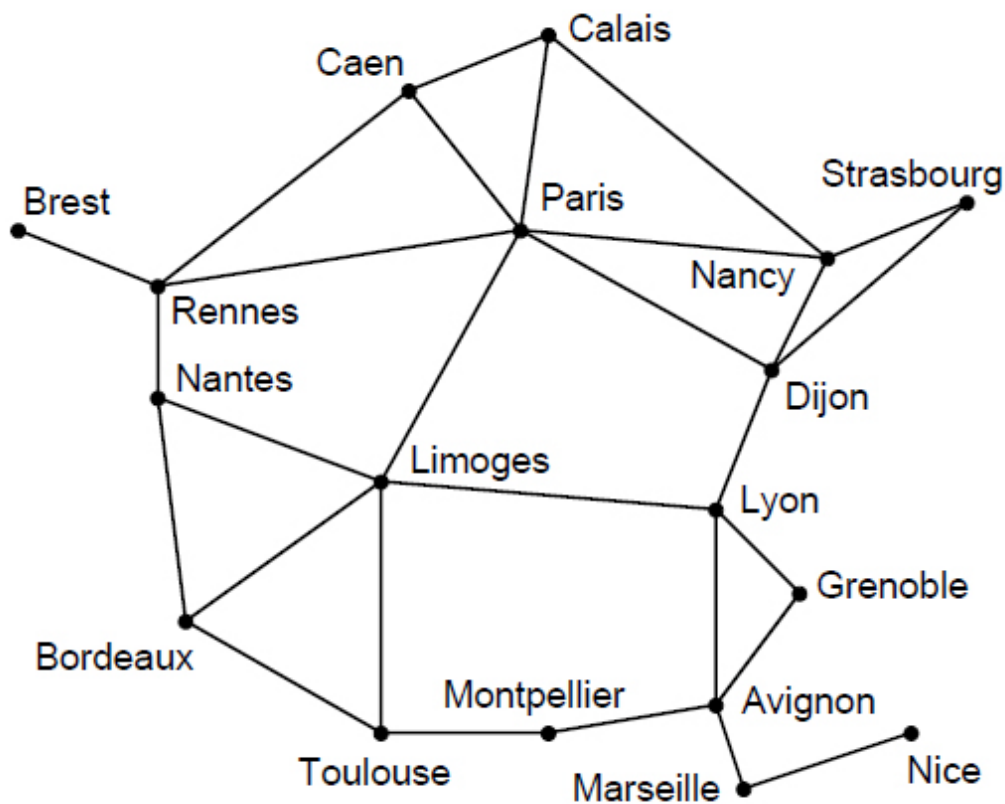


Danh sách đề tài đồ án môn Trí Tuệ Nhân Tạo

(Mỗi nhóm gồm từ 2 đến 3 sinh viên)

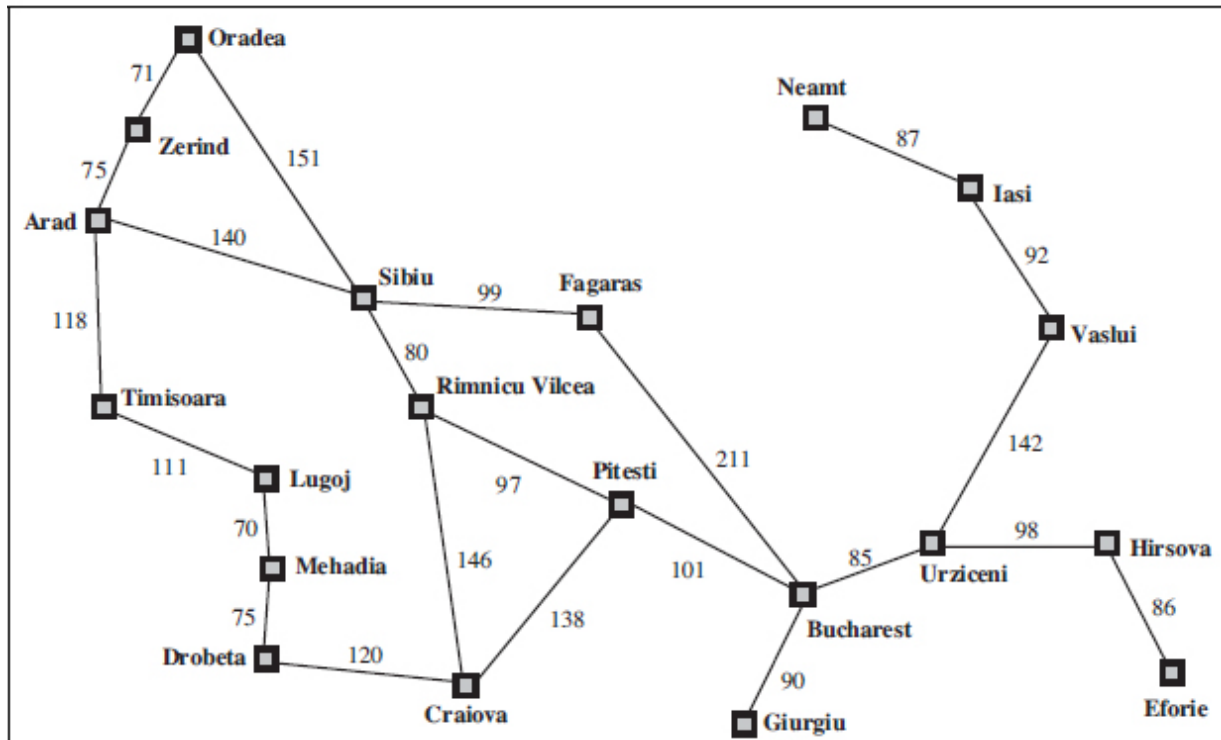
1. Sử dụng giải thuật A* để giải bài toán đồ 8-Miếng (8-puzzle).
2. Sử dụng giải thuật best-first-search để giải bài toán đồ 8-Miếng (8-puzzle). Hàm ước lượng h đánh giá độ tốt của trạng thái hiện hành tính bằng số ô mà trạng thái hiện hành sai biệt so với trạng thái đích.
3. Sử dụng giải thuật Depth-first-search và breadth-first-search để giải bài toán tìm đường đi từ thành phố Rennes đến Avignon.

Cho bản đồ thành phố ở nước Pháp như sau:



4. Sử dụng giải thuật A* để giải bài toán tìm đường đi ngắn nhất từ thành phố Arad đến Bucharest.

Cho bản đồ thành phố ở Romania như sau:

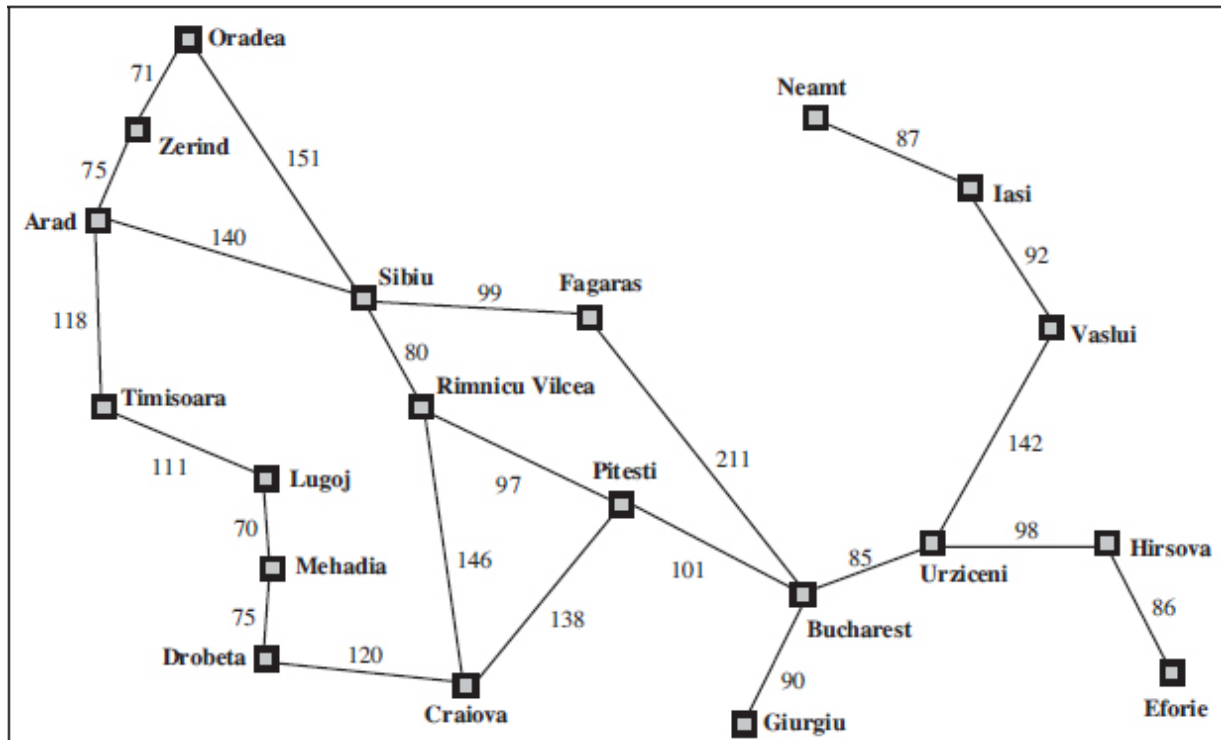


Và khoảng cách ước lượng từ mỗi thành phố đến Bucharest như sau:

Arad	366	Mehadia	241
Bucharest	0	Neamt	234
Craiova	160	Oradea	380
Drobeta	242	Pitesti	100
Eforie	161	Rimnicu Vilcea	193
Fagaras	176	Sibiu	253
Giurgiu	77	Timisoara	329
Hirsora	151	Urziceni	80
Iasi	226	Vasui	199
Lugoj	244	Zerind	374

5. Sử dụng giải thuật Best-first-search để giải bài toán tìm đường đi ngắn nhất từ thành phố Arad đến Bucharest.

Cho bản đồ thành phố ở Romania như sau:

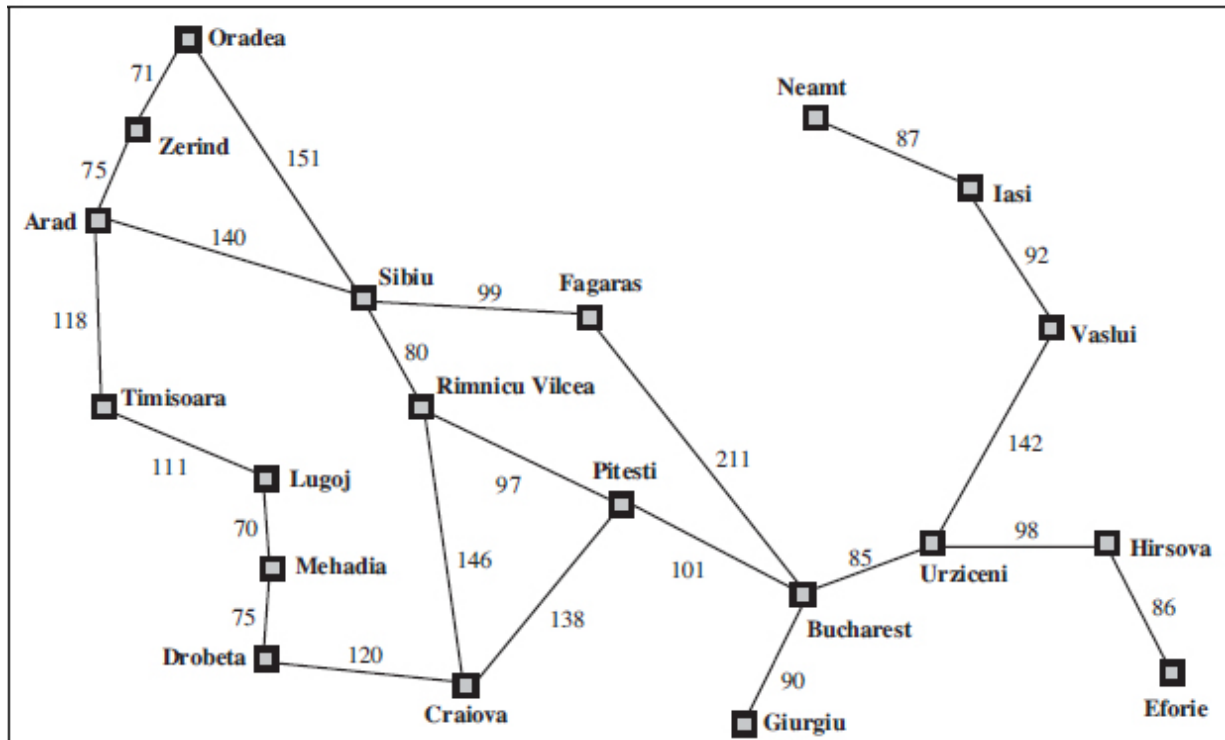


Và khoảng cách ước lượng từ mỗi thành phố đến Bucharest như sau:

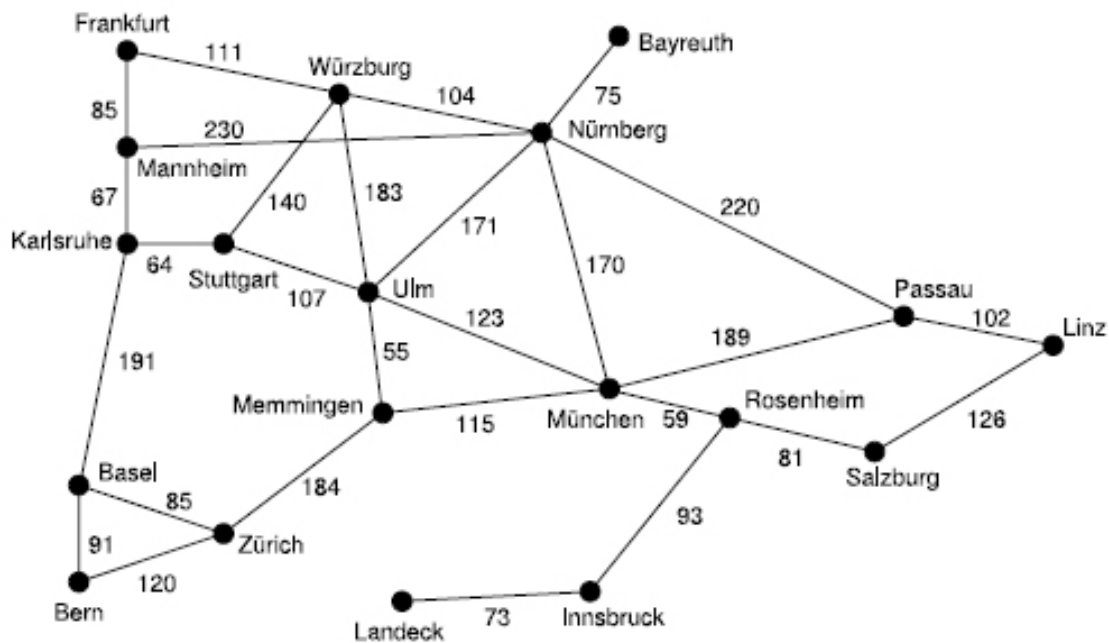
Arad	366	Mehadia	241
Bucharest	0	Neamt	234
Craiova	160	Oradea	380
Drobeta	242	Pitesti	100
Eforie	161	Rimnicu Vilcea	193
Fagaras	176	Sibiu	253
Giurgiu	77	Timisoara	329
Hirsova	151	Urziceni	80
Iasi	226	Vasui	199
Lugoj	244	Zerind	374

6. Sử dụng giải thuật Uniform-cost search để giải bài toán tìm đường đi ngắn nhất từ thành phố Arad đến Bucharest.

Cho bản đồ thành phố ở Romania như sau:



7. Sử dụng giải thuật A* để giải bài toán tìm đường đi ngắn nhất từ thành phố Linz đến Ulm.
Cho bản đồ thành phố ở Đức như sau:

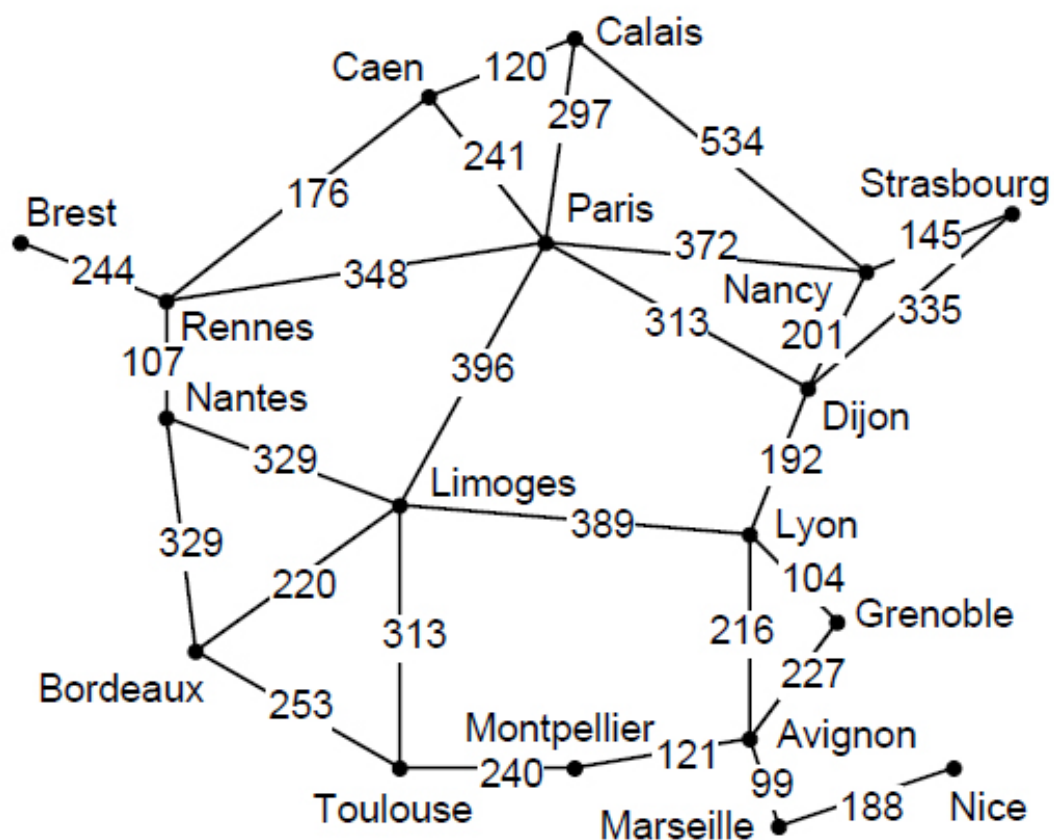


Và khoảng cách ước lượng từ mỗi thành phố đến Ulm như sau:

Bassel	204	Mannheim	164
Bayreuth	207	Memmingen	47
Bern	247	Nurnberg	132
Frankfurt	215	Passau	257
Innbruck	163	Rosenheim	168
Karlsruhe	137	Stuttgart	75
Landeck	143	Salzburg	236
Linz	318	Wurzburg	153
Munchen	120	Zurich	157

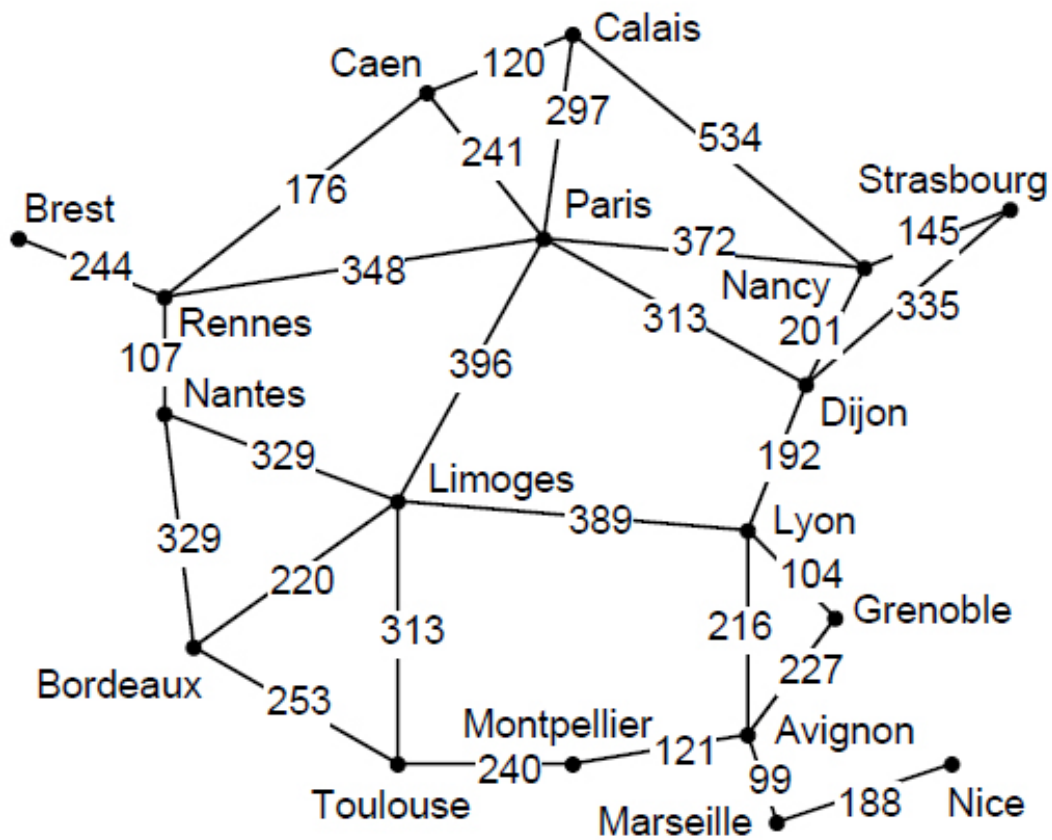
8. Sử dụng giải thuật Uniform-cost search để giải bài toán tìm đường đi ngắn nhất từ thành phố Rennes đến Avignon.

Cho bản đồ thành phố ở Pháp như sau:



9. Sử dụng giải thuật A* để giải bài toán tìm đường đi ngắn nhất từ thành phố Rennes đến Avignon.

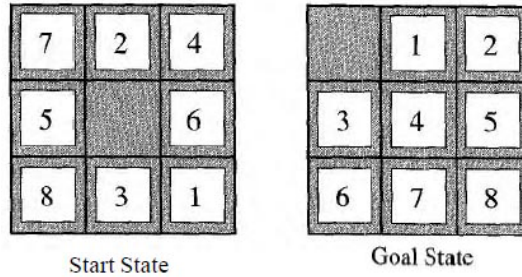
Cho bản đồ thành phố ở Pháp như sau:



Và khoảng cách ước lượng từ mỗi thành phố đến Avignon như sau:

Bordeaux	54	Montpellier	12
Brest	93	Nancy	14
Caen	52	Nantes	64
Calais	30	Nice	25
Dijon	3	Paris	25
Grenoble	9	Rennes	65
Limoges	36	Strasbourg	29
Lyon	0	Toulouse	34
Marseille	5		

10. Sử dụng giải thuật A* để giải bài toán đồ 8-Miếng (8-puzzle) với hàm ước lượng h_2 được định nghĩa như sau. Giả sử ta có hai trạng thái khởi đầu và đích như trong hình sau:

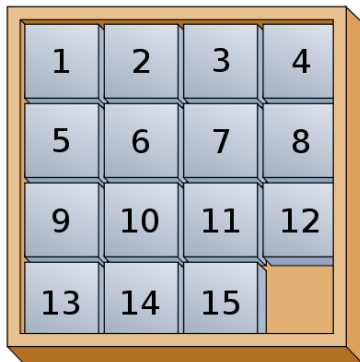


Hàm ước lượng h_2 của trạng thái khởi đầu được tính bằng tổng khoảng cách của từng ô trong trạng thái này so với với từng ô tương ứng trong trạng thái đích. Khoảng cách ở đây là tổng số lần dịch chuyển ô theo chiều ngang và chiều dọc. Khoảng cách này còn được gọi là khoảng cách Manhattan.

ô trị1	ô trị2	ô trị3	ô trị4	ô trị5	ô trị6	ô trị7	ô trị8
--------	--------	--------	--------	--------	--------	--------	--------

$$h_2 = 3 + 1 + 2 + 2 + 2 + 3 + 3 + 2 = 18$$

11. Sử dụng giải thuật A* để giải bài toán đồ 15-Miếng (15-puzzle). Cấu trúc của 15-miếng như sau:



12. Leo đồi (Hill-Climbing) là một giải thuật tìm kiếm heuristic được mô tả như sau:

function HILL-CLIMBING(*problem*) **returns** a state that is a local maximum

input: *problem*

local variables: *current, neighbor* // *current* is a node and *neighbor* is a node

current ← MAKE-NODE(INITIAL-STATE[*problem*])

loop do

neighbor ← a highest-valued successor of *current*

if VALUE[*neighbor*] > VALUE[*current*] **then returns** STATE[*current*]

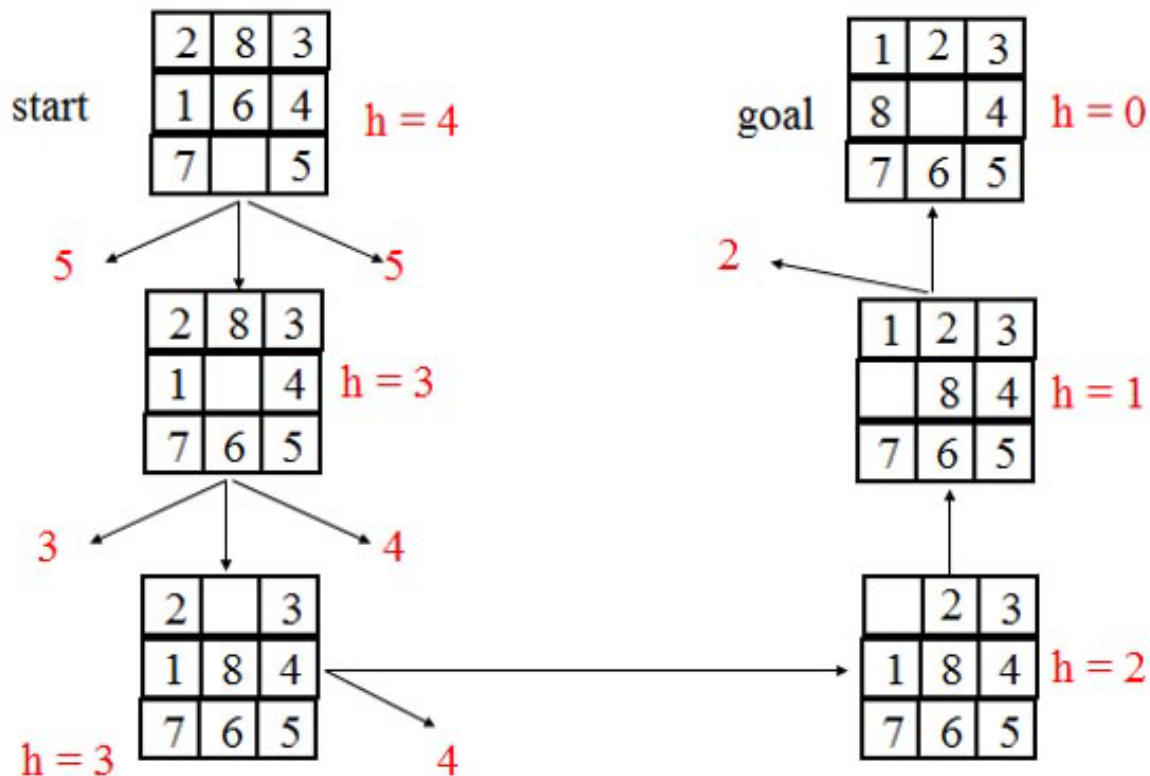
current ← *neighbor*

// VALUE is the evaluation function, smaller is better

Hãy sử dụng giải thuật Leo-đồi để giải bài toán đồ 8-Miếng (8-puzzle).

Hướng dẫn: xem hình vẽ sau:

Hill climbing: 8-puzzle example



Lưu ý: Hàm ước lượng h đánh giá độ tốt của trạng thái hiện hành bằng số ô mà trạng thái hiện hành sai biệt so với trạng thái đích.

13. Tìm hiểu giải thuật phân lớp k-lân cận gần nhất của Scikit-Learn để phân lớp bộ dữ liệu mẫu Liver (bệnh gan) lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

14. Tìm hiểu giải thuật phân lớp k-lân cận gần nhất của Scikit-Learn để phân lớp bộ dữ liệu mẫu Parkinson's Disease lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

15. Tìm hiểu giải thuật phân lớp k-lân cận gần nhất của Scikit-Learn để phân lớp bộ dữ liệu mẫu Heart-Cleveland (bệnh tim) lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

16. Tìm hiểu giải thuật phân lớp k-lân cận gần nhất của Scikit-Learn để phân lớp bộ dữ liệu mẫu Wine Quality lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

17. Tìm hiểu các giải thuật *lựa chọn thuộc tính* của phần mềm Scikit-Learn để áp dụng vào bộ dữ liệu Customer Churn (khách hàng rời bỏ công ty). Sau khi lựa chọn được những thuộc tính có ý nghĩa cho ứng dụng này, áp dụng giải thuật ANN (Artificial Neural Network) của Scikit-Learn để giải bài toán dự đoán khách hàng rời bỏ công ty (Customer Churn prediction).

Bộ dữ liệu Churn lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

18. Tìm hiểu giải thuật phân lớp Cây Quyết Định của Scikit-Learn để phân lớp bộ dữ liệu mẫu Car lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

19. Tìm hiểu giải thuật phân lớp Cây Quyết Định của Scikit-Learn để phân lớp bộ dữ liệu mẫu Lenses lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

20. Tìm hiểu giải thuật phân lớp Cây Quyết Định của Scikit-Learn để phân lớp bộ dữ liệu mẫu Titanic Survival lấy từ trang web của UCI:

http://archive.ics.uci.edu/ml/datasets

21. Tìm hiểu giải thuật phân lớp Naïve Bayes của Scikit-Learn để phân lớp bộ dữ liệu mẫu Car lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

22. Tìm hiểu giải thuật phân lớp Naïve Bayes của Scikit-Learn để phân lớp bộ dữ liệu mẫu Lenses lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

23. Hiện thực giải thuật Bagging với các bộ phân lớp thành phần là k-lân cận gần nhất để phân lớp bộ dữ liệu Breast cancer lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

So sánh hiệu quả phân lớp của giải thuật Bagging với bộ phân lớp đơn lẻ k-lân cận gần nhất.

24. Hiện thực giải thuật Hồi quy tuyến tính (linear regression) để dự báo:

- dữ liệu Thermostat Sales gồm 52 điểm dữ liệu và

TABLE 8.1 Weekly Thermostat Sales

206	189	172	255
245	244	210	303
185	209	205	282
169	207	244	291
162	211	218	280
177	210	182	255
207	173	206	312
216	194	211	296
193	234	273	307
230	156	248	281
212	206	262	308
192	188	258	280
162	162	233	345

- dữ liệu Air Passengers gồm 144 điểm dữ liệu sau đây:

Bộ dữ liệu Air Passengers

112, 118, 132, 129, 121, 135, 148, 148, 136, 119,
104, 118, 115, 126, 141, 135, 125, 149, 170, 170,
158, 133, 114, 140, 145, 150, 178, 163, 172, 178,
199, 199, 184, 162, 146, 166, 171, 180, 193, 181,
183, 218, 230, 242, 209, 191, 172, 194, 196, 196,
236, 235, 229, 243, 264, 272, 237, 211, 180, 201,
204, 188, 235, 227, 234, 264, 302, 293, 259, 229,
203, 229, 242, 233, 267, 269, 270, 315, 364, 347,
312, 274, 237, 278, 284, 277, 317, 313, 318, 374,
413, 405, 355, 306, 271, 306, 315, 301, 356, 348,
355, 422, 465, 467, 404, 347, 305, 336, 340, 318,
362, 348, 363, 435, 491, 505, 404, 359, 310, 337,
360, 342, 406, 396, 420, 472, 548, 559, 463, 407,
362, 405, 417, 391, 419, 461, 472, 535, 622, 606,
508, 461, 390, 432.

25. Hiện thực giải thuật Hồi quy tuyến tính (linear regression) để dự báo tỉ giá ngoại tệ.

(Bộ dữ liệu mẫu tỉ giá ngoại tệ sẽ được cung cấp).

26. Hiện thực giải thuật Hồi quy tuyến tính (linear regression) để dự báo giá cổ phiếu của một công ty.

(Bộ dữ liệu mẫu giá cổ phiếu sẽ được cung cấp).

27. Tìm hiểu giải thuật gom cụm k-Means của Scikit-Learn để gom cụm bộ dữ liệu mẫu Iris lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

Lưu ý: Iris là bộ dữ liệu đã được phân lớp sẵn, nên thông số k của giải thuật k-means phải chọn bằng số lớp của bộ dữ liệu và phải dùng các *tiêu chí đánh giá ngoại* để đánh giá chất lượng gom cụm.

28. Tìm hiểu giải thuật gom cụm k-Means của Scikit-Learn để gom cụm bộ dữ liệu mẫu Image Segmentation lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

29. Tìm hiểu giải thuật gom cụm k-Means của Scikit-Learn để gom cụm bộ dữ liệu mẫu Glass lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

30. Hiện thực giải thuật gom cụm k-Means có kết hợp phương pháp khởi tạo các centroid ban đầu của Al Daoud để gom cụm bộ dữ liệu mẫu Glass lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

31. Hiện thực giải thuật gom cụm k-Means++ để gom cụm bộ dữ liệu mẫu Glass lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

32. Tìm hiểu giải thuật gom cụm k-Means của Scikit-Learn để gom cụm bộ dữ liệu mẫu Wholesale Customers lấy từ trang web của UCI:

<http://archive.ics.uci.edu/ml/datasets/>

33. Tìm hiểu giải thuật gom cụm k-Means của Scikit-Learn để gom cụm bộ dữ liệu chuỗi thời gian chứng khoán Stock.

(Bộ dữ liệu mẫu Stock sẽ được cung cấp).

34. Tìm hiểu mạng nơ ron học sâu Convolutional Neural Network + phần mềm Keras và áp dụng vào xây dựng chương trình phân lớp hình ảnh digit 0-9.

Lưu ý: Bộ dữ liệu mẫu có thể lấy từ MNIST database.

35. Tìm hiểu mạng nơ ron học sâu LSTM + phần mềm Keras và áp dụng vào dự báo dữ liệu tỉ giá ngoại tệ.

(Bộ dữ liệu mẫu tỉ giá ngoại tệ sẽ được cung cấp).

Các bộ dữ liệu học máy thông dụng (UCI)

<u>Tên bộ dữ liệu</u>	<u>Số mẫu</u>	<u>Số thuộc tính</u>	<u>Loại dữ liệu</u>	<u>Số lớp</u>	<u>Ghi Chú</u>
Breast cancer	669	5	số nguyên (1-10)	2	clustering, classification
BMW-Browsers	31	7	rời rạc	2	classification
Car	1728	6	rời rạc	4	classification
Churn	3333	20	số thực, rời rạc	2	ANN
CMC	1473	10	rời rạc	3	classification
CPU	209	8	số nguyên		clustering
Customer	60	4	số		clustering
Glass	214	9	số thực	7	clustering
Heart (Cleveland)	303	14	số thực	4	kNN
Iris	150	4	số	3	classification, clustering
Image Segmentation	2310	19	số thực	7	clustering
Lenses	24	4	rời rạc	3	Decision-tree
Liver	345	7	số thực	2	k-NN
Mushroom	8124	22	rời rạc	2	clustering
Parkinson's Disease	756	24	số thực	2	classification
Pima	768	8	số nguyên, số thực	2	clustering
Ruspini	75	2	số nguyên	4	clustering
Soybean	307	35	rời rạc	4	clustering
Titanic	2200	11	rời rạc	2	clustering, classification
Votes	435	16	rời rạc	2	clustering
Wine	1599/4898	11+1	số thực	3	ANN
Wholesale Customers	440	8	số thực		clustering