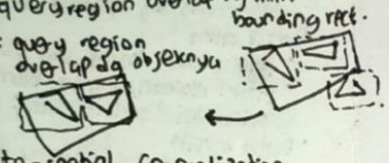
- Fitur kontinu
  data curah hujan & pusat kota
- Spatial Relationship
1. Topological Relations — disjoint, contains
2. Metric relation
3. Direction relation
- Spatial Data Type
1. Field-based model
   → raster data
   → framework: partition of space
2. object-based model
   → point, line, polygon, object, attrb.
- Spatial Query Lang.
  → data types : point, line, polygon
  → operations : overlap, distance
- Query Processing
  → filter : query region overlap dg min. bounding rect.
  → refine : query region overlap dg objeknya



- Spatial-to-spatial Generalization
→ merging of a set of geographic areas by spatial operations
→ dissolve, merge, clip, intersect, union
- Spatial co-location
→ association rule → explore spatial auto-corr.
- Spatial autocorr.
"everything is related to everything, but nearby things are more related than distant things."
- Spatial Classification Method
1. Spatial auto-regression (SAR)
2. Markov random field (MRF)
- Spatial Outlier Detection
1. Graphical test — variogram clouds, moran scatterplots
2. Quantitative test — scatterplots, spatial statistics, lbh akurat

Sequential Pattern Mining
- diberikan himpunan sekuens & support threshold, temukan himpunan lengkap dr frequent subseq.
- sebuah basis data sekuen terdiri dr elemen xx / kejadian xx terurut.

SID	Seq
10	<a(abc)(ac)d(cf)>
20	<(ad)c(bc)(ae)>
30	<(ef)(ab)(df)cb>
40	<eg(af)cbc>

- Aplikasi
  1. urutan pembelanjaan konsumen
  2. perawatan medis, bencana alam
  3. pola panggilan telepon, weblog click streams
  4. sekuens DNA & struktur gen

- Metode
1. Basis Apriori
   - GSP (Generalized Sequential Pattern)
     - initial candidate L=1
     - scan db hitung support
     - L++ candidate L=2
       ↳ 2 kejadian & 1 kejadian
     - scan DB hitung support
     - sampai sudah < minsup semua / L = psg maks. DB
   - SPADE (Sequential Pattern Discovery using Equivalence class)
     - vertical data layout

SID	EID	Items
⋮	⋮	⋮

     - inisialisasi item

SID	EID
⋮	⋮

     - dst

items / ab

SID	EID(a)	EID(b)
⋮	⋮	⋮

2. Pattern-Growth
   - FreeSpan
   - PrefixSpan
     - tentukan semua pola sekuen dg psg 1
     <a> <b> <c> ...
     4 3 ...
     - bagi ruang pencarian

<a> suffix ....
10 ...
30 ...
50 ...
90 ...

<f> suffix
⋮ ⋮

- length = 2 → cek suffix di <a>

efisiensi
- tdk ada kandidat yg perlu dibangun kan
- projected DB berukuran kecil

- Proyeksi bi-level
  membagi ruang pencarian berdasarkan pola sekuens dg psg 2, lg membentuk projected DB & melakukan recursive mining pd bilevel DB.
- kendala SPM
  kendala item, panjang, super-pattern, agregat, ekspresi regular, durasi, gap
- Pseudo-projection
  menghindari penyalinan postfix ser fisik
  → menggunakan pointers
- Periodicity Analysis
1. Full p. → setiap titik pd waktunya berkontribusi utk periodicity
2. Partial p. → notasi umum yg lbh general, lg. beberapa segmen yg berpartisipasi
3. Cyclic association rule
   ↳ Metode : 1 → FFT, analisis statistik
   2, 3 → apriori-like

Text & web Mining
- text data : kata yg mendeskripsikan objek
- WWW : teks tdk terstruktur & data multimedia
- Teks mining : proses mengekstrak informasi yg belum diketahui dr repositori dokumen tekstual
- Damin vs Teksmin
- numerik & kategorik — teks
- RDBMS — free from text
- prediksi — temu kembali informasi
- machine learning — indexing, ontologis
- Karakteristik Data Teks
  berukuran besar, dimensi tinggi, beberapa mode input, ketergantungan, ambiguitas, noisy data, tdk terstruktur.
- Analisis Teks semantik / sintaktik
  ↳ Part of speech tagging → tentukan pos setiap kata
  2. Word sense disambiguation
     → berbasis konteks / proximity
  3. Parsing → parse tree utk setiap kalimat
- Rembangkitan Fitur : Bag of words
  - stemming : root kata, flying → fly
  - Stop words : kata umum → a, an
- Seleksi fitur
  - mengurangi dimensi
  - menghapus yg tdk relevan
- Web Mining : teknik damin utk mencari & mengekstrak scr otomatis informasi dr dokumen / layanan web.
- Why web?
  - kaya akan informasi tekstual
  - banyaknya data pd pola akses pengguna
  - possible to retrieve "previously unknown" information
- Fitur unik pd web
  - informasi hyperlink
  - informasi akses & penggunaan
  - web sgt dinamis
- Pendekatan Web Content Mining
  1. Information Retrieval → berdasarkan yg diminta
  2. Basis data → kueri kompleks & keyword
- Web Structure Mining
  - hlm web yg relevan & kualitas tinggi
  - hyperlink
- Web Usage / Log Mining
  - analisis statistik dr pengguna → psg path, view hm, byknya page views
  - analisis statistik dr site → hlm yg paling sering dilihat, invalid URL plg umum
  - aplikasi
    → kostumer potensial
    → perform a sistem web server
    → deteksi fraud / intrusion