

Seminar: Multimediale Informationssysteme

SQL/MM

SQL Multimedia and Application Packages

Christian Müller

12.07.2002

Übersicht

- 1 Einleitung
- 2 Full Text
- 3 Spatial
- 4 Still Image
- 5 Zusammenfassung

Einleitung

- neue Datentypen für die Speicherung und Verarbeitung multimedialer Daten
- hier behandelt: Text, Bilder, geographische Daten
- kein Audio, Video
- auf Basis der Typerweiterbarkeit von SQL:1999
- UDT (benutzerdefinierte Datentypen) mit eigenen Funktionen und Methoden

Full Text

- zur Speicherung und Verarbeitung von Texten
- Texte als semantische Einheit (Wörter, Sätze, Absätze)
- Zuordnung von Sprache zu den Texten
- zur Suche in den Texten mehrere Methoden
- verschiedene Arten von Suchmustern (Pattern)

Methoden zur Suche

```
SELECT name, b.Text.Score(`SQL`)  
FROM Beispiel b  
WHERE b.Text.Contains(`Standard%`);
```

- Contains: prüft auf Existenz des Pattern im Text
- Score: gibt Auskunft darüber wie gut das Pattern zum Text passt
- NumberOfMatches: Anzahl der Vorkommen des Patterns

einfache Pattern

- Wortsuche:

BeispielText.Contains(' "Standard" ')

- Phrasensuche

BeispielText.Contains(' "internationaler Standard" ')

- Suche nach mehreren Wörtern / Phrasen

BeispielText.Contains(' "Standard_" ')

BeispielText.Contains(' "internationaler % Standard" ')

Pattern

Pattern zur Suche von Ausdrücken, welche in bestimmten Zusammenhängen mit dem Startausdruck stehen:

- ähnlicher Klang
- ähnliche Buchstabenfolge
- Synonyme
- allgemeinere Ausdrücke
- speziellere Ausdrücke

Pattern

Suche nach Synonymen von *Liste* im Thesaurus von *Informatik*

```
BeispielText.Contains( `THESAURUS "Informatik"  
EXPAND SYNONYM TERM OF "Liste" `)
```


Pattern

Suche nach Stammformen

~STEMMED FORM OF "Mäuse" ~ wird zu

~("Maus" , "Mäuse")~

~STEMMED FORM OF „Internationale Standards“ ~ wird zu

~("International Standard" , "Internationaler Standard" ,
"Internationales Standard" , ...)~

Pattern

proximity expansion pattern

```
`(("Standard", "International")  
NEAR "Sprache" WITHIN 0 SENTENCES IN ORDER)`
```

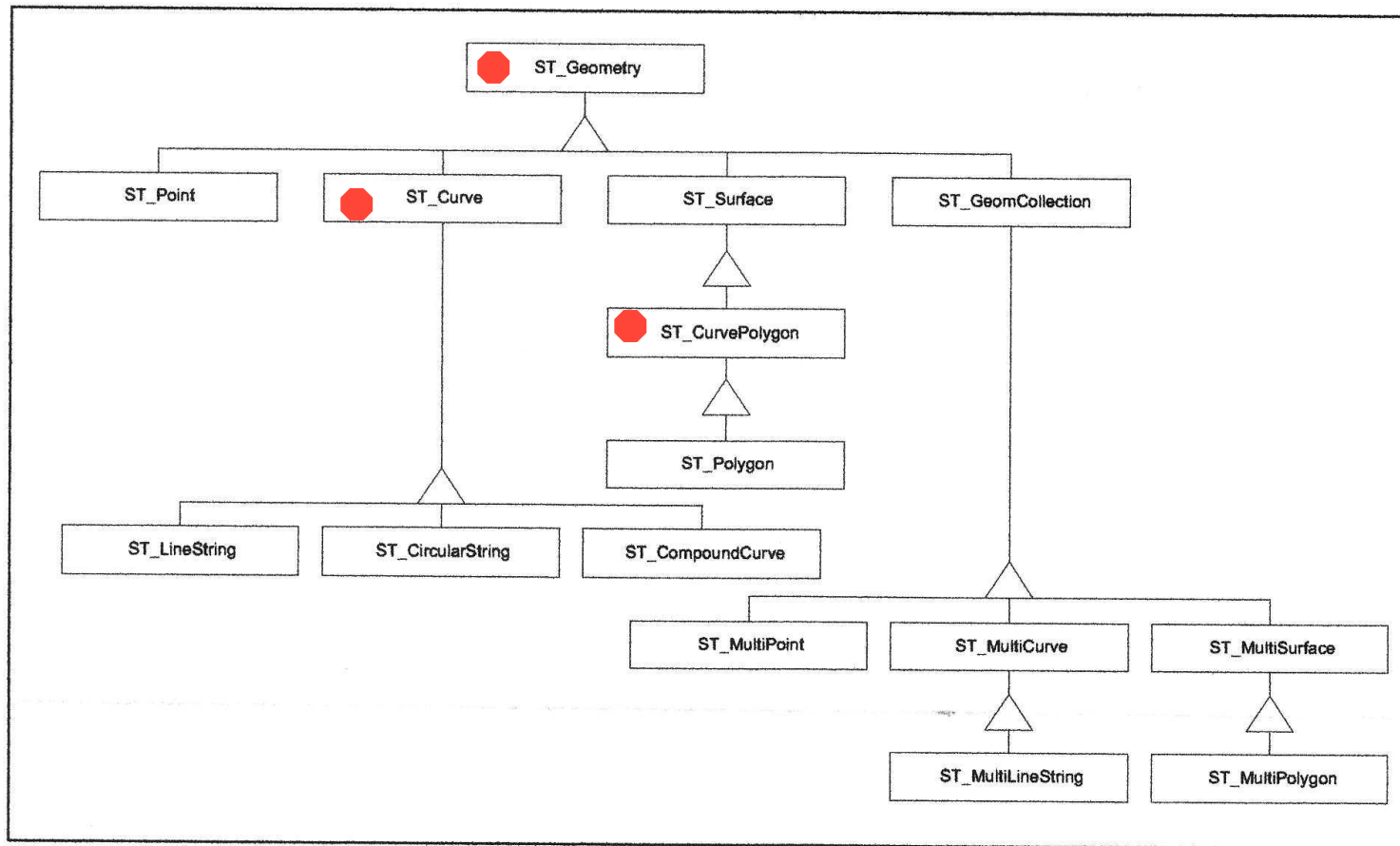
context condition pattern

```
`("Standard", "International")  
IN SAME SENTENCE AS "Spache"`
```

Spatial

- Datentypen zur Speicherung von geographischen Daten
- 0-, 1- oder 2-dimensionale Daten
- Methoden für die Berechnung von Lagebeziehungen
- Einsatzmöglichkeiten bei Stadtplanung, Landaufteilung und militärischer Bereich

Datentyphierarchie

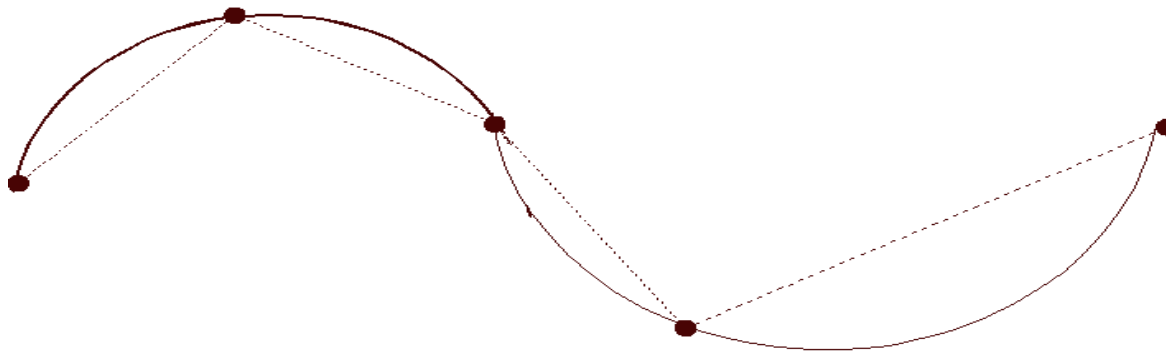


ST_Geometry

- Basistyp für alle anderen Typen
- stellt Standardmethoden bereit (Dimension, Typ,...)
- weiterführende Methoden zwischen ST_Geometry-Werten, wie
 - ST_Intersection
 - ST_Union
 - ST_Difference
 - ST_SymDifference
 - ST_Distance

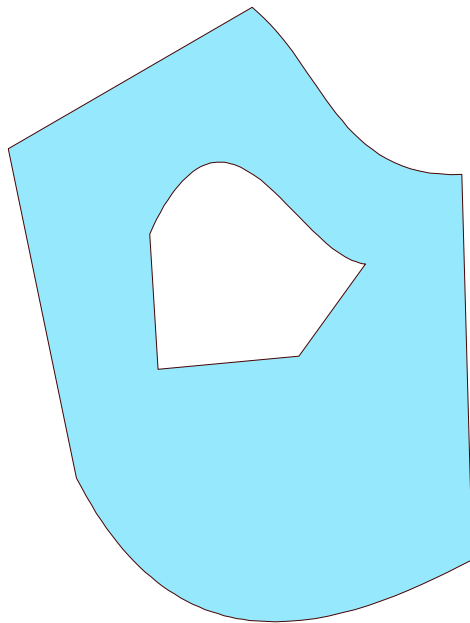
ST_Curve

- eindimensionale Geometrie
- Folge von Punkten
- Unterscheidung der Untertypen (ST_LineString, ST_CircularString)
- ST_CompoundCurve



ST_CurvePolygon

- stellen Flächen dar
- aus Kurven zusammengesetzt



Testen von Lagebeziehungen

- Testen auf Überlappung, Disjunktheit, Überschneidung, Berührungen, ... von zwei Geometrien
- die Methoden zum Testen dieser Beziehungen lassen sich auf eine Methode ST_Relate zurückführen

a.ST_Relate(b, `TF*012***`)

	Interior	Boundary	Exterior
Interior	1: (Interior(a) \cap Interior(b)). ST_Dimension()	2: (Interior(a) \cap Boundary(b)). ST_Dimension()	3: (Interior(a) \cap Exterior(b)). ST_Dimension()
Boundary	4: (Boundary(a) \cap Interior(b)). ST_Dimension()	5: (Boundary(a) \cap Boundary(b)). ST_Dimension()	6: (Boundary(a) \cap Exterior(b)). ST_Dimension()
Exterior	7: (Exterior(a) \cap Interior(b)). ST_Dimension()	8: (Exterior(a) \cap Boundary(b)). ST_Dimension()	9: (Exterior(a) \cap Exterior(b)). ST_Dimension()

ST_Disjoint

$a.ST_Disjoint(b) \Leftrightarrow$

$((Interior(a) \cap Interior(b) = \emptyset) \wedge$

$(Interior(a) \cap Boundary(b) = \emptyset) \wedge$

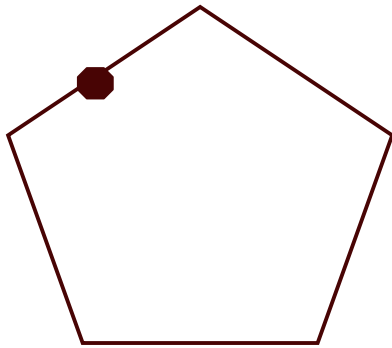
$(Boundary(a) \cap Interior(b) = \emptyset) \wedge$

$(Boundary(a) \cap Boundary(b) = \emptyset)) \Leftrightarrow$

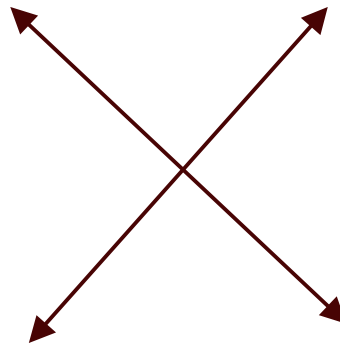
$a.ST_Relate(b, \text{'FF*FF*FF*FF'})$

Auf ähnliche Weise könne auch andere Lagebeziehungen getestet werden.

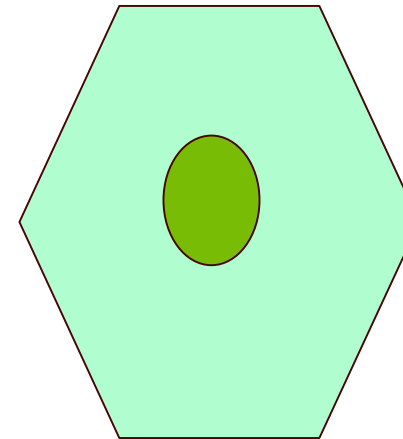
Beispiele für Lagebeziehungen



Berührung
ST_Polygon/ST_Point



Überschneidung
ST_LineString/ST_LineString



Beinhalten
ST_Polygon/ST_CurvePolygon

externe Datenformate

zum Ein- bzw. Auslesen der Daten gibt es drei Formate

- Binärrepräsentation
- Textrepräsentation
- XML-Repräsentation (GML)

```
<Point gid="P1">
```

```
  <coord><X>5.0</X><Y>0.57</Y></coord>
```

```
</Point>
```

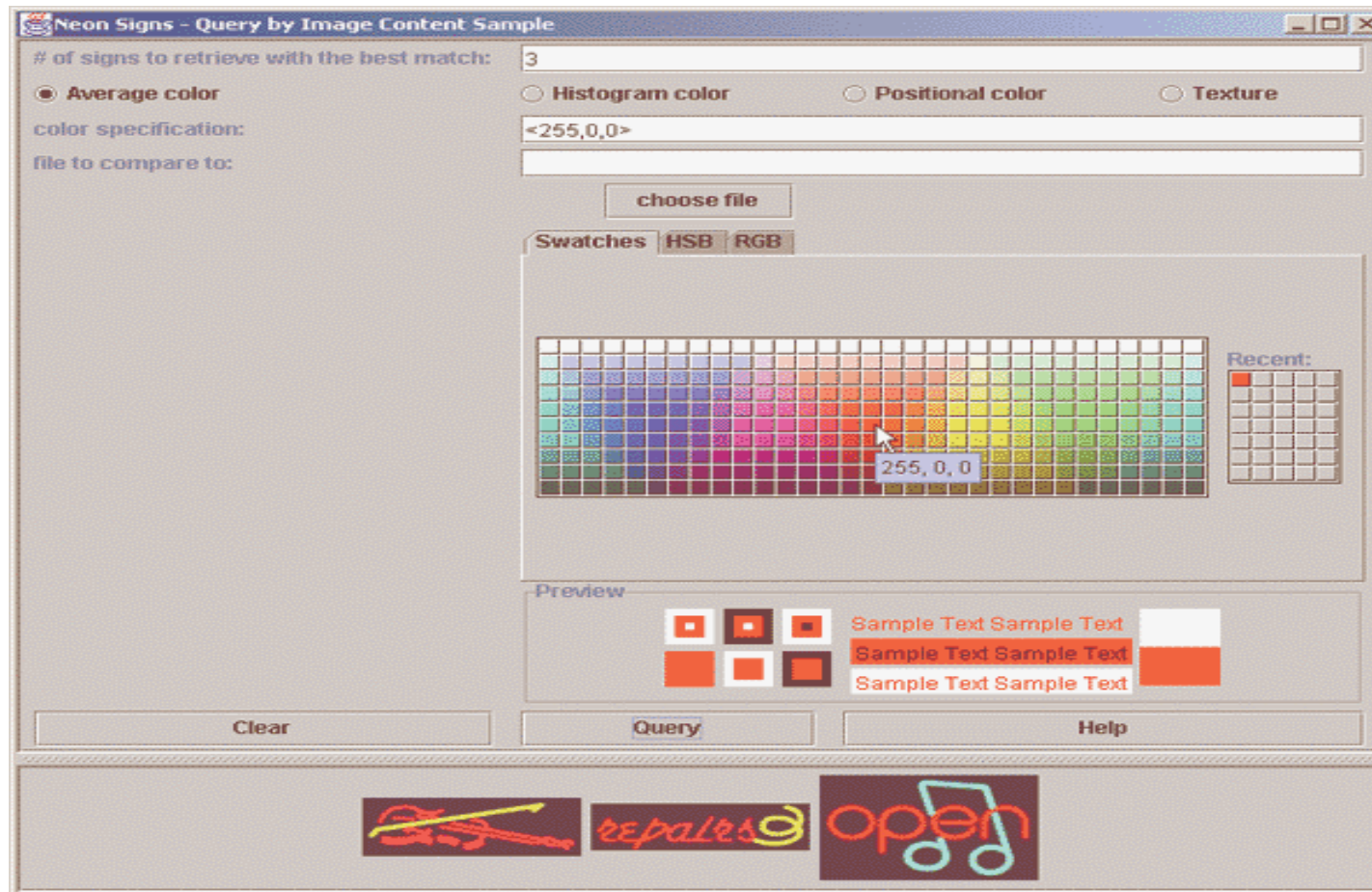
Still Image: Einleitung

- zur Speicherung und Verarbeitung von Bildern in Datenbanken
- Suche nach Bildern mit Hilfe von „visuellen“ Prädikaten
- dazu Definition von Bildmerkmalen:
 - SI_AverageColor
 - SI_ColorHistogram
 - SI_PositionalColor
 - SI_Texture
 - SI_FeatureList: Kombination

SI_AverageColor

- durchschnittliche Farbe
- wird berechnet durch:
 - Aufsummierung der einzelnen Farbwerte jedes Pixels (z.B. rot, grün, blau)
 - Division durch Anzahl der Pixel
- unterschiedliche Bilder können gleiche Werte besitzen

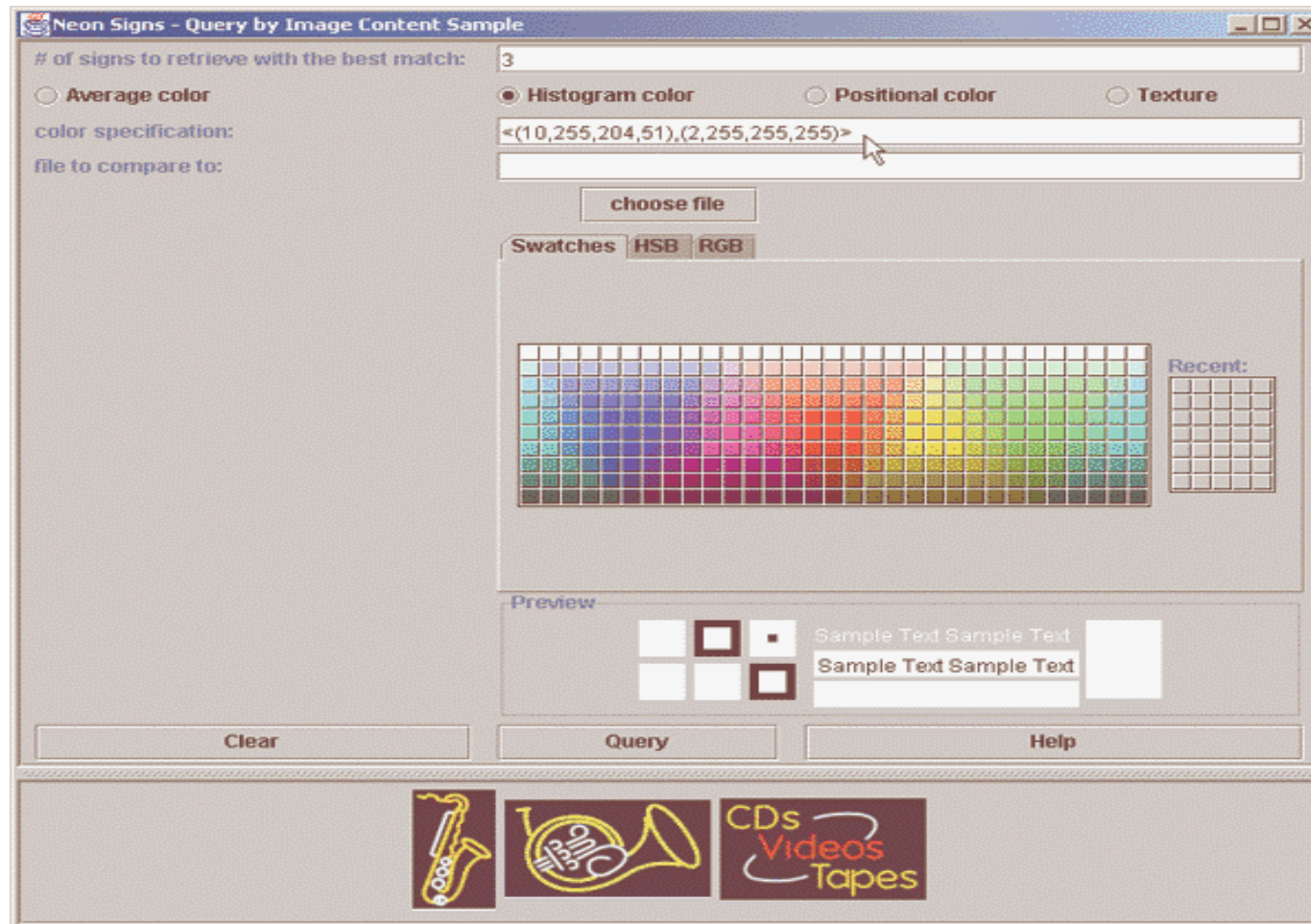
SI_AverageColor



SI_ColorHistogram

- relative Frequenz der Farben
- Zerlegung des Farbraumes
- zu jedem Teil des Farbraums werden die Pixel gezählt
- Division durch Anzahl der Pixel

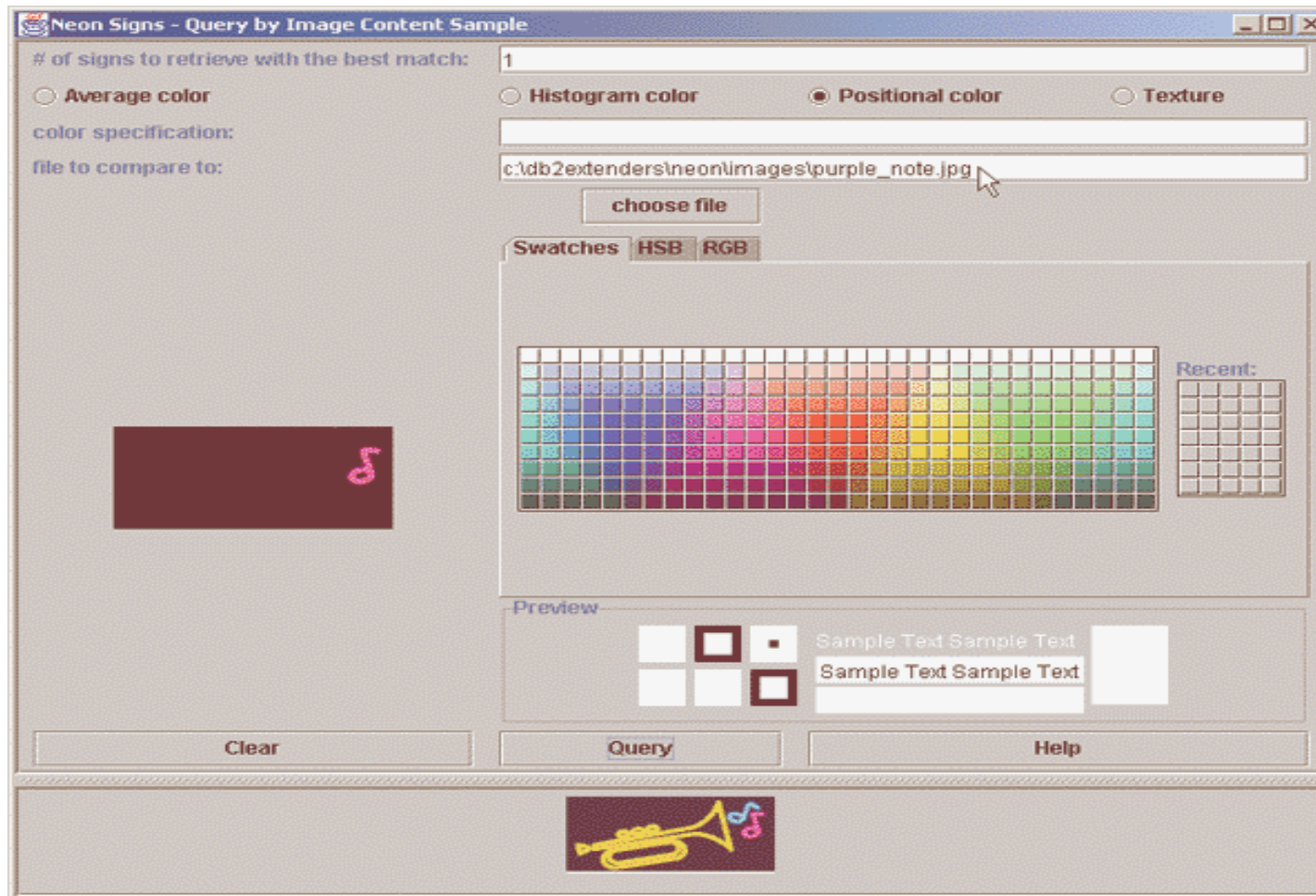
SI_ColorHistogram



SI_PositionalColor

- durchschnittliche Farbe für Teil des Bildes
- Bild wird in n mal m Rechtecke eingeteilt
- Berechnung für diese Bereiche analog zu SI_AverageColor

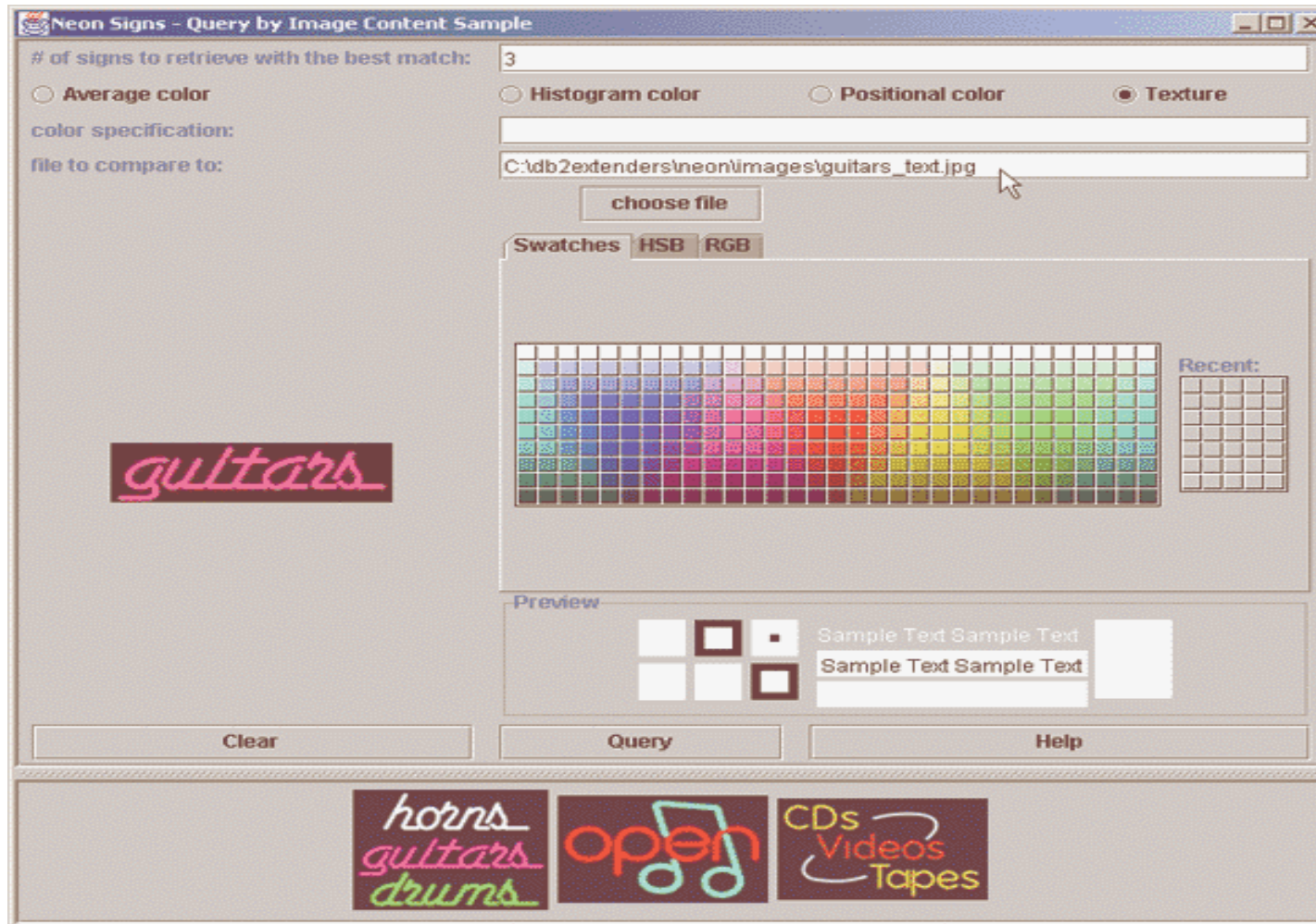
SI_PositionalColor



SI_Texture

- Charakterisierung anhand Grobheit und Kontrast
- keine genaueren Angaben im Standard
- große Variationen möglich

SI_Texture



SQL/MM: Zusammenfassung

- neue Datentypen für Full Text, Spatial und Still Image
- bessere vereinheitlichte Möglichkeiten zur Suche in Texten, geographischen Daten und Bildern
- Relevanz in kommerziellen DBS

Ende