



Bài 5

PHÂN TÍCH DỮ LIỆU ĐỊA LÝ

Ths. Trần Mạnh Trường
truongtm@gmail.com

Hà Nội – 2020





Các chủ đề nghiên cứu

1

Giới thiệu về GIS

2

Biểu diễn dữ liệu địa lý

3

Định vị và tham chiếu không gian

4

Thu thập dữ liệu không gian

5

Phân tích dữ liệu không gian

6

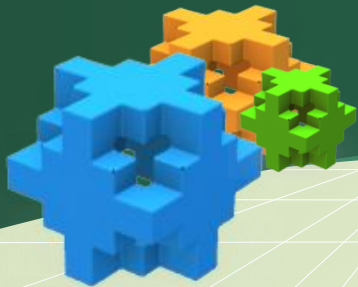
Hiển thị dữ liệu địa lý

7

Các chủ đề GIS nâng cao

8

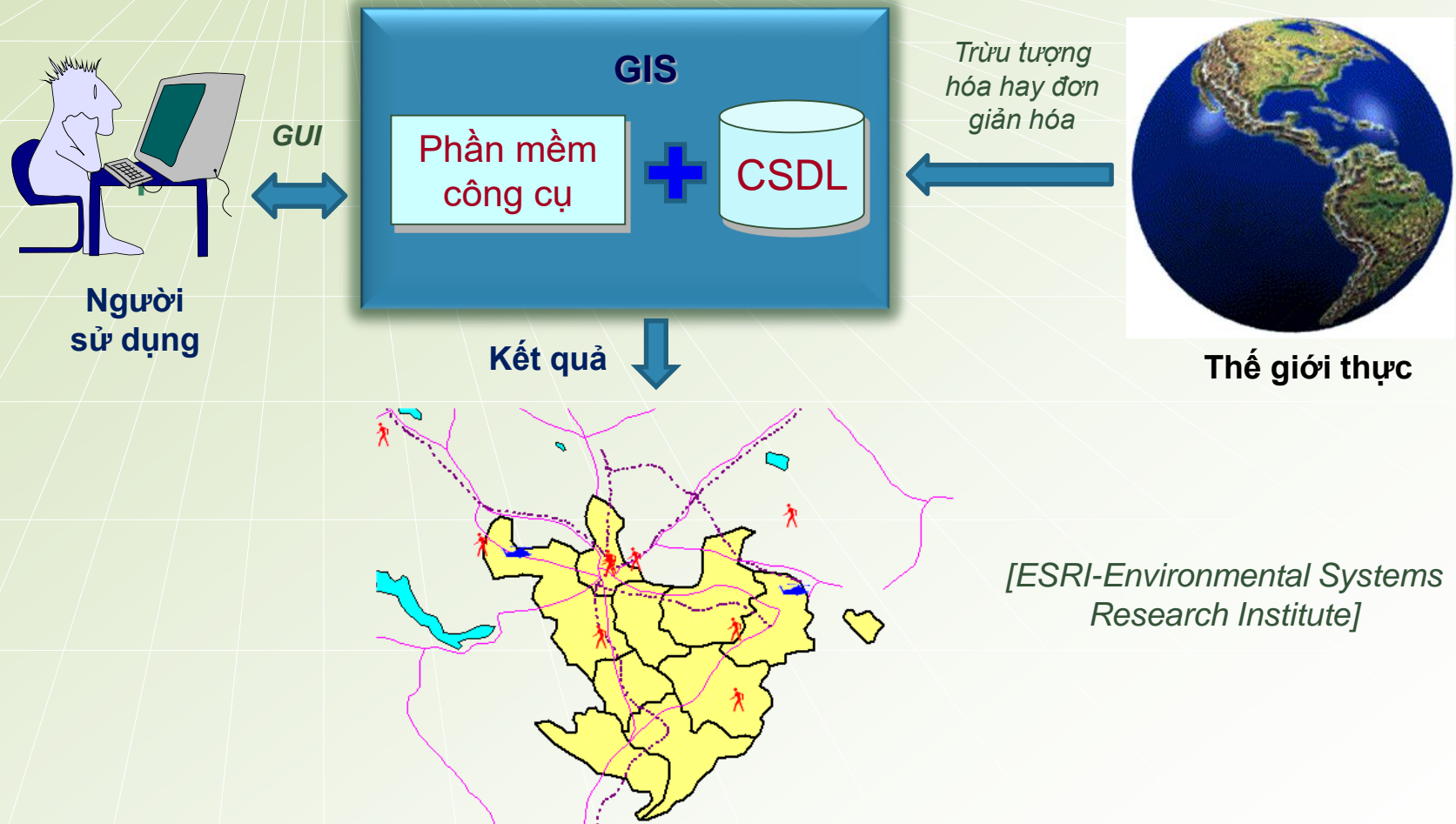
9

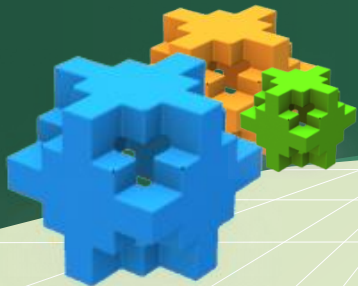


Nội dung

- ❖ Giới thiệu
- ❖ Khái niệm phân tích không gian
- ❖ Truy vấn CSDL địa lý
- ❖ Các thuật toán cơ sở
- ❖ Các thuật toán phân tích không gian
- ❖ Tổng kết bài

Nhắc lại kiến trúc GIS





1. Giới thiệu

- ❖ Các phương pháp truy vấn giúp người sử dụng tương tác với CSDL bằng chuột hay bàn phím. Kết quả được trình diễn theo các khung nhìn chuẩn
- ❖ Các phương pháp đo đạc xác định độ dài, diện tích, hình dạng, độ dốc và các thuộc tính khác của đối tượng
- ❖ Biến đổi dữ liệu không gian tạo ra các thông tin mới bằng cách chế tác hình học trên các đối tượng trong CSDL



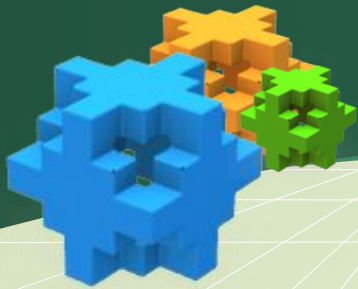
2. Phân tích không gian

- ❖ Khái niệm phân tích không gian hay được sử dụng thay cho phân tích địa lý
 - Các phương pháp, kỹ thuật nghiên cứu ở đây được áp dụng cho dữ liệu ở bất kỳ không gian nào, không chỉ trong không gian địa lý.
- ❖ Phân tích không gian là tiến trình biến đổi dữ liệu thô thành các thông tin hữu ích
 - Phân tích không gian bao gồm các phép biến đổi, chế tác và các phương pháp khác tác động trên dữ liệu địa lý để sinh ra giá trị mới, hỗ trợ ra quyết định và phát hiện các mẫu, dị thường mà con người không quan sát trực tiếp được.
- ❖ Một số phương pháp phân tích không gian được thực hiện thủ công, bằng thước đo... từ khi GIS chưa ra đời.



Phân tích không gian

- ❖ Các phương pháp phân tích không gian có thể rất phức tạp, nhưng có thể cũng rất đơn giản
- ❖ Các kỹ thuật có thể là
 - Sử dụng phương pháp toán học rất phức tạp
 - Mắt và bộ não người có thể nhận biết nhanh một số mẫu hay dị thường trong tập dữ liệu địa lý
- ❖ Giải pháp hiệu quả nhất là sử dụng kết hợp các phần mềm và phương pháp toán học phức tạp với khả năng của con người
 - Tuy nhiên phân tích không gian hiệu quả đòi hỏi sự thông minh của người sử dụng, không phải chỉ có máy tính mạnh.



Các loại phân tích không gian

❖ Truy vấn

- Là thao tác phân tích cơ bản nhất, trong đó GIS trả lời câu hỏi đơn giản của người sử dụng
- Ví dụ: *How many houses are found within 1 km of this point?*

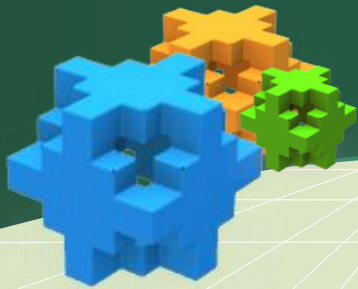
❖ Đo

- Các giá trị số mô tả các khía cạnh của dữ liệu địa lý
- Bao gồm, độ dài, diện tích, hình dạng của đối tượng và quan hệ giữa cặp đối tượng như khoảng cách, hướng.

❖ Biến đổi

- Sử dụng các quy tắc hình học, logic và số học để làm thay đổi tập dữ liệu như tổ hợp, so sánh để tạo ra tập mới
- Bao gồm cả thao tác biến đổi raster sang vector và ngược lại
- Tạo ra các fields từ tập đối tượng.

❖



Các loại phân tích không gian

❖

❖ Tóm tắt mô tả

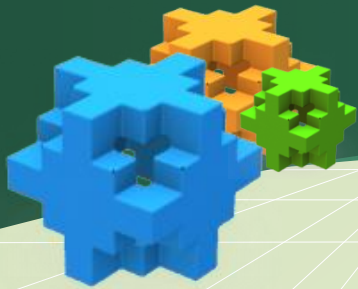
- Biểu diễn phần cơ bản của tập dữ liệu bởi 1 hoặc 2 numbers
- Tương đương với phân tích thống kê dữ liệu mô tả như lấy trung bình, sai số.

❖ Tối ưu hóa

- Lựa chọn vị trí lý tưởng cho đối tượng theo tiêu chí biết trước
- Ví dụ: Ứng dụng trong nghiên cứu thị trường, xây dựng công nghiệp...

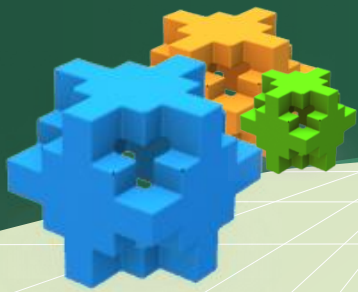
❖ Kiểm thử giả thuyết

- Tập trung vào suy diễn từ các mẫu kết quả để khái quát hóa
- Đây là vấn đề khó.



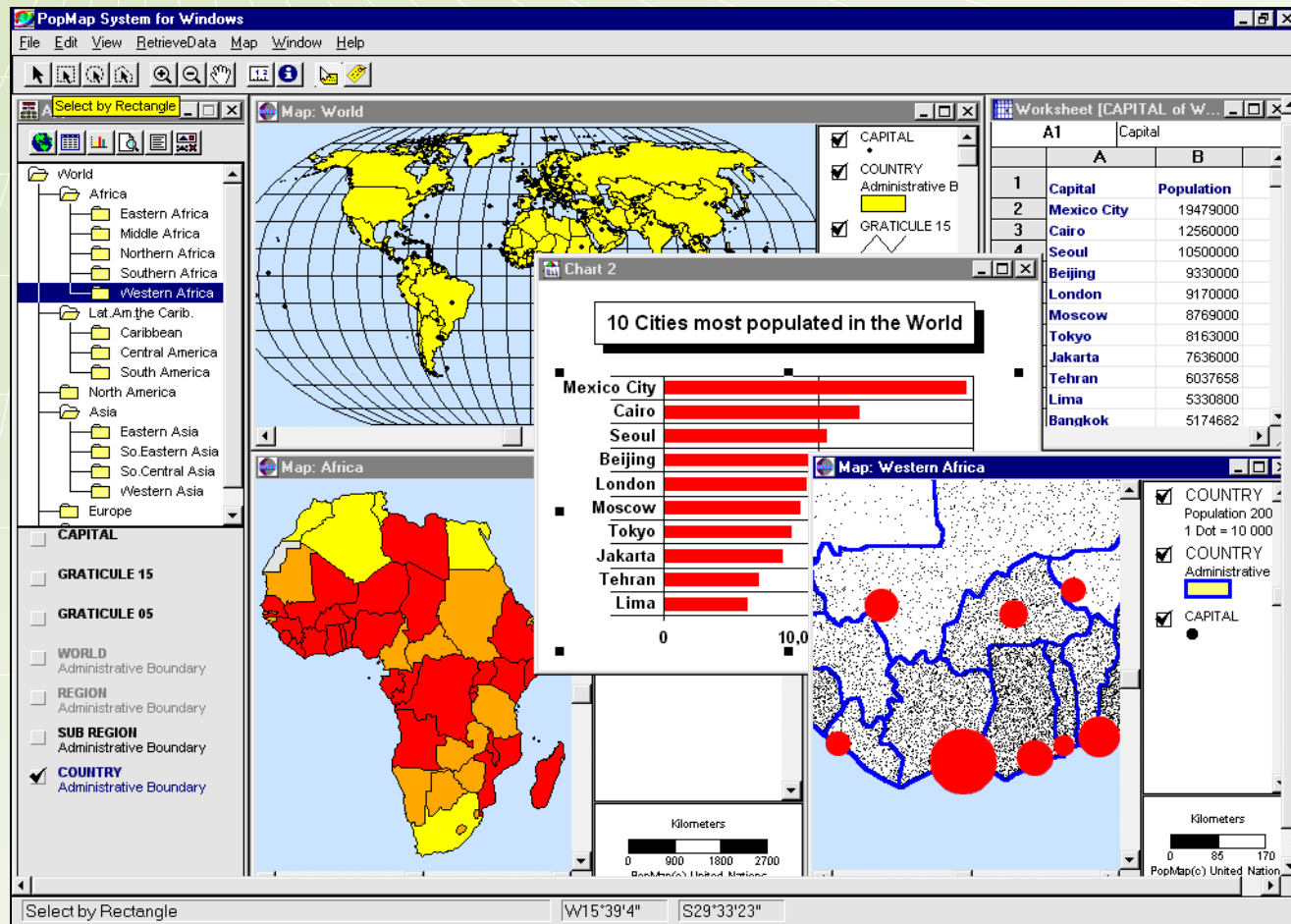
3. Truy vấn CSDL

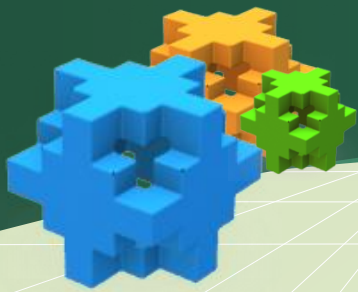
- ❖ Truy vấn là phép toán phân tích cơ sở, trong đó GIS trả lời các câu hỏi của người sử dụng
 - Tương tác với hệ thống có thể bằng thiết bị trỏ, bàn phím nhập câu hỏi, chọn thực đơn, nhấn Buttons để gửi câu truy vấn SQL đến CSDL
 - HCI tạo ra giao diện mạnh, ví dụ cho phép tương tác với hệ thống bằng tiếng nói... Hiệu quả khi điều hành xe cộ
- ❖ Loại truy vấn đơn giản nhất là tương tác giữa người sử dụng với các khung nhìn khác nhau, ví dụ:
 - Khung nhìn Catalog view hiển thị nội dung CSDL
 - Khung nhìn Map view hiển thị bản đồ để người sử dụng truy vấn vị trí trên bất cứ nơi đâu của bản đồ
 - Khung nhìn Table view hiển thị bảng dữ liệu thuộc tính gắn với đối tượng (điểm, đường, vùng)



Truy vấn CSDL

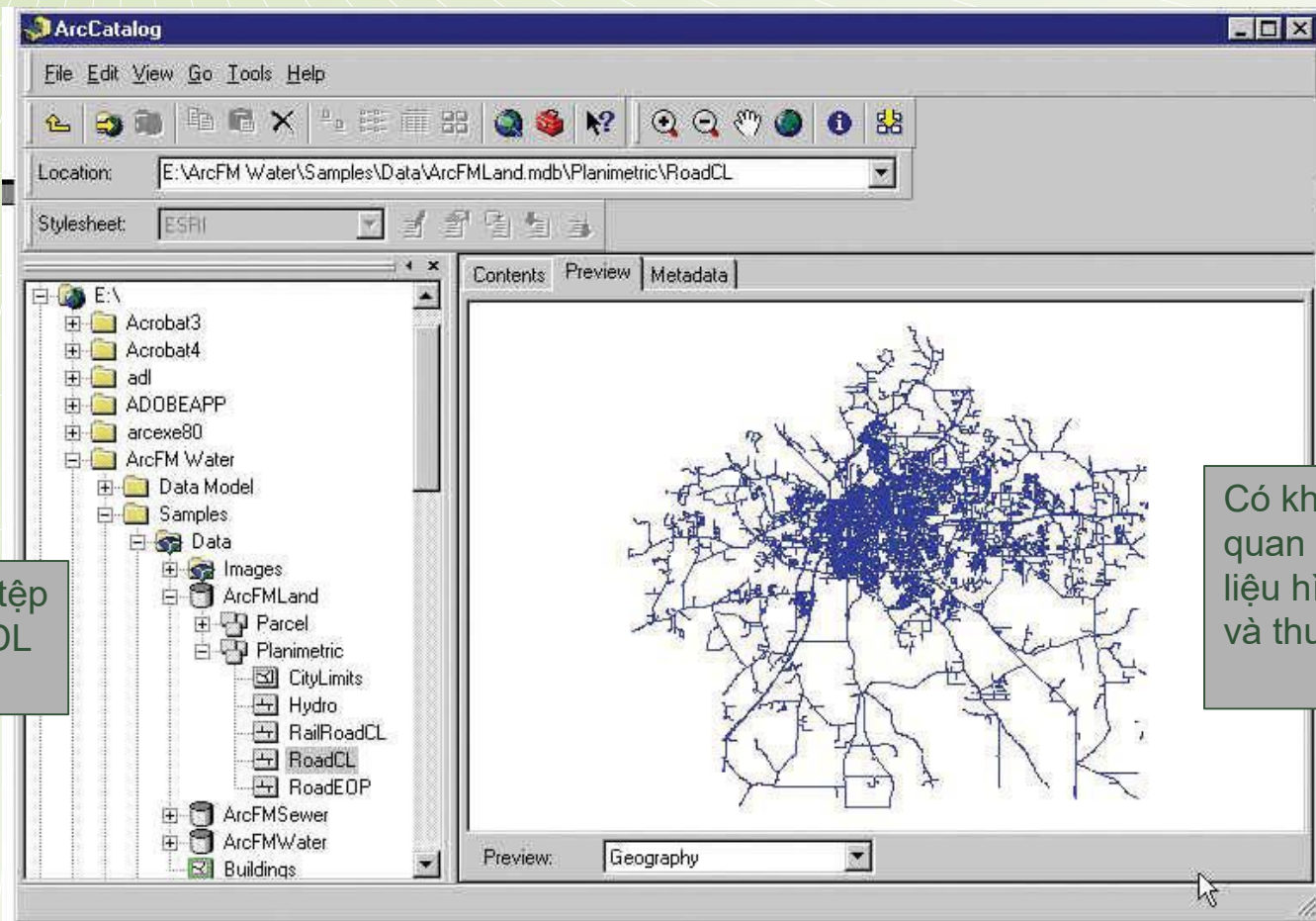
❖ Phần mềm PopMap (UNSTAT-IOIT)





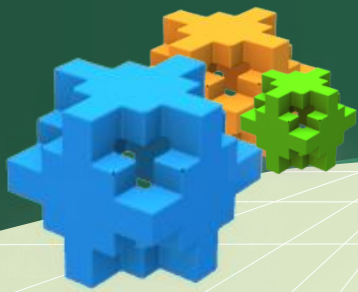
Truy vấn CSDL

❖ Phần mềm ArcCatalog của ESRI



Cấu trúc tệp
của CSDL

Có khả năng
quan sát dữ
liệu hình học
và thuộc tính

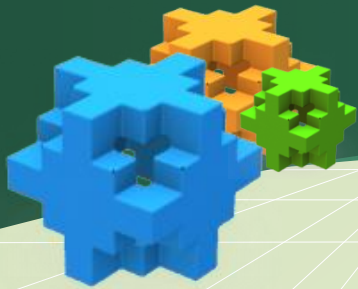


Truy vấn CSDL

❖ Phần mềm ArcMap của ESRI

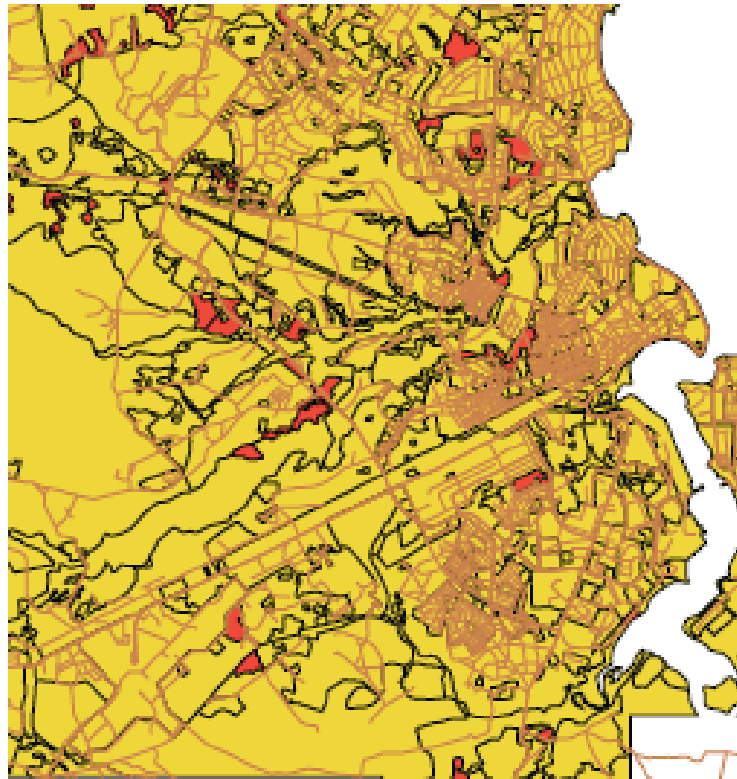
The screenshot displays the ArcMap application window titled 'Untitled - ArcMap'. The main map area shows a network of blue lines representing streets. The 'Layers' panel on the left shows the 'street_arc' layer is selected. The 'Attributes of street_arc' table is open, showing a list of records with columns: FID_1, Shape, NAME, Shape_Length, and Type. The table contains 23 records, with the first 12 highlighted in yellow. The status bar at the bottom indicates the current record is 13 out of 2102, with 2102 out of 2000 records selected.

FID_1	Shape	NAME	Shape_Length	Type
1	Polyline		136.500014433901	4
2	Polyline		267.503534104294	4
3	Polyline		403.25717951764	4
4	Polyline		806.22123346338	4
5	Polyline		403.006998608611	3
6	Polyline		799.93784833492	4
7	Polyline		800.741232233492	5
8	Polyline		53.400023458049	5
9	Polyline		398.754707416395	5
10	Polyline		123.104556726918	5
11	Polyline		95.6591882595931	5
12	Polyline		40.0053830742638	5
13	Polyline		135.607670712223	5
14	Polyline		90.6600280511209	5
15	Polyline		30.5051829119346	5
16	Polyline		91.3370072599947	5
17	Polyline		158.455728217768	5
18	Polyline		108.500000101048	5
19	Polyline		266.572276546268	5
20	Polyline		269.274931124029	5
21	Polyline		88.7588024628139	5
22	Polyline		134.601737966072	5
23	Polyline		82.8980220064282	5

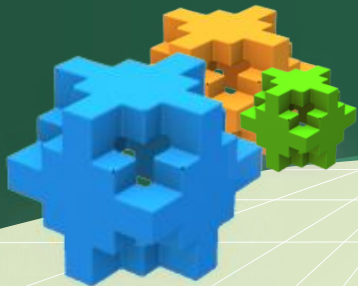


Truy vấn CSDL

- ❖ Ví dụ lựa chọn theo điều kiện thuộc tính
 - Hiển thị các đặc trưng thỏa mãn “*Area < 400.000 km² on land use areas in Dares Salaam, Tanzania*”



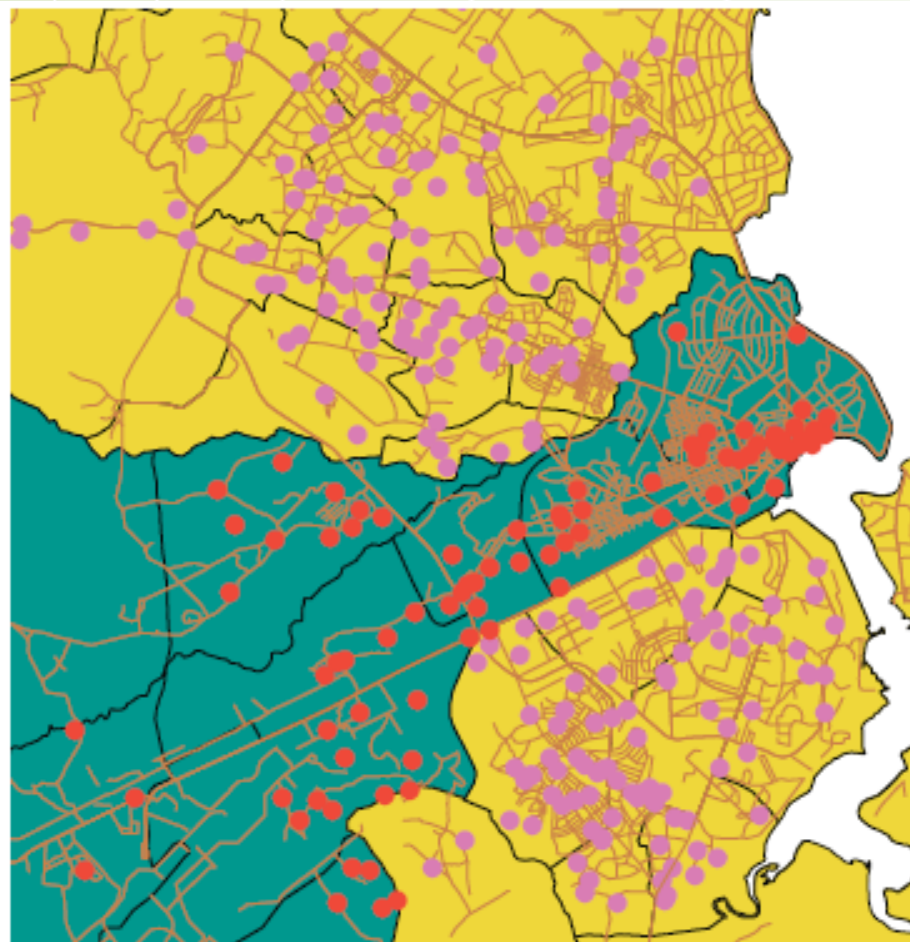
Area	IDs	Land Use
174308.7000	2	30
2066475.000	3	70
214582.5000	4	80
29313.8600	5	80
73328.0800	6	80
53303.3000	7	80
614530.1000	8	20
1637161.000	9	80
156357.4000	10	70
59202.2000	11	20
83289.5900	12	80
225642.2000	13	20
28377.3300	14	40
228930.3000	15	30
986242.3000	16	70

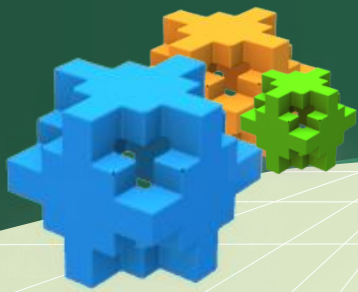


Truy vấn CSDL

❖ Ví dụ lựa chọn theo quan hệ topo

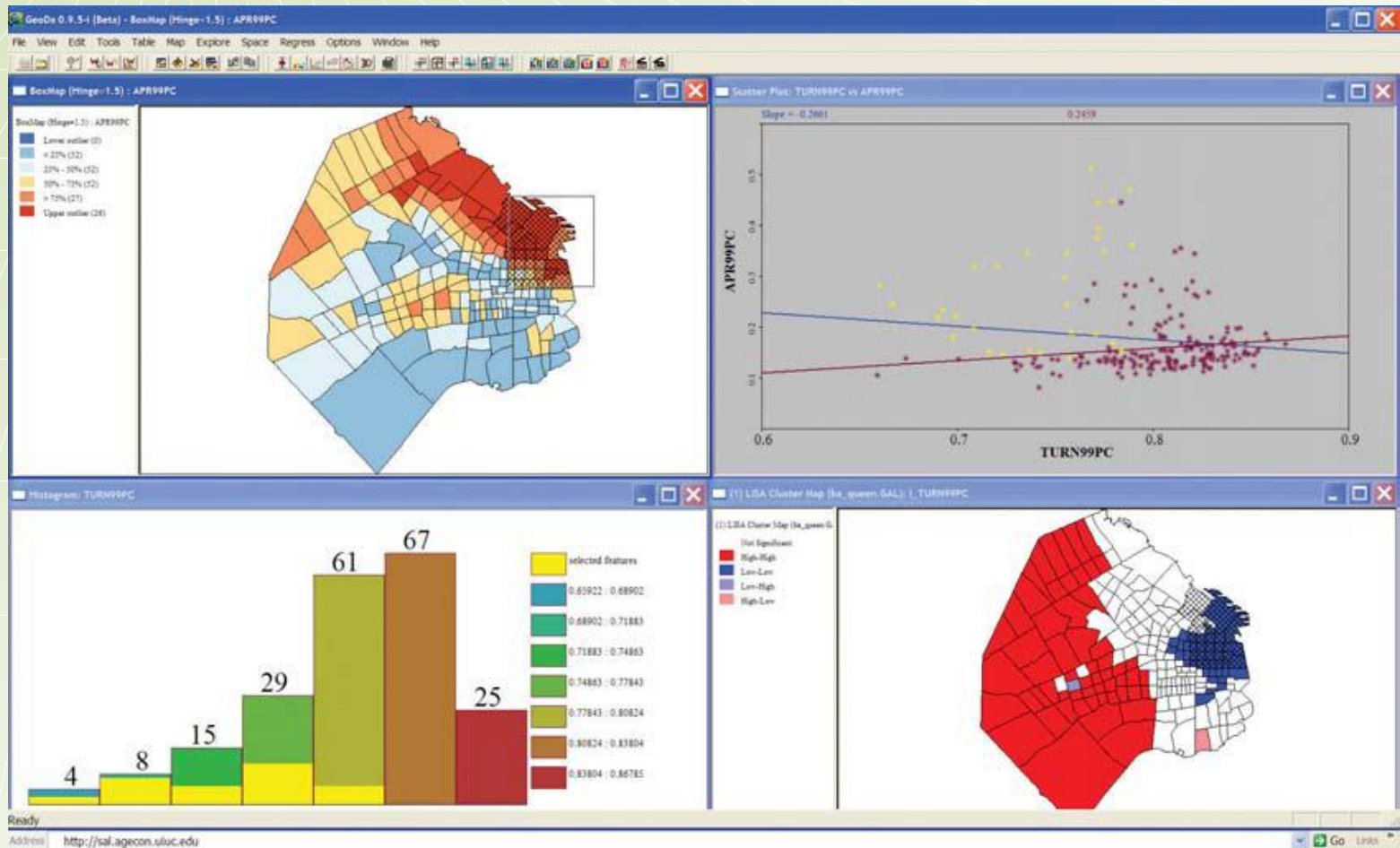
- Hiện thị đặc trưng nằm trong đối tượng chọn trước
- Hiện thị đặc trưng mà nó cắt đối tượng chọn trước
- Hiện thị đặc trưng liền kề đối tượng chọn trước
- Hiện thị đặc trưng trên cơ sở khoảng cách cho trước giữa chúng

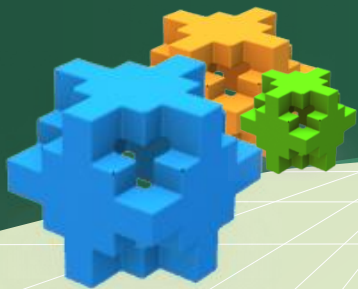




Truy vấn CSDL

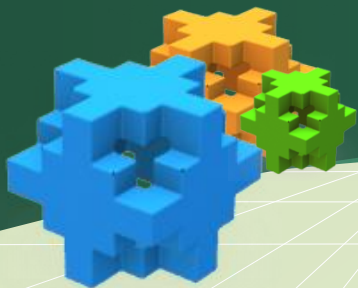
❖ Histogram view trong ArcMap của ESRI





4. Các thuật toán cơ sở

- ❖ Các thuật toán sử dụng trong các tiến trình phức tạp của GIS được hình thành từ các thuật toán đơn giản.
- ❖ Thuật toán tìm giao của hai đường thẳng
 - Ứng dụng trong GIS: xếp chồng đa giác, trộn, làm tan đa giác và đoạn thẳng, điểm trong đa giác, loại bỏ đa giác lạ...
 - Ví dụ: đường thẳng đi từ (4,2) đến (2,0) có giao với đoạn thẳng (0,4), (4,0)?
 - Giải pháp
 - phương trình đường thẳng qua 2 điểm $y=ax+b$, trong đó $b=(y_2-y_1)/(x_2-x_1)$
 - từ ví dụ trên có: $y=-2+x$ và $y=4-x$ giao điểm tại (3,1)
 - Tổng quát: nếu ta có $y=a_1+b_1x$ và $y=a_2+b_2x$
thì giao điểm sẽ ở tại: $x_i=-(a_1-a_2)/(b_1-b_2)$; $y_i=a_1+b_1x_i$
- ❖ Giao đoạn thẳng:
 - Đòi hỏi kiểm tra xem tọa độ giao đường thẳng có nằm trong các đoạn thẳng?



Các thuật toán cơ sở

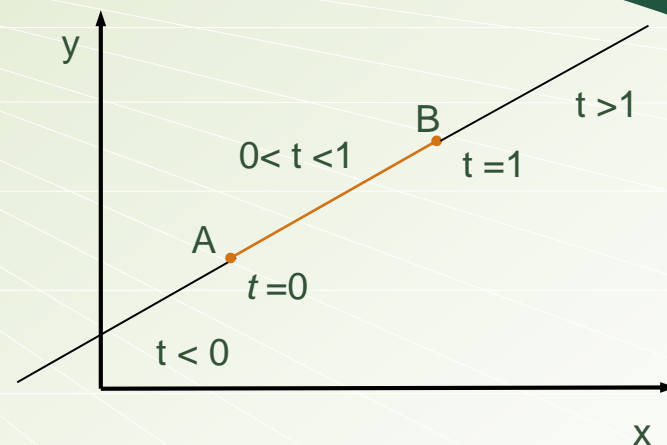
❖ Trong GIS chúng ta thường làm việc với đoạn thẳng thay cho đường thẳng

❖ Phương pháp biểu diễn đoạn thẳng bằng tham số:

- Đoạn thẳng 1 qua (x_A, y_A) và (x_B, y_B)
- Đoạn thẳng 2 qua (x_C, y_C) và (x_D, y_D)
- Giao của 2 đoạn thẳng tại t, s như sau:

$$t = \frac{(x_C - x_A)(y_C - y_D) - (x_C - x_A)(y_C - y_A)}{(x_B - x_A)(y_C - y_D) - (x_C - x_D)(y_B - y_A)}$$

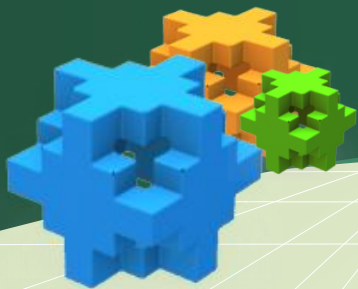
$$s = \frac{(x_B - x_A)(y_C - y_A) - (x_C - x_A)(y_B - y_A)}{(x_B - x_A)(y_C - y_D) - (x_C - x_D)(y_B - y_A)}$$



$$\begin{aligned} x &= x_A + t(x_B - x_A) \\ y &= y_A + t(y_B - y_A) \end{aligned}$$

$$\begin{aligned} x &= x_C + s(x_D - x_C) \\ y &= y_C + s(y_D - y_C) \end{aligned}$$

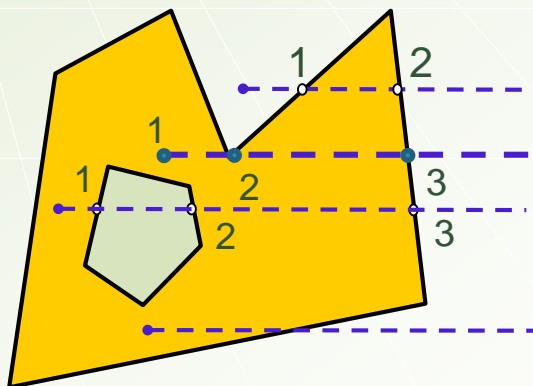
$$0 \leq t \leq 1 \text{ và } 0 \leq s \leq 1$$

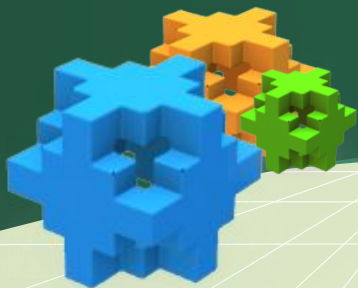


Các thuật toán cơ sở

❖ Điểm trong đa giác: Định lý nửa đường thẳng của Jordan

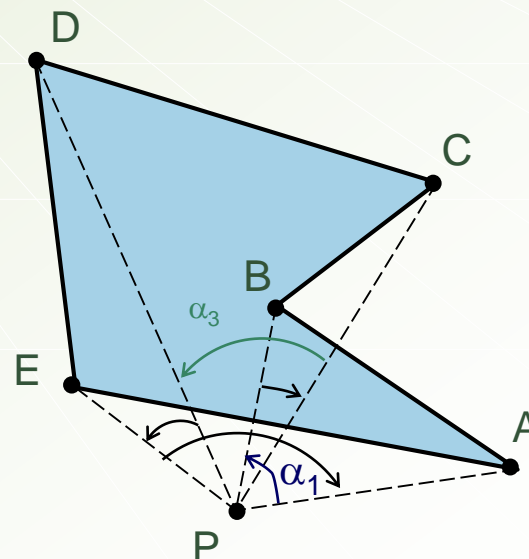
- Từ điểm cho trước, hãy vẽ tia ra ngoài tận cùng các cạnh đa giác
- Tính tổng giao điểm của tia với các cạnh đa giác
 - Nếu tổng số điểm là lẻ thì điểm đó nằm trong đa giác, ngược lại tổng số điểm chẵn thì điểm nằm ngoài đa giác
- Chú ý khi tia đi qua đỉnh đa giác hay trùng với cạnh đa giác.





Các thuật toán cơ sở

- ❖ Điểm trong đa giác: Phương pháp kiểm tra góc
 - Ví dụ, xét điểm P cho trước có ở trong đa giác ABCDE?
 - Từ điểm P nối với các đỉnh đa giác để tạo thành các góc theo thứ tự ngược chiều kim đồng hồ
 - Các góc này có giá trị dương hoặc âm tùy theo hướng đo
 - Tính tổng các góc
 - Nếu tổng các góc bằng 0 thì P nằm ngoài đa giác
 - Nếu tổng các góc bằng 360° thì P nằm trong đa giác

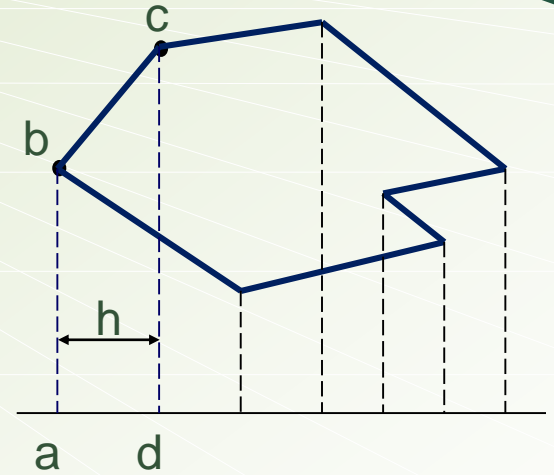


5. Đo đạc

❖ Tính diện tích đa giác

- Giả sử các đỉnh đa giác xếp đặt theo chiều quay của kim đồng hồ
- Vẽ các đường vuông góc từ đỉnh đa giác xuống trục nằm ngang để tạo các hình thang
 - Chiều cao hình thang bằng hiệu hai tọa độ x của hai đỉnh liên tiếp. Chúng có giá trị âm hoặc dương.
- Diện tích đa giác sẽ bằng tổng diện tích của các hình thang này

$$A_p = \sum_{i=1}^{n-1} (x_{i+1} - x_i) \frac{(y_i - y_b) + (y_{i+1} - y_b)}{2}$$



Vì y_b có giá trị bất kỳ kể cả 0, cho nên diện tích hình thang sẽ được tính như sau:

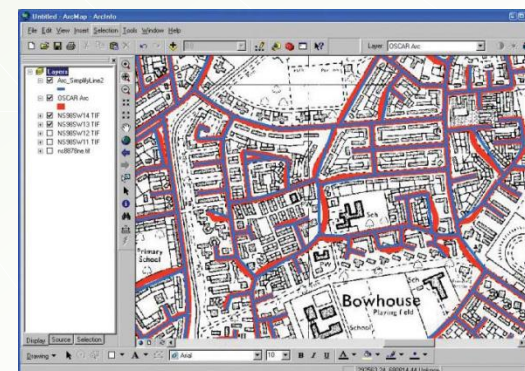
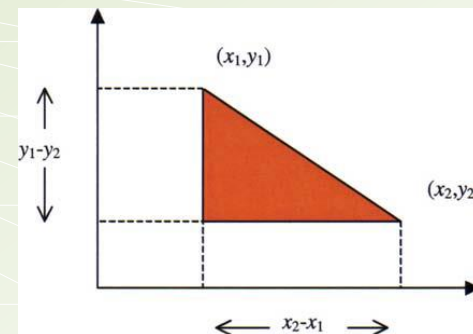
$$A_p = \frac{1}{2} \left[(x_n y_1 - x_1 y_n) + \sum_{i=1}^{n-1} (x_i y_{i+1} - x_{i+1} y_i) \right]$$

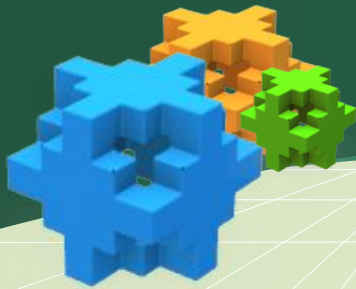


Đo đạc

❖ Khoảng cách và độ dài

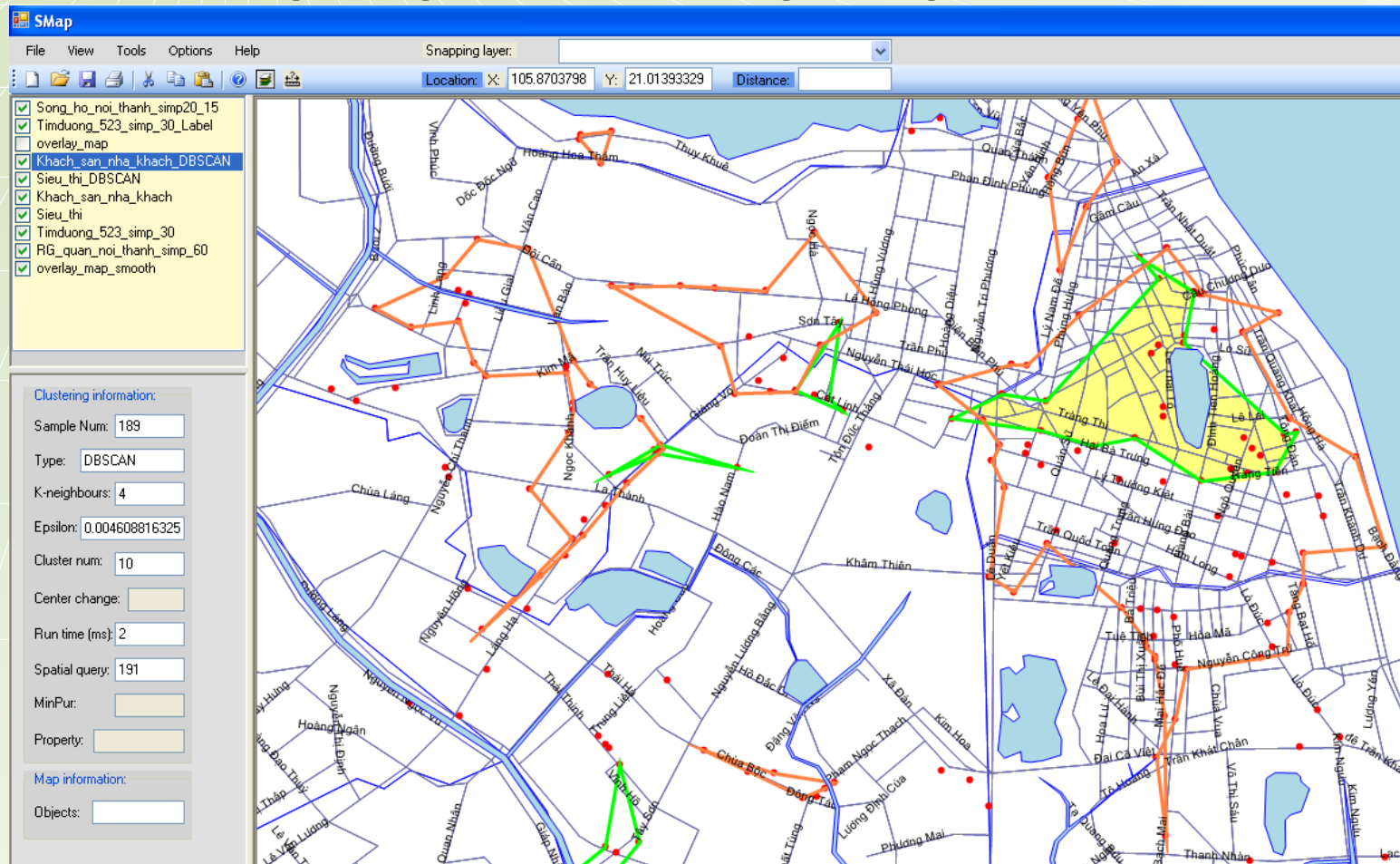
- Sử dụng độ đo Pythagore để tính khoảng cách hai điểm trên mặt phẳng. Không làm việc với kinh, vĩ độ
- Sử dụng phương pháp tính khoảng cách trên vòng tròn lớn với đơn vị đo kinh, vĩ độ (Bài 3)
- Lý do sai số tính độ dài
 - Xấp xỉ đường phố bằng các đoạn thẳng
 - Sai số của phép chiếu khi tính toán khoảng cách trên mặt phẳng.



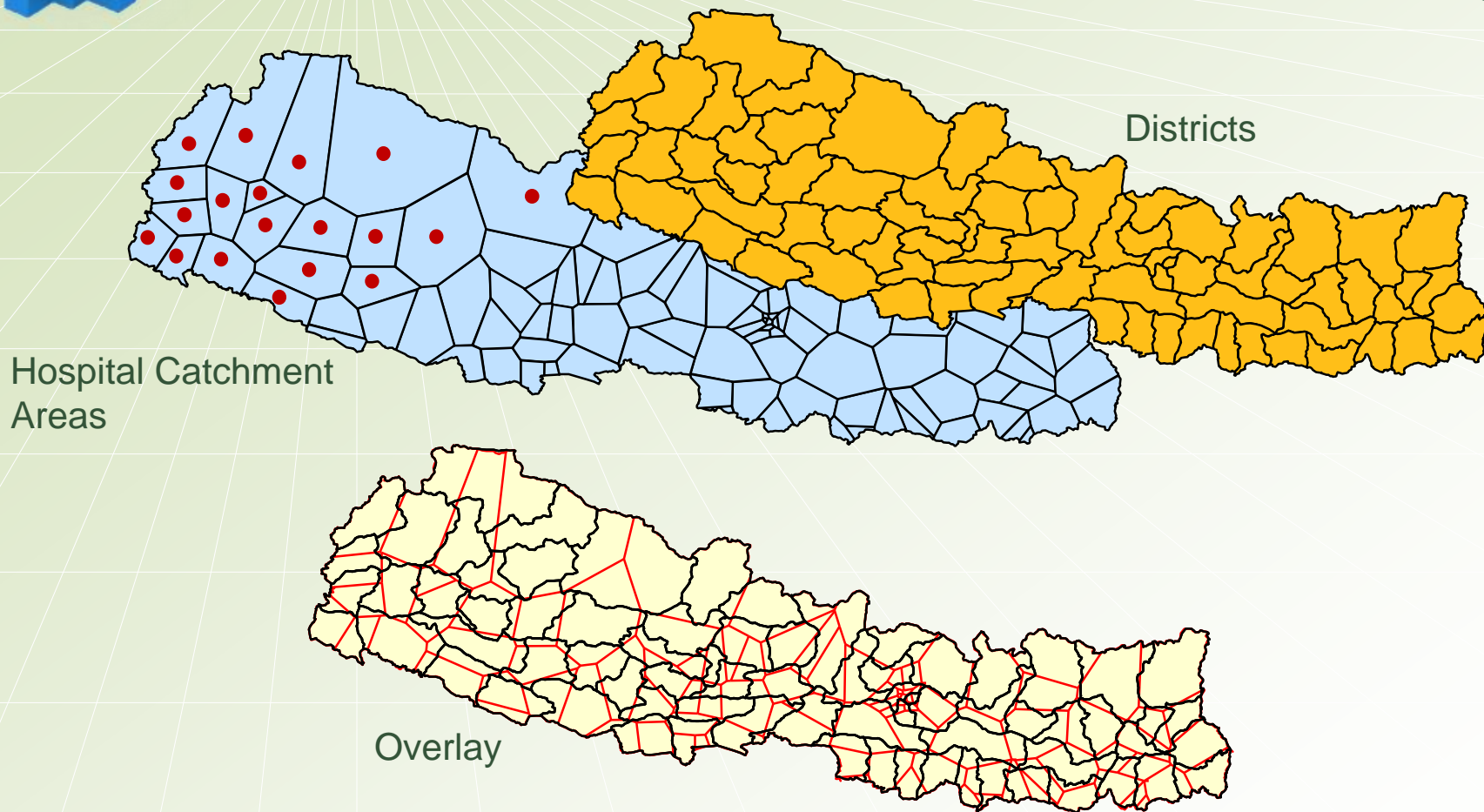
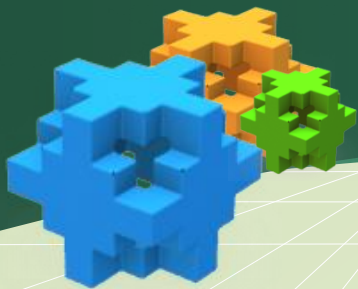


Biến đổi

❖ Xếp chồng đa giác: Ví dụ ứng dụng



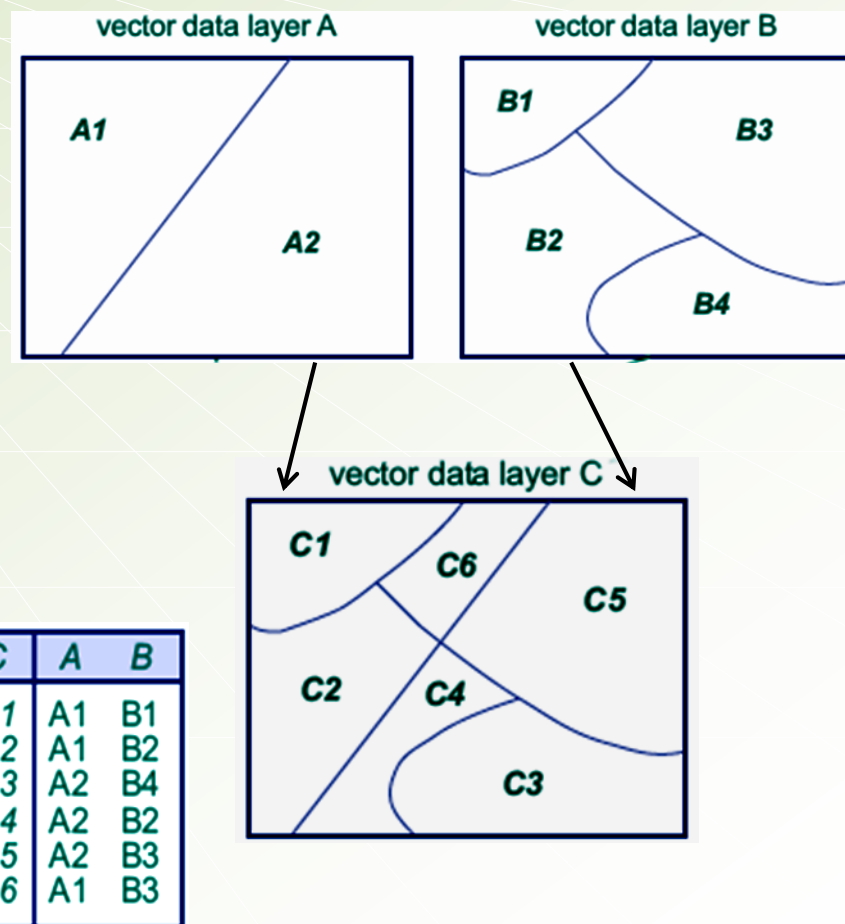
Ví dụ Overlay



Biến đổi

❖ Xếp chồng (Overlaying) đa giác

- Xây dựng các đối tượng vùng mới từ hai vùng cho trước.
- Các đa giác mới là phần chung hay phần riêng của các đa giác cho trước.
- Thuật toán là khá phức tạp.





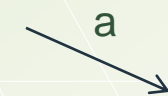
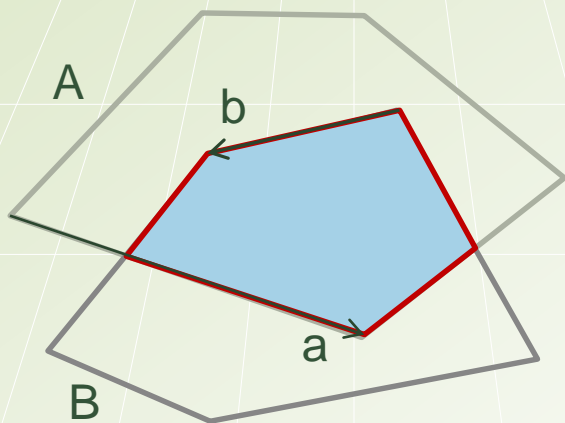
Biến đổi

❖ Thuật toán tìm giao của hai đa giác

■ Kiểm tra xem hai đa giác có giao nhau?

- Hai đa giác đơn P và Q cắt nhau nếu thỏa 1 trong các điều kiện sau:
 - Một cạnh của P cắt 1 cạnh của Q
 - P nằm trong Q
 - Q nằm trong P

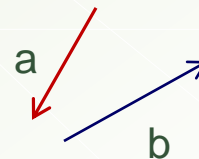
■ Tìm giao của hai đa giác lồi



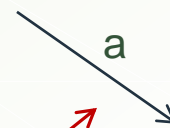
(a)



(b)

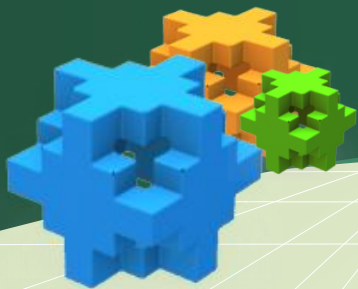


(c)



(d)

» Trường hợp (b) chọn b vì theo quy tắc ngược chiều kim đồng hồ



Các thuật toán phân tích

❖ Giao của hai đa giác lồi (tt)

■ Thuật toán

- Giả thiết cạnh đa giác sắp xếp theo thứ tự ngược chiều kim đồng hồ.
- Thuật toán duyệt theo các cạnh đa giác, bắt đầu từ cặp cạnh a, b và tiếp tục với cạnh có xu thế cắt cạnh của đa giác kia (các cạnh màu đỏ trong trang trước)

CONVEX_POLY_INTER(a, b): a convex polygon

begin

Let a and b be the first edge of A and B

repeat

if (a intersect b) **then**

Report the intersection $a \cap b$

Memorize which boundary is "inside"

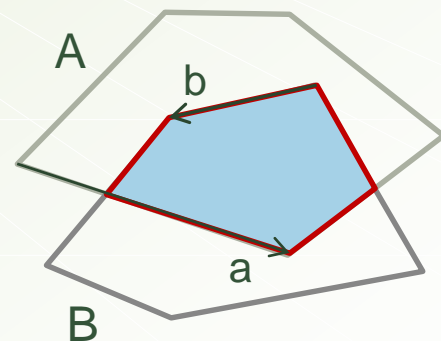
endif

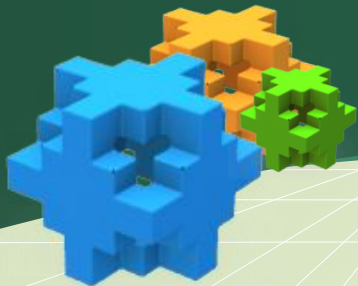
Advance on A or B (report the points on the inner boundary)

until (A and B have been scanned)

if no intersection has been found, test $A \subset B$ and $B \subset A$

end



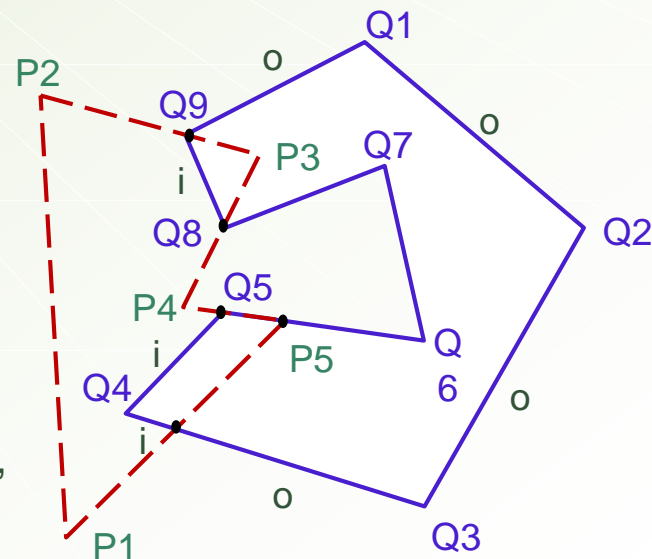


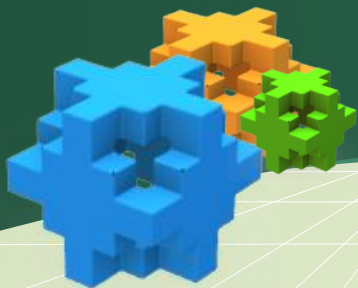
Các thuật toán phân tích

- ❖ Giao của hai đa giác bất kỳ (Clamer Schutte, Đại học Delft, Hà lan)
 - Cho trước hai đa giác P và Q không có lỗ hổng, không tự cắt và đỉnh của chúng được sắp xếp theo chiều kim đồng hồ
 - Hãy tìm đa giác thuộc tập $P \cup Q$, $P \setminus Q$ và $Q \setminus P$
 - Các bước của thuật toán
 - Phân lớp các đỉnh của hai đa giác vào 2 danh sách: gán vào mỗi đỉnh giá trị **i**(nside), **o**(utside) hay **b**(oundary) phụ thuộc vào vị trí của nó so với đa giác kia

$P_v = \langle (P1, o), (P2, o), (P3, i), (P4, o), (P5, b) \rangle$

$Q_v = \langle (Q1, o), (Q2, o), (Q3, o), (Q4, i), (Q5, b), (Q6, o), (Q7, o), (Q8, b), (Q9, b) \rangle$





Các thuật toán phân tích

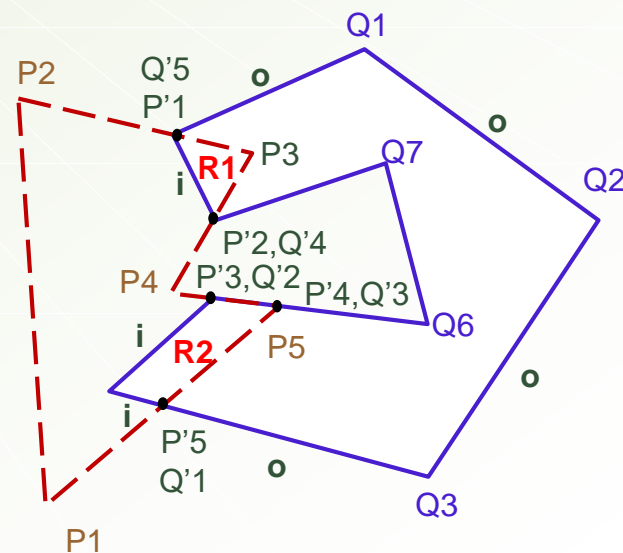
❖ Giao của hai đa giác bất kỳ (tt)

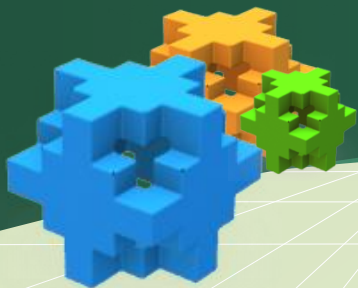
- Tìm giao của các cạnh của hai đa giác P, Q.
- Mỗi giao điểm được xen vào Pv hay Qv và đánh dấu **b**(oundary) để có danh sách mới

$P_{ef} = \langle (P1, o), (P2, o), (P'1, b), (P3, i), (P'2, b), (P4, o), (P'3, b), (P5, b), (P'4, b), (P'5, b) \rangle$
 $Q_{ef} = \langle (Q1, o), (Q2, o), (Q3, o), (Q'1, b), (Q4, i), (Q5, b), (Q'2, b), (Q'3, b), (Q6, o), (Q7, o), (Q8, b), (Q'4, b), (Q'5, b), (Q9, b) \rangle$

- Lựa chọn các cạnh mới cho đa giác kết quả

- Cạnh đa giác mới là một phần cạnh đa giác gốc, nằm hoàn toàn trong hay hoàn toàn ngoài đa giác kia
- Ví dụ: chọn các đoạn nằm trong khi đầu cuối nằm trong, nếu cả hai đầu mút nằm trên cạnh đa giác thì kiểm tra điểm giữa của nó xem có nằm trong đa giác?





Các thuật toán phân tích

❖ Giao của hai đa giác bất kỳ (tt)

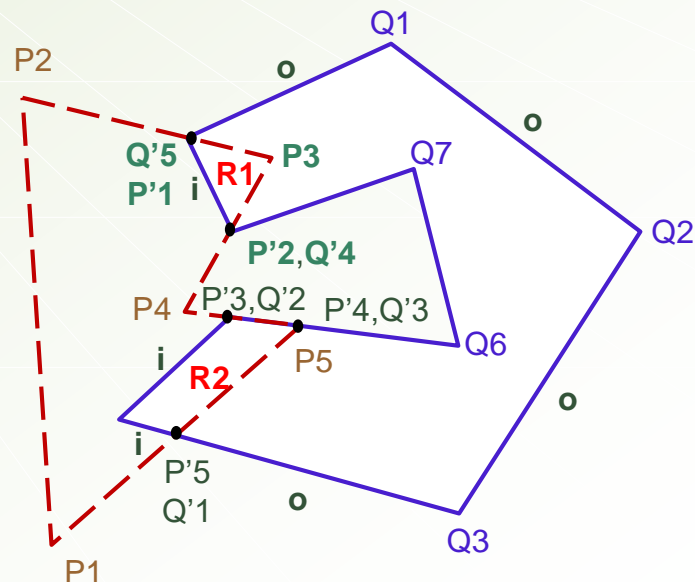
■ Danh sách Pef và Qef:

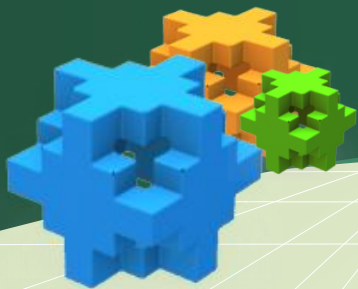
Pef = $\langle (P1, o), (P2, o), (P'1, b), (P3, i), (P'2, b), (P4, o), (P'3, b), (P5, b), (P'4, b), (P'5, b) \rangle$

Qef = $\langle (Q1, o), (Q2, o), (Q3, o), (Q'1, b), (Q4, i), (Q5, b), (Q'2, b), (Q'3, b), (Q6, o), (Q7, o), (Q8, b), (Q'4, b), (Q'5, b), (Q9, b) \rangle$

■ Tách các đa giác kết quả

- Tách từng đa giác kết quả. Bắt đầu từ đoạn bất kỳ có dấu i hay b sau đó tìm trong 2 danh sách các đoạn tiếp theo có đầu mút khớp với điểm cuối của cạnh trước. Lặp cho đến khi trở lại điểm ban đầu
- Ví dụ: Hãy bắt đầu từ đoạn $(P'1, P3)$ trong danh sách Pef. Sau đó tìm đoạn (i, b) có P3 là điểm mút, ta thấy $(P3, P'2)$. Tiếp theo từ P'2, ta tìm trong Qef thấy đoạn Q'4Q'5. Hình thành đa giác R1.

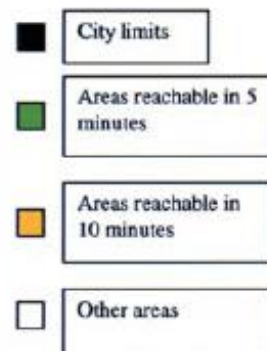
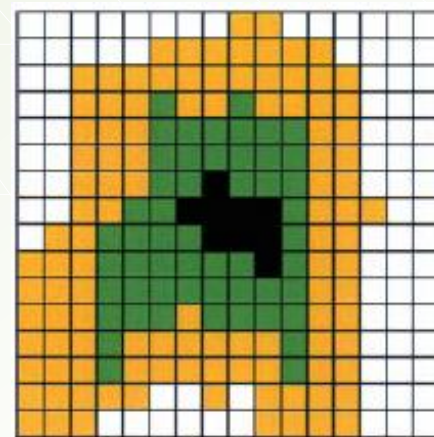
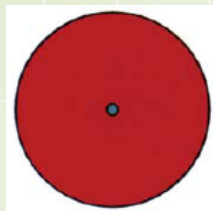
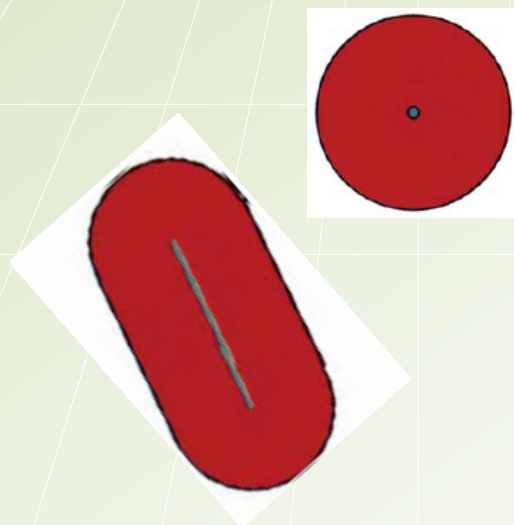




Các thuật toán phân tích

❖ Xây dựng vùng đệm (*Buffering*)

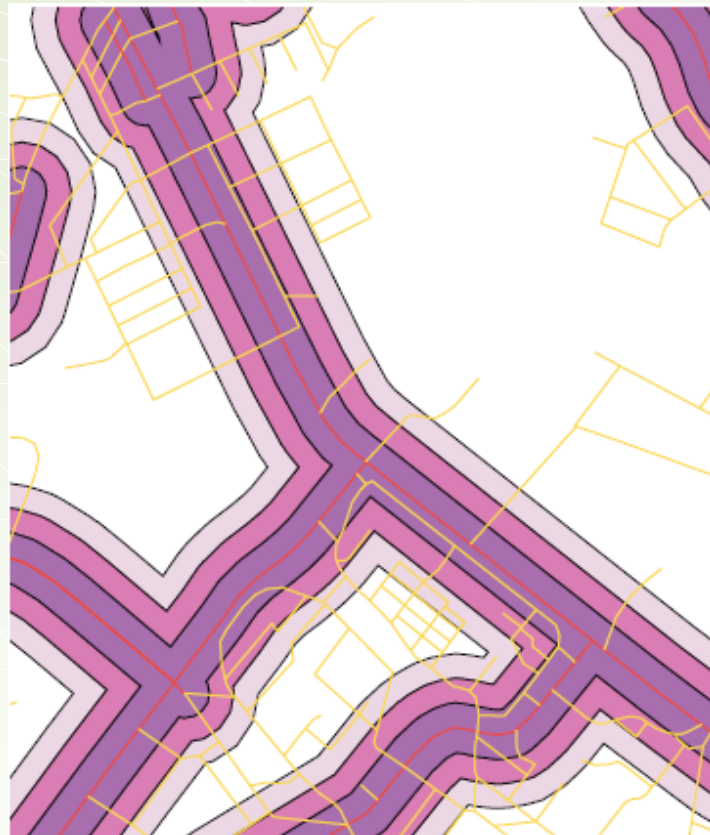
- Thao tác Buffer là một trong những thao tác quan trọng nhất của GIS
- Áp dụng thao tác *buffer* cho cả raster và vector GIS
- Cho trước đối tượng điểm, dạng đường hay vùng, nhiệm vụ là xây dựng đối tượng mới tạo bởi vùng mà mọi điểm của nó có khoảng cách nhất định tới đối tượng gốc.

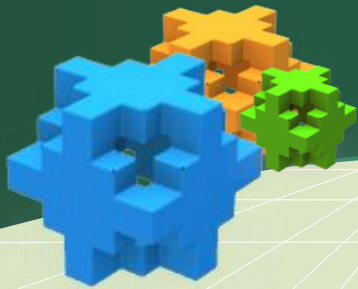




Các thuật toán phân tích

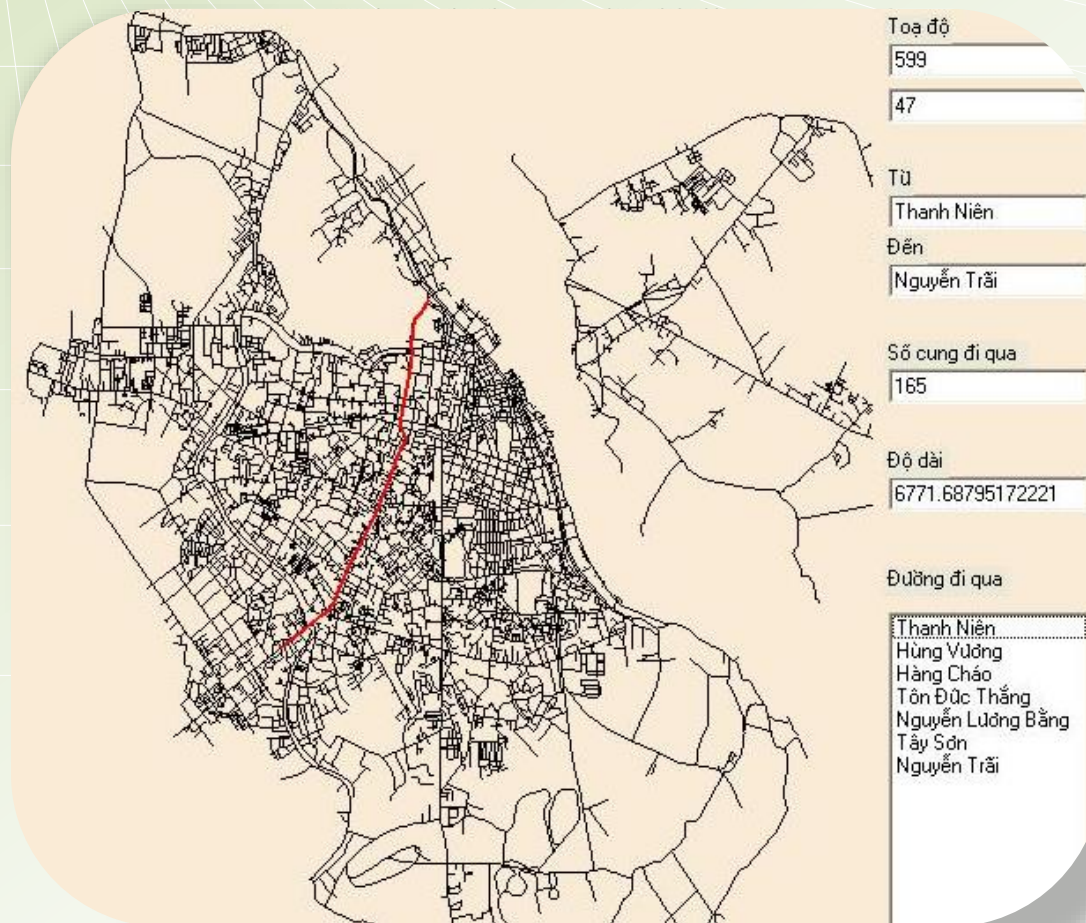
❖ Ví dụ Buffering





Các thuật toán phân tích

❖ Tìm đường đi ngắn nhất

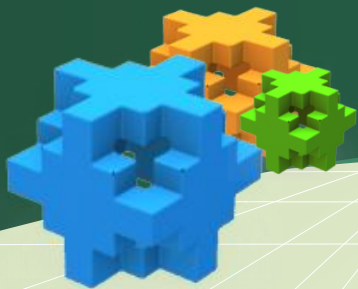




Các thuật toán phân tích

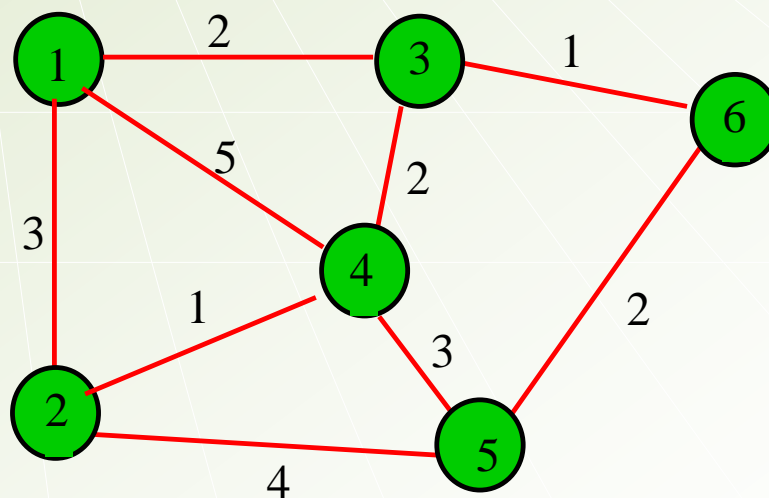
❖ Thuật toán tìm đường đi ngắn nhất

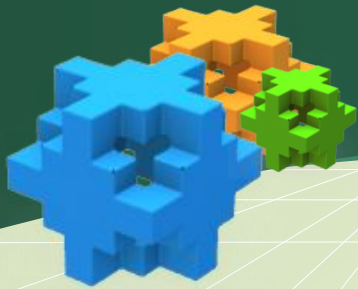
- Chuyển bài toán tìm đường đi ngắn nhất trên bản đồ thành bài toán tìm đường đi ngắn nhất giữa hai đỉnh của một đồ thị liên thông.
- Chuyển định dạng bản đồ sang đồ thị: $G = (V, E)$
 - Mỗi điểm sẽ được chuyển thành một đỉnh. Mỗi đường là một cung trong đồ thị. Mỗi cung có một giá trị (trọng số)
- Các thuật toán tiêu biểu
 - Thuật toán Ford-Bellman
 - Áp dụng trong trường hợp tìm đường đi từ một đỉnh tới tất cả các đỉnh còn lại trong đồ thị, trọng số của cung có thể âm
 - Thuật toán Dijkstra
 - Tìm đường đi tối ưu từ đỉnh s đến các đỉnh còn lại của đồ thị.
 - Thuật toán A^*
 - Tìm một đường đi từ một nút khởi đầu tới một nút đích cho trước.



Ví dụ thuật toán Dijkstra

- ❖ Tiệm cận: Greedy
- ❖ Input: Đồ thị có trọng số $G=\{E,V\}$, đỉnh khởi đầu $v \in V$, trọng số của các cạnh đồ thị không âm
- ❖ Output: Độ dài của đường đi ngắn nhất từ đỉnh khởi đầu $v \in V$ đến mọi đỉnh còn lại của đồ thị
- ❖ Độ phức tạp về thời gian là $O(|V|^2 + |E|)$





Ví dụ thuật toán Dijkstra

❖ Pseudocode

```
dist[s] ← 0  
for all v ∈ V - {s}  
do dist[v] ← ∞  
S ← ∅  
Q ← V  
while Q ≠ ∅  
do u ← mindistance(Q, dist)  
S ← S ∪ {u}  
for all v ∈ neighbors[u]  
do if dist[v] > dist[u] + w(u, v)  
then dist[v] ← dist[u] + w(u, v)  
return dist
```

(distance to source vertex is zero)

(set all other distances to infinity)

(S, the set of visited vertices is initially empty)

(Q, the queue initially contains all vertices)

(while the queue is not empty)

(select the element of Q with the min. distance)

(add u to list of visited vertices)

(if new shortest path found)

(set new value of shortest path)

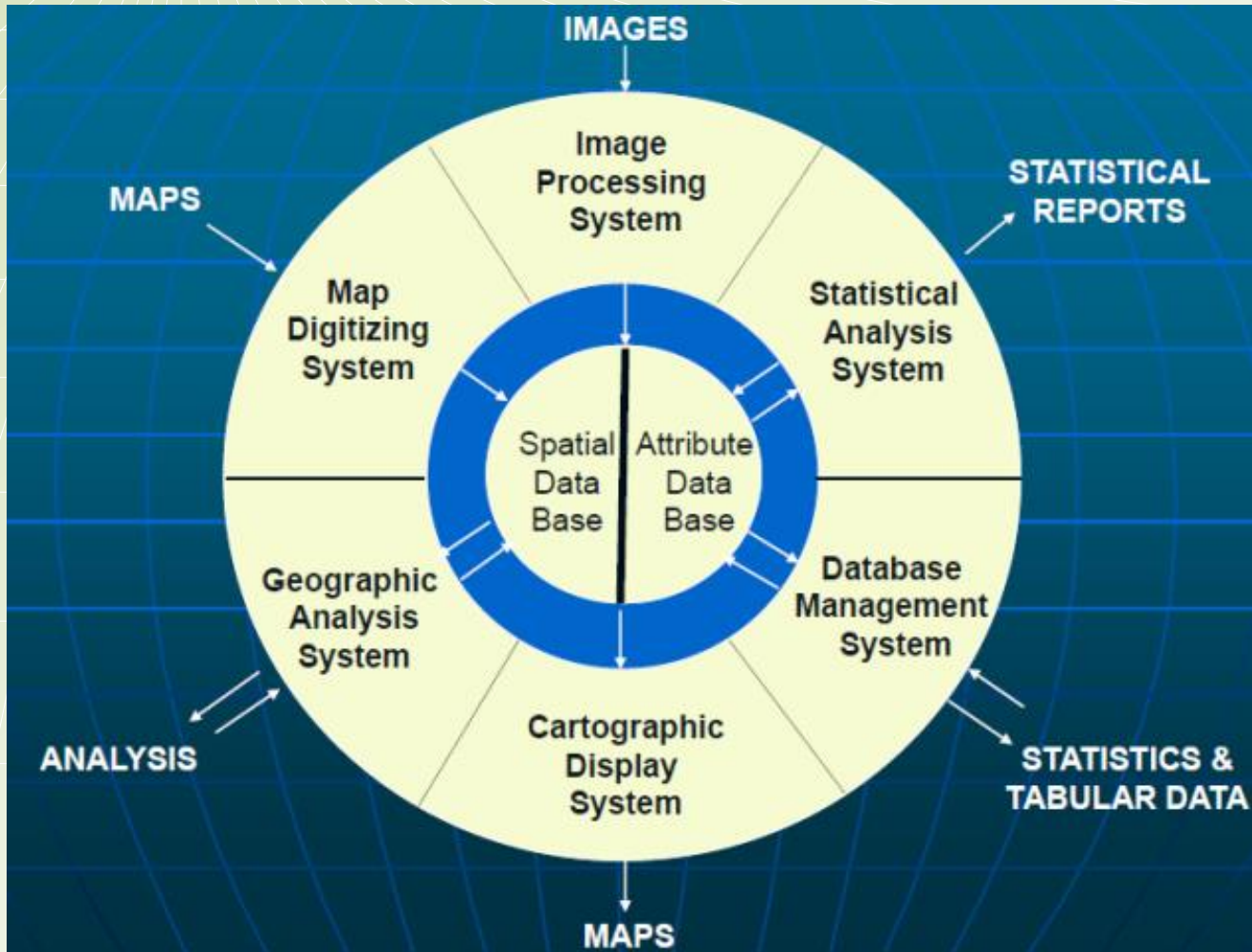
❖ Demo



7. Tổng kết bài

- ❖ Định nghĩa phân tích địa lý
- ❖ Khả năng truy vấn của GIS
- ❖ Các phương pháp đo độ dài, diện tích
- ❖ Biến đổi để tạo ra đối tượng mới và xác định quan hệ hình học giữa các đối tượng
- ❖ Tìm đường đi ngắn nhất

Tổng kết bài



Câu hỏi?