

# Large Scale Databases

Assignment Project Exam Help

S <https://eduassistpro.github.io/>  
els  
Add WeChat edu\_assist\_pro

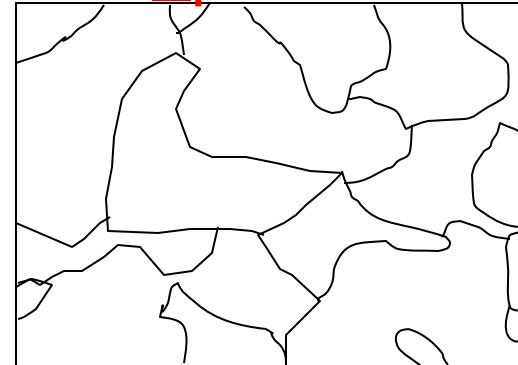
Chris Jones

# Spatial Data Models



2 main types of spatial data model:

- Field-based or location-based
  - 'raster' <https://eduassistpro.github.io/>
  - Add WeChat [edu\\_assist\\_pro](#)
- object-based
  - 'vector model'



# Raster Model

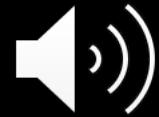


- Divide ground surface into regular grid cells
- record what is present in each cell [Assignment](#) [Project](#) [Exam](#) [Help](#)
  - class of land <https://eduassistpro.github.io/> assured property
- individual raster is usually -based
  - i.e. represents particular class of data
- may be several raster 'layers' in a database
- store each raster as matrix (array) of numeric data values

# Raster model



- Often derived/interpreted from scanning device images
  - e.g. multispectral scanners or *rasterised* from vector data
- Good for:
  - representing geographical phenomena
  - fast computation (especially operations, to find where multiple classes of phenomena coincide in space)
- Poor for
  - precisely surveyed objects
    - As it has fixed resolution
    - And lacks explicit object structure



# Vector Model

- Describes the location of discrete objects using geometry
- precise locations
- can record bohttps://eduassistpro.github.io/ of complete objects
- good for determining spatial relationships between whole objects
- requires more complex geometric computation to determine spatial relations

# Geometric Representation of Locations (with coordinates) in Vector Model



- Primitive geometric objects
  - points, lines, areas, surfaces, volumes

⋮ ⋮ ⋮ ⋮

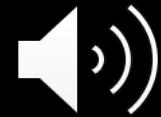
Assignment Project Exam Help



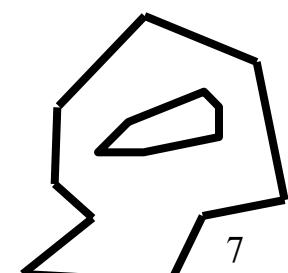
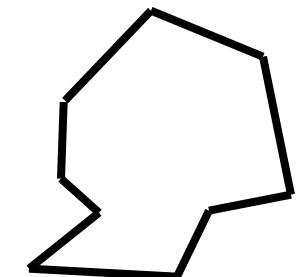
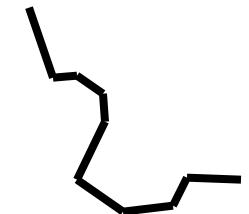
<https://eduassistpro.github.io/>

- Facilitate measurement (~~Add WeChat~~ ~~edu\_assist<sub>pro</sub>~~, length, height, area, volume, orientation)

# 2D Geometry Data Types



- *point*
  - defined by X, Y, (Z) coordinates
  - e.g. telegraph post, town at small scale
- *line (polyline / line string)*
  - defined by lis  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
  - e.g. roads an <https://eduassistpro.github.io/>
- *simple polygon*
  - defined by closed list of points
  - e.g. building, administrative district, lake
- *complex polygon*
  - defined by exterior polygon and one or more interior polygons (holes)



# Simple Feature Geometry Class Hierarchy from Open Geospatial Consortium



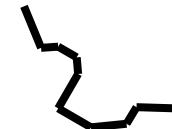
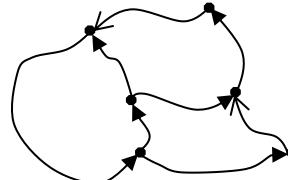
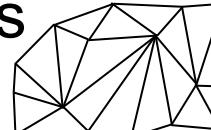
Assignment Project Exam Help

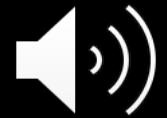
<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

# Open Geospatial Consortium (OGC)

## Geometry Types

- Point, Curve, Surface are types of Geometry
- LineString is a Curve = sequence of points with linear interpolation between the points
- Line is an open LineString 
- LinearRing is closed 
- Polygon is a plan <https://eduassistpro.github.io/> or more Linear Rings
- Triangle is a Polygon 
- Polyhedral surface is polygons that share boundaries
- Triangulated Irregular Network (TIN) consists of Triangles 
- TIN is a PolyhedralSurface
- MultiPolygon is a Multisurface is a GeometryCollection
- MultiLineString is a MultiCurve is a GeometryCollection



# Large Scale Databases

Assignment Project Exam Help

Location <https://eduassistpro.github.io/> Systems

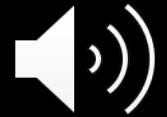
Add WeChat edu\_assist\_pro

Chris Jones

# Spatial Information



- Describing Location : qualitative vs quantitative
- Spatial Data Models (conceptual models)
- Geometry Data Types
- Spatial Relationships
  - Topological relations <https://eduassistpro.github.io/>
  - Orientation relations [Add WeChat edu\\_assist\\_pro](#)
  - Proximal relations
- *Shape and Pattern*
- *Correlations*
- *Interactions*
- *Routes*



# Location

## Qualitative vs Quantitative

- Qualitative: Place names, post codes / zip codes  
e.g. Cardiff, Queens Street, CF24 3AA
- Quantitative : Coordinates  
(lat/long vs map grid)  
e.g. (52.3, -2.3) (24)

Assignment Project Exam Help

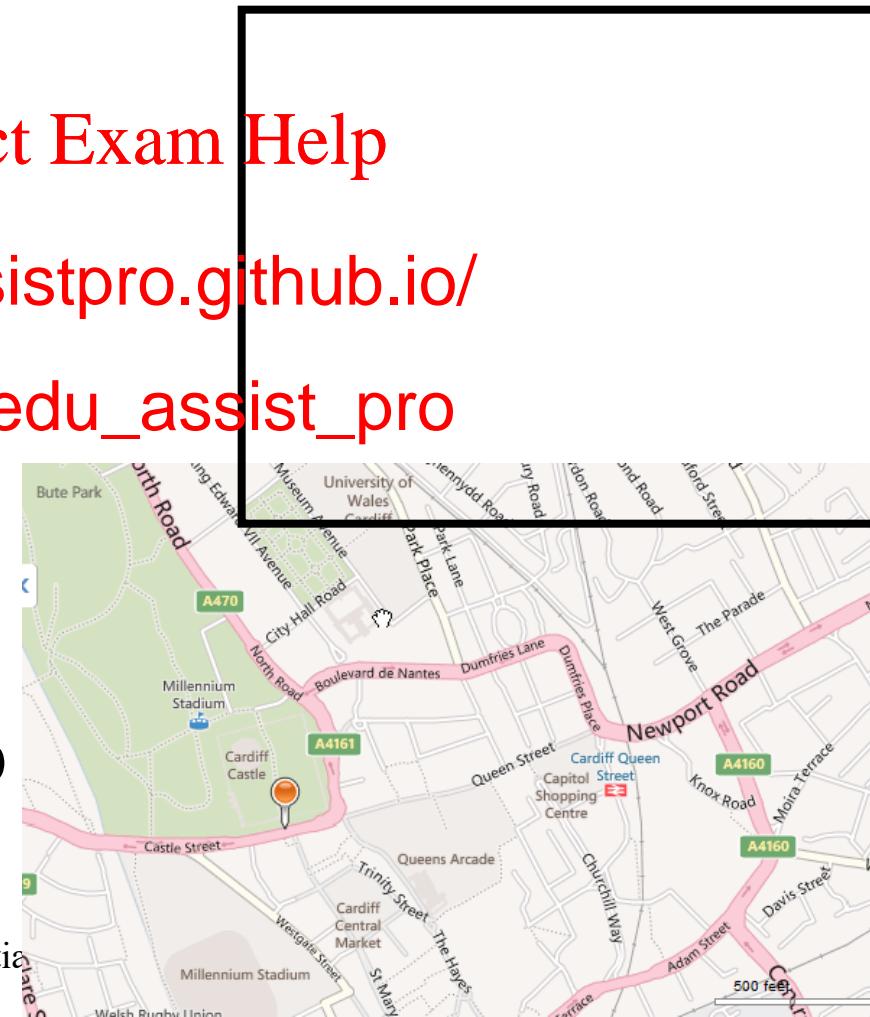
For GIS the two are <https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

## Absolute vs Relative

- Absolute : place name or coordinates, e.g. Taffs Well
- Relative : spatial relationship to absolute position

e.g. North of Cardiff / Near Cardiff



# Coordinate Systems: Latitude and longitude (global)



Assignment Project Exam Help

P

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

Latitude  $\Phi$  at point P is the angle between the line perpendicular to the Earth's surface at P and the plane of the equator

Longitude  $\lambda$  is angle between the plane passing through P and the Earth's poles and the plane of the prime meridian (also passes through poles)

# A Datum



For purposes of specifying coordinates and map projections, the Earth is approximated in shape by a spheroid – an ellipse rotated about its vertical axis.

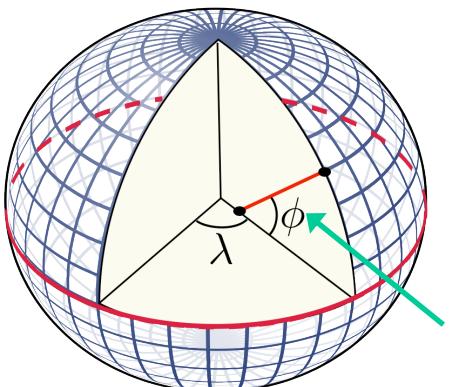
Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat [edu\\_assist\\_pro](#)

The spheroid is called a **Datum** and is defined by the lengths of two radii.

The datum used for GPS coordinates is called WGS84



Geodetic latitude

Spatial-3-Locations & Coordinate Systems

# Latitude & Longitude Coordinate Systems



Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

A **global** coordinate system.

Treating latitude and longitude as Cartesian (rectangular ) coordinates (above right) results in major distortions of areas and angles → hence need local map projections

# Map Projections and Map Grids



Project rays from centre of globe to a surface (cylinder, cone, plane) that can be unwrapped to, or is, a plane.

The resulting 2D planar coordinate system – in metres (not angles) - is referred to as a Map Grid or Map Projection

**Assignment Project Exam Help**



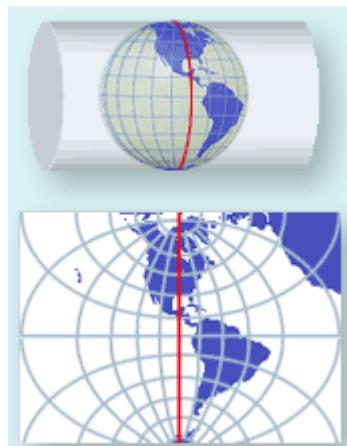
Cy

Pr

<https://eduassistpro.github.io/>

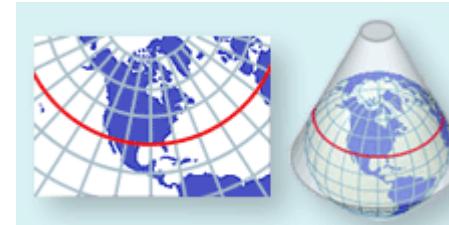
Azimuthal  
Projection

Add WeChat [edu\\_assist\\_pro](#)



Transverse  
Cylindrical  
Projection

e.g. British  
National Grid



Conical  
Projection

# Discrete Global Grids, e.g. HealPix (the previous map grids are local)

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

DGG divide surface of Earth into approx equal size cells (at different resolutions).

HealPix uses rectangular cells.  
Others DGGs use triangulations of the Earth's surface

Nasa Jet Propulsion Lab.

<http://healpix.jpl.nasa.gov/healpixBackgroundPurpose.shtml>

Assignment Project Exam Help  
Larg ases

<https://eduassistpro.github.io/>

Spatial Rel Add WeChat edu\_assist\_pro

# Significance of Spatial Relationships to Spatial Databases



- Queries to spatial databases often employ spatial relationships to express what data are to be retrieved

<https://eduassistpro.github.io/>

For example:

- *which objects lie inside a particular region?*
- *which regions are neighbouring another region?*
- *which areas does a region overlap?*
- *what objects a particular class are within some particular distance of another object?*

# Spatial Relationships: *topological, proximal, orientation*



## Topological relations (qualitative)

- Different types of connectivity between objects (e.g. containment, touching, overlap..)
- Invariant to rotation and scaling

## Proximal relations (qualitative)

- Refer to distance between objects

## Orientation relations (qualitative / quantitative)

- Refer to the direction/orientation between objects



# Topological relations between points, lines and areas

The main topological relations are:

- **meet (or touch)**
- **inside / contained** ([Assignment](#) | [Project](#) | [Exams](#) | [Help](#))
- **overlap** <https://eduassistpro.github.io/>
- **separate / disjoint** (not con  
Add WeChat edu\_assist\_pro
- **equal**
- **Other relations are covers and covered by**

# Topological relation: meets (touch) 🔊

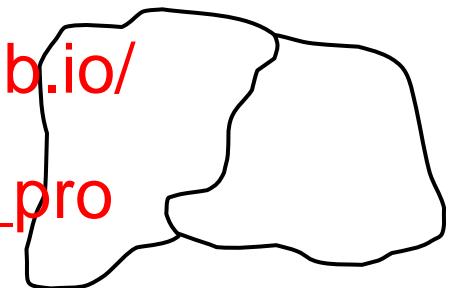
## meets (or touch)

*Two geometry objects meet/touch if there is some intersection between A and B but their interiors do not intersect*

Assignment Project Exam Help

### Examples

<https://eduassistpro.github.io/>



Two **regions** A and B meet if part of the boundary of A coincides with part of the boundary of B, but they do not otherwise overlap.

E.g. county A shares boundary with county B

# Meets / touch continued

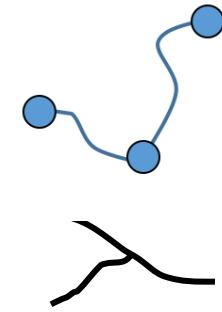


Two **lines** touch if their end points (boundary) intersect or if the boundary of one intersects the interior of the other

e.g. road A meets another road B

(the boundary of road A intersects the boundary of road B)

(the remainder is the interior of road A)

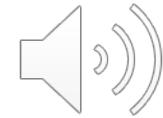


A **point** and a **line** touch if the point coincides with the boundary of the line

(NB a point has no extent and so is not regarded as having a boundary)



# Topological relation: Inside / Contained



**inside / contains** (converse relations)

A is inside (contained by) B

(i.e. B contains A)

if all parts of A co

or of B and no

parts of their bou

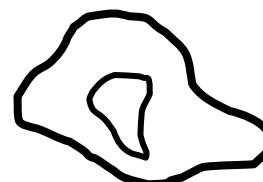
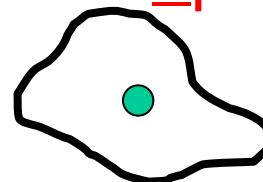
<https://eduassistpro.github.io/>

[= “proper part” but see cover]

[Add WeChat](#) [edu\\_assist\\_pro](#)

e.g.

town A inside county B



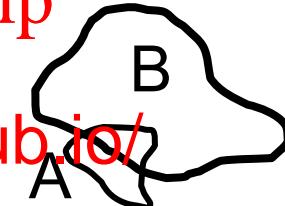
nation X contains province Y

# Topological relation: Overlap 🔊

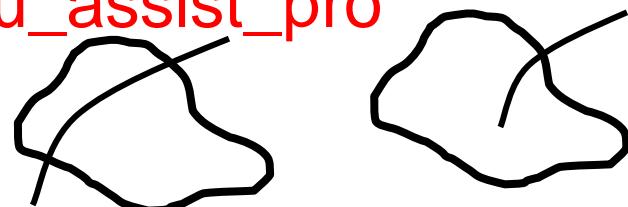
*Two objects A and B **overlap** if part of the interior of A intersects with part of the interior of B*

*while part of the interior of A intersects with the exterior of B*

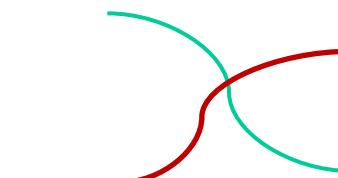
E.g. national park <https://eduassistpro.github.io/>



Add WeChat edu\_assist\_pro



E.g. road A crosses county X



E.g. road A crosses road B

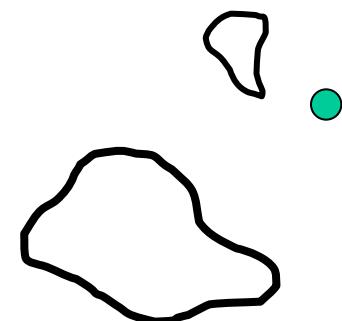
# Topological relation: Disjoint 🔊

**separate / disjoint** (not connected)

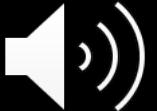
A is disjoint from B if no part of A  
coincides with <https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

e.g. lake X is separate from the city Y



# Topological relation: Equal



## Equal

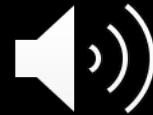
Two objects A and B are equal if each (and every) part of A coincides with a part of B and each part of B coincides with a part of A.

[Assignment](#) [Project](#) [Exam](#) [Help](#)

<https://eduassistpro.github.io/Bio/>

[Add WeChat](#) [edu\\_assist\\_pro](#)

# Topological relation: Covers / Covered By

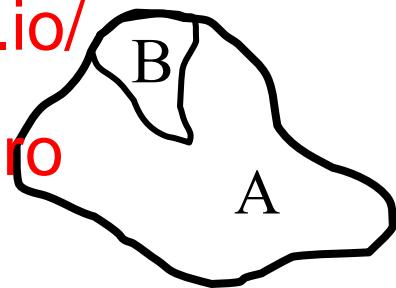


## Covers / covered by

Object A covers object B if A

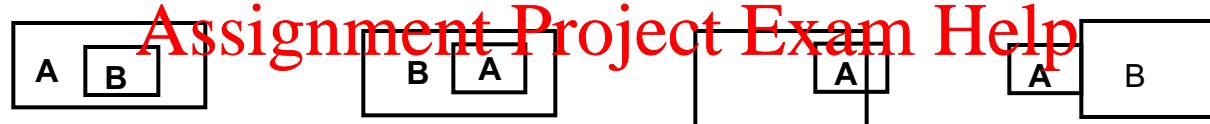
contains B and coincides with B.

In that case, B is covered by A.



e.g. England covers the county of Kent (the coastline of Kent is the coastline of England)

# Summary of Topological Relations



**A contains B** <https://eduassistpro.github.io/> **A touches B**

Add WeChat edu\_assist\_pro



**A covers B** **A coveredBy B** **A equals B** **A disjoint B**

# 9-Intersection model to define topological relations between regions

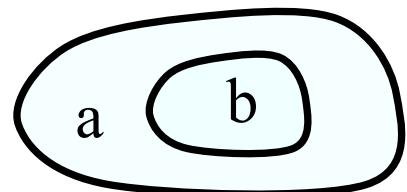


Treats spatial objects as regions having 3 components:

Boundary (B), Interior (I) and Exterior (E)



The topological relations between two regions can be characterised by which of these components intersect each other



For Example:

a contains b

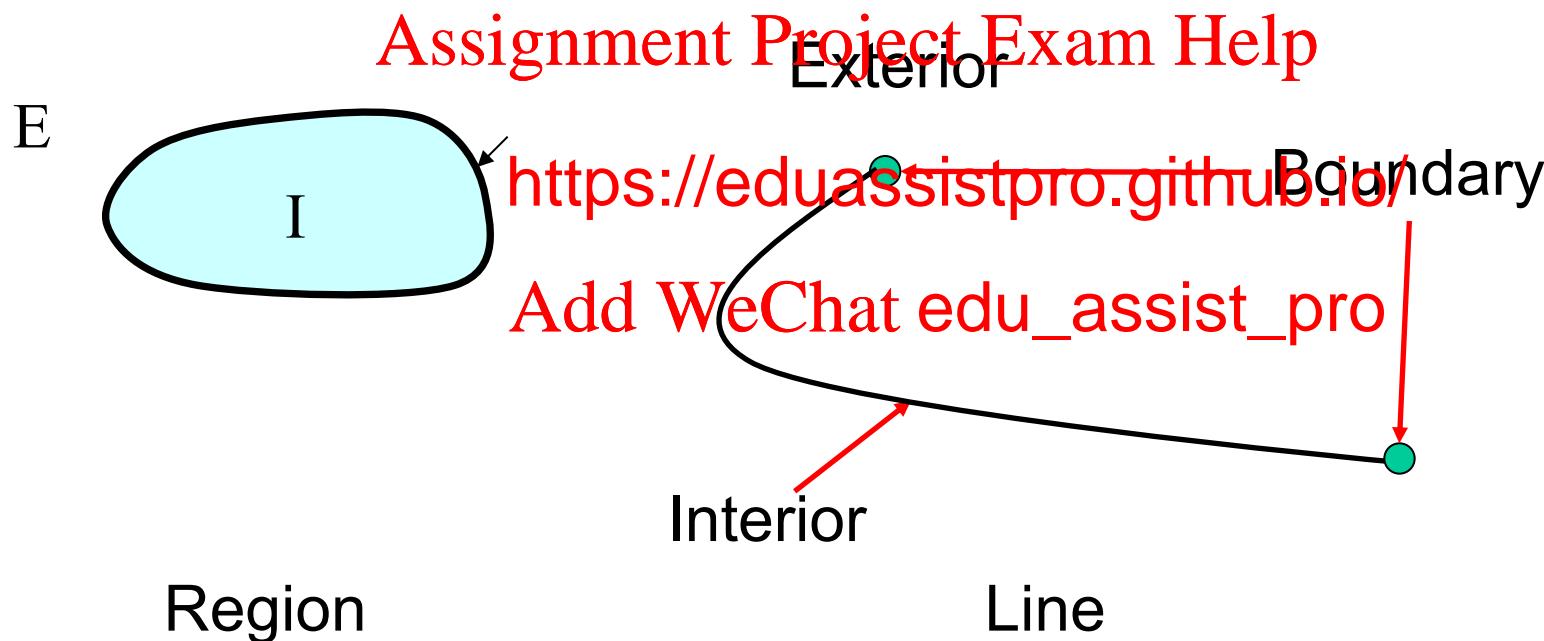
a \ b	Interior	Boundary	Exterior
Interior	$I(a) \cap I(b)=1$	$I(a) \cap B(b)=1$	$I(a) \cap E(b)=1$
Boundary	$B(a) \cap I(b)=0$	$B(a) \cap B(b)=0$	$B(a) \cap E(b)=1$
Exterior	$E(a) \cap I(b)=0$	$E(a) \cap B(b)=0$	$E(a) \cap E(b)=1$

# Applying the 9-intersection model to lines



Treats spatial objects as having 3 components:

Boundary (B), Interior (I) and Exterior (E)

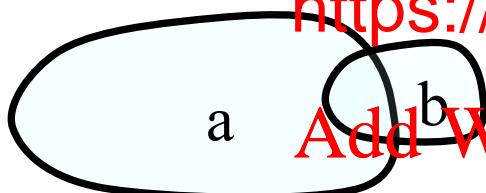


A Point has no boundary.

Exercise: define the intersection matrix for  
a Overlaps b relation  
where a and b are regions



Assignment Project Exam Help



<https://eduassistpro.github.io/>

Add WeChat [edu\\_assist\\_pro](https://edu_assist_pro)

	terior	Boundary	Exterior
Interior			
Exterior			

# Orientation relations



- *quantitative*: measured angular bearing, relative to north, or some other direction
  - e.g. person on bearing 45 degrees from lamp post (measured cf <https://eduassistpro.github.io/>; but bearing could also be measured in degrees)
- *qualitative*: e.g. north, south, east, west, above, below, in front.....
  - e.g. person is north east of the lamp post

# Proximal relations



- quantitative
  - measured distance
    - e.g. city <https://eduassistpro.github.io/> city B
    - Add WeChat edu\_assist\_pro
- qualitative
  - “near”, “close”, “far”, “distant” etc.

# References for the 9-intersection model and RCC (Region Connection Calculus)

M.J. Egenhofer, R. Franzosa (1991) Point-set topological spatial relations.  
International Journal of Geographical Information Systems, 5 (2), pp. 161-174

M.J. Egenhofer, J. Herring (1991), Categorizing binary topological relations between regions, lines and points in geographic databases, Technical Report, Department of Surveying Engineering, University of Maine, 1991

## Assignment Project Exam Help

E. Clementini, P. Di Felice,  
Topological Relationships      Small Set of Formal  
International Symposium o      Proceedings of the Third  
titles on Google.

## Add WeChat edu\_assist\_pro

You can obtain copies of these articles by searching titles on Google.

Wikipedia (<https://en.wikipedia.org/wiki/DE-9IM>) has a reasonable explanation of the dimensionally extended 9-intersection model as introduced in Clementini

G. Cohn; B. Bennett; J. Gooday; M. M. Gotts (1997). "Qualitative Spatial Representation and Reasoning with the Region Connection Calculus". *GeoInformatica*. 1 (3): 275–316

Randell, D. A., Cui, Z. and Cohn, A. G. "A spatial logic based on regions and connection," *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning*,<sup>35</sup> Morgan Kaufmann, San Mateo, pp. 165–176, 1992.

Spatial Database Management Assignment Project Exam Help

of G <https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro  
Chris J

# Modelling Geographical Information



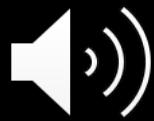
- Geographical Objects have
  - identity
    - name, uniq
  - classification
    - one or more classes
  - thematic attributes / properties / statistics
    - various (non-spatial) data types
  - spatial representation
    - vector or raster structured data

# Example geographical objects



County Object	City Object	Street Object
Name	Assignment Project Exam Help	Name
Capital City	<a href="https://eduassistpro.github.io/">https://eduassistpro.github.io/</a>	City
Population	Add WeChat edu_assist_pro	Surface-material
<b>Boundary</b>	Date of ori <b>Boundary</b>	Width <b>Centre-Line</b>

# Data Types



- Identity, class, thematic attributes
  - standard database types such as string, number, date
- Spatial representation (e.g. Boundary)
  - specialised geospatial models

<https://eduassistpro.github.io/>  
Add WeChat edu\_assist\_pro

# Example data for a set of counties

## County records

Name	Capital	Population	Boundary
Bigshire	Upton	4002	A
Mereshire	Downton	67358	B
Bottleshire	Fizzton	108	C
....	.....	.....	

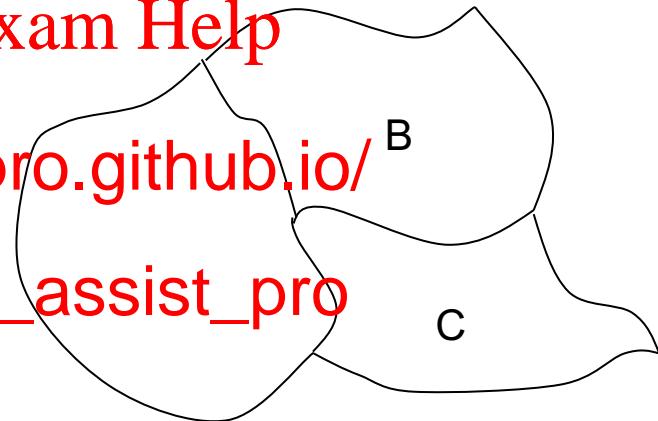
Assignment Project Exam Help

<https://eduassistpro.github.io/>

Boundary geometry records

Add WeChat edu\_assist\_pro

Boundary	Coordinates
A	x1,y1,x2,y2.....xm, ym
B	x1,y1,x2,y2.....xn, yn
C	x1,y1,x2,y2.....xo, yo



Alternative ways in which the geometry could be stored :  
Boundary as sequence of line segments - segments shared

e.g.

40

# Fixed and variable length fields

- County records
  - all fields consist of single data item, one of which is geometry
- Boundary geo <https://eduassistpro.github.io/>
- Coordinates = variable len
  - Order (sequence) of coordinates must be maintained

# Applying relational databases to spatial data



- The need for Object-Relational Databases and OO Databases with spatial capability
  - [ Normalisation of the geometric data would result in very inefficient storage and performance ]
- Complex data types
  - to represent geometry <https://eduassistpro.github.io/>
- Spatial functions and operators
  - to compute properties of and spatial relations between geometry objects
- Spatial indexes
  - to provide efficient execution of spatial query operators such as for topological and proximity (distance) relations.

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat [edu\\_assist\\_pro](https://edu_assist_pro)

# Object-Relational Spatial Databases



- Examples of DBMS with spatial functionality

Oracle

<https://eduassistpro.github.io/>

PostgreSQL

Add WeChat edu\_assist\_pro

Microsoft SQL Server

IBM DB2 (Spatial Extender)

MySQL

*[[ + NoSQL databases such as MongoDB. ]]]*

# Example data type declarations



```
CREATE TYPE Point AS OBJECT (
    X NUMBER,
    Y NUMBER,
    Z NUMBER,
    MEMBER FUNCTION Distance(P1 IN Point)
        RETURN NUMBER );
https://eduassistpro.github.io/
```

```
CREATE TYPE LineType AS VA ) OF Point;
    Add WeChat edu_assist_pro
```

```
CREATE TYPE LineString AS OBJECT(
    Number_of_Points INT,
    Geometry LineType,
    MEMBER FUNCTION Length (SELF IN LineString)
        RETURN NUMBER );
```

# Example Table Declarations



- CREATE TABLE Settlements (  
    Name Varchar(30),  
    Population Number,  
    Location  
    <https://eduassistpro.github.io/>
- CREATE TABLE Roads (  
    Name Varchar(30),  
    Type Varchar(30),  
    Location LineString );

# Note on Oracle Spatial Data Types



- Has (very complex) geometry object called **SDO\_GEOGRAPHY**
    - Can represent points, lines, polygons etc.
    - Multiple parts
    - Is associated with SDO\_GEOGRAPHY
- Assignment Project Exam Help**  
**Add WeChat edu\_assist\_pro**
- <https://eduassistpro.github.io/>

Also: **SDO\_POINT\_TYPE**

- Just for single points
- Can be used as part of the SDO\_GEOGRAPHY object

# Oracle SDO\_GEOmetry object



```
CREATE TYPE sdo_geometry AS OBJECT (
```

```
    SDO_GTYPE NUMBER,
```

```
    SDO_SRID NUMBER,
```

```
    SDO_POINT SDO_POINT_TYPE,
```

```
    SDO_ELEM_INFO SDO_ELEM_INFO_ARRAY,
```

```
    SDO_ORDINATES SDO_ORDINATE_ARRAY);
```

Assignment Project Exam Help

```
CREATE TYPE sdo_pohttps://eduassistpro.github.io/
```

```
    X NUMBER,
```

```
    Y NUMBER,
```

```
    Z NUMBER);
```

Add WeChat edu\_assist\_pro

```
CREATE TYPE sdo_elem_info_array AS VARRAY (1048576) of NUMBER;
```

```
CREATE TYPE sdo_coordinate_array AS VARRAY (1048576) of NUMBER;
```

*See Exercise for examples of using SDO\_GEOmetry*

Spatial Database Management Assignment Project Exam Help      Query of

Ge https://eduassistpro.github.io/  
ation

Add WeChat edu\_assist\_pro  
Chris J

# Querying Spatial Databases



- Proximity / distance -based
  - find objects within a distance of a specified feature
- Topological relations
  - Retrieve objects related by topological relations  
<https://eduassistpro.github.io/>
- Intersection (spatial)
  - Retrieve geometry at intersections of spatial objects
- Region Containment / Spatial Filter
  - retrieve data including geometry in a specified region – Filter is approximate and results could be at least partly outside

# Proximity / Distance Queries

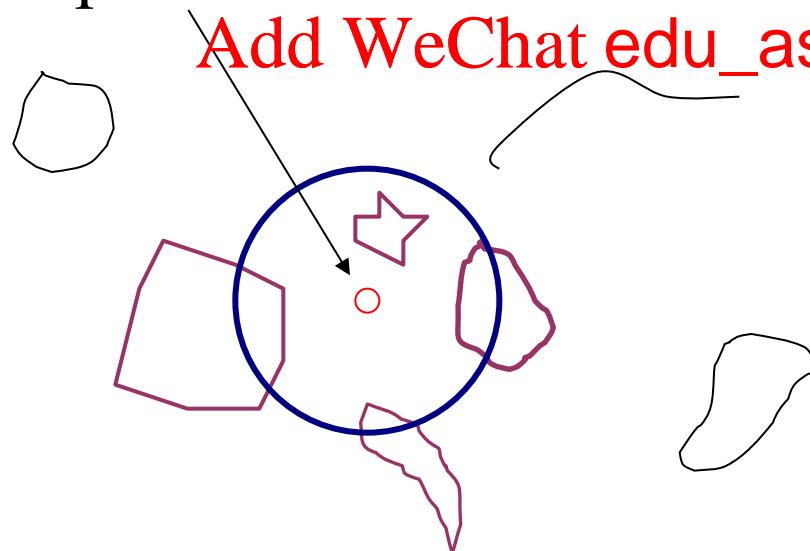


## Within distance query

- Find properties of features within specified distance of specified feature
  - relative to poi

<https://eduassistpro.github.io/>

Add WeChat `edu_assist_pro`



# Generic Spatial SQL operators for Within Distance



Find Names of villages within 5km of Cardiff:

```
Select S2.name  
From Settlements S1, Settlements S2  
Where S1.name = https://eduassistpro.github.io/  
       distance(S1.location, S2.location) < 5000  
       and S2.type = "village"
```

Function **distance** returns a real number (the distance) and takes as parameters two geometric objects

NB: Oracle has a distance function SDO\_GEOGRAPHICAL\_DISTANCE that computes the distance. It also has a within distance operator SDO\_WITHIN\_DISTANCE

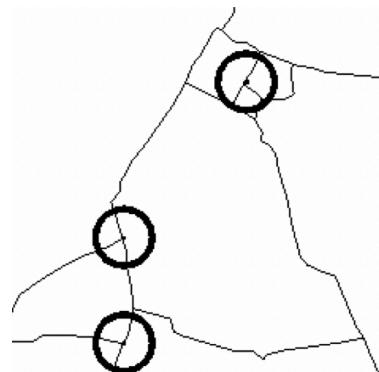
# Distance Buffers



- Generate a region of space that is some distance from a specified geometry object.
- The buffers are then used in intersection queries to retrieve the data later slides.  
<https://eduassistpro.github.io/>

Assignment Project Exam Help  
Oracle has a Buffer function (SDO\_BUFFER)

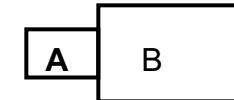
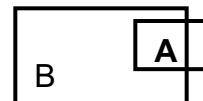
Point buffers



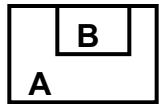
Line buffers



# Topological relations typically supported (OpenGIS standard)



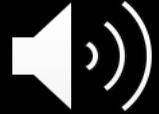
Assignment Project Exam Help



<https://eduassistpro.github.io/>

Add WeChat [edu\\_assist\\_pro](https://eduassistpro.github.io/)

# Generic Spatial SQL queries for topological relations



e.g. Names of woodlands greater than 2000 square metres inside, or partially inside, Dyfed:

Select Woodlands.name

From Woodlands, Counties

Where

Assignment Project Exam Help

inside(Woodl ties.boundary )

or https://eduassistpro.github.io/

overlap(Woodland.bounda ties.boundary)

and Add WeChat edu\_assist\_pro

Woodlands.area > 2000 a

Counties.name = "Dyfed";

inside and overlap return a Boolean and take as parameters two geometric objects that could be lines or polygons

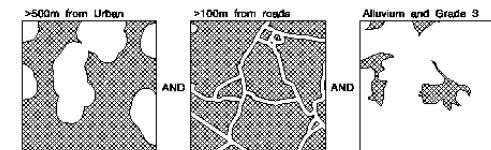


# Intersection (also called overlay or spatial join)

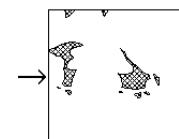
- Returns geometry of locations where a set of two or more spatial features coincide or intersect
- The spatial features could be, for example
  - regions with particular thematic classification or identity
  - regions that are within specified distance of other objects, i.e. buffer zones

Alluvium and Grade <https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro



Large Scale Databases



# Buffer queries (intersection)

- *Retrieve geometry* within distance of another geometry objects (as opposed to measuring distance or finding objects within a distance)

Assignment Project Exam Help

Retrieve geometry <https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

# Buffer queries (continued)



Retrieve geometry  
within distance of a  
line

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat ~~edu\_assist\_pro~~  
Retrieve geometry  
within distance of a  
polygon

Note that when Oracle Buffer function SDO\_GEOGRAPHICAL\_BUFFER is applied to lines it operates separately on each individual segment  
So in the above examples lines would need to be continuous

# Generic Spatial SQL query for intersection

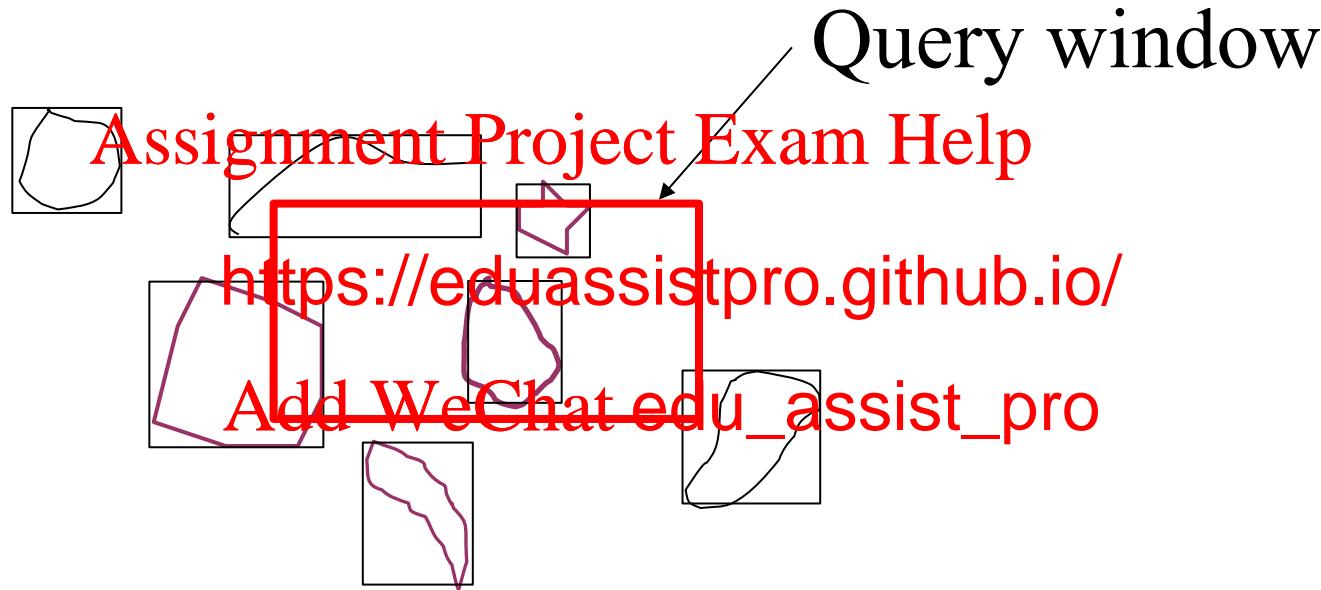


**Total area of woodlands in county of Dyfed:**

```
Select      Assignment Project Exam Help  
sum(area(in      ds           .boundary,  
          https://eduassistpro.github.io/       ) )  
From    Woodlands, County  
Where   County.name = 'Dyf'Add WeChat edu_assist_pro
```

intersection returns the geometry resulting from the intersection of two geometry objects given as parameters.

# The filter operation for region containment

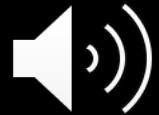


All objects (entire geometry) whose bounding rectangles intersect the query rectangle are retrieved, i.e. not strict containment or overlap

# Generic Spatial SQL

## Rectangular Region Containment query

Filter version (a low cost operation)



```
Select S.name, S.geometry
```

```
From Settlements S
```

Assignment Project Exam Help

```
Where Filter(
```

```
Re https://eduassistpro.github.io/
```

) ;

Add WeChat edu\_assist\_pro

S.geometry represents the geometry of the settlements

Rectangle defines the geometry of a rectangular query region

Here the settlement geometry is tested for intersection with a rectangular query region.

If the settlements are area features, in query they would be approximated by bounding rectangles (see figure)

60

# Spatio-Textual Indexing: Geoparsing 1

Assignment Project Exam Help  
To ition

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro  
Chris J

Cardiff University, UK

# Geographical Information Retrieval

*Retrieval of geo-information from text documents*

(or weakly-structured sources)

- Social media messages / posts (. current events; disasters; health of people; public gatherings; wildlife etc.)
- Web pages: many from <https://eduassistpro.github.io/>
  - Many queries are place name
  - Relevant pages usually found, as normal, by matching the place name to document text
  - Some web content is explicitly spatially indexed (with coordinates) usually “points of interest” (POI), e.g. restaurants, shops, banks, cinemas, hotels, museums.....
- Digital libraries, scientific documents....

volcanic eruptions in new zealand

Latest People Photos Videos

ABC News @ABC · Dec 11, 2019

LATEST: At least six people have now died in wake of volcanic eruptions that occurred on small island in New Zealand on Monday, officials say.

Death toll from New Zealand volcano eruption rises to 6, while 8 still ...  
At least six people have now died in the wake of volcanic eruptions that occurred on a small island in New Zealand on Monday, officials ...  
abcnews.go.com

Search: gardens in auckland city

Google

All Maps Images Shopping News More

About 54,900,000 results (0.82 seconds)

Rating Hours

Hours or services may differ

**Myers Park**  
4.4 ★★★★ (733) · Park  
72 Greys Ave  
Urban park with a playground & sculpture

**Albert Park**  
4.6 ★★★★ (3,108) · Park  
33-43 Princes St  
Scenic, landscaped park with a fountain

**Auckland Domain**  
4.6 ★★★★ (7,299) · Park  
Park Rd  
Large city park in a volcanic crater

More places

www.aucklandforkids.co.nz · gardens-and-parks-in-a... · Gardens and Parks in Auckland | Auckland for Kids

Gardens and Parks, Albert Park, Princes Street, Auckland Central, Auckland Domain, Park Road, Grafton, Central Auckland. Dove-Meyer Robinson Park - Parnell Rose Gardens, Eden Gardens, Michael Joseph Savage Memorial, Mount Eden - Mangerehau, Mount Wellington, One Tree Hill Domain and Cornwall Park.

# How is text information geographically referenced?



- Some social media have coordinates from GPS
- POIs are associated with explicit coordinates attached
- Otherwise place names in documents indicate geographical locations
- The place names in documents allow them to be indexed spatially

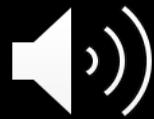
[www.sciencelearn.org.nz/](http://www.sciencelearn.org.nz/)

Assignment Project Exam Help  
Add WeChat edu\_assist\_pro  
<https://eduassistpro.github.io/volcano>  
magma  
Underlying  
diffuse pool of magma  
that occasionally finds its  
way to the surface. Unlike  
a 'classic' volcano – such  
as Mt Taranaki or Mt  
Ngāuruhoe with a single  
vent through the crust – in  
Auckland, the magma

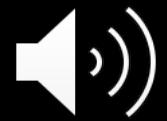
Large Scale Databases - C.B. Jones

The screenshot shows a Twitter post from ABC News (@ABC) dated December 11, 2019, stating: "At least six people have now died in wake of volcanic eruptions that occurred on small island in New Zealand on Monday, officials say." Below the tweet is a photograph of a volcano erupting. To the right is a Google search results page for "parks in auckland city". The search bar shows the query. The results include a map of Auckland with several park locations marked, and a list of three parks with ratings and descriptions: Myers Park (4.4 stars), Albert Park (4.6 stars), and Auckland Domain (4.6 stars). There are also links to "More places" and "Gardens and Parks in Auckland | Auckland for Kids".

# Some challenges in GIR



- Finding relevant documents for a geospatial query
- Indexing documents with respect to text and space
- Recognition of geographic references in text documents (toponym recognition)  
**Assignment Project Exam Help**
- Disambiguation  
(toponym resolution) <https://eduassistpro.github.io/> of documents
- Extracting structured geographical information from natural language  
**AddWeChat@edu\_assist\_pro**
- Understanding vague spatial language
  - Vague / vernacular place names  
*(city centre, the south,)*
  - Vague / imprecise language (*near, beside, on, north of, in front of, between...*)



# Geoparsing

*Identifying and geocoding geographic references in text (place names, addresses...)*

- Toponym recognition:

- Many words from the assignment project exam help used in a non-geographic sense (<https://eduassistpro.github.io/things>) = non-geo - geo ambiguity

Add WeChat edu\_assist\_pro

- Toponym resolution

- Many different places are referred to by same name = geo - geo ambiguity
- Need to find the correct coordinates for a toponym.  
Usually use gazetteers (lists of place names + coordinates + feature type)

# Ambiguity in toponym recognition



Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

The words in red are toponyms that need to be recognized.  
On the right are all words that can be used as toponyms –  
but some of them (coloured magenta) are not actually  
toponyms in this text. This is non-geo - geo ambiguity

# Named entity recognition



- The task of recognizing place names in text is part of the natural language processing method of Named Entity Recognition (NER)
- NER software such as <https://eduassistpro.github.io/> people, location, organisations, numbers  
Add WeChat edu\_assist\_pro

Jack Hagel, Staff Writer . Redevelopment of the World Trade Center site in New York is getting some input from a Raleigh real - estate maven .

PERSON                    FAC                    GPE

GPE = country, city or state;

FAC = built structures: (Buildings, airports, highways, bridges etc)

# NER with the Allen NLP software



Demo at

<https://demo.allennlp.org/named-entity-recognition/>

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

ORG = organization

GPE = country, city or state;

FAC = built structures: (Buildings, airports, highways, bridges etc)

LOC = other locations including natural features

# What could distinguish a place name from other entities?



For example, *the presence or the absence* of :

- Initial capital letter
- Preceded by a spatial preposition / suffix such as in, near  
**Assignment Project Exam Help**
- Can be found in a gazetteer of place names
- Preceded by initial **<https://eduassistpro.github.io/>**
- Preceded by “Mr”, “Miss”, “Dr”  
**Add WeChat edu\_assist\_pro**
- A sequence of numeric digits
- Can be found in a list of first names
- Is preceded or followed by a word that refers to type of place (city, lake, river...)

Which of the above apply to place names in English?



# Methods of automating toponym recognition

- Use machine learning (ML) with features (items of evidence) such as those listed in previous slide.
- Classifiers, such as Decision Tree (e.g. Random Forest); Support Vector Machine (SVM) or a deep learning method (neural network).  
**Assignment Project Exam Help**
- Input features could rds to the left and the right of the word or <https://eduassistpro.github.io/> (e.g. 3 words to left and 3 to right).  
**Add WeChat edu\_assist\_pro**

**Club Plaza in Kansas City, Mo. the center**

-Kansas City had been identified as a noun phrase or

proper noun and hence a potential place name

ML Features either a one-hot vector or all words in collection – set elements to 1 if word in the target sentence. Or similar vector of tf-idf (text frequency / inverse document frequency) values

# Word Embeddings as features for machine learning



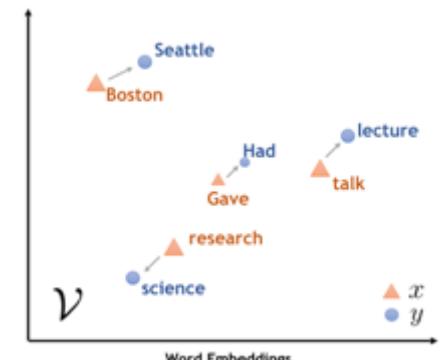
- Rather than the actual words, can use *word embeddings* of neighbouring words as input features

**Club Plaza in Kansas City, Mo the center**  
**Assignment Project Exam Help**

- Represent a word as
- Each dimension repr of the word.
- Vector obtained with dimensionality re techniques based on the co-occurrence of target word and other words commonly used in the context of the word.
- Words with similar meaning are closer in the vector space
- word embedding methods, e.g. GloVe, Word2Vec, FastText, Bert

<https://eduassistpro.github.io/>  
Add WeChat edu\_assist\_pro

Image: Zenun Kastrati WET: Word Embedding-Topic distribution vectors for MOOC video lectures dataset



Large Scale Databases C.H. Jones "Navigating themes in restaurant reviews with Word Mover's Distance"

# References

Purves R.S, P.D. Clough, C.B. Jones, M.M. Hall, M. Murdoch (2018) “Geographic Information Retrieval: Progress and challenges in spatial search of text”. Foundations and Trends in Information Retrieval. Vol. 12, No. 2-3, pp 164-318.

F. Melo and B. Martins (2017). Automated geocoding of textual documents: A survey of current approaches. *Transactions on GIS*, 20(1), 3–38.

J. L. Leidner and M. D. Li phical references in the form of

place names and associate <https://eduassistpro.github.io/> THAL Special, 3(2):5–11.

Geonames Gazetteer - [https://www.geonames.o](https://www.geonames.org)

AllenNLP NER demo - [http://demo.allennlp.org/NamedEntityRecognition/](http://demo.allennlp.org/NamedEntityRecognition)

Nominatum Geocoding API - [https://nominatim.org/](https://nominatim.org)

GeoPy Python library for geocoding - <https://geopy.readthedocs.io/en/1.22.0/>

SpaCy guidance on NER (and other NLP) <https://spacy.io/usage/linguistic-features>

Stanford Named Entity Recognizer - <https://nlp.stanford.edu/software/CRF-NER.html>  
and <https://corenlp.run/>

# Spatio-Textual Indexing: Geoparsing 2

Assignment Project Exam Help  
To

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

# Toponym Resolution (Geocoding or georeferencing)



- Having identified a word as a probable place name, we need to find its coordinates
- Can look up the name in a **gazetteer** but in many cases there are several instances of the same name
- Geonames gazetteer <https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

# Toponym geo-geo disambiguation



- Ambiguous names include:  
Raleigh, Kansas City.

Various strategies (see next slides)

- Several methods look for associations between names in same

Nearby places

geographical hierarchy  
e.g. New York and World Trade Center

-and Manhattan and World Trade Center

or between names and other text:  
(language models)

<https://eduassistpro.github.io/>

Jack Hagel, Staff Writer  
Redevelopment of the **World Trade Center** site in **New York** is getting some input from a **Raleigh** real-estate maven.

York Properties President Smedes York was chairman of an Urban Land Institute panel at the **World Trade Center** and **Lower Manhattan** summit last month.

He was chairman of the Urban Land Institute, a **Washington** nonprofit organization, from 1989 to 1991. His dad, J.W. "Willie" York, joined the Urban Land Institute in 1947. That's where he met J.C. Nichols, the developer of **Country Club Plaza** in **Kansas City, Mo.** - the center that inspired Willie York to build **Raleigh's Cameron Village**, the **Southeast's** first shopping center.

# Computing methods for toponym resolution



- Default sense – choose the biggest / best known instance of the place, e.g. New York in USA as opposed to the village in UK
- If an ambiguous name has a hierarchical relation with an unambiguous name in the document then choose the instance of the name that has that relation:

“Kansas City, Mo “<https://eduassistpro.github.io/> n Missouri (Mo)

“Raleigh’s Cameron Village” implies a link to a page that contains Cameron Village which is in North Carolina

- If an instance of an ambiguous place is geographically close to instances of unambiguous places in the text, choose that instance (see slide on spatial minimality heuristic).
- Use language modelling methods : locations are represented by the words with which they are associated. See later slide.



# Disambiguation with Default Sense

- choose the biggest / best known instance of the place

Look up place name in  
a gazetteer

Assignment Project Exam Help

<https://eduassistpro.github.io/>

has largest p  
Add WeChat [edu\\_assist\\_pro](#)

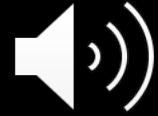
It works here for  
*New York, Kansas City  
and Raleigh*

Jack Hagel, Staff Writer  
Redevelopment of the **World Trade Center** site in **New York** is getting some input from a **Raleigh** real-estate maven.

York Properties President Smedes York was chairman of an Urban Land Institute panel at the **World Trade Center** and **Lower Manhattan** Summit last month.

York was chairman of the Urban Land Institute, a **Washington** nonprofit organization, from 1989 to 1991. His dad, J.W. "Willie" York, joined the Urban Land Institute in 1947. That's where he met J.C. Nichols, the developer of **Country Club Plaza** in **Kansas City, Mo.** - the center that inspired Willie York to build **Raleigh's Cameron Village**, the **Southeast**'s first shopping center.  
77

# Disambiguation with Hierarchical Relations



If text includes a geographical parent or child of a candidate instance of place:

– assign

ambiguous topography Assignment Project Exam Help

Hierarchical relati

<https://eduassistpro.github.io/>

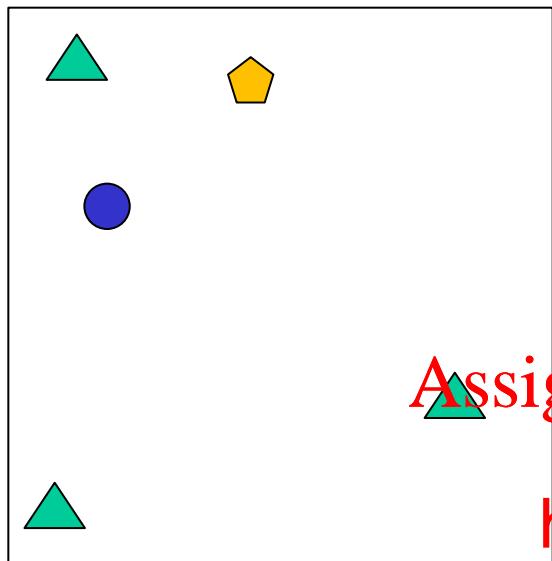
- Kansas City → Mo. [Mo. = Missou
  - Cameron Village → Raleigh
  - World Trade Center → New York
  - Lower Manhattan → New York
  - Country Club Plaza → Kansas City
- [ Raleigh → Southeast (vernacular)]

Jack Hagel, Staff Writer  
Redevelopment of the **World Trade Center** site in **New York** is getting some input from a **Raleigh** real-estate maven.

York Properties President Smedes York was chairman of an Urban Land Institute panel at the **World Trade Center** and **Lower Manhattan** Summit last month.

York was chairman of the Urban Land Institute, a **Washington** nonprofit organization, from 1989 to 1991. His dad, J.W. "Willie" York, joined the Urban Land Institute in 1947. That's where he met J.C. Nichols, the developer of **Country Club Plaza** in **Kansas City, Mo.** - the center that inspired Willie York to build **Raleigh's Cameron Village**, the **Southeast**'s first shopping center.

# Resolving ambiguity with spatial minimality



Toponyms with coordinates

▲ Ambiguous – 3 candidate locations

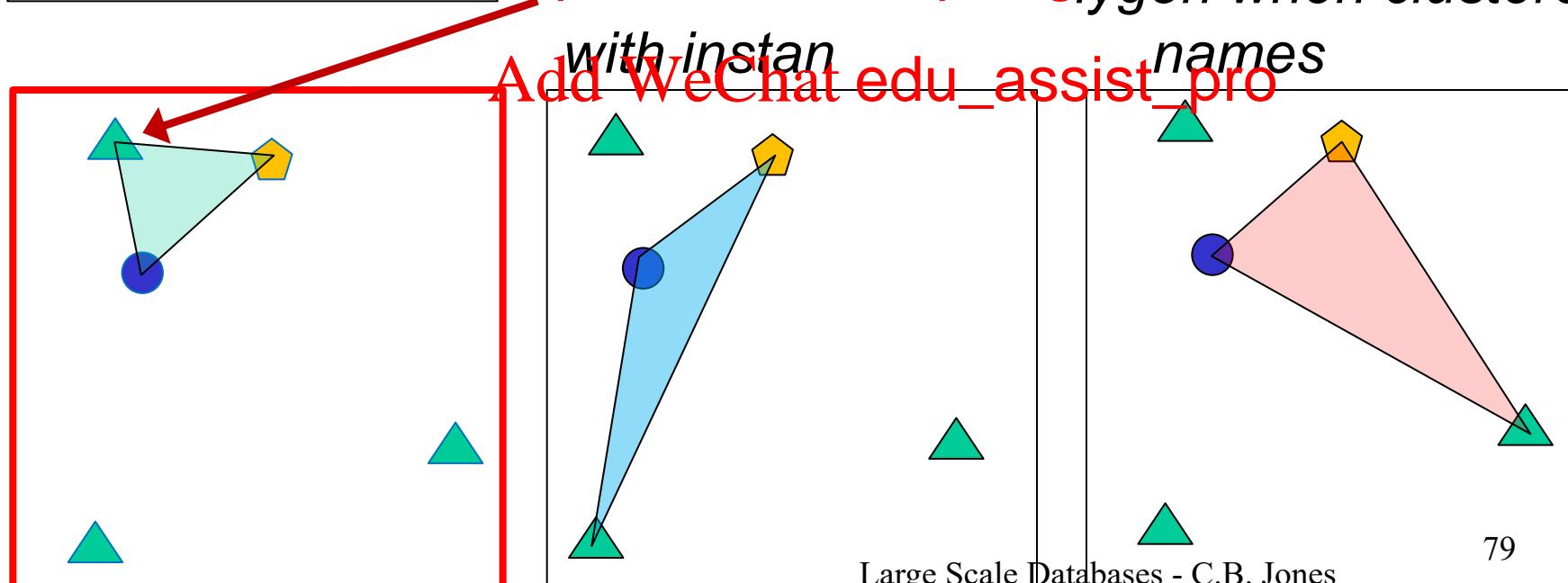
● Not ambiguous – 1 location

◆ Not ambiguous – 1 location

Assignment Project Exam Help

nym select candidate  
lygon when clustered

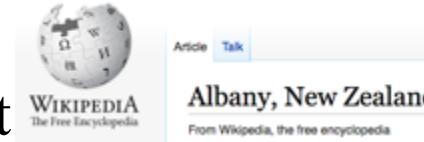
with instant names  
Add WeChat edu\_assist\_pro



# Language modelling approach for toponym resolution



- Use all the associated words to resolve ambiguity – might or might not include place names
- E.g. *Albany, New Zealand* can be represented by the words in Wikipedia article that mentions Westfield centre, Massey University etc.
- *Albany, New York* is associated with different words – hence a different language model
- Language model is a weighted set of words associated with toponym that is being modelled



## Albany, New Zealand

Albany is one of the northernmost suburbs of the contiguous [Auckland](#) metropolitan area in New Zealand. It is located to the north of the [Waitematā Harbour](#).

15 kilometres (9.3 mi) northwest of the Auckland city centre. The suburb is in the Albany ward, one of the thirteen administrative divisions of [Auckland Council](#). One of the city's newest suburbs, it was until relatively recently a town in its own right, and still has a feeling of not being truly a part of the city, which lies predominantly to the southeast of it. Much of the land to the north of Albany is still semi-rural.

The [Māori](#) name for the area was Okahukura (literally, 'place of rainbows' or 'place of butterflies'). The town was By 1890 it was a fruit-growing area and in that year it was renamed 'Albany' d ['Albany' in Australia](#), pronounced with a short 'a' as in [Albert](#).<sup>[4][2]</sup> The name for Scotland) and its Latinisation.

"happy mix of businesses, hotels, shops, apartments, and entertainment (including) an environment of parks and cycleways linking to the new [park-and-ride bus station](#) and the [rapid-ay to downtown Auckland](#)", according to a newspaper report.<sup>[3]</sup> It would wanted soundproof apartments against outside noise.<sup>[3]</sup> Initial plans swimming pool as well as the headquarters for the North Shore City nt has proceeded accordingly, but the 2008–09 economic downturn has blunted some of this activity.

## Demographics

The population was 3,057 in the 2013 Census, an increase of 888 from the 2006 Census.<sup>[4]</sup> There were 1,092 occupied dwellings in Albany in 2013,<sup>[4]</sup> and demographic makeup was 73% European, 5% Maori, 2% Pacific peoples, 22% Asian, 3% Middle Eastern/Latin American/African, and 1% other.<sup>[4]</sup> The median income of \$32,600 was higher than for the Auckland Region of \$29,600.<sup>[4]</sup> Unemployment in Albany was 7.0%, lower than the Auckland average of 8.1%.<sup>[4]</sup> 91% had internet access and 88% had cell phones.<sup>[4]</sup> Cars were prevalent.<sup>[4]</sup> A near majority (48%) were born overseas.<sup>[4]</sup> Ethnically, in keeping with the wider North Shore, Albany was predominantly [Pakeha](#) and Asian, and had a relatively high proportion of recent migrants from both elsewhere in New Zealand and overseas.

## Retail and commercial activity

Albany has become, in some respects, a substantial shopping and retail zone within the northwestern Auckland area.

The area (the future 'Albany Town Centre') is fast-growing in terms of its population and the development of the built environment, following planning decisions and land sales made by central and local governments in the 1980s and 1990s. Through the 1990s industrial and retail areas were rapidly produced, predominantly owned and occupied by local and foreign corporate capital. A major [shopping centre](#) hub was opened in the late 1990s and has since expanded, with [Westfield Albany](#) becoming New Zealand's largest shopping centre. The so-called supermall opened in August 2007 on McKinnon Drive costing \$210 million with 142 shops built by over 3500 workers, which features 1800 cinema seats and an indoor area of 7ha.<sup>[5]</sup> There is parking for 2300 vehicles.<sup>[5]</sup> [Kmart](#), [Farmers](#), and [New World](#) stores are anchors.<sup>[5]</sup> The mall claimed it provides "free space for community organisations for awareness and fundraising activities" but one volunteer claimed he was ejected from the premises while trying to raise money for veterans because of a dispute with mall management.<sup>[6]</sup>

# Disambiguation with language models



Given an ambiguous name in a document:

- Match the words in the document to the language models each instance of ambiguous toponym
- Assign ambiguous toponym to the instance that has language model most similar to the text in the document.

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

Westfield's new Albany Mall is their first NZ

situated 20.8ha s Albany. one of the larger malls in the Southern Hemisphere

## Albany, New Zealand

From Wikipedia, the free encyclopedia

For the electorate, see Albany (New Zealand electorate).

For other places named Albany, see Albany.

Albany is one of the northernmost suburbs of the contiguous Auckland metropolitan area in New Zealand. It is located to the north of the Waitematā Harbour,

15 kilometres (9.3 mi) northwest of the Auckland city centre. The suburb is in the Albany ward, one of the thirteen administrative divisions of Auckland Council. One of the city's newest suburbs, it was until relatively recently a town in its own right, and still has a feeling of not being truly a part of the city, which lies predominantly to the southeast of it. Much of the land to the north of Albany is still semi-rural.

The Māori name for the area was Okahukura (literally, 'place of rainbows' or 'place of butterflies'). The town was originally known as Lucas Creek. By 1890 it was a fruit-growing area and in that year it was renamed 'Albany' after the fruit-growing district called 'Albany' in Australia, pronounced with a short 'a' as in Albert.<sup>[1]</sup> The name Albany derives from Alba (Gaelic for Scotland) and its Latinisation.

Contents [hide]
1.1 Demographics
1.2 Retail and commercial activity
1.3 Residential real estate
1.4 Sports
1.5 Transportation
1.6 Parks

## Albany, New York

From Wikipedia, the free encyclopedia

For other uses, see Albany (disambiguation).

Albany (<sup>1</sup>əlˈbənē/ <sup>2</sup>əlˈbe-nē) is the capital of the U.S. state of New York and the seat and largest city of Albany County. Albany is located on the west bank of the Hudson River approximately 10 miles (16 km) south of its confluence with the Mohawk River and approximately 135 miles (220 km) north of New York City.<sup>[2]</sup>

Albany is known for its rich history, commerce, culture, architecture, and institutions of higher education. Albany constitutes the economic and cultural core of the Capital District of New York State, which comprises the Albany–Schenectady–Troy, NY Metropolitan Statistical Area, including the nearby cities and suburbs of Troy, Schenectady, and Saratoga Springs. With a 2013 Census-estimated population of 1.1 million<sup>[3]</sup> the Capital District is the third-most populous metropolitan region in the state. As of the 2010 census, the population of Albany was 97,856.

The area that later became Albany was settled by Dutch colonists who, in 1614, built Fort Nassau for fur trading and, in 1624, built Fort Orange. In 1664, the English took over the Dutch settlements, renaming the city as Albany, in honor of the then Duke of Albany, the future James II of England and James VII of Scotland. The city was officially chartered in 1686 under English rule. It became the capital of New York in 1797 following formation of the United States. Albany is one of the oldest surviving settlements of the original British thirteen colonies, and is the longest continuously chartered city in the United States.<sup>[10]</sup>

During the late 18th century and throughout most of the 19th, Albany was a center

# References

Purves R.S, P.D. Clough, C.B. Jones, M.M. Hall, M. Murdoch (2018) “Geographic Information Retrieval: Progress and challenges in spatial search of text”. Foundations and Trends in Information Retrieval. Vol. 12, No. 2-3, pp 164-318.

F. Melo and B. Martins (2017). Automated geocoding of textual documents: A survey of current approaches. *Transactions GIS*, 20(1), 3–38.

J. L. Leidner and M. D. Li

place names and associate <https://eduassistpro.github.io/>

G. DeLozier, J. Baldridge, L. London (2015) G  
Using Geographic Word Profiles. AAAI: P  
Conference on Artificial Intelligence: 2382-238

Geonames Gazetteer - <https://www.geonames.org>

AllenNLP NER demo - <https://demo.allennlp.org/named-entity-recognition/>

Nominatum Geocoding API - <https://nominatim.org/>

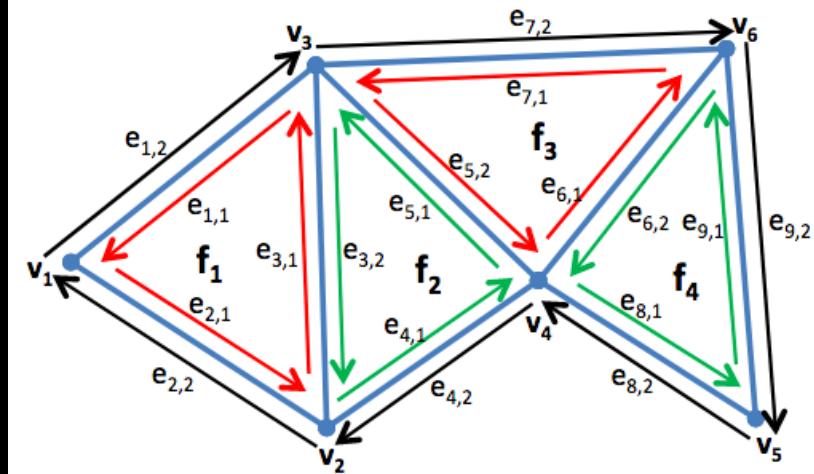
GeoPy Python library for geocoding - <https://geopy.readthedocs.io/en/1.22.0/>

SpaCy guidance on NER (and other NLP) - <https://spacy.io/usage/linguistic-features>

physical references in the form of  
TAL Special, 3(2):5–11.

dent Toponym Resolution  
Twenty-Ninth AAAI

Add WeChat [edu\\_assist\\_pro](https://edu_assist_pro)



# Large Scale Databases

<https://eduassistpro.github.io/>

Topologically Structured Data  
Add WeChat [edu\\_assist\\_pro](http://edu_assist_pro)

Chris Jones

# Topologically Structured Vector Data



- records explicitly what geometry is connected to what (independent of distance and orientation),  
e.g..
  - adjacency between land parcels,
  - roads and rivers <https://eduassistpro.github.io/>
- Assignment Project Exam Help  
Add WeChat edu\_assist\_pro
- many variations on what information is stored, depending on purpose
  - e.g. polygon map, network, triangulated network

# Topological primitives for 2D space



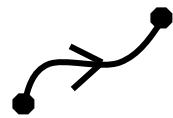
Node - isolated point or terminus of arc  
0-dimensional;



Arc (link) - (1-dimensional);

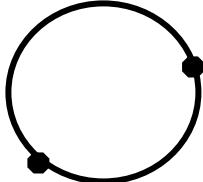
**Assignment Project Exam Help**

– 1 by two nodes;



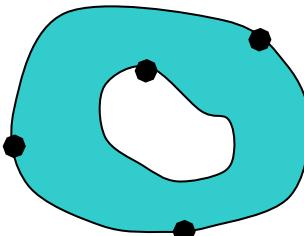
Dir <https://eduassistpro.github.io/> (1-dimensional);

– **AddWeChat edu\_assist\_pro**



Ring (1-dimensional);

– circuit of one or more arcs



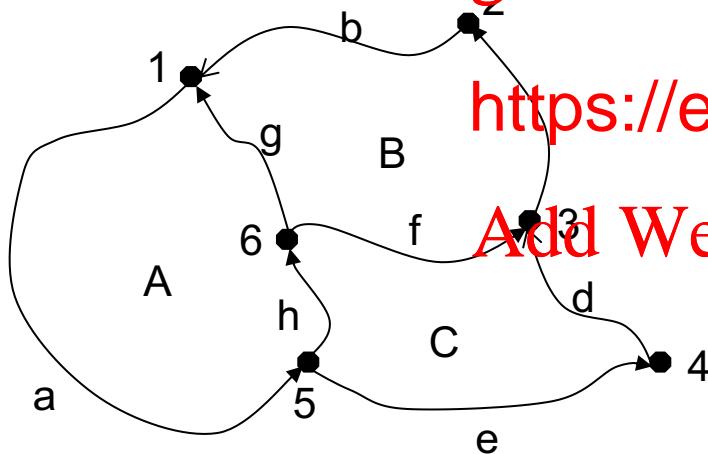
Face (2-dimensional)

– region bounded by rings (one outer and 0 or more inner rings, i.e. holes)

# Polygon Map



could be used, for example, to record administrative boundaries or land use categories



Assignment Project Exam Help  
<https://eduassistpro.github.io/>  
Add WeChat edu\_assist\_pro

Polygon (face)	Bounding arcs
A	a h g
B	f c b -g
C	e d -f -h

Arc			Left face	Right face
a	1	5	A	
b	2	1	B	
c	3	2	B	
d	4	3	C	
e	5	4	C	
f	6	3	B	C
g	6	1	A	B
h	5	6	A	C

*A simple approach to representing the relations.  
Can be more complex - see next slide*

# Doubly Connected Edge List (DCEL)



Assignment Project Exam Help

<https://eduassistpro.github.io/>

$f_5$

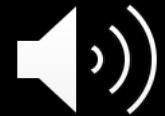
<http://www.cs.sfu.ca/~binary/813.2011/DCEL.pdf> Add WeChat edu\_assist\_pro

Half-edge	Origin	Twin	Incident Face	Next	Previous
$e_{3,1}$	$v_2$	$e_{3,2}$	$f_1$	$e_{1,1}$	$e_{2,1}$
$e_{3,2}$	$v_3$	$e_{3,1}$	$f_2$	$e_{4,1}$	$e_{5,1}$
$e_{4,1}$	$v_2$	$e_{4,2}$	$f_2$	$e_{5,1}$	$e_{3,2}$
$e_{4,2}$	$v_4$	$e_{4,1}$	$f_5$	$e_{2,2}$	$e_{8,2}$
....	....	....	....	....	....

Vertex	Coordinates (approx.)	Incident Edge
$v_1$	0,3	$e_{1,2}$
$v_2$	3,0	$e_{2,2}$
$v_3$	3,5	$e_{7,2}$
4	6,3	$e_{4,2}$
....	....	....

e	Outer Component	Inner Components
$f_1$	$e_{1,1}$	nil
$f_2$	$e_{4,1}$	nil
$f_3$	$e_{5,2}$	nil
$f_4$	$e_{6,2}$	nil
$f_5$	nil	$e_{1,2}$

# Linear Network



For finding routes and  
shortest paths

Assignment Project Exam Help  
defined by

<https://eduassistpro.github.io/>  
rt node; end

Node Table

Node id	connected nodes (arcs)
1	2(b) 5(a) 6(g)
2	3(c) 1(b)
3	4(d) 2(c) 6 (f)
4	5(e) 3(d)
5	4(e) 6(h) 1(a)
6	3(f) 1(g) 5(h)

Add WeChat edu\_assist\_pro  
no d by

node id.;

ordered list of connected nodes  
or of connected links.

Links may be accompanied by cost  
/ impedance values such as ?,  
and by direction restrictions

# Large Scale Databases

Assignment Project Exam Help:  
Spatial Databases:

G <https://eduassistpro.github.io/> ns

Add WeChat edu\_assist\_pro

Chris Jones

# Who uses GIS?

GIS are used by range of public, commercial and industrial services as well as private individuals – you and me - who want to find out about places and things that are in geographic space.

<https://eduassistpro.github.io/>

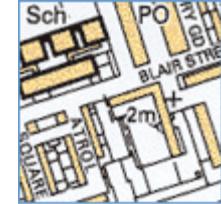
In practice they are used by  who needs access to geographical info

# Topographic and Land Survey Organisations



Provide base mapping for wide range of applications

- major users are local government and utilities



## Types of information represented

- buildings; land boundaries; rivers, coastline, lakes; spot heights; roads, railways, canals; forests, woodlands; tourist facilities

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

## Example organisations:

- Ordnance Survey  
Vector Digital maps (1:1250, 1:2500, 1:10,000, 1:250000, 1:625000 ) + raster maps; terrain models
- AA - Vector and raster products
- Harper Collins/Bartholomews
- USGS (United States Geological Survey);
- US Bureau of Census: TIGER/Line files (urban)



# Ordnance Survey Digital Maps at scales

1:625,000; 1:250,000

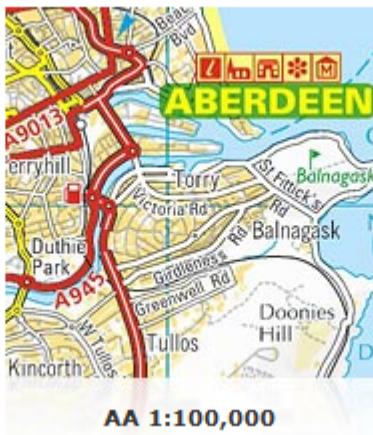
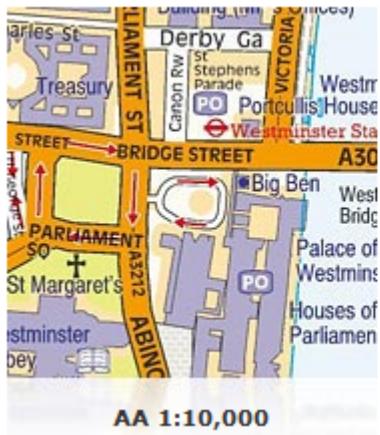
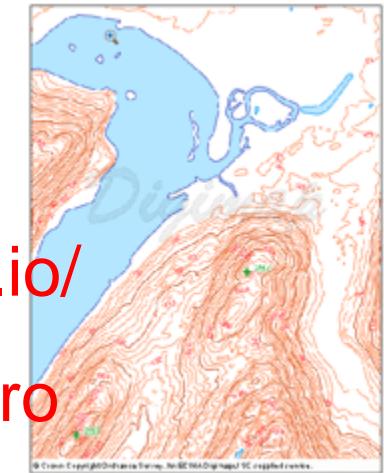


1:10,000, 1:1250 + Terrain data

## Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro



AA Digital map  
data at  
different scales

92<sup>92</sup>

# Topographic and Land Survey Organisations



## Example GIS Activities

- Update existing commercial map products to reflect changes
- Create maps ~~Assignment Project Exam Help~~ (could require “general” <https://eduassistpro.github.io/>)
- Create a detailed terrain model ~~Add WeChat edu\_assist\_pro~~ sed development area





# Web 2.0 sources of geo-base data

## Open Street Map ([www.openstreetmap.org](http://www.openstreetmap.org))

User-contributed map data from GPS and aerial photography

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

## GeoNames Gazetteer ([www.geonames.org](http://www.geonames.org))

User-contributed information on place names,  
alternative names, country, coordinates (footprint), feature type<sup>94</sup>

# Environmental / Geoscientific Survey Agencies



- Uses of GIS include:

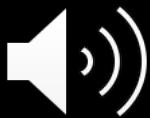
- creating databases of sample data and classified data
- interpretation, modelling and analysis of sample data
- creation of maps and reports

Assignment Project Exam Help

Add WeChat edu\_assist\_pro



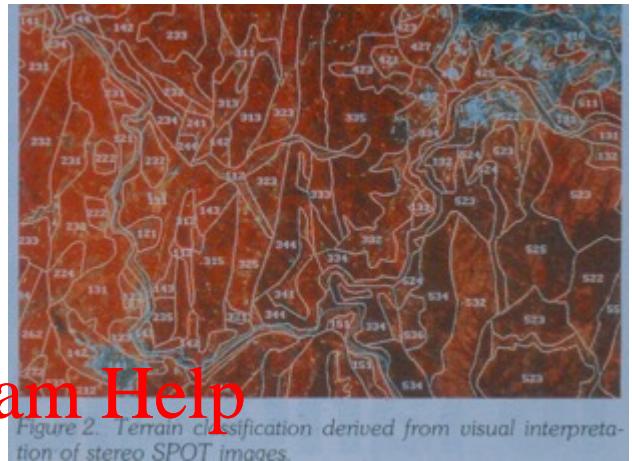
Large Scale Databases : C.B. Jones



- Soils
    - erosion susceptibility, crop suitability

- Geology

- outcrop maps; earthquake events; **Assignment Project Exam Help** 3D models of borehole, seis <https://eduassistpro.github.io/> and resistivity studies



**Figure 2.** Terrain classification derived from visual interpretation of stereo SPOT images.

# Ecology



- identify environmentally sensitive areas;
- relationships between flora and fauna and environmental factors



Assignment Project Exam Help

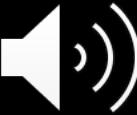
<https://eduassistpro.github.io/>



Hydrology (River Coastal Zones) Add WeChat edu\_assist\_pro

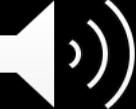
- create valid terrain models
- analyses: e.g. drainage basins, runoff, pollutant monitoring

# Government Administration and Planning



- Data on property type, ownership, valuations, transactions;
  - *which properties do we own and what is their value?*
- Maintenance of roads, buildings, parks
- Example analysis
  - development
  - deprivation / needs
  - emergency response strategies
    - *where is a good place to create a new industrial site?*
    - *which properties are most at risk if the chemical plant emits toxic gases?*

# Utilities / Facilities Management

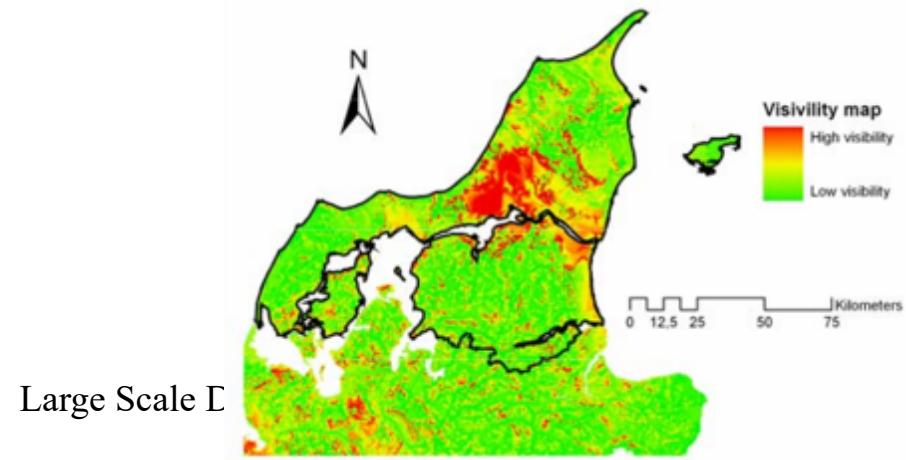
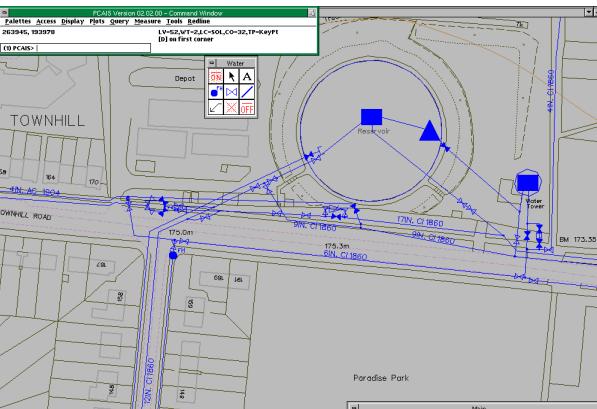


gas, electricity, water, sewers, telecomms

- databases for maintenance, customers, valuation
  - *where should we dig in Mary Street to find the source of the gas leak?*

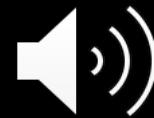
## Assignment Project Exam Help

- Example analyse
  - *where are the best properties available* [https://eduassistpro.github.io/ines?](https://eduassistpro.github.io/)
  - ...which properties [Add WeChat edu\\_assist\\_pro](#)



Large Scale L

# Navigation



- Route maps
  - Static / mobile
- Driving instructions
- Analyses of optimal routes (various)

*Assignment Project Exam Help  
What requirements for  
turning the data?*

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro





- analysis: store siting, targeting sales
  - *will it be profitable to build a Morrisons in Bogford?*
  - *to which Assignment Project Exam Help centre send adver*
- produce 'mo  
classifying  
economic status

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

*Where does the  
data come from?*

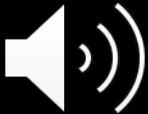
# Healthcare



- analyses:
  - assessment of needs
  - epidemiological studies
    - identification of clusters
    - correlation and socio-e
  - location of health centres
- Data sources
  - census, topographic survey,
  - medical records from hospitals
  - GP records



# Crime and Law Enforcement



- Analyses : identify hotspots, in order to plan policing / detection activities;
- Correlation of crime with demographic factors
- Monitoring tagged individuals with ~~Assignment Project Exam Help~~

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro



- E.g. Geo-tagging photos with location (e.g. Instagram, Flickr) – finding pictures about particular places

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

# Web 2.0 / Social Web thematically specific data



**“Volunteered geographic information”**

Add features to existing base maps using Google Maps,  
GoogleEarth etc

e.g.

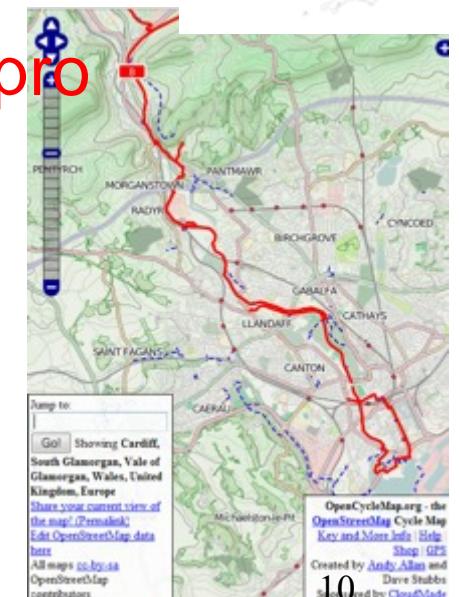
Assignment Project Exam Help

<https://eduassistpro.github.io/>

- Bird species distrib  
(citizen science)

[Add WeChat edu\\_assist\\_pro](#)

- Cycle maps (with OpenStreetMap)  
<http://www.opencyclemap.org/>



- Running and cycling : <https://www.strava.com/>

# Large Scale Databases

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro  
Regular



# Indexing Coordinate-based Spatial Data

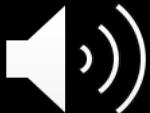
## Requirement

- fast access to spatial data relating to a location specified by coordinates in 2 dimensions

## Problem

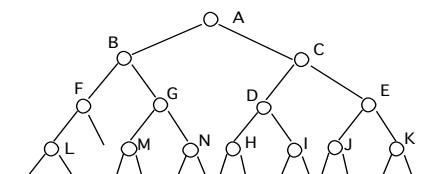
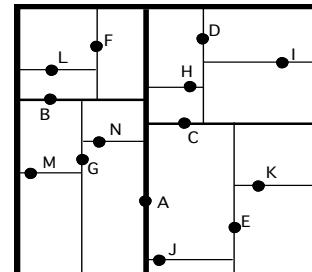
Assignment Project Exam Help

- Conventional (one-dimensional index) <https://eduassistpro.github.io/> structures) that allow fast access to data specified by a scalar such as a name or a unique identifier
- OK for indexing on place names, but geometric data has at least 2 dimensions
- Create new methods or adapt the existing ones
  - both done in practice



# 2 approaches to spatial indexing

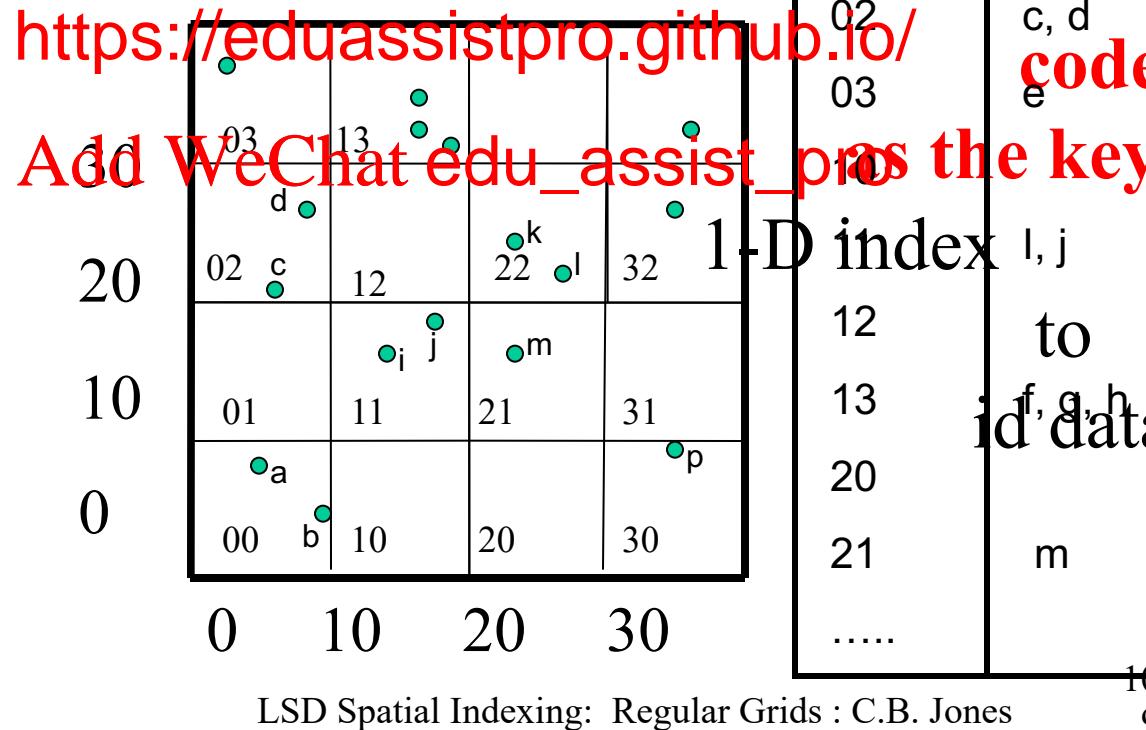
- Space-directed
  - divide the region of interest into uniform cells and record what is in each cell (analogous to raster data model). Index the content of cells. Cell size
  - e.g. regular grid,
- Object-directed
  - place bounding boxes around objects and index the content of the boxes, e.g. R-Tree
  - Various other approaches, e.g. k-D tree in which the locations of points split space vertically and horizontally – but less suited to DBMS





# Regular Grid

- Each cell is given unique numerical id i.e.  
a ***location code***, derived typically from the coordinates of its lower left corner
- Entries in each index record refers geometry in corr
- **Use location : cell\_id** for method (B-tree) access geometry



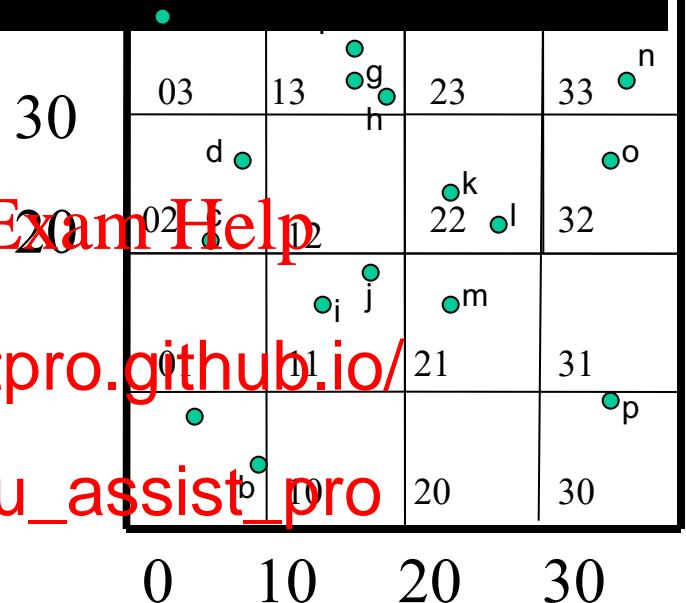
# Location Codes: x, y



- **concatenation**

codes is to concatenate the x and the y coordinates

- alternative is bit interleaving  
(see Quadtree Lect  
<https://eduassistpro.github.io/>



- Allocate a fixed length to the x and y coordinate values.

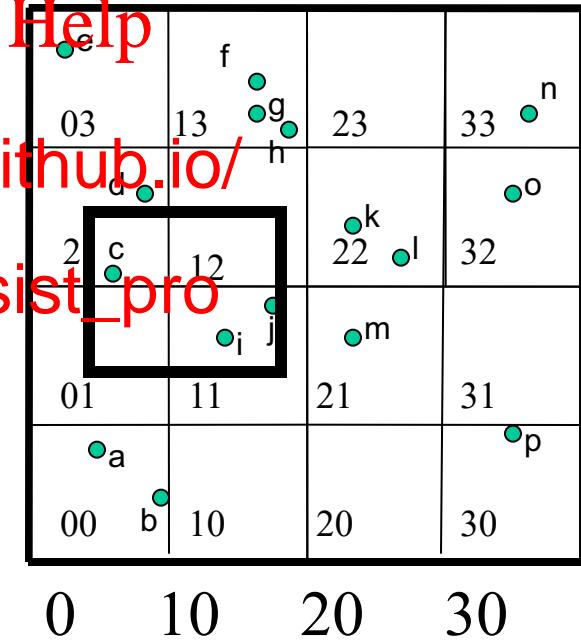
In example, concatenation involves initially dividing the coordinates by grid size – as they are always multiples of 10

More generally: set fixed length to x and y and pad to left with zeros, before concatenation.

# Retrieving data inside a rectangular window from regular grid indexed data



- Given a search window, calculate which cells it overlaps.  
example : 01,02,11,12
- Retrieve corresponding spatial objects ( <https://eduassistpro.github.io/> )
- Might need to perform a second access to the file(s) containing the geometric data.
- Omit geometry outside window (*refinement stage*)
- To insert data, determine containing cells and add data to indexed records



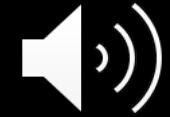
# Large Scale Databases

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro  
Quadt

# Quadtree Indexing

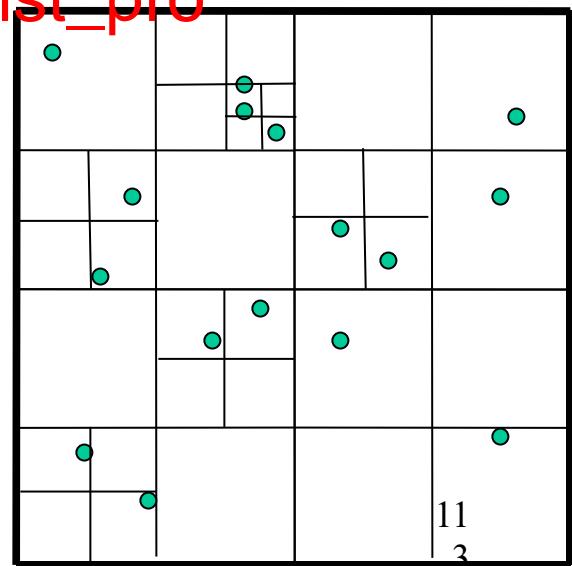
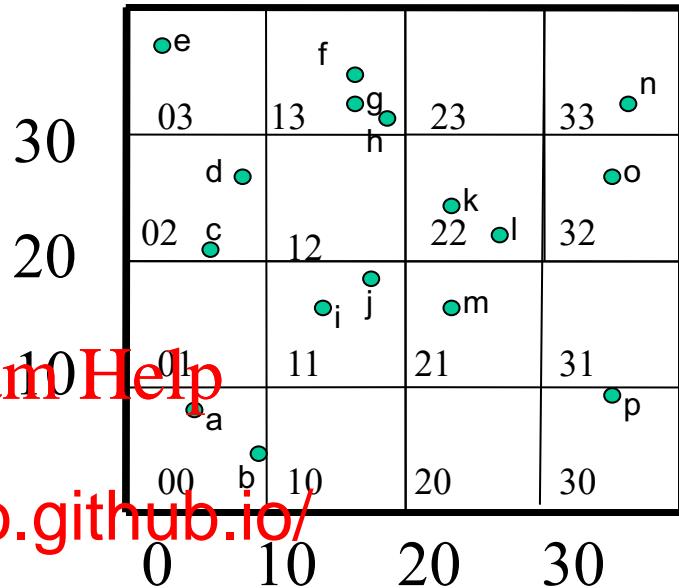


- Issues with regular grid
  - *Data per cell?*
  - *Consequences for storage and processing?*

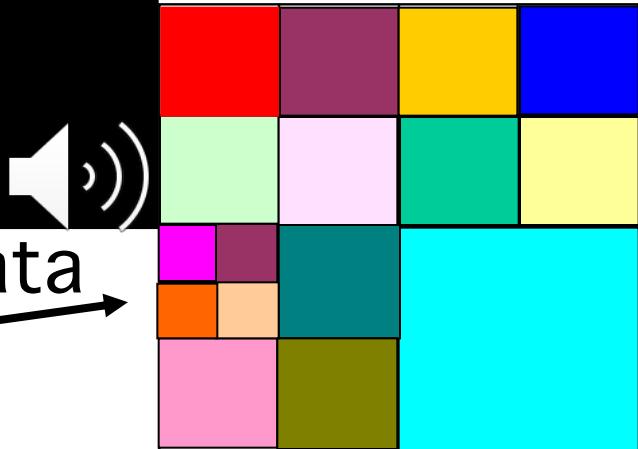
Assignment Project Exam Help

- Quadtree address <https://eduassistpro.github.io/> problems by splitting a cell recursively into 4 if some storage threshold is exceeded.

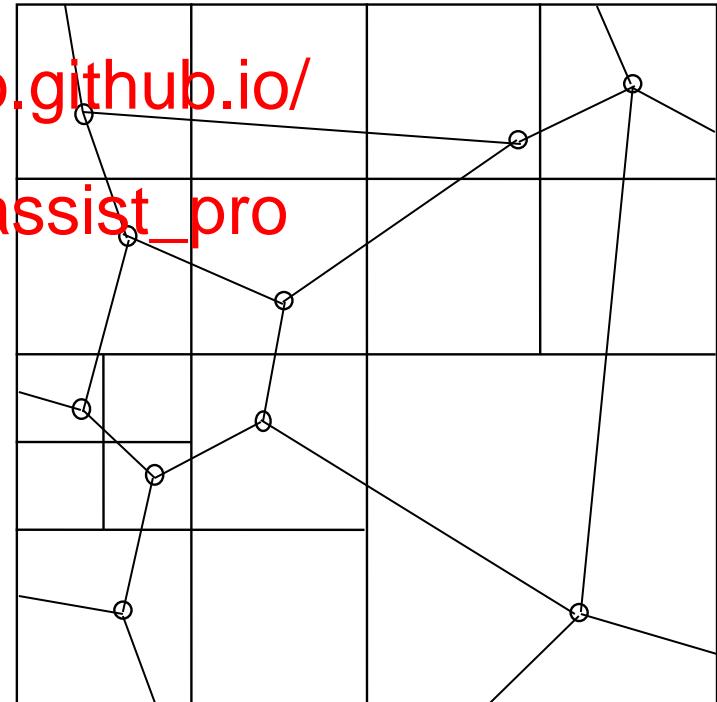
In this example the threshold is 1 point per cell



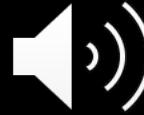
# Example types of quadtree



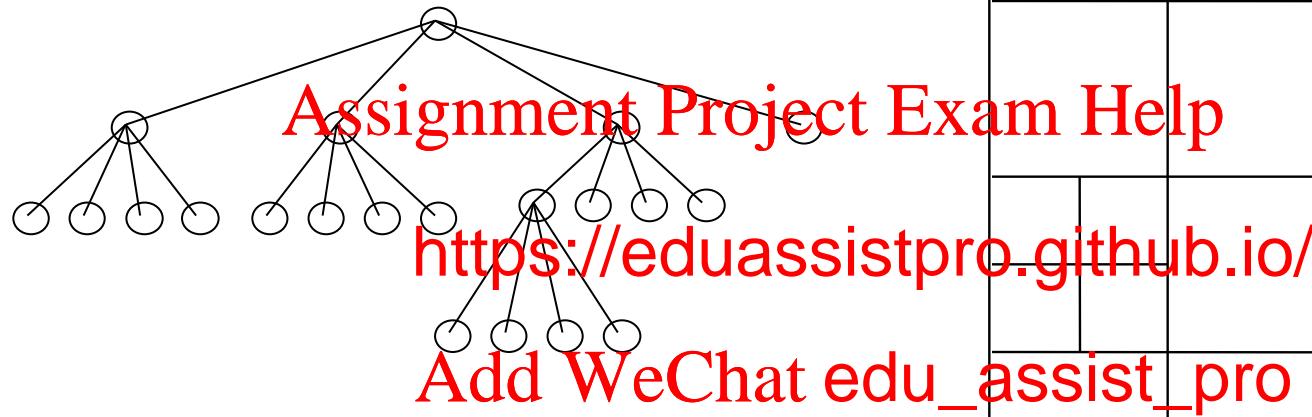
- *Region quadtrees* store raster data
  - cell has uniform pixel value
- *PR quadtrees* store points
  - cell stores Assignment (one point or more if higher threshold)
- *PM quadtree* <https://eduassistpro.github.io/>
  - cells refer to
  - thresholds: one complete intersecting edge per cell or one vertex per cell (plus its incident edges)
  - geometry stored separate from quadtree cell index
- Lots of other quadtrees



# Pointer-based quadtrees vs (Pointerless) linear quadtrees



- earliest quadtrees were all pointer-based tree structures



- not well adapted to database storage (e.g. variable number of levels)
- Alternative is to use location codes to identify cells
- Index cells with the location codes (=keys) - > linear quadtrees (can adapt to standard database indexing methods)

# Linear quadtree



Cells numbered with Morton Numbers (location codes) according to their ancestry in the quadtree.

Cells (identified with location codes) indexed with B-tree or similar

Two possible approaches to create numbers (there are others):

Decide on extent of quadtree as power of 2 (determines length of codes – e.g. one bas

<https://eduassistpro.github.io/>

Hierarchically based on quadtree

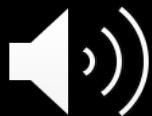
Add WeChat [edu\\_assist\\_pro](https://edu_assist_pro)

assume quadrants numbered 0 1 2  
in order, e.g. lower left, upper left,  
lower right, upper right. (base 4).

Add respective digit for each  
sub-quadrant and follow with trailing  
digits. See diagram in which  
zeros are used as trailing digits.

113	131	133	311	313	331	333	
130	310	312	330	332			
130	132	310	312	330	332		
00	300	300	320	322			
101	103	121	123	301	303	321	323
100	102	120	122	300	302	320	322
011	013	031	033	211	213	231	233
010	012	030	032	210	212	230	232
000	000	000	000	200	202	220	222
001	003	021	023	201	203	221	223
000	002	020	022	200	202	220	222

# Order of Morton codes / numbers



[Here using decimal numbers - *These are not the actual Morton numbers – just their order*]

5	7	13	15
4	6	12	14
1	3	9	11
0	2	8	10

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat `edu_assist_pro`

21	23	29	31	53	55	61	63
	8	30	52	54	60	62	
	5	27	49	51	57	59	
	26	48	50	56	58		
5	7	13	15	37	39	45	47
4	6	12	14	36	38	44	46
1	3	9	11	33	35	41	43
0	2	8	10	32	34	40	42



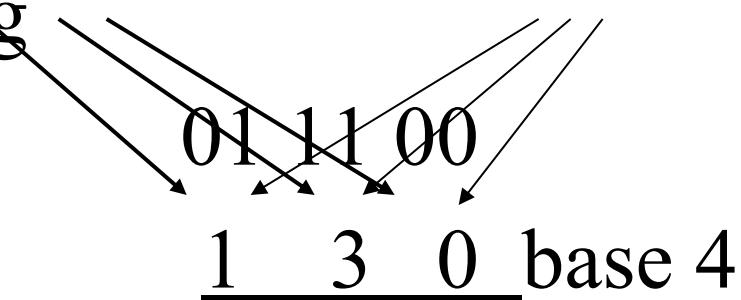
# Bit Interleaving

- Transforms a pair of coordinates into a single number
- Convert x and y coordinates of a point to binary numbers
- Create new number by alternating bits from each coordinate, starting with X coordinate

Assignment Project Exam Help

- Eg. Assuming <https://eduassistpro.github.io/> (width = 8)  
 $X : 2 = 010$       ~~Y : 6 = 110~~  $\rightarrow$  WeChat edu\_assist pro

interleaving





# Linear Quadtree addressing

Problem with the interleaved (base 4) numbers.

*Are they unique?*

Solve either by:

adding cell size to addresses;

or use base 5 numbers for quadrants,  
i.e. numbering cells 1,2,3,4 and use 0 for trailing digits.

- add 1 to each of t

<https://eduassistpro.github.io/>

base 4 digits - except ~~Add trailing digits~~ [edu\\_assist\\_pro](https://edu_assist_pro)

01 11 00

= 1 3 0

= 2 4 0 (if cell is level L = 2 )

= 2 4 1 (if cell is level L = 3 )

*Level L in quadtree is given by n-k where width of tree is  $2^n$  and there are k (= n-L) trailing digits*

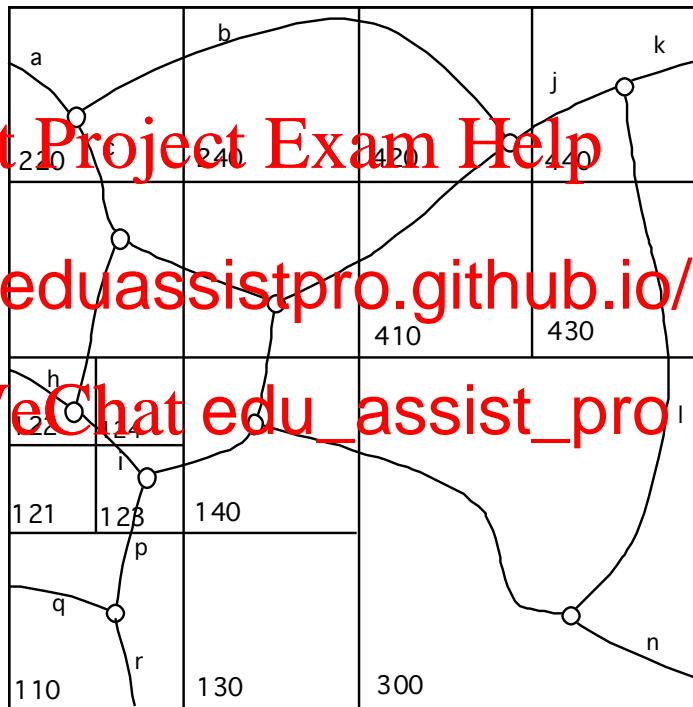
111	113	131	133	311	313	331	333
110	112	130	132	310	312	330	332
		100			300		
101	103	121	123	301	303	321	323
100	102	120	122	300	302	320	322
		100	120	300	320		
011	013	031	033	211	213	231	233
010	012	030	032	210	212	230	232
		030		210		230	
		000			200		
003	021	023		201	203	221	223
000	020	022		200	202	220	222
002							

224	242	244		422	424	442	444
221	223	241	243	421	423	441	443
		240		420		440	
		200			400		
212	214	232	234	412	414	432	434
210		230		410		430	
211	213	231	233	411	413	431	433
122	124	142	144	322	324	342	344
120		140		320		340	
121	123	141	143	321	323	341	343
		100			300		
112	114	132	134	312	314	332	334
110		130		310		330	
111	113	131	133	311	313	331	333

# Example quadtree for storing linear features



- Quadtree is of size 8X8  
→ n=3
- Here each cell references entire features that intersect
- Cells subdivide if more than one vertex or if no vertex and more than one line feature



QUADTREE INDEX FILE	
Quadtree address	Geometry ids
110	r,p,q
121	
122	h,f,i
123	i,p,o
124	i
130	
140	o,e,m
210	f,c,g
220	a,b,c
230	g,e,d
240	b
300	m,n,l
410	d
420	b,d,j
430	l
440	j,k,l



# Exercise with location codes

- Use bit interleaving to derive the base 5 code of the highlighted cell

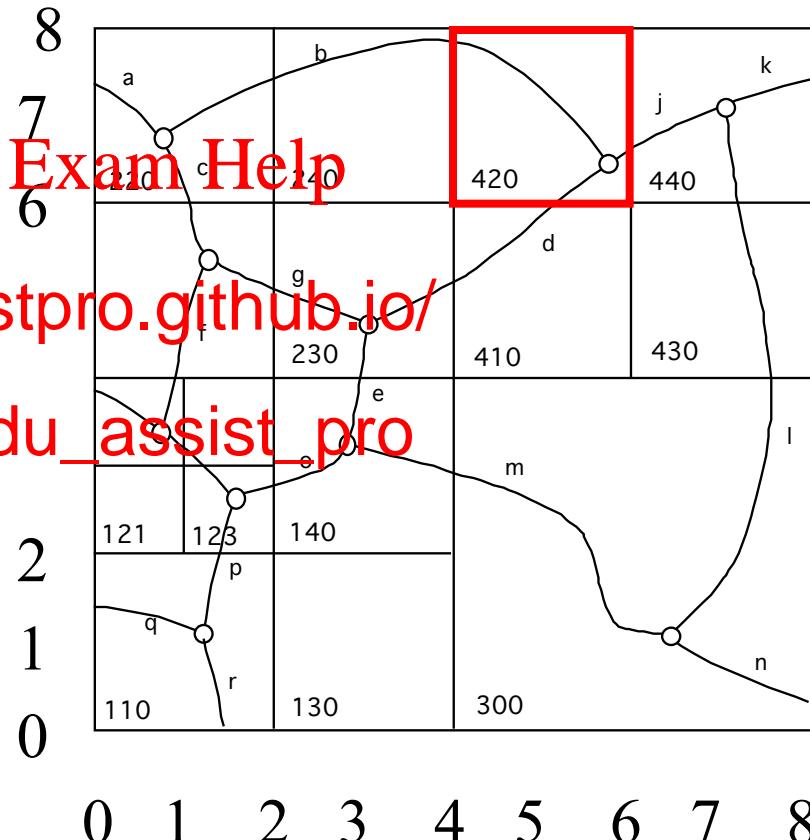
(second level of subdivision so  $L = 2$ )

Assignment Project Exam Help  
<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

Note : 1<sup>st</sup> level of subdivision is  $L = 1$

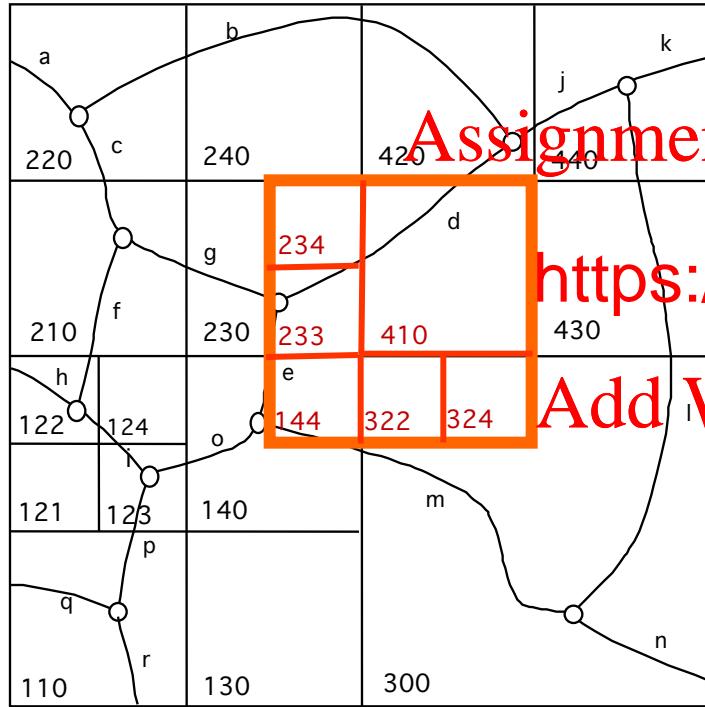
Maximum level of subdivision (smallest cell size) here is  $L=3$



# Querying Linear Quadtrees with query quadtree blocks



- Convert Query window to maximal quadtree cells (“blocks”)
- Match the query cells with the data cells
- inside / equal / contains



<https://eduassistpro.github.io/>

322 →  
324 →

Add WeChat [edu\\_assist\\_pro](#)

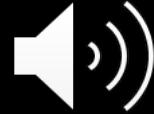
410 → 410 → d

Then retrieve referenced geometry objects  
for each data cell and test against the query  
window

*Note: this is just one possible  
approach to querying  
quadtrees*

N.B. Check for duplicate data cells to avoid  
retrieving same data

# Ideal cell numbering systems?



Cells that are adjacent in space should have numbers that are adjacent.

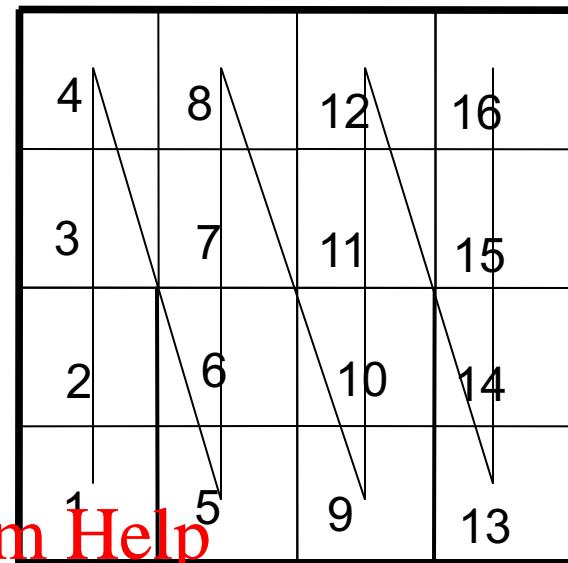
and numbers that are adjacent should refer to locations that are close in space

Assignment Project Exam Help

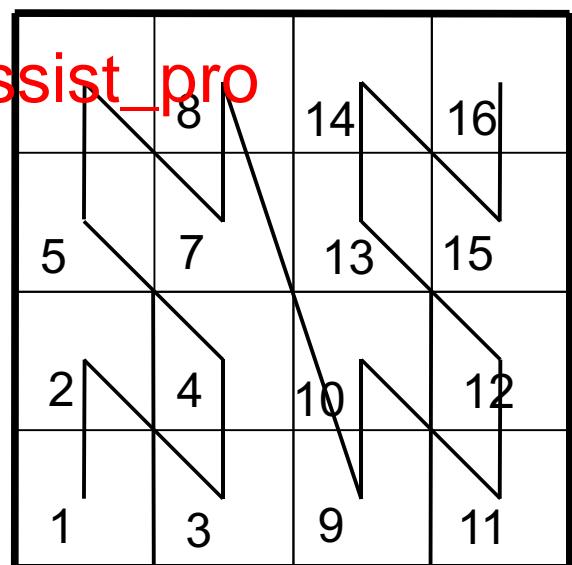
<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro  
data on disk that is sorted by the cell numbers will then tend to be adjacent on disk / storage (hence minimise search between different parts of storage)

Ideally a query window will only need to access a single region of the storage



Key order based on x,y concatenation



Key order based on x,y bit interleaving

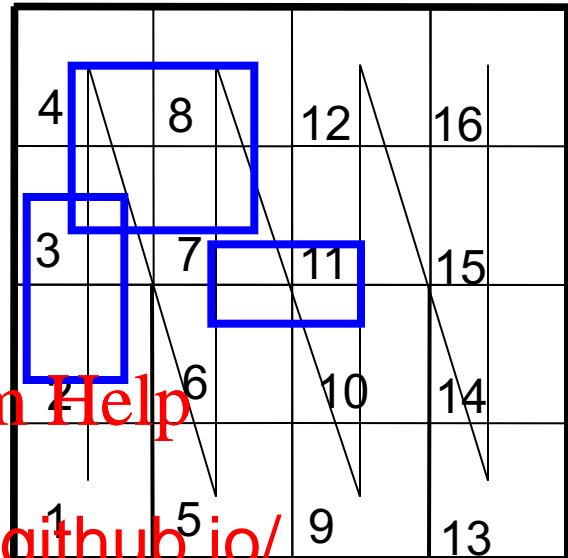
# Space filling curves



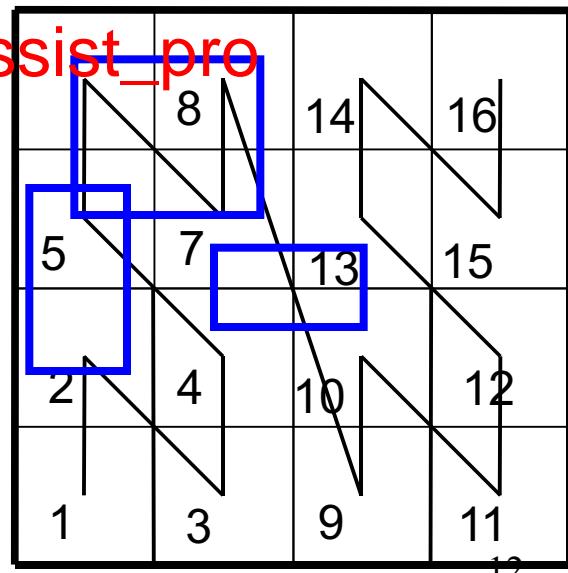
- column-order numbering
  - in general two cells adjacent in a column will have adjacent numbers, but cells adjacent in rows will not normally have adjacent numbers.

Assignment Project Exam Help

<https://eduassistpro.github.io/>



- Morton Numbers Add Z number WeChat edu\_assist\_pro  
(bit interleaving)
  - increases the chances of spatially adjacent cells having similar numbers, but does not solve the problem entirely.





# Hilbert Curve

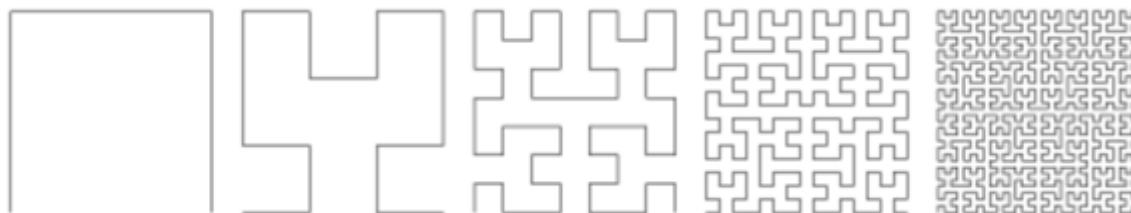
## Hilbert Curve

Improves on Morton numbers

/ Z-curve, though more complicated to

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro



<http://mathworld.wolfram.com/HilbertCurve.html>

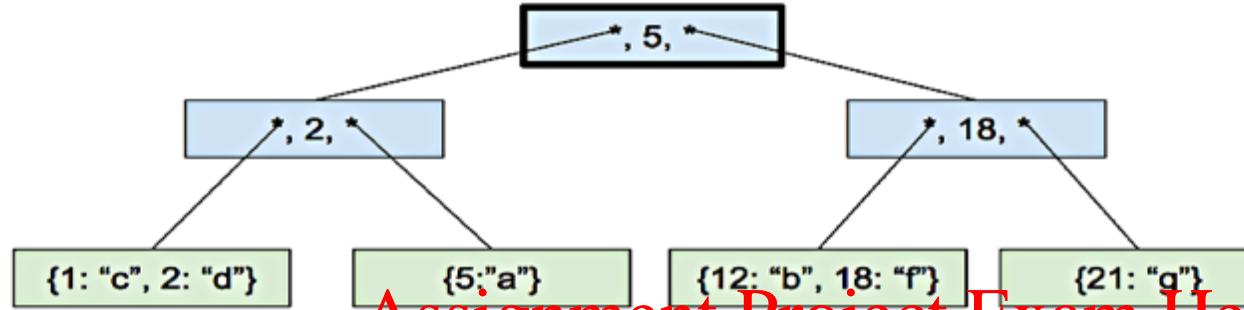
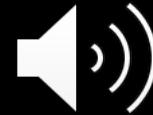
# Large Scale Databases

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat [edu\\_assist\\_pro](#)  
Object-directed Spatial Indexing  
with R-trees

# Revision of B-tree / B<sup>+</sup>-tree



Assignment Project Exam Help

- Balanced, i.e. all
- Each non-leaf no  
minimum of  $\lceil m/2 \rceil$  children
- Each non-leaf node has maximum  $m-1$  keys
- Leaf nodes can store  $m$  keys and  $m$  data values
- B<sup>+</sup>-tree non-leaf nodes store key + child pointers  
Leaf nodes store entries of <key + value>
- Root has at least 2 children

Example  
B<sup>+</sup>-tree  
[B-tree stores  
values in non-  
leaf as well as  
leaf nodes]

1

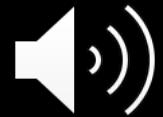
<https://eduassistpro.github.io/>

= order) and a



# Object-directed indexing with R-Trees

- Multi-dimensional indexing methods which find data within specified ranges of coordinates on the axis of each dimension.
- 2D R-Tree indexes rectangles = bounding boxes of spatial objects (i.e. minimum bounding rectangles **MBR**)
- Extension of B<sup>+</sup>  
**Assignment Project Exam Help**
- The R-tree creates a  
of  
MBRs of objects. <https://eduassistpro.github.io/>
- The leaf nodes contain  
the entries w  
**Add We Chat edu\_assist\_pro**  
(R7-R16) that enclose the  
original  
geometries.
- At higher levels of the hierarchy  
MBRs enclose sets of  
lower level MBRs.



# R-trees properties

Each entry in a node/page stores a rectangle and a pointer or ID

$$|\text{children}| = |\text{keys}|$$

Each non-leaf node

stores a maximum  
[Assignment](#) [Project](#) [Exam](#) [Help](#)

of  $M$  references to

child MBRs

( $M=3$  in example)

<https://eduassistpro.github.io/>

Add WeChat [edu\\_assist\\_pro](#)

Each node

(apart from root)

has minimum of

( $\lceil M/2 \rceil$ ) entries

Root has at least 2  
entries

Non-leaf node entries  
reference groups of  
child rectangles

Leaf node entries  
reference geometry  
data records



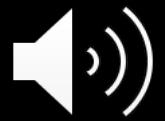
# Example from R\*Tree:

[https://en.wikipedia.org/wiki/R-tree#/media/File:R\\*-tree\\_built\\_using\\_topological\\_split.png](https://en.wikipedia.org/wiki/R-tree#/media/File:R*-tree_built_using_topological_split.png)

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro



# Searching R-trees

- To retrieve data from the R-tree : start at the root and test the search window for overlap with the MBRs of each entry in the root.
- Search descends recursively down every branch for which there is ov

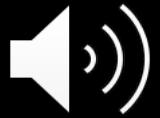
<https://eduassistpro.github.io/>

Qu

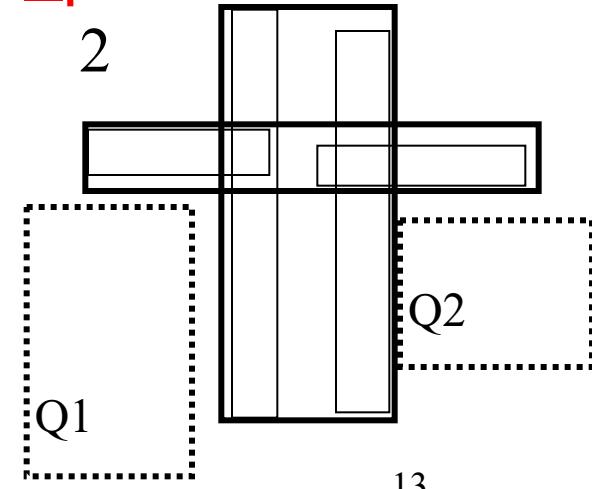
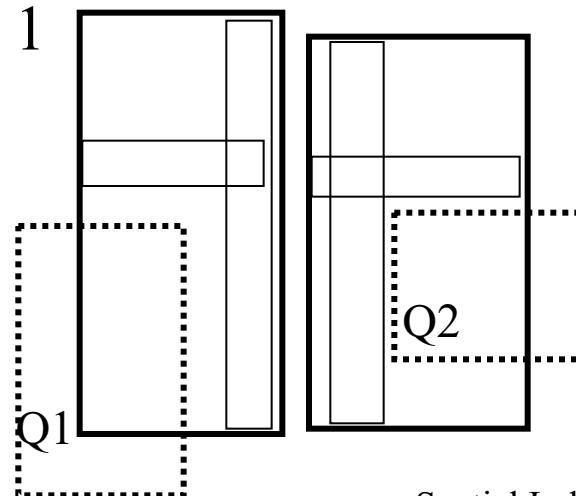
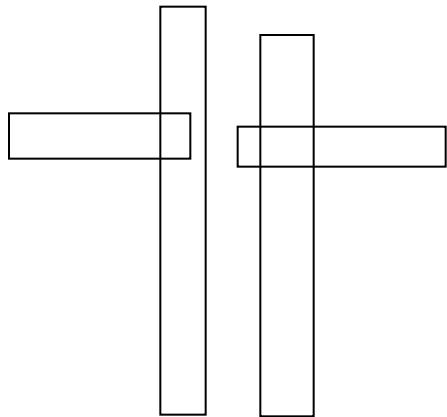
Add WeChat edu\_assist\_pro



# Efficient R-tree structure



- Objective : minimise the number of fruitless searches down branches of the tree
- Figure shows two alternatives for constructing bounding rectangles around 4 MBRs, with  $M=3$   
Assignment Project Exam Help  
In option 1, query rectangle Q1 would result in fruitless searches.  
<https://eduassistpro.github.io/>  
In option 2, these searches would be avoided.  
Add WeChat edu\_assist\_pro

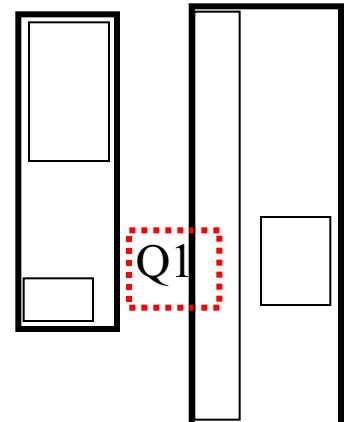
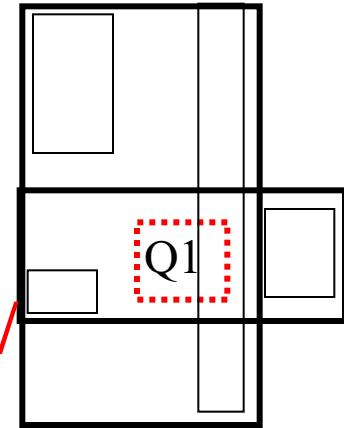


# Methods to improve efficiency

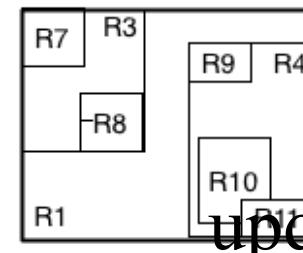


- Minimise area (*coverage*) of MBRs
  - reduces chances of query rectangle covering empty regions of an MBR
- Minimise *overlap* of MBRs
  - query windows covering the overlapping areas will always result in the tree - even though the objects.
- Minimise *size of tree*
  - fewer branches to traverse
- BUT the methods may counteract each other

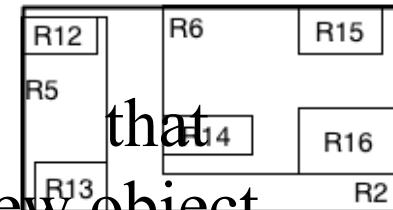
Assignment Project Exam Help  
<https://eduassistpro.github.io/>  
Add WeChat edu\_assist\_pro



# Building R-Trees



updates



that  
new object

Assignment Project Exam Help

strategy

- If leaf node is full
- Various strategies to minimise coverage and overlap

Add WeChat [edu\\_assist\\_pro](#)

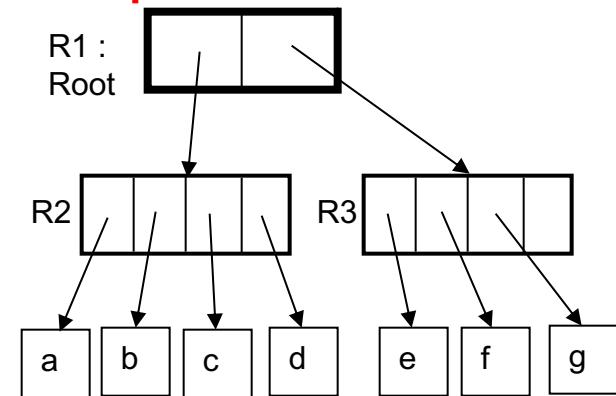
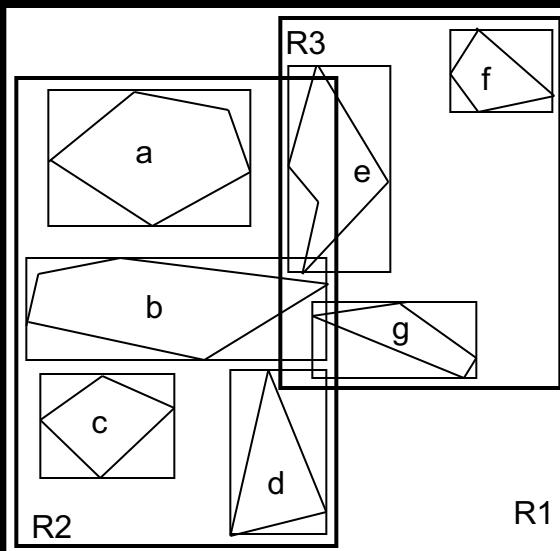
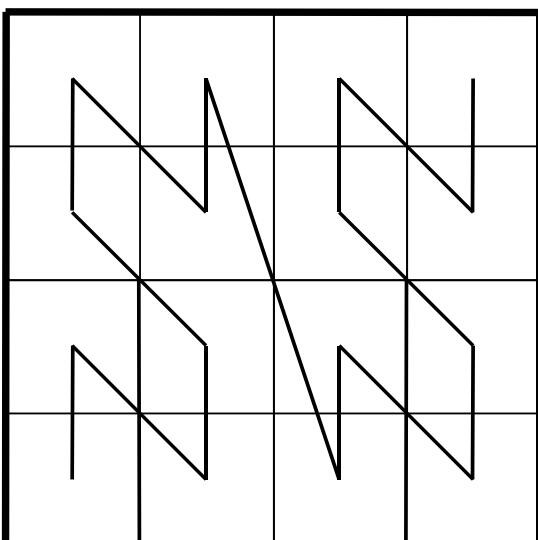
- *Static* methods rebuild the entire tree when a new node is inserted

- they typically result in a **packed** R-tree
- Grouping of rectangles can be based on spatial ordering (such as Morton numbers of their centroids)



# Morton codes for static R-tree update

- A rectangle is given an ordering based on the Morton code of its centroid (i.e. by bit-interleaving the x and y coordinates of the centre point of the rectangle).
- Use the ordering to create successive groups of M rectangles
- In a packed node method of update the lowest level non-leaf nodes are first filled **Assignment Project Exam Help**
- Nodes at the remaining levels are constructed by grouping the lower Morton numbers of the centroids of the MBRs at the <https://eduassistpro.github.io/>,  
Add WeChat **edu\_assist\_pro**



13

# R-Trees references

- "R-Trees: A Dynamic Index Structure for Spatial Searching". Proceedings ACM SIGMOD 1984.

This is the original paper on R-trees. It presents several methods for node splitting strategies for updating the tree. Given a set of rectangles to be allocated to two new nodes, following node overflow, it attempts to find the two subsets of rectangles that are most similar to each other, which form two distinct clusters. It tends to minimize overlap.

- Beckmann, N., Kriegel H.P., Schneider, R., Seeger, M. (1990) "The R\*-tree: An efficient and robust access method for rectangles". Proceedings ACM SIGMOD 1990.

This paper presents a combination of methods to reduce overlap and coverage. When inserting a new rectangle in leaf nodes, choose an entry with free space that will both minimize increase in size and minimize overlap. Also when inserting, re-insert some of the existing data, as the order in which rectangles were inserted previously affects the quality of the R-tree.

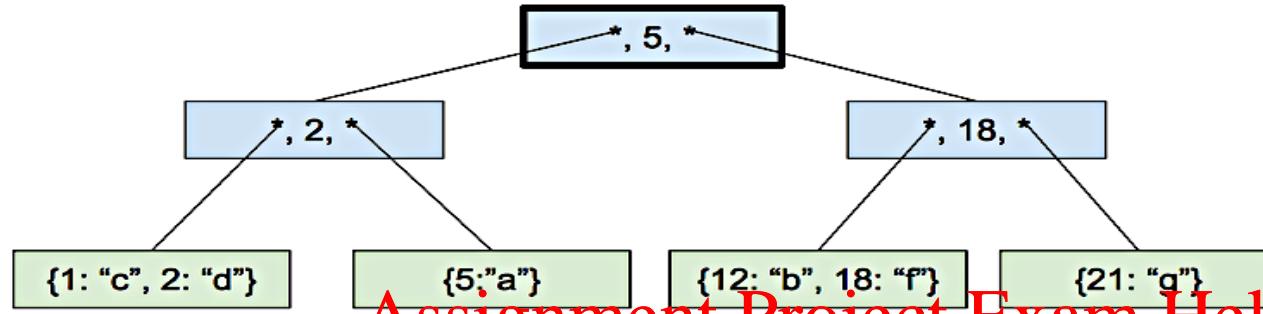
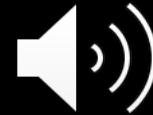
# Large Scale Databases

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat `edu_assist_pro`  
Object-directed Spatial Indexing  
with R-trees

# Revision of B-tree / B<sup>+</sup>-tree



Assignment Project Exam Help

- Balanced, i.e. all el
- Each non-leaf no <https://eduassistpro.github.io/> children = order) and a minimum of  $\lceil m/2 \rceil$
- Each non-leaf node has maximum  $m-1$  keys
- Leaf nodes can store  $m$  keys and  $m$  data values
- B<sup>+</sup>-tree non-leaf nodes store key + child pointers  
Leaf nodes store entries of <key + value>
- Root has at least 2 children



# Object-directed indexing with R-Trees

- Multi-dimensional indexing methods which find data within specified ranges of coordinates on the axis of each dimension.
- 2D R-Tree indexes rectangles = bounding boxes of spatial objects (i.e. minimum bounding rectangles **MBR**)
- Extension of B<sup>+</sup>  
**Assignment Project Exam Help**
- The R-tree creates a  
of  
MBRs of objects. <https://eduassistpro.github.io/>
- The leaf nodes contain  
the entries w  
**Add We Chat edu\_assist\_pro**  
(R7-R16) that enclose the  
original  
geometries.
- At higher levels of the hierarchy  
MBRs enclose sets of  
lower level MBRs.



# R-trees properties

Each entry in a node/page stores a rectangle and a pointer or ID

$$|\text{children}| = |\text{keys}|$$

Each non-leaf node

stores a maximum  
[Assignment](#) [Project](#) [Exam](#) [Help](#)

of  $M$  references to

child MBRs

( $M=3$  in example)

<https://eduassistpro.github.io/>

Add WeChat [edu\\_assist\\_pro](#)

Each node

(apart from root)

has minimum of

( $\lceil M/2 \rceil$ ) entries

Root has at least 2  
entries

Non-leaf node entries  
reference groups of  
child rectangles

Leaf node entries  
reference geometry  
data records



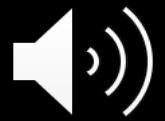
# Example from R\*Tree:

[https://en.wikipedia.org/wiki/R-tree#/media/File:R\\*-tree\\_built\\_using\\_topological\\_split.png](https://en.wikipedia.org/wiki/R-tree#/media/File:R*-tree_built_using_topological_split.png)

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro



# Searching R-trees

- To retrieve data from the R-tree : start at the root and test the search window for overlap with the MBRs of each entry in the root.
- Search descends recursively down every branch for which there is ov

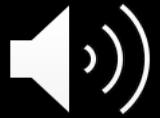
<https://eduassistpro.github.io/>

Qu

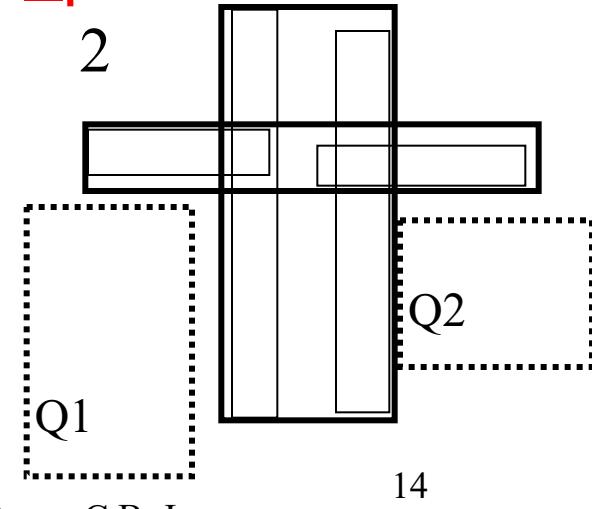
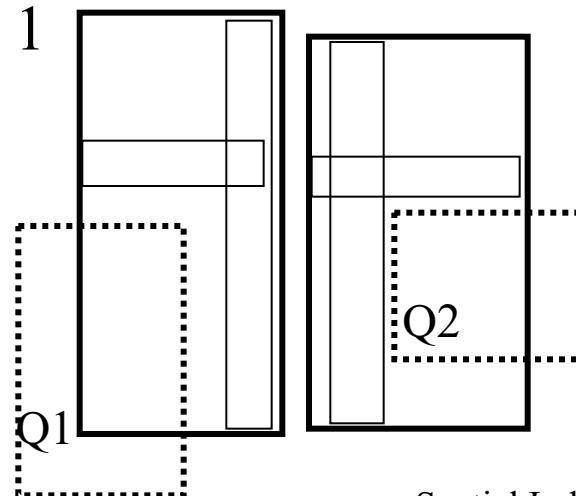
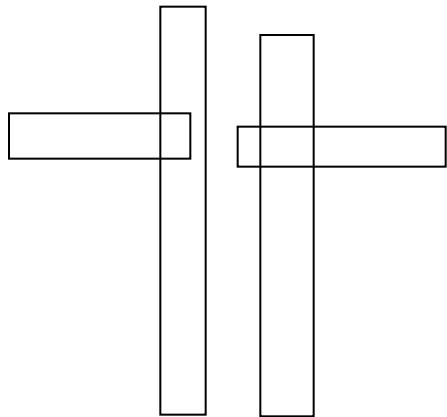
Add WeChat edu\_assist\_pro



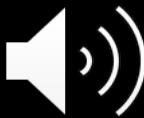
# Efficient R-tree structure



- Objective : minimise the number of fruitless searches down branches of the tree
- Figure shows two alternatives for constructing bounding rectangles around 4 MBRs, with  $M=3$   
Assignment Project Exam Help  
In option 1, query rectangle Q1 would result in fruitless searches.  
<https://eduassistpro.github.io/>  
In option 2, these searches would be avoided.  
Add WeChat edu\_assist\_pro

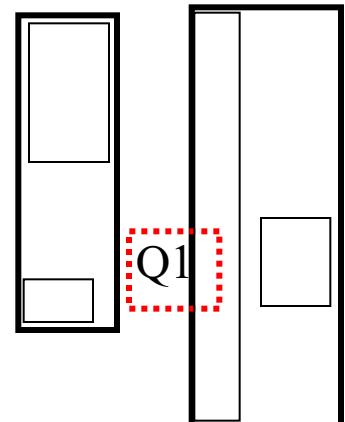
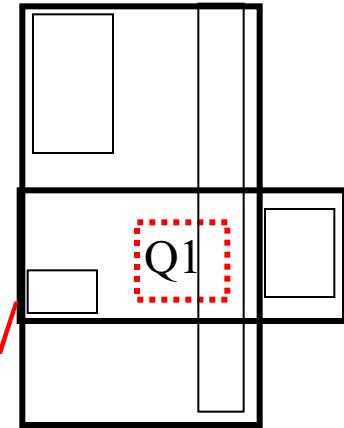


# Methods to improve efficiency

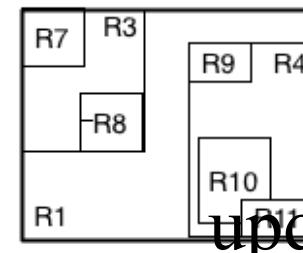


- Minimise area (*coverage*) of MBRs
  - reduces chances of query rectangle covering empty regions of an MBR
- Minimise *overlap* of MBRs
  - query windows covering the overlapping areas will always result in the tree - even though the objects.
- Minimise *size of tree*
  - fewer branches to traverse
- BUT the methods may counteract each other

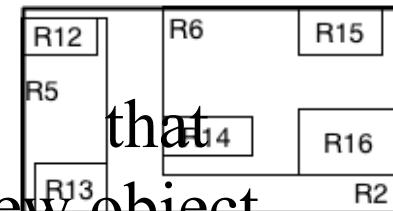
Assignment Project Exam Help  
<https://eduassistpro.github.io/>  
Add WeChat edu\_assist\_pro



# Building R-Trees



updates



that  
new object

Assignment Project Exam Help

strategy

If leaf node is f

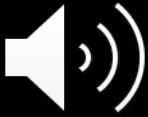
Various strategies to minimise coverage  
and overlap

Add WeChat edu\_assist\_pro

- *Static* methods rebuild the entire tree when a new node is inserted

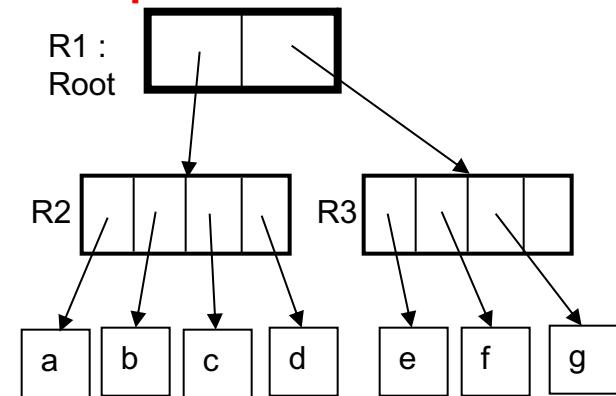
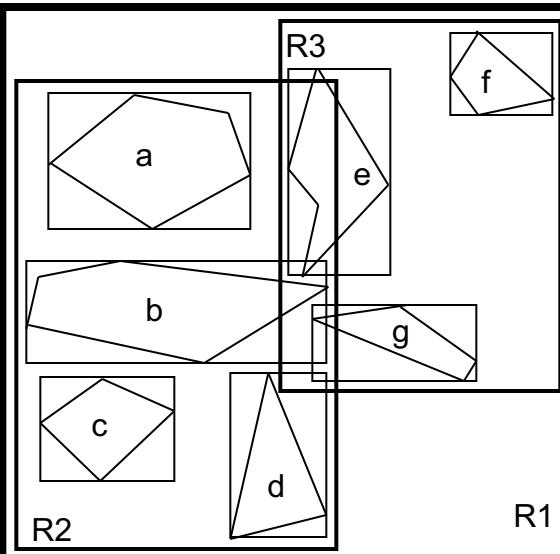
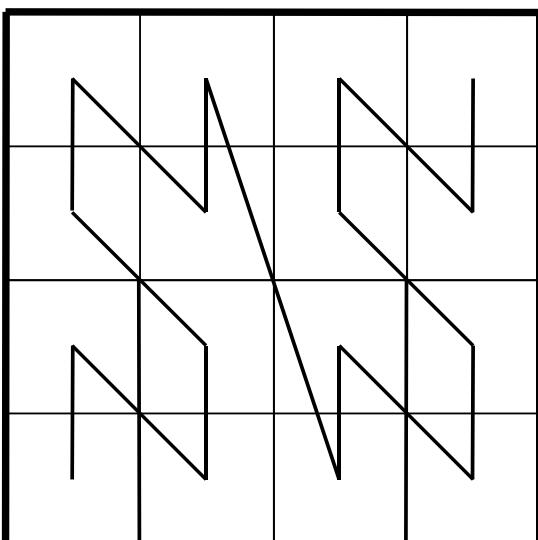
– they typically result in a **packed** R-tree

– Grouping of rectangles can be based on spatial ordering  
(such as Morton numbers of their centroids)



# Morton codes for static R-tree update

- A rectangle is given an ordering based on the Morton code of its centroid (i.e. by bit-interleaving the x and y coordinates of the centre point of the rectangle).
- Use the ordering to create successive groups of M rectangles
- In a packed node method of update the lowest level non-leaf nodes are first filled **Assignment Project Exam Help**
- Nodes at the remaining levels are constructed by grouping the lower Morton numbers of the centroids of the MBRs at the <https://eduassistpro.github.io/>,  
Add WeChat **edu\_assist\_pro**



# R-Trees references

- "R-Trees: A Dynamic Index Structure for Spatial Searching". Proceedings ACM SIGMOD 1984.

This is the original paper on R-trees. It presents several methods for node splitting strategies for updating the tree. Given a set of rectangles to be allocated to two new nodes, following node overflow, it attempts to find the two subsets of rectangles that are most similar to each other, which form two distinct clusters. It tends to minimize overlap.

- Beckmann, N., Kriegel H.P., Schneider, R., Seeger, M. (1990) "The R\*-tree: An efficient and robust access method for rectangles". Proceedings ACM SIGMOD 1990.

This paper presents a combination of methods to reduce overlap and coverage. When inserting a new rectangle in leaf nodes, choose an entry with free space that will both minimize increase in size and minimize overlap. Also when inserting, re-insert some of the existing data, as the order in which rectangles were inserted previously affects the quality of the R-tree.

Assignment Project Exam Help

Spatio-textu <https://eduassistpro.github.io/> indexing

Add WeChat edu\_assist\_pro

# Motivation: Find geographically-specific resources on the web



Form of query:

Something  $sp\_related\_to$  Somewhere  
Assignment Project Exam Help

e.g. castles *near*



Something : Concept terms

Somewhere : Placename

$sp\_related\_to$  = in, near, within Xkm, north\_of...etc.

# Matching queries to documents



Document composed of

Concept (text) terms

+ Spatial “document footprint”

Assignment Project Exam Help

- Document footprint <https://eduassistpro.github.io/> eo-locations of places mentioned ~~Add your location~~ [AddWeChat edu\\_assist\\_pro](https://eduassistpro.github.io/)
- Query requires matching :  
Query concept terms and Query footprint  
to  
Document terms and Document footprint



# Geo-Parsing : true & false references

Geo-parsing: Detecting geo-references (mostly toponyms) and resolving them (geo-coding) to coordinates

Actual place names, e.g.

New York, Raleigh,

Lower Manhattan Assignment Project Exam Help

Kansas City, Mo,

Washington

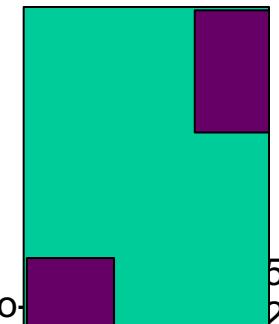
<https://eduassistpro.github.io/>

- All words highlighted in yellow are names of places *somewhere*

# Document Footprints



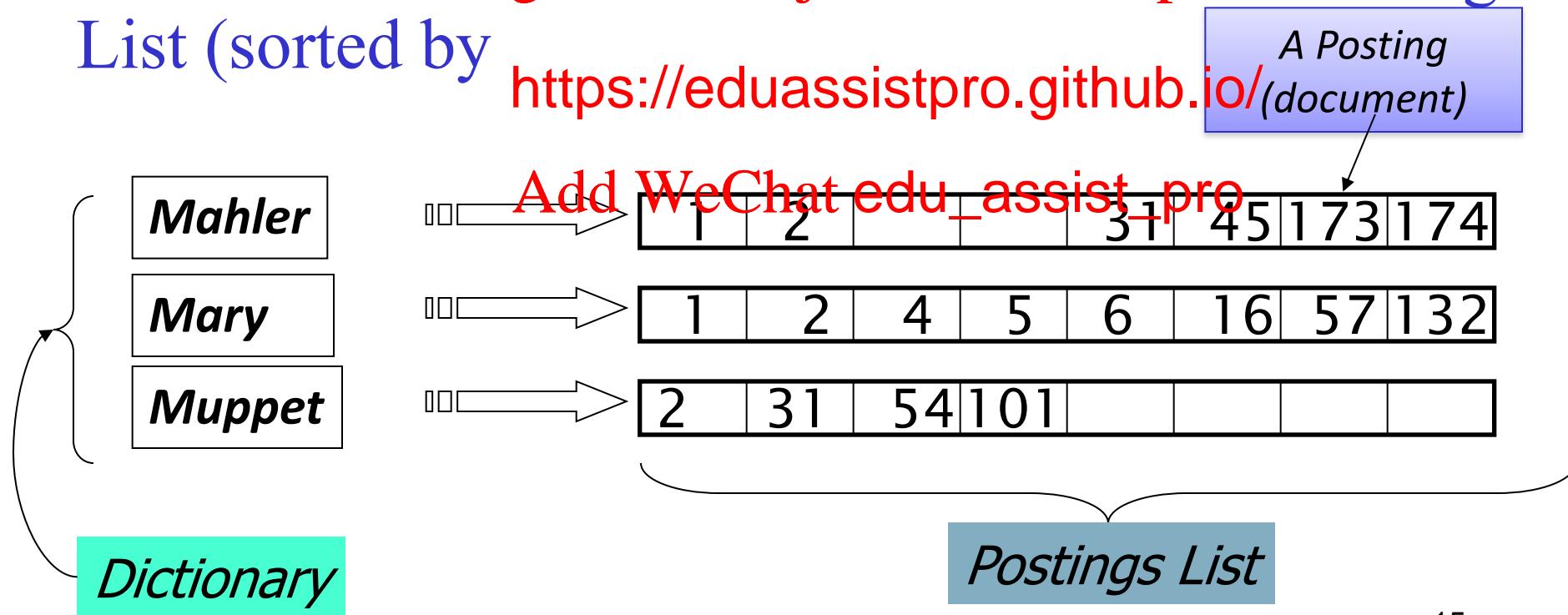
- Document footprints often approximated by one or more minimum bounding rectangles (MBR) or by points
- One MBR could cover multiple geo-references in a document - but poor results if geo-references are far apart
- Could represent each document as a set of MBRs for the main geo-references



# Text Indexing: Inverted File Index



- The set of terms is referred to as a **Dictionary**
- The Dictionary (lexicon) is sorted alphabetically
- The list of ~~Assignment Project Exam Help~~ Postings  
List (sorted by <https://eduassistpro.github.io/>)



# Inverted File Index Simple Example



Documents  $T_i = \text{DocIDs}$

T1: The cow jumped over the moon

T2: How now blue cow

T3: Now the moon shines on the blue cow

Assignment Project Exam Help

<https://eduassistpro.github.io/>

To find the document

the terms “blue” “moon” “cow” “now” “the” “over” “shines” “on” “jumped” “help” “assignment” “project” “exam” “edu\_assist\_pro” “WeChat” “Add” “GitHub”

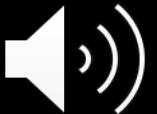
intersect the respective postings lists of

$$\{2, 3\} \cap \{1, 3\} \cap \{1, 2, 3\} = \{3\}$$

Thus document 3 is returned

Terms (Dictionary)	DocIDs Postings List
blue	{2, 3}
cow	{1, 2, 3}
how	{2}
jumped	{1}
oon	{1, 3}
w	{2, 3}
on	{3}
over	{1}
shines	{3}
the	{1, 3}

# Combining Text Indexing with Spatial Indexing



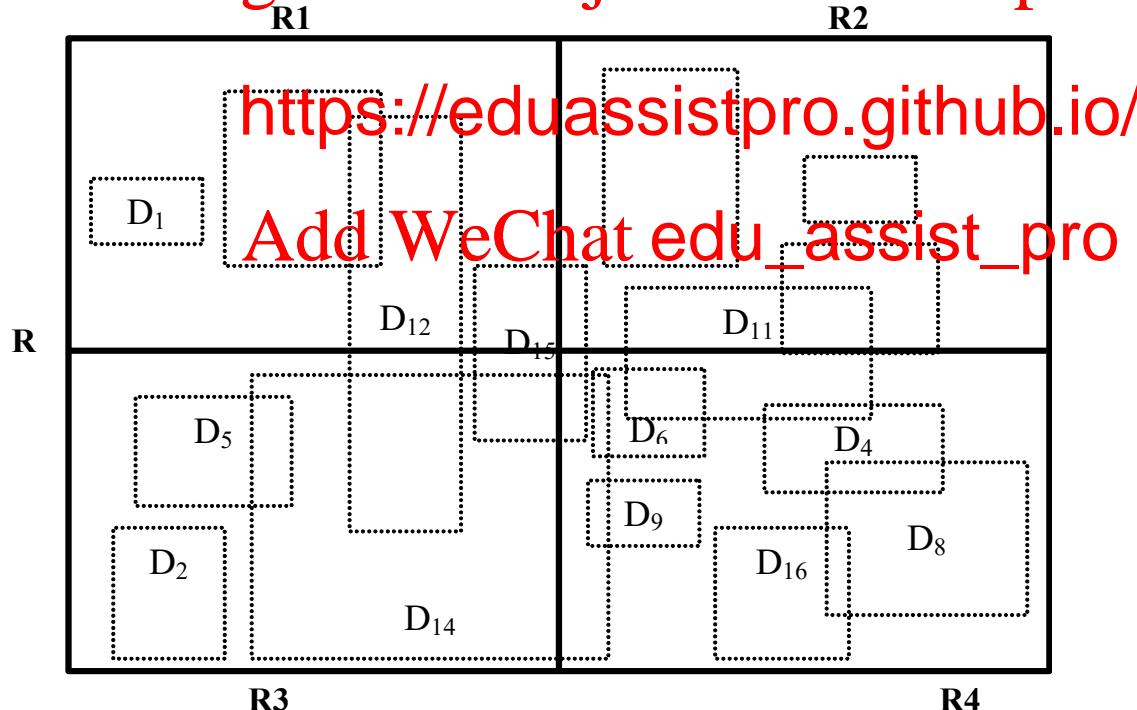
1. Separate ~~Assignment Project Exams Help~~
2. Space first / <https://eduassistpro.github.io/ex>
3. Text first / ~~Add WeChat~~ `edu_assist_pro`

# Spatial Indexing of Web Documents with a Regular Grid



- Derive document footprint(s)  $D_i$  for each document
- For each cell, record documents whose footprints intersect it

Assignment Project Exam Help



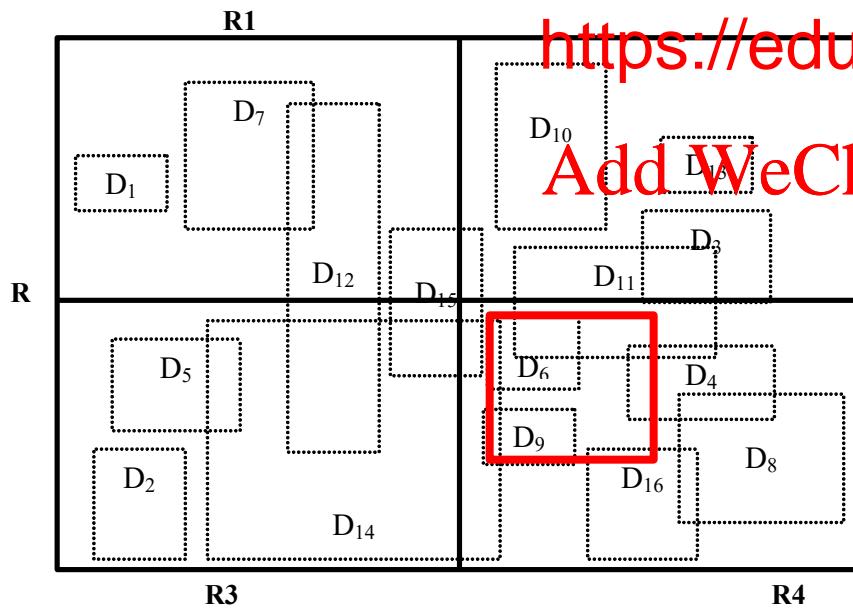
# Separate text and spatial indexes



query <terms + query footprint>

Term1	D1, D2, D6, D23, ...
Term2	D2, D6, D9, D11, ...
Term3	D27, D85, Assignment Project Exam Help

1. Find documents containing query term(s)



<https://eduassistpro.github.io/>

Find documents intersecting query footprint

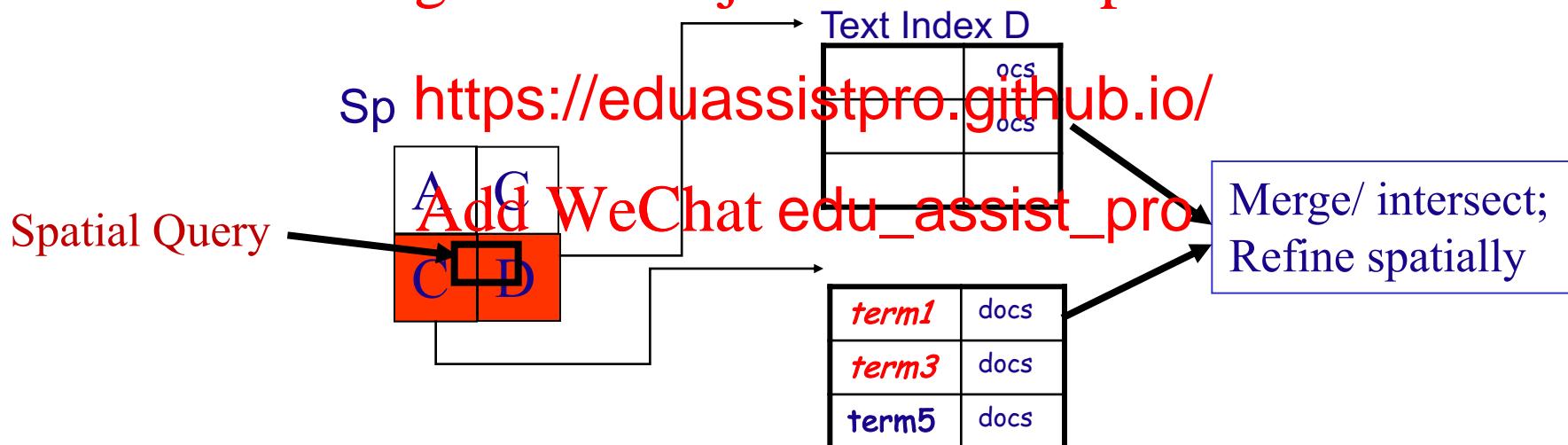
3. Find intersection of the two sets (here D6, D9, D11)

# Integrated approaches: Space-first methods



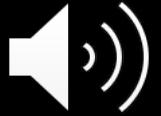
- Spatial cells of index point to inverted file (text) indexes of documents whose footprints intersect the cell
- Below illustrated with a regular grid
- [See next slide for an R-tree (with more detail)]

Assignment Project Exam Help



- Find cells that intersect query footprint / window;
- For each such cell, search associated inverted index to find documents that contain all query terms;
- Test (filter) resulting document footprints against query footprint.

# Space First with R-tree



Each entry in a leaf node points to an inverted file index of documents whose footprints are inside the corresponding rectangle

**Assignment Project Exam Help**

<https://eduassistpro.github.io/>

Inverted  
(text) indexes

**Add WeChat edu\_assist\_pro**

Documents  
with  
footprints  
inside parent  
rectangle

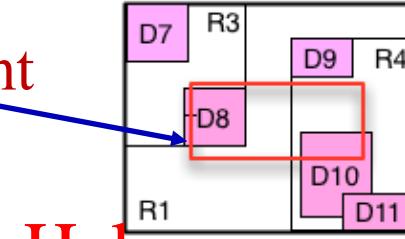


# Space First with R-tree (cont.)



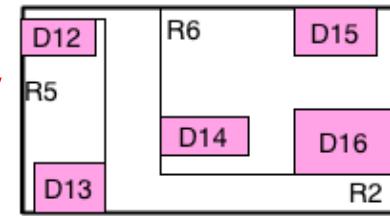
Query text: t1, t3

Query footprint



Assignment Project Exam Help

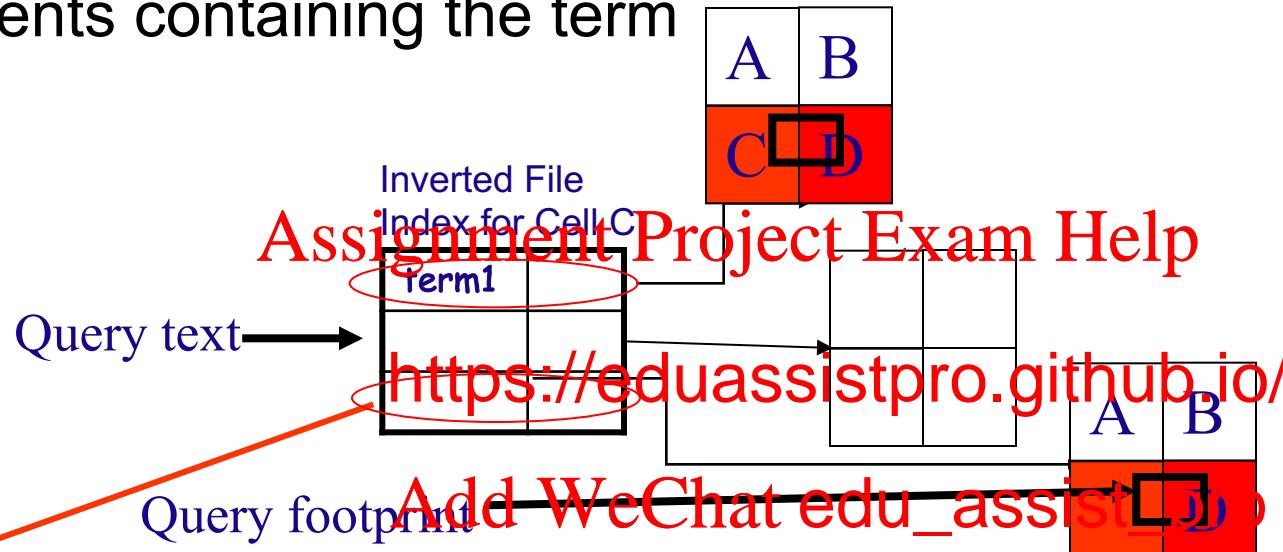
- Find leaf cells that intersect query footprint: R3, R4
- Find documents that contain all query term https://eduassistpro.github.io/ D8, D9, D10, D11
- Test document footprints Add WeChat edu\_assist\_pro against query footprint
- -> D8, D10





# Text first spatio-textual indexing

For each term in inverted file index: store spatial index of documents containing the term



Index Entry (for regular grid/quadtree): term3 : cellC(D<sub>1</sub>, D<sub>7</sub>); cellD(D<sub>3</sub>, D<sub>11</sub>, D<sub>13</sub>)...

Retrieval:

- For each query term, access associated spatial index to retrieve ids of documents in spatial cells intersecting the query footprint
- Intersect results to find documents that contain all query terms
- Filter those documents against the query footprint

# References

Fernando Melo and Bruno Martins. Automated geocoding of textual documents: A survey of current approaches. *Transactions in GIS*, 21(1):3–38, 2017.

Dingming Wu, Gao Cong, and Christian S. Jensen. A framework for efficient spatial web object retrieval. *The VLDB Journal*, 21(6):797–822, December 2012.

Yinghua Zhou, Xing Xie, <https://eduassistpro.github.io/> and Wei-Ying Ma. Hybrid index structures for geographical search on the web. In CIKM, pages 155–162, 2005.

Add WeChat edu\_assist\_pro

S. Vaid, C.B. Jones, H. Joho, and M. Sanderson. Spatio-textual indexing for geographical search on the web. In SSTD, pages 218–235, 2005.

Chengyuan Zhang, Ying Zhang, Wenjie Zhang, and Xuemin Lin. Inverted linear quadtree: Efficient top k spatial keyword search. In ICDE, pages 901–912. IEEE Computer Society, 2013.

# Large Scale Databases

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro  
Spatial D S

Revision

# Class Test structure



4 Questions (with sub-parts)

- **ANSWER ALL QUESTIONS**

Assignment Project Exam Help

10 marks per <https://eduassistpro.github.io/>

Add WeChat edu\_assist\_pro

2 questions on Alia Abdelmety's material

2 questions on Chris Jones's material

# Spatial Databases Questions in the Class Test

Focus for *Spatial Databases* questions is

on problem solving  
Assignment Project Exam Help

This requires of concepts  
but look especially examples of  
questions in exercises that require  
solving a problem of some sort

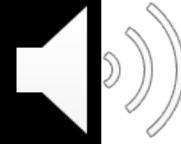
# databases

Subject of questions can relate to:



- Construction of spatial indexes given a set of objects  
Requires understand principles of different spatial indexing methods
  - ~~Assignment Project Exam Help~~ :  
how t coordinates;  
how they r <https://eduassistpro.github.io/>ow they represent  
Add WeChat `edu_assist_pro`
- 9-intersection model for representing topological relations with area objects and linear objects
  - SQL queries given a set of data tables generic spatial SQL as illustrated in lecture slides is fine,  
i.e. not necessary to memorise the exact names

# Spatial Database Topics Covered



- Geographical information characteristics
  - Spatial data models, Locations,
  - Spatial relations between spatial objects
- Object Relational data <https://eduassistpro.github.io/>
- Spatial SQL [Add WeChat edu\\_assist\\_pro](#)
- Spatial indexing
  - Space vs object directed
  - Regular grid, quadtrees & location codes, R-tree
  - Spatio-textual indexing

# Geographical information characteristics



- Location

- Qualitative: names, postcodes, absolute vs relative
- Coordinates: lat/long vs map grids. Datums

- Geometric objects

Assignment Project Exam Help

- Points, lines, areas.... OGC Simple Feature geometry model
- Topological pri <https://eduassistpro.github.io/>

- Spatial relations between objects

Add WeChat edu\_assist\_pro

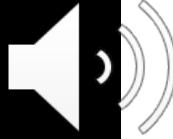
- Topological :

- Inside, contains, overlap, touch (meets), equal, disjoint, covers, covered by
- 9-intersection model of each topological relation

- Proximal (distance) : qualitative vs quantitative

- Direction / orientation : qualitative vs quantitative

# Spatial data models



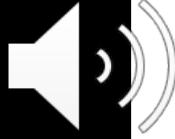
- Raster / location-based
  - record what is present within regular grid cells
- Vector / object-based
  - Geometric primitives: point, lineString, https://eduassistpro.github.io/Assignment Project Exam Help  
linearRing), s)...
  - Topological primitives: node (start node / end node), ring (arcs), face (rings), polyhedron (3D)
  - Topologically structured maps : polygon map, network map, triangulated irregular network

# Object-Relational technologies



- Complex data types to represent geometry
  - SQL Create : geometry objects from basic data types (numbers, arrays etc)
- Spatial query language <https://eduassistpro.github.io/>
  - Spatial properties : area, distance..
  - Spatial relations : topology
  - Geometric intersection (returns geometry)
- Spatial indexing

# Spatial Indexing



- Space-directed vs Object-directed
- Spatio-textual

Assignment Project Exam Help

## Space-direc

<https://eduassistpro.github.io/>

- Regular grid
  - Location codes based on left corner of cells : concatenation of x/y (vs bit-interleaving)
    - Index location codes with B-tree
  - Use for processing queries and updates
  - Limitations of regular grid

# Space-directed indexing cont.



- **Quadtrees**

- Recursive subdivision of space into quadrants
- Create location codes for cells using Morton codes / Z-n <https://eduassistpro.github.io/>
  - Bit-interleaving of x, y, z
- Issue of unique codes (base 4 vs base 5)
- Query processing with ‘block’ location codes

## Space-filling curves

- desirable properties of location code numbering systems

# Object-directed indexing with R-trees



R-trees extend/modify some aspects of B-trees

Surround geometry objects with minimum bounding rectangles (MBRs) and Create hierarchy of MBRs

- Entry in tree node = MBR + pointer to another node  
or (from leaf n <https://eduassistpro.github.io/>)
- Search by comparing query with the MBRs in the node entries
- Desirable properties of R-Tree
- Dynamic vs static update methods (static → packed)

# Spatio-textual / Spatial keyword indexing



## Spatial indexing of documents

- Geoparsing → document footprints

Assignment Project Exam Help

## Inverted file documents

- dictionary + <https://eduassistpro.github.io/>

## Combining spatial and file indexing

- separate indexes and merge results
- space first
- text first