụng được cho các lĩnh vực khác bất kỳ và có thể cần sử dụng công nghệ so khớp tên.

Trong những năm 1990, liệu sự cải tiến kỹ thuật là sự cải tiến phần cứng (ví dụ, sự cải tiến đối với cấu trúc mạch, chẳng hạn như diot, tranzito, hoặc chuyển mạch)

hoặc là sự cải tiến phần mềm (sự cải tiến đối với quy trình phương pháp) có thể được phân biệt một cách rõ ràng. Tuy nhiên, khi các công nghệ phát triển, các sự cải tiến hiện tại đối với nhiều quy trình phương pháp có thể được xem xét là các sự cải tiến trực tiếp đối với các cấu trúc mạch phần cứng. Nhà thiết kế thường lập trình quy trình phương pháp được cải thiện thành mạch phần cứng, để thu được cấu trúc mạch phần cứng tương ứng. Do đó, quy trình phương pháp có thể được cải thiện bằng cách sử dụng mô đun thực thể phần cứng. Ví dụ, thiết bị logic lập trình được (PLD - Programmable Logic Device) (ví dụ, mảng cổng lập trình được dạng trường (FPGA - Field Programmable gate Array)) là mạch tích hợp như vậy, và chức năng logic của PLD được xác định bởi người sử dụng thông qua thiết bị lập trình. Nhà thiết kế thực hiện lập trình để "tích hợp" hệ thống kỹ thuật số vào PLD mà không yêu cầu nhà sản xuất chip thiết kế và sản xuất chip mạch tích hợp chuyên dụng. Ngoài ra, hiện tại, thay vì tự sản xuất chip mạch tích hợp, việc lập trình như vậy chủ yếu được thực hiện bằng cách sử dụng "trình biên dịch logic" phần mềm. Trình biên dịch logic phần mềm này tương tự với trình biên dịch phần mềm được sử dụng để phát triển và viết chương trình. Mã gốc cần được viết bằng ngôn ngữ lập trình cụ thể để biên dịch. Ngôn ngữ này được gọi là ngôn ngữ mô tả phần cứng (HDL - Hardware Description Language). Có nhiều HDL, chẳng hạn như Ngôn ngữ biểu thức Boole nâng cao (ABEL - Advanced Boolean Expression Language), Ngôn ngữ mô tả phần cứng Altera (AHDL - Altera Hardware Description Language), Confluence, Ngôn ngữ lập trình đại học Cornell (CUPL - Cornell University Programming Language), HDCal, Ngôn ngữ mô tả phần cứng Java (JHDL - Java Hardware Description Language), Lava, Lola, MyHDL, PALASM, và Ngôn ngữ mô tả phần cứng Ruby (RHDL - Ruby Hardware Description Language). Ngôn ngữ mô tả phần cứng mạch tích hợp rất cao (VHDL - Very-high-speed Integrated Circuit Hardware Description Language) và Verilog được sử dụng phổ biến nhất. Người có hiểu biết trung bình trong lĩnh vực kỹ thuật cũng nên biết rằng mạch phần cứng để thực hiện quy trình phương pháp logic có thể dễ dàng thu được khi quy trình phương pháp được lập trình logic bằng cách sử dụng một vài ngôn ngữ mô tả phần cứng được mô tả và được lập trình vào mạch tích hợp.

Bộ điều khiển có thể được thực hiện bằng cách sử dụng phương pháp phù hợp bất kỳ. Ví dụ, bộ điều khiển có thể là bộ vi xử lý hoặc bộ xử lý và vật ghi đọc được bằng máy tính lưu trữ mã chương trình đọc được bằng máy tính (chẳng hạn như phần

mềm hoặc phần sụn) mà có thể được thực thi bởi bộ vi xử lý hoặc bộ xử lý, cổng logic, chuyển mạch, mạch tích hợp chuyên dụng (ASIC - mạch tích hợp chuyên dụng), bộ điều khiển logic lập trình được, hoặc bộ vi điều khiển được cài đặt sẵn. Các ví dụ của bộ điều khiển bao gồm nhưng không bị giới hạn ở các bộ vi xử lý sau đây: ARC 625D, Atmel AT91SAM, Microchip PIC18F26K20, và Silicone Labs C8051F320. Bộ điều khiển bộ nhớ cũng có thể được thực hiện như một phần của logic điều khiển của bộ nhớ này. Người có hiểu biết trung bình trong lĩnh vực kỹ thuật cũng nên biết rằng, ngoài việc thực hiện bộ điều khiển bằng cách sử dụng mã chương trình đọc được bằng máy tính, lập trình logic có thể được thực hiện đối với các bước của phương pháp để cho phép bộ điều khiển thực hiện các chức năng tương tự dưới các dạng của cổng logic, chuyển mạch, mạch tích hợp chuyên dụng, bộ điều khiển logic lập trình được, và bộ vi điều khiển được cài đặt sẵn. Do đó, bộ điều khiển có thể được coi là thành phần phần cứng, và thiết bị được tạo cấu hình để thực hiện các chức năng khác nhau trong bộ điều khiển cũng có thể được coi là cấu trúc trong thành phần phần cứng. Hoặc thiết bị được tạo cấu hình để thực hiện các chức năng khác nhau thậm chí có thể được coi là cả môđun phần mềm thực hiện phương pháp và cấu trúc trong thành phần phần cứng.

Hệ thống, thiết bị, môđun, hoặc khối được minh họa theo các phương án thực hiện trên đây có thể được thực hiện bằng cách sử dụng chip máy tính hoặc thực thể, hoặc có thể được thực hiện bằng cách sử dụng sản phẩm có chức năng nhất định. Thiết bị thực hiện điển hình là máy tính. Máy tính có thể là, ví dụ, máy tính cá nhân, máy tính xách tay, điện thoại di động, điện thoại camera, điện thoại thông minh, trợ lý kỹ thuật số cá nhân, trình phát đa phương tiện, thiết bị định vị, thiết bị thư điện tử, bảng điều khiển trò chơi, máy tính dạng bảng, thiết bị đeo được hoặc sự kết hợp bất kỳ của các thiết bị này.

Để đơn giản cho việc mô tả, thiết bị ở trên được mô tả bằng cách chia các chức năng thành các khối khác nhau. Chắc chắn, khi sáng chế được thực hiện, các chức năng của các khối có thể được thực hiện trong một hoặc nhiều mảnh của phần mềm và/hoặc phần cứng.

Người có hiểu biết trung bình trong lĩnh vực kỹ thuật nên hiểu rằng phương án thực hiện của sáng chế có thể được đề xuất làm phương pháp, hệ thống, hoặc sản phẩm

chương trình máy tính. Do đó, sáng chế có thể sử dụng dạng của các phương án thực hiện chỉ của phần cứng, các phương án thực hiện chỉ của phần mềm, hoặc các phương án thực hiện với sự kết hợp của phần mềm và phần cứng. Ngoài ra, sáng chế có thể sử dụng dạng của sản phẩm chương trình máy tính mà được thực hiện trên một hoặc nhiều vật ghi lưu trữ sử dụng được bằng máy tính (bao gồm nhưng không bị giới hạn ở bộ nhớ đĩa, CD-ROM, bộ nhớ quang, v.v.) chứa mã chương trình sử dụng được bằng máy tính.

Sáng chế được mô tả dựa vào các lưu đồ và/hoặc các sơ đồ khối của phương pháp, thiết bị (hệ thống), và sản phẩm chương trình máy tính dựa trên các phương án thực hiện theo sáng chế. Đáng lưu ý là các lệnh chương trình máy tính có thể được sử dụng để thực hiện mỗi quá trình và/hoặc mỗi khối trong các lưu đồ và/hoặc các sơ đồ khối và sự kết hợp của quá trình và/hoặc khối trong các lưu đồ và/hoặc các sơ đồ khối này. Các lệnh chương trình máy tính này có thể được tạo ra cho máy tính đa năng, máy tính chuyên dụng, bộ xử lý được cài đặt sẵn, hoặc bộ xử lý của thiết bị xử lý dữ liệu lập trình được khác để tạo ra máy, sao cho các lệnh này được thực thi bởi máy tính hoặc bộ xử lý của thiết bị xử lý dữ liệu lập trình được khác để tạo ra thiết bị để thực hiện các chức năng được chỉ định trong một hoặc nhiều quá trình trong các lưu đồ và/hoặc trong một hoặc nhiều khối trong các sơ đồ khối.

Các lệnh chương trình máy tính này có thể được lưu trữ trong bộ nhớ đọc được bằng máy tính mà có thể lệnh cho máy tính hoặc thiết bị xử lý dữ liệu lập trình được khác làm việc theo cách cụ thể, sao cho các lệnh được lưu trữ trong bộ nhớ đọc được bằng máy tính tạo ra thành phần lạ bao gồm thiết bị lệnh. Thiết bị lệnh thực hiện lệnh được chỉ định trong một hoặc nhiều quá trình trong các lưu đồ và/hoặc trong một hoặc nhiều các khối trong các sơ đồ khối.

Các lệnh chương trình máy tính này có thể được tải vào máy tính hoặc thiết bị xử lý dữ liệu lập trình được khác, sao cho chuỗi các hoạt động và các hoạt động và các bước được thực hiện trên máy tính hoặc thiết bị lập trình được khác, nhờ đó tạo ra quá trình xử lý được thực hiện bằng máy tính. Do đó, các lệnh được thực thi trên máy tính hoặc thiết bị lập trình được khác tạo ra các bước để thực hiện chức năng được chỉ định trong một hoặc nhiều quá trình trong các lưu đồ và/hoặc trong một hoặc nhiều khối trong các sơ đồ khối.

Theo cấu hình điển hình, thiết bị tính toán bao gồm một hoặc nhiều bộ xử lý (CPU), một hoặc nhiều giao diện đầu vào/đầu ra, một hoặc nhiều giao diện mạng, và một hoặc nhiều bộ nhớ.

Bộ nhớ có thể bao gồm bộ nhớ không liên tục, bộ nhớ truy cập ngẫu nhiên (RAM - Random Access memory), bộ nhớ lâu dài, và/hoặc dạng khác trong vật ghi đọc được bằng máy tính, ví dụ, bộ nhớ chỉ đọc (ROM - Read-Only Memory) hoặc bộ nhớ flash (flash RAM). Bộ nhớ là ví dụ của vật ghi đọc được bằng máy tính.

Vật ghi đọc được bằng máy tính bao gồm các vật ghi liên tục, không liên tục, có thể di chuyển và không thể di chuyển có thể lưu trữ thông tin bằng cách sử dụng phương pháp hoặc công nghệ bất kỳ. Thông tin này có thể là lệnh đọc được bằng máy tính, cấu trúc dữ liệu, mô đun chương trình, hoặc dữ liệu khác. Các ví dụ của vật ghi lưu trữ bằng máy tính bao gồm nhưng không bị giới hạn ở bộ nhớ đổi pha (PRAM - Phase-change Random Access memory), bộ nhớ truy cập ngẫu nhiên tĩnh (SRAM - Static Random Access memory), bộ nhớ truy cập ngẫu nhiên động (DRAM - Dynamic Random Access Memory), loại khác của bộ nhớ truy cập ngẫu nhiên (RAM - Random Access Memory), bộ nhớ chỉ đọc (ROM - read-only Memory), bộ nhớ chỉ đọc lập trình được xóa được bằng tín hiệu điện (EEPROM - Electrically Erasable Programmable Read-Only Memory), bộ nhớ flash hoặc công nghệ bộ nhớ khác, bộ nhớ chỉ đọc dùng đĩa nén (CD-ROM - Compact Disc Read-Only Memory), đĩa đa năng kỹ thuật số (DVD - Digital Versatile Disc) hoặc thiết bị lưu trữ quang khác, băng từ cassette, thiết bị lưu trữ đĩa băng từ/từ hoặc thiết bị lưu trữ từ khác, hoặc vật ghi không truyền khác bất kỳ. Vật ghi lưu trữ bằng máy tính có thể được sử dụng để lưu trữ thông tin truy cập được bởi thiết bị điện toán. Dựa trên phần mô tả trong bản mô tả này, vật ghi đọc được bằng máy tính không bao gồm vật ghi đọc được bằng máy tính khả biến (vật ghi khả biến) chẳng hạn như tín hiệu dữ liệu điều chế và bộ vận chuyển.

Đáng lưu ý thêm rằng, các thuật ngữ "bao gồm", "chứa", hoặc các biến thể khác bất kỳ của chúng nhằm bao hàm sự bao gồm không độc quyền, để quá trình, phương pháp, sản phẩm, hoặc thiết bị bao gồm danh sách phần tử không chỉ bao gồm các phần tử đó mà còn bao gồm các phần tử khác mà không được liệt kê rõ ràng, hoặc còn bao gồm các phần tử vốn có của quá trình, phương pháp, sản phẩm, hoặc thiết bị như vậy. Không có nhiều ràng buộc, phần tử đứng trước bởi "bao gồm ..." không loại trừ sự tồn

tại của các phần tử giống nhau bổ sung trong quá trình, phương pháp, sản phẩm, hoặc thiết bị bao gồm phần tử này.

Người có hiểu biết trung bình trong lĩnh vực kỹ thuật nên hiểu rằng phương án thực hiện của sáng chế có thể được đề xuất làm phương pháp, hệ thống, hoặc sản phẩm chương trình máy tính. Do đó, sáng chế có thể sử dụng dạng của các phương án thực hiện chỉ của phần cứng, các phương án thực hiện chỉ của phần mềm, hoặc các phương án thực hiện với sự kết hợp của phần mềm và phần cứng. Ngoài ra, sáng chế có thể sử dụng dạng của sản phẩm chương trình máy tính mà được thực hiện trên một hoặc nhiều vật ghi lưu trữ sử dụng được bằng máy tính (bao gồm nhưng không bị giới hạn ở bộ nhớ đĩa, CD-ROM, bộ nhớ quang, v.v.) chứa mã chương trình sử dụng được bằng máy tính.

Sáng chế có thể được mô tả trong ngữ cảnh chung của các lệnh thực thi được bằng máy tính được thực thi bởi máy tính, ví dụ, môđun chương trình. Nói chung, môđun chương trình bao gồm thường trình, chương trình, đối tượng, thành phần, cấu trúc dữ liệu, v.v. thực thi tác vụ cụ thể hoặc thực hiện loại dữ liệu tóm tắt cụ thể. Sáng chế cũng có thể được thực hiện trong môi trường điện toán phân tán. Trong các môi trường điện toán phân tán, các tác vụ được thực hiện bởi các thiết bị xử lý từ xa được nối thông qua mạng truyền thông. Trong môi trường điện toán phân tán, môđun chương trình có thể được bố trí trong cả vật ghi lưu trữ bằng máy tính cục bộ và từ xa bao gồm các thiết bị lưu trữ.

Các phương án thực hiện của bản mô tả này được mô tả theo cách lũy tiến. Đối với các phần giống hoặc tương tự của các phương án thực hiện, sự tham chiếu có thể được thực hiện cho các phương án thực hiện. Mỗi phương án thực hiện tập trung vào sự khác nhau với các phương án thực hiện khác. Đặc biệt, phương án thực hiện của hệ thống về cơ bản tương tự với phương án thực hiện của phương pháp, và do đó được mô tả một cách ngắn gọn. Đối với các phần liên quan, sự tham chiếu có thể được thực hiện cho các mô tả liên quan theo phương án thực hiện của phương pháp.

Phần mô tả trên đây là các phương án thực hiện theo sáng chế, và không nhằm giới hạn sáng chế. Người có hiểu biết trung bình trong lĩnh vực kỹ thuật có thể thực hiện các cải biến và thay đổi khác nhau của sáng chế. Cải biến, sự thay thế tương

đương, hoặc cải tiến bất kỳ được thực hiện mà không lệch khỏi nguyên lý của sáng chế nên nằm trong phạm vi bảo hộ của các điểm yêu cầu bảo hộ trong sáng chế.

YÊU CẦU BẢO HỘ

1. Phương pháp được thực hiện bằng máy tính để so khớp tên, phương pháp này bao gồm các bước:

nhận đầu vào bao gồm tên cần được so khớp, tên này bao gồm các từ;

xác định tập hợp tên thứ nhất tương ứng với tên này, tập hợp tên thứ nhất này bao gồm các phần tử;

thực hiện so khớp phân đoạn từ đối với tên cần được so khớp, bước này bao gồm các bước:

thu được chỉ số của mỗi tên nằm trong tập hợp tên thứ nhất, trong đó chỉ số này bao gồm từ được thiết lập trước liên quan đến tên này, và

phân đoạn tên cần được so khớp để thu được mỗi từ nằm trong tên cần được so khớp;

thực hiện so khớp tương tự đối với mỗi từ nằm trong tên cần được so khớp và mỗi chỉ số để thu được tập hợp con tương ứng của tập hợp tên thứ nhất, bằng cách thực hiện các kiểu thuật toán so khớp chuỗi và, đối với mỗi tổ hợp của kiểu thuật toán so khớp chuỗi và từ nằm trong tên cần được so khớp, thu được tập hợp con tương ứng, mỗi tập hợp con bao gồm các tên này trong tập hợp tên thứ nhất bao gồm từ tương tự khớp với từ tương ứng của tên cần được so khớp theo kiểu thuật toán so khớp chuỗi tương ứng;

xác định tập hợp tên chuẩn từ các tập hợp con, bước này bao gồm các bước:

(i) lấy bộ kết hợp của N tập hợp con đối với mỗi từ trong tên cần được so khớp để xác định M tập hợp kết hợp, trong đó M là số lượng từ trong tên cần được so khớp và N là số lượng kiểu thuật toán so khớp chuỗi, và chọn mỗi tên có tổng số lần xuất hiện trong M tập hợp kết hợp vượt quá ngưỡng đã định xác định, là tập hợp tên chuẩn, các tên được chọn dựa vào tổng số lần xuất hiện, hoặc

(ii) thu được N tập hợp con, một tập hợp cho mỗi kiểu thuật toán so khớp chuỗi, giao giữa N tập hợp con, và lấy tập hợp làm tập hợp tên chuẩn, hoặc trong đó tập hợp tên chuẩn chỉ bao gồm các tên tương tự đối với ít nhất một từ và theo ít nhất một thuật toán so khớp chuỗi cho tên cần được so khớp; hoặc

(iii) chọn làm các tên trong tập hợp tên chuẩn là một phần của các tên xuất hiện trong ít nhất hai trong số M tập hợp kết hợp;

xác định rằng tên này đồng nghĩa với ít nhất một tên chuẩn trong tập hợp tên chuẩn mặc dù một hoặc nhiều ký tự tên của tên này khác một hoặc nhiều ký tự tên chuẩn của tên chuẩn, bước này bao gồm bước thực hiện ít nhất một trong số dò tìm từ viết tắt, dò tìm thuật ngữ xung hô, dò tìm đa ngôn ngữ, hoặc dò tìm bí danh tên cần được so khớp để xác định xem liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên chuẩn trong tập hợp tên chuẩn hay không; và

để đáp lại việc xác định rằng tên này đồng nghĩa với ít nhất một tên chuẩn, tạo ra kết quả so khớp tên bao gồm tổng có trọng số của các độ so khớp của mỗi trong số các từ, các độ so khớp bao gồm các độ so khớp văn bản, độ trùng khớp ngữ âm, và độ so khớp chuỗi.

2. Phương pháp được thực hiện bằng máy tính theo điểm 1, phương pháp này còn bao gồm các bước:

gọi ra tập hợp tên định trước mà không cần được so khớp; và

xác định rằng tên cần được so khớp không nằm trong tập hợp tên định trước này.

3. Phương pháp được thực hiện bằng máy tính theo điểm 1, trong đó bước lấy bộ kết hợp của N tập hợp con đối với mỗi từ trong tên cần được so khớp để xác định M tập hợp kết hợp tạo ra kết quả so khớp toàn diện của mỗi từ.

4. Phương pháp được thực hiện bằng máy tính theo điểm 1, trong đó thuật toán so khớp chuỗi bao gồm một hoặc nhiều trong số các thuật toán sau đây: thuật

toán so khớp cây tiền tố, thuật toán so khớp cây từ điển, thuật toán so khớp tương tự chuỗi, và thuật toán so khớp tương tự cách phát âm.

5. Phương pháp được thực hiện bằng máy tính theo điểm 1, phương pháp này còn bao gồm các bước:

căn chỉnh, là tập hợp tên chuẩn được căn chỉnh, mỗi từ nằm trong tên trong tập hợp tên chuẩn với mỗi từ nằm trong tên cần được so khớp dựa vào sự tương tự giữa mỗi từ nằm trong tên chuẩn trong tập hợp tên chuẩn và mỗi từ nằm trong tên cần được so khớp; và

trong đó bước xác định xem liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên chuẩn hay không bao gồm bước:

dò tìm tên cần được so khớp dựa vào tập hợp tên chuẩn được căn chỉnh để xác định xem liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên chuẩn trong tập hợp tên chuẩn hay không.

6. Phương pháp được thực hiện bằng máy tính theo điểm 1, trong đó bước thực hiện dò tìm từ viết tắt đối với tên cần được so khớp bao gồm các bước:

gọi ra dữ liệu kết hợp đối chiếu từ viết tắt, trong đó mỗi sự kết hợp đối chiếu từ viết tắt phản ánh mối quan hệ ánh xạ từ viết tắt giữa ít nhất một từ và từ viết tắt của từ này;

dò tìm xem liệu từ nằm trong tên cần được so khớp có mối quan hệ ánh xạ từ viết tắt với từ nằm trong tên chuẩn trong tập hợp tên chuẩn hay không dựa vào dữ liệu kết hợp đối chiếu từ viết tắt; và

xác định, dựa vào kết quả dò tìm, xem liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên chuẩn trong tập hợp tên chuẩn hay không.

7. Phương pháp được thực hiện bằng máy tính theo điểm 1, trong đó bước thực hiện dò tìm thuật ngữ xưng hô đối với tên cần được so khớp bao gồm các bước:

gọi ra dữ liệu thuật ngữ xưng hô;

dò tìm xem liệu tên cần được so khớp có bao gồm thuật ngữ xung hô hay không, trong đó được xem là thuật ngữ xung hô này không ảnh hưởng đến ý nghĩa của tên chứa thuật ngữ xung hô này; và

xác định, dựa vào kết quả dò tìm, xem liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên chuẩn trong tập hợp tên chuẩn hay không.

8. Phương pháp được thực hiện bằng máy tính theo điểm 1, trong đó bí danh bao gồm biệt danh của tên tương ứng với bí danh này hoặc tên đồng nghĩa của tên tương ứng với bí danh này trong các lĩnh vực khác nhau.

9. Phương pháp được thực hiện bằng máy tính theo điểm 1, trong đó bước thực hiện dò tìm bí danh đối với tên cần được so khớp bao gồm ít nhất một trong số các bước sau đây:

dò tìm biệt danh của tên cần được so khớp, hoặc dò tìm tên đồng nghĩa của tên cần được so khớp trong các lĩnh vực khác nhau.

10. Phương pháp được thực hiện bằng máy tính theo điểm 1, trong đó bước thực hiện dò tìm đa ngôn ngữ đối với tên cần được so khớp bao gồm các bước:

xác định ngôn ngữ tương ứng với tên cần được so khớp;

gọi ra ít nhất một trong số quy tắc đồng nghĩa biến dạng chính tả hoặc quy tắc đồng âm biến dạng chính tả của ít nhất một trong số ngôn ngữ này hoặc ngôn ngữ khác; và

dò tìm tên cần được so khớp theo ít nhất một trong số quy tắc đồng nghĩa biến dạng chính tả hoặc quy tắc đồng âm biến dạng chính tả để xác định xem liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên chuẩn trong tập hợp tên chuẩn hay không.

11. Phương pháp được thực hiện bằng máy tính theo điểm 1, trong đó bước thực hiện dò tìm bí danh đối với tên cần được so khớp bao gồm các bước:

gọi ra dữ liệu kết hợp đối chiếu bí danh, trong đó mỗi dữ liệu kết hợp đối

chiếu bí danh phản ánh mối quan hệ ánh xạ bí danh giữa ít nhất một từ và bí danh của từ này;

dò tìm xem liệu từ nằm trong tên cần được so khớp có mối quan hệ ánh xạ bí danh với từ nằm trong tên chuẩn trong tập hợp tên chuẩn dựa vào dữ liệu kết hợp đối chiếu bí danh hay không; và

xác định, dựa vào kết quả dò tìm, xem liệu tên cần được so khớp có đồng nghĩa với ít nhất một tên chuẩn trong tập hợp tên chuẩn hay không.

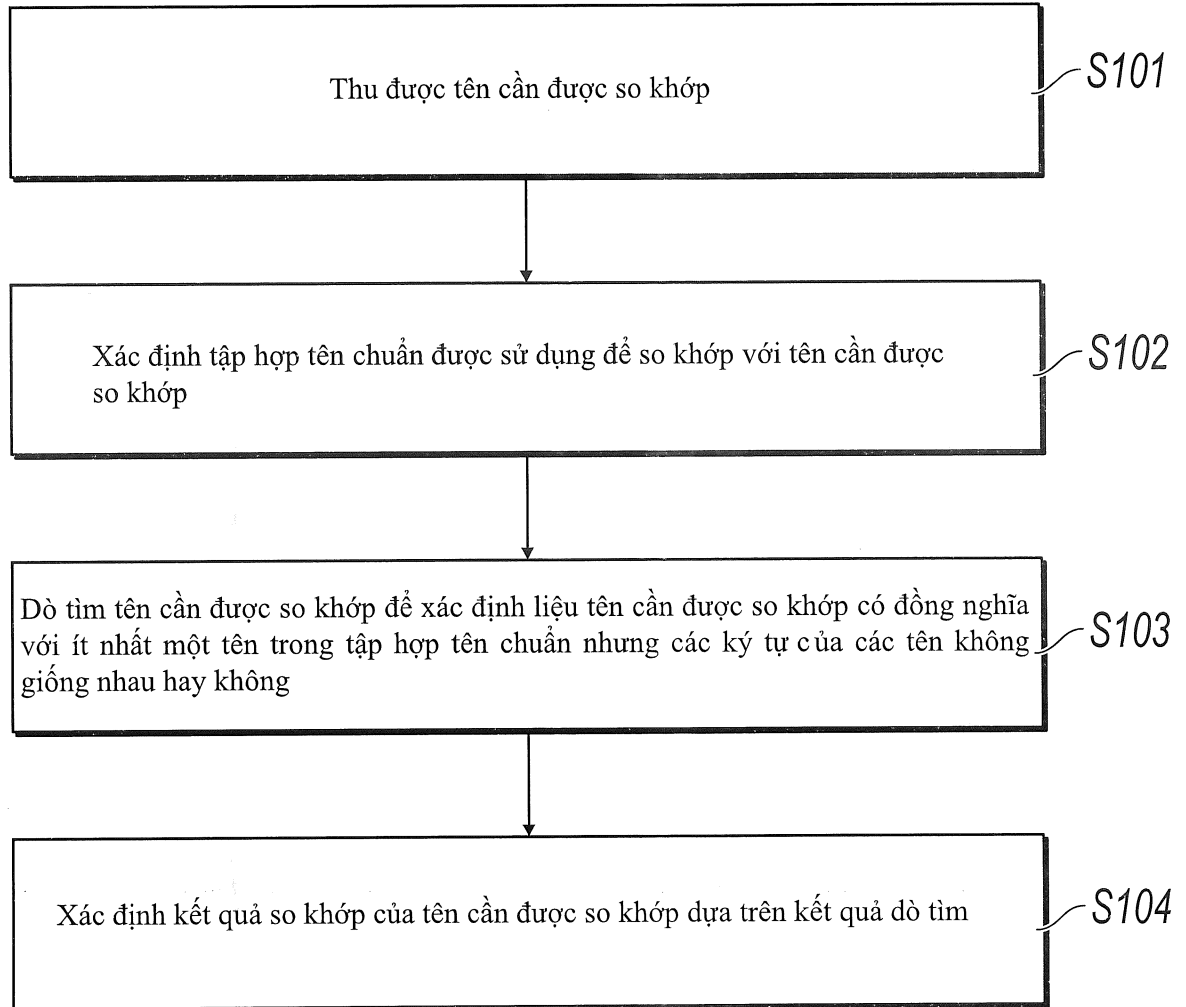
12. Phương pháp được thực hiện bằng máy tính theo điểm 11, trong đó bước tạo ra kết quả so khớp bao gồm các bước:

để đáp lại việc xác định rằng tên cần được so khớp đồng nghĩa với ít nhất một tên chuẩn trong tập hợp tên chuẩn xác định ít nhất một tên chuẩn là kết quả so khớp tên cần được so khớp; hoặc

để đáp lại việc xác định rằng tên cần được so khớp không đồng nghĩa với ít nhất một tên chuẩn trong tập hợp tên chuẩn, xác định kết quả so khớp của tên cần được so khớp bằng cách so khớp với tên cần được so khớp với tên chuẩn trong tập hợp tên chuẩn bằng cách sử dụng một hoặc nhiều thuật toán tương tự.

13. Phương pháp được thực hiện bằng máy tính theo điểm 12, trong đó thuật toán tương tự bao gồm ít nhất một trong số: thuật toán được sử dụng để tính toán độ trùng khớp văn bản, thuật toán được sử dụng để tính toán độ trùng khớp ngữ âm, hoặc thuật toán được sử dụng để tính toán độ trùng khớp chuỗi.

14. Thiết bị so khớp tên, thiết bị này bao gồm các môđun được tạo cấu hình để thực hiện phương pháp theo điểm bất kỳ từ 1 đến 13.

**FIG. 1**

2 / 5

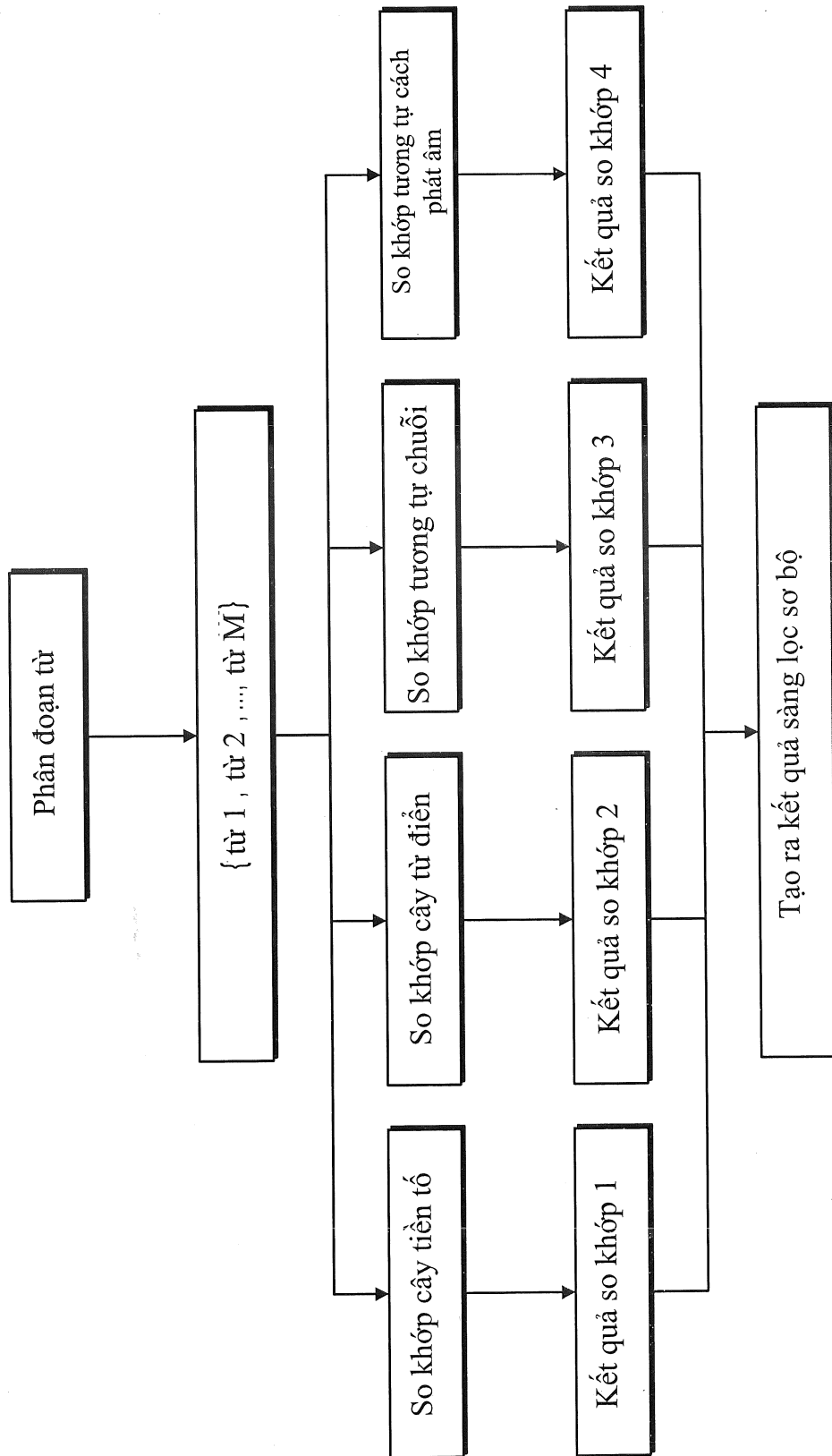


FIG. 2

3 / 5

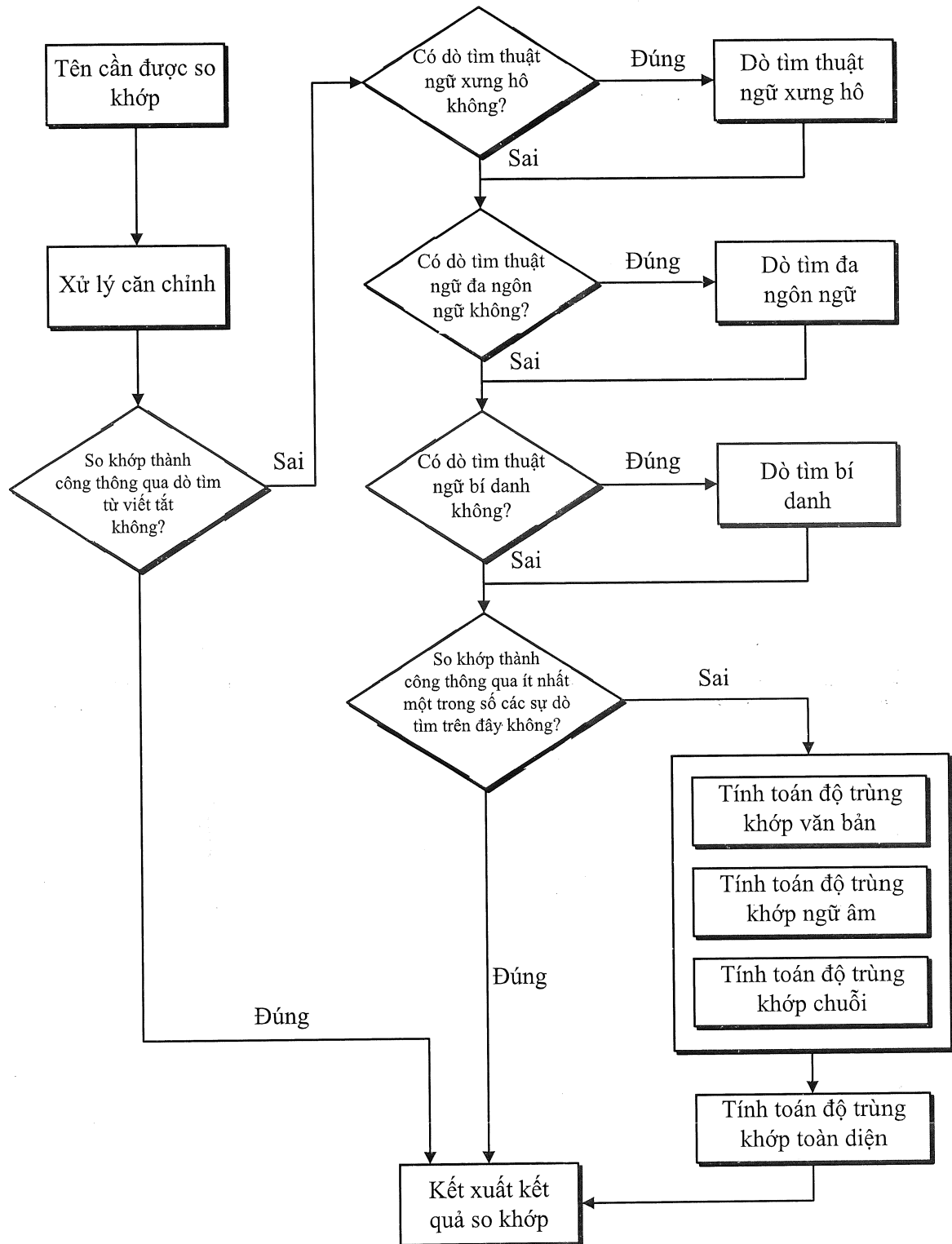
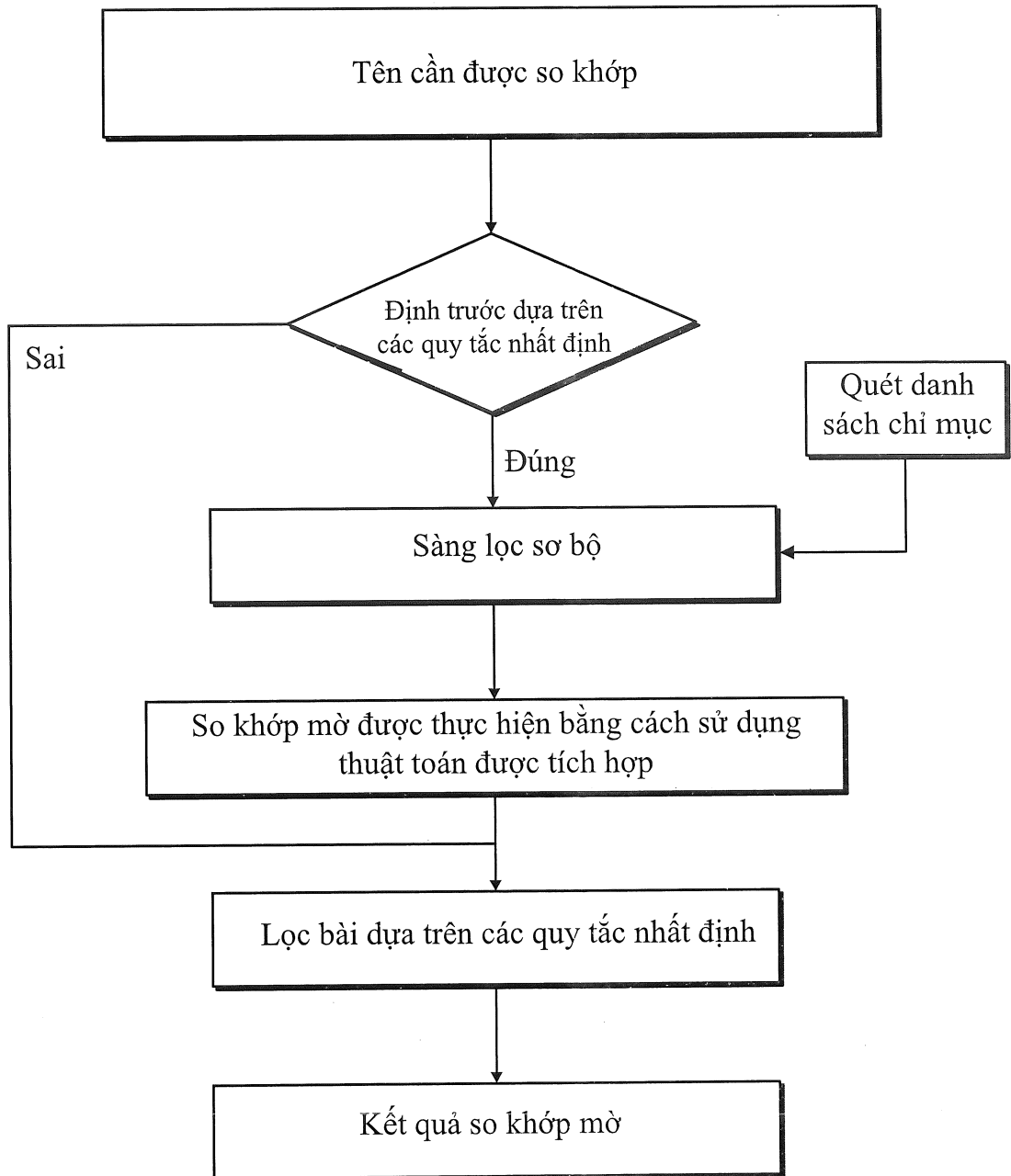
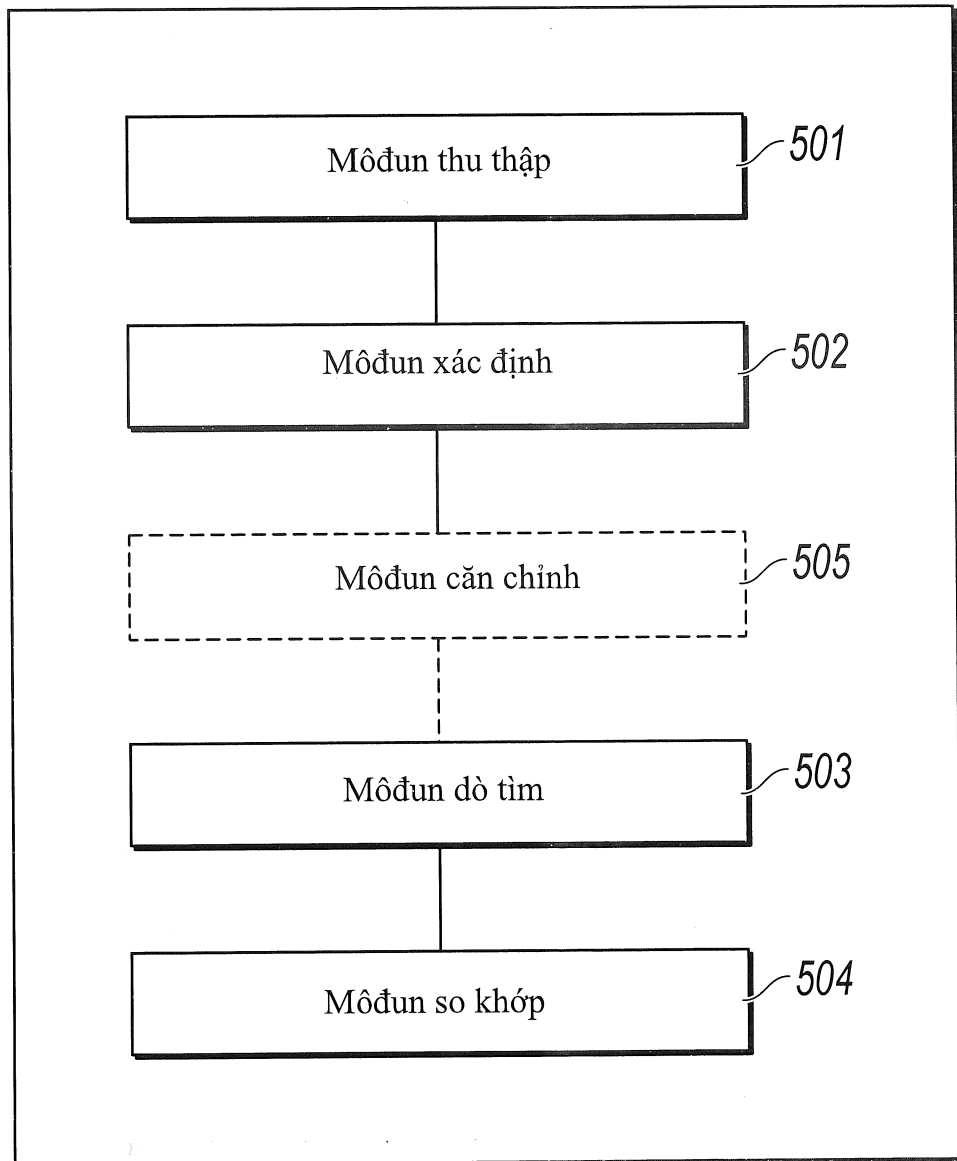


FIG. 3

**FIG. 4**

**FIG. 5**