Tutorial 7: Spatial Data Structure

INFS3200/7907
Advanced Database Systems

Question

- How is a spatial object (e.g. a square, triangle or rectangle) represented numerically, stored and managed to a database ?
- We need algorithms such as Quadtree to transform spatial objects to numerical representations that are stored into the database.

Quadtree

A **quadtree**

- is a tree data structure in which each internal node has up to four children.
- is most often used to partition a two dimensional space by recursively subdividing it into four quadrants.

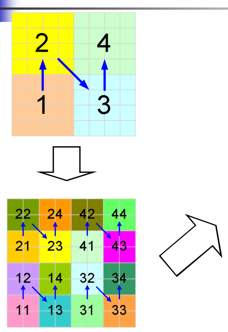
Question1

Consider the following objects in a space. The space is subdivided into 64 regions

I							H
						G	
					F		
				E			
			D				
		C					
	B						
A							

Question1

Consider the following objects in a space. The space is subdivided into 64 regions



222	224	242	244	422	424	442	444
221	223	241	243	421	423	441	443
212	214	232	234	412	414	432	434
211	213	231	233	411	413	431	433
122	124	142	144	322	324	342	344
121	123	141	143	321	323	341	343
112	114	132	134	312	314	332	334
111	113	131	133	311	313	331	333

Q1

a) Encoding the regions using Morton order codes
b) Order the objects by Morton order

I 222	224	242	244	422	424	442	H 444
221	223	241	243	421	423	G 441	443
212	214	232	234	412	F 414	432	434
211	213	231	233	E 411	413	431	433
122	124	142	D 144	322	324	342	344
121	123	C 141	143	321	323	341	343
112	B 114	132	134	312	314	332	334
A 111	113	131	133	311	313	331	333

A: 111
B: 114
C: 141
D: 144
I: 222
E: 411
F: 414
G: 441
H: 444

Q1c

Are there any pair of objects adjacent in the list which are not near each order in space

I 222	224	242	244	422	424	442	H 444
221	223	241	243	421	423	G 441	443
212	214	232	234	412	F 414	432	434
211	213	231	233	E 411	413	431	433
122	124	142	D 144	322	324	342	344
121	123	C 141	143	321	323	341	343
112	B 114	132	134	312	314	332	334
A 111	113	131	133	311	313	331	333

A: 111, B: 114, C: 141, D: 144, I: 222, E: 411, F: 414, G: 441, H: 444

D144 and I222 are far apart spatially, as are I222 and E411

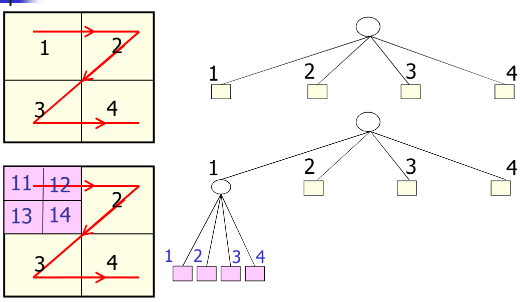
Q1d

You can tell by looking at only the immediately adjacent objects in the list that there are no objects geographically adjacent to object I. How?

I 222	224	242	244	422	424	442	H 444
221	223	241	243	421	423	G 441	443
212	214	232	234	412	F 414	432	434
211	213	231	233	E 411	413	431	433
122	124	142	D 144	322	324	342	344
121	123	C 141	143	321	323	341	343
112	B 114	132	134	312	314	332	334
A 111	113	131	133	311	313	331	333

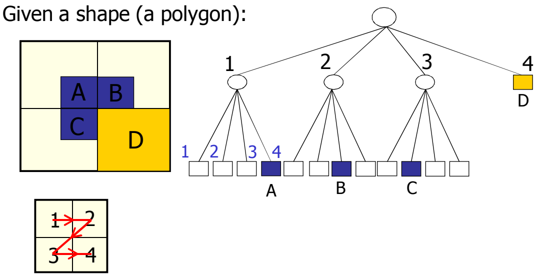
The geographical *neighbours* are 221, 223 and 224, none of which are in the list: A: 111, B: 114, C: 141, D: 144, I: 222, E: 411, F: 414, G: 441, H: 444

Quadtree – Mapping a spatial object to a quadtree

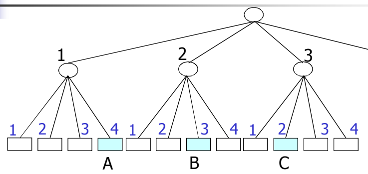


Q2- Find a quadtree representation

Given a shape (a polygon):



Q2 (b)

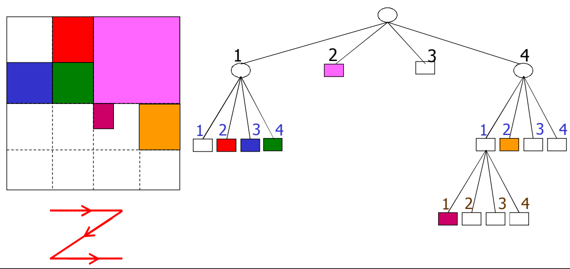


base 5

A	$(1\ 4)_5$
B	$(2\ 3)_5$
C	$(3\ 2)_5$
D	$(4)_5$

Q3

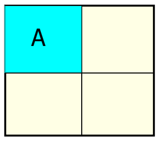
Draw the shape described by the following location codes:
12, 13, 14, 2, 411, 42
Assume that they are concatenation of base 5 digits.



Q4 (a)

A location code describes a region of space. What properties are held between two regions of space if their location codes are A and B?

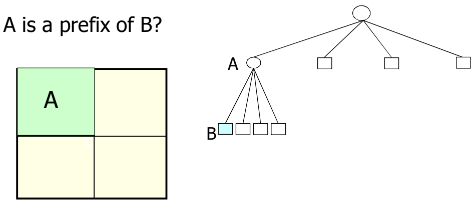
(a) A = B?



The region is identical.

Q4 (b)

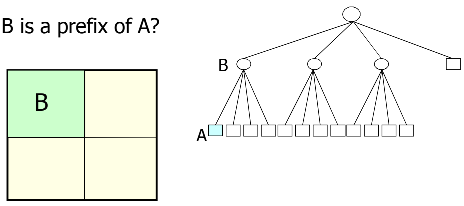
(b) A is a prefix of B?



Region A completely covers region B.

Q4 (c)

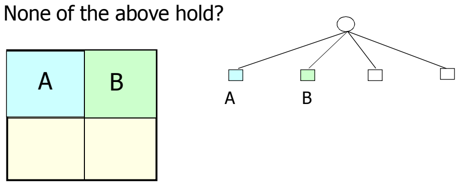
(c) B is a prefix of A?



Region B completely covers region A.

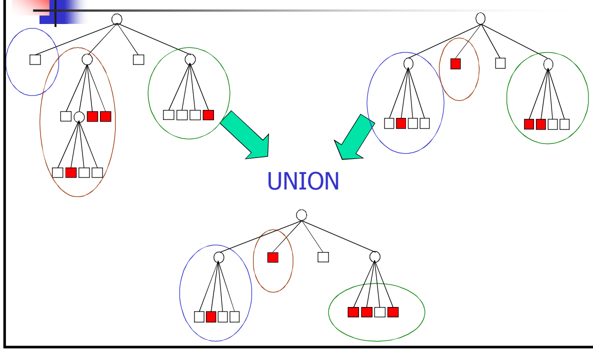
Q4 (d)

None of the above hold?

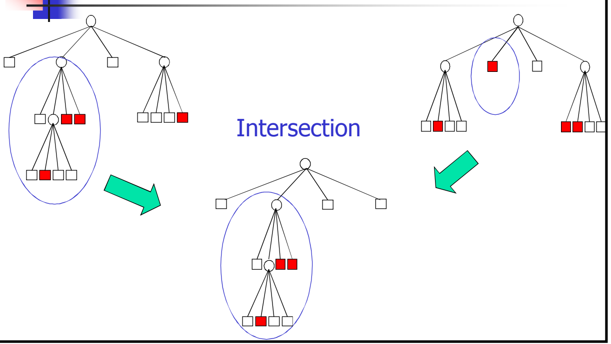


Regions A and B are disjoint.

Question 5 – Union of Quadtrees



Question 5 – Intersection of Quadtrees



Question 6

- You are trying to find out whether the public transport network for the Olympics is adequate.
- You have
  - A collection S of public transport stops {si}
  - A collection V = {vi} of venues, each of which has a collection of one or more entries Ei = {ei}.
- The network is adequate if at least one entry for each venue is within 100 metres of at least one public transport stop. All objects have location coordinates in the same system.

Q6A

Develop a simple algorithm that will associate each entry for each venue with the public transport with 100 metres, and will report if the network is adequate. Assume a predicate that is true if two objects are within 100 metres of each other.

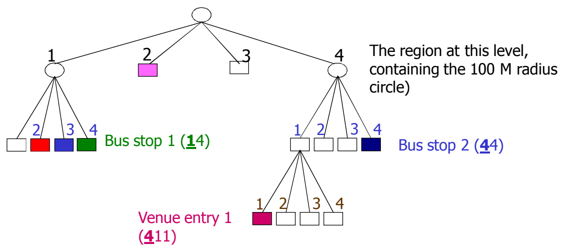
```
adequate:=true
for each venue
    venueServed:=false
    for each entry in venue
        for each stop
            if within100M(entry, stop) then
                venueServe:=true
                    output venue, entry, stop
adequate:= adequate and venueServed
```

- This algorithm requires comparing every entry for every venue with every stop

Q6B

- Assume we have a quadtree index on the region containing all the objects of interest. Each entry is indexed with a most specific location code. Each public transport stop is indexed with a location code entirely containing the 100 metre radius circle with the object at its centre. These codes are of varying specificity, and are generally less specific than the location codes for the entries.
- Sketch a modification of the algorithm of Q6a which uses the quadtree to eliminate most of the entry to public transport stop comparisons.

Q6B Quadtree Index

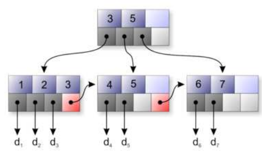


Q6B

```
adequate:=true
for each venue
    venueServed:=false
    for each entry in venue
        for each stop with location code a prefix of the entry location code
            if within100M(entry, stop) then
                venueServe:=true
                    output venue, entry, stop
adequate:= adequate and venueServed
```

Q6B

- Furthermore, the modification can be implemented with a B-tree index search.
  - **B-trees** are tree data structures that are most commonly found in databases and filesystems.
  - B-trees keep data sorted and allow amortized logarithmic time insertions and deletions.



Q6B

- It can be expected to greatly reduce the size of the collection of stops checked for each entry
- Why?
- The cost of binary tree search in average is Ω ($\log_2 n$) while the one of linear search is Ω (n)
  - $n > 0$ and $n \neq 1$, $n > \log_2 n$
  - e.g. $\log_2 256 = 8$

Questions