**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**------**oOo**------**

**ĐỒ ÁN MÔN HỌC**

Môn học: Biểu diễn tri thức và suy luận

*Học kỳ I - Năm Học 2022-2023*

**ĐỀ TÀI:**

**XÂY DỰNG HỆ THỐNG HỖ TRỢ TÌM KIẾM CÁC CĂN HỘ CHUNG CƯ THEO CÁC THÔNG SỐ DO NGƯỜI DÙNG NHẬP VÀO**

**Giảng Viên: NGUYỄN ĐÌNH HIỂN**

**NHÓM THỰC HIỆN:**

**1**. DƯƠNG NGỌC HÙNG (*18520792)*

**2***.* NGUYỄN NHẬT HUY (*19520112)*

**3***.* TÔ THANH HIỀN *(19521490)*

**MỤC LỤC**

[TỔNG QUAN 3](#_Toc122346329)

[1. Giới thiệu bài toán: 3](#_Toc122346330)

[2. Nội dung báo cáo đồ án: 3](#_Toc122346331)

[CƠ SỞ LÝ THUYẾT 4](#_Toc122346332)

[1. Biểu diễn tri thức 4](#_Toc122346333)

[2. Biễu diễn tri thức dựa trên luật: 4](#_Toc122346334)

[a) Ưu điểm của phương pháp biểu diễn tri thức dựa trên luật: 5](#_Toc122346335)

[b) Nhược điểm của phương pháp biểu diễn tri thức dựa trên luật. 5](#_Toc122346336)

[3. Suy diễn tiến: 6](#_Toc122346337)

[a) Ưu điểm suy diễn tiến: 7](#_Toc122346338)

[b) Nhược điểm suy diễn tiến: 7](#_Toc122346339)

[4. Suy diễn lùi: 7](#_Toc122346340)

[5. Biểu diễn tri thức dựa trên mạng ngữ nghĩa: 9](#_Toc122346341)

[a) Đặc điểm: 9](#_Toc122346342)

[b) Ưu điểm: 9](#_Toc122346343)

[c) Nhược điểm: 9](#_Toc122346344)

[d) Biểu diễn tri thức bằng mạng ngữ nghĩa: 10](#_Toc122346345)

[GIẢI QUYẾT BÀI TOÁN 12](#_Toc122346346)

[1. Đặt vấn đề: 12](#_Toc122346347)

[2. Thu thập tri thức: 12](#_Toc122346348)

[3. Nguồn tri thức - Bảng thu thập tri thức: 13](#_Toc122346349)

[4. Biểu diễn tri thức: 18](#_Toc122346350)

[a) Gợi ý các sản phẩm khác theo các tiêu chí do hệ thống tự định trước: 20](#_Toc122346351)

[5. Thuật giải suy diễn tiến: 21](#_Toc122346352)

[CÀI ĐẶT – DEMO CHƯƠNG TRÌNH 24](#_Toc122346353)

[1. Cài đặt 24](#_Toc122346354)

[2. Demo 24](#_Toc122346355)

[ĐÁNH GIÁ – KẾT LUẬN 25](#_Toc122346356)

[1. Đánh giá 25](#_Toc122346357)

[2. Kết luận 25](#_Toc122346358)

[TÀI LIỆU THAM KHẢO 26](#_Toc122346359)

TỔNG QUAN

1. Giới thiệu bài toán:

Nhu cầu tìm kiếm căn hộ của người dân ở TP.HCM tăng cao trong thời gian qua [1] đi kèm với đa dạng mô hình căn hộ cho từng phân khúc thị trường đã dẫn đến việc xây dựng một hệ thống hỗ trợ tìm kiếm các căn hộ chung cư theo các thông tin cá nhân cho từng người dùng cụ thể là cần thiết. Tuy nhiên, việc thị trường căn hộ không tồn tại kết quả khớp nhất với nhu cầu của 1 người dùng là rất dễ xảy ra do nguồn cung bất động sản nói chung và căn hộ nói riêng ở TP.HCM đang dần khan hiếm và thu nhập của người dân không theo kịp đà tăng giá của bất động sản [2]. Trong trường hợp này, hệ thống hỗ trợ tìm kiếm cần phải có chức năng gợi ý những sản phẩm hiện có tiệm cận nhất với nhu cầu của người dùng để có thể tăng tính trải nghiệm cho người dùng và khả năng giao dịch xảy ra cao hơn. Đây là cơ hội để ấp dụng các phương pháp biểu diễn tri thức và suy luận để giải quyết bài toán gợi ý mặt hàng dựa trên tri thức từ người dùng và từ nguồn chuyên gia. Đây cũng chính là nội dung của đồ án nhóm em - xây dựng hệ thống hỗ trợ tìm kiếm và gợi ý các căn hộ theo thông tin từ người dùng và từ nguồn chuyên gia.

1. Nội dung báo cáo đồ án:

Báo cáo đồ án của nhóm sẽ trình bày về:

* Cơ sở lý thuyết (Biểu diễn tri thức và suy diễn tiến/lùi, mạng ngữ nghĩa).
* Giải quyết bài toán (Thu thập tri thức, biễu diễn tri thức và suy diễn tiến).
* Cài đặt, demo chương trình.
* Kết luận và đánh giá.

CƠ SỞ LÝ THUYẾT

1. Biểu diễn tri thức

Biểu diễn tri thức (knowledge representation) là một lĩnh vực nghiên cứu quan trọng của Trí tuệ nhân tạo, nhằm phát triển các phương pháp, cách thức biểu diễn tri thức và các công cụ hỗ trợ việc biểu diễn tri thức. Có rất nhiều phương pháp biểu diễn tri thức và phải phụ thuộc vào tri thức cụ thể của từng bài toán ta mới chọn được một phương pháp biểu diễn hợp lý nhất. Các phương pháp biểu diễn tiêu biểu:

* Biểu diễn tri thức bằng luật sản xuất (Production rules)
* Biểu diễn tri thức bằng khung (Frames)
* Biểu diễn tri thức sử dụng mạng ngữ nghĩa(Semantic networks )
* Biểu diễn tri thức bằng ontology
* Các mô hình xác suất

Khi biểu diễn tri thức cho một bài toán, việc biểu diễn này được coi là tốt khi chúng thỏa mãn được 4 yếu tố sau

* Tính hoàn chỉnh (Completeness)
* Tính ngắn gọn (Conciseness)
* Tính hiệu quả về tính toán( Computational efficiency)
* Tính rõ ràng, dễ hiểu(Transparency)

Với bài toán cụ thể đang xét, hệ thống hỗ trợ tìm kiếm các căn hộ chung cư theo các thông số do người dùng nhập vào, việc **biểu diễn tri thức sử dụng luật** là phương pháp biểu diễn phù hợp nhất.

1. Biễu diễn tri thức dựa trên luật:

Phương pháp biểu diễn tri thức bằng luật sinh được phát minh bởi Newell và Simon trong lúc hai ông đang cố gắng xây dựng một hệ giải bài toán tổng quát. Đây là một kiểu biểu diễn tri thức có cấu trúc, và được sử dụng phổ biến nhất trong các hệ cơ sở tri thức. Ý tưởng cơ bản là tri thức có thể được cấu trúc bằng một cặp điều kiện – hành động: “NẾU điều kiện xảy ra THÌ hành động sẽ được thi hành”.

Trong biểu diễn tri thức sử dụng luật, ta cần xác định:

* Tập các sự kiện F (Facts): { A1, A2, A3, …… An, B1, B2, B3, …… Bm }
* Tập các luật R(rules) áp dụng trên các sự kiện có dạng sau: IF “A1 AND A2 AND … AND An” THEN B

Trong đó:

* (A1, A2, …… An, B) là các sự kiện trong F.
* A1, A2, A3, …… An: được gọi là điều kiện (conditions).
* B: là kết luận (conclusion).

Trong biễu diễn tri thức dựa trên luật, cơ chế suy luận trên các luật sản xuất bao gồm hai cơ chế : suy diễn tiến, và suy diễn lùi.

* + - ***Suy diễn tiến***: là quá trình suy luận xuất phát từ một số sự kiện ban đầu, xác định các sự kiện có thể sinh ra từ sự kiện này.
    - ***Suy diễn lùi***: là quá trình suy luận ngược xuất phát từ một số sự kiện ban đầu, ta tìm kiếm các sự kiện đã "sinh" ra sự kiện này.

Vấn đề tối ưu luật: Tập các luật trong một cơ sở tri thức rất có khả năng thừa, trùng lặp hoặc nguy hiểm hơn là mâu thuẫn nhau. Dĩ nhiên là hệ thống có thể đổ lỗi cho người dùng về việc đưa vào hệ thống những tri thức như vậy. Tuy việc tối ưu một cơ sở tri thức về mặt tổng quát là một thao tác khó (vì giữa các tri thức thường có quan hệ không tường minh), nhưng trong giới hạn cơ sở tri thức dưới dạng luật, ta vẫn có một số thuật toán đơn giản để loại bỏ các vấn đề này:

* Rút gọn vế phải: A ∧ B 🡪 A ∧ C sẽ trở thành A ∧ B 🡪 C
* Rút gọn vế trái: (L1) A, B 🡪 C (L2) A 🡪 X (L3) X 🡪 C sẽ trở thành A 🡪 C do đó L1 bị dư thừa có thể loại bỏ
* Phân rã và kết hợp luật: A ∧ B 🡪 C sẽ trở thành A 🡪 C, B 🡪 C
* Luật thừa: Một luật là thừa nếu có thể suy ra từ luật khác ví dụ A 🡪 B, B 🡪 C, A 🡪 C thì luật thứ 3 bị thừa.

1. Ưu điểm của phương pháp biểu diễn tri thức dựa trên luật:

Biểu diễn tri thức bằng luật đặc biệt hữu hiệu trong những tình huống hệ thống cần đưa ra những hành động dựa vào những sự kiện có thể quan sát được. Nó có những ưu điểm chính yếu sau đây :

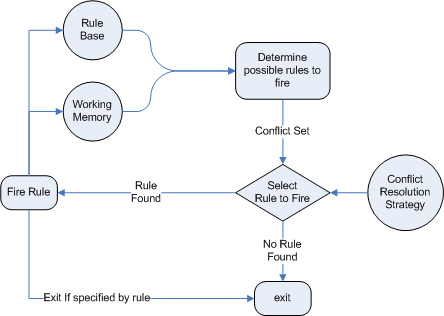
* Các luật rất dễ hiểu nên có thể dễ dàng dùng để trao đổi với người dùng (vì nó là một trong những dạng tự nhiên của ngôn ngữ).
* Có thể dễ dàng xây dựng được cơ chế suy luận và giải thích từ các luật.
* Việc hiệu chỉnh và bảo trì hệ thống là tương đối dễ dàng.
* Có thể cải tiến dễ dàng để tích hợp các luật mờ.

1. Nhược điểm của phương pháp biểu diễn tri thức dựa trên luật.

* Các tri thức phức tạp đôi lúc đòi hỏi quá nhiều (hàng ngàn) luật sinh. Điều này sẽ làm nảy sinh nhiều vấn đề liên quan đến tốc độ lẫn quản trị hệ thống.
* Thống kê cho thấy, người xây dựng hệ thống trí tuệ nhân tạo thích sử dụng luật sinh hơn tất cả phương pháp khác (dễ hiểu, dễ cài đặt) nên họ thường tìm mọi cách để biểu diễn tri thức bằng luật sinh cho dù có phương pháp khác thích hợp hơn! Đây là nhược điểm mang tính chủ quan của con người.
* Cơ sở tri thức luật sinh lớn sẽ làm giới hạn khả năng tìm kiếm của chương trình điều khiển. Nhiều hệ thống gặp khó khăn trong việc đánh giá các hệ dựa trên luật sinh cũng như gặp khó khăn khi suy luận trên luật sinh.

1. Suy diễn tiến:

Là một trong hai cơ chế cơ chế suy luận trên các luật sản xuất trong biểu diễn tri thức dựa trên luật. Bài toán cần chứng minh: Với một tập các mệnh đề giả thiết (cơ sở tri thức) KB, cần suy ra mệnh đề kết luận Q. Suy diễn tiến là suy diễn dựa trên dữ liệu nên rất phù hợp với các bài toán đưa ra quyết đinh, nhận dạng đối tượng…

* + *Ý tưởng của suy diễn tiến*: Lặp lại 2 bước sau cho đến khi suy ra được kết luận:
    - Áp dụng các luật có mệnh đề giả thiết được thỏa mãn trong KB
    - Bổ sung kết luận của các luật đó vào trong KB.
  + *Sơ đồ thuật toán:*
  + *Những điểm cần chú ý khi cài đặt giải thuật suy diễn tiến:*
    - Tránh các vòng lặp bằng cách kiểm tra xem các mệnh đề mới đã có trong danh sách các mệnh đề cần chứng minh chưa? Nếu rồi thì không bổ sung lại nữa!
    - Tránh việc chứng minh lặp lại đối với một mệnh đề. Mệnh đề này có thể đã được chứng minh là đúng ở trước đó hoặc đã được chứng minh ở trước đó là không thể thỏa mãn được trong KB.

1. Ưu điểm suy diễn tiến:
   * Ưu điểm chính của suy diễn tiến là làm việc tốt khi bài toán về bản chất đi thu thập thông tin rồi thấy điều cần suy diễn.
   * Suy diễn tiến cho ra khối lượng lớn các thông tin từ một số thông tin ban đầu. Nó sinh ra nhiều thông tin mới.
   * Suy diễn tiến là tiếp cận lý tưởng đối với loại bài toán cần giải quyết các nhiệm vụ như lập kế hoạch, điều hành điều khiển và diễn dịch.
2. Nhược điểm suy diễn tiến:
   * Một nhược điểm chính của hệ thống suy diễn tiến là không xác định được rằng chỉ một vài thông tin là quan trọng. Hệ thống hỏi một loạt các câu hỏi mà không biết rằng chỉ cần một số ít câu đã đi đến kết luận được.
   * Hệ thống có thể hỏi những câu không liên quan. Có thể các câu trả lời cũng quan trọng, nhưng làm người dùng lúng túng khi phải trả lời các câu không dính đến chủ đề, từ đó làm giảm trải nghiệm người dùng của hệ thống.
3. Suy diễn lùi:

Là quá trình suy luận ngược xuất phát từ một số sự kiện ban đầu, ta tìm kiếm các sự kiện đã "sinh" ra sự kiện này. Ví dụ: Trong ví dụ ở phần “Suy diễn tiến”, nếu ban đầu ta cần tìm C. Ta xem trong các luật sinh ra C để tìm sự kiện nào đã có trong đề bài. Nếu tìm được thì kết thúc còn không tìm được thì lại truy ngược lên đối với các sự kiện đã sinh ra C. Ở đây nhờ luật R1 ta tìm ra được sự kiện A, B mà đề bài đã cho trước.

Thuật giải suy diễn lùi:

B1: Giả sử mục tiêu đúng

B2: Phát sinh các mục tiêu con

B3: Kiểm tra các mục tiêu con

If mục tiêu đáp ứng then goto B8

B4: Tìm luật có thể phát sinh sự kiện mới

B5: If không tìm được luật then

Dừng không tìm được lời giải

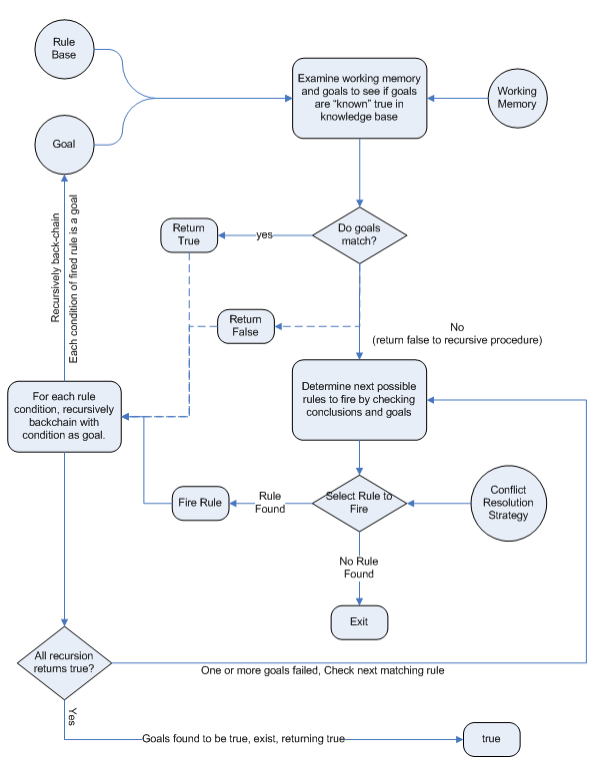
B6: If B4 thành công then

Ghi nhận thông tin về luật vào lời giải và sự kiện mới vào giả thiết được phát sinh từ các luật.

B7: Goto B4

B8: Tìm được lời giải trong danh sách luật solution.

Sơ đồ thuật toán:



1. Biểu diễn tri thức dựa trên mạng ngữ nghĩa:
2. Đặc điểm:

Mạng ngữ nghĩa là một phương pháp biểu diễn tri thức, được xây dựng dựa trên phương pháp đồ thị để biểu diễn các mối liên hệ giữa các tri thức tổng quát, các khái niệm, các sự việc...

Do mạng ngữ nghĩa là một loại đồ thị cho nên ta có thể dùng những thuật toán của đồ thị trên mạng ngữ nghĩa như thuật toán tìm liên thông, tìm đường đi ngắn nhất,… để thực hiện các cơ chế suy luận. Điểm đặc biệt của mạng ngữ nghĩa so với đồ thị thông thường chính là việc gán một ý nghĩa cho các cung. Cung nối giữa hai đỉnh cho biết giữa hai khái niệm tương ứng có sự liên hệ như thế nào. Việc gán ngữ nghĩa vào các cung của đồ thị đã giúp giảm bớt được số lượng đồ thị cần phải dùng để biễu diễn các mối liên hệ giữa các khái niệm.

Một đặc điểm quan trọng của mạng ngữ nghĩa là tính kế thừa. Chính đặc tính kế thừa của mạng ngữ nghĩa đã cho phép ta có thể thực hiện được rất nhiều phép suy diễn từ những thông tin sẵn có trên mạng.

Cơ chế suy diễn áp dụng trong mạng ngữ nghĩa là thực hiện theo thuật toán loang truyền đơn giản theo hai bước sau:

* + Kích hoạt các đỉnh đã cho ban đầu (các đỉnh đã có giá trị)
  + Nếu một đỉnh chưa xác định nối với n đỉnh khác (thông qua những mối liên hệ) và trong đó có n-1 đỉnh đã xác định thì đỉnh đó cũng được xác định. Lặp lại bước này cho đến khi xác định được tất cả các đỉnh.

1. Ưu điểm:

* Mạng ngữ nghĩa rất linh động, có thể thêm vào mạng các đỉnh hoặc cung mới để bổ sung các tri thức cần thiết
* Mạng ngữ nghĩa có tính trực quan cao nên rất dễ hiểu
* Mạng ngữ nghĩa cho phép các đỉnh có thể kế thừa các tính chất từ các đỉnh khác thông qua các cung loại “là’ từ đó có thể tạo ra các liên kết “ngầm” giữa những đỉnh không có liên kết trực tiếp với nhau.
* Mạng ngữ nghĩa hoạt động khá tự nhiên theo cách thức con người ghi nhận thông tin.

1. Nhược điểm:

* Vẫn chưa có một chuẩn nào quy định các giới hạn cho các đỉnh và cung của mạng. Điều đó đồng nghĩa với người dùng có thể gắn bất kỳ khái niệm nào cho đỉnh hoặc cung.
* Tính thừa kế trong mạng có thể dẫn đến khả năng mâu thuẩn tri thức

1. Biểu diễn tri thức bằng mạng ngữ nghĩa:

Khi biểu diễn một mạng ngữ nghĩa, các đỉnh của đồ thị là các đối tượng (khái niệm, tri thức, sự việc) nào đó, còn các cung giữa các đỉnh thể hiện các mối liên hệ giữa các đối tượng (khái niệm, tri thức, sự việc) này với nhau.

******

*Hình 1.1 Ví dụ về mạng ngữ nghĩa tiêu biểu*

Trong ví dụ trên, các yếu tố như “Xe máy, Xe, Động cơ, Xăng, Đường” được xem là các đối tượng của mạng ngữ nghĩa. Trong khi đó các yếu tố “Là, Di chuyển trên, chạy bằng hay có” là các mối liên hệ giữa các đối tượng.



*Hình 1.2 ví dụ về mạng ngữ nghĩa kế thừa*

Trong mạng ngữ nghĩa này, ta có thể thấy được các mối quan hệ như sau:

* Hình vuông là hình chữ nhật + là tứ giác + có 4 góc. Từ đó ta có thể suy luận được là hình vuông có 4 góc.
* Hình chữ nhật là hình bình hành + hình bình hành có hai cặp cạnh bằng nhau. Từ đó có thể suy ra hình chữ nhật có hai cặp cạnh bằng nhau.
* Dù không có đường liên hệ trực tiếp từ đối tượng “Hình vuông” đến đối tượng “4 góc” nhưng thông qua tính chất kế thừa ta có thể xác định được là đối tượng “Hìnhvuông” có liên hệ “có” với đối tượng “4 góc”.
* Tương tự với trường hợp của đối tượng “Hình chữ nhật” và đối tượng “Hai cặp cạnh bằng nhau”.

Tuy mạng ngữ nghĩa là một kiểu biểu diễn trực quan đối với con người nhưng khi đưa vào máy tính, các đối tượng và mối liên hệ giữa chúng thường được biểu diễn dưới dạng những phát biểu động từ (như vị từ). Hơn nữa, các thao tác tìm kiếm trên mạng ngữ nghĩa thường khó khăn (đặc biệt đối với những mạng có kích thước lớn). Do đó, mô hình mạng ngữ nghĩa được dùng chủ yếu để phân tích vấn đề. Sau đó, nó sẽ được chuyển đổi sang dạng luật hoặc frame để thi hành hoặc mạng ngữ nghĩa sẽ được dùng kết hợp với một số phương pháp biểu diễn khác.

GIẢI QUYẾT BÀI TOÁN

1. Đặt vấn đề:

Mục đích của đề tài là xây dựng hệ thống hỗ trợ tìm kiếm căn hộ chung cư. Người dùng sẽ nhập các thông số của căn hộ mà họ đang tìm kiếm. Từ các thông tin này, hệ thống sẽ ra đưa các căn hộ chung cư có thông số trùng với thông số người dùng nhập.

Trường hợp không có kết quả trả về, hệ thống sẽ yêu cầu người dùng cung cấp thêm thông tin cá nhân thông qua các câu hỏi bổ sung từ hệ thống. Từ các thông tin này, chương trình sẽ mở rộng miền tri thức tìm kiếm và gợi ý thêm các căn hộ nằm ngoài miền tìm kiếm lúc đầu mà vẫn đáp ứng nhu cầu của khách hàng.

Với bài toán ứng dụng biểu diễn tri thức, bước đầu tiên là cần phải thu thập tri thức, đó chính là bộ dữ liệu về chung cư căn hộ để người dùng có thể tìm kiếm và các tri thức để có thể mở rộng miền tri thức. Tri thức có vai trò rất quan trọng đối với độ tin cậy của hệ thống. Nếu bộ dữ liệu sai thì hệ thống không thể đưa ra kết quả tìm kiếm chính xác cho người, trong khi nếu tri thức mở rộng không hợp lý thì hệ thống hỗ trợ tìm kiếm sẽ không đưa ra gợi ý thỏa mõa mong muốn của người mua.

Tiếp theo, chúng ta sẽ xây dựng cơ sở tri thức hỗ trợ tìm kiếm bằng mạng ngữ nghĩa bao gồm các đỉnh và cung. Cơ chế suy diễn mạng ngữ nghĩa sẽ được thực hiện . Các đỉnh đã được kích thoạt từ ban đầu sẽ đi kích hoạt các đỉnh nối liền khác nếu đủ số đỉnh đã được kích hoạt trong luật. Lặp lại bước này cho đến khi xác định được tất cả các đỉnh. Dựa vào tất cả các đỉnh đã kích hoạt, hệ thống sẽ mở rộng miền tri thức tìm kiếm và trả về các kết quả phù hợp.

1. Thu thập tri thức:

Thu thập tri thức là bước quan trọng mở đầu cho mỗi bài toán, đặc biệt đối với bài toán tư vấn ra quyết định. Các tri thức cụ thể cho bài toán của nhóm như sau:

* + Bộ dữ liệu về căn hộ chung cư trên Kaggle [3] lấy từ chotot.vn bao gồm các thuộc tính: giá tiền, diện tích, vị trí, số phòng ngủ, tên chung cư.
  + Dữ liệu về người sử dụng: tuổi tác, nghề nghiệp, giới tính, sở thích, thế hệ Z, thế hệ Y, độc thân, cặp đôi, gia đình, con cái còn nhỏ.
  + Các gợi ý được hỏi bổ sung trong trường hợp cần mở rộng miền tri thức:
* vay 30% giá trị căn nhà muốn mua kỳ hạn dài
* có thể chịu làm xa
* Từ một phòng ngủ
* Từ hai phòng ngủ
* Từ ba phòng ngủ
* Số phòng ngủ bằng số thành viên của gia đình + 1 \* trẻ em và phải ở trung tâm và rộng rãi
* vay 50% giá trị căn nhà nhưng giá nhà cao
* Có thể chọn ngoại ô, từ 2 phòng ngủ
  + Tri thức: các dữ liệu về căn hộ phù hợp với các dữ liệu của người sử dụng. Ví dụ: Những người có thu nhập cao có thể vay ngân hàng thêm tiền để mua.

Cách thức thu thập thông tin:

* + Dựa trên tìm kiếm thông tin: trên internet dataset về căn hộ chung cư (kaggle), có thể crawl dữ liệu trên các trang web bất động sản
  + Dựa trên hỏi đáp: về dữ liệu người sử dụng (tuổi tác, nghề nghiệp, giới tính, sở thích…)
  + Dựa trên thống kê để đưa ra các tri thức.

1. Nguồn tri thức - Bảng thu thập tri thức:

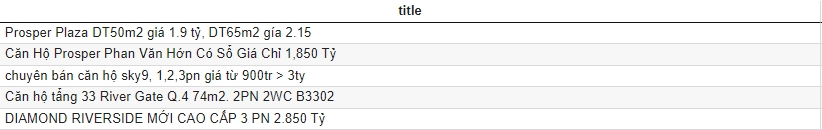
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Thay đổi phạm vi tìm kiếm | Nguồn | Nội dung | Keyphrase | Ưu tiên |
| vay 30% giá trị căn nhà muốn mua kỳ hạn dài | [Lương tháng 10 triệu có nên đi vay mua nhà?](https://vietnamnet.vn/luong-thang-10-trieu-co-nen-di-vay-mua-nha-2032337.html?fbclid=IwAR0BiJlUQUtHdg4m0LETinVbfI0o0FGFsypdvhIw75cN1nhSDBOidakRIWs)  [4] | TS. Trịnh Thị Phan Lan, Đại học Kinh tế - ĐHQGHN – Vay thời gian càng ngắn thì lãi suất và gốc trả hàng tháng cũng sẽ cao, những gói vay này phù hợp người có thu nhập hàng tháng cao. Những người có thu nhập ổn định nhưng không cao thì nên chọn gói vay có thời gian dài 10-25 năm. | Thu nhập không cao. Thu nhập ổn định | 2 |
| có thể chịu làm xa | [Gen Z "đói" việc chấp nhận đi làm xa nhà hơn 20km, làm thế nào để không muốn vứt bỏ tất cả sau một ngày dài kiệt quệ?](https://kenh14.vn/gen-z-doi-viec-chap-nhan-di-lam-xa-nha-hon-20km-lam-the-nao-de-khong-muon-vut-bo-tat-ca-sau-mot-ngay-dai-kiet-que-20220530163227245.chn)  [5] | Gen Z "đói" việc chấp nhận đi làm xa nhà hơn 20km, làm thế nào để không muốn vứt bỏ tất cả sau một ngày dài kiệt quệ? | Thế hệ Z | 2 |
| Từ một phòng ngủ | [HOW MANY BEDROOMS DO YOU TRULY NEED?](https://www.levine-homes.com/how-many-bedrooms-do-you-truly-need/)  [6] | Single buyers can usually skate by with just one room, but it’s also important to consider the potential for roommates. While you might not need anything large, having an extra office space can be nice for those who have hobbies or work from home. A house with one to three bedrooms should be large enough | Độc thân | 2 |
| Từ hai phòng ngủ | [HOW MANY BEDROOMS DO YOU TRULY NEED?](https://www.levine-homes.com/how-many-bedrooms-do-you-truly-need/)  [6] | Couples should also have some flexibility. While a single-bedroom home can be nice and cozy, it is also convenient to have a space where you can get some much-needed alone time. A two- to three-bedroom home is often suitable | Cặp đôi | 2 |
| Từ ba phòng ngủ | [HOW MANY BEDROOMS DO YOU TRULY NEED?](https://www.levine-homes.com/how-many-bedrooms-do-you-truly-need/)  [6] | How many bedrooms do you need when you have a whole family to think of? It really depends on the size of your family. It is convenient to allow each child to have his or her own bedroom, or you can have them share rooms. A family of four should fit comfortably in a house with three to four bedrooms. Consider an additional room for every one to two children or if you plan on your family growing | Cặp đôi, gia đình | 3 |
| Số phòng ngủ bằng số thành viên của gia đình + 1 \* trẻ em và phải ở trung tâm và rộng rãi | [HOW MANY BEDROOMS DO YOU TRULY NEED?](https://www.levine-homes.com/how-many-bedrooms-do-you-truly-need/)  [6] | Are you planning to add to your family, and if so will any of your children be sharing rooms? Younger kids can share very happily, but as they reach the teenage years they may want more space. The age gap between your kids, and whether they’re the same gender, will also come into the equation | gia đình, con cái còn nhỏ | 4 |
| vay 50% giá trị căn nhà nhưng giá nhà cao | [The Impact Of Buyers' Demography On Property Purchasing](https://www.researchgate.net/publication/322020386_The_Impact_Of_Buyers'_Demography_On_Property_Purchasing)  [7] | High income people would also prefer  to buy high-cost house (McCarthy, 1976 Borsch-Supan et al., 2001; Ariffin, 2010) , Vay thời gian càng ngắn thì lãi suất và gốc trả hàng tháng cũng sẽ cao, những gói vay này phù hợp người có thu nhập hàng tháng cao | thu nhập cao | 2 |
| Có thể chọn ngoại ô, từ 2 phòng ngủ | [XU HƯỚNG MUA NHÀ CỦA GIỚI TRẺ HIỆN NAY](https://phumyhungcorp.com.vn/tin-tuc/xu-huong-mua-nha-cua-gioi-tre)  [8] | Bên cạnh đó, với tình trạng giá nhà đất ngày càng leo thang, thế hệ trẻ ngày nay không tìm kiếm những căn hộ có diện tích quá lớn. Họ thường tìm kiếm các căn vừa phải, với giá hợp lý nhưng cần được thiết kế thông minh, đủ công năng hiện đại với 2 phòng ngủ, một phòng khách, không gian bếp riêng biệt… | Thế hệ Y | 2 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Câu hỏi | Trả lời | Nội dung | nguồn | Keyphrase  Được matching | |
| Trong câu hỏi | Trong câu trả lời |
| Thế nào dược coi là có thu nhập cao ? | lương từ 30 triệu | Lương cao không nói lên bạn là người giàu và có dư dả về tài chính. Như đã đề cập ở trên, mức lương còn phù hợp với mức sống tại khu vực bạn ở. Nhiều người sở hữu mức lương mơ ước từ 20-30 triệu/ 1 tháng” | [Lương bao nhiêu là cao và ổn định? Cách gia tăng thu nhập ai cũng nên biết](https://yuanta.com.vn/tin-tuc/luong-bao-nhieu-la-cao-va-on-dinh-cach-gia-tang-thu-nhap-ai-cung-nen-biet)  [9] | thu nhập cao | lương |
| Thế nào được coi là thu nhập không cao ? | lương từ 10 triệu | Ngược lại với nhiều người dù mức lương chỉ có mức trung bình từ 10-15 triệu, tuy nhiên mức chi tiêu không cao và không phải tốn quá nhiều chi phí thuê nhà, ăn uống, đi lại,… Điều này giúp họ tiết kiệm tối đa khoản lương nhận được, vì vậy luôn họ luôn có một cuộc sống ổn định về tài chính. | [Lương bao nhiêu là cao và ổn định? Cách gia tăng thu nhập ai cũng nên biết](https://yuanta.com.vn/tin-tuc/luong-bao-nhieu-la-cao-va-on-dinh-cach-gia-tang-thu-nhap-ai-cung-nen-biet)  [9] | thu nhập không cao | lương |
| Ai được coi là thế hệ Y ? | sinh từ 1981-1996 | Các nhà nghiên cứu và truyền thông phổ biến xem khoảng thời gian từ đầu thập niên 1980 đến giữa thập niên 1990 – đầu thập niên 2000 (hoặc từ năm 1981 đến 1996) là khoảng thời gian được sinh ra của nhóm này. | [Thế hệ Millennials](https://vi.wikipedia.org/wiki/Th%E1%BA%BF_h%E1%BB%87_Millennials)  [10] | thế hệ Y | Tuổi |
| Ai được coi là thế hệ Z ? | sinh từ 1997 - 2012 | Các nhà nghiên cứu và các phương tiện truyền thông phổ biến nhận định khoảng thời gian được sinh ra của thế hệ này là từ năm 1997 đến năm 2012 theo một nghiên cứu đến từ trung tâm nghiên cứu Pew | [Thế hệ Z](https://vi.wikipedia.org/wiki/Th%E1%BA%BF_h%E1%BB%87_Z)  [11] | thế hệ Z | Tuổi |
| Thế nào được coi là một gia đình ? | Là một cặp đôi và đã đính hôn hoặc cưới | So what is a family? A husband, a wife, and two children — maybe even a pet — served as the model for the traditional Canadian family for most of the 20th century. | [Chapter 14. Marriage and Family](https://opentextbc.ca/introductiontosociology2ndedition/chapter/chapter-14-marriage-and-family/)  [12] | gia đình | cặp đôi, cưới hoặc đính hôn |
| Quận nào được coi là ngoại ô | Hóc Môn,Bình Chánh,Nhà Bè,Cần Giờ | HCM City’s suburban districts of Hoc Mon, Binh Chanh and Nha Be will become urban districts or cities prior to 2025, while two other suburban districts of Cu Chi and Can Gio will be turned into urban ones in the 2025-2030 period. | [HCM City plans to turn five suburban districts into urban districts, cities](https://vietnamnet.vn/en/hcm-city-plans-to-turn-five-suburban-districts-into-urban-districts-cities-720292.html)  [13] |  |  |
| Giá nhà thế nào được coi là đắt đó | 200 triệu/m2 | Đầu năm 2020, giá bất động sản TP.HCM thiết lập mặt bằng giá mới, giá căn hộ trung tâm Quận 1 lên đến 500 triệu/m2, khu vực trung tâm Quận 2 lên đến 200 triệu/m2, các quận/huyện vùng ven cũng tăng kỷ lục. | [Giá bất động sản TP.HCM tăng cao chưa từng có](https://nhabanquan7.vn/gia-bat-dong-san-tp-hcm-tang-cao-chua-tung-co.html)  [14] |  |  |
| Nhà rộng rãi là thế nào ? | 111m2 | Three-bedroom apartments are large enough to accommodate a family of six or up to three roommates, especially if the apartment has multiple bathrooms. In some cases, a 3-bedroom apartment will be found in a townhome or duplex community, which contains a second level or loft area and larger living space on the main level. | [What Size Apartment Should I Get?](https://www.southernmanagement.com/blog/what-size-apartment-should-i-get/)  [15] |  |  |
| Trung tâm thành phố là ở đâu | Quận 1, Quận 3, Quận 4, Quận 5, Quận 6, Quận 8, Quận 10, Quận 11, Quận Phú Nhuận, Quận Bình Thạnh, Quận Tân Phú, Quận Tân Bình, Quận Gò Vấp | Tp.HCM với 5 khu đô thị lớn gồm:  Khu đô thị Trung Tâm Tp.HCM (Quận 1, Quận 3, Quận 4, Quận 5, Quận 6, Quận 8, Quận 10, Quận 11, Quận Phú Nhuận, Quận Bình Thạnh, Quận Tân Phú, Quận Tân Bình, Quận Gò Vấp)  Khu đô thị Đông Tp.HCM (Quận 9, Quận 2 & Thủ Đức)  Khu đô thị Tây Tp.HCM (Quận Bình Tân, 1 phần Huyện Bình Chánh)  Khu đô thị Nam Tp.HCM (Quận 7, Huyện Nhà Bè, Huyện Bình Chánh, Huyện Cần Giờ)  Khu đô thị Bắc Tp.HCM (Quận 12, Huyện Hóc Môn & Huyện Củ Chi) – Tại khu vực Thạnh Xuân – Quận 12 có Dự án PiCity High Park rộng 8,6 Ha lớn nhất khu vực. Thanh toán trước 600 triệu cho đến khi nhận căn hộ, với giá bán chỉ 1.7 tỷ/căn, Mặt tiền đường Thạnh Xuân 13. Dự án đã có GPXD vào ngày 19/06/2020 và đang tiến hàng xây dựng hầm móng vào ngày 01/07/2020 | [BẢN ĐỒ HÀNH CHÍNH TPHCM VÀ 24 QUẬN HUYỆN MỚI NHẤT 2021](https://danhkhoireal.vn/ban-do-tphcm-va-24-quan-huyen-moi-nhat/)  [16] |  |  |

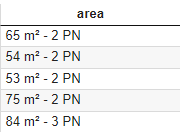
1. Biểu diễn tri thức:

Trong các giải quyết bài toán của nhóm, chúng em sử dụng mạng ngữ nghĩa để biểu diễn các tri thức. Với cơ sở tri thức là bộ dữ liệu từ Kaggle, các đỉnh chính là các thuộc tính trong bảng dữ liệu:

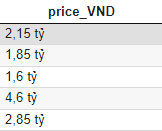
* Tên chung cư/căn hộ:



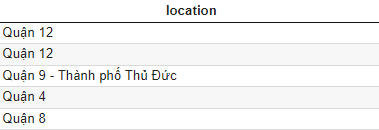
* Diện tích căn hộ - số phòng ngủ:



* Giá tiền:



* Vị trí (quận/huyện):



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **title** | **price\_VND** | **location** | **area\_m2** | **Number\_of\_bedrooms** |
| 0 | Prosper Plaza DT50m2 giá 1.9 tỷ, DT65m2 gía 2.15… | 2150000000 | 12 | 65.0 | 2 |
| 1 | Căn Hộ Prosper Phan Văn Hớn Có Sổ Giá Chỉ 1,85… | 1850000000 | 12 | 54.0 | 2 |
| 2 | chuyên bán căn hộ sky9, 1,2,3pn giá từ 900tr >… | 1600000000 | 9 | 53.0 | 2 |
| 3 | Căn hộ tẩng 33 River Gate Q.4 74m2. 2PN 2WC B... | 4600000000 | 4 | 75.0 | 2 |
| 4 | DIAMOND RIVERSIDE MỚI CAO CẤP 3 PN 2.850 Tỷ | 2850000000 | 8 | 84.0 | 3 |

1. Gợi ý các sản phẩm khác theo các tiêu chí do hệ thống tự định trước:

Trong trường hợp hệ thống không trả về kết quả nào, mô hình mạng ngữ nghĩa và các tri thức đã thu thập được sẽ mở rộng các miền tìm kiếm. Khi đó, mô hình mạng ngữ nghĩa sẽ gồm 13 đỉnh và 14 cung:

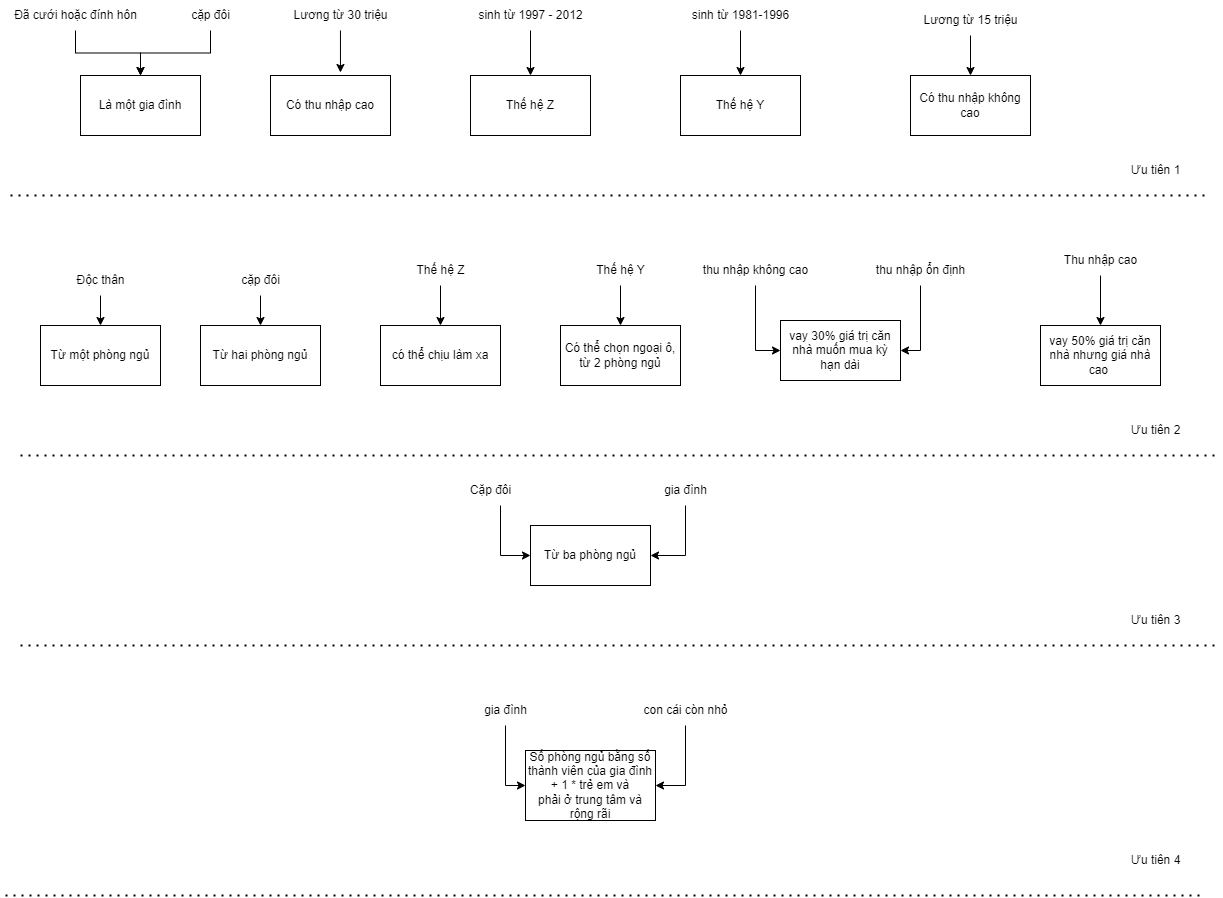
* Các đỉnh được gán một giá trị ưu tiên dựa vào việc mô tả người dùng chi tiết thế nào
* Thông tin của người dùng nhập vào ban đầu được gán độ ưu tiên thấp nhất, đối với những trường dữ liệu mà người dùng không có yêu cầu thì đặt độ ưu tiên cao nhất.
* Khi một đỉnh được kích hoạt, xét theo từng trường dữ liệu:
* Nếu độ ưu tiên của đỉnh trên trường dữ liệu hiện tại > độ ưu tiên của trường dữ liệu được lưu

**HOẶC**

* Nếu độ ưu tiên của đỉnh trên trường dữ liệu hiện tại == độ ưu tiên của trường dữ liệu được lưu và phạm vi tìm kiếm của đỉnh trên trường dữ liệu hiện tại > phạm vi tìm kiếm của trường dữ liệu được lưu

**THỰC HIỆN**:

* Cập nhật phạm vi tìm kiếm trên trường dữ liệu được lưu
* Cập nhật lại độ ưu tiên trên trường dữ liệu được lưu



1. Thuật giải suy diễn tiến:

* Người dùng nhập input là các trường từ tập dữ liệu Kaggle [3] như: “Diện tích”, “Số phòng ngủ”, “Giá tiền”, “Vị trí (quận)”.
* Dựa trên các thông số đó từ người dùng, hệ thống lọc các kết quả khớp với các nhu cầu trên.
* Nếu hệ thống tìm thấy kết quả thì trả về cho giao diện người dùng.
* Hỏi xem nếu người dùng có cần gợi ý thêm những căn hộ tiềm năng không.
* Nếu muốn gợi ý, hệ thống sẽ bước vào giai đoạn mở rộng miền tri thức và đưa ra gợi ý dựa trên miền mới này:
  + Thực hiện theo phần “*4a) Gợi ý các sản phẩm khác theo các tiêu chí do hệ thống tự định trước*” ở mục “GIẢI QUYẾT BÀI TOÁN”
  + Lặp lại việc hỏi xem người dùng có muốn gợi ý thêm những căn hộ khác không.
* Nếu người dùng không muốn gợi ý thì xuất các kết quả hiện có ra giao diện.
  + Hỏi người dùng có muốn so sánh giữa 2 kết quả hiện có không.
  + Nếu có thì lấy các kết quả người dùng muốn so sánh và trả về giao diện người dùng.
* Chương trình dừng lại

Dưới đây là flowchart biểu diễn thuật toán giải quyết bài toán của nhóm.

Diagram

Description automatically generated

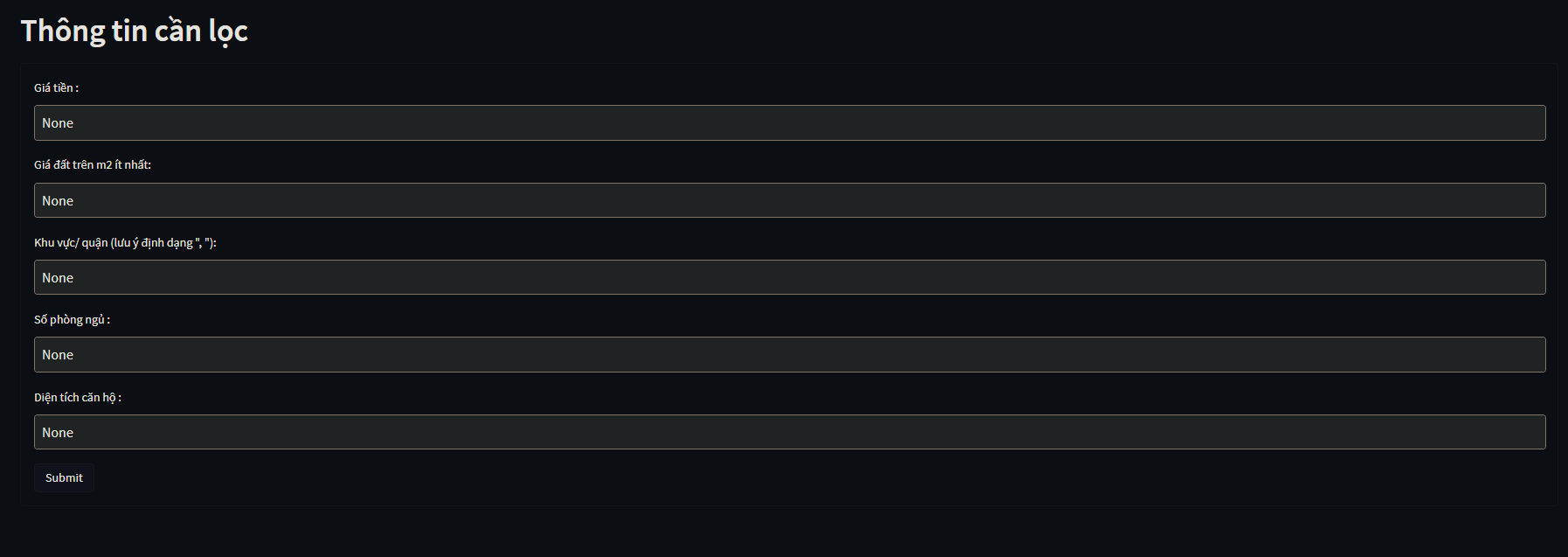
CÀI ĐẶT – DEMO CHƯƠNG TRÌNH

1. Cài đặt

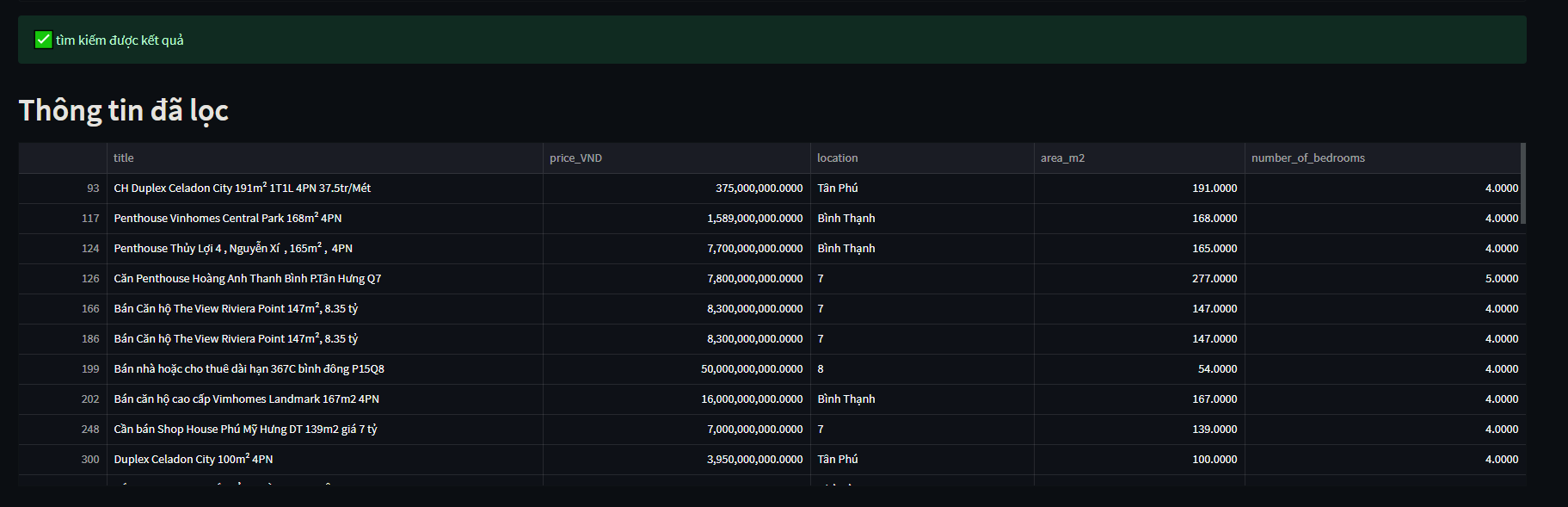
Chương trình được cài đặt bằng ngôn ngữ python 3, sử dụng thư viện pandas để truy xuất, cập nhật dữ liệu, framework streamlit để xây dựng UI và tạo ra ứng dụng web. Được host trên streamlit cloud, ứng dụng có thể được sử dụng thông qua địa chỉ [github](https://github.com/HuyRand/Knownledge_Representation) mà không cần phải cài đặt

1. Demo

Bảng nhập thông tin ban đầu

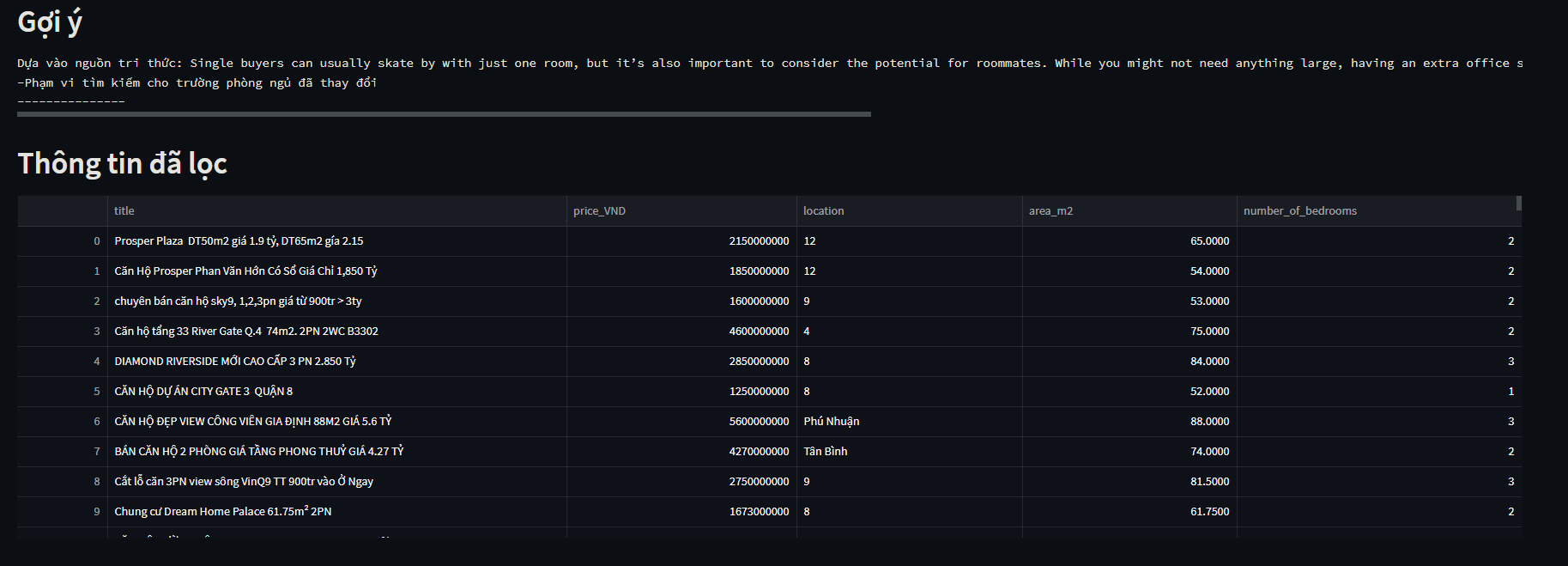


Hệ thống có thể trả về kết quả dựa theo yêu cầu người



Trường hợp hệ thống không thể trả về kết quả theo yêu cầu người dùng -> Hỏi thêm thông tin về người dùng -> Hệ thông gợi ý cho người dùng





Sau khi có được kết quả được lọc, hệ thống cho phép người dùng so sánh thông tin giữa 2 căn hộ dựa theo giá trị index được người dùng chọn. Nếu người dùng chọn giá trị index không có trong thông tin đã được lọc -> báo lỗi. Nếu người dùng chọn đúng -> biểu diễn bảng so sánh





ĐÁNH GIÁ – KẾT LUẬN

1. Đánh giá

Với việc sử dụng mạng ngữ nghĩa, mô hình tương đối dễ hiểu và đơn giản nhưng vẫn còn tính trừu tượng trong việc gán tính ưu tiên cho các đỉnh làm và khả năng mâu thuẫn giữa các miền tri thức làm cho việc scale mô hình còn khó khăn, cần nhưng chuyên gia trong lĩnh vực đang làm việc tính chính thống để có thể tạo ra mô hình trực quan và chính xác. Dữ liệu còn được lưu theo dạng excel và được truy xuất từ đó, có thể được đưa lên thành database.

1. Kết luận

Bằng cách sử dụng mạng ngữ nghĩa, ta có thể tạo được mô hình gợi ý dựa trên những tri thức được hệ thống hóa theo các kĩ thuật biểu diễn tri thức, giúp cho việc hiểu được hệ thống hoạt động như thế nào và tại sao lại dẫn đến kết quả đó. Vẫn còn những khó khăn trong việc thu thập, chuẩn hóa và hệ thống hóa các tri thức nhưng phương pháp đã thể hiện được tính trực quan và thiết thức khi so sánh với những phương pháp khác.

TÀI LIỆU THAM KHẢO

[1]: <https://tuoitre.vn/nhu-cau-nha-o-khong-lo-o-khu-dong-tp-hcm-20220914170452531.htm>

[2]: <https://www.vietnamplus.vn/nguon-cung-can-ho-tai-tphcm-chi-dap-ung-52-nhu-cau-theo-ke-hoach/825082.vnp>

[3]: <https://www.kaggle.com/datasets/hoandan/apartment-prices-in-the-city-ho-chi-minh-city>

[4]: <https://vietnamnet.vn/luong-thang-10-trieu-co-nen-di-vay-mua-nha-2032337.html?fbclid=IwAR0BiJlUQUtHdg4m0LETinVbfI0o0FGFsypdvhIw75cN1nhSDBOidakRIWs>

[5]: <https://kenh14.vn/gen-z-doi-viec-chap-nhan-di-lam-xa-nha-hon-20km-lam-the-nao-de-khong-muon-vut-bo-tat-ca-sau-mot-ngay-dai-kiet-que-20220530163227245.chn>

[6]: <https://www.levine-homes.com/how-many-bedrooms-do-you-truly-need/>

[7]:<https://www.researchgate.net/publication/322020386_The_Impact_Of_Buyers'_Demography_On_Property_Purchasing>

[8]: <https://phumyhungcorp.com.vn/tin-tuc/xu-huong-mua-nha-cua-gioi-tre>

[9]: <https://yuanta.com.vn/tin-tuc/luong-bao-nhieu-la-cao-va-on-dinh-cach-gia-tang-thu-nhap-ai-cung-nen-biet>

[10]: <https://vi.wikipedia.org/wiki/Th%E1%BA%BF_h%E1%BB%87_Millennials>

[11]: <https://vi.wikipedia.org/wiki/Th%E1%BA%BF_h%E1%BB%87_Z>

[12]: <https://opentextbc.ca/introductiontosociology2ndedition/chapter/chapter-14-marriage-and-family/>

[13]: <https://vietnamnet.vn/en/hcm-city-plans-to-turn-five-suburban-districts-into-urban-districts-cities-720292.html>

[14]: <https://nhabanquan7.vn/gia-bat-dong-san-tp-hcm-tang-cao-chua-tung-co.html>

[15]: <https://www.southernmanagement.com/blog/what-size-apartment-should-i-get/>

[16]: <https://danhkhoireal.vn/ban-do-tphcm-va-24-quan-huyen-moi-nhat/>