Machine Learning supported Integration of heterogenous Database Eco-Systems

Leonard Traeger
(TH Köln 11110185)
(UMBC CZ05211)

01.04.2022

Master thesis presentation
Summer 2022
Prof. Dr. Behrend
Prof. Dr. Karabatis

Technology
Arts Sciences
TH Köln

UMBC

Content

1. Problem Statement
2. Related Work
3. Inteplato
4. Performance Evaluation
5. Conclusion
6. Limitations and Future Work

Problem Statement

"Just because different pieces of data share the same syntax does not mean that they are automatically interoperable." - Pollock and Hodgson (2004)

Data Discovery

- Growth of data
- Domain-independent consolidation
- Heterogeneous architectures
- Limited human resources

Technical

- Schematic and semantic conflicts
- Scalable algorithms $O(N^N) \rightarrow O(2N)$
- Human-in-the-loop

Related Work in DI: schema mapping

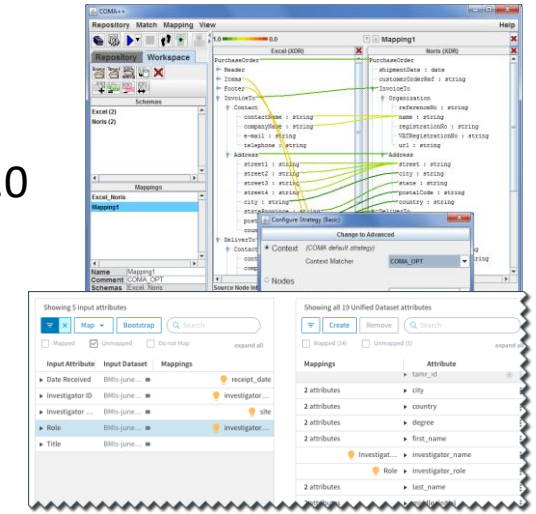
Driven by DWH, MDM, Semantic Web and Big Data Vs

Machine Learning

- Labeled data through unsupervised methods
- Instance transformation capabilities
- Meta / instance multi-dimensional representations
- Learn schema mapping, duplicate detection, and data fusion jointly
- Neural network scepsis

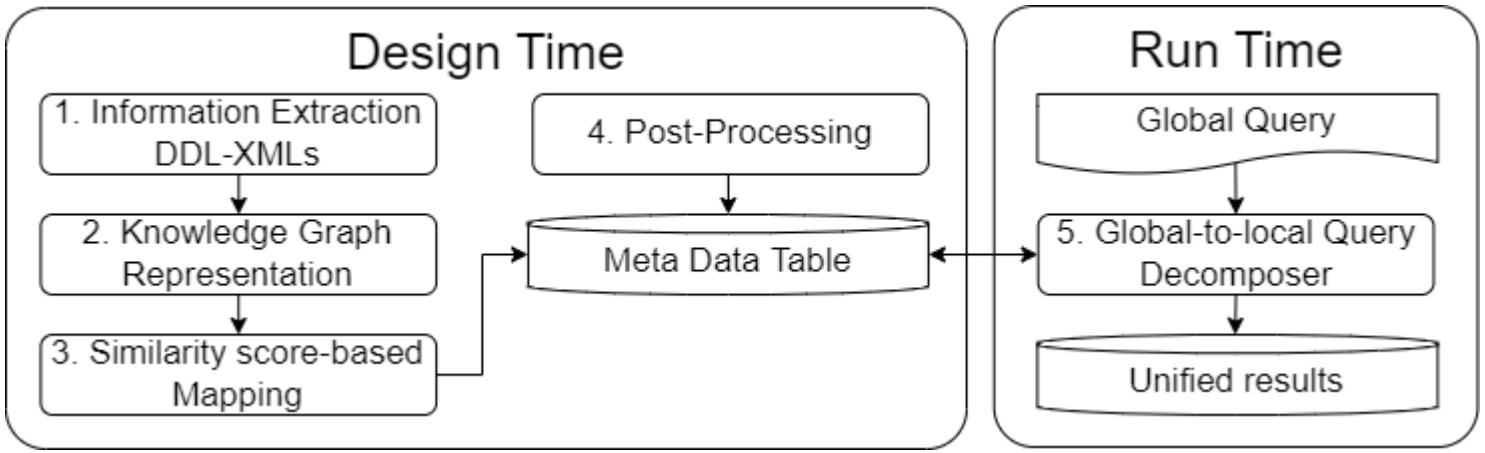
Existing Applications

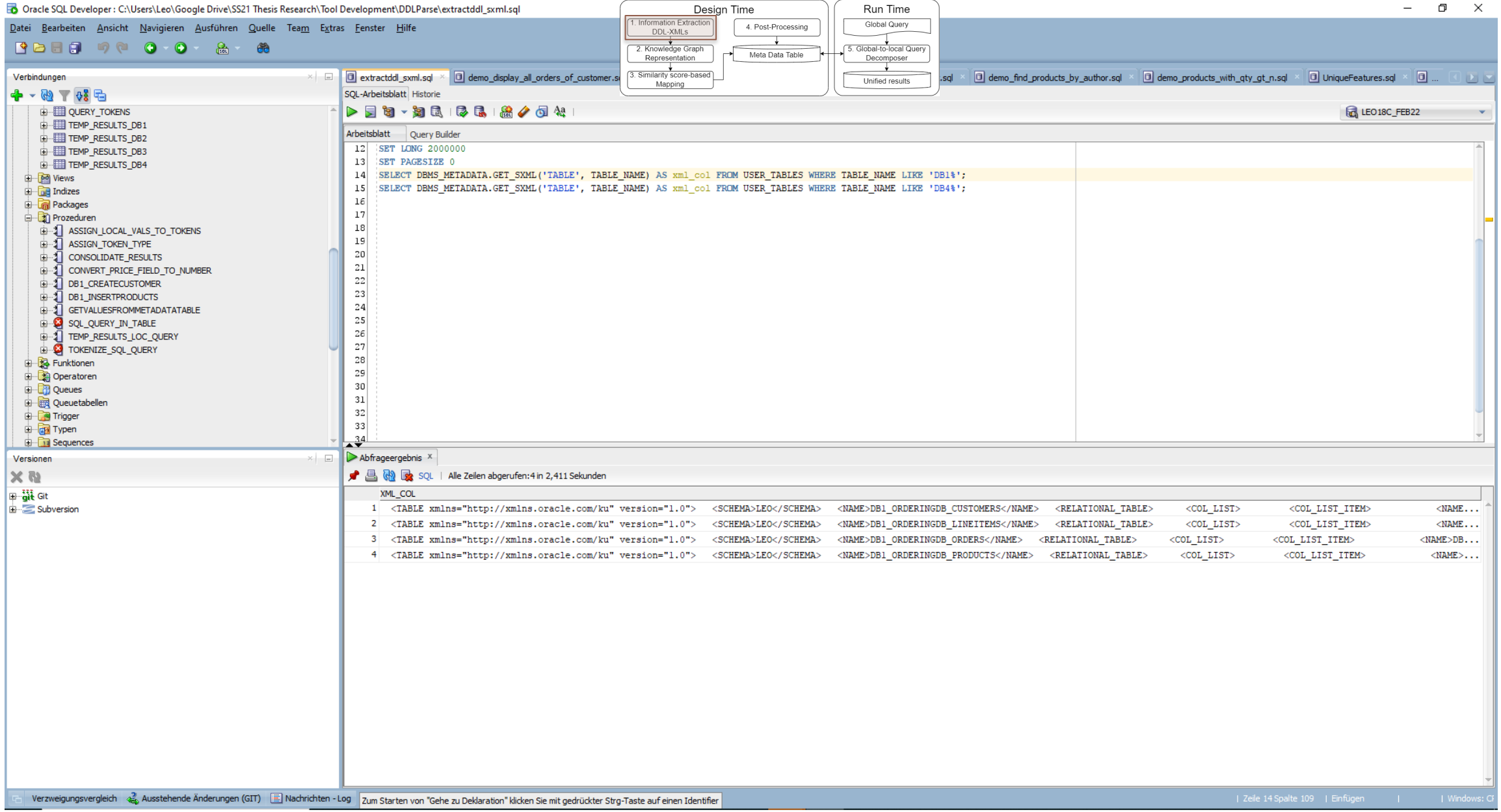
- Yearly benchmark challenges drive new algorithms but **less applications**
- COMA 3.0
- Tamr

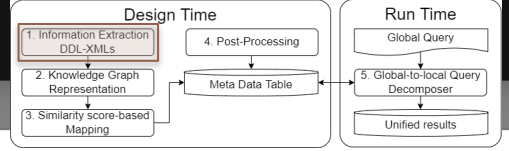


Map schemas with Inteplato

*A **web-based application** mapping **multiple schemas** represented in a **knowledge graph**.*
*The framework provides **explainable recommendations** through a **similarity scoring-system***
*and an **interactive human-in-the-loop** processing with linkage to a **global-to-local-query decomposer**.*




Oracle SQL Developer : C:\Users\Leo\Google Drive\SS21 Thesis Research\Tool Development\DDLParse\extractddl_sxml.sql
Datei Bearbeiten Ansicht Navigieren Ausführen Quelle Team Extras Fenster Hilfe
Design Time
1. Information Extraction DDL-XMLs
2. Knowledge Graph Representation
3. Similarity score-based Mapping
4. Post-Processing
Meta Data Table
Run Time
Global Query
5. Global-to-local Query Decomposer
Unified results
Verbindungen
QUERY_TOKENS
TEMP_RESULTS_DB1
TEMP_RESULTS_DB2
TEMP_RESULTS_DB3
TEMP_RESULTS_DB4
Views
Indizes
Packages
Prozeduren
ASSIGN_LOCAL_VALS_TO_TOKENS
ASSIGN_TOKEN_TYPE
CONSOLIDATE_RESULTS
CONVERT_PRICE_FIELD_TO_NUMBER
DB1_CREATECUSTOMER
DB1_INSERTPRODUCTS
GETVALUESFROMMETADATATABLE
SQL_QUERY_IN_TABLE
TEMP_RESULTS_LOC_QUERY
TOKENIZE_SQL_QUERY
Funktionen
Operatoren
Queues
Queuetabellen
Trigger
Typen
Sequences
Versionen
Git
Subversion
extractddl_sxml.sql
demo_find_products_by_author.sql
demo_products_with_qty_gt_n.sql
UniqueFeatures.sql
SQL-Arbeitsblatt
Historie
LEO18C_FEB22
Arbeitsblatt
Query Builder
SET LONG 2000000
SET PAGESIZE 0
SELECT DBMS_METADATA.GET_SXML('TABLE', TABLE_NAME) AS xml_col FROM USER_TABLES WHERE TABLE_NAME LIKE 'DB1%';
SELECT DBMS_METADATA.GET_SXML('TABLE', TABLE_NAME) AS xml_col FROM USER_TABLES WHERE TABLE_NAME LIKE 'DB4%';
Abfrageergebnis
SQL | Alle Zeilen abgerufen: 4 in 2,411 Sekunden
XML_COL
1 <TABLE xmlns="http://xmlns.oracle.com/ku" version="1.0"> <SCHEMA>LEO</SCHEMA> <NAME>DB1_ORDERINGDB_CUSTOMERS</NAME> <RELATIONAL_TABLE> <COL_LIST> <COL_LIST_ITEM> <NAME...
2 <TABLE xmlns="http://xmlns.oracle.com/ku" version="1.0"> <SCHEMA>LEO</SCHEMA> <NAME>DB1_ORDERINGDB_LINEITEMS</NAME> <RELATIONAL_TABLE> <COL_LIST> <COL_LIST_ITEM> <NAME...
3 <TABLE xmlns="http://xmlns.oracle.com/ku" version="1.0"> <SCHEMA>LEO</SCHEMA> <NAME>DB1_ORDERINGDB_ORDERS</NAME> <RELATIONAL_TABLE> <COL_LIST> <COL_LIST_ITEM> <NAME>DB...
4 <TABLE xmlns="http://xmlns.oracle.com/ku" version="1.0"> <SCHEMA>LEO</SCHEMA> <NAME>DB1_ORDERINGDB_PRODUCTS</NAME> <RELATIONAL_TABLE> <COL_LIST> <COL_LIST_ITEM> <NAME>...
Verzweigungsvergleich
Ausstehende Änderungen (GIT)
Nachrichten - Log
Zum Starten von "Gehe zu Deklaration" klicken Sie mit gedrückter Strg-Taste auf einen Identifier
Zeile 14 Spalte 109
Einfügen



Inteplato

Sign out

SCHEMA UNIVERSE OPTIONS

Upload new schema

Delete schema

GLOBAL SCHEMA OPTIONS

New global table

New global attribute

CONCEPT FILTER OPTIONS

Tables

Attributes

PATH FILTER OPTIONS

Table to attributes

Global to local

Dashboard

Info | Save | Options | Concept List | Local Max Matcher | Metadata Table

Concept Detailview

Name

Type

Data Type

Constraint

Table Source

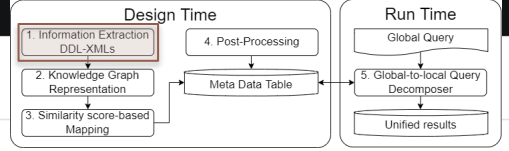
Schema Source

Upload new schema

XML DDL 1 | XML DDL2 | Global

<WHOLESCHEMA> <TABLE xmlns="http://xmlns.oracle.com/ku" version="1.0"> <SCHEMA>LEO</SCHEMA>
<NAME>DB1_ORDERINGDB_CUSTOMERS</NAME> <RELATIONAL_TABLE> <COL_LIST> <COL_LIST_ITEM> <NAME>DB1_CUSTOMER_ID</NAME>
<DATATYPE>NUMBER</DATATYPE> <SCALE>0</SCALE> </COL_LIST_ITEM> <COL_LIST_ITEM> <NAME>DB1_CUSTOMER_NAME</NAME>
<DATATYPE>VARCHAR2</DATATYPE> <LENGTH>255</LENGTH> <COLLATE_NAME>USING_NLS_COMP</COLLATE_NAME> </COL_LIST_ITEM>
<COL_LIST_ITEM> <NAME>DB1_CELLNUMBER</NAME> <DATATYPE>NUMBER</DATATYPE> <PRECISION>10</PRECISION> <SCALE>0</SCALE>
</COL_LIST_ITEM> <COL_LIST_ITEM> <NAME>DB1_EMAIL_ID</NAME> <DATATYPE>VARCHAR2</DATATYPE> <LENGTH>50</LENGTH>
<COLLATE_NAME>USING_NLS_COMP</COLLATE_NAME> </COL_LIST_ITEM> <COL_LIST_ITEM> <NAME>DB1_SHIPPING_ADDRESS</NAME>
<DATATYPE>VARCHAR2</DATATYPE> <LENGTH>255</LENGTH> <COLLATE_NAME>USING_NLS_COMP</COLLATE_NAME> </COL_LIST_ITEM>
</COL_LIST> <CHECK_CONSTRAINT_LIST> <CHECK_CONSTRAINT_LIST_ITEM> <CONDITION>db1_email_id like </CONDITION>
</CHECK_CONSTRAINT_LIST_ITEM> <CHECK_CONSTRAINT_LIST_ITEM> <CONDITION>LENGTH(db1_cellNumber)=10</CONDITION>
</CHECK_CONSTRAINT_LIST_ITEM> </CHECK_CONSTRAINT_LIST> <PRIMARY_KEY_CONSTRAINT_LIST> <PRIMARY_KEY_CONSTRAINT_LIST_ITEM>
<COL_LIST> <COL_LIST_ITEM> <NAME>DB1_CUSTOMER_ID</NAME> </COL_LIST_ITEM> </COL_LIST> <USING_INDEX> <INDEX_ATTRIBUTES>
<PCTFREE>10</PCTFREE> <INITRANS>2</INITRANS> <MAXTRANS>255</MAXTRANS> <STORAGE> <INITIAL>65536</INITIAL>
<NEXT>1048576</NEXT> <MINEXTENTS>1</MINEXTENTS> <MAXEXTENTS>2147483645</MAXEXTENTS> <PCTINCREASE>0</PCTINCREASE>
<FREELISTS>1</FREELISTS> <FREELIST_GROUPS>1</FREELIST_GROUPS> <BUFFER_POOL>DEFAULT</BUFFER_POOL>
<FLASH_CACHE>DEFAULT</FLASH_CACHE> <CELL_FLASH_CACHE>DEFAULT</CELL_FLASH_CACHE> </STORAGE>
<TABLESPACE>USERS</TABLESPACE> <LOGGING>Y</LOGGING> </INDEX_ATTRIBUTES> </USING_INDEX>
</PRIMARY_KEY_CONSTRAINT_LIST_ITEM> </PRIMARY_KEY_CONSTRAINT_LIST>
<DEFAULT_COLLATION>USING_NLS_COMP</DEFAULT_COLLATION> <PHYSICAL_PROPERTIES> <HEAP_TABLE> <SEGMENT_ATTRIBUTES>
<SEGMENT_CREATION_IMMEDIATE></SEGMENT_CREATION_IMMEDIATE> <PCTFREE>10</PCTFREE> <PCTUSED>40</PCTUSED>

Close | Import DDL into Concept List



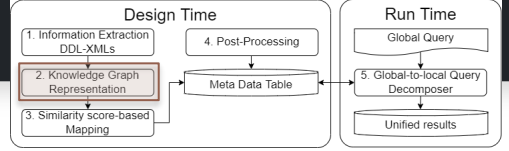
Concept List

Id	Concept Name	Manipulated Name	Synonyms	Parent #	Type	Schema	Data Type	Constraint	Constraint ref #	Global #
CONCEPT_1	global				schema					
CONCEPT_2	Customers	Customers Client		CONCEPT_1	table	global				
CONCEPT_3	Customer_Id	Customer Id						PRIMARY KEY		
CONCEPT_4	Customer_Name	Customer Name		CONCEPT_2	attribute	global	VARCHAR2			
CONCEPT_5	Customer_Phone_number	Customer Phone nu		CONCEPT_2	attribute	global	VARCHAR2			
CONCEPT_6	Customer_Mail_Id	Customer Mail Id		CONCEPT_2	attribute	global	VARCHAR2			
CONCEPT_7	Customer_Address	Customer Address		CONCEPT_2	attribute	global	VARCHAR2			
CONCEPT_8	Customer_Title	Customer Title		CONCEPT_2	attribute	global	VARCHAR2			
CONCEPT_9	Customer_Credit_Card	Customer Credit Car		CONCEPT_2	attribute	global	VARCHAR2			
CONCEPT_10	Customer_CVC	Customer CVC		CONCEPT_2	attribute	global	VARCHAR2			
CONCEPT_11	Products	Products		CONCEPT_1	table	global				
CONCEPT_12	Product_Id	Product Id		CONCEPT_11	attribute	global	NUMBER	PRIMARY KEY		
CONCEPT_13	Product_Name	Product Name		CONCEPT_11	attribute	global	VARCHAR2			
CONCEPT_14	Product_Category	Product Category		CONCEPT_11	attribute	global	VARCHAR2			
CONCEPT_15	Product_Description	Product Description		CONCEPT_11	attribute	global	VARCHAR2			
CONCEPT_16	Product_Price	Product Price		CONCEPT_11	attribute	global	NUMBER			
CONCEPT_17	Product_Quantity	Product Quantity		CONCEPT_11	attribute	global	NUMBER			
CONCEPT_18	Product_Quantity_Per_Unit	Product Quantity Pe		CONCEPT_11	attribute	global	NUMBER			
CONCEPT_19	Product_Sold	Product Sold		CONCEPT_11	attribute	global	NUMBER			

CUSTOMERS CLIENT,CLIENT,CUSTOMERS,CUSTOMER,GUEST,NODE,CLIENT

Inteplato

Sign out



Dashboard

Info | Save | Options | Concept List | Local Max Matcher | Metadata Table

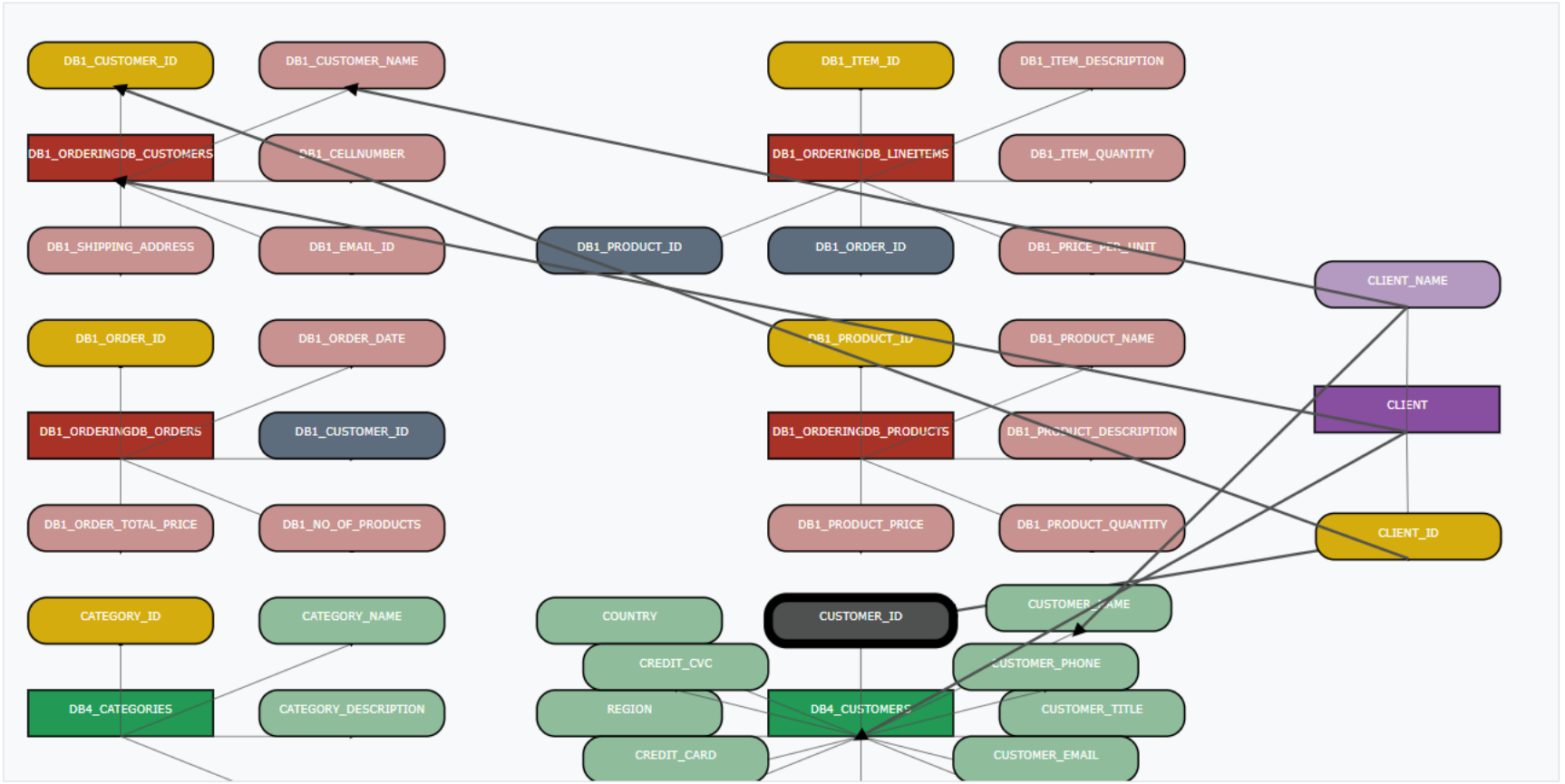
SCHEMA UNIVERSE OPTIONS
- Upload new schema
- Delete schema

GLOBAL SCHEMA OPTIONS
- New global table
- New global attribute

CONCEPT FILTER OPTIONS
- Tables
- Attributes

PATH FILTER OPTIONS
- Table to attributes
- Global to local

Schema Universe



Concept Detailview

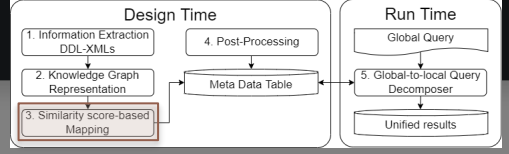
Name	CLIENT_ID
Type	attribute
Data Type	NUMBER
Constraint	PRIMARY KE
Table Source	CLIENT
Schema Source	global

Global-to-local relationships

Local Schema 1

DB1_CUSTOMER_ID

Local Schema 2



Inteplato

Sign out

SCHEMA UNIVERSE OPTIONS

Upload new schema

Delete schema

GLOBAL SCHEMA OPTIONS

New global table

New global attribute

CONCEPT FILTER OPTIONS

Tables

Attributes

PATH FILTER OPTIONS

Table to attributes

Global to local

Dashboard

Info | Save | Options | Concept List | Local Max Matcher | Metadata Table

Local concept finder

Mapping suggestions for global **table CUSTOMER (CONCEPT_121)**:
Retrieved Synonyms (2) for manipulated name **CUSTOMER** are **CLIENT,CUSTOMER**

id	name	manipulated_name	schema	synonyms	fuzzy search	#datatypes	tot score table	tot score attribute	tot score	set
CONCEPT_42	DB1_ORDERINGDB_CUSTOMERS	CUSTOMERS	SCHEMA1	1/2	0.999	2/2	2.499	1	3.499	
CONCEPT_73	DB4_CUSTOMERS	CUSTOMERS	SCHEMA2	1/2	0.999	2/2	2.499	1	3.499	
CONCEPT_61	DB1_ORDERINGDB_PRODUCTS	PRODUCTS	SCHEMA1	0	0	2/2	1	0.7219	1.7219	
CONCEPT_48	DB1_ORDERINGDB_LINEITEMS	LINEITEMS	SCHEMA1	0	0	2/2	1	0.6105	1.6105	
CONCEPT_108	DB4_PRODUCTS	PRODUCTS	SCHEMA2	0	0	2/2	1	0.2493	1.2493	
CONCEPT_86	DB4_LINE_ITEMS	LINE ITEMS	SCHEMA2	0	0	2/2	1	0.2376	1.2376	
CONCEPT_94	DB4_ORDERS	ORDERS	SCHEMA2	0	0	2/2	1	0.2161	1.2161	
CONCEPT_116	DB4_SHIPPERS	SHIPPERS	SCHEMA2	0	0	2/2	1	0.2055	1.2055	
CONCEPT_68	DB4_CATEGORIES	CATEGORIES	SCHEMA2	0	0	2/2	1	0	1	
CONCEPT_55	DB1_ORDERINGDB_ORDERS	ORDERS	SCHEMA1	0	0	1/2	0.5	0	0.5	

Close

Concept Detailview

Name	CUSTOMER
Type	table
Data Type	
Constraint	
Table Source	
Schema Source	global

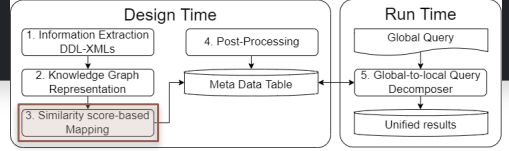
Global-to-local relationships

Local Schema 1

Local Schema 2

Inteplato

Sign out



SCHEMA UNIVERSE OPTIONS

- Upload new schema
- Delete schema

GLOBAL SCHEMA OPTIONS

- New global table
- New global attribute

CONCEPT FILTER OPTIONS

- Tables
- Attributes

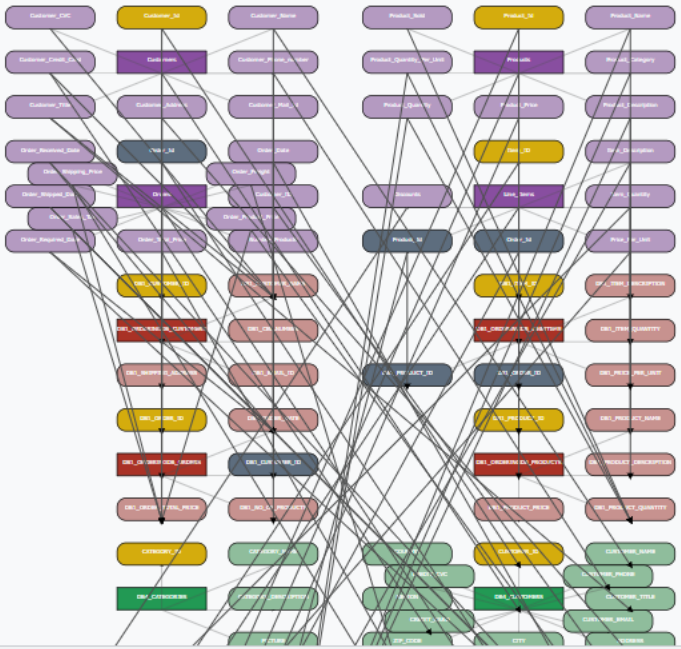
PATH FILTER OPTIONS

- Table to attributes
- Global to local

Dashboard

Info | Save | Options | Concept List | Local Max Matcher | Metadata Table

Schema Universe



Concept Detailview

Name	Customers
Type	table
Data Type	
Constraint	
Table Source	
Schema Source	global

Global-to-local relationships

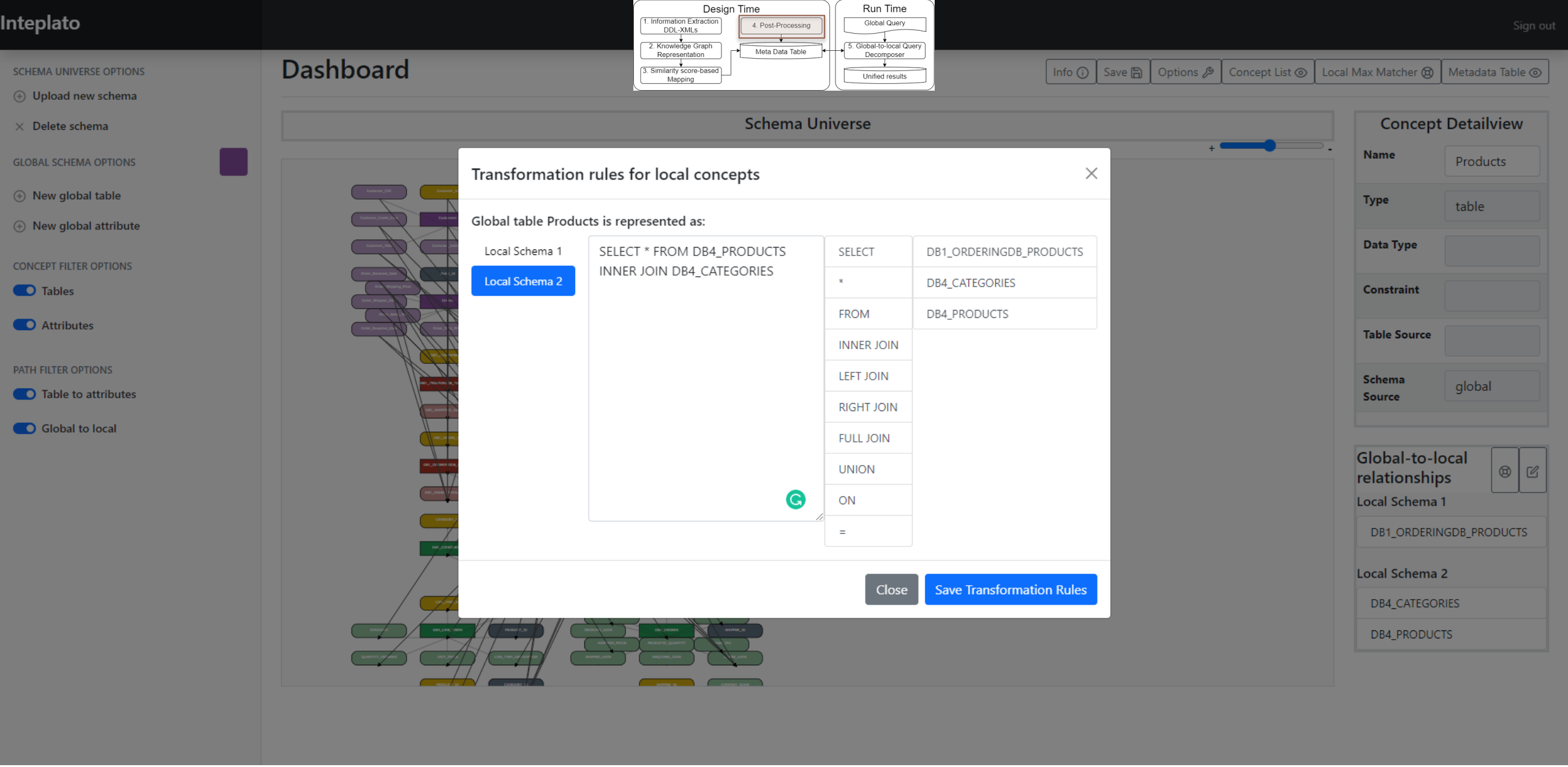
Local Schema 1

Local Schema 2

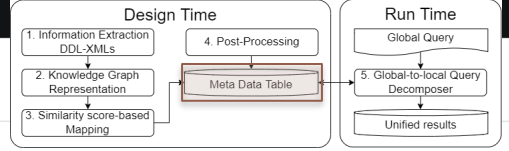
Elements | Console | Recorder | Sources | Network | Performance | Memory | Application | Security | Lighthouse | ResourcesSaver

top | Filter | Default levels | 1 Issue: 1

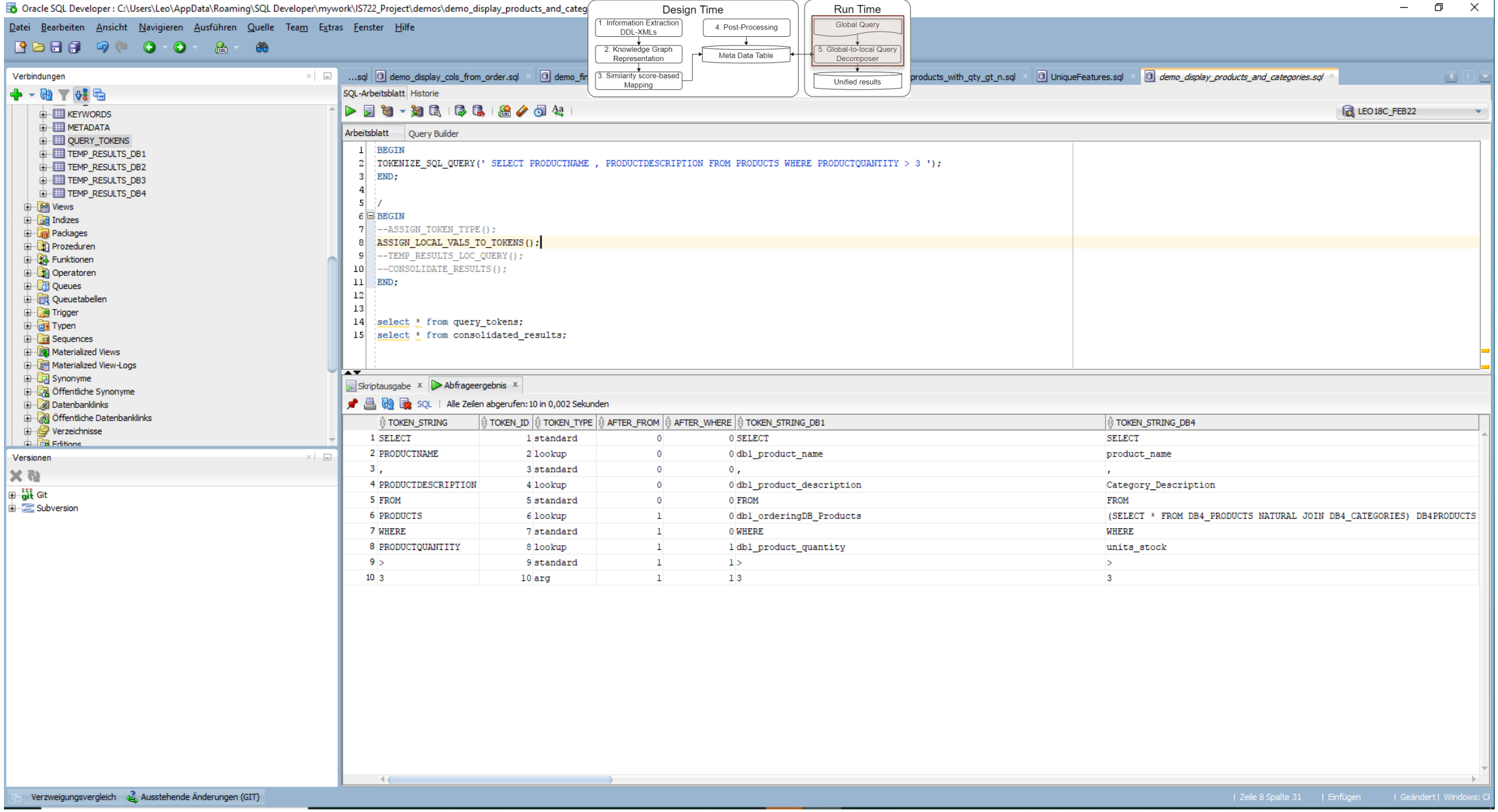
```
CONCEPT_14 has similarly scored local concepts. Human support required for SCHEMA1   index.html:747
CONCEPT_16 has similarly scored local concepts. Human support required for SCHEMA1   index.html:747
CONCEPT_19 has similarly scored local concepts. Human support required for SCHEMA2   index.html:747
CONCEPT_24 has similarly scored local concepts. Human support required for SCHEMA2   index.html:747
CONCEPT_29 has similarly scored local concepts. Human support required for SCHEMA1   index.html:747
CONCEPT_40 has similarly scored local concepts. Human support required for SCHEMA1   index.html:747
```

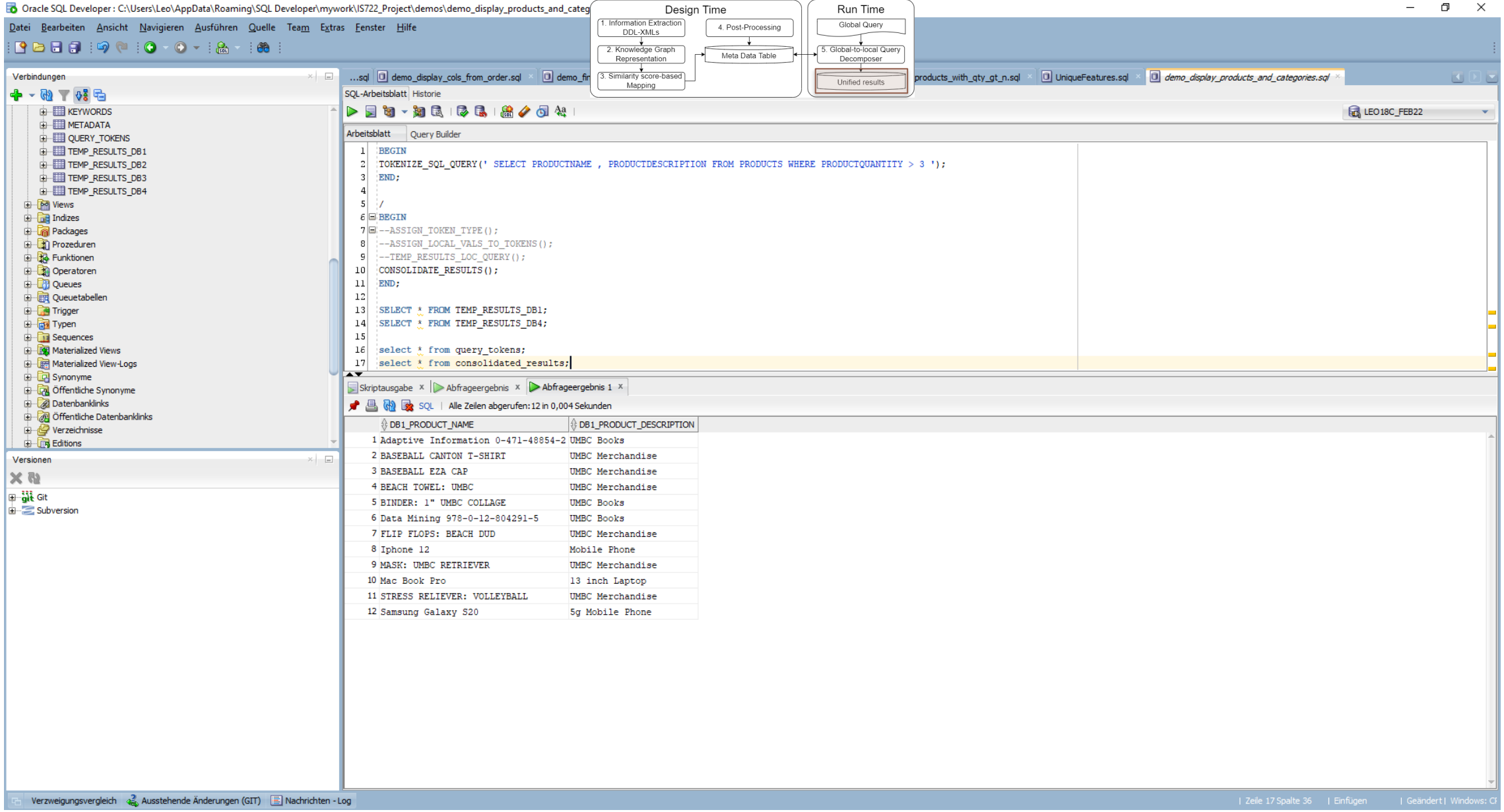

Inteplato
Sign out
SCHEMA UNIVERSE OPTIONS
Upload new schema
Delete schema
GLOBAL SCHEMA OPTIONS
New global table
New global attribute
CONCEPT FILTER OPTIONS
Tables
Attributes
PATH FILTER OPTIONS
Table to attributes
Global to local
Dashboard
Design Time
1. Information Extraction DDL-XMLs
2. Knowledge Graph Representation
3. Similarity score-based Mapping
4. Post-Processing
Meta Data Table
Run Time
Global Query
5. Global-to-local Query Decomposer
Unified results
Info
Save
Options
Concept List
Local Max Matcher
Metadata Table
Schema Universe
Transformation rules for local concepts
Global table Products is represented as:
Local Schema 1
Local Schema 2
SELECT * FROM DB4_PRODUCTS
INNER JOIN DB4_CATEGORIES
SELECT
DB1_ORDERINGDB_PRODUCTS
*
DB4_CATEGORIES
FROM
DB4_PRODUCTS
INNER JOIN
LEFT JOIN
RIGHT JOIN
FULL JOIN
UNION
ON
=
Close
Save Transformation Rules
Concept Detailview
Name
Products
Type
table
Data Type
Constraint
Table Source
Schema Source
global
Global-to-local relationships
Local Schema 1
DB1_ORDERINGDB_PRODUCTS
Local Schema 2
DB4_CATEGORIES
DB4_PRODUCTS

Metadata Table



#	Global Concept Name	Transformation rule for schema 1	Transformation rule for schema 2
CONCEPT_2	Customers	DB1_ORDERINGDB_CUSTOMERS	DB4_CUSTOMERS
CONCEPT_3	Customer_Id	DB1_CUSTOMER_ID	CUSTOMER_ID
CONCEPT_4	Customer_Name	DB1_CUSTOMER_NAME	CUSTOMER_NAME
CONCEPT_5	Customer_Phone_number	DB1_CUSTOMER_NAME	CUSTOMER_PHONE
CONCEPT_6	Customer_Mail_Id	DB1_EMAIL_ID || '@googlemail.com'	CUSTOMER_EMAIL
CONCEPT_7	Customer_Address	DB1_SHIPPING_ADDRESS	ADDRESS
CONCEPT_8	Customer_Title	DB1_CUSTOMER_NAME	CUSTOMER_TITLE
CONCEPT_9	Customer_Credit_Card	DB1_CUSTOMER_NAME	CREDIT_CARD
CONCEPT_10	Customer_CVC	DB1_CUSTOMER_NAME	
CONCEPT_11	Products	DB1_ORDERINGDB_PRODUCTS	SELECT * FROM DB4_PRODUCTS INNER JOIN DB4_CATEGORIES
CONCEPT_12	Product_Id	DB1_PRODUCT_ID	PRODUCT_ID
CONCEPT_13	Product_Name	DB1_PRODUCT_NAME	PRODUCT_NAME
CONCEPT_14	Product_Category		PRODUCT_NAME
CONCEPT_15	Product_Description	DB1_PRODUCT_DESCRIPTION	PRODUCT_NAME
CONCEPT_16	Product_Price		UNIT_PRICE
CONCEPT_17	Product_Quantity	DB1_PRODUCT_QUANTITY	QUANTITYPERUNIT
CONCEPT_18	Product_Quantity_Per_Unit	DB1_PRODUCT_QUANTITY	QUANTITYPERUNIT
CONCEPT_19	Product_Sold	DB1_PRODUCT_QUANTITY	
CONCEPT_20	Orders	DB1_ORDERINGDB_ORDERS	DB4_ORDERS
CONCEPT_21	Order_Id	DB1_ORDER_ID	ORDER_ID
CONCEPT_22	Order_Date	DB1_ORDER_DATE	ORDER_DATE
CONCEPT_23	Customer_ID	DB1_CUSTOMER_ID	CUSTOMER_ID
CONCEPT_24	Number_Products	DB1_NO_OF_PRODUCTS	
CONCEPT_25	Order_Total_Price	DB1_ORDER_TOTAL_PRICE	TOTAL_PRICE
CONCEPT_26	Order_Required_Date	DB1_ORDER_DATE	ORDER_DATE
CONCEPT_27	Order_Shipped_Date	DB1_ORDER_DATE	ORDER_DATE
CONCEPT_28	Order_Received_Date	DB1_ORDER_DATE	ORDER_DATE


Design Time
1. Information Extraction DDL-XMLs
2. Knowledge Graph Representation
3. Similarity score-based Mapping
4. Post-Processing
Meta Data Table
Run Time
Global Query
5. Global-to-local Query Decomposer
Unified results
Oracle SQL Developer : C:\Users\Leo\AppData\Roaming\SQL Developer\mywork\IS722_Project\demos\demo_display_products_and_categ
Datei Bearbeiten Ansicht Navigieren Ausführen Quelle Team Extras Fenster Hilfe
Verbindungen
KEYWORDS
METADATA
QUERY_TOKENS
TEMP_RESULTS_DB1
TEMP_RESULTS_DB2
TEMP_RESULTS_DB3
TEMP_RESULTS_DB4
Views
Indizes
Packages
Prozeduren
Funktionen
Operatoren
Queues
Queuetabellen
Trigger
Typen
Sequences
Materialized Views
Materialized View-Logs
Synonyme
Öffentliche Synonyme
Datenbanklinks
Öffentliche Datenbanklinks
Verzeichnisse
Editions
Versionen
Git
Subversion
demo_display_cols_from_order.sql
products_with_qty_gt_n.sql
UniqueFeatures.sql
demo_display_products_and_categories.sql
SQL-Arbeitsblatt
Historie
LEO18C_FEB22
Arbeitsblatt
Query Builder
BEGIN
TOKENIZE_SQL_QUERY(' SELECT PRODUCTNAME , PRODUCTDESCRIPTION FROM PRODUCTS WHERE PRODUCTQUANTITY > 3 ');
END;
/
BEGIN
--ASSIGN_TOKEN_TYPE();
ASSIGN_LOCAL_VALS_TO_TOKENS();
--TEMP_RESULTS_LOC_QUERY();
--CONSOLIDATE_RESULTS();
END;
select * from query_tokens;
select * from consolidated_results;
Skriptausgabe
Abfrageergebnis
SQL | Alle Zeilen abgerufen: 10 in 0,002 Sekunden
TOKEN_STRING TOKEN_ID TOKEN_TYPE AFTER_FROM AFTER_WHERE TOKEN_STRING_DB1 TOKEN_STRING_DB4
1 SELECT 1 standard 0 0 SELECT SELECT
2 PRODUCTNAME 2 lookup 0 0 db1_product_name product_name
3 , 3 standard 0 0 , ,
4 PRODUCTDESCRIPTION 4 lookup 0 0 db1_product_description Category_Description
5 FROM 5 standard 0 0 FROM FROM
6 PRODUCTS 6 lookup 1 0 db1_orderingDB_Products (SELECT * FROM DB4_PRODUCTS NATURAL JOIN DB4_CATEGORIES) DB4PRODUCTS
7 WHERE 7 standard 1 0 WHERE WHERE
8 PRODUCTQUANTITY 8 lookup 1 1 db1_product_quantity units_stock
9 > 9 standard 1 1 > >
10 3 10 arg 1 1 3 3
Verzweigungsvergleich
Ausstehende Änderungen (GIT)
Zeile 8 Spalte 31
Einfügen
Geändert
Windows: CF


Design Time
1. Information Extraction DDL-XMLs
2. Knowledge Graph Representation
3. Similarity score-based Mapping
4. Post-Processing
Meta Data Table
Run Time
Global Query
5. Global-to-local Query Decomposer
Unified results
Oracle SQL Developer : C:\Users\Leo\AppData\Roaming\SQL Developer\mywork\IS722_Project\demos\demo_display_products_and_categ
Datei Bearbeiten Ansicht Navigieren Ausführen Quelle Team Extras Fenster Hilfe
Verbindungen
KEYWORDS
METADATA
QUERY_TOKENS
TEMP_RESULTS_DB1
TEMP_RESULTS_DB2
TEMP_RESULTS_DB3
TEMP_RESULTS_DB4
Views
Indizes
Packages
Prozeduren
Funktionen
Operatoren
Queues
Queuetabellen
Trigger
Typen
Sequences
Materialized Views
Materialized View-Logs
Synonyme
Öffentliche Synonyme
Datenbanklinks
Öffentliche Datenbanklinks
Verzeichnisse
Editions
Versionen
Git
Subversion
demo_display_cols_from_order.sql
products_with_qty_gt_n.sql
UniqueFeatures.sql
demo_display_products_and_categories.sql
SQL-Arbeitsblatt
Historie
LEO18C_FEB22
Arbeitsblatt
Query Builder
BEGIN
TOKENIZE_SQL_QUERY(' SELECT PRODUCTNAME , PRODUCTDESCRIPTION FROM PRODUCTS WHERE PRODUCTQUANTITY > 3 ');
END;
/
BEGIN
--ASSIGN_TOKEN_TYPE();
--ASSIGN_LOCAL_VALS_TO_TOKENS();
--TEMP_RESULTS_LOC_QUERY();
CONSOLIDATE_RESULTS();
END;
SELECT * FROM TEMP_RESULTS_DB1;
SELECT * FROM TEMP_RESULTS_DB4;
select * from query_tokens;
select * from consolidated_results;
Skriptausgabe
Abfrageergebnis
Abfrageergebnis 1
SQL | Alle Zeilen abgerufen: 12 in 0,004 Sekunden
DB1_PRODUCT_NAME
DB1_PRODUCT_DESCRIPTION
1 Adaptive Information 0-471-48854-2 UMBC Books
2 BASEBALL CANTON T-SHIRT UMBC Merchandise
3 BASEBALL EZA CAP UMBC Merchandise
4 BEACH TOWEL: UMBC UMBC Merchandise
5 BINDER: 1" UMBC COLLAGE UMBC Books
6 Data Mining 978-0-12-804291-5 UMBC Books
7 FLIP FLOPS: BEACH DUD UMBC Merchandise
8 Iphone 12 Mobile Phone
9 MASK: UMBC RETRIEVER UMBC Merchandise
10 Mac Book Pro 13 inch Laptop
11 STRESS RELIEVER: VOLLEYBALL UMBC Merchandise
12 Samsung Galaxy S20 5g Mobile Phone
Verzweigungsvergleich
Ausstehende Änderungen (GIT)
Nachrichten - Log
Zeile 17 Spalte 36
Einfügen
Geändert
Windows: C

Performance Evaluation

Goal: **mapping** a **global schema** with **two local schemas**

Evaluation**: human experts** versus **Inteplato's similarities and objective function**

Data: **two independently** created **logistics** and **single mediator schemas**

Inteplato Configuration	Local schema(1) *	Local schema(2)	Average Accuracy
No synonyms retrieval and distance parameter(0.95)	17/25	17/39	55.80%
No synonyms retrieval and distance parameter(0.975)	17/25	17/39	55.80%
No synonyms retrieval and distance parameter(1)	20/25	21/39	66.92%
Synonyms retrieval and distance parameter(0.95)	21/25	23/39	71.49%
Synonyms retrieval and distance parameter(0.975)	22/25	26/39	77.33%
Synonyms retrieval and distance parameter(0.99)	23/25	28/39	**81.90%**
Synonyms retrieval and distance parameter(1)	23/25	28/39	**81.90%**

Limitations

- Global schema is given
- Global schema is context-dependent
- Experimental data is self-made
  → benchmarks

* Non-existent global concept correspondences (14 out of 39) were cleaned before evaluation.

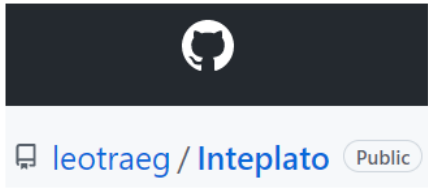
Conclusion

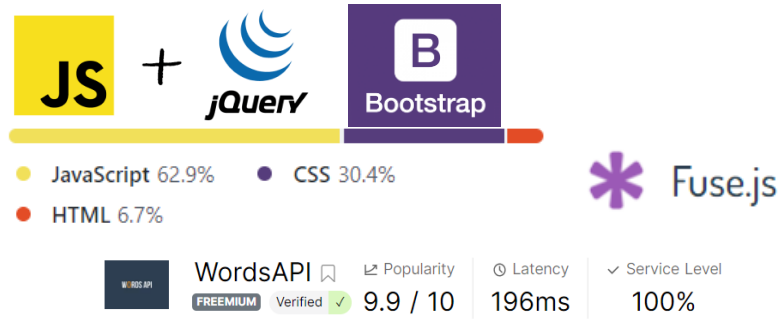
Mastering the process of schema mapping with algorithmic support and traceability for data experts

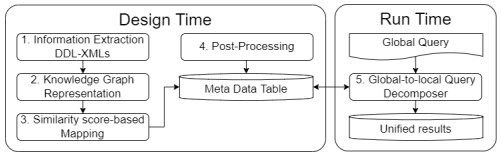
- Scientific directions

Inteplato

- Web-based schema mapping application
- Human-centric and scalable framework
- Technical documentation
- Performance evaluation







#	Similarity function	Concept Global Attribute	Concept Global Table	Range per Concept Local
1	Fuzzy Search	X	X	0..1
2	Synonyms Retrieval & Intersection	X	X	0..1
3	Datatype Similarity	X		0..1
4	Constraint Similarity	X		0..1
5	Attribute Score	$\bowtie \{1,2,3,4\}$		$\in \mathbb{R}+$
6	Attribute Table Score	mapping to 9		$\in \mathbb{R}+$
7	Total Attribute Score	$\bowtie \{5,6\}$		$\in \mathbb{R}+$
8	Datatypes Intersection		X	0..1
9	Table Score		$\bowtie \{1,2,8\}$	$\in \mathbb{R}+$
10	Table Attribute Score		sum max mapping to 5	schema normalized to 0..1
11	Total Table Score		$\bowtie \{9,10\}$	$\in \mathbb{R}+$

$$\text{local max matcher(distance)} = \begin{cases} \text{Table}_x.map(max(\text{total table score}(\text{Table}_y) \in \text{Schema}_j) \\ \frac{second\,max(\text{total table score}(\text{Table}_y) \in \text{Schema}_j)}{max(\text{total table score}(\text{Table}_y) \in \text{Schema}_j)} < distance \\ \text{Attribute}_y.map(max(\text{total attribute score}(\text{Attribute}_y) \in \text{Schema}_j) \\ \frac{second\,max(\text{total attribute score}(\text{Attribute}_y) \in \text{Schema}_j)}{max(\text{total attribute score}(\text{Table}_y) \in \text{Schema}_j)} < distance \\ \text{alert biased mapping attempt} \end{cases} \quad \text{otherwise}$$

$$\text{where } 0 < distance < 1$$

$$\forall\, x \in \text{Table}_{Schema_{global}}$$

$$\forall\, y \in \text{Attribute}_{Schema_{global}}$$

$$\forall\, j \in \text{Schema}_{local}$$

Limitations and Future Work

- Word-embeddings
- Rule-based conflict resolution
- Data preparation
- Field intersections

- Competitions
- Bridge run & design time
- Plug-and-play system

Thank you for your attention!

Q&A