

ATURAN ASOSIASI SPASIAL PADA BASIS DATA SPASIAL

Magdalena Karismariyanti¹, Dhinta Darmantoro², Dana Sulisty Kusumo³

Jurusan Teknik Informatika, Sekolah Tinggi Teknologi Telkom
Jln. Telekomunikasi, Dayeuh Kolot, Bandung, 40257, Indonesia
e-mail: ¹magda_ellen@yahoo.com, ²ddd@stttelkom.ac.id, ³dana@stttelkom.ac.id

ABSTRAKSI

Data mining sebagai pengolah data telah diketahui bahwa data mining dapat menemukan suatu pola pada basis data yang besar. Spatial data mining merupakan proses menemukan pola yang menarik dan potensial untuk dicari informasinya berdasarkan pola dari spatial dataset.

Spatial data mining memerlukan neighbour object untuk mencari keterhubungan obyek itu sendiri dengan obyek-obyek lain yang mempengaruhi obyek tersebut. Setelah neighbour object diidentifikasi maka association rule dapat dicari.

Pencarian neighbour object dilakukan dengan melakukan buffer dari obyek pencarian, sehingga ditemukan obyek disekitar obyek tersebut. Analisis spatial association rule menggunakan metode Apriori.

Parameter ketetanggaan/kedekatan dari obyek adalah berdasarkan radius atau jumlah neighbour object. Semakin banyak neighbour object, semakin banyak frequent item yang dibangkitkan, dan semakin banyak pula rule-nya.

Kata kunci: neighbour object, spatial association rule, spatial data mining.

1. PENDAHULUAN

1.1 Latar Belakang

Data mining adalah penemuan *trend* yang menarik atau pola pada *dataset* yang besar, dengan tujuan untuk menuntun pada keputusan tentang kegiatan kedepan[10]. Salah satu pembahasan yang ada pada data mining adalah spatial data mining.

Spatial data mining merujuk kepada *extraction of knowledge, spatial relationship*, atau pola yang menarik lainnya yang tidak secara eksplisit disimpan dalam basis data spasial[3]. Basis data spasial menyimpan *space-related data*, misalnya peta, *space-related data* yang dimaksud adalah data topologi atau data jarak.

Basis data spasial terdiri dari obyek dengan *spatial* dan *non-spatial* data. Proses penemuan untuk *spatial data* lebih kompleks dari data relasional. Hal ini terkait dengan efisiensi algoritma dan kompleksitas pola yang ditemukan dalam basis data spasial. *Spatial data mining* harus mempertimbangkan *neighbour object*, hal ini diperlukan karena atribut *neighbour* dari beberapa obyek mempengaruhi obyek itu sendiri[1].

Sebagaimana dalam data mining terdapat *association rule*, pada *spatial data mining* juga terdapat *spatial association rule* yang mendeskripsikan asosiasi antara obyek berdasarkan *spatial neighbourhood relation*. Permasalahan yang muncul adalah bagaimana mencari keterhubungan data spasial dari *neighbour object* dan bagaimana *association rules*-nya dari hasil keterhubungan tersebut. Pada penelitian ini algoritma *association rule* yang digunakan adalah Apriori.

1.2 Batasan Masalah

Batasan masalah yang dalam penelitian ini adalah:

- Basis data spasial yang dapat dijalankan di desktop.
- Basis data spasial yang digunakan adalah peta kota Bandung.
- Pencarian obyek berdasarkan tipe obyek *point* (titik).
- Association rule* yang akan digunakan hanya sebatas untuk menemukan *rule* dari hasil pencarian *neighbour object*.
- Data mining association rule* yang digunakan adalah apriori.

2. LANDASAN TEORI

2.1 Basis Data

Basis Data adalah sekumpulan data, yang mendeskripsikan kegiatan dari satu atau lebih organisasi yang saling berhubungan.

Basis data ada beberapa jenis, yaitu [3]:

- Basis data relasional (*Relational Database*)
- Basis data transaksi (*Transactional Database*)
- Sistem Basis data Lanjut (*Advanced Database System*), contohnya basis data spasial.

2.2 Mining Spatial Databases

Basis data spasial menyimpan *space-related data* yang sangat besar, seperti map, data *medical image* dan VLSI *chiplayout* data. Basis data spasial memiliki kerumitan dibandingkan dengan basis data relasional. Basis data spasial memiliki informasi topologi/jarak, biasanya diatur menggunakan *multidimensional spatial indexing structure* yang diakses dengan metode akses data spasial dan sering membutuhkan *spatial reasoning, geometric computation*, dan *spatial knowledge representation techniques*.

Ada beberapa tantangan yang berhubungan dengan pembangunan dan utilitas pada data spasial. Tantangan pertama adalah integrasi dari spatial data dari sumber yang heterogen dan sistem yang heterogen. Format data tidak hanya *structure-specifics* (misalnya, *raster-based spatial data* dengan *vector-based spatial data*, *object-oriented* dengan model-rasional), tetapi *vendor-specific* (misalnya, ESRI, MapInfo, Intergraph, dan lain-lain). Tantangan yang kedua adalah analisis yang cepat dan fleksibel dari data spasial.

2.3 Association Rule

Association rule mining menemukan asosiasi atau korelasi diantara kumpulan data item yang banyak. Dengan banyaknya data yang terus-menerus dikumpulkan dan disimpan, banyak industri yang menjadi tertarik untuk *mining association rule* dari data yang dimiliki. Penemuan dari relasi asosiasi yang menarik dari data transaksi yang sangat besar dapat membantu proses pengambilan keputusan, seperti *catalog design*, *cross marketing* dan *loss-leader analysis*.

2.4 Konsep dasar

Ada sebuah transaksi D memiliki *itemset* $J = \{i_1, i_2, \dots, i_m\}$. Masing-masing transaksi memiliki identitas yaitu TID. Pada setiap TID yang dimiliki D, memiliki *itemset* T dimana $T \subseteq J$. A dikatakan sebuah transaksi D jika dan hanya jika $A \subseteq T$. *Association rule* adalah bentuk implikasi $A \Rightarrow B$ dimana $A \subset J$, $B \subset J$, dan $A \cap B = \emptyset$. Sebuah *rule* $A \Rightarrow B$ memiliki **support** s, dimana s adalah persentase transaksi di D yang terdiri AuB dengan probabilitas $P(AuB)$. Rule $A \Rightarrow B$ memiliki **confidence** c, dimana c adalah persentase dari transaksi pada D yang meliputi A dan juga meliputi B [3].

$$\text{Support } (A \Rightarrow B) = P(AuB)$$

$$\text{Confidence } (A \Rightarrow B) = P(B|A)$$

Rule yang memenuhi kedua *minimum support threshold* (*min_sup*) dan *minimum confidence threshold* (*min_conf*) disebut **strong**. Dengan ketetapan, penulisan nilai *support* dan *confidence* antara 0% dan 100%.

Set of item dinamakan *itemset*. Sebuah *itemset* terdiri dari k *items* adalah *k-itemset*. Set $\{\text{computer, financial_management_software}\}$ adalah 2-itemset. Itemset memenuhi *minimum support* jika frekuensi kemunculan *itemset* lebih besar sama dengan perkalian *min_sup* dan total jumlah transaksi di $D \geq (\text{min_sup} * \text{total transaksi})$. Jumlah transaksi dibutuhkan *itemset* untuk memenuhi *minimum support* maka dari itu merujuk ke **minimum support count**. Jika itemset memenuhi *minimum support* maka disebut **frequent itemset**. Sekumpulan *frequent k-itemset* dilambangkan dengan L_k .

2.5 The Apriori Algorithm

Apriori adalah *influential Algorithm* untuk *me-mining frequent itemset* untuk boolean association rule. Nama dari algoritma ini berdasarkan fakta bahwa algoritma menggunakan *prior knowledge* dari properti *frequent itemset*. Apriori menggunakan pendekatan *level-wise search*, dimana *k-itemset* yang digunakan untuk mengeksplor $(k+1)$ -*itemset*. Pertama, set dari *frequent 1-itemset* ditemukan. Dinotasikan dengan L_1 . L_1 digunakan untuk menemukan L_2 , set dari *frequent 2-itemset*, yang digunakan untuk menemukan L_2 dan seterusnya sampai tidak ada lagi *k-itemset* ditemukan. Untuk menemukan masing-masing L_k memerlukan penelusuran ke seluruh basis data.

“Bagaimana property apriori digunakan dalam algoritma?” ada dua langkah yang meliputi join dan prune.

- Join Step: untuk menemukan L_k , sebuah set candidate *k-itemset* dibuat dengan join L_{k-1} dengan L_{k-1} . Set candidate dinotasikan dengan C_k . Misal l_1 dan l_2 menjadi itemset L_{k-1} . Notasi $l_i [j]$ merujuk item ke-j di l_i (misal, $l_1[k-2]$ menunjuk pada ke-2 terakhir dari l_1).
- Prune Step: C_k adalah super set dari L_k , anggotanya boleh atau tidak boleh frequent, tapi semua frequent *k-itemset* dimasukkan kedalam C_k . Scan basis data menunjukkan jumlah masing-masing kandidat di C_k menghasilkan determinasi L_k (misalnya semua kandidat berjumlah tidak kurang dari jumlah minimum support adalah frequent, maka dari itu menjadi miliki L_k) [3].

2.6 Pembangkitan Association Rule dari Frequent Itemset

Setelah *frequent itemset* dari transaksi pada database D ditemukan, lanjut ke membuat *strong association rule* (*strong association rule* memenuhi *minimum support* dan *minimum confidence*), dimana kondisi probabilitas diekspresikan pada bagian itemset support count [3]:

$$\text{Confidence } (A \Rightarrow B) = P(B|A) = \frac{\text{Support_count}(AuB)}{\text{Support_count}(A)}$$

Keterangan:

- $\text{support_count}(A \cup B)$ adalah jumlah transaksi meliputi itemset $A \cup B$
- $\text{support_count}(A)$ adalah jumlah transaksi yang meliputi itemset A

Berdasarkan persamaan, *asoiation rule* dibentuk sebagai berikut:

Untuk setiap *frequent itemset* l, bangkitkan subset dari l. Misalnya $l = \{i_1, i_2, i_5\}$

Rule yang dibangkitkan:

$$i_1 \wedge i_2 \Rightarrow i_5$$

$$i_1 \wedge i_5 \Rightarrow i_2$$

$$\begin{aligned}
 i_2 \wedge i_5 &\Rightarrow i_1 \\
 i_1 &\Rightarrow i_2 \wedge i_5 \\
 i_2 &\Rightarrow i_1 \wedge i_5 \\
 i_5 &\Rightarrow i_1 \wedge i_2
 \end{aligned}$$

Karena *rules* digenerate dari *frequent itemset*, masing-masing secara otomatis memenuhi minimum support.

2.7 Spatial Association Analysis

“Bagaimana dengan me-mining spatial association rule?” sama dengan me-mining association rule di basis data transaksional dan relasional, spatial association rule dapat di-mining pada basis data spasial. **Spatial Association Rule** berbentuk $A \Rightarrow B [s\%, c\%]$.

Keterangan:

- A dan B adalah sekumpulan predikat spasial atau non-spasial.
- s% adalah support of the rule.
- c% adalah confidence of the rule.

Contoh:

$$\text{is_a}(X, \text{school}) \wedge \text{close_to}(X, \text{sport_center}) \Rightarrow \text{close_to}(X, \text{park}) [0.5\%, 80\%]$$

Artinya adalah:

80% school close to (yang dekat) sport center, close to (dekat) parks.

0.5% dari data memiliki rule seperti di atas.

Spatial predicate dapat membangun spatial association rule.

Tabel 1. Beberapa operation type dan operation name

Operator Type	Operation Name
Distance information	Close_to Far_away
Topological relation	Intersects Overlap Disjoint
Spatial orientation	Left_of West_of

Untuk me-mining *spatial association* tergantung dari *spatial predicate close_to*, pertama dapat dikumpulkan kandidat yang menghasilkan *minimum support threshold* dengan:

- menerapkan algoritma evaluasi spasial kasar, misalnya, menggunakan minimum struktur *bounding rectangle* (yang terhubung hanya pada dua *spatial point* daripada *polygon* yang kompleks)
- evaluasi *relaxed spatial predicate*, *g_close_to*, yang merupakan generalisasi *close_to*

2.8 Spatial Neighbourhood relationship dan Operasinya

Ada tiga tipe dasar *spatial relation*: topologi, *distance* (jarak) dan *direction relations* yang dikombinasikan dengan operator logik untuk menunjukkan *neighbourhood relations*. *Spatial objek* seperti *points* (titik), *lines* (garis), *polygon* atau *polyhedron* direpresentasikan oleh beberapa titik. Contoh, poligon direpresentasikan dengan *edge (vector representation)* atau dengan *point (raster representation)*.

Topological relation didasarkan pada *boundary*, *interior* dan komplemen dari dua objek yang berelasi dengan variannya menggunakan transformasi kontinu, one-one, onto dan inversnya adalah kontinu. Relasi adalah A *disjoint* B, A *meets* B, A *overlaps* B, A *equal* B, A *covers* B, A *covered-by* B, A *contain* B, A *inside* B.

3. ANALISIS DAN PERANCANGAN

3.1 Analisis Sistem

3.1.1 Kebutuhan Fungsional

Sistem yang akan dibangun bertujuan untuk menganalisis *neighbour object* dari data peta. Setelah *neighbour object* ditemukan akan dibangkitkan *association rule*nya. Untuk itu, sistem harus memiliki kemampuan untuk:

- menentukan satu obyek yang akan dicari relasi dengan obyek lain
- menentukan luas area yang akan dianalisis berdasarkan obyek tertentu
- membangkitkan keterhubungan antarobyek (*neighbourhood object relationship*)
- menentukan minimum support dan minimum confidence
- membangkitkan aturan asosiasi berdasarkan *neighbourhood object relationship*

3.1.2 Analisis Proses

Proses-proses sistem ini adalah sebagai berikut:

Proses 1: **Pengaturan**, proses untuk menentukan tipe obyek yang berguna untuk membantu pencarian *neighbour object* dan minimum/maksimum radius berdasarkan skala pada peta.

Proses 2: **Pembangkitan Neighbour**, proses untuk mencari obyek-obyek disekitar obyek yang dimasukkan oleh pengguna. Object *neighbour* yang akan dicari berada di area dengan radius yang telah dimasukkan oleh pengguna. Hasil dari proses ini disimpan dalam data *neighbour*.

Proses 3: **Pembangkitan Association Rule**, proses pengolahan data Object *neighbour relationship* menjadi aturan asosiasi berdasarkan minimum support dan minimum confidence yang dimasukkan

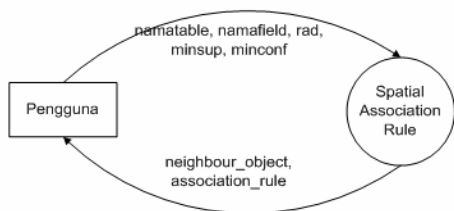
oleh pengguna. Hasil dari proses ini disimpan dalam data rule.

Proses 4: **Pelaporan**, proses penampilan hasil implementasi berupa data neighbour dan data rule.

3.2 Perancangan Sistem

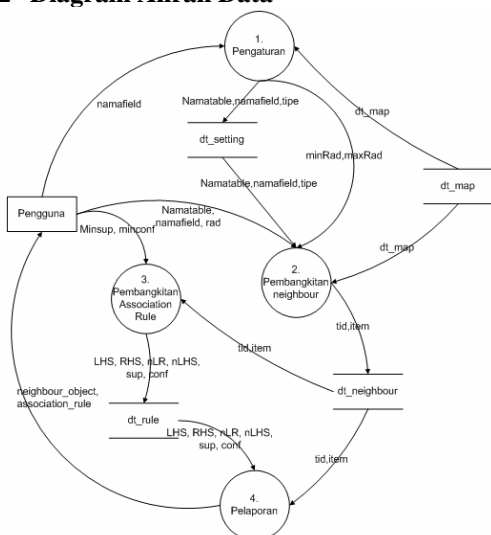
Model Perancangan terstruktur menggunakan Diagram Aliran Data (DAD) yang menggambarkan aliran data dan input dan output dari sebuah proses.

3.2.1 Diagram Konteks



Gambar 1. Diagram Konteks

3.2.2 Diagram Aliran Data



Gambar 2. DAD Level 1

3.2.3 Struktur Table

Basis data yang digunakan ada dua macam yaitu basis data relasional dan basis data spasial. Perancangan tabel mendeskripsikan isi table dan atribut yang dimiliki.

a. Basis Data Relasional

- Tabel dt_frequent
Table dt_frequent adalah table berisi Lk. Itemset dimana subsetnya adalah frequent pada Lk-1 dan memiliki SupCount >= minsup
- Tabel dt_neighbour
Table dt_neighbour berisi data obyek dengan obyek tetangganya berdasarkan radius yang diberikan.
- Tabel dt_rule

Table dt_rule berisi association rule yang dihasilkan dari dt_frequent.

Skema dt_rule = { LHS+RHS+nLR+nLHS+sup+conf }

LHS menyatakan dan RHS menyatakan itemset sebelah kanan (consequent).

- Tabel dt_setting
Table dt_setting berisi data yang diambil dari dt_map dan inputan user. Dt_setting digunakan untuk membantu proses penentuan neighbour obyek.

b. Basis Data Spasial

Dt_map adalah basis data spasial berupa peta. Pada basis data peta ini memiliki beberapa table yang dideskripsikan seperti di berikut:

- Tabel hotel
- Tabel jalan
- Tabel kecamatan
- Tabel objek_wisata
- Tabel perumahan
- Tabel rumah_sakit
- Tabel stasiun_KA
- Tabel tempat_belanja
- Tabel tempat_makan
- Tabel terminal

4. IMPLEMENTASI DAN UJI COBA

4.1 Implementasi

Tahap implementasi merupakan tahap penerapan hasil perancangan ke bahasa pemrograman yang dimengerti komputer.

4.2 Uji Coba

4.2.1 Pengujian Keluaran Manual

Obyek cari : stasiun_ka
Jumlah transaksi : 6
Jumlah Itemset : 17
Minimum support (min_sup) : 30% x 6 transaksi = 2

Table 2. Data Transaksi (D) berdasarkan pencarian Obyek Cari dan Neighbour Obyeknya

TID	ItemSet
st.hall	Hotel,rumah_sakit,stasiun_ka,tempat_belanja,tempat_makan,terminal
Ciroyom	obyek_wisata,stasiun_ka,terminal
Andir	stasiun_ka,terminal
Cimindi	stasiun_ka,terminal
Gede bage	stasiun_ka
Kiara condong	stasiun_ka,tempat_belanja,tempat_makan

C1

itemset	SC
hotel	1
obyek wisata	1
rumah_sakit	1
stasiun_ka	6
tempat_belanja	2
tempat_makan	2
terminal	4

Cari ItemSet yang Supcount \geq min_support.

L1

Itemset	Supcount
stasiun_ka	6
tempat_belanja	2
tempat_makan	2
terminal	4

C2

itemset	SC
stasiun_ka, tempat_belanja	2
stasiun_ka, tempat_makan	2
stasiun_ka, terminal	4
tempat_belanja, tempat_makan	2
tempat_belanja, terminal	1
tempat_makan, terminal	1

Cari ItemSet yang Supcount \geq min_support.

L2

itemset	SC
stasiun_ka, tempat_belanja	2
stasiun_ka, tempat_makan	2
stasiun_ka, terminal	4
tempat_belanja, tempat_makan	2

C3

itemset	SC
stasiun_ka, tempat_belanja, tempat_makan	2
stasiun_ka, tempat_belanja, terminal	1
stasiun_ka, tempat_makan, terminal	1
stasiun_ka, tempat_makan, tempat_belanja	2

Cari ItemSet yang Supcount \geq min_support.

L3

itemset	SC
stasiun_ka, tempat_belanja, tempat_makan	2
stasiun_ka, tempat_makan, tempat_belanja	2

L3 \blacktriangleright \blacktriangleleft L3 untuk membangkitkan C4. Karena hasil join-nya tidak memenuhi kombinasi ke 4, maka algoritma apriori berhenti.

4.2.2 Pengujian Keluaran Sistem

Berdasarkan pengujian yang dilakukan pada uji coba manual dan ujicoba sistem diperoleh hasil yang sama. Dengan ini dibuktikan bahwa *output* sistem adalah benar.

4.3 Analisis Data Implementasi

4.3.1 Analisis Close_to

Pendefinisian *close_to* yang digunakan adalah sebagai berikut:

1. *Close_to* didefinisikan berdasarkan radius.
2. *Close_to* didefinisikan berdasarkan jumlah *neighbour object*-nya.

4.3.2 Analisis Jumlah Neighbour object

- Semakin banyak titik pada obyek cari maka semakin banyak pula *neighbour object* yang ditemukan.
- Radius mempengaruhi lebar area pencarian, semakin luas area pencarian maka semakin

banyak *neighbour object* yang ditemukan atau sama dengan area pencarian sebelumnya.

4.3.3 Analisis Jumlah Frequent item dan Rule berdasarkan Radius dan Minimum support

Radius berpengaruh terhadap banyaknya *neighbour object* yang dihasilkan. Oleh karena itu, perbandingan radius dengan *frequent itemset* dan *rule* adalah berbanding lurus. Semakin besar radius, semakin banyak *frequent itemset* dan *rule* yang dihasilkan. Radius akan menghasilkan *frequent item* dan *rule* yang sama dengan radius sebelumnya pada saat *neighbour object* yang ditemukan tidak memberikan semantik yang berbeda.

4.3.4 Analisis Waktu Pembangkitan Neighbour object, Frequent Itemset, Rule

Semakin banyak Neighbour object, Frequent Itemset, Rule yang dibangkitkan semakin banyak waktu yang diperlukan.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

- a. Pencarian *neighbour object* berdasarkan radius per km (kilometer) pada skala peta.
- b. Pencarian *neighbour object* berdasarkan rata-rata *neighbour object* yang diinginkan user.
- c. Definisi *close_to* tidak hanya tergantung dari luas daerah dari obyek pusat, namun *close_to* dapat didefinisikan dari jumlah obyek yang ada di daerah tersebut.
- d. Penambahan radius dari 1 menjadi 2, mengakibatkan penambahan *frequent item* menjadi rata-rata 4 kali lipat untuk *minimum support* 10%. Semakin besar *minimum support* semakin sedikit *frequent itemset* yang dibangkitkan. Semakin banyak *frequent item*, semakin banyak *rule* yang dibangkitkan.
- e. Spatial association *rule* didapat dari association *rule* dimana LHS (left hand side) *rule*-nya mengandung obyek yang menjadi pusat pencarian (obyek cari).

5.2 Saran

- a. Menggunakan data peta yang lebih besar untuk mendapatkan keterhubungan (relasi) antara setiap tipe obyek.
- b. Menganalisis aturan asosiasi berdasarkan data spasial dan data non-spasial sekaligus.

PUSTAKA

[1] Ester, Martin. Krigel, Hans-Peter. Dan Sander, Jörg. 2001. *Algorithm and Application for Spatial Data Mining*. URL: http://www.dbs.informatik.uni-muenchen.de/Publikationen/Papers/Chapter7_revised.pdf

- [2] Güting, Ralf Hartmut. Spatial Database System. URL: <http://www.informatik.fernuni-hagen.de/import/p14/tutorial-neu.pdf>
- [3] Han, Jiawei. Kamber, Micheline. 2001. *Data Mining: Concept and Techniques*. Morgan Kaufmann: USA.
- [4] Hardiyanto, Romi. 2002. *Basis Data dan Query Spasial pada RDBMS POSTGRESQL/POSTGIS*. Penelitian, Departemen Teknik Geodesi, Fakultas Teknik Sipil dan Perencanaan, Institut Teknologi Bandung.
- [5] Koperski, Krzysztof. 1997. *Method Exploring Spatial Association*. URL: <http://db.cs.sfu.ca/GeoMiner/Survey/html/node10.html> . Last update 9 Januari 1997 15:48:06. Download: 08 Januari 2006
- [6] Koperski, Krzyzstof and Han, Jiawei. Discovery od Spatial Association Rule in Geographic Information Database. URL: <http://www.dsi.unive.it/~dm/ssd95.pdf>
- [7] Malerba, Donato. and Lisi, Francesca A. *An ILP Method for Spatial Association Rule Mining*. URL: <http://www.informatik.uni-freiburg.de/~ml/ecmlpkdd/WS-Proceedings/w06/lisi.pdf>
- [8] Malerba, Donato. ESPOSITO, Floriana and Lisi, Francesca A. *Mining Spatial Association Rules In Census Data*. URL: <http://www.di.uniba.it/~malerba/publications/n-tts-spada.pdf>
- [9] Openshaw, Stan. *Geographical data mining: keys design issues*. URL: http://www.geovista.psu.edu/sites/geocomp99/Gc99/051/gc_051.htm Download: 08 Januari 2006.
- [10] Peuquet, Donna J. And Guo, Diansheng. *Mining Spatial Data using An Interactive Rule-Based Approach*. URL: <http://www.giscience.org/GIScience2000/papers/232-Peuquet.pdf>
- [11] Shekhar, Shashi. dan Zhang, Phuseng. 2004. *Spatial Data Mining: Accomplishments and Research Needs*. URL: <http://www-users.cs.umn.edu/~shekhar/talk/giscience04keynote.pdf>