

Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ!

Geometrie unaccent FuzzyStrMatch

Fazit

Ortsuche

mit PostgreSQL und PostGIS

Stephan Wagner

ITos GmbH, CH-9642 Ebnat-Kappel

FOSSGIS 13. Juni 2013



Ziel

Ortsuche

Preamb

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm





Konzept

Ortsuche

Preamb

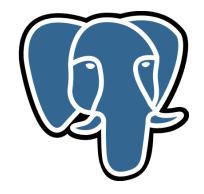
SQLpostgres Scharfe Suche

Suche

Geometrie

FuzzyStrMat Trigramm

Fazit



Hinweise an den Mahaut



Preambel Abhängigkeit Ortsbezeichner

Ortsuche

Preambel

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ

Geometrie unaccent

FuzzyStrMate Trigramm FTS

Fazit

Ortskoordinaten

- Koordinatensystem
- Einheiten

Ortsname

- Sprache
- Schreibweise
- Rechtschreibung
- Wie hei[ss,ß]t das Ding schon wieder



Preambel Abhängigkeit Ortsbezeichner

Ortsuche

Preambel

SQLpostgres
Scharfe Suche
Unscharfe

PostgreSQ

Geometrie unaccent FuzzyStrMate

Fazit

Ortskoordinaten

- Koordinatensystem
- Einheiten
- Ortsname
 - Sprache
 - Schreibweise
 - Rechtschreibung
 - Wie hei[ss,ß]t das Ding schon wieder?



Preambel Beispiel Ortsbezeichnung

Ortsuche

Preambel

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ

Geometrie unaccent FuzzyStrMato Trigramm

- Geographisch: (PostgreSQL oder PostGIS-Geometrietypen)
 - Koordinate Ortszentrum (Punkt)
 - Umhüllende des Ortes (Linie)
 - Ortsfläche
- Sprachlich:
 - Fribourg (französisch)
 - Frybùrg (alemannisch)
 - Friboua, Fribôrg (franko-provenzalisch
 - Freiburg, Freiburg U(e)chtland
 - Freiburg, Freiburg im Breisgau, Friburg im Brisgau (alemannisch)
- Stichworte: Nähe, Radius, Abstand, Nearest Neighbour,
 kNN-Gist-Index ⇒ Nach Abstand sortiert



Preambel Beispiel Ortsbezeichnung

Ortsuche

Preambel

SQLpostgres Scharfe Suche Unscharfe

PostgreSQ

Geometrie unaccent FuzzyStrMatc Trigramm

Fazit

 Geographisch: (PostgreSQL oder PostGIS-Geometrietypen)

- Koordinate Ortszentrum (Punkt)
- Umhüllende des Ortes (Linie)
- Ortsfläche
- Sprachlich:
 - Fribourg (französisch)
 - Frybùrg (alemannisch)
 - Friboua, Fribôrg (franko-provenzalisch)
 - Freiburg, Freiburg Ü(e)chtland
 - Freiburg, Freiburg im Breisgau, Friburg im Brisgau (alemannisch)
- Stichworte: Nähe, Radius, Abstand, Nearest Neighbour, kNN-Gist-Index ⇒ Nach Abstand sortiert



Ortsuche

Preambel

SQLpostgres Scharfe Suche Unscharfe

PostgreSQ

Geometrie unaccent FuzzyStrMatcl Trigramm

Fazit

Preambel Beispiel Ortsbezeichnung

- Geographisch: (PostgreSQL oder PostGIS-Geometrietypen)
 - Koordinate Ortszentrum (Punkt)
 - Umhüllende des Ortes (Linie)
 - Ortsfläche
- Sprachlich:
 - Fribourg (französisch)
 - Frybùrg (alemannisch)
 - Friboua, Fribôrg (franko-provenzalisch)
 - Freiburg, Freiburg Ü(e)chtland
 - Freiburg, Freiburg im Breisgau, Friburg im Brisgau (alemannisch)
- Stichworte: Nähe, Radius, Abstand, Nearest Neighbour, kNN-Gist-Index ⇒ Nach Abstand sortiert



Preambel Hinweise

Ortsuche

Preambel

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatc Trigramm

- Spracheinstellung und gewähltes Wörterbuch beeinflussen Suchergebnis
- ② Einbettung (PL/pgSQL-)Funktionen
- Autocomplete



Preambel Hinweise

Ortsuche

Preambel

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQI

Geometrie unaccent FuzzyStrMatch Trigramm

- Spracheinstellung und gewähltes Wörterbuch beeinflussen Suchergebnis
- Einbettung (PL/pgSQL-)Funktionen
- Autocomplete



Preambel Hinweise

Ortsuche

Preambel

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQI

Geometrie unaccent FuzzyStrMatc Trigramm

- Spracheinstellung und gewähltes Wörterbuch beeinflussen Suchergebnis
- Einbettung (PL/pgSQL-)Funktionen
- 4 Autocomplete



Preambel

PostgreSQL-Erweiterungen einbinden

Ortsuche

Preambel

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQL
Geometrie
unaccent
FuzzyStrMatch
Trigramm
FTS

- ls /usr/share/postgresql/9.2/extension oder SELECT * FROM pg_available_extensions listet bereits installierte Erweiterungen.
- CREATE EXTENSION 'ExtensionName' WITH SCHEMA 'Schema' VERSION 'Version' FROM 'AlteVersion'
- **Debian:** aptitude install postgresql-contrib-9.2
- Ansonsten kompilieren (besp. PostGIS 2.x)



Preambel Versionen

Ortsuche

Preambel

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQ!

Geometrie unaccent FuzzyStrMatcl Trigramm

- PostgreSQL 9.2
- PostGIS 2.0.3
- pg_trgm 1.0
- fuzzystrmatch 1.0
- unaccent 1.0



Preambel Referent

Ortsuche

Preambel

SQLpostgres Scharfe Suche Unscharfe

PostgreSQI

Geometrie unaccent FuzzyStrMatcl Trigramm

Fazit

Stephan Wagner

- Geschäftsführer ITos GmbH
- Web-(GIS)-Lösungen
- PostgreSQL Django (SQL, Python, HTML, CSS, JS)







Preambel Referent

Ortsuche

Preambel

SQLpostgres Scharfe Suche Unscharfe

PostgreSQI

Geometrie unaccent FuzzyStrMatch Trigramm

Fazit

Stephan Wagner

- Geschäftsführer ITos GmbH
- Web-(GIS)-Lösungen
- PostgreSQL Django (SQL, Python, HTML, CSS, JS)







Ortsuche

Preambel

Scharfe Suche

Preambel Referent

- Stephan Wagner
- Geschäftsführer ITos GmbH
- Web-(GIS)-Lösungen
- PostgreSQL Django (SQL, Python, HTML, CSS, JS)
- ⇒ System-Architektur und -Administration







Themen

Ortsuche

Preambel

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatcl Trigramm FTS

razit

Inhaltsverzeichnis

- Preambel
- 2 SQLpostgres
 - Scharfe Suche
 - Unscharfe Suche
- PostgreSQL
 - Geometrie
 - unaccent
 - FuzzyStrMatch
 - Trigramm
 - FTS
- Fazit



SQLpostgres Suche nach Spalteninhalt

Ortsuche

Preambe

SQLpostgres Scharfe Suche

Suche PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm FTS

- SELECT * FROM myTable;
- SELECT row2, row1 * 1.5 AS row1new FROM myTable;



SQLpostgres Einschränkende Suche: WHERE

Ortsuche

Preambe

SQLpostgres
Scharfe Suche
Unscharfe
Suche

 ${\sf PostgreSQ}$

Geometrie unaccent FuzzyStrMatch Trigramm FTS

_ .

- SELECT row1 FROM myTable WHERE row2 =
 'OrtsName';
- SELECT row1 FROM myTable WHERE row2 > 'OrtsName1' AND row2 < 'OrtsName2';
- SELECT row2 FROM myTable WHERE row1 < (SELECT ABS(AVG(row1) - row1) FROM myTable;



SQLpostgresSuche Beschleunigen: INDEX

Ortsuche

Preambe

SQLpostgres
Scharfe Suche

Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm

Fazit

Auf Suchspalten INDEX legen

- Ab PostgreSQL 9.1 GiST kNN-INDEX
- Ab PostgreSQL 9.3 SELECT nur auf INDEX möglich
- Beachte:
 - Planer-Ausgabe analysieren (EXPLAIN ANALYSE
 - impliziter SOR1
 - RAM f
 ür INDEX, SORT (PG tuning



Ortsuche

Preambe

SQLpostgres
Scharfe Suche

Unscharfe Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm

razıt

Auf Suchspalten INDEX legen

- Ab PostgreSQL 9.1 GiST kNN-INDEX
- Ab PostgreSQL 9.3 SELECT nur auf INDEX möglich
- Beachte:

SQLpostgres

Suche Beschleunigen: INDEX

- Planer-Ausgabe analysieren (EXPLAIN ANALYSE
- impliziter SORT
- RAM für INDEX, SORT (PG tuning)



SQLpostgresSuche Beschleunigen: INDEX

Ortsuche

Preambe

SQLpostgres
Scharfe Suche

Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm

- Auf Suchspalten INDEX legen
- Ab PostgreSQL 9.1 GiST kNN-INDEX
- Ab PostgreSQL 9.3 SELECT nur auf INDEX möglich
- Beachte:
 - Planer-Ausgabe analysieren (EXPLAIN ANALYSE
 - impliziter SOR1
 - RAM f
 ür INDEX, SORT (PG tuning)



SQLpostgres Suche Beschleunigen: INDEX

Ortsuche

Preambe

SQLpostgres Scharfe Suche

Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm

- Auf Suchspalten INDEX legen
- Ab PostgreSQL 9.1 GiST kNN-INDEX
- Ab PostgreSQL 9.3 SELECT nur auf INDEX möglich
- Beachte:
 - Planer-Ausgabe analysieren (EXPLAIN ANALYSE)
 - impliziter SORT
 - RAM für INDEX, SORT (PG tuning)



Ortsuche

Preambe

SQLpostgres Scharfe Suche

Jucile

PostgreSQ

Geometrie unaccent FuzzyStrMatcl Trigramm

Fazit

Auf Suchspalten INDEX legen

- Ab PostgreSQL 9.1 GiST kNN-INDEX
- Ab PostgreSQL 9.3 SELECT nur auf INDEX möglich
- Beachte:

SQLpostgres

Suche Beschleunigen: INDEX

- Planer-Ausgabe analysieren (EXPLAIN ANALYSE)
- impliziter SORT
- RAM für INDEX, SORT (PG tuning)



Ortsuche

Preambe

SQLpostgres
Scharfe Suche
Unscharfe

Unscharfe Suche

Geometrie unaccent FuzzyStrMatch Trigramm

Fazit

SQLpostgres Einschränkende Suche: CASE WHEN

SELECT CASE WHEN row1='OrtNameFR' THEN 'OrtNameDE'
WHEN row2='OrtNameDE' THEN 'OrtNameDE

WHEN row2='OrtNameDE' THEN 'OrtNameDE' ELSE NONE

END

FROM myTable;



SQLpostgres Minimal Unscharf: (I)LIKE

Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

Geometrie unaccent FuzzyStrMatch Trigramm FTS

115

Mustersuche auf Begriff:

- SELECT * FROM myTable WHERE row2 LIKE
 'Ortsna%';
- SELECT row1 FROM myTable WHERE row2 ILIKE '%rtsNa%';
- SELECT row1 FROM myTable WHERE lower(row2)
 LIKE '%rtsna%';



SQLpostgres Minimal unscharf: SIMILAR TO

Ortsuche

Unscharfe Suche

Mustersuche auf ganze Zeichenkette:

- SELECT * FROM myTable WHERE row1 SIMLAR TO 'Ortsna%':
- SELECT * FROM myTable WHERE row2 SIMLAR TO '(Ortlort)%':
- SELECT * FROM myTable WHERE row2 SIMLAR TO ',[0,o]rt%';



Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ

Geometrie unaccent FuzzyStrMato Trigramm FTS

Fazit

SQLpostgresMinimal unscharf: Reguläre Ausdrücke (POSIX)

Vergleichsoperator	Beschreibung
!	Negation
\sim	keysensitiver Vergleich
$\sim *$	key in sensitiver Vergleich

Suche in etwa 'Karlstrasse', Strasse in der Schreibweise B:

SELECT * FROM myTable WHERE row2
'[C,c,K,k]arl+[ß]+\$';



PostgreSQL Über SQL hinaus

Ortsuche

Preambe

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatcl Trigramm

Fazit

SQLpostgres

PostgreSQL



GIS

Ortsuche

Preamb

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQ

Geometrie

FuzzyStrMatch Trigramm FTS

Fazit

Koordinatensystem

Einheiten

Koordinaten-Distanz

- PostgreSQL:
 - Punkt (point), Gerade (lseg), Linie (path), Polygon (polygon), Rechteck (box), Kreis (circle)
- PostGIS:
 - GiST-kNN-Index mit <-> als Distanzoperator
 - st_setsrid(st_makepoint(-90,40),4326) LIMIT 1
 - ullet \Rightarrow Die 10 zum Testpunkt nächstgelegenen Punkte



Koordinaten-Distanz

GIS

Ortsuche

Preamb

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQ

Geometrie

FuzzyStrMatc Trigramm FTS

razıt

Koordinatensystem

- Einheiten
- PostgreSQL:
 - Punkt (point), Gerade (lseg), Linie (path), Polygon (polygon), Rechteck (box), Kreis (circle)
- PostGIS:
 - GiST-kNN-Index mit <-> als Distanzoperator
 - SELECT ort FROM ortsnamen ORDER BY geom <-> st_setsrid(st_makepoint(-90,40),4326) LIMIT 10;
 - ullet \Rightarrow Die 10 zum Testpunkt nächstgelegenen Punkte



unaccent unaccent

Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQL Geometrie

unaccent FuzzyStrMatch Trigramm FTS

Fazit

Extension aus dem Umfeld von FTS

- Datei mit Konvertierungsregeln unter \$SHAREDIR/tsearch_data/
- Filter, welcher Lexeme (~ Begriffe) von Akzenten befreit
- Abfrage: SELECT unaccent('Freiburg im Üechtland');
 - ⇒ Freiburg im Uechtland



unaccent

Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQI

unaccent FuzzyStrMatch Trigramm FTS

Fazit

Extension aus dem Umfeld von FTS

- Datei mit Konvertierungsregeln unter \$SHAREDIR/tsearch_data/
- Filter, welcher Lexeme (~ Begriffe) von Akzenten befreit
- Abfrage: SELECT unaccent('Freiburg im Üechtland');
 - ⇒ Freiburg im Uechtland



FuzzyStrMatch PG-Extension

Ortsuche

Preambo

SQLpostgres
Scharfe Suche

Suche PostgreSQ

Geometrie

FuzzyStrMatch Trigramm

- Soundex
- (Double-)Metaphone
- Levenshtein



FuzzyStrMatch

Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

Geometrie

FuzzyStrMatch Trigramm

- Algorithmus: Robert Russell und Margaret Odel, 1918
- Phonetische Indizierung nach englischsprachigem Klangmuster
- Ergibt auch für Deutsch und Französich brauchbare Ergebnisse
- Code: Erster Buchstabe plus drei Ziffern, bsp. W-213 für Wikipedia.
 - Wort mit mehr Buchstaben als benötigt ⇒ Schnittt
 - Wort mit wenig Buchstaben als benötigt ⇒ Nullen
- SELECT SOUNDEX('Fribourg') ⇒ F616
- 'Fryburg' ⇒ F616 ← 'Freiburg im Üechtland'
- SELECT * FROM ort WHERE difference(ort.nm, 'Friburg') > 2;
- Britney Spears \Rightarrow B635 S162 \Leftarrow Bewährten Superzicke \odot



FuzzyStrMatch

Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQL Geometrie

FuzzyStrMatch Trigramm

Fazi

- Algorithmus: Robert Russell und Margaret Odel, 1918
- Phonetische Indizierung nach englischsprachigem Klangmuster
- Ergibt auch für Deutsch und Französich brauchbare Ergebnisse
- Code: Erster Buchstabe plus drei Ziffern, bsp. W-213 für Wikipedia.
 - Wort mit mehr Buchstaben als benötigt ⇒ Schnitt
 - ullet Wort mit wenig Buchstaben als benötigt \Rightarrow Nullen
- SELECT SOUNDEX('Fribourg') ⇒ F616
- 'Fryburg' ⇒ F616 ← 'Freiburg im Üechtland'
- SELECT * FROM ort WHERE difference(ort.nm,
 'Friburg') > 2;
- Britney Spears \Rightarrow B635 S162 \Leftarrow Bewährten Superzicke \odot



FuzzyStrMatch

Ortsuche

Preambe

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQL Geometrie

FuzzyStrMatch Trigramm

- Algorithmus: Robert Russell und Margaret Odel, 1918
- Phonetische Indizierung nach englischsprachigem Klangmuster
- Ergibt auch für Deutsch und Französich brauchbare Ergebnisse
- Code: Erster Buchstabe plus drei Ziffern, bsp. W-213 für Wikipedia.
 - Wort mit mehr Buchstaben als benötigt ⇒ Schnitt
 - ullet Wort mit wenig Buchstaben als benötigt \Rightarrow Nullen
- SELECT SOUNDEX('Fribourg') ⇒ F616
- 'Fryburg' ⇒ F616 ← 'Freiburg im Üechtland'
 - SELECT * FROM ort WHERE difference(ort.nm, 'Friburg') > 2;
- Britney Spears \Rightarrow B635 S162 \Leftarrow Bewährten Superzicke \odot



FuzzyStrMatch

Ortsuche

Preambo

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQL Geometrie

FuzzyStrMatch Trigramm

- Algorithmus: Robert Russell und Margaret Odel, 1918
- Phonetische Indizierung nach englischsprachigem Klangmuster
- Ergibt auch für Deutsch und Französich brauchbare Ergebnisse
- Code: Erster Buchstabe plus drei Ziffern, bsp. W-213 für Wikipedia.
 - Wort mit mehr Buchstaben als benötigt ⇒ Schnitt
 - ullet Wort mit wenig Buchstaben als benötigt \Rightarrow Nullen
- SELECT SOUNDEX('Fribourg') ⇒ F616
- 'Fryburg' ⇒ F616 ← 'Freiburg im Üechtland'
- SELECT * FROM ort WHERE difference(ort.nm, 'Friburg') > 2;
- Britney Spears \Rightarrow B635 S162 \Leftarrow Bewährten Superzicke \odot



Ortsuche

Preambe

SQLpostgres
Scharfe Suche

PostgreSQ

Geometrie

FuzzyStrMatch

FTS

Fazıt

- Algorithmus: Lawrence Phillips, 2000
- Phonetische Indizierung nach englischsprachigem Klangmuster
- Absichtlich phonetisch unscharfe Darstellung (Sprechweisen)
- Unterscheidet feingranularer als SOUNDEX
- Gewichtung der Operationen möglich
- SELECT metaphone('Fryburg im Üechtland', 7); ⇒
 FRBRKMX"
- SELECT dmetaphone('Fryburg im Üechtland'); ⇒
 FRPR"



Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe

PostgreSQ

unaccent

FuzzyStrMatch Trigramm

Fazit

- Phonetische Indizierung nach englischsprachigem Klangmuster
- Absichtlich phonetisch unscharfe Darstellung (Sprechweisen)
- Unterscheidet feingranularer als SOUNDEX
- Gewichtung der Operationen möglich
- SELECT metaphone('Fryburg im Üechtland', 7); ⇒
 FRBRKMX"
- SELECT dmetaphone('Fryburg im Üechtland'); ⇒
 FRPR"



Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe

PostgreSQ

Geometrie

FuzzyStrMatch Trigramm

Fazit

- Phonetische Indizierung nach englischsprachigem Klangmuster
- Absichtlich phonetisch unscharfe Darstellung (Sprechweisen)
- Unterscheidet feingranularer als SOUNDEX
- Gewichtung der Operationen möglich
- SELECT metaphone('Fryburg im Üechtland', 7); ⇒
 FRBRKMX"
- SELECT dmetaphone('Fryburg im Üechtland'); ⇒
 FRPR"



Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe

PostgreSQ

Geometrie

FuzzyStrMatch Trigramm

Fazit

- Phonetische Indizierung nach englischsprachigem Klangmuster
- Absichtlich phonetisch unscharfe Darstellung (Sprechweisen)
- Unterscheidet feingranularer als SOUNDEX
- Gewichtung der Operationen möglich
- SELECT metaphone('Fryburg im Üechtland', 7); ⇒
 FRBRKMX"
- SELECT dmetaphone('Fryburg im Üechtland'); ⇒
 FRPR"



(Double-)Mataphone

FuzzyStrMatch

Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe

PostgreSQ

unaccent FuzzyStrMatch

Trigramm FTS

Fazit

- Phonetische Indizierung nach englischsprachigem Klangmuster
- Absichtlich phonetisch unscharfe Darstellung (Sprechweisen)
- Unterscheidet feingranularer als SOUNDEX
- Gewichtung der Operationen möglich
- SELECT metaphone('Fryburg im Üechtland', 7); ⇒ FRBRKMX"
- SELECT dmetaphone('Fryburg im Üechtland'); ⇒
 FRPR"



Ortsuche

Preambo

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQI Geometrie

FuzzyStrMatch Trigramm

Fazit

• Algorithmus: Wladimir Levenshtein, 1965

- **Distanz:** Anzahl von Einzeloperationen, um Zeichenkette 1 auf 2 umzuwandeln
- Haupteinsatzbereich: Duplikatserkennung, Rechtschreibeprüfung
- Bei grössen Datensätzen langsam: O(NxM)
- SELECT levenshtein('Fribourg', 'Freibourg'); ⇒
 1
- SELECT levenshtein('Fribourg', 'Fryburg im Üechtland'); ⇒ 15



Ortsuche

Preamb

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQI Geometrie

FuzzyStrMatch Trigramm

Fazit

• Algorithmus: Wladimir Levenshtein, 1965

- **Distanz:** Anzahl von Einzeloperationen, um Zeichenkette 1 auf 2 umzuwandeln
- Haupteinsatzbereich: Duplikatserkennung, Rechtschreibeprüfung
- Bei grössen Datensätzen langsam: O(NxM)
- SELECT levenshtein('Fribourg', 'Freibourg'); ⇒
 1
- SELECT levenshtein('Fribourg', 'Fryburg im Üechtland'); ⇒ 15



Ortsuche

Preambo

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQI Geometrie

FuzzyStrMatch Trigramm FTS

- Algorithmus: Wladimir Levenshtein, 1965
- **Distanz:** Anzahl von Einzeloperationen, um Zeichenkette 1 auf 2 umzuwandeln
- Haupteinsatzbereich: Duplikatserkennung, Rechtschreibeprüfung
- Bei grössen Datensätzen langsam: O(NxM)
- SELECT levenshtein('Fribourg', 'Freibourg'); ⇒
 1
- SELECT levenshtein('Fribourg', 'Fryburg im Üechtland'); ⇒ 15



Ortsuche

Preamb

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQI Geometrie

FuzzyStrMatch Trigramm

- Algorithmus: Wladimir Levenshtein, 1965
- **Distanz:** Anzahl von Einzeloperationen, um Zeichenkette 1 auf 2 umzuwandeln
- Haupteinsatzbereich: Duplikatserkennung, Rechtschreibeprüfung
- Bei grössen Datensätzen langsam: O(NxM)
- SELECT levenshtein('Fribourg', 'Freibourg'); ⇒
 1
- SELECT levenshtein('Fribourg', 'Fryburg im Üechtland'): ⇒ 15



Ortsuche

Preamb

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQI Geometrie

FuzzyStrMatch Trigramm FTS

razıt

- Algorithmus: Wladimir Levenshtein, 1965
- **Distanz:** Anzahl von Einzeloperationen, um Zeichenkette 1 auf 2 umzuwandeln
- Haupteinsatzbereich: Duplikatserkennung, Rechtschreibeprüfung
- Bei grössen Datensätzen langsam: O(NxM)
- SELECT levenshtein('Fribourg', 'Freibourg'); ⇒
 1
- SELECT levenshtein('Fribourg', 'Fryburg im Üechtland'); $\Rightarrow 15$



Trigramm Begriffe

Ortsuche

Preamb

SQLpostgres Scharfe Suche Unscharfe Suche

Geometrie
unaccent
FuzzyStrMatch

Trigramm FTS

Fazit

Module: pg_trgm

• **Trigramm:** Zerlegung eines Textes in 3-Zeichen-Fragmente (n-Gramm)

 Trigramme: Zur Weissagung dienende, altchinesiche Symbole

• **Dice-Koeffizient** d: Anteil übereinstimmender N-Gramme zweier Terme $(0 \le d \le 1)$



Trigramm Begriffe

Ortsuche

Preamb

SQLpostgres
Scharfe Suche
Unscharfe
Suche

Geometrie unaccent FuzzyStrMatch

Trigramm FTS

Fazit

Module: pg_trgm

• **Trigramm:** Zerlegung eines Textes in 3-Zeichen-Fragmente (n-Gramm)

 Trigramme: Zur Weissagung dienende, altchinesiche Symbole

• **Dice-Koeffizient** d: Anteil übereinstimmender N-Gramme zweier Terme $(0 \le d \le 1)$



Trigramm Trigramm-Dice

Ortsuche

Preamb

SQLpostgres Scharfe Suche Unscharfe

Postaro\$0

Geometrie

unaccent FuzzyStrMa

Trigramm

Fazit

• $d(a,b) = \frac{2|T(a) \cap T(b)|}{|T(a)| + |T(b)|}$

Beispiel:

Term a = "wirk", Term b = "work"

• $T(a) = \S\S w, \S wi, wir, irk, rk\S, k\S\S$

• $T(b) = \S\S w, \S wo, wor, ork, rk\S, k\S\S$

• $T(a) \cap T(b) = \S\S w, k\S\S, rk\S$

• $d(wirk, work) = \frac{2 \cdot 3}{6+6} = \frac{1}{2}$

d.h., Die Ähnlichkeit oder Distanz beträgt 0.5



Ortsuche

Preambo

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ!

Geometrie unaccent FuzzyStrMatch Trigramm FTS

Fazit

Seit PostgreSQL 8.3 fester Bestandteil

- Resultatmenge: Dokumente, welche mit dem Suchbegriff möglichst nah Übereinstimmen
- Basis-Funktionalität:
 - ① Tokens auf Lexeme reduziert, Stopwords gefiltert: Bréil sur Royas ⇒ breil roya (Datentyp tsvector oder tsvector)
 - @ GIN oder kNN-GiST-Index auf Spalte Lexem
 - ③ Suchanfrage mit to_tsquery() und dem Matching-Operator &&
 - 4 Ausgabe des Suchergebnisses nach Relevanz



Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQL
Geometrie
unaccent
FuzzyStrMatch
Trigramm

FTS

Seit PostgreSQL 8.3 fester Bestandteil

- Resultatmenge: Dokumente, welche mit dem Suchbegriff möglichst nah Übereinstimmen
- Basis-Funktionalität:
 - ① Tokens auf Lexeme reduziert, Stopwords gefiltert: Bréil sur Royas ⇒ breil roya (Datentyp tsvector oder tsvector)
 - Q GIN oder kNN-GiST-Index auf Spalte Lexem
 - ③ Suchanfrage mit to_tsquery() und dem Matching-Operator &&
 - 4 Ausgabe des Suchergebnisses nach Relevanz



Ortsuche

Preamb

SQLpostgres Scharfe Suche Unscharfe Suche

Geometrie unaccent FuzzyStrMatch Trigramm

FTS

• Seit PostgreSQL 8.3 fester Bestandteil

- Resultatmenge: Dokumente, welche mit dem Suchbegriff möglichst nah Übereinstimmen
- Basis-Funktionalität:
 - ① Tokens auf Lexeme reduziert, Stopwords gefiltert: Bréil sur Royas ⇒ breil roya (Datentyp tsvector oder tsvector)
 - 2 GIN oder kNN-GiST-Index auf Spalte Lexem
 - 3 Suchanfrage mit to_tsquery() und dem Matching-Operator
 - 4 Ausgabe des Suchergebnisses nach Relevanz



Ortsuche

Preamb

SQLpostgres Scharfe Suche Unscharfe Suche

Geometrie unaccent FuzzyStrMatch Trigramm

Fazit

• Seit PostgreSQL 8.3 fester Bestandteil

- Resultatmenge: Dokumente, welche mit dem Suchbegriff möglichst nah Übereinstimmen
- Basis-Funktionalität:
 - ① Tokens auf Lexeme reduziert, Stopwords gefiltert: Bréil sur Royas ⇒ breil roya (Datentyp tsvector oder tsvector)
 - 2 GIN oder kNN-GiST-Index auf Spalte Lexem
 - ③ Suchanfrage mit to_tsquery() und dem Matching-Operator &&
 - 4 Ausgabe des Suchergebnisses nach Relevanz



Ortsuche

Preamb

SQLpostgres Scharfe Suche Unscharfe Suche

Geometrie unaccent FuzzyStrMatch Trigramm

FTS

• Seit PostgreSQL 8.3 fester Bestandteil

- Resultatmenge: Dokumente, welche mit dem Suchbegriff möglichst nah Übereinstimmen
- Basis-Funktionalität:
 - ① Tokens auf Lexeme reduziert, Stopwords gefiltert: Bréil sur Royas ⇒ breil roya (Datentyp tsvector oder tsvector)
 - 2 GIN oder kNN-GiST-Index auf Spalte Lexem
 - Suchanfrage mit to_tsquery() und dem Matching-Operator &&
 - 4 Ausgabe des Suchergebnisses nach Relevanz



FTS Sprachspezifische Wörterbücher

Ortsuche

Preambo

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ!

Geometrie unaccent FuzzyStrMatch Trigramm FTS

Fazit

 Simple, Synomym, Thesaurus, Ispell, Snowball (Stemming-Algorithmus)

- Is /usr/share/postgresql/9.2/tsearch_data
- CREATE TEXT SEARCH DISCTIONARY ...
- Synonym: Ur-Begriff ⊔ Synonym
- Thesaurus: Ur-Satz : Synonym-Satz



FTS Sprachspezifische Wörterbücher

Ortsuche

Preamb

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm

FTS

 Simple, Synomym, Thesaurus, Ispell, Snowball (Stemming-Algorithmus)

- Is /usr/share/postgresql/9.2/tsearch_data
- CREATE TEXT SEARCH DISCTIONARY ...
- **Synonym:** Ur-Begriff ⊔ Synonym
- Thesaurus: Ur-Satz : Synonym-Satz



FTS Sprachspezifische Wörterbücher

Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ|

Geometrie unaccent FuzzyStrMatch Trigramm

FTS

 Simple, Synomym, Thesaurus, Ispell, Snowball (Stemming-Algorithmus)

Is /usr/share/postgresql/9.2/tsearch_data

• CREATE TEXT SEARCH DISCTIONARY ...

• **Synonym:** Ur-Begriff ⊔ Synonym

• Thesaurus: Ur-Satz : Synonym-Satz



Ortsuche

Preambe

SQLpostgres Scharfe Suche Unscharfe Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm

Fazit

Distanz verbindet Ortsnamensuche und Ortskoordinatensuche

- Ortsnamensuche mit Standard-SQL quasi statisch
- Opening PostgreSQL bietet Alternativen:
 - Soundex, Metaphone
 - Levenshtein
 - Trigramm
 - FTS
 - unaccen



Ortsuche

Preambe

- SQLpostgres
 Scharfe Suche
 Unscharfe
 Suche
- PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm

- Distanz verbindet Ortsnamensuche und Ortskoordinatensuche
- Ortsnamensuche mit Standard-SQL quasi statisch
- Opening PostgreSQL bietet Alternativen:
 - Soundex, Metaphone
 - Levenshtein
 - Trigramm
 - FTS
 - unaccen



Ortsuche

Preambe

- SQLpostgres
 Scharfe Suche
 Unscharfe
 Suche
- **PostgreSQ**

Geometrie unaccent FuzzyStrMatch Trigramm

- Distanz verbindet Ortsnamensuche und Ortskoordinatensuche
- 2 Ortsnamensuche mit Standard-SQL quasi statisch
- OstgreSQL bietet Alternativen:
 - Soundex, Metaphone
 - Levenshtein
 - Trigramm
 - FTS
 - unaccent



Ortsuche

Preamb

- SQLpostgres Scharfe Suche Unscharfe Suche
- **PostgreSQ**

Geometrie unaccent FuzzyStrMatch Trigramm

- Distanz verbindet Ortsnamensuche und Ortskoordinatensuche
- Ortsnamensuche mit Standard-SQL quasi statisch
- OstgreSQL bietet Alternativen:
 - Soundex, Metaphone
 - Levenshtein
 - Trigramm
 - FTS
 - unaccent



Ortsuche

Fazit

Preambe

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQ

Geometrie unaccent FuzzyStrMatch Trigramm FTS

- Distanz verbindet Ortsnamensuche und Ortskoordinatensuche
- 2 Ortsnamensuche mit Standard-SQL quasi statisch
- OstgreSQL bietet Alternativen:
 - Soundex, Metaphone
 - Levenshtein
 - Trigramm
 - FTS
 - unaccent



Ortsuche

Preambe

- SQLpostgres
 Scharfe Suche
 Unscharfe
 Suche
- **PostgreSQ**

Geometrie unaccent FuzzyStrMatch Trigramm

- Distanz verbindet Ortsnamensuche und Ortskoordinatensuche
- 2 Ortsnamensuche mit Standard-SQL quasi statisch
- OstgreSQL bietet Alternativen:
 - Soundex, Metaphone
 - Levenshtein
 - Trigramm
 - FTS
 - unaccent



Ortsuche

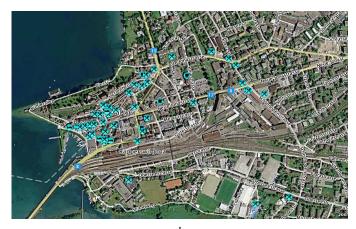
Preambo

SQLpostgres
Scharfe Suche
Unscharfe
Suche

PostgreSQ

Geometrie

Trigramm FTS



oder FOSSGIS e.V. Sektempfang