



Räumliche Informationsrecherche am Beispiel des Umweltinformationsnetzes Deutschland *gein*[®]

Dr. Thomas Vögele
Koordinierungsstelle UDK/GEIN

GIS Nachmittag
11. Mai 2005
GIS Zentrum
Universität Hannover

- Die Koordinierungsstelle UDK/GEIN (KUG)
- Das Umweltinformationsnetz Deutschland (*gein*[®])
- Räumliche Informationsrecherche,
- Was ist räumliche Relevanz?
- 3 Methoden der räumlichen Informationsrecherche
 - Methode 1: Koordinaten, Bounding Boxes (Standard)
 - Methode 2: Räumliche Indizes auf der Basis von regulären Referenzgittern (Geothesaurs Umweltbundesamt, *gein*[®])
 - Methode 3: Räumliche Indizes auf der Basis polygonaler Referenz tessellierungen (Schlieder, Vögele)
- Zusammenfassung



Rahmenbedingungen:

- Koordinierungsstelle UDK/GEIN (KUG) besteht seit Anfang 2003
- Angesiedelt im Niedersächsischen Umweltministerium
- Wird über Bund-Länder Verwaltungsvereinbarung finanziert

Aufgaben:

- Inhaltliche und technische Betreuung von Umweltdatenkatalog (UDK) und Umweltinformationsnetz Deutschland *gein*[®]
- Weiterentwicklung und Verbesserung von UDK und *gein*[®]
→ **Portal-U** (ab 2006)



- „German *Environmental Information Network*“
- Entwickelt im Auftrag des Umweltbundesamtes
- online seit Expo 2000 (www.gein.de)
- Behördliche Umweltinformationen aus Bund und Ländern
- Über 400.000 Webseiten im *gein*[®]
- 8 FIS/Metadatenbanken mit über 500.000 Einträgen
- Umweltportal
 - Themenseiten, Aktuelle Meldungen, Messwerte, und
 - **Suche**

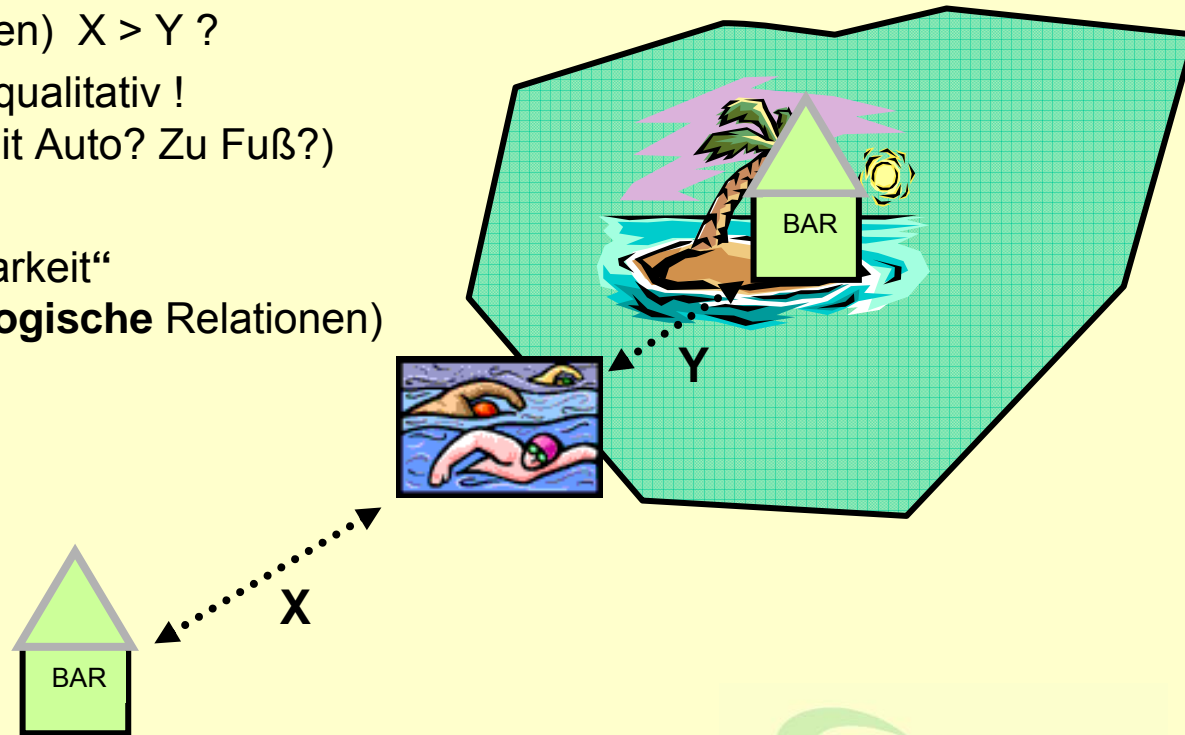
„**Naturschutzgebiet**“ in der „**Lüneburger Heide**“ ?

→ Naturschutzgebiet, das räumlich relevant ist zur Lüneburger Heide

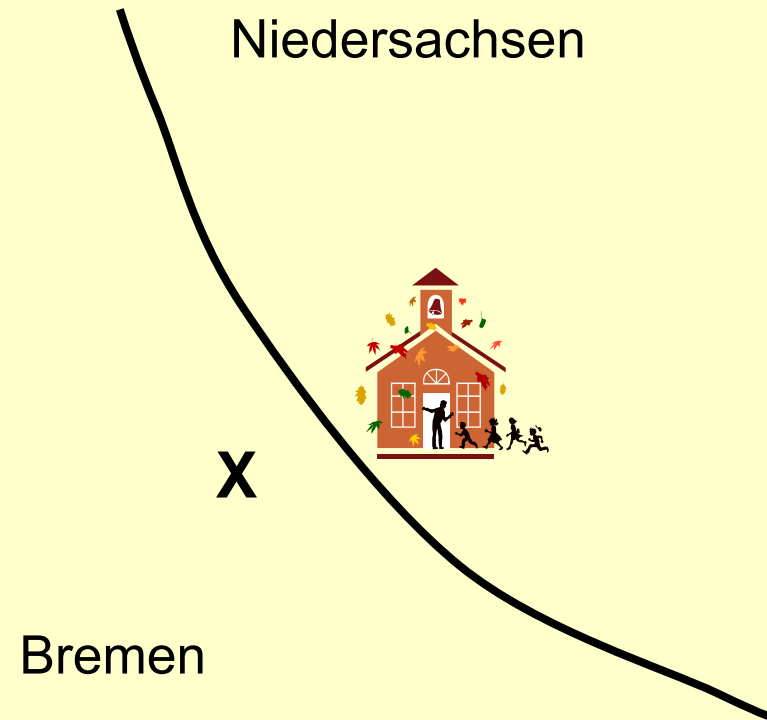
Offene Fragen:

- Was ist räumliche Relevanz?
- ~~Wie können Webseiten etc. (automatisch) georeferenziert werden?~~
- Wie kann die räumliche Relevanz der Anfrage-Location („Lüneburger Heide“) zu einer Referenz-Location (z.B. „Landkreis SFA“ bzw. Koordinaten) berechnet werden?
- Wie kann diese Berechnung performant (on-the-fly) durchgeführt werden?
- Wie können Naturräume und andere Regionen, deren Grenzen nicht eindeutig definiert sind, modelliert werden?

- First Law of Geography:
„In geographic space, everything is related to everything else, but **near** things are more related than **distant** things“ (Tobler, 1970)
- „Entfernung“ bzw. „Nähe“
(**metrische** Relationen) $X > Y$?
- Problem: „Nähe“ oft qualitativ !
(„5 min entfernt“ → mit Auto? Zu Fuß?)
- Aber auch „Erreichbarkeit“
(Verknüpfung, **topologische** Relationen)

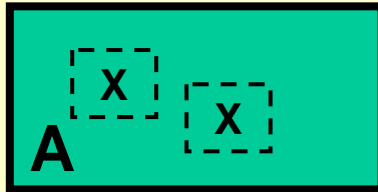


- Menschen organisieren räumliche Information hierarchisch, als **Partonomien** (Hirtle, 1985; Timpf, 1999)
- Beispiel: Verwaltungseinheiten
- Zugehörigkeit zu einer Partonomie
 - „X in Y“ (räumlich), aber auch
 - „X ist Teil von, gehört zu Y“ (partonomisch/mereologisch)
- Partonomische Relationen können wichtiger sein als metrische oder topologische Relationen
- Hängt ab vom Kontext der Suchanfrage, Typ des angefragten Objekts



Wie können diese Relationen zur Berechnung der räumlichen Relevanz für eine räumliche Suchanfrage berechnet werden?

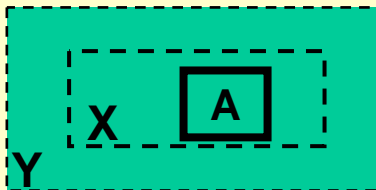
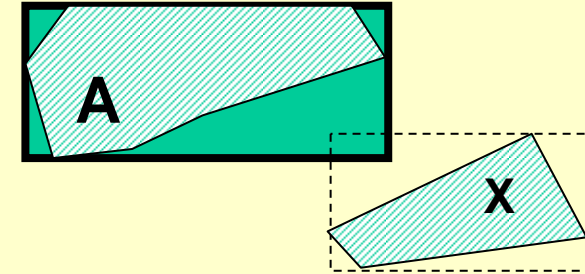
- Methode 1: Bounding Boxes, Koordinaten
- Methode 2: Räumliche Indizes mit regulären Referenzgittern
- Methode 3: Räumliche Indizes mit irregulären Referenzgittern (graph-basierte polygonale Tessellierungen)



Welche X sind in A
enthalten?



Welche X werden
von A geschnitten?



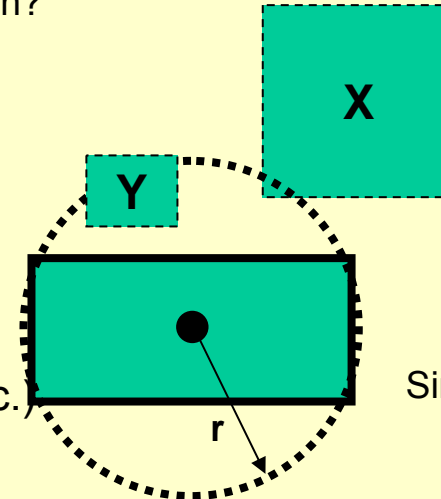
A ist enthalten in X, Y?

Vorteile:

- Geometrien vorhanden (außer: Naturräume etc.)
- Weniger (Rechen)aufwand als Polygone

Nachteile:

- Sehr grobe Repräsentation
- Problematisch insbesondere für lineare Objekte (Flüsse etc.)
- „koordinatentypische“ Probleme (z.B. Raumbezugssystem)
- Nur quantitative Betrachtung (Entfernungen, keine partonomische Hierarchie)



Sind X und Y im
Suchradius r?

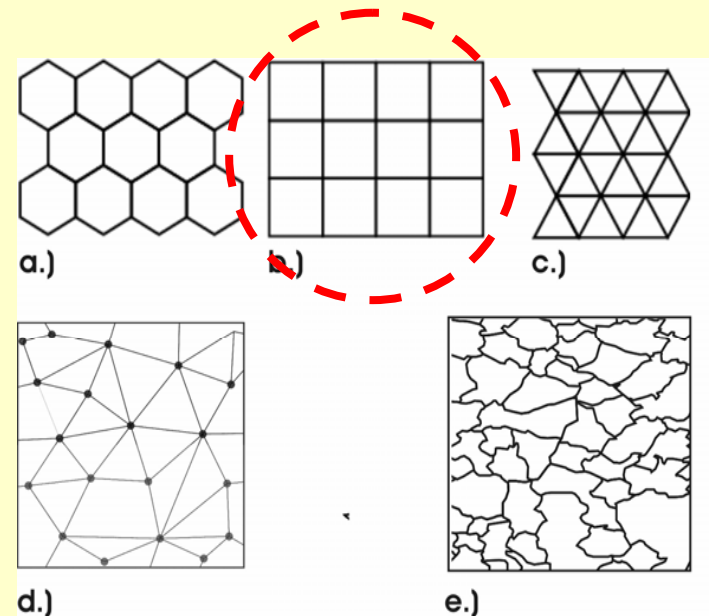
Methode 2: Referenzgitter und räumliche Indizes

Prinzip:

- Statt kontinuierlichem Raum (Koordinaten)
→ diskreter Raum (Referenzraster) und
räumliche Indizes
 - Bsp: Stadtplan

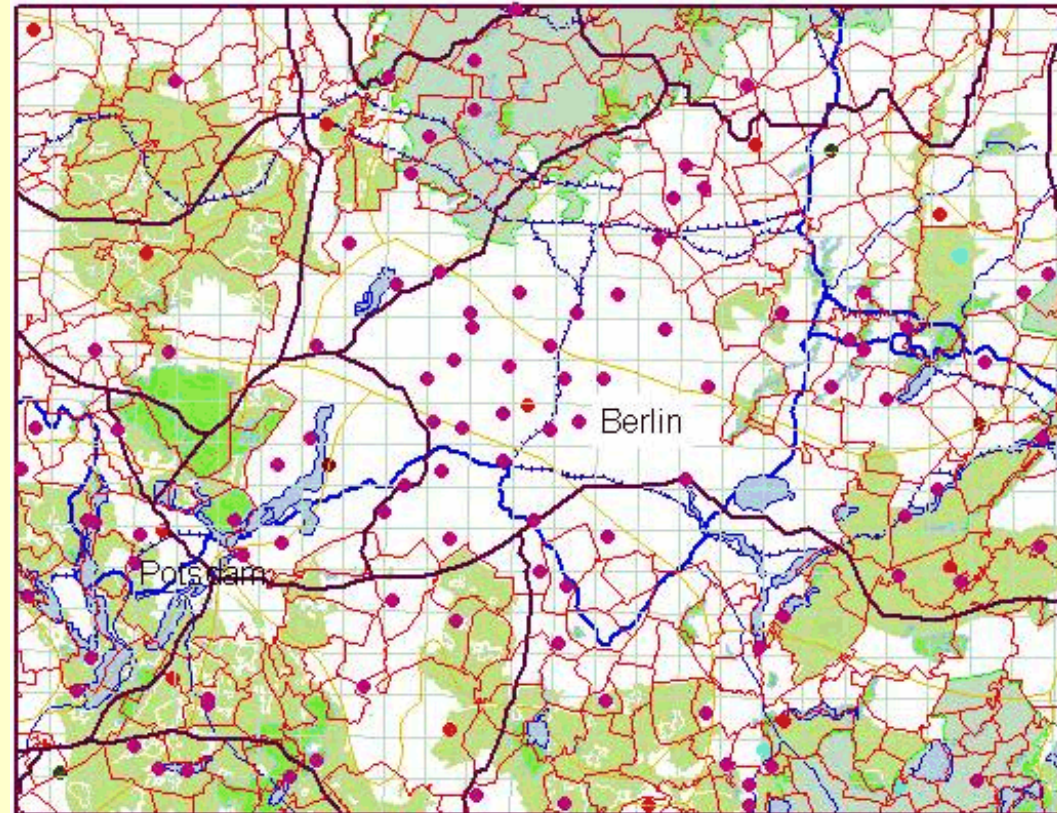
Vorteile:

- Genauigkeit besser als BB
- Aufwand geringer als Polygone
- Relationen (topologisch, metrisch) können
einfach berechnet werden
- Besser geeignet für Repräsentation
unscharf definierter Regionen
(z.B. Naturräume)

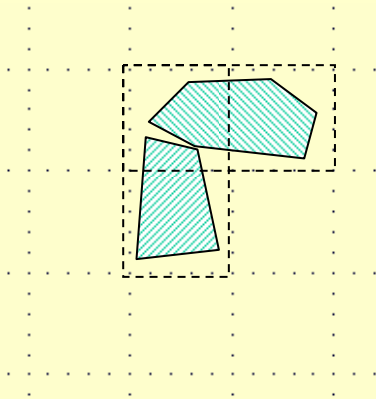


Reguläre und
irreguläre
Tessellierungen

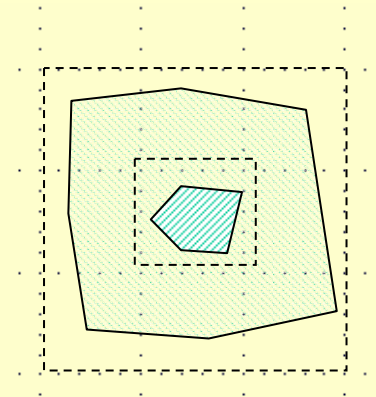
- UBA Geothesaurs (SNS)
- Räumliche Suche in gein
- 3x3km Referenzgitter
- Über 48000 geografische Namen
- Darunter fast 3500 Naturräume, Flüsse, Einzugsgebiete etc.



- Räumliche Relevanz wird als Funktion der Überlappung zweier Regionen auf Basis der überdeckten Rasterzellen berechnet
- $\text{Relevanz}(A,B) = \text{Überlappung} / \text{Gesamtfläche}$
- Metrische Relationen werden nicht berücksichtigt



Gesamtfläche = 3
 Überlappung = 1
 Relevanz = 1/3



Gesamtfläche = 9
 Überlappung = 1
 Relevanz = 1/9

The screenshot shows the search interface of the 'DAS PORTAL für Umweltfragen' website. The page has a green and blue color scheme. At the top left is the 'gein' logo. The main title is 'DAS PORTAL für Umweltfragen' with a subtitle 'Umweltinformation von Bund und Ländern'. A navigation bar contains links for 'Home', 'Über gein', 'Mitmachen', 'Hilfe/FAQ', 'english', and 'Impressum'. The search section is titled 'SUCHE' and features a search box containing the text 'Naturschutzgebiet in der Lüneburger Heide'. Below the search box is a 'Früher verwendete Suchtexte' dropdown menu with an 'OK' button. To the right of the search box is a 'Suche mit Recherche Assistent' section with a 'Suchvorschläge in' dropdown menu set to 'deutsch' and a 'Weiter' button. Below this is a 'Volltextsuche' section with a description: 'Einfache Suche mit dem eingegebenen Suchtext'. At the bottom of the search section is a 'Suchen' button and a dropdown menu for 'Ergebnisse finden...' set to 'mit allen Suchbegriffen(und)'. The bottom of the page is divided into three columns: 'AKTUELLES' with a 'Neueste Meldung vom: 03.05.2005 14:47:59 Uhr' and links for 'News' and 'Messwerte'; 'THEMENSEITEN' with a link to 'Blättern in 761 ausgewählten Themenseiten' and a list of topics including 'Allgemeine Umweltfragen', 'Abfall', 'Boden', 'Chemikalien', 'Energie/Rohstoffe', 'Ernährung/Landwirtschaft', 'Gentechnik', 'Gesundheit', 'Lärm/Erschütterungen', 'Luft', 'Natur/Landschaft', 'Strahlung', 'Umweltrecht', 'Wirtschaft', and 'Wasser'; and 'UMWELTKALENDER' with a link to 'Blättern im Umweltkalender' and a description: 'Wann war noch... "Tschernobyl", wann die "Konferenz von Rio"? Stöbern Sie hier...'. The footer contains the 'udk gein' logo and the text 'Koordinierungsstelle'.

RECHERCHE ASSISTENT

Eingegeben: Naturschutzgebiet in der Lüneburger Heide
Untersucht: [Naturschutzgebiet, Lüneburger, Heide] Zurück zur Sucheingabe

Aus Ihrem Suchtext wurden sinnentsprechende Begriffe ermittelt und den Kategorien Thema, Raum und Zeit zugeordnet:

Thema	Raum	Zeit
<p>Angezeigt: 2 von 2 Angaben</p> <p>Bitte wählen Sie ein oder mehrere Schlagwörter für die Suchanfrage aus:</p> <p><input type="checkbox"/> Naturschutzgebiet <input type="checkbox"/> Rechtsverordnung</p> <p>Suchwort hinzufügen: Groß-/Kleinschreibung! <input style="width: 100%;" type="text"/> <input type="button" value="Hinzufügen"/></p> <p>Ergebnisse finden... <input type="text" value="mit allen Begriffen"/></p>	<p>Angezeigt: 10 von 46 Angaben</p> <p>Bitte wählen Sie einen oder mehrere Namen für die Suchanfrage aus:</p> <p><input type="checkbox"/> Klietzer Heide Landschaft <input type="checkbox"/> Ückerländer Heide Naturraum <input type="checkbox"/> Tornau vor der Heide Gemeinde <input type="checkbox"/> Dübener Heide Landschaft <input type="checkbox"/> Hohe Heide Landschaft <input type="checkbox"/> Reicherskreuzer Heide und Schwansee Schutzgebiet <input type="checkbox"/> Muskauer Heide Landschaft <input type="checkbox"/> Diedersdorfer Heide und Großbeerener Grabe Schutzgebiet <input checked="" type="checkbox"/> Lüneburger Heide Naturpark <input type="checkbox"/> Lieberoser Heide und Schlaubegebiet Naturraum</p> <p style="text-align: center;"><input type="button" value="Alle Anzeigen"/></p> <p>Name hinzufügen: Groß-/Kleinschreibung! <input style="width: 100%;" type="text"/> <input type="button" value="Hinzufügen"/></p>	<p>Keine Zeitangaben gefunden.</p> <p>Sie haben folgende Möglichkeiten: - Ändern Sie Ihre Sucheingabe - Verwenden Sie die Volltextsuche - Fügen Sie eine Zeitangabe hinzu - Hilfe/FAQ enthält weitere Hinweise</p> <p>Zeitbezug hinzufügen: z.B. 12.3.2000, 2003, 05.2001 ... <input style="width: 100%;" type="text"/> <input type="button" value="Hinzufügen"/></p> <p>Ergebnisse finden... <input type="text" value="Zeitraum / Zeitpunkt"/></p>



FACHBEGRIFFE

aus der SNS-Topic Map.

[zurück zum Recherche Assistenten](#)

Geografischer Name (Quelle: UBA Geo-Thesaurus Umwelt)

[Zurück zum Seitenanfang](#) ↑

<i>Name:</i>	Lüneburger Heide
<i>Geo-Typ</i>	Naturpark
<i>Ihre Eingabe</i>	Naturschutzgebiet in der Lüneburger Heide
<i>Status</i>	wurde bereits vorgeschlagen.

Lagebeziehungen

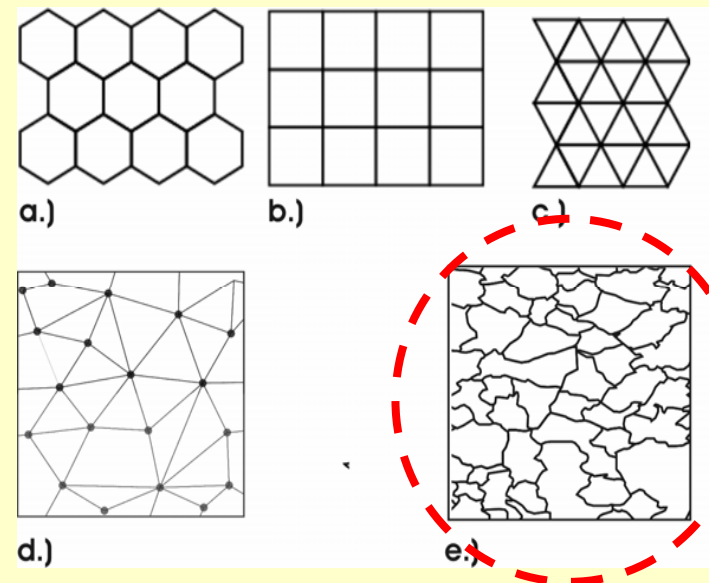
[Zurück zum Seitenanfang](#) ↑

[Undeloh](#) (50%, Gemeinde)
[Bispingen](#) (50%, Gemeinde)
[Schneeverdingen](#) (50%, Gemeinde)
[Egestorf](#) (30%, Gemeinde)
[Hanstedt](#) (30%, Gemeinde)
[Hohe Heide](#) (30%, Naturraum)
[Wümme](#) (20%, Fluß)
[Luhe](#) (20%, Fluß)
[Seeve](#) (20%, Fluß)
[Este](#) (20%, Fluß)
[Buchholz in der Nordheide](#) (20%, Gemeinde)
[Handeloh](#) (20%, Gemeinde)
[Welle](#) (20%, Gemeinde)
[Harburg](#) (20%, Kreis)
[Soltau-Fallingb.ostel](#) (20%, Kreis)
[Wümme](#) (20%, Wassereinzugsgebiet)
[Luhe](#) (20%, Wassereinzugsgebiet)
[Elbe von Lühe bis Oste](#) (20%, Wassereinzugsgebiet)
[Wilseder Berg](#) (10%, Berg)
[Böhme](#) (10%, Fluß)
[Veerse](#) (10%, Fluß)
[Asendorf](#) (10%, Gemeinde)
[Jesteburg](#) (10%, Gemeinde)
[Garlstorf](#) (10%, Gemeinde)
[Gödenstorf](#) (10%, Gemeinde)
[Soltau](#) (10%, Gemeinde)
[Lüneburger Heide](#) (10%, Landschaft)
[Staader Geest](#) (10%, Naturraum)
[Wümmeniederung](#) (10%, Naturraum)
[Lüneburger Heide](#) (10%, Naturraum)
[Südheide](#) (10%, Naturraum)
[Böhme](#) (10%, Wassereinzugsgebiet)

Methode 3: Polygonale Tessellierungen

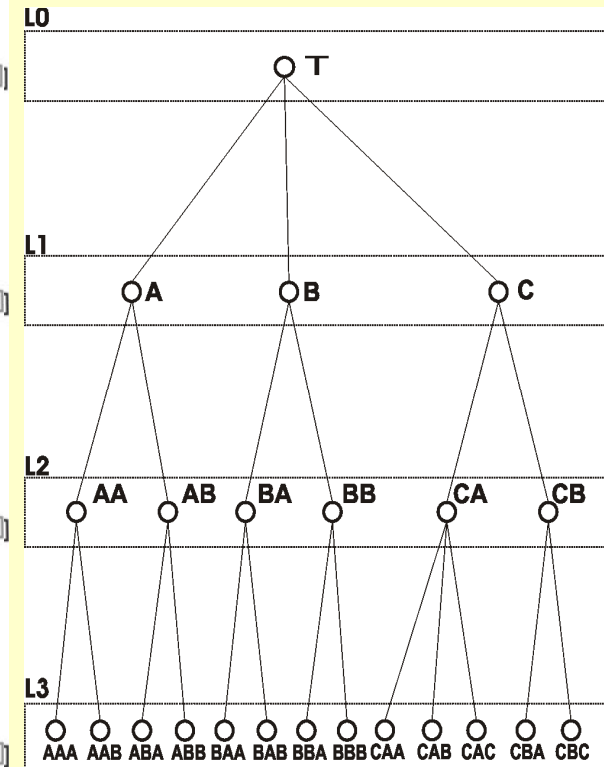
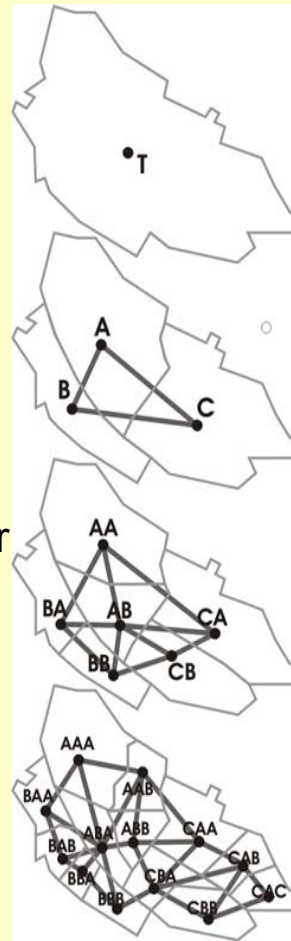
- ▶ Nutzung standardisierter polygonaler Tessellierungen als Referenzsystem
 - ▶ Administrative Einheiten, Postleitzahlbezirke
 - ▶ Standardisierte Ortsnamen
 - ▶ Intuitiv: Werden vielfach bereits zur indirekten Georeferenzierung genutzt
 - ▶ Polygonale Repräsentation in vielen Organisationen vorhanden

- ▶ Aber: Enthalten redundante Information, können weiter vereinfacht werden
 - ▶ Umrisse (Polygone) und absolute Ortsangaben (Koordinaten) überflüssig
 - ▶ Benötigt werden nur Nachbarschaftsbeziehungen
 - ▶ Deshalb: Graphbasierte Abstraktion

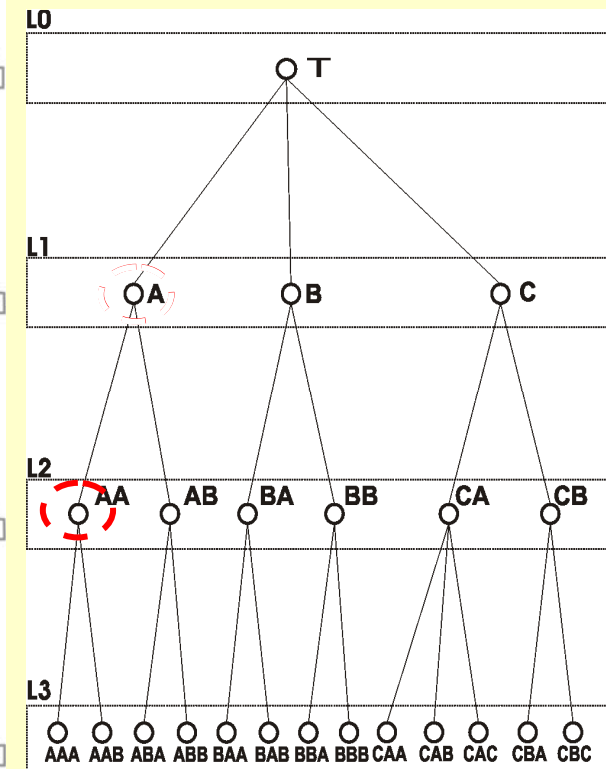
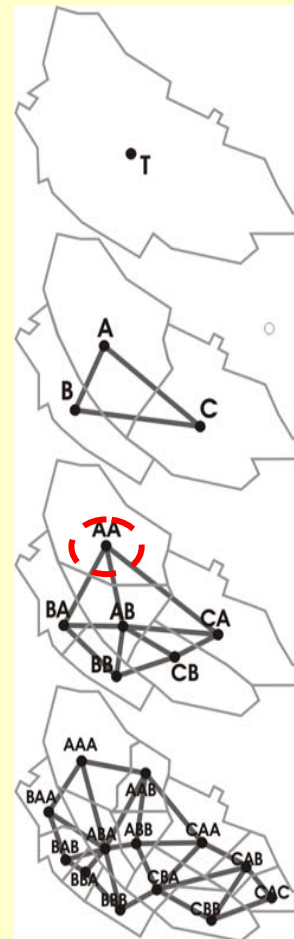


Reguläre und
irreguläre
Tessellierungen

- ▶ Basis: Tessellierung administrativer Einheiten
 - ▶ z.B. Staat, Länder, Landkreise, Gemeinden
- ▶ „Decomposition Tree“
 - ▶ Rekursive partonomische Zerlegung der Tessellierung
- ▶ Nachbarschaftsgraph
 - ▶ Graphbasierte Abstraktion einer Tessellierung auf der Dekompositionsebene L
 - ▶ Connection (Nachbarschafts) Graph (Schlieder et al. 2001)



- Räumliche Relevanz als Funktion der Graphentfernungen
- „Vertikale Entfernung“
- Enthalten in A?
 - AA Kind von A $\rightarrow \sigma_{AA} = \sigma_A (= 1)$
 - BA nicht Kind von A $\rightarrow (\sigma_A = \sigma_{AA}) > \sigma_{BA}$
- AA ist Teil von ...?
 - A ist Vater von AA $\rightarrow \sigma_A < \sigma_{AA}$
- „Horizontale Entfernung“
- „In der Nähe“ von AA?
 - AB dir. Nachbar von AA
BA dir. Nachbar von AA $\rightarrow (\sigma_{AB} = \sigma_{BA}) < (\sigma_{AA})$
 - AB dir. Nachbar von AA;
BB dir. Nachbar von AB $\rightarrow \sigma_{BB} < \sigma_{AB} < \sigma_{AA}$

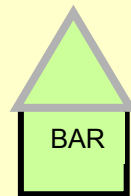


Kumulative Distanz zweier Referenzeinheiten

Horizontale Distanz Vertikale Distanz

$$\varepsilon(r_i, r_q) = \alpha \delta(r_i, r_q) + (1 - \alpha) \nu(r_i, r_q)$$

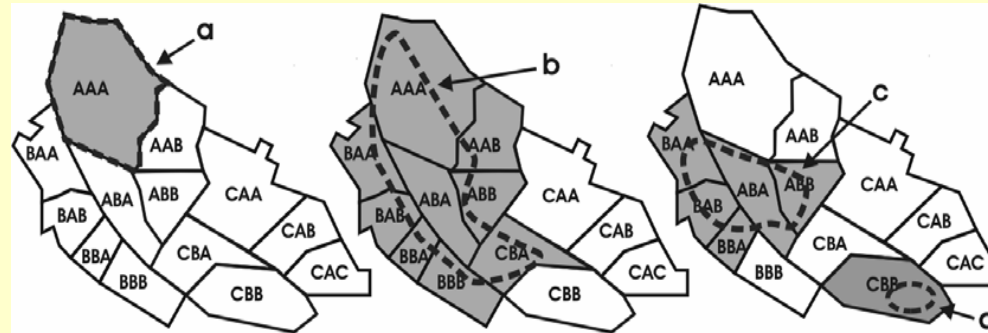
- Kumulative Entfernung als lineare Verknüpfung von δ und ν
- Gewichtung über Faktor α ist kontextabhängig



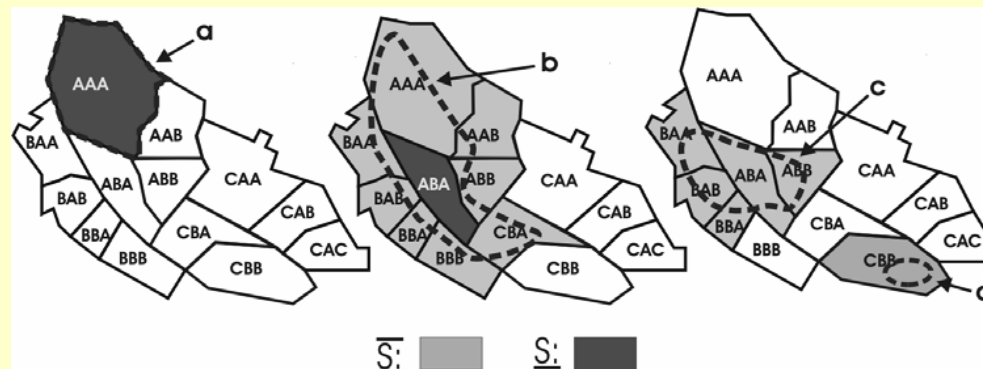
$\alpha = 1$



$\alpha = 0$



„simple“
spatial footprint



„complex“
spatial footprint

- ▶ Regionen werden als Mengen von „Referenzeinheiten“ modelliert
 - ▶ $c = \{ABA, ABB, BAA, BAB, BBA\}$
- ▶ Räumliche Relevanz zweier Regionen wird als Funktion für die jeweiligen Referenzeinheiten berechneten Relevanzen ermittelt.

- Raumbezogene Suche spielt in der Praxis eine wichtige Rolle
- Verschiedene Ansätze möglich
- Polygone
 - Genau, aber „teuer“ (rechenintensiv, datenintensiv)
 - Ungenau definierte Regionen oft nicht polygonal vorhanden
- Umfassende Rechtecke (bounding boxes):
 - Einfacher, aber ungenau (falsche Überschneidungen etc.)
- Koordinatenbasierte Ansätze generell:
 - Hohe Anforderungen an Datenqualität
 - Rechenintensiv
 - Hohes Maß an Standardisierung erforderlich (Referenzsysteme)
 - Qualitative Relationen (qual. Entfernung, Position in Partonomie) werden nicht berücksichtigt

- Räumliche Indizes generell:
 - Genauigkeit der Repräsentation zwischen BB und Polygon
 - Relativ wenig Rechenaufwand
 - Gut geeignet um ungenau definierte Regionen (z.B. Naturregionen) abzubilden
- Methode 2: Geothesaurs mit regulärer Tessellierung (Referenzgitter)
 - Umfangreicher Geothesaurs vorhanden
 - Begrenzte und festgesetzte Auflösung, standardisiertes Gitter notwendig
 - Relevanzmaß bildet „Enthaltensein“ nur bedingt ab
 - „Nähe“ von Regionen wird nicht berechnet
- Methode 3: Graphbasierte polygonale Tessellierungen
 - Effiziente Berechnung möglich
 - Semi-qualitatives Maß für Entfernungsrelation
 - Partonomische und räumliche Entfernung werden berücksichtigt
 - Standardisierte Referenz tessellierungen existieren bereits (admin. Einheiten)
 - Bis jetzt kein umfangreicher Geothesaurs vorhanden

Vielen Dank für Ihre Aufmerksamkeit !

Wenn Sie Fragen haben wenden Sie sich bitte an die Webseite der
Koordinierungsstelle

www.udk-gein.de

www.gein.de

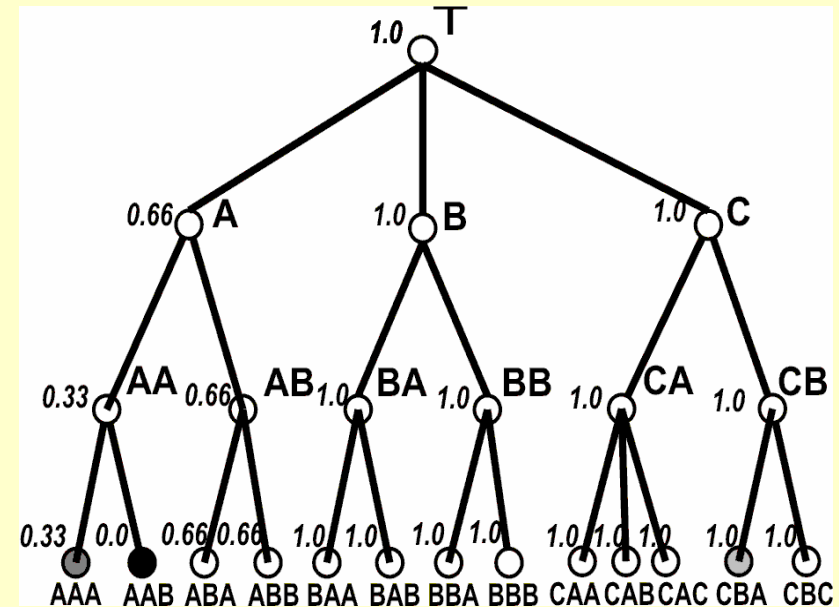
www.umweltdatenkatalog.de

oder kontaktieren Sie uns per e-mail

kug@numis.niedersachsen.de

- Vertikale Distanz \mathcal{D} wird im Decomposition Tree berechnet
- Ist ein Maß für die partonomische Nähe zweier Regionen
- Kürzester Pfad zum nächsten gemeinsamen Vaterknoten
- Normalisiert mit max. Tiefe des Baumes

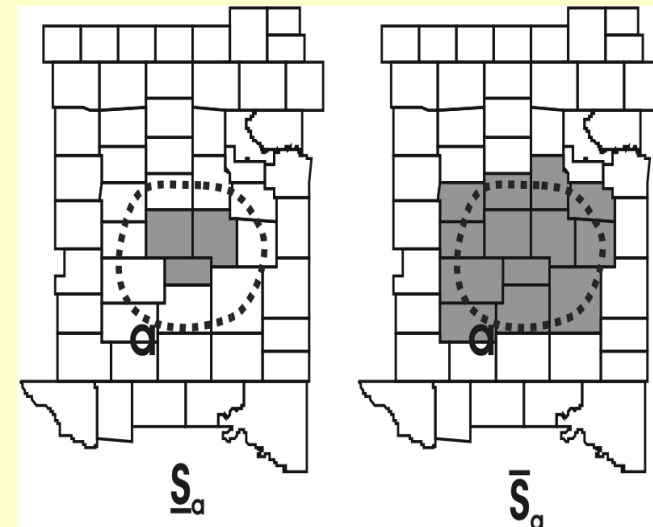
- $D = 3$
- $\mathcal{D}(AAB, AAA) = 1/3$
- $\mathcal{D}(AAB, ABA) = 2/3$
- $\mathcal{D}(AAB, CBA) = 3/3$



$$v_{norm}(r_i, r_q) = \frac{SP(r_P, r_q)}{D}$$

- ▶ Komplexer (Complex) Spatial Footprint
 - ▶ Nutzt Ansatz von Worboys, 1998
 - ▶ Untere Approximation = minimale Ausdehnung
 - ▶ Obere Approximation = maximale Ausdehnung

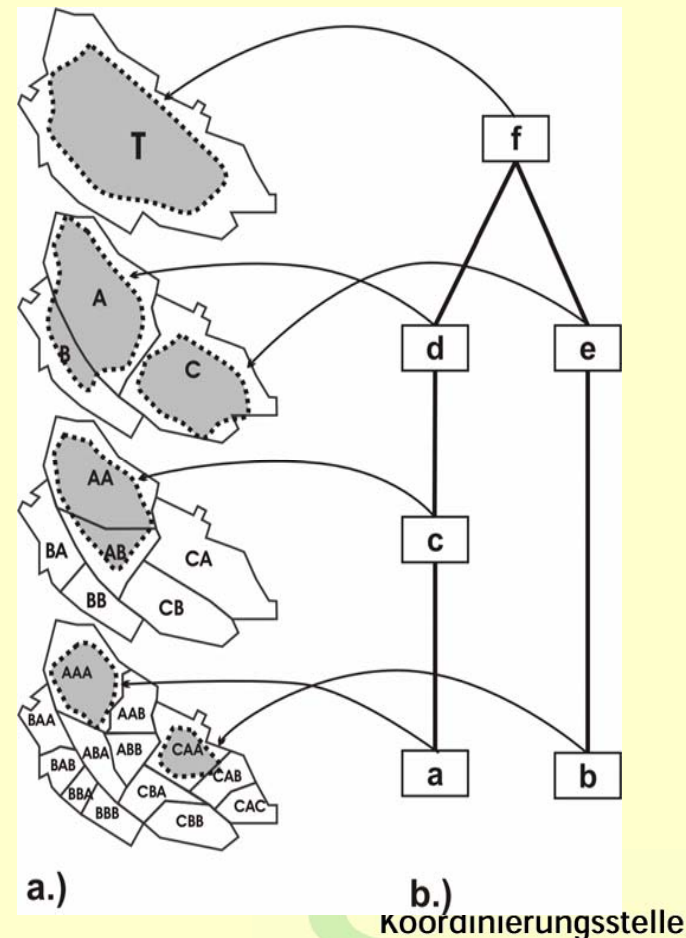
$$S_p = (\underline{S}_p, \bar{S}_p), \text{ where } \underline{S}_p \subseteq \bar{S}_p$$



$$\underline{L}(p) = \underline{S}_p = \begin{cases} \{r_i\} & : EQ(p, r_i) \\ \{\dots, r_i, \dots\} & : PP^{-1}(p, r_i) \\ \{\} & : PP(p, r_i) \vee PO(p, r_i) \vee DR(p, r_i) \end{cases}$$

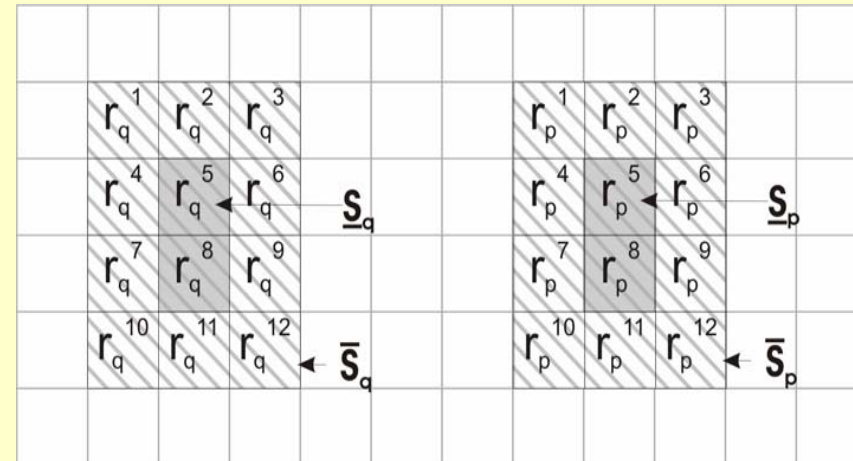
$$\bar{L}(p) = \bar{S}_p = \begin{cases} \{r_i\} & : EQ(p, r_i) \\ \{\dots, r_i, \dots\} & : PP^{-1}(p, r_i) \vee PP(p, r_i) \vee PO(p, r_i) \\ \{\} & : DR(p, r_i) \end{cases}$$

- ▶ Ortsnamensstruktur (PNS) = Konzeptionalisierung eines geografischen Raumes
- ▶ Definition eines Ortsnames in PNS:
 - Extensional auf Basis eines qualitative spatial reference model, oder
 - Intensional als part-of anderer Ortsnamen(regionen)
- ▶ Für jeden Ortsnamen können Extension bzw. Position in Partonomie berechnet werden
- ▶ Ziel: Ein intuitives und benutzerfreundliches Tool um geografische Räume zu modellieren



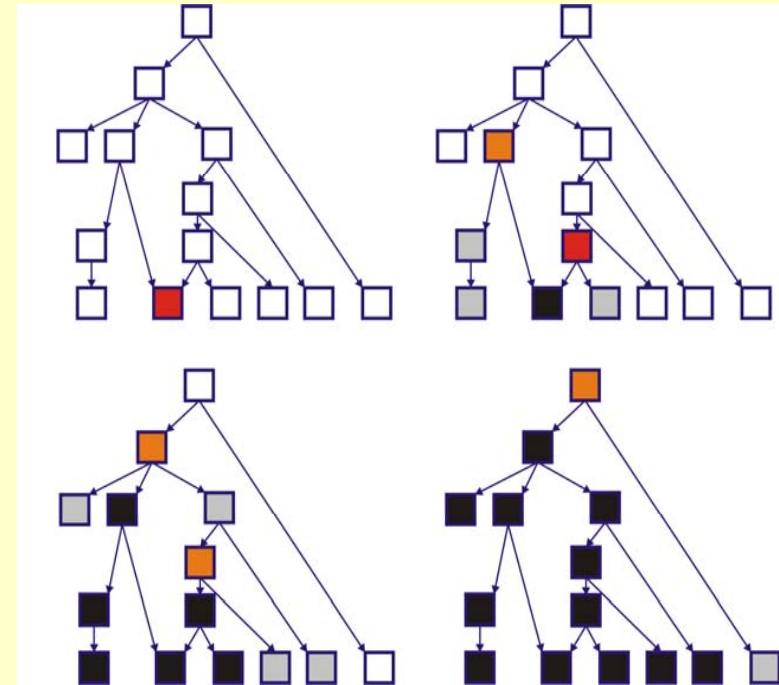
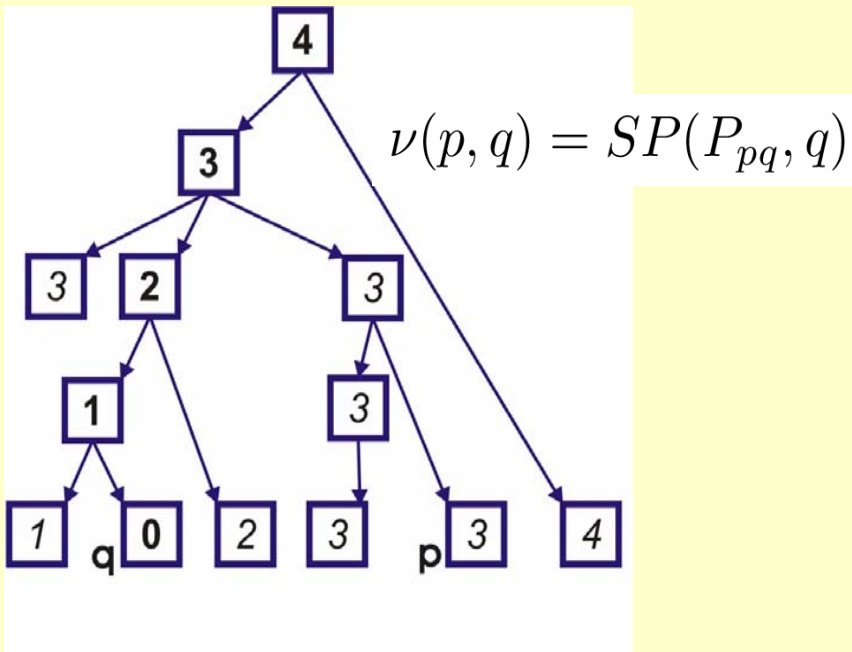
Berechnung der horizontale Distanz zweier Regionen

- Für jedes Element r_q in S_q wird eine „Distanzmatrix“ auf $pSRT$ berechnet
- Für jedes Element r_p in S_p
 - Integration der durch die Distanzmatrizen zugeordneten Werte (Default: arithm. Mittel)
- Für komplexe SPF:
 - Distanzmatrizen für obere/ untere Approximation von S_q
 - Separate Berechnung für obere/untere Approx. von S_p :
 - Integration obere/untere Approx.



$$\delta(p, q) = f(\delta(S_p, S_q))$$

$$\delta(p, q) = E(\delta(S_p \times S_q)) = E \begin{pmatrix} \delta(\bar{S}_p, \bar{S}_q) & \delta(\bar{S}_p, \underline{S}_q) \\ \delta(\underline{S}_p, \bar{S}_q) & \delta(\underline{S}_p, \underline{S}_q) \end{pmatrix}$$



- Berechnung analog zur vertikalen Distanz im räumlichen Referenzmodell

- ▶ Sequenzielle Berechnung für Ortsnamen mit mehreren Elternknoten