



(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)



1-0026638

(51)⁷ G06F 17/28; G10L 15/00

(13) B

(21) 1-2015-03927

(22) 14/03/2014

(86) PCT/US2014/029244 14/03/2014

(87) WO/2014/144716 18/09/2014

(30) 61/793,044 15/03/2013 US

(45) 25/12/2020 393

(43) 25/02/2016 335A

(73) THE DUN & BRADSTREET CORPORATION (US)

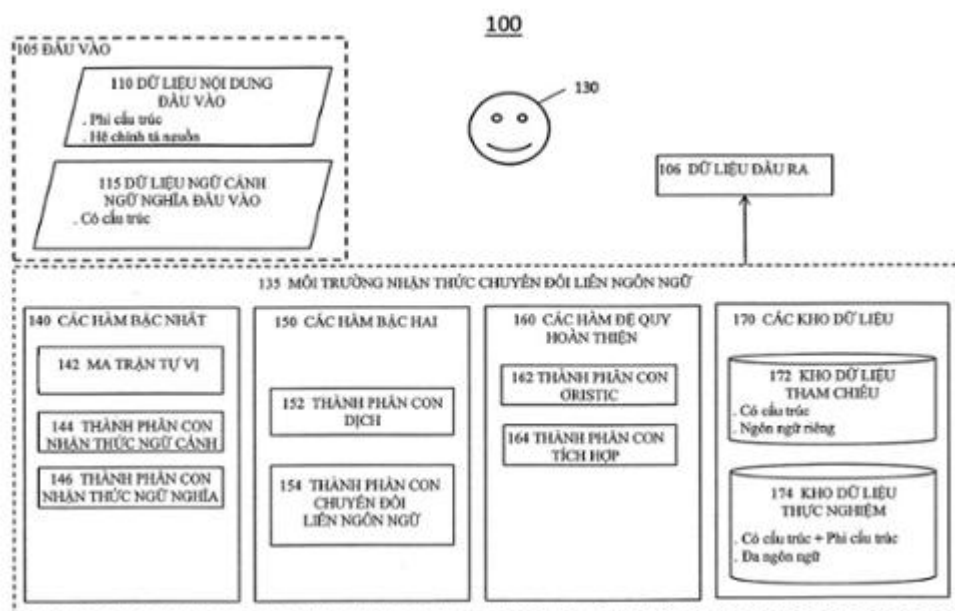
103 JFK Parkway, Short Hills, NJ 07078, USA

(72) SCRIFFIGNANO, Anthony, J. (US); KLEIN, Michael (US); CAROLAN, Sean (US); MATTHEWS, Warwick (AU).

(74) Công ty TNHH T&T INVENMARK Sở hữu trí tuệ Quốc tế (T&T INVENMARK CO., LTD.)

(54) PHƯƠNG PHÁP VÀ HỆ THỐNG CHUYỂN ĐỔI NGÔN NGỮ, THIẾT BỊ LƯU TRỮ

(57) Sáng chế đề cập đến phương pháp chuyển đổi ngôn ngữ bao gồm các bước phân tích cú pháp chuỗi ký tự thành các tự vị của nó, và tạo lập mẫu các ký tự biểu diễn dạng trừu tượng của các tự vị. Sáng chế còn đề cập đến hệ thống thực hiện phương pháp này, thiết bị lưu trữ chứa các lệnh để điều khiển bộ xử lý thực hiện phương pháp này.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực chuyển đổi ngôn ngữ, trong đó thông tin được chuyển đổi giữa hai hoặc nhiều ngôn ngữ hay hệ thống ghi, tạo ra các dạng biểu diễn bậc hai, bậc ba và nhiều bậc của thông tin gốc.

Tình trạng kỹ thuật của sáng chế

Các giải pháp được mô tả trong đoạn này là các giải pháp có thể được tìm kiếm, nhưng không nhất thiết là các giải pháp đã được nghĩ đến hoặc tìm kiếm trước đây. Do đó, trừ khi có quy định khác, các giải pháp này không thể là giải pháp kỹ thuật đã biết và không được thừa nhận là giải pháp kỹ thuật đã biết do đưa vào đoạn này.

Sáng chế đề cập đến lĩnh vực tự động chuyển đổi ngôn ngữ của dữ liệu, và tập trung cụ thể vào kỹ thuật chuyển đổi giữa các hệ chính tả khác nhau (như hệ thống chữ Kirin của Nga sang hệ thống chữ Latin) trong các ngữ cảnh cụ thể (như tên thực thể doanh nghiệp).

Các kỹ thuật đã biết không chuyển đổi thỏa đáng các phần khác nhau của tên ở ngôn ngữ thứ nhất sang tên ở ngôn ngữ thứ hai. Trong ngữ cảnh này, “các phần khác nhau” được dùng để chỉ các phần tử ngữ nghĩa như tên, tên địa lý, danh từ chung, tính từ mô tả, hậu tố kết hợp, v.v.. Ví dụ, có thể cần chuyển đổi tên của một doanh nghiệp ở nước Nga, vốn được viết bằng chữ Kirin, sang hệ thống chữ Latin “có thể hiểu được” đối với thánh giả nói tiếng Đức. Các kỹ thuật đã biết thường tiếp cận vấn đề này bằng cách thực hiện phép ánh xạ 1-1 và/hoặc dịch trực tiếp. Trong ngữ cảnh này, “ánh xạ 1-1” được dùng để chỉ việc lưu trữ và truy tìm một từ đơn trong ngôn ngữ đích đã được ánh xạ sang một từ trong dữ liệu nguồn (tên). Trong ngữ cảnh này, “dịch trực tiếp” được dùng để chỉ việc dịch ý nghĩa của một từ (hoặc

toàn bộ tên) từ ngôn ngữ nguồn sang ngôn ngữ đích. Do vậy, các kỹ thuật đã biết có được các bản chuyển đổi “có thể phát âm” nhưng không, ví dụ, chuyển đổi phần mô tả của tên ‘doanh nghiệp’ sang ngôn ngữ mà người nói tiếng Đức có thể hiểu.

Vấn đề khác với các kỹ thuật đã biết là trong trường hợp một kỹ thuật tạo ra bản dịch hoặc bản chuyển đổi sai, kỹ thuật này không có phương pháp tự động cải thiện chất lượng của bản dịch hoặc bản chuyển đổi. Tức là, các kỹ thuật đã biết không xét đến việc học hỏi từ thực nghiệm và tận dụng kinh nghiệm.

Bản chất kỹ thuật của sáng chế

Sáng chế đề cập đến phương pháp chuyển đổi ngôn ngữ bao gồm bước phân tích cú pháp chuỗi ký tự thành các tự vị của nó, và tạo lập mẫu các ký tự biểu diễn dạng trừu tượng hóa của các tự vị. Sáng chế còn đề cập đến hệ thống thực hiện phương pháp này, và thiết bị lưu trữ chứa các lệnh để điều khiển bộ xử lý thực hiện phương pháp này.

Mô tả vắn tắt các hình vẽ

Fig.1 là sơ đồ khối cấu trúc logic của quy trình tự động chuyển đổi ngôn ngữ của dữ liệu.

Fig.2 là sơ đồ khối cấu trúc logic của kho dữ liệu tham chiếu được sử dụng bởi quy trình trên Fig.1.

Fig.3 là sơ đồ khối cấu trúc logic của kho dữ liệu thực nghiệm được sử dụng bởi quy trình trên Fig.1.

Fig.4 là sơ đồ khối cấu trúc logic của các hàm bậc một của quy trình trên Fig.1.

Fig.5 là sơ đồ khối cấu trúc logic của các hàm bậc hai của quy trình trên Fig.1.

Fig.6 là sơ đồ khối cấu trúc logic của các hàm đệ quy hoàn thiện của quy trình trên Fig.1.

Fig.7 là lưu đồ của hoạt động làm ví dụ của các hàm bậc một của quy trình

trên Fig.1.

Fig.7A là hình vẽ chi tiết của một đoạn trên Fig.7, và minh họa lưu đồ của hoạt động làm ví dụ được thực hiện bởi quy trình ma trận tự vị.

Fig.7B là hình vẽ chi tiết của một đoạn trên Fig.7, và minh họa lưu đồ của hoạt động làm ví dụ được thực hiện bởi quy trình nhận thức ngữ cảnh.

Fig.7C là hình vẽ chi tiết của một đoạn trên Fig.7, và minh họa lưu đồ của hoạt động làm ví dụ được thực hiện bởi quy trình nhận thức ngữ nghĩa.

Fig.8 là lưu đồ của hoạt động làm ví dụ của các hàm bậc hai của quy trình trên Fig.1.

Fig.8A là hình vẽ chi tiết của một đoạn trên Fig.8, và minh họa sự tương tác của động cơ quy tắc và dịch vụ phối âm với kho quy tắc phối âm biến đổi.

Fig.8B minh họa quy trình xử lý bằng cách dịch và chuyển đổi liên ngôn ngữ với ví dụ chữ Kirin của Nga.

Fig.9 là lưu đồ của hoạt động làm ví dụ của các hàm đệ quy hoàn thiện.

Fig.9A là hình vẽ chi tiết của một đoạn trên Fig.9, và minh họa dạng biểu diễn tượng trưng của quy trình oristic và các kho dữ liệu được tham chiếu bởi các thành phần con của quy trình oristic.

Fig.9B là hình vẽ chi tiết của một đoạn trên Fig.9, và minh họa quy trình tích hợp và các kho dữ liệu được tham chiếu bởi các thành phần con của quy trình tích hợp.

Fig.9C minh họa dạng biểu diễn tượng trưng của động cơ quy tắc và dịch vụ phối âm.

Fig.10 là sơ đồ khối của hệ thống sử dụng các phương pháp được mô tả ở đây.

Thành phần hoặc dấu hiệu chung cho nhiều hơn một hình vẽ được biểu thị bằng cùng một số chỉ dẫn trong mỗi hình vẽ này.

Mô tả chi tiết sáng chế

Thuật ngữ “liên ngôn ngữ”, được sử dụng ở đây, và thuật ngữ “kiểu từ điển”, được sử dụng trong đơn yêu cầu cấp patent Mỹ tạm thời số 61/793.044, đều

có nghĩa là “giữa hoặc liên quan đến hai hoặc nhiều ngôn ngữ”.

Fig.1 là sơ đồ khối cấu trúc logic của quy trình 100 để tự động chuyển đổi ngôn ngữ của dữ liệu. Quy trình 100 thu từ người dùng 130, có thể là con người hoặc hệ thống gọi, đầu vào 105 được cấp cho môi trường nhận thức chuyển đổi liên ngôn ngữ 135, và tạo ra dữ liệu đầu ra 106, dữ liệu đầu ra là một phiên bản của đầu vào 105 đã được chuyển đổi giữa hai hoặc nhiều ngôn ngữ hay hệ thống ghi. Quy trình 100 tạo ra các dạng biểu diễn bậc hai, bậc ba và nhiều bậc của đầu vào 105, và nhờ vậy giúp cho người dùng 130 sự nhận thức vượt quá bản phiên âm chữ cái giữa hệ chính tả nguồn và hệ chính tả đích.

Quy trình 100 cung cấp cho người dùng 130 sự nhận thức, bao gồm nhưng không chỉ giới hạn ở, suy luận tương tự, trong một miền cụ thể của liên ngôn ngữ, tức là giữa các ngôn ngữ, hoặc liên chính tả, tức là giữa các hệ thống ghi, chuyển đổi hoặc dịch liên ngôn ngữ theo ngữ nghĩa và phi ngữ nghĩa, theo ngữ cảnh và phi ngữ cảnh. Quy trình 100 cung cấp cho người dùng 130 khả năng nhận biết, phân tích, so sánh, đối chiếu hoặc thu thập thông tin chứa trong đầu vào 105 với nhiều hình thái đồng thời, tức là, thông tin được biểu diễn bằng một hoặc nhiều ngôn ngữ hoặc hệ thống ghi, để phiên âm đầu vào 105 giữa các ngôn ngữ khác nhau, các hệ thống chữ hoặc các hệ thống ghi (các hình thái) khác nhau, bằng cách, ngoài những việc khác, nhận dạng các phần tử thực hoặc chỉ định dấu hiệu cho đầu vào 105. Các phần tử thực này được dùng làm các từ cùng gốc, cho phép so sánh ý nghĩa của dữ liệu bắt nguồn ở các hình thái khác loại.

Đầu vào 105 bao gồm dữ liệu nội dung nhập 110 và dữ liệu ngữ cảnh ngữ nghĩa nhập 115.

Dữ liệu nội dung nhập 110 là dữ liệu chủ đề của chính đầu vào đang đến, thường sẽ là tên của doanh nghiệp, được biểu diễn bằng một ngôn ngữ và hệ thống ghi (hệ chính tả) cụ thể. Dữ liệu nội dung nhập 110 là “phi cấu trúc” vì không có sẵn hướng dẫn cho nội dung của dữ liệu nội dung nhập 110 để hỗ trợ thi hành quy trình 100.

Dữ liệu ngữ cảnh ngữ nghĩa nhập 115 là dữ liệu ngữ cảnh có thể được phát hiện hoặc suy ra từ, ngoài những việc khác, việc phân tích đầu vào 105, ngữ cảnh,

lịch sử hoặc hoàn cảnh mà đầu vào 105 được cung cấp, hoặc siêu dữ liệu của đầu vào 105. Dữ liệu ngữ cảnh ngữ nghĩa nhập 115 được coi là “có cấu trúc” vì nó là siêu dữ liệu về dữ liệu nội dung nhập 110, ví dụ, nguồn của dữ liệu nội dung nhập 110, ngày tháng nhận được dữ liệu nội dung nhập 110, và hệ thống đã truyền dữ liệu nội dung nhập 110 đến hệ thống thực hiện quy trình 100.

Quy trình 100 gồm chức năng trên một số miền con hoặc một số tập con chức năng trong môi trường nhận thức chuyển đổi liên ngôn ngữ 135, tức là các hàm bậc một 140, các hàm bậc hai 150, và các hàm đệ quy hoàn thiện 160. Môi trường nhận thức chuyển đổi liên ngôn ngữ 135 còn có các kho dữ liệu 170.

Các kho dữ liệu 170 là các phương tiện lưu trữ dữ liệu, và bao gồm kho dữ liệu tham chiếu 172 và kho dữ liệu thực nghiệm 174. Kho dữ liệu thực nghiệm 174 được cập nhật dựa vào kinh nghiệm nhận được khi thi hành quy trình 100. Kho dữ liệu tham chiếu 172 được cập nhật theo các quy tắc mục tiêu và các tiêu chuẩn thay vì dựa trên kinh nghiệm nhận được khi thi hành quy trình 100. Việc phân tách các kho dữ liệu 170 thành kho dữ liệu tham chiếu 172 và kho dữ liệu thực nghiệm 174 chỉ để thuận tiện giải thích, và không nhất thiết phản ánh sự phân chia vật lý của các kho có liên quan.

Các hàm bậc một 140 là tập hợp các hàm xử lý dữ liệu đến đầu vào, tức là, đầu vào 105, và bao gồm ba thành phần con, tức là ma trận tự vị 142, thành phần con nhận thức ngữ cảnh 144, và thành phần con nhận thức ngữ nghĩa 146.

Các hàm bậc hai 150 là tập hợp các hàm và các quy trình thao tác trên tổ hợp của đầu vào 105 và các đầu ra của các hàm bậc một 140. Hàm bậc hai 150 bao gồm hai thành phần con, tức là thành phần con dịch 152, và thành phần con chuyển đổi liên ngôn ngữ 154.

Các hàm đệ quy hoàn thiện 160 là tập hợp các hàm thao tác trên các kết quả nhận được từ các hàm bậc một 140 và các hàm bậc hai 150, cũng như các đầu vào khác suy ra được từ quá trình nhận biết và phân tích sự thực hiện của quy trình 100, để cải thiện hiệu suất và hiệu quả của quy trình 100. Việc phân tích này bao gồm hệ thống hóa và tổng hợp dữ liệu tham chiếu có trong kho dữ liệu thực nghiệm 174. Hàm đệ quy hoàn thiện 160 bao gồm hai thành phần con, tức là thành phần con

oristic 162 và thành phần con tích hợp 164.

Fig.2 là sơ đồ khối cấu trúc logic của kho dữ liệu tham chiếu 172. Kho dữ liệu tham chiếu 172 bao gồm:

(a) kho từ đồng nghĩa 205, lưu trữ các tập hợp từ đồng nghĩa và các mục nhập thay thế cho các từ cụ thể hoặc các thành phần con ngôn ngữ khác;

(b) kho kiểu dáng 210, chứa thông tin và dữ liệu định tính, như các trọng số hoặc điểm số tương đối, về các dạng kiểu dáng của ngôn ngữ ghi;

(c) kho tiêu chuẩn hóa 215, chứa các quy tắc và từ điển để hỗ trợ cho việc tiêu chuẩn hóa các từ, các cụm từ hoặc các thành phần con ngôn ngữ khác;

(d) từ điển dịch 220, chứa các quy tắc để dịch các từ cụ thể, các cụm từ hoặc các thành phần con ngôn ngữ khác từ hệ chính tả nguồn sang hệ chính tả đích, và bản dịch tiềm năng giữa hai hệ chính tả (tức là, hệ chính tả nguồn và hệ chính tả đích);

(e) kho chuyển đổi liên ngôn ngữ 225, chứa các quy tắc để chuyển đổi liên ngôn ngữ các từ cụ thể, các cụm từ hoặc các thành phần con ngôn ngữ khác từ hệ chính tả nguồn sang hệ chính tả đích, và các bản chuyển đổi liên ngôn ngữ tiềm năng giữa hai hệ chính tả (tức là, hệ chính tả nguồn và hệ chính tả đích).

(f) kho loại nguồn 230, chứa thông tin về các loại nguồn dữ liệu;

(g) kho quy tắc phối âm biến đổi 235, chứa các quy tắc phối âm; và

(h) kho quy tắc tối ưu hóa 240, chứa các quy tắc để tối ưu hóa toàn hệ thống.

Fig.3 là sơ đồ khối cấu trúc logic của kho dữ liệu thực nghiệm 174. Kho dữ liệu thực nghiệm 174 bao gồm:

(a) kho phân tích tự vị 305, chứa các từ điển và các quy tắc để phân tích cú pháp và phân tích các tự vị;

(b) kho tần số mẫu tự vị 310, chứa các số đếm tần số của các tự vị;

(c) kho thông tin tính duy nhất 315, chứa các quy tắc, các bảng tần số và các từ điển liên quan đến tính duy nhất của các từ, các cụm từ và các thành phần con ngôn ngữ khác;

(d) kho nguồn 320, chứa thông tin về các nguồn dữ liệu cụ thể;

(e) kho thông tin thống kê 325, chứa thông tin thống kê được tạo ra khi thi

hành các thành phần con của môi trường nhận thức chuyển đổi liên ngôn ngữ 135;

(f) kho lưu trữ 330, chứa toàn bộ các đầu ra của quy trình xử lý đầu vào 105 bằng môi trường nhận thức chuyển đổi liên ngôn ngữ 135;

(g) kho tín hiệu 335, chứa, ngoài những thông tin khác, các cờ hiệu và các điểm số suy ra được từ quá trình thực hiện của môi trường nhận thức chuyển đổi liên ngôn ngữ 135;

(h) kho thay thế 340, chứa các mục nhập thay thế cho các từ cụ thể, các cụm từ và các thành phần con ngôn ngữ khác; và

(i) kho dữ liệu thực hiện 345, chứa thông tin thống kê liên quan đến các chỉ báo thực hiện chủ chốt đối với môi trường nhận thức chuyển đổi liên ngôn ngữ 135.

Fig.4 là sơ đồ khối cấu trúc logic của các hàm bậc một 140. Như nêu trên, hàm bậc một 140 bao gồm ma trận tự vị 142, thành phần con nhận thức ngữ cảnh 144, và thành phần con nhận thức ngữ nghĩa 146.

Ma trận tự vị 142 là tập hợp các quy trình con thành phần thao tác trên đầu vào 105 ở các mức ngữ nghĩa cơ bản nhất, rút gọn đầu vào 105 về các tự vị cơ bản của nó chẳng hạn. Ma trận tự vị 142 bao gồm bộ phân tích cú pháp & phân tích tự vị 405, bộ ánh xạ mẫu tự vị 410, và bộ lập mô hình mẫu tự vị 415.

Bộ phân tích cú pháp & phân tích tự vị 405 phân tích cú pháp và phân tích đầu vào 105, bằng cách sử dụng, ngoài những thông tin khác, các từ điển và siêu dữ liệu lưu trữ trong kho phân tích tự vị 305, để nhận dạng và chỉ định các phần tử ngữ nghĩa sẽ được cấp cho các quy trình khác.

Bộ ánh xạ mẫu tự vị 410 sử dụng đầu ra của bộ phân tích cú pháp & phân tích tự vị 405, và khử nhập nhằng, tức là, giải cấu trúc, các mẫu ngữ nghĩa của nội dung đầu vào 105 theo cách tượng trưng. Đầu ra của bộ ánh xạ mẫu tự vị 410 là mẫu tượng trưng, tức là, dạng trừu tượng hóa, bộc lộ cấu trúc của nội dung đầu vào 105. Ví dụ về việc khử nhập nhằng như vậy có thể là chuyển đổi “Jim’s Mowing Springvale” thành “PN-CD-GL”, trong đó “PN (Proper Noun)” là danh từ riêng, “CD (Commercial Description)” là mô tả thương mại và “GL (Geographic Location)” là vị trí địa lý.

Bộ lập mô hình mẫu tự vị 415 nhận đầu ra của bộ ánh xạ mẫu tự vị 410, và

sử dụng dữ liệu trong kho tần số mẫu tự vị 310, để phân biệt các mẫu tương tự như các tự vị cấu thành của đầu vào 105. Việc kiểm tra tính tương tự nhiều hơn so với tính tương tự mẫu bề ngoài.

Thành phần con nhận thức ngữ cảnh 144 là tập hợp các quy trình con thành phần thao tác trên đầu vào 105 ở mức ngữ cảnh. Tức là, chúng phân tích đầu vào 105 có xem xét đến các thuộc tính và dấu hiệu đến từ nguồn gốc, định thời và nội dung của đầu vào 105, nhưng cao hơn phép phân tích ngữ nghĩa mức tự vị cơ bản. Thành phần con nhận thức ngữ cảnh 144 bao gồm bộ phân tích ngữ cảnh 420, bộ phân loại nguồn 425, bộ phân tích tính duy nhất 430 và bộ tạo dữ liệu thay thế 435.

Bộ phân tích ngữ cảnh 420 phân tích đầu vào 105 bằng cách phân tích nội dung của nó ở mức tập trung vào ý nghĩa tổng thể của nội dung, cũng như các thuộc tính được tạo bởi bộ ánh xạ mẫu tự vị 410 và bộ lập mô hình mẫu tự vị 415. Việc phân tích này bao gồm phân tích nội dung của đầu vào 105 để tìm, ngoài những thông tin khác, “các thời kỳ nghệ thuật” và “biệt ngữ”, và có thể tham chiếu đến các hàm như các bộ mã hóa địa lý, tức là, các dịch vụ giải quyết tính đồng nhất của các thực thể địa lý, và các từ điển công nghiệp, ví dụ, các danh mục chữ viết tắt riêng trong công nghiệp đối với quốc gia cụ thể theo ngôn ngữ cụ thể. Đầu ra chính của bộ phân tích ngữ cảnh 420 là siêu dữ liệu về phép phân tích được thực hiện bởi bộ phân tích ngữ cảnh 420, tức là, phân loại và mô tả đặc điểm của nội dung đầu vào 105.

Có thể có nhiều lần lặp lại quy trình xử lý giữa bộ phân tích ngữ cảnh 420 và các thành phần của ma trận tự vị 142 vì các phân loại và các mẫu được lọc.

Bộ phân loại nguồn 425 phân tích siêu dữ liệu nguồn về đầu vào 105 được cấp trong dữ liệu ngữ cảnh ngữ nghĩa nhập 115, có tham chiếu đến dữ liệu lịch sử về các nguồn và các loại nguồn của các đầu vào trước đó vào quy trình 100, dữ liệu lịch sử được lưu trữ trong kho nguồn 320 và kho loại nguồn 230. Đầu ra của bộ phân loại nguồn 425 là dữ liệu mô tả về cấu trúc (như dữ liệu về kiểu dáng, âm điệu và cấu trúc ngữ pháp) và các dạng định tính (như độ xác thực, độ trung thực, độ biến đổi, tính đầy đủ và độ phức tạp) của các đầu vào điển hình từ các nguồn giống hoặc tương tự như nguồn của đầu vào 105.

Bộ phân tích tính duy nhất 430 phân tích đầu vào 105 về tính duy nhất, ở mức từ (hoặc thành phần con ngôn ngữ khác), cũng như tính duy nhất của các nhóm gồm các từ hoặc các cụm từ tương đối so với các đường cơ sở khác nhau, có xem xét đến kho thông tin tính duy nhất 315 làm tham chiếu. Đầu ra của bộ phân tích tính duy nhất 430 là các điểm số mô tả tính duy nhất tương đối của đầu vào 105 và các phần cấu thành của nó.

Bộ tạo dữ liệu thay thế 435 tạo ra các từ thay thế (hoặc các dạng biểu diễn tự vị khác), các cụm từ và các tên cho các phần cấu thành của đầu vào 105. Các bản thay thế này là ngữ cảnh (tức là chúng không dựa vào các bảng tần số chuẩn hoặc các quy tắc mức ngôn ngữ), dựa vào kinh nghiệm, và có nguồn gốc từ kho dữ liệu thay thế 340.

Thành phần con nhận thức ngữ nghĩa 146 là tập hợp các quy trình con thành phần thao tác trên đầu vào 105 ở mức ngôn ngữ. Tức là, chúng làm việc ở mức ngữ cảnh ngôn ngữ giả định của đầu vào (ví dụ, chữ Kirin của Nga hoặc một “ngôn ngữ” cụ thể hơn như chữ Kirin của Nga để đặt tên cho các cơ quan chính phủ). Thành phần con nhận thức ngữ nghĩa 146 bao gồm quy trình tiêu chuẩn hóa 440, bộ phân tích từ đồng nghĩa 445, và bộ phân tích kiểu dáng 450.

Quy trình tiêu chuẩn hóa 440 thực hiện các quy trình làm sạch, phân tích cú pháp và tiêu chuẩn hóa đối với đầu vào 105 để tạo ra “tổng quan tiêu chuẩn tốt nhất” về nội dung của nó. Quy trình tiêu chuẩn hóa 440 sẽ sử dụng dữ liệu trong kho tiêu chuẩn hóa 215.

Bộ phân tích từ đồng nghĩa 445 phân tích các từ, hoặc các thành phần con ngôn ngữ khác, của đầu vào 105 để có các từ đồng nghĩa dưới dạng các từ thay thế cho ngôn ngữ cụ thể của đầu vào 105. Bộ phân tích từ đồng nghĩa 445 sử dụng kho từ đồng nghĩa 205.

Bộ phân tích kiểu dáng 450 phân tích kiểu dáng của ngôn ngữ đầu vào 105 (bao gồm quan sát âm điệu, hình thức, biệt ngữ, từ viết tắt, chữ viết tắt, v.v.), và tính toán các điểm số và các chỉ báo để biểu diễn kiểu dáng được chỉ định. Bộ phân tích kiểu dáng 450 sẽ sử dụng dữ liệu trong kho kiểu dáng 345. Các đầu ra của bộ phân tích kiểu dáng 450 là các điểm số và các dấu hiệu mô tả tính chất kiểu dáng

của đầu vào 105.

Fig.5 là sơ đồ khối cấu trúc logic của các hàm bậc hai 150. Như nêu trên, hàm bậc hai bao gồm thành phần con dịch 152 và thành phần con chuyển đổi liên ngôn ngữ 154. Hàm bậc hai 150 sử dụng động cơ quy tắc 525 và dịch vụ phối âm 530.

Động cơ quy tắc 525 sử dụng các quy tắc có trong kho quy tắc phối âm biến đổi 235.

Dịch vụ phối âm 530 là hệ thống chuỗi công việc sử dụng các chuỗi công việc và logic quyết định có trong kho quy tắc phối âm biến đổi 235.

Động cơ quy tắc 525 và dịch vụ phối âm 530 làm việc phối hợp giữa các hàm bậc hai 150, tức là, qua thành phần con dịch 152 và thành phần con chuyển đổi liên ngôn ngữ 154, để ghép nối các đầu ra của các phần thành phần của các hàm bậc một 140, để thiết lập sắp đặt đầu vào 105 và các phần cấu thành của nó.

Các chuỗi công việc và các tập hợp quy tắc được thi hành bởi động cơ quy tắc 525 và dịch vụ phối âm 530 được chứa trong kho quy tắc phối âm biến đổi 235. Các chuỗi công việc và các quy tắc này sẽ khai thác các dấu hiệu, các điểm số và dữ liệu khác để tạo nên các đầu ra của hàm bậc một 140.

Thành phần con dịch 152 bao gồm các quy trình con để chuyển đổi các từ (hoặc các thành phần con ngôn ngữ khác), từ đầu vào 105 giữa các ngôn ngữ. Về vấn đề này, thành phần con dịch 152 bao gồm quy trình dò tìm từ dành riêng đã dịch 505 và quy trình dịch 510.

Quy trình dò tìm từ dành riêng đã dịch 505 là quy trình nhờ đó các phần của đầu vào 105, bao gồm siêu dữ liệu và các biến thể được tạo bởi các hàm bậc một 140, được phân tích bằng cách sử dụng từ điển dịch 220 để tạo ra các bản dịch chuyên dùng hoặc dựa vào ‘thời kỳ nghệ thuật’ dự bị của cùng một phần đầu vào.

Quy trình dịch 510 dịch các phần của đầu vào 105, bao gồm siêu dữ liệu và các biến thể được tạo bởi các hàm bậc một 140, giữa các ngôn ngữ, ví dụ, tiếng Nga và tiếng Anh. Quy trình dịch 510 có thể cần phải gọi ra các dịch vụ mạng, các ứng dụng hoặc các hệ thống khác để thực hiện các chức năng dịch.

Thành phần con chuyển đổi liên ngôn ngữ 154 bao gồm một vài quy trình

con để dịch các từ (hoặc các thành phần con ngôn ngữ khác), từ đầu vào 105 giữa các ngôn ngữ. Thành phần con chuyển đổi liên ngôn ngữ 154 bao gồm quy trình dò tìm từ dành riêng liên từ điển 515, và quy trình chuyển đổi 520.

Quy trình dò tìm từ dành riêng liên từ điển 515 là quy trình nhờ đó các phần của đầu vào 105, bao gồm siêu dữ liệu và các biến thể được tạo bởi các hàm bậc một 140, được phân tích bằng cách sử dụng từ điển chuyển tự 220 để tạo ra các bản dịch dự bị là các bản chuyển đổi chuyên dùng hoặc dựa trên ‘thời kỳ nghệ thuật’ của đầu vào 105 hoặc các phần của nó.

Quy trình chuyển đổi 520 chuyển tự các phần của đầu vào 105 giữa các hệ thống chữ, (ví dụ, từ hệ thống chữ Hy Lạp sang hệ thống chữ Latin). Quy trình chuyển đổi 520 có thể cần phải gọi ra các dịch vụ mạng, các ứng dụng hoặc các hệ thống khác để thực hiện chức năng chuyển tự.

Fig.6 là sơ đồ khối cấu trúc logic của các hàm đệ quy hoàn thiện 160. Như nêu trên, hàm đệ quy hoàn thiện 160 bao gồm thành phần con oristic 162 và thành phần con tích hợp 164. Động cơ quy tắc 525 và dịch vụ phối âm 530, được sử dụng bởi các hàm bậc hai 150, cũng được sử dụng bởi các hàm đệ quy hoàn thiện 160, và làm việc phối hợp giữa thành phần con oristic 162 và thành phần con tích hợp 164 để thực hiện tối ưu hóa và cải thiện đối với hiệu quả và hiệu suất của quy trình 100.

Như nêu trên, dịch vụ phối âm 530 là hệ thống chuỗi công việc, trong ngữ cảnh của các hàm đệ quy hoàn thiện 160, sử dụng các chuỗi công việc và logic quyết định có trong kho quy tắc tối ưu hóa 240.

Thành phần con oristic 162 là tập hợp các quy trình con thành phần để liên tục phân tích đầu ra (có trong kho lưu trữ 330) của tất cả các thành phần con của môi trường nhận thức chuyển đổi liên ngôn ngữ 135, cũng như dữ liệu đầu ra 106 được lưu trữ trong kho lưu trữ 330, để tối ưu hóa sự thực hiện của quy trình 100 theo hoạt động quan sát được. Quy trình 100, bằng thành phần con oristic 162, đang tự hoàn thiện. Tức là, thành phần con oristic 162 học hỏi từ thực nghiệm, và thay đổi hoặc sắp xếp lại thứ tự các chuỗi công việc được thi hành trong quy trình 100 để tạo ra kết quả tối ưu nhất hoặc kết quả cuối cùng. Thành phần con oristic 162 bao gồm bộ phân tích thống kê 605, bộ sắp thứ tự từ điển 610, và bộ tính điểm

số tín hiệu 615.

Bộ phân tích thống kê 605 thực hiện các phép phân tích thống kê, như phân tích tần số của các từ, các cụm từ hoặc các thành phần con ngôn ngữ khác của đầu vào 105 và thước đo của xu hướng chính qua dữ liệu lịch sử của các đầu vào và các đầu ra lịch sử của quy trình 105 của môi trường nhận thức chuyển đổi liên ngôn ngữ 135 có trong kho lưu trữ 330 và kho dữ liệu thực nghiệm 174, để tạo lập điểm số và các dấu hiệu khác lưu trữ trong kho thông tin thống kê 325 và sẽ có thể được dùng làm tài nguyên khi điều hướng các hàm bậc một 140 và các hàm bậc hai 150 bởi thành phần con tích hợp 164.

Bộ sắp thứ tự từ điển 610 sử dụng, ngoài những thông tin khác, các đầu ra của bộ phân tích thống kê 605, để tạo lập hoặc cập nhật các chuỗi công việc nhằm sắp xếp lại thứ tự của các từ điển, bao gồm kho chuyển đổi liên ngôn ngữ 225, kho từ đồng nghĩa 205, kho tần số mẫu tự vị 310, kho thông tin tính duy nhất 315, và kho dữ liệu thay thế 340, để cho các mục nhập cuối cùng hoặc tối ưu nhất hoặc được gửi trả về bởi các quy trình sử dụng các kho này (ví dụ, truy tìm dữ liệu từ kho tần số mẫu tự vị 310).

Bộ tính điểm số tín hiệu 615 thi hành các thường trình để gán các thuộc tính ưu tiên cho các dấu hiệu và các metric khác nhau suy ra được khi thực hiện quy trình 100, và truyền các điểm số này đến kho tín hiệu 335.

Thành phần con tích hợp 164 là tập hợp các quy trình con thành phần sử dụng, ngoài những thông tin khác, các đầu ra của thành phần con oristic 162, và sau đó cung cấp đầu vào cho các chuỗi công việc được thi hành bởi động cơ quy tắc 525 và dịch vụ phối âm 530 để tiến hành thay đổi các quy trình và các thường trình trong các thành phần con của môi trường nhận thức chuyển đổi liên ngôn ngữ 135, nhằm gia tăng hiệu suất và hiệu quả thực hiện của quy trình 100. Các thay đổi này được ghi dưới dạng các mục nhập trong kho quy tắc tối ưu hóa 240. Về vấn đề này, thành phần con tích hợp 164 có bộ phân tích dữ liệu lịch sử 620, và bộ tối ưu hóa 625.

Bộ phân tích dữ liệu lịch sử 620 phân tích các dấu hiệu thực hiện (bao gồm nhưng không chỉ giới hạn ở các thuộc tính thời gian thi hành, sử dụng tài nguyên,

sử dụng kho dữ liệu, chất lượng và tính xác thực, và điểm số phản hồi của người dùng) của môi trường nhận thức chuyển đổi liên ngôn ngữ 135 khi thi hành quy trình 100. Các dấu hiệu thực hiện được ghi trong kho thực hiện 345 bởi bộ phân tích dữ liệu lịch sử 620, và các dấu hiệu thực hiện này sẽ được đọc bởi bộ tối ưu hóa 625 để lựa chọn các quy trình để cập nhật hoặc sửa đổi.

Bộ tối ưu hóa 625 sử dụng, ngoài những thông tin khác, các dấu hiệu thực hiện được tạo bởi bộ phân tích dữ liệu lịch sử 620 và cập nhật kho quy tắc tối ưu hóa 240, và bắt đầu thực hiện các thường trình tối ưu hóa trong động cơ quy tắc 525 và dịch vụ phối âm 530.

Fig.7 là lưu đồ hoạt động của các hàm bậc một 140, với ví dụ về chữ Kirin của Nga.

Fig.7A là hình vẽ chi tiết của một đoạn trên Fig.7, và minh họa lưu đồ của hoạt động đang được thực hiện bởi ma trận tự vị 142, với ví dụ về chữ Kirin của Nga.

Fig.7B là hình vẽ chi tiết của một đoạn trên Fig.7, và minh họa lưu đồ của hoạt động đang được thực hiện bởi thành phần con nhận thức ngữ cảnh 144, với ví dụ về chữ Kirin của Nga.

Fig.7C là hình vẽ chi tiết của một đoạn trên Fig.7, và minh họa lưu đồ của hoạt động đang được thực hiện bởi thành phần con nhận thức ngữ nghĩa 146, với ví dụ về chữ Kirin của Nga.

Theo Fig.7, các hàm bậc một 140 nhận đầu vào 105, theo ví dụ này là chữ Kirin của Nga “Невское Электромонтажное Общество”, tức là, dữ liệu nội dung văn tin 110, từ “Hệ thống thu thập đối tác” nguồn vào ngày 01 tháng Một 2014, tức là, dữ liệu ngữ cảnh ngữ nghĩa nhập 115. Các hàm bậc một 140 tạo ra đầu ra tạm thời 760, sẽ được lưu trữ trong kho lưu trữ 330.

Theo Fig.7A, bộ phân tích cú pháp & phân tích tự vị 405 phân tích cú pháp đầu vào 105 và có tham chiếu kho phân tích tự vị 305, gán các phân loại cho các phần cấu thành của dữ liệu nội dung văn tin 110 (các tự vị, các từ, các cụm từ, v.v.).

Theo ví dụ này, bộ phân tích cú pháp & phân tích tự vị 405 phân tích dữ liệu

nội dung nhập 110, và phân loại nó như được thể hiện trong Bảng 1.

Bảng 1

Phần đầu vào	Siêu dữ liệu: loại	Siêu dữ liệu: sắp thứ tự
Невское	Danh từ - Tên	N1
Электромонтажное	Tính từ - bộ mô tả	N2
Общество	Danh từ – loại thực thể thương mại	S3

Bộ ánh xạ mẫu tự vị 410 nhận dữ liệu nội dung nhập 110 và siêu dữ liệu (được thể hiện trên đây trong các cột 2 và 3 của Bảng 1) được tạo bởi bộ phân tích cú pháp & phân tích tự vị 405, và tạo lập “Mẫu Tự vị”, là dạng trừu tượng hóa của cấu trúc ngữ pháp và ngữ nghĩa của dữ liệu nội dung văn tin 110.

Theo ví dụ này, bộ ánh xạ mẫu tự vị 410 tạo ra mẫu: 1N:GN-2N:CD-3S:IN biểu thị rằng đầu vào được tạo thành từ 2 phần chính, tên (N - Name) và hậu tố (S - Suffix), và cả ba phần chi tiết: tên địa lý (GN - Geographic Name), mô tả thương mại (CD - Commercial Description) và hậu tố kết hợp (IN - Incorporation Suffix).

Bộ lập mô hình mẫu tự vị 415 nhận đầu ra của bộ ánh xạ mẫu tự vị 410 và thực hiện tìm kiếm trên kho tần số mẫu tự vị 310 để tìm ra các mẫu tương tự một cách đáng kể.

Bảng 2 thể hiện ví dụ của một số mẫu tìm được bởi bộ lập mô hình mẫu tự vị 415. Thực tế, các mẫu khác cũng có thể tìm được, như các mẫu được thể hiện trên Fig.7A, ở trong bộ lập mô hình mẫu tự vị 415.

Bảng 2

Mẫu gốc (được tạo bởi bộ ánh xạ mẫu tự vị 410)	Các mẫu tìm được (tìm được bởi bộ lập mô hình mẫu tự vị 415)
1N:GN-2N:CD-3S:IN	GN-CD-*
	PN-GN-CD-IN
	CD-GN-IN
	IN-CD-GN

Theo Fig.7B, bộ phân loại nguồn 425 phân tích siêu dữ liệu nguồn của đầu vào (dữ liệu ngữ cảnh ngữ nghĩa nhập 115). Theo ví dụ này, “hệ thống thu thập đối tác” nguồn được tìm kiếm với từ khóa “PCS (Partner Collection System)” và bộ phân loại nguồn 425 truy tìm từ 320 siêu dữ liệu kho nguồn (SOURCES STORE) như được thể hiện trong Bảng 3.

Bảng 3

Mã nguồn	Siêu dữ liệu
PCS	Điểm số có mặt hậu tố kết hợp: 10
	Điểm số về độ thay đổi: 33
	Điểm số về độ trung thực: 62

Danh mục siêu dữ liệu trong Bảng 2 chỉ là ví dụ, và không là tập hợp đóng.

Bộ phân tích ngữ cảnh 420 nhận đầu vào 105 và các đầu ra của ma trận tự vị 142, và có tham chiếu bộ mã hóa địa lý và từ điển thương mại, tạo ra các phân loại chi tiết của các phần cấu thành (các từ và các cụm từ, hoặc dạng tương đương) của nội dung của đầu vào 105. Theo ví dụ này, các phân loại chi tiết được thể hiện trong Bảng 4.

Bảng 4

Phần đầu vào	Phân tích ngữ cảnh
Невское	Danh từ riêng – PLACE, ST. PETERSBURG; METRO SUBURB, RIVER; Nhóm ngữ nghĩa - MARITIME
Электромонтажное	Tính từ – INDUSTRIAL; NICHE - ENGINEERING; NICHE - ELECTRICAL
Общество	Danh từ – Loại kết hợp; DOMESTIC; GENERIC; MID-SIZE

Siêu dữ liệu mới, (tức là, ngữ cảnh phân tích được thể hiện trong Bảng 4) có thể được lưu trữ dưới dạng các mã hoặc thẻ khóa để cho các thành phần khác sử dụng có hiệu quả.

Bộ tạo dữ liệu thay thế 435 nhận đầu vào 105 và, có tham chiếu kho dữ liệu thay thế 340, tạo ra dữ liệu thay thế như được thể hiện trong Bảng 5.

Bảng 5

Phần đầu vào	Dữ liệu thay thế
Невское	1. Санкт-Петербург, St Petersburg 2. Река Нева, Neva River
Электромонтажное	1. электрических подрядчика
Общество	[không có]

Bộ phân tích tính duy nhất 430 nhận đầu vào 105 (kể cả đầu ra của các phần khác của các hàm bậc một 140), và có tham chiếu kho thông tin tính duy nhất 315, tạo ra các điểm số về tính duy nhất, tức là điểm số biểu thị tính duy nhất của các phần của đầu vào 105. Điểm số về tính duy nhất với ví dụ này được thể hiện trong Bảng 6.

Bảng 6

Phần đầu vào	Mẫu	Tính duy nhất
Невское Электромонтажное Общество	GN-CD-IN	100
Невское Электромонтажное	GN-CD	86
Невское	GN	15
Электромонтажное	CD	6
Общество	IN	1

Bộ phân tích tính duy nhất 430, khi tạo lập các điểm số về tính duy nhất, cũng tính đến dữ liệu thay thế được tạo bởi bộ tạo dữ liệu thay thế 435.

Theo Fig.7C, quy trình tiêu chuẩn hóa 440 chuẩn hóa nội dung của đầu vào 105 (kể cả các bản thay thế được tạo bởi bộ tạo dữ liệu thay thế 43) bằng cách sử dụng các quy tắc riêng của từ điển (trong trường hợp này từ điển thích hợp có thể là “các tên thương mại chữ Kirin của Nga”). Theo ví dụ này, đầu vào có thể được tiêu chuẩn hóa như được thể hiện trong Bảng 7.

Bảng 7

Phần đầu vào	Tiêu chuẩn hóa
Невское	НЕВА
санкт-петербург	САНКТ-ПЕТЕРБУРГ
Река Нева	НЕВА
Электромонтажное	ЭЛЕКТРИЧЕСКИЙ
электрических подрядчик	ЭЛЕКТРИЧЕСКИЙ ПОДРЯДЧИК
Общество	КОМПАНИЯ, ООО

Bộ phân tích từ đồng nghĩa 445 tìm kiếm đầu vào 105 và các phần của nó trong kho từ đồng nghĩa 205, để tạo ra các từ đồng nghĩa như được thể hiện, ví dụ, trong Bảng 8, cho các phần cấu thành của đầu vào 105 và các bản thay thế được tạo bởi bộ tạo dữ liệu thay thế 435.

Bảng 8

Phần đầu vào	Các từ đồng nghĩa
Невское Электромонтажное Общество	1. Невский Электромонтажное Общество 2. Невский Кримпер Общество
Электромонтажное	[nil]
Общество	1. Компания 2. Фирма 3. Общество 4. рота

Bộ phân tích kiểu dáng 450 phân tích kiểu dáng của đầu vào 105 qua một số cỡ, và tạo lập siêu dữ liệu để biểu diễn kết quả phân tích này. Theo ví dụ này, bộ phân tích kiểu dáng 450 tạo ra đầu ra như được thể hiện trong Bảng 9.

Bảng 9

Đầu vào	Kiểu dáng siêu dữ liệu
Невское Электромонтажное Общество	Điểm số về hình thức: 88 Điểm số về từ viết tắt: 0 Điểm số về biệt ngữ: 15

Trong Bảng 9, cột Kiểu dáng siêu dữ liệu mô tả, theo một ví dụ, việc gán cho đầu vào “Невское Электромонтажное Общество” các điểm số dựa trên kiểu dáng để phân loại kiểu dáng của đầu vào 105 dựa vào các khía cạnh như sự sử dụng biệt ngữ riêng của ngôn ngữ, sử dụng từ viết tắt và hình thức ngữ pháp của cấu trúc của đầu vào 105.

Fig.8 là lưu đồ của hoạt động làm ví dụ của các hàm bậc hai 150, sử dụng đầu vào là chữ Kirin của Nga làm ví dụ.

Fig.8A là hình vẽ chi tiết của một đoạn trên Fig.8, và mô tả sự tương tác của động cơ quy tắc 525 và dịch vụ phối âm 530 với kho quy tắc phối âm biến đổi 235, để minh họa mối quan hệ của chúng với quy trình xử lý ví dụ chữ Kirin của Nga trên Fig.8.

Theo Fig.8A, đầu vào 105 và tất cả các đầu ra của các hàm bậc một 140 được dùng bởi động cơ quy tắc 525, có tham chiếu toàn bộ dữ liệu này và kho quy tắc phối âm biến đổi 235, tạo ra các quy tắc điều khiển dữ liệu (“các chuỗi công việc”) sẽ được lưu trữ trong kho quy tắc phối âm biến đổi 235, các chuỗi công việc này xác định chuỗi các bước tiếp theo trong quy trình 100 để sắp đặt đầu vào 105.

Dịch vụ phối âm 530, thi hành các chuỗi công việc được quy định bởi động cơ quy tắc 525 và được lưu trữ trong kho quy tắc phối âm biến đổi 235.

Bảng 10 lập danh mục, với ví dụ này, các bước của chuỗi công việc được quy định bởi động cơ quy tắc 525.

Bảng 10

Phần đầu vào	Các bước công việc
Невское	1. Gửi tới quy trình dò tìm từ dành riêng liên từ điển 515 2. Gửi tới quy trình chuyển đổi 520
Электромонтажное	3. Gửi tới quy trình dò tìm từ dành riêng đã dịch 505 4. Gửi tới quy trình dịch 510
Общество	5. Gửi tới quy trình dò tìm từ dành riêng liên từ điển 515
Невское Электромонтажное	6. Gửi tới quy trình chuyển đổi 520

Các bước của chuỗi công việc được thể hiện trong Bảng 10 chỉ là một tập hợp con nhỏ của các lệnh có thể cần cho ví dụ này. Tập hợp đầy đủ có thể bao gồm các tác động trên nhiều biến thể của đầu vào 105 và các phần chỉ định của nó được tạo bởi các hàm bậc một 140.

Fig.8B là hình vẽ chi tiết của một đoạn trên Fig.8, và mô tả dữ liệu làm ví dụ được xử lý bởi thành phần con dịch 152 và thành phần con chuyển đổi liên ngôn ngữ 154 với ví dụ chữ Kirin của Nga trên Fig.8. Về vấn đề này, thành phần con dịch 152 tạo ra dữ liệu dịch 860, và thành phần con chuyển đổi liên ngôn ngữ 154 tạo ra dữ liệu chuyển đổi 870.

Quay trở lại Fig.8, dữ liệu dịch 860 và dữ liệu chuyển đổi 870 được kết hợp để tạo ra dữ liệu đầu ra 106. Cụ thể hơn, động cơ quy tắc 525 thi hành các quy tắc dựa vào các kết quả của thành phần con dịch 152 và thành phần con chuyển đổi liên ngôn ngữ 154 để hợp dịch kết quả cuối cùng, tức là, dữ liệu đầu ra 106, sẽ được lưu trữ trong kho lưu trữ 330.

Bảng 11 thể hiện nội dung mẫu của dữ liệu dịch 860 và dữ liệu chuyển đổi 870.

Bảng 11

Phần đầu vào	Dữ liệu dịch 860	Dữ liệu chuyển đổi 870
Невское	[nil]	NEVSKOYE, NEVA
Электромонтажное	WIRING, ELECTRICAL	ELEKTROMONTAZHNOYE
Общество	ASSOCIATION, SOCIETY, ENTERPRISES	OBSHCHESTVO, OOO, LLC

Bảng 12 thể hiện phiên bản tổng hợp cuối cùng sẽ trở thành dữ liệu đầu ra 106.

Bảng 12

Phần đầu vào	Tổng hợp để tạo dữ liệu đầu ra 106
Невское	NEVA
Электромонтажное	ELECTRICAL
Общество	LLC

Như vậy “Невское Электромонтажное Общество” được chuyển đổi thành “NEVA ELECTRICAL LLC”.

Fig.9 là lưu đồ của hoạt động làm ví dụ của các hàm đệ quy hoàn thiện 160. Fig.9 mô tả dạng biểu diễn tượng trưng của các hàm đệ quy hoàn thiện 160 và các kho dữ liệu được tham chiếu bởi các thành phần con của hàm đệ quy hoàn thiện 160.

Fig.9A là hình vẽ chi tiết của một đoạn trên Fig.9, và mô tả thành phần con oristic 162 và bộ phân tích thống kê các thành phần con của nó 605, bộ tính điểm số tín hiệu 615, và bộ sắp thứ tự từ điển 610, cũng như các kho dữ liệu tương tác với các thành phần con này.

Fig.9B là hình vẽ chi tiết của một đoạn trên Fig.9, và mô tả thành phần con tích hợp 164 và các kho dữ liệu được tham chiếu bởi các tập hợp con của thành phần con tích hợp 164.

Fig.9C là hình vẽ chi tiết của một phần Fig.9, và mô tả sự tương tác của động cơ quy tắc 525 và dịch vụ phối âm 530 với kho quy tắc tối ưu hóa 240.

Như vậy, quy trình 100 là phương pháp bao gồm các bước:

(a) phân tích cú pháp đầu vào 105, tức là, chuỗi ký tự, thành các tự vị của nó (xem Fig.7A, bộ phân tích cú pháp & phân tích tự vị 405); và

(b) tạo lập mẫu các ký tự biểu diễn dạng trừu tượng hóa của các tự vị (xem Fig.7A, bộ ánh xạ mẫu tự vị 410).

Mẫu các ký tự là nhóm các ký tự tương ứng với một tự vị trong các tự vị của đầu vào 105. Ví dụ, trên Fig.7A, mẫu “1N:GN-2N:CD-3S:IN” bao gồm nhóm ký tự “CD”, tương ứng với ký hiệu “mô tả thương mại”, và được ánh xạ bởi bộ ánh xạ mẫu tự vị 410 có tham chiếu kho phân tích tự vị 305.

Quy trình 100 còn bao gồm bước truy tìm từ nguồn dữ liệu, thông tin về tự vị. Ví dụ, xem Fig.7B, bộ phân tích ngữ cảnh 420.

Quay trở lại Fig.7A, bộ ánh xạ mẫu tự vị 410, mẫu ký tự bao gồm dãy gồm nhóm ký tự thứ nhất, ví dụ, GN, và nhóm ký tự thứ hai, ví dụ, CD. GN tương ứng với tự vị thứ nhất trong các tự vị của đầu vào 105, và CD tương ứng với tự vị thứ hai trong các tự vị của đầu vào 105. Theo Fig.8A, quy trình 100 còn bao gồm các bước (a) chọn, dựa vào dãy ký tự, một trong số các quy trình, và (b) thi hành quy trình này trên chuỗi ký tự.

Quy trình 100 mang lại sự kết hợp giữa việc dịch, tức là, dữ liệu dịch 860, và chuyển đổi liên ngôn ngữ, tức là, dữ liệu chuyển đổi 870, của chuỗi ký tự.

Fig.10 là sơ đồ khối của hệ thống 1000 sử dụng các phương pháp được mô tả ở đây. Hệ thống 1000 có máy tính 1005 nối với mạng truyền thông dữ liệu, tức là, mạng 1030, như mạng Internet.

Máy tính 1005 có giao diện người dùng 1010, bộ xử lý 1015, và bộ nhớ 1020. Mặc dù máy tính 1005 được thể hiện ở đây dưới dạng thiết bị độc lập, nhưng không chỉ giới hạn như vậy, mà còn có thể được ghép nối với các thiết bị khác (không được thể hiện) trong hệ thống xử lý phân tán.

Giao diện người dùng 1010 bao gồm thiết bị nhập, như bàn phím hoặc hệ thống con nhận dạng tiếng nói, để cho phép người dùng 130 truyền thông thông tin và sự lựa chọn chỉ thị \$ đến bộ xử lý 1015. Giao diện người dùng 1010 còn bao gồm thiết bị đầu ra như màn hình hoặc máy in. Cơ cấu điều khiển con trỏ như

xoay, hoặc cần điều khiển trò chơi, cho phép người dùng 130 điều khiển con trỏ trên màn hình để truyền thông tin bổ sung và sự lựa chọn chỉ thị đến bộ xử lý 1015.

Hệ thống 1000 còn bao gồm thiết bị người dùng 1045 được ghép nối truyền thông với máy tính 1005 mạng cục 1030. Người dùng 130 có thể tương tác với máy tính 205 qua thiết bị người dùng 1045, là cách khác để làm như vậy qua giao diện người dùng 1010.

Bộ xử lý 1015 là thiết bị điện tử được tạo cấu hình của mạch logic để đáp lại và thi hành các lệnh.

Bộ nhớ 1020 là vật ghi đọc được bằng máy tính phi nhất thời được mã hóa với chương trình máy tính. Về vấn đề này, bộ nhớ 1020 lưu trữ dữ liệu và lệnh có thể đọc và thi hành bởi bộ xử lý 1015 để điều khiển hoạt động của bộ xử lý 1015. Bộ nhớ 1020 có thể được lắp đặt trong bộ nhớ truy nhập ngẫu nhiên (RAM - Random Access Memory), ổ cứng, bộ nhớ chỉ đọc (ROM - Read Only Memory), hoặc tổ hợp của chúng. Một trong số các thành phần của bộ nhớ 1020 là môđun chương trình 1025.

Môđun chương trình 1025 chứa các lệnh để điều khiển bộ xử lý 1015 thi hành phương pháp được mô tả ở đây. Ví dụ, theo sự điều khiển của môđun chương trình 1025, bộ xử lý 1015 thi hành quy trình 100. Thuật ngữ "môđun" như được sử dụng ở đây để chỉ hoạt động chức năng có thể được thể hiện dưới dạng thành phần độc lập hoặc dưới dạng cấu hình tích hợp của nhiều thành phần phụ. Do vậy, môđun chương trình 1025 có thể được thực hiện dưới dạng một môđun hoặc dưới dạng nhiều môđun hoạt động phối hợp với nhau. Hơn nữa, mặc dù môđun chương trình 1025 được mô tả ở đây dưới dạng được lắp đặt trong bộ nhớ 1020, và do đó được cài đặt trong phần mềm, nhưng nó cũng có thể được thực hiện ở bất kỳ một trong số phần cứng (ví dụ, mạch điện tử), phần sụn, phần mềm, hoặc tổ hợp của chúng.

Bộ xử lý 1015 nhận đầu vào 105, qua mạng 1030 hoặc qua giao diện người dùng 1010. Đầu vào 105 có thể được cấp cho máy tính 1005, và do vậy được cấp cho quy trình 100, bởi người dùng 130 qua giao diện người dùng 1010 hoặc thiết bị

người dùng 1045. Đầu vào 105 còn có thể được cấp cho quy trình tự động, ví dụ như được suy ra từ các tệp tin được đệ trình bằng cách sử dụng các tính năng của máy xử lý theo bó, hoạt động trong máy tính 1005 hoặc trên thiết bị từ xa (không được thể hiện) ghép nối với máy tính 1005 qua mạng 1030. Các kho dữ liệu 170 có thể là các thành phần của máy tính 1005, ví dụ, được lưu trữ trong bộ nhớ 1020, hoặc có thể nằm bên ngoài máy tính 1005, ví dụ, trong cơ sở dữ liệu 1040, hoặc trong cơ sở dữ liệu (không được thể hiện) mà máy tính 1005 truy nhập qua mạng cục bộ (không được thể hiện) hoặc mạng 1030. Bộ xử lý 1015 gửi trả về dữ liệu đầu ra 106 qua mạng 1030 hoặc giao diện người dùng 1010.

Dù được thể hiện dưới dạng đã được nạp sẵn vào bộ nhớ 1020, nhưng mô đun chương trình 1025 cũng có thể được tạo cấu hình trên thiết bị lưu trữ 1035 để sau đó nạp vào bộ nhớ 1020. Thiết bị lưu trữ 1035 còn là vật ghi đọc được bằng máy tính phi nhất thời được mã hóa với chương trình máy tính, và có thể là thiết bị lưu trữ thông thường bất kỳ để lưu trữ mô đun chương trình 1025 trong đó. Ví dụ về thiết bị lưu trữ 1035 bao gồm đĩa mềm, đĩa compac, băng từ, bộ nhớ chỉ đọc, phương tiện nhớ quang, ổ nhớ tác động nhanh bus nối tiếp đa năng (USB - Universal Serial Bus), đĩa đa năng số, hoặc ổ đĩa zip. Thiết bị lưu trữ 1035 còn có thể là bộ nhớ truy nhập ngẫu nhiên, hoặc kiểu lưu trữ điện tử khác, nằm trong hệ thống lưu trữ từ xa và được ghép nối với máy tính 1005 qua mạng 1030.

Các lợi ích kỹ thuật của quy trình 100 và hệ thống 1000 bao gồm độ chính xác của các đầu ra được cải thiện và khả năng mở rộng hoạt động gia tăng, cũng như đề xuất các quy trình học tập vòng lặp đóng để cho phép quy trình 100 thi hành với độ chính xác gia tăng theo thời gian.

Các kỹ thuật được mô tả ở đây để làm ví dụ và không được hiểu là có ý nghĩa hạn chế cụ thể nào đối với sáng chế. Cần phải hiểu rằng, các phương án thay thế, các tổ hợp và các sửa đổi có thể được tìm ra bởi người có hiểu biết trung bình trong lĩnh vực kỹ thuật này. Ví dụ, các bước gắn với các quy trình được mô tả ở đây có thể được thực hiện theo thứ tự bất kỳ, trừ khi có quy định khác hoặc được chỉ rõ bởi chính các bước này. Sáng chế dự định bao hàm tất cả các phương án thay thế, sửa đổi và cải biến như vậy, chúng đều trong phạm vi của yêu cầu bảo hộ kèm

theo.

Thuật ngữ "bao gồm " hoặc "gồm có" được hiểu là xác định sự có mặt của các dấu hiệu, số nguyên, bước hoặc thành phần đã nêu, nhưng không loại trừ sự có mặt của một hoặc nhiều dấu hiệu, số nguyên, bước hoặc thành phần khác hoặc các tổ hợp của chúng.

YÊU CẦU BẢO HỘ

1. Phương pháp chuyển đổi ngôn ngữ bao gồm các bước sau:

nhận dữ liệu nhập mà bao gồm (a) chuỗi các ký tự trong ngôn ngữ thứ nhất, và (b) dữ liệu ngữ cảnh ngữ nghĩa liên quan đến nguồn dữ liệu nhập nêu trên;

phân tích cú pháp chuỗi ký tự thành các tự vị của nó;

tạo lập mẫu các ký tự để biểu diễn dạng trừu tượng của các tự vị nêu trên;

phân tích dữ liệu ngữ cảnh ngữ nghĩa nêu trên và mẫu các ký tự nêu trên theo các quy tắc, để tạo ra sự chuyển đổi liên ngôn ngữ có thể của mẫu các ký tự nêu trên;

chuyển đổi chuỗi các ký tự nêu trên từ ngôn ngữ thứ nhất thành ngôn ngữ thứ hai phù hợp với chuyển đổi liên ngôn ngữ có thể, theo đó thu được việc chuyển đổi;

phân tích chỉ số hiệu quả về việc chuyển đổi; và

cập nhật các quy tắc nêu trên dựa trên chỉ số hiệu quả nêu trên.

2. Phương pháp chuyển đổi ngôn ngữ theo điểm 1, trong đó mẫu ký tự bao gồm nhóm ký tự tương ứng với một tự vị trong các tự vị này và được ánh xạ vào nguồn dữ liệu.

3. Phương pháp chuyển đổi ngôn ngữ theo điểm 2, trong đó phương pháp này còn bao gồm bước truy tìm từ nguồn nêu trên, thông tin về tự vị này.

4. Phương pháp chuyển đổi ngôn ngữ theo điểm 1, trong đó:

mẫu ký tự bao gồm dãy gồm nhóm ký tự thứ nhất và nhóm ký tự thứ hai, nhóm ký tự thứ nhất tương ứng với tự vị thứ nhất trong các tự vị này, nhóm ký tự thứ hai tương ứng với tự vị thứ hai trong các tự vị này, và phương pháp này còn bao gồm bước:

chọn, dựa vào dãy nêu trên, một trong số các quy trình; và

thi hành quy trình này trên chuỗi ký tự.

5. Phương pháp chuyển đổi ngôn ngữ theo điểm 4, trong đó quy trình này mang lại sự kết hợp giữa dịch và chuyển đổi liên ngôn ngữ của chuỗi ký tự.

6. Phương pháp chuyển đổi ngôn ngữ theo điểm 1,

trong đó bước phân tích dữ liệu ngữ cảnh ngữ nghĩa và mẫu các ký tự nêu trên cũng được thực hiện theo các ưu tiên và thuộc tính, và thu được (a) các nghĩa ứng tuyển hoặc các chiến lược bố trí liên ngôn ngữ ứng tuyển của các mẫu ký tự nêu trên,

và (b) thông tin để thông báo việc lặp lại trong tương lai phương pháp nêu trên, và trong đó phương pháp này còn bao gồm bước, sau khi phân tích dữ liệu ngữ cảnh ngữ nghĩa nêu trên và mẫu các ký tự nêu trên, và trước khi chuyển đổi:

chọn chiến lược bố trí liên ngôn ngữ tối ưu cho các phần cấu thành nên mẫu các ký tự liên quan đến (a) việc phân tích dữ liệu ngữ cảnh ngữ nghĩa và mẫu các ký tự nêu trên, và (b) các quy tắc, độ ưu tiên và thuộc tính nêu trên, và được thông báo bằng cách học từ việc lặp lại trước đó phương pháp nêu trên.

7. Hệ thống chuyển đổi ngôn ngữ bao gồm:

bộ xử lý; và

bộ nhớ nối truyền thông với bộ xử lý và chứa các lệnh có thể đọc được bằng bộ xử lý để lệnh cho bộ xử lý thực hiện các thao tác bao gồm:

nhận dữ liệu nhập mà bao gồm (a) chuỗi các ký tự trong ngôn ngữ thứ nhất, và (b) dữ liệu ngữ cảnh ngữ nghĩa liên quan đến nguồn dữ liệu nhập nêu trên;

phân tích cú pháp chuỗi ký tự thành các tự vị của nó;

tạo lập mẫu các ký tự để biểu diễn dạng trừu tượng của các tự vị nêu trên;

phân tích dữ liệu ngữ cảnh ngữ nghĩa nêu trên và mẫu các ký tự nêu trên theo các quy tắc, để tạo ra sự chuyển đổi liên ngôn ngữ có thể của mẫu các ký tự nêu trên;

chuyển đổi chuỗi các ký tự nêu trên từ ngôn ngữ thứ nhất thành ngôn ngữ thứ hai phù hợp với chuyển đổi liên ngôn ngữ có thể, theo đó thu được việc chuyển đổi;

phân tích chỉ số hiệu quả về việc chuyển đổi; và

cập nhật các quy tắc nêu trên dựa trên chỉ số hiệu quả nêu trên.

8. Hệ thống chuyển đổi ngôn ngữ theo điểm 7, trong đó mẫu ký tự bao gồm nhóm ký tự tương ứng với một tự vị trong các tự vị này và được ánh xạ vào nguồn dữ liệu.

9. Hệ thống chuyển đổi ngôn ngữ theo điểm 8, trong đó các lệnh còn lệnh cho bộ xử lý thực hiện thao tác:

truy tìm từ nguồn dữ liệu, thông tin về tự vị này.

10. Hệ thống chuyển đổi ngôn ngữ theo điểm 7, trong đó:

mẫu ký tự bao gồm dãy gồm nhóm ký tự thứ nhất và nhóm ký tự thứ hai,

nhóm ký tự thứ nhất tương ứng với tự vị thứ nhất trong các tự vị này,

nhóm ký tự thứ hai tương ứng với tự vị thứ hai trong các tự vị này, và

các lệnh còn lệnh cho bộ xử lý thực hiện các thao tác bao gồm:
 chọn, dựa vào dãy nêu trên, một trong số các quy trình; và
 thi hành quy trình này trên chuỗi ký tự.

11. Hệ thống chuyển đổi ngôn ngữ theo điểm 10, trong đó quy trình này mang lại sự kết hợp giữa việc dịch và chuyển đổi liên ngôn ngữ cho chuỗi ký tự này.

12. Hệ thống chuyển đổi ngôn ngữ theo điểm 7,

trong đó bước phân tích dữ liệu ngữ cảnh ngữ nghĩa và mẫu các ký tự nêu trên cũng được thực hiện theo các ưu tiên và thuộc tính, và thu được (a) các nghĩa ứng tuyển hoặc các chiến lược bố trí liên ngôn ngữ ứng tuyển của các mẫu ký tự nêu trên, và (b) thông tin để thông báo việc lặp lại trong tương lai phương pháp nêu trên, và

trong đó phương pháp này còn bao gồm bước, sau khi phân tích dữ liệu ngữ cảnh ngữ nghĩa nêu trên và mẫu các ký tự nêu trên, và trước khi chuyển đổi:

chọn chiến lược bố trí liên ngôn ngữ tối ưu cho các phần cấu thành nên mẫu các ký tự liên quan đến (a) việc phân tích dữ liệu ngữ cảnh ngữ nghĩa và mẫu các ký tự nêu trên, và (b) các quy tắc, độ ưu tiên và thuộc tính nêu trên, và được thông báo bằng cách học từ việc lặp lại trước đó phương pháp nêu trên.

13. Thiết bị lưu trữ chứa các lệnh đọc được bằng bộ xử lý để lệnh cho bộ xử lý thực hiện các thao tác bao gồm:

nhận dữ liệu nhập mà bao gồm (a) chuỗi các ký tự trong ngôn ngữ thứ nhất, và (b) dữ liệu ngữ cảnh ngữ nghĩa liên quan đến nguồn dữ liệu nhập nêu trên;

phân tích cú pháp chuỗi ký tự thành các tự vị của nó;

tạo lập mẫu các ký tự để biểu diễn dạng trừu tượng của các tự vị nêu trên;

phân tích dữ liệu ngữ cảnh ngữ nghĩa nêu trên và mẫu các ký tự nêu trên theo các quy tắc, để tạo ra sự chuyển đổi liên ngôn ngữ có thể của mẫu các ký tự nêu trên;

chuyển đổi chuỗi các ký tự nêu trên từ ngôn ngữ thứ nhất thành ngôn ngữ thứ hai phù hợp với chuyển đổi liên ngôn ngữ có thể, theo đó thu được việc chuyển đổi;

phân tích chỉ số hiệu quả về việc chuyển đổi; và

cập nhật các quy tắc nêu trên dựa trên chỉ số hiệu quả nêu trên.

14. Thiết bị lưu trữ theo điểm 13, trong đó mẫu ký tự bao gồm nhóm ký tự tương ứng với một tự vị trong các tự vị này và được ánh xạ vào nguồn dữ liệu.

15. Thiết bị lưu trữ theo điểm 14, trong đó các lệnh còn lệnh cho bộ xử lý thực hiện thao tác:

truy tìm từ nguồn dữ liệu, thông tin về tự vị này.

16. Thiết bị lưu trữ theo điểm 13, trong đó:

mẫu ký tự bao gồm dãy gồm nhóm ký tự thứ nhất và nhóm ký tự thứ hai, nhóm ký tự thứ nhất tương ứng với tự vị thứ nhất trong các tự vị này, nhóm ký tự thứ hai tương ứng với tự vị thứ hai trong các tự vị này, và các lệnh còn lệnh cho bộ xử lý thực hiện các thao tác bao gồm:

chọn, dựa vào dãy nêu trên, một trong số các quy trình; và

thi hành quy trình này trên chuỗi ký tự.

17. Thiết bị lưu trữ theo điểm 16, trong đó quy trình này thực hiện kết hợp giữa dịch và chuyển đổi liên ngôn ngữ cho chuỗi ký tự này.

18. Thiết bị lưu trữ theo điểm 13,

trong đó bước phân tích dữ liệu ngữ cảnh ngữ nghĩa và mẫu các ký tự nêu trên cũng được thực hiện theo các ưu tiên và thuộc tính, và thu được (a) các nghĩa ứng tuyển hoặc các chiến lược bố trí liên ngôn ngữ ứng tuyển của các mẫu ký tự nêu trên, và (b) thông tin để thông báo việc lặp lại trong tương lai phương pháp nêu trên, và

trong đó phương pháp này còn bao gồm bước, sau khi phân tích dữ liệu ngữ cảnh ngữ nghĩa nêu trên và mẫu các ký tự nêu trên, và trước khi chuyển đổi:

chọn chiến lược bố trí liên ngôn ngữ tối ưu cho các phần cấu thành nên mẫu các ký tự liên quan đến (a) việc phân tích dữ liệu ngữ cảnh ngữ nghĩa và mẫu các ký tự nêu trên, và (b) các quy tắc, độ ưu tiên và thuộc tính nêu trên, và được thông báo bằng cách học từ việc lặp lại trước đó phương pháp nêu trên.

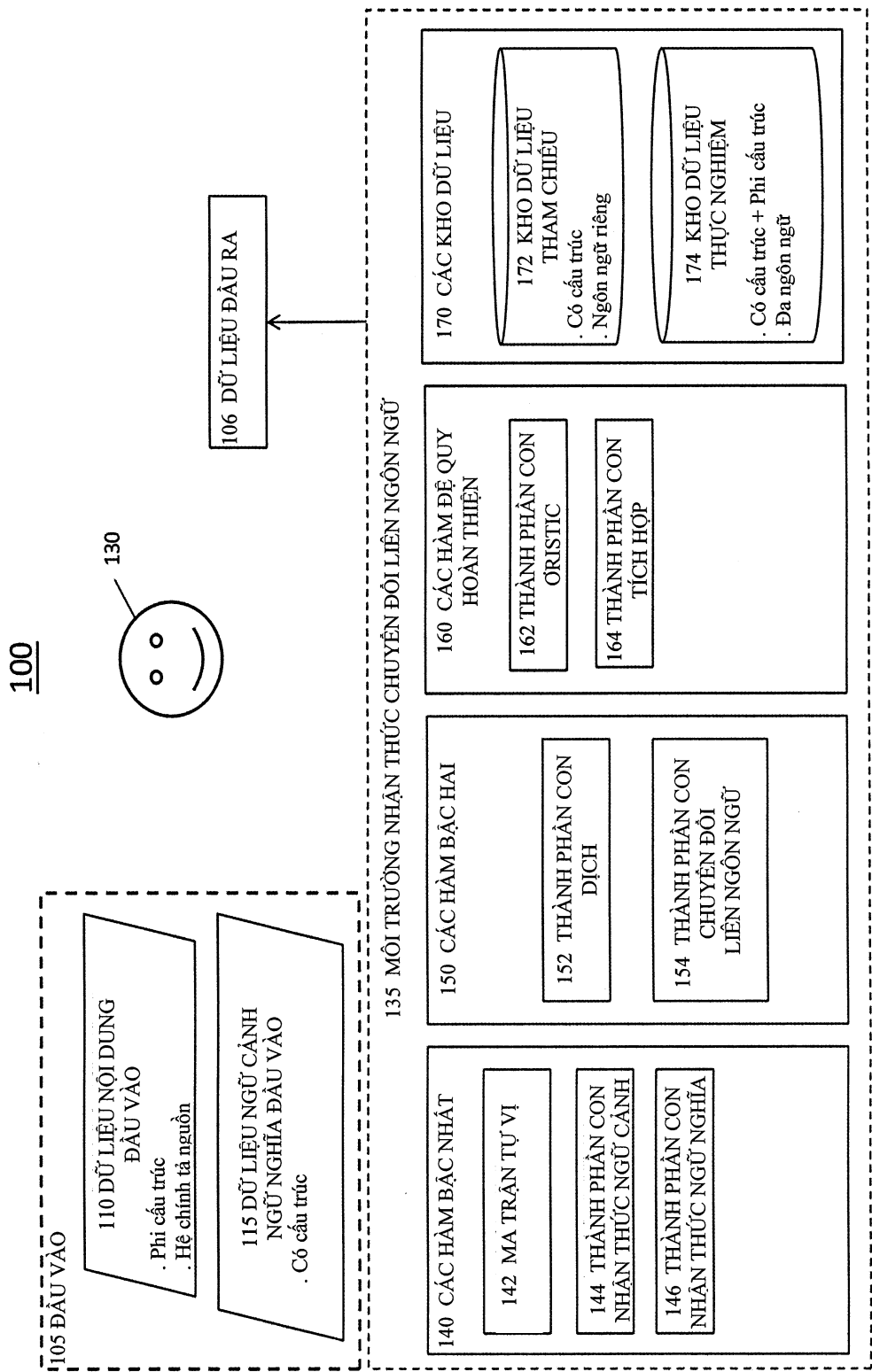


FIG. 1

172 KHO DỮ LIỆU THAM CHIẾU

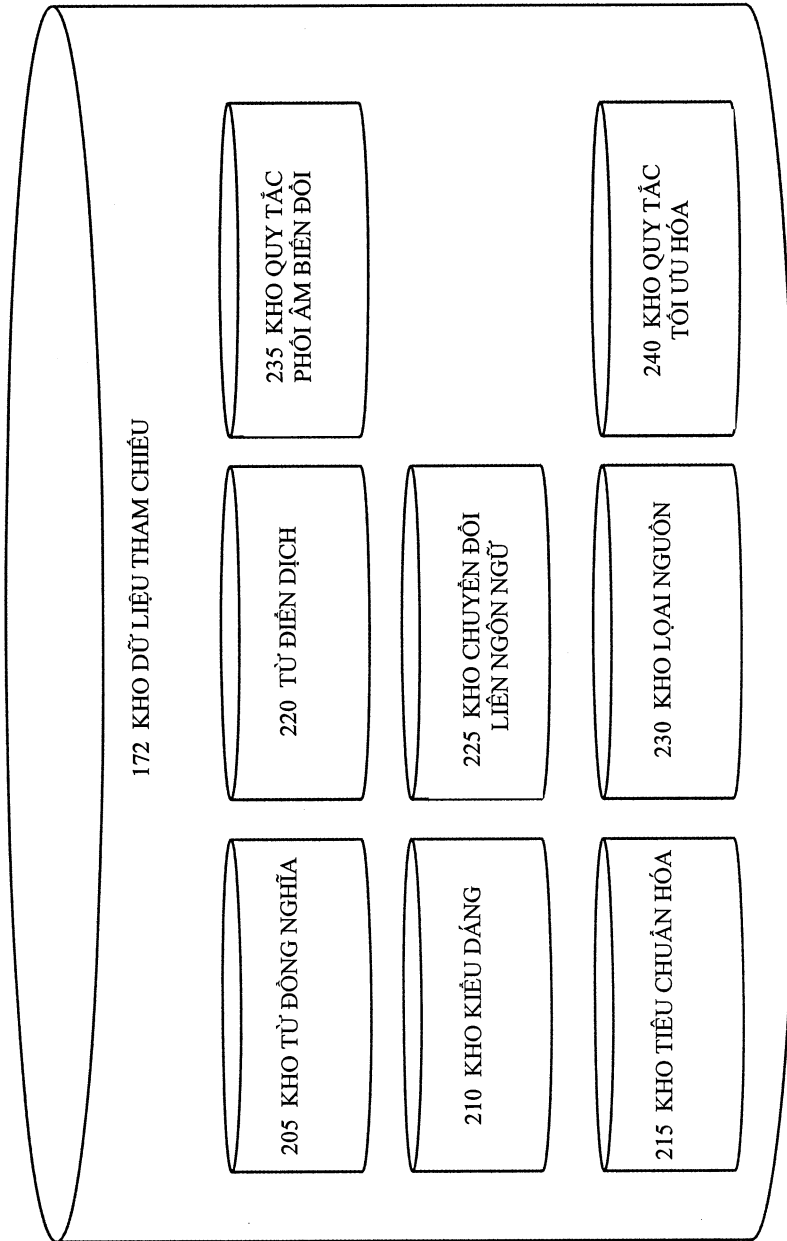


FIG. 2

174 KHO DỮ LIỆU THỰC NGHIỆM

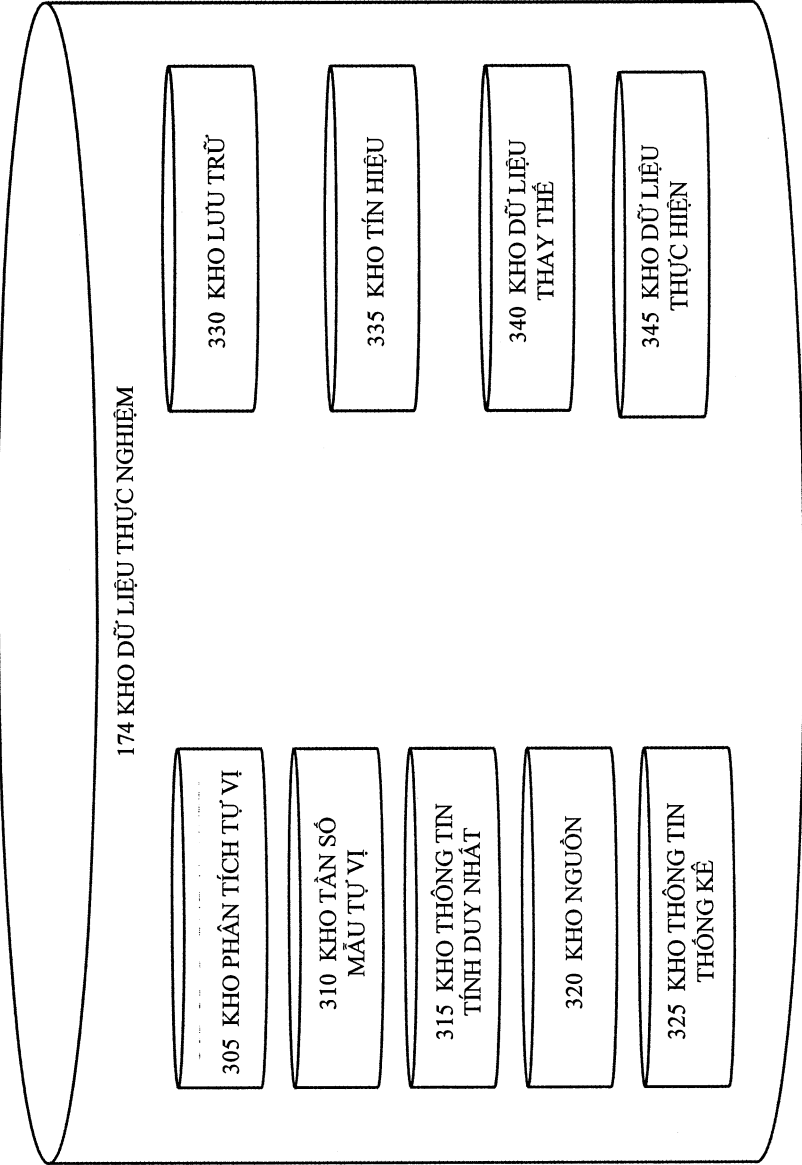


FIG. 3

140 CÁC HÀM BẬC MỘT

142 - MÀ TRẬN TỰ VỊ
405 - BỘ PHÂN TÍCH CÚ PHÁP & PHÂN TÍCH TỰ VỊ
410 - BỘ ÁNH XẠ MẪU TỰ VỊ
415 - BỘ LẬP MÔ HÌNH MẪU TỰ VỊ

144 - THÀNH PHẦN CON NHẬN THỨC NGỮ CẢNH
420 - BỘ PHÂN TÍCH NGỮ CẢNH
425 - BỘ PHÂN LOẠI NGUỒN
430 - BỘ PHÂN TÍCH TÍNH DUY NHẤT
435 - BỘ TẠO DỮ LIỆU THAY THẾ

146 - THÀNH PHẦN CON NHẬN THỨC NGỮ NGHĨA
440 - QUY TRÌNH TIÊU CHUẨN HÓA
445 - BỘ PHÂN TÍCH TỪ ĐỒNG NGHĨA
450 - BỘ PHÂN TÍCH KIỂU DẠNG

150 CÁC HÀM BẬC HAI

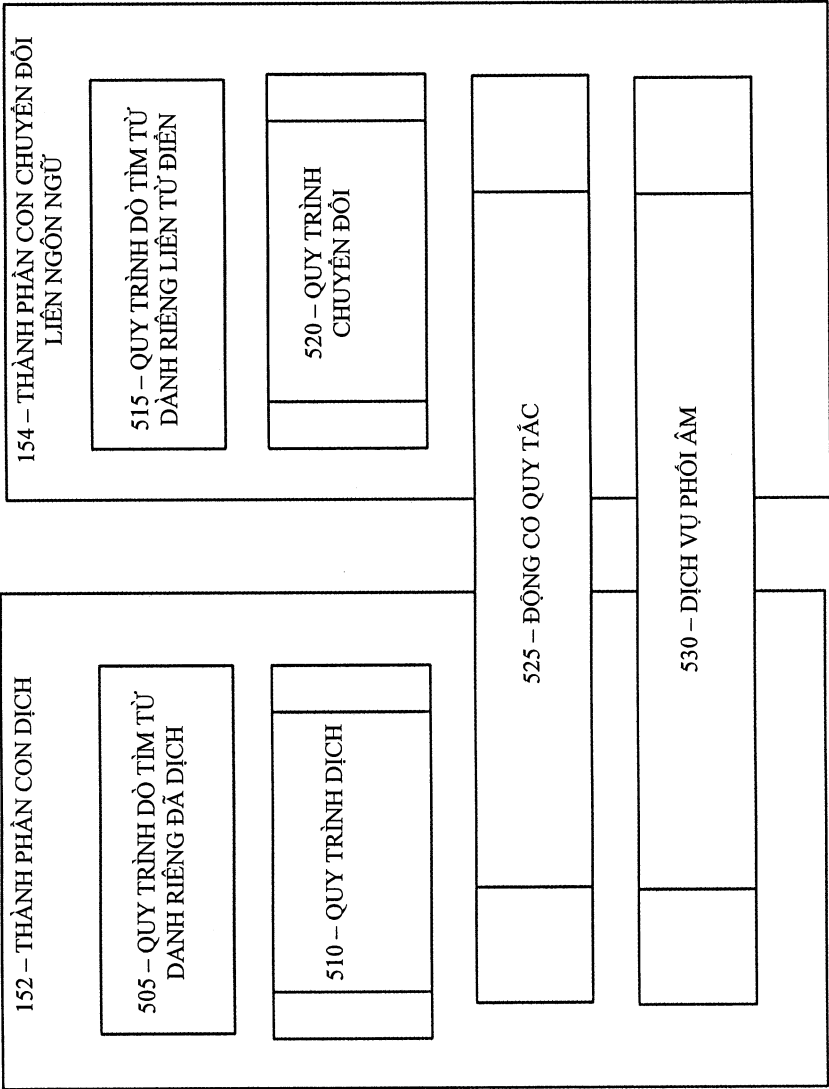


FIG. 5

160 CÁC HÀM ĐỀ QUY HOÀN THIÊN

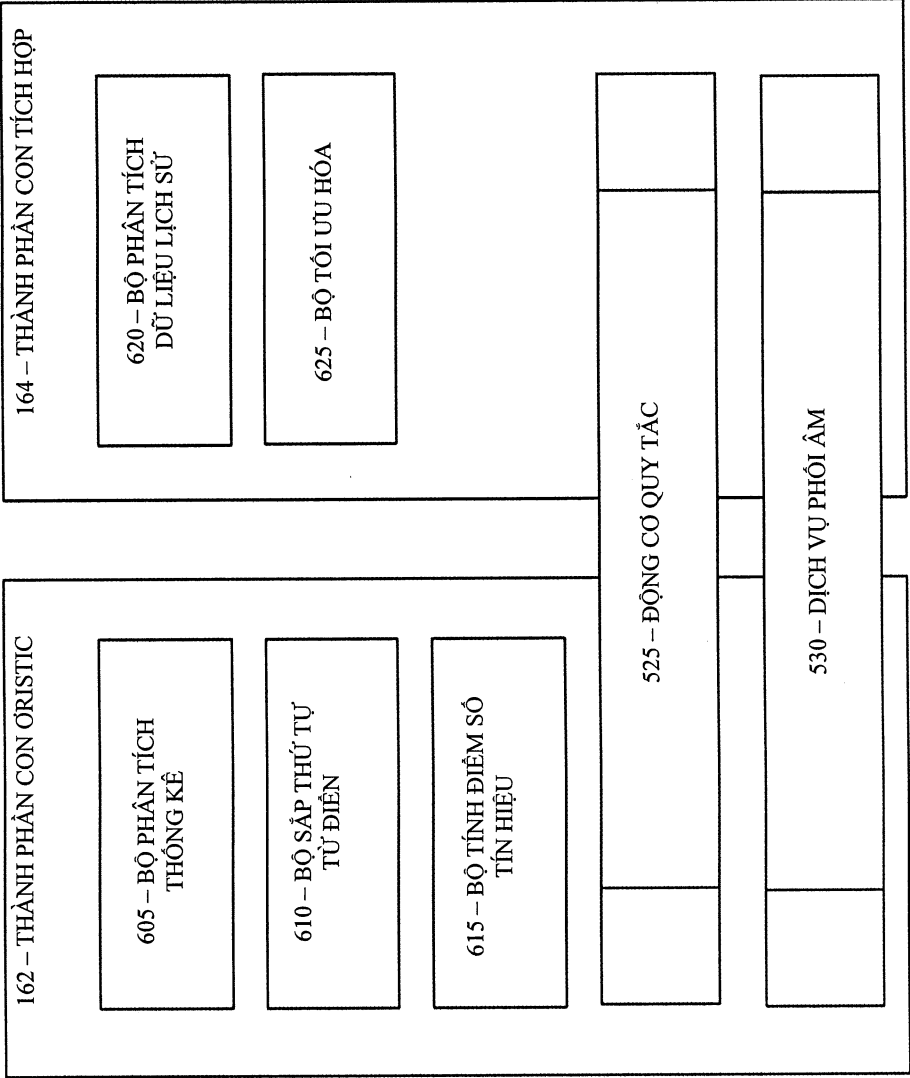
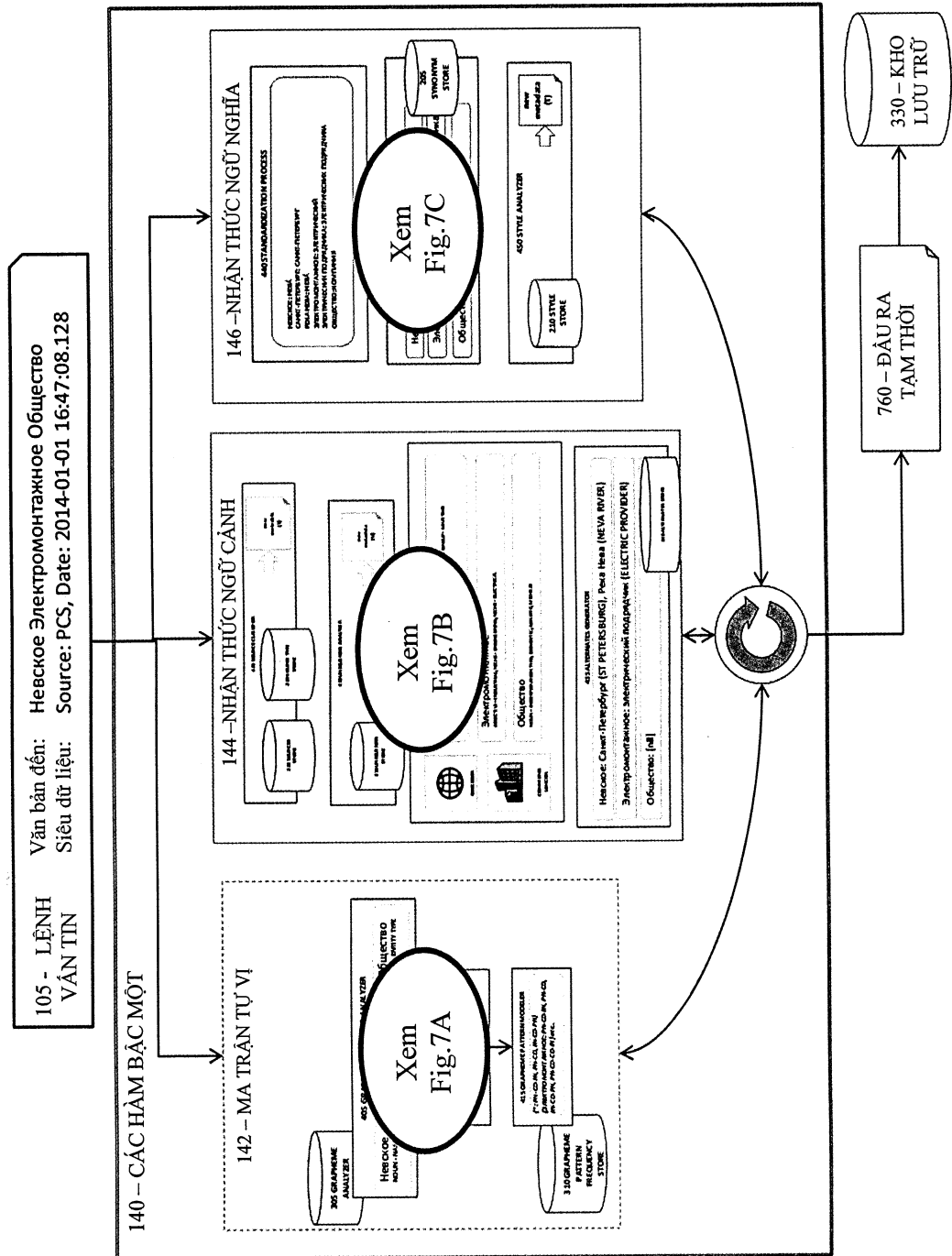


FIG. 6



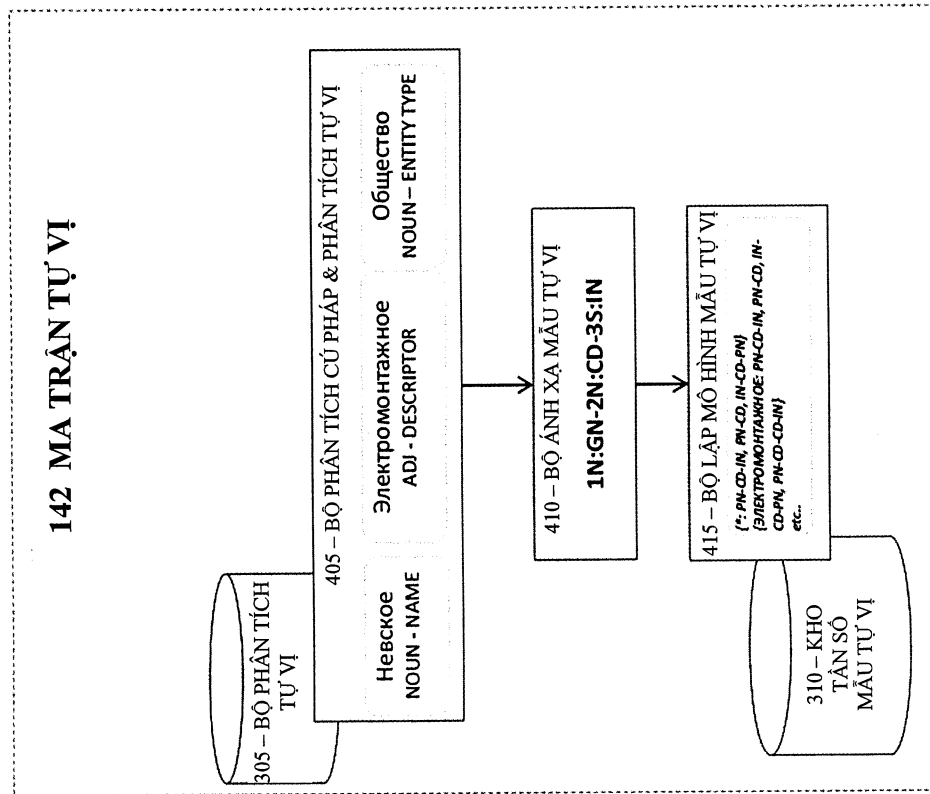


FIG. 7A

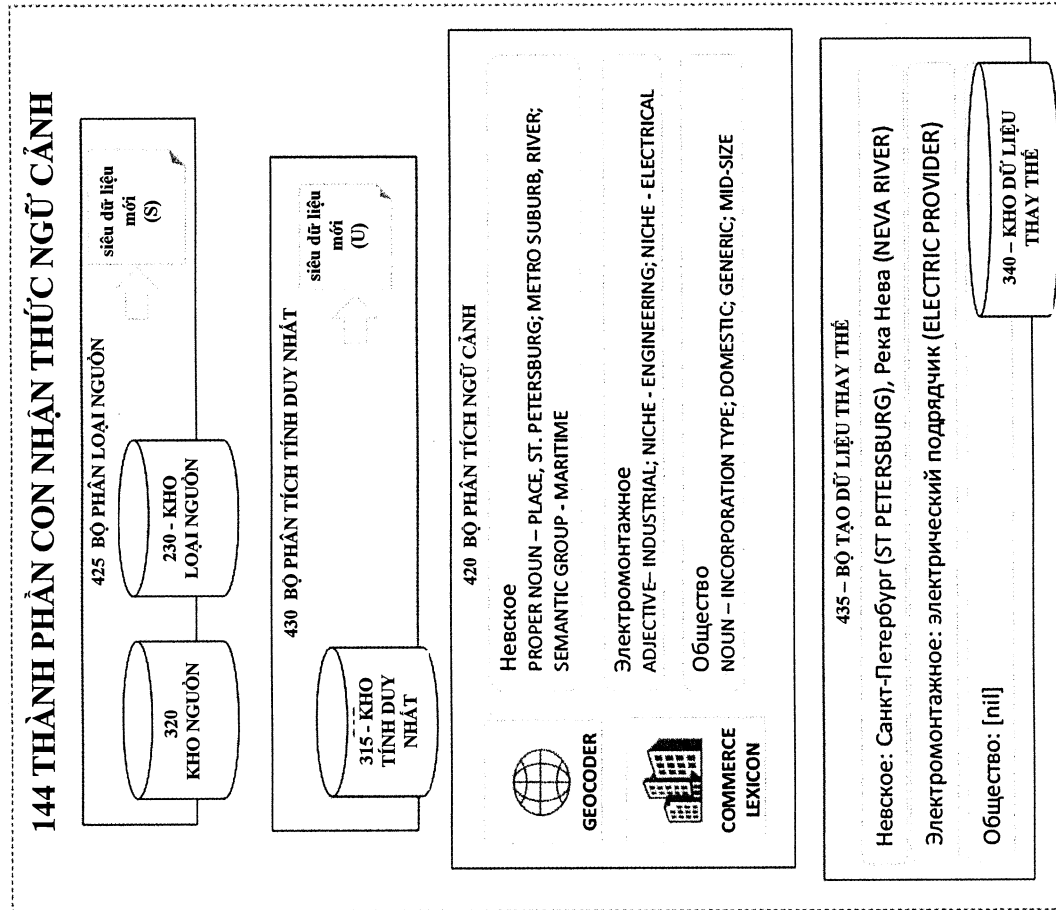


FIG. 7B

146 THÀNH PHẦN CON NHẬN THỨC NGŨ NGHĨA

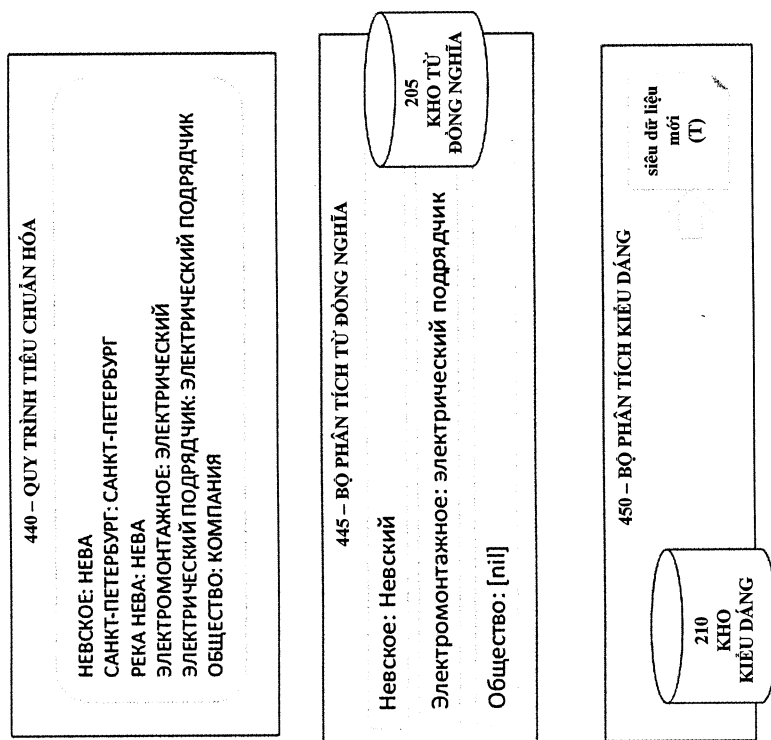


FIG. 7C

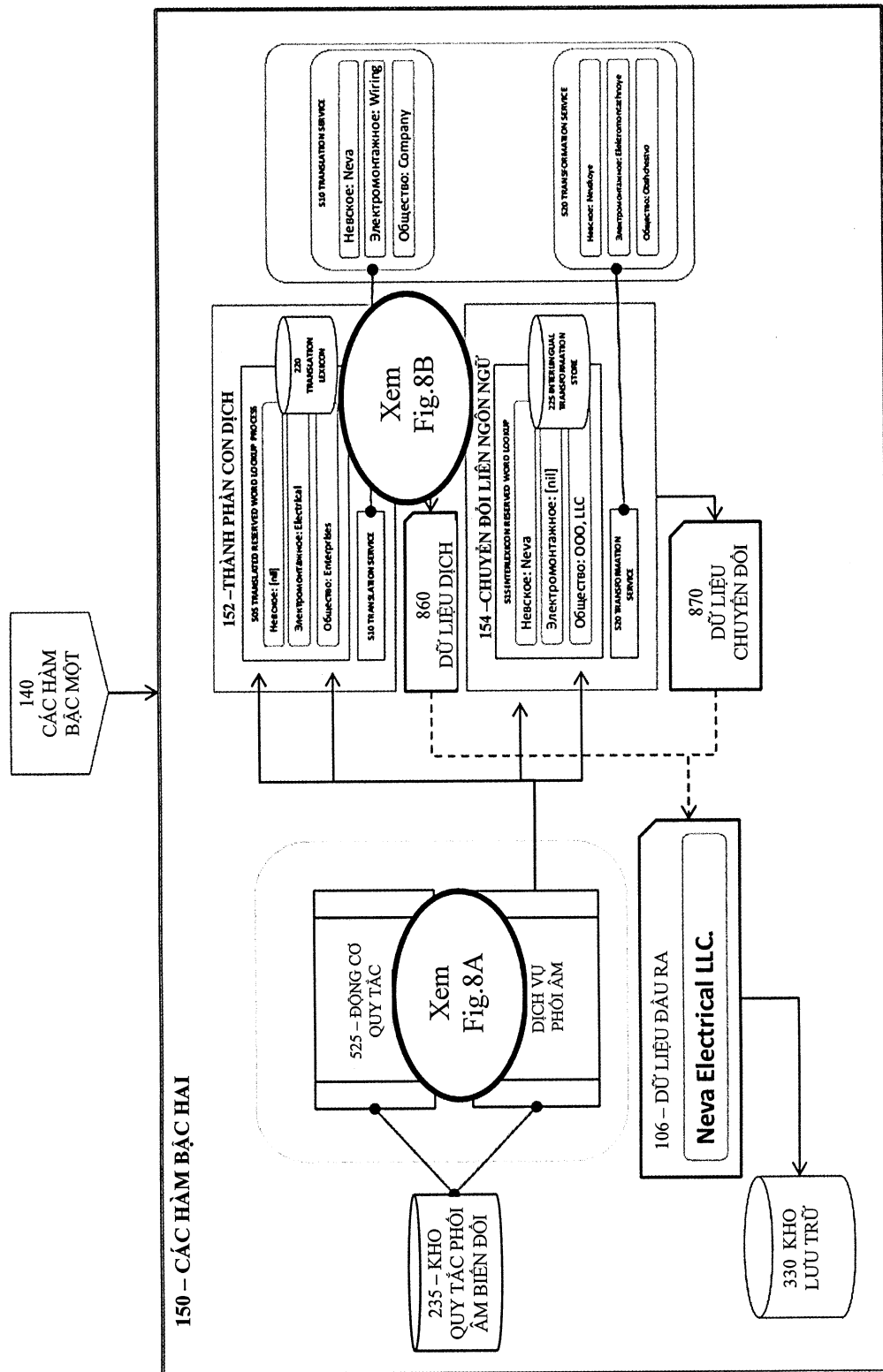


FIG. 8

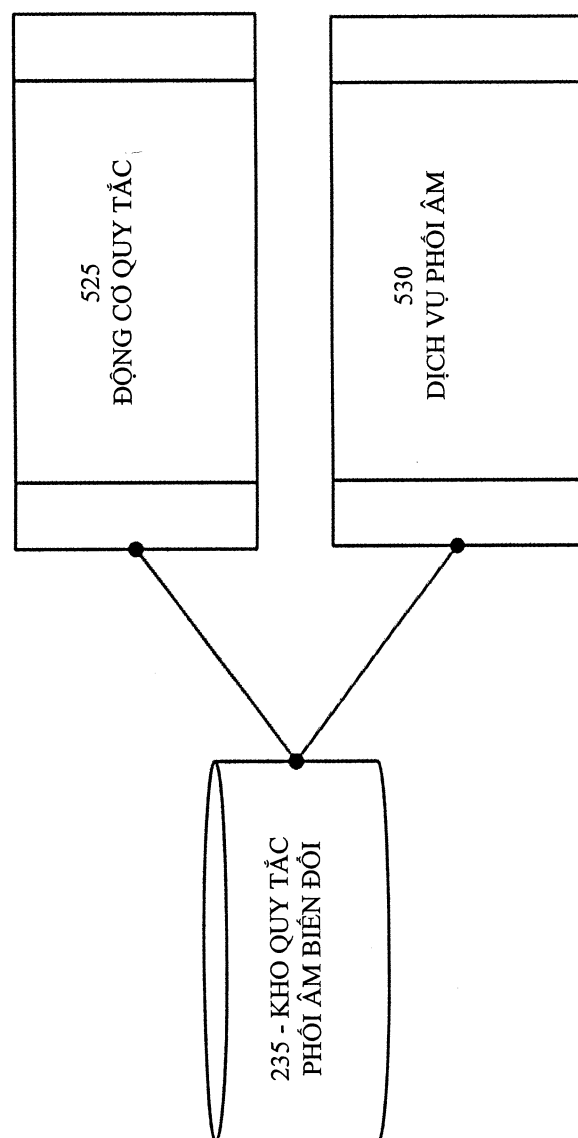


FIG. 8A

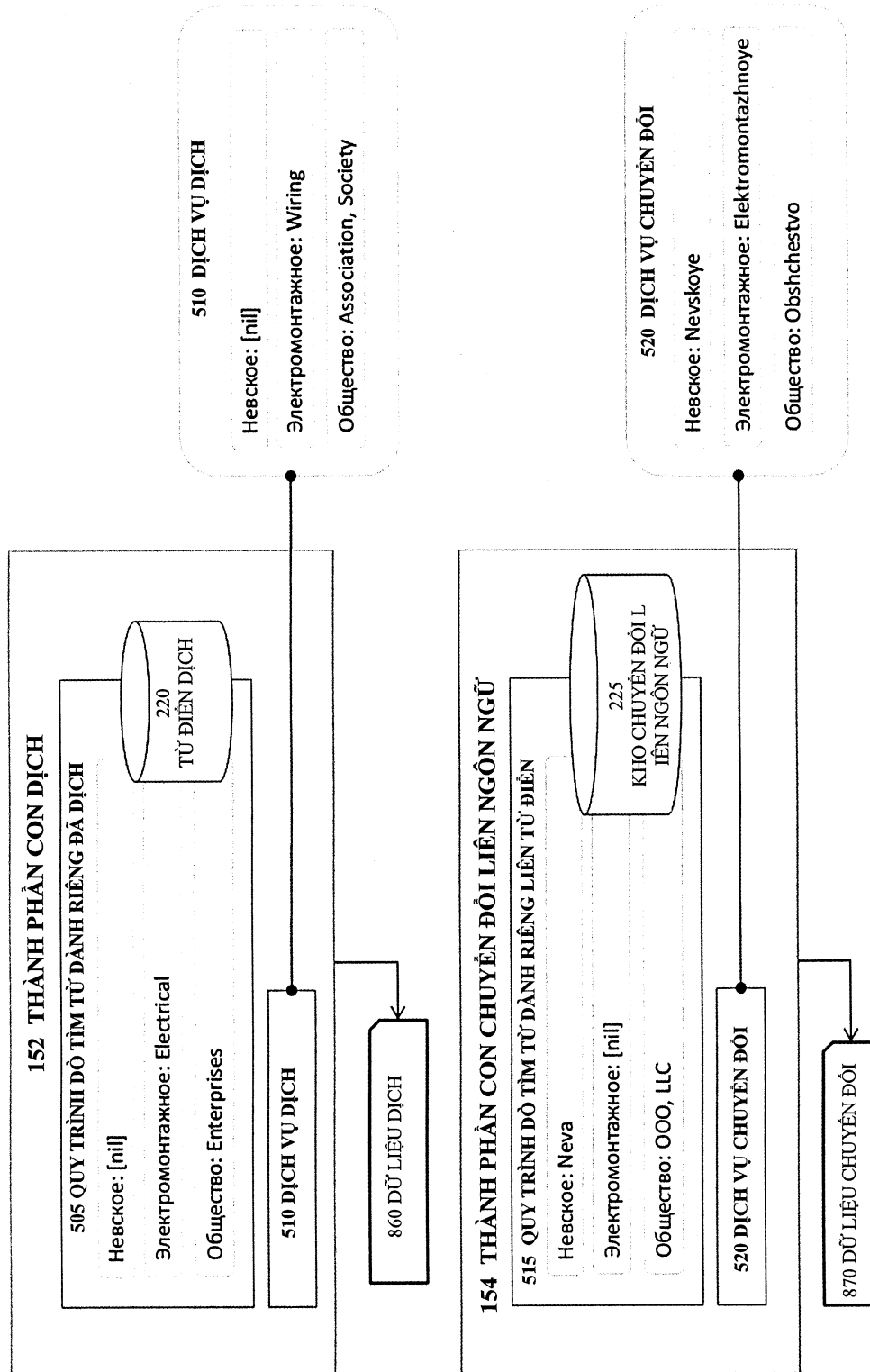


FIG. 8B

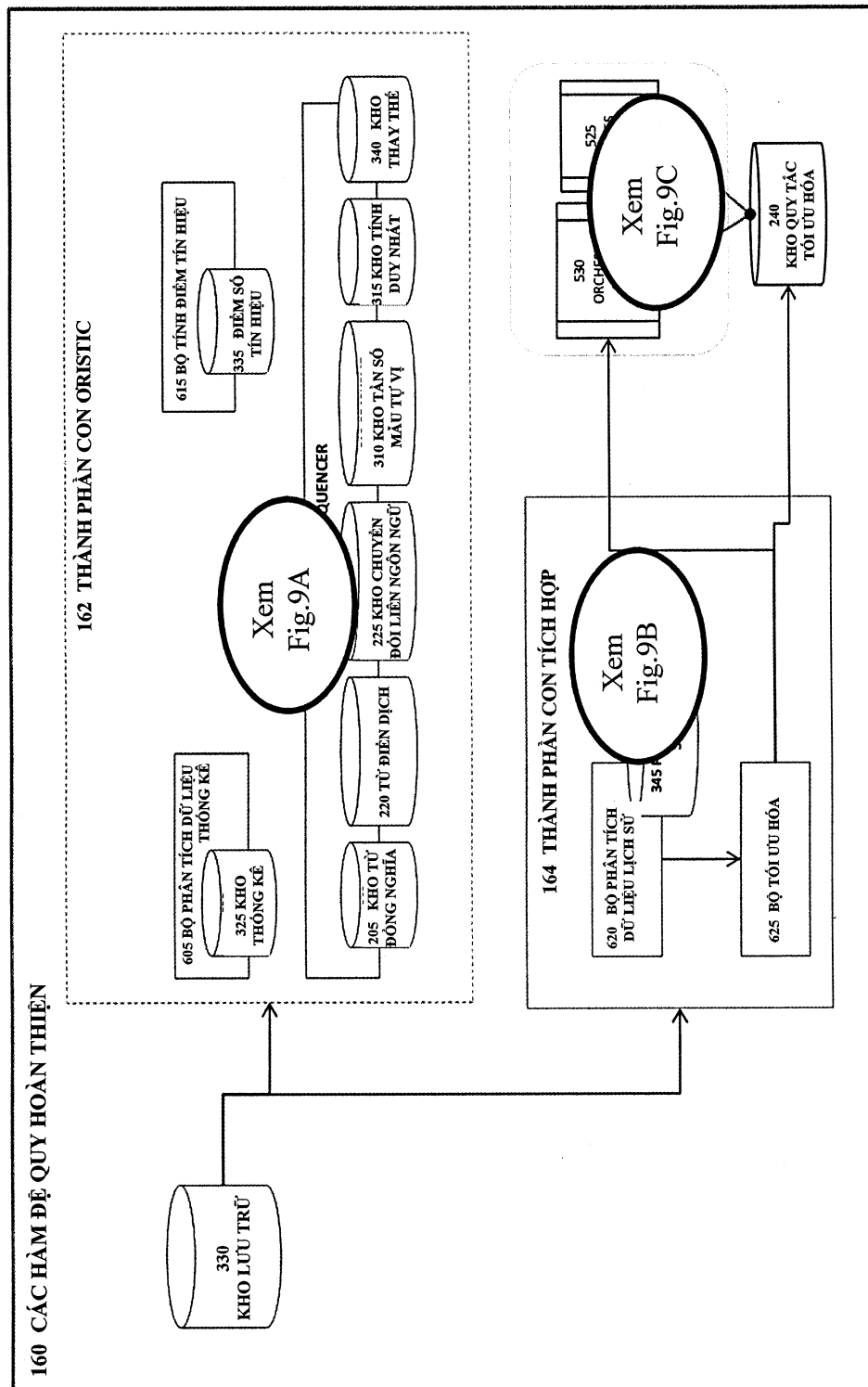


FIG. 9

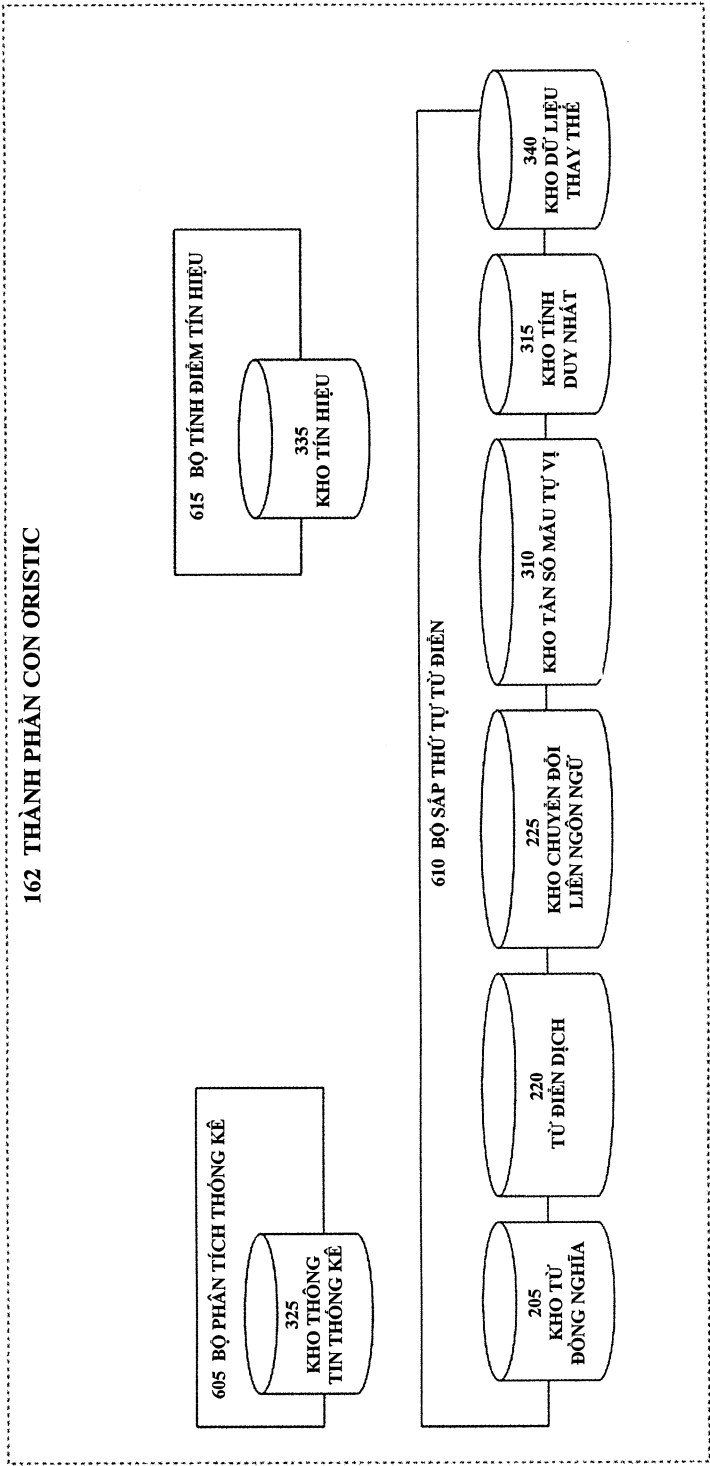


FIG. 9A

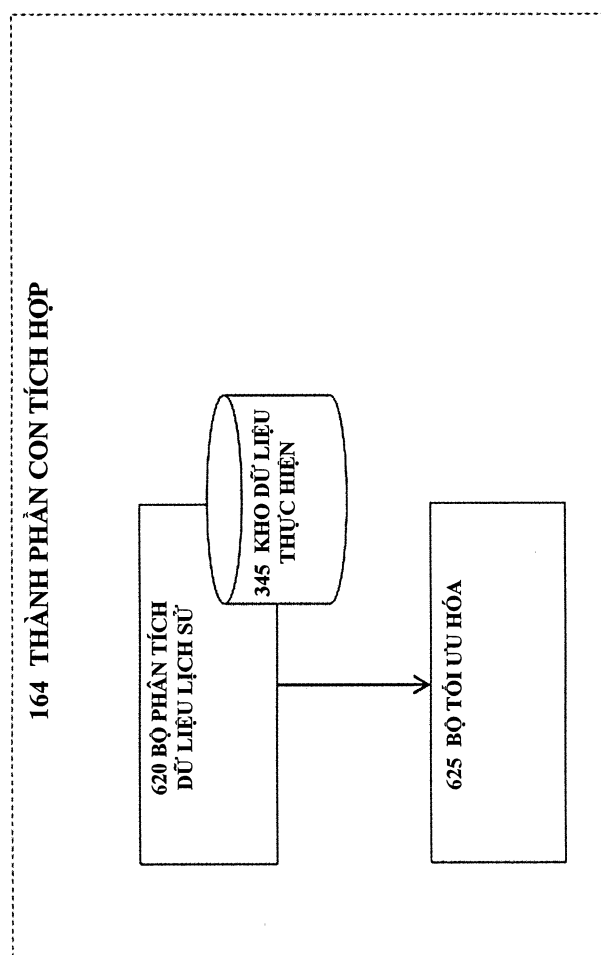


FIG. 9B

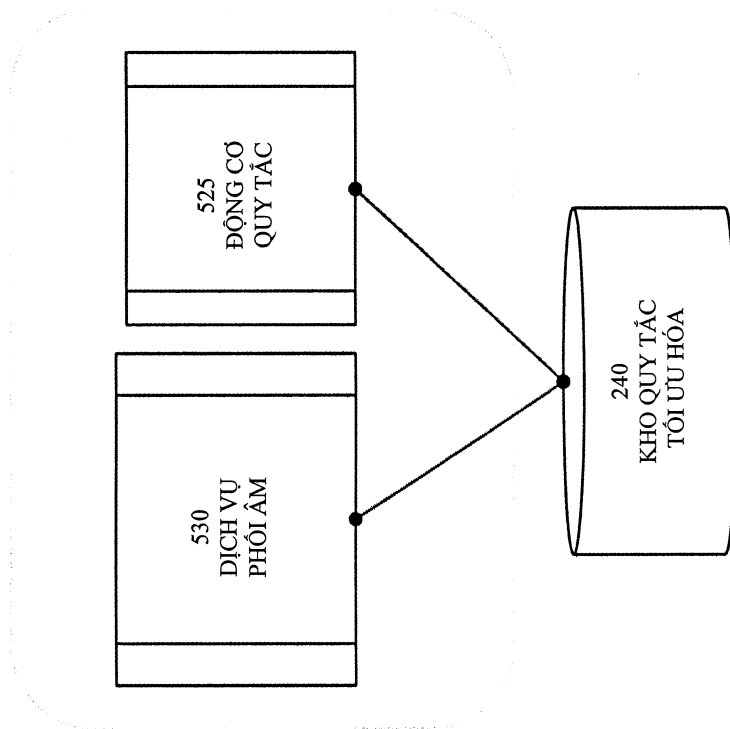


FIG. 9C

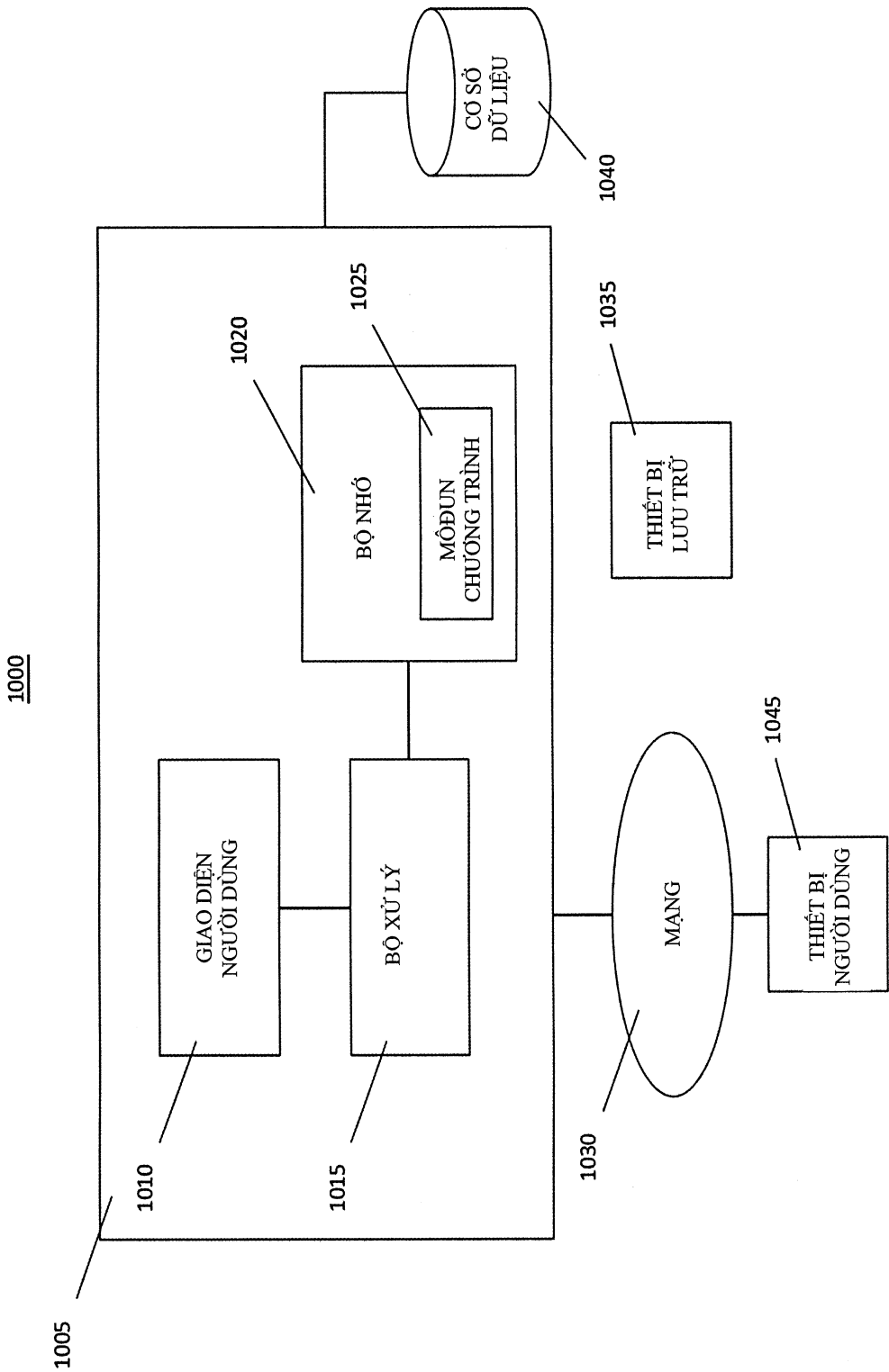


FIG. 10